Klinikum rechts der Isar Technische Universität München

Aus der Klinik und Poliklinik für Neurologie
des Klinikums rechts der Isar der
Technischen Universität München

Direktor: Prof. Dr. Bernhard Hemmer

Einfluss genetischer Faktoren auf Störungen des Zentralnervensystems

Influence of genetic factors on disorders of the central nervous system

Zusammenstellung wissenschaftlicher Veröffentlichungen

zur Erlangung der Lehrbefähigung für das Fach

Molekulare Medizin

an der Fakultät für Medizin der Technischen Universität München

vorgelegt von

Dr. rer. nat. Till F. M. Andlauer

Mai 2021

Mit Genehmigung der Fakultät für Medizin
der Technischen Universität München

Fachmentorat

Prof. Dr. Bernhard Hemmer (Vorsitzender)

Prof. Dr. Bertram Müller-Myhsok

PD Dr. Alexander Hapfelmeier

Dekan: Prof. Dr. Bernhard Hemmer

Tag des Kolloquiums:
3. Dezember 2019

Nec laudibus nec timore

Clemens August Graf von Galen

Table of Contents

Introduction

Polygenic disorders

Common, multifactorial disorders like multiple sclerosis (MS), major depressive disorder (MDD), and bipolar disorder (BD) are polygenic. In such polygenic disorders, hundreds to thousands of genetic factors contribute to disease susceptibility, with each genetic variant exerting only a very small effect on disorder risk. The same variants may, in concert with further, independent genetic factors, also shape the clinical presentation, severity, and course of the disorders. These genetic effects occur in addition to and in interaction with environmental risk factors that also modulate disease liability and expression (Figure 1). Genome-wide association studies (GWAS) have identified robustly associated genetic variants for most common disorders during the last fifteen years (Andlauer et al., 2018). Many of these variants influence several disorders or traits – especially in the case of psychiatric disorders (Brainstorm et al., 2018; PGC-CDG et al., 2019).

For some polygenic traits, several thousand independently associated genetic variants have already been described (Yengo et al., 2018). Accordingly, it was recently proposed to extend the polygenic inheritance framework to an omnigenic model for disorders like schizophrenia (Boyle et al., 2017).

Genetic variants associated with highly polygenic disorders like MDD show maximum odds ratios (ORs) below 1.05 (Wray et al., 2018). Nevertheless, the effect sizes of some risk loci may significantly surpass the polygenic background consisting of small effects. For example, the well-established MS risk allele *HLA-DRB1*15:01* shows an odds ratio (OR) of 2.9, while the ORs of most other MS risk variants are below 1.3 (Andlauer et al., 2016; IMSGC, 2019a).

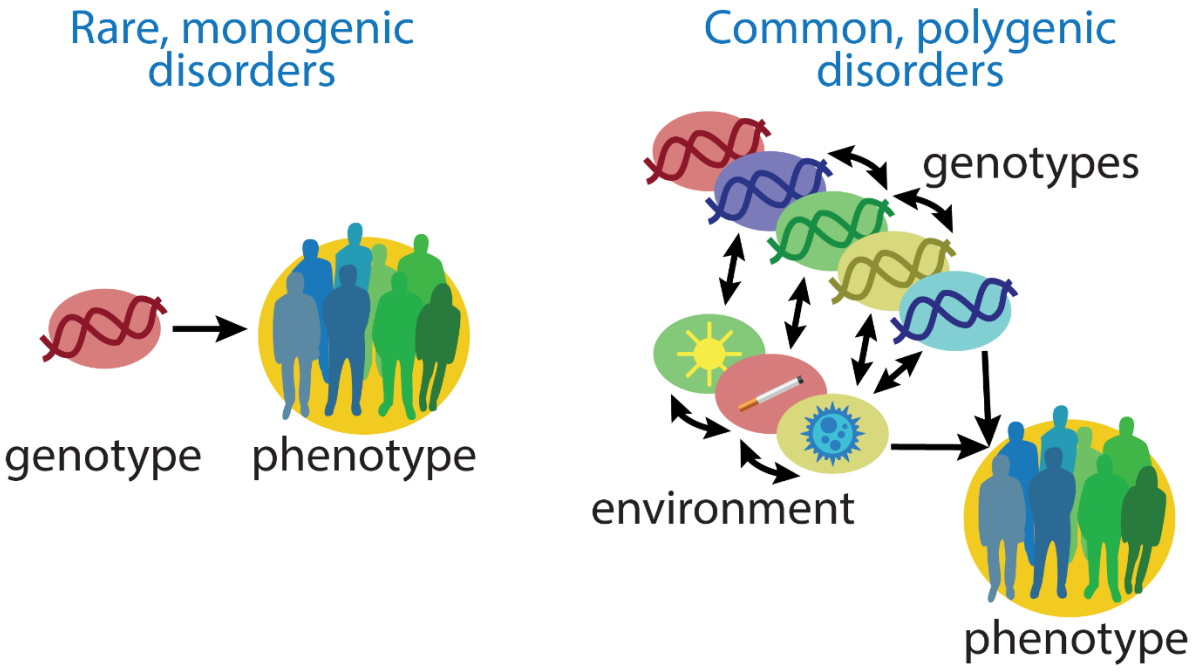


Figure 1: Monogenic *vs.* polygenic disorders.
In rare monogenic disorders, a single mutation can causally trigger the illness. In frequently occurring polygenic disorders, many genetic variants contribute to disorder risk both in an additive manner and in interaction with each other. These disorders are multifactorial: environmental risk and protective factors modify disorder risk both in an additive manner and also in interaction with each other – as well as in interaction with genetic risk factors. Modified after (Andlauer and Nöthen, 2020).

Genome-wide association studies of selected neuropsychiatric disorders

Multiple sclerosis

MS is an autoimmune disease affecting the central nervous system (CNS). The prevalence in Germany is estimated to be 0.3%, with a female-male ratio of 2.5 (Daltrozzo et al., 2018; Holstiege et al., 2017). The first genetic variant robustly associated with MS, later specified as the allele *HLA-DRB1*15:01*, was already identified in 1973, long before the advent of GWAS (Hillert, 1994; Jersild et al., 1973). Human leukocyte antigen (HLA) alleles, mapping to the major histocompatibility complex (MHC) region on chromosome 6, constitute the strongest MS-associated variants in the genome (Patsopoulos et al., 2013). Possibly, HLA alleles like *HLA-DRB1*15:01* may confer risk to MS by increasing the probability of autoreactive, off-target T cell responses. These T cells might recognize both viral peptides and, with low specificity, myelin-derived peptides (Dendrou et al., 2018).

GWASs analyze the association of common single nucleotide polymorphisms (SNPs), also called single nucleotide variants (SNVs), showing minor allele frequencies (MAF) of 1% or higher, with a disease or trait of interest (Andlauer et al., 2018) (Figure 2). Low frequency and rare variants are typically analyzed using specialized methods with more statistical power (Dankowski et al., 2015; IMSGC et al., 2018; Maaser et al., 2018).

Since 2007, MS case/control GWAS conducted by consortia like the International MS Genetics Consortium have identified more than 230 disease-associated variants (Andlauer et al., 2016; IMSGC, 2019a; Patsopoulos, 2018; Sawcer et al., 2014) (Figure 3A).

These variants map to genes that regulate pathways connected to various immune cells, including microglia, *i.e.*, immune cells of the brain (IMSGC, 2019a). An especially prominent pathway affected by MS risk genes is T helper cell differentiation (IMSGC et al., 2011). Taken together, the known MS-associated common variants currently explain about 40% of the SNP-based heritability of the disorder (IMSGC, 2019a).

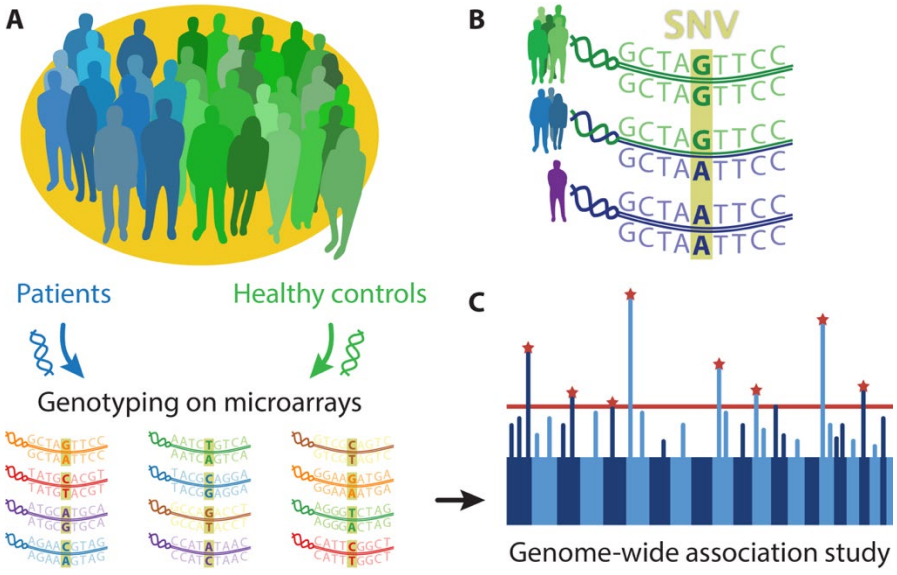


Figure 2: Genome-wide association studies.

A: In order to conduct a case/control genome-wide association study (GWAS), patients and healthy controls are genotyped on microarrays. **B:** Patients and controls may differ in the minor allele frequency (MAF) of single nucleotide variants (SNV). **C:** For a GWAS, all available SNVs across the genome are typically analyzed using logistic or linear regression models. SNVs are considered as significantly associated if their *p*-values are below the genome-wide significance threshold (*i.e.*, a *p*-value $<5\times10^{-8}$; red line).

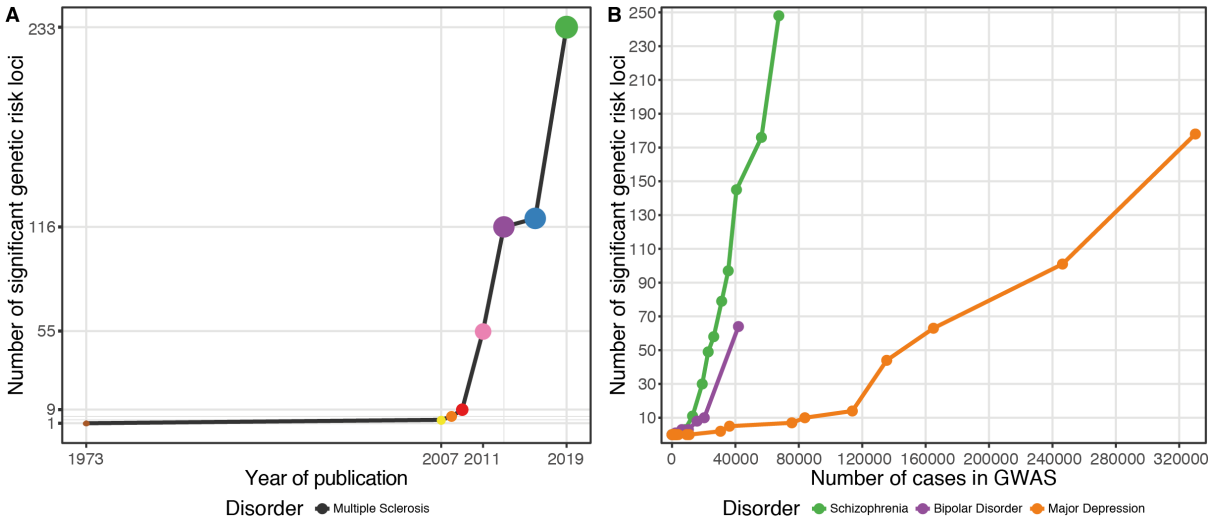


Figure 3: Development of significant GWAS loci.
A: The number of significant genetic risk loci relative to the year of publication for MS. Colors are arbitrary. The dot sizes are proportional to the number of significant risk loci per study. **B:** The number of significant genetic risk loci relative to the number of cases for schizophrenia, BD, and MDD. Modified after (Kendall et al., 2021).

Major depressive disorder

Major depressive disorder (MDD) is a highly prevalent mood disorder negatively affecting both the quality and length of life. As the first published GWAS result, Sullivan and colleagues identified the gene *PCLO* as potentially associated with MDD (Sullivan et al., 2009). This gene, subsequently confirmed as an MDD risk factor (Wray et al., 2018), codes for the presynaptic active zone component Piccolo. We have characterized Piccolo in mice as a regulator of presynaptic F-actin assembly (Waites et al., 2011).

With the help of private genotyping companies and biobanks, the Psychiatrics Genomics Consortium was able to identify 102 genome-wide significant variants (Howard et al., 2019; Kendall et al., 2021; Wray et al., 2018) (Figure 3B). Analyses of data from the Million Veteran Program subsequently increased this list to 178 MDD-associated loci (Levey et al., 2020). MDD-associated genes play roles in synaptic formation and function, neuronal differentiation, and inflammation (Wray et al., 2018).

Although depression is more frequently observed among MS patients than in the general population (Feinstein et al., 2014; Heitmann et al., 2020), these two disorders are genetically not correlated (Brainstorm et al., 2018). Also Mendelian randomization analyses did not support a causal relationship between both disorders (Harroud et al., 2021).

More than it does for other disorders, the phenotype definition matters when studying MDD. Variants identified in GWAS using less strict MDD phenotype definitions, *i.e.*, not requiring a structured clinical interview following the Diagnostic and Statistical Manual of Mental Disorders, are not necessarily specific to MDD. Instead, they might constitute risk loci shared by several psychiatric disorders (Cai et al., 2020).

Schizophrenia and bipolar disorder

GWAS on schizophrenia (SCZ) identified 270 associated genetic loci (PGC-SCZ, 2014; PGC-SCZ et al., 2020). These are more independent associations than reported for MDD, although the SCZ GWAS analyzed smaller samples (Figure 3B). MDD GWAS are possibly more challenging to conduct because depression shows, compared to SCZ, a higher prevalence and a lower heritability (Levinson et al., 2014). MDD patients used for the GWAS might also have been more heterogeneous than the SCZ cases were.

For bipolar disorder (BD), 64 risk loci have been described to date (Mullins et al., 2020;

Stahl et al., 2019), at sample sizes comparable to SCZ GWAS (Figure 3B). Interestingly, many psychiatric disorders, especially SCZ and BD, are genetically correlated with each other (Brainstorm et al., 2018; PGC-CDG et al., 2019). This means that the same genetic variants increase the risk for several disorders at the same time.

Many of the genes influencing various psychiatric disorders stress the importance of pre- and, especially, postsynaptic plasticity in the etiology of these disorders. These findings highlight the benefit of studying the function of pre- and postsynaptic proteins in different model systems, as we have done repeatedly (Andlauer et al., 2014; Beck et al., 2015; Faber et al., 2020; Fulterer et al., 2018; Waites et al., 2011). Eventually, such studies in animal models will contribute to our understanding of the biological underpinnings of psychiatric and neurological disorders.

Polygenic (risk) scores

Most of the variants associated with polygenic disorders do not locate to protein-coding exons of genes but, instead, map to intergenic regions. Here, they may influence the transcription-factor binding efficacy to long-range enhancers and, thereby, of gene expression. However, these effects are often subtle, leading to minor differences in gene expression that only exert a noticeable effect over a long time or influence a phenotype in combination with other risk factors. Such small, often indirect effects render quantification of the genetic risk for polygenic disorders challenging.

The polygenic burden of such variants associated with a disorder or trait can be quantified using polygenic scores (PGS) (Andlauer and Nöthen, 2020). PGS were developed as an efficient tool to reliably assess the cumulative, additive risk contributed by genetic variants across the genome. PGS typically use effect sizes and *p*-values from published GWAS as training data to calculate the polygenic burden in independent test datasets. When calculating PGS, the number of disease-associated alleles per variant is multiplied by the GWAS effect size (Figure 4A). Subsequently, the weighted allele counts of all variants are summed to compute the final PGS.

The selection and weighting of variants can be improved using linkage disequilibrium (LD) information, as, e.g., implemented in the Bayesian method PRS-CS (Ge et al., 2019). The prediction performance of PGS calculated using PRS-CS is typically higher than that of traditional PGS (Ni et al., 2020).

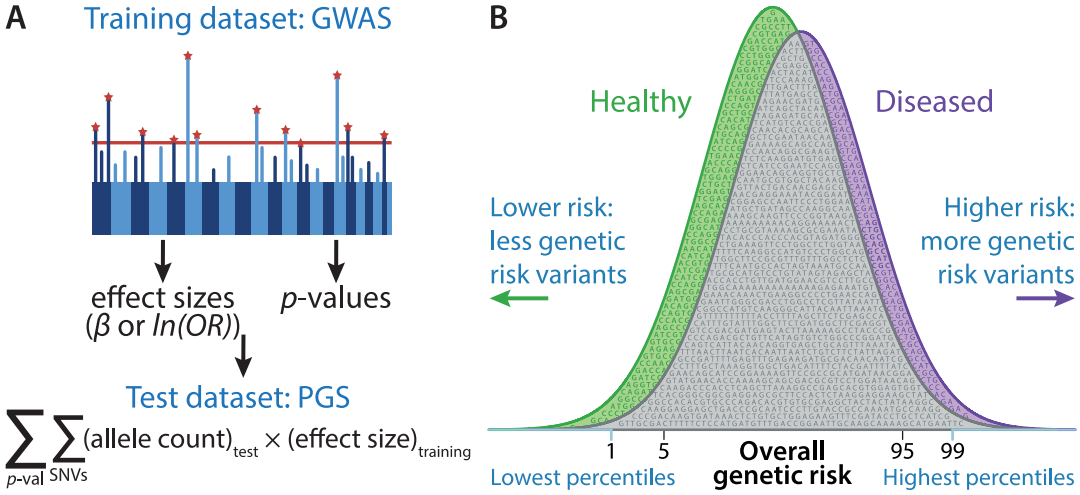


Figure 4: Calculation of polygenic (risk) scores

A: For PGS, weighted allele counts of uncorrelated SNVs are summarized. Variants are weighted by β coefficients or the natural logarithms of odds ratios. **B:** PGS follow a normal distribution on a population level. Only PGS at the extremes of the distribution differ significantly. Modified after (Andlauer and Nöthen, 2020).

Presentation of selected studies

Discovery of Novel Multiple Sclerosis Susceptibility Loci

Andlauer TFM, Buck D, …, Hemmer B, Müller-Myhsok B (2016):
Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. **Science Advances** 2(6) e1501678

Further studies relevant for this section:

Arloth J, Eraslan G, **Andlauer TFM**, …, Theis FJ, Binder EB, Mueller NS (2020): *DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning.* **PLOS Computational Biology** doi:10.1371/journal.pcbi.1007616

Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, Beecham AH, James T, Replogle J, Vlachos IS, McCabe C, Pers TH, Brandes A, White C, Keenan B, Cimpean M, Winn P, Panteliadis IP, Robbins A, **Andlauer TFM**, …, De Jager PL (2019): *Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility.* **Science** 365(6460):eaav7188

Madireddy L, Patsopoulos NA, Cotsapas C, Bos SD, Beecham A, McCauley J, Kim K, Jia X, Santaniello A, Caillier SJ, **Andlauer TFM**, …, De Jager P, Sawcer S, Oksenberg JR, Baranzini SE (2019): *A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis.* **Nature Communications** 10:2236

Mitrovič M, Patsopoulos NA, Beecham AH, Dankowski D, Goris A, …, **Andlauer TFM**, …, De Jager PL, Kockum I, Hafler DA, Cotsapas C (2018): *Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk.* **Cell** 175(6):1679-87

Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, James T, Ewing E, Lindén M, Górnikiewicz B, Aeinehband S, Stridh P, Link J, **Andlauer TFM**, …, Jagodic M (2018): *DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis.* **Nature Communications 9**(1):2397

Dankowski T, Buck D, **Andlauer TFM**, …, Hemmer B, Ziegler A (2015): *Successful Replication of GWAS Hits for Multiple Sclerosis in 10,000 Germans Using the Exome Array.* **Genetic Epidemiology** 39(8):601-8

Nischwitz S, Wolf C, **Andlauer TFM**, …, Mueller-Myhsok B, Weber F (2015): *MS susceptibility is not affected by single nucleotide polymorphisms in the MMP9 gene.* **Journal of Neuroimmunology** 279:46-9

Most major GWAS on MS have been conducted by the International MS Genetics Consortium (IMSGC). To increase the sample size, the IMSGC combines samples from diverse European and Northern American ancestral backgrounds. Between two major IMSGC GWAS, published in 2013 and 2019 (IMSGC, 2019a; IMSGC et al., 2013), the German Competence Network Multiple Sclerosis (KKNMS) conducted a GWAS on German MS patients and healthy controls (Andlauer et al., 2016). Here, the aim was to examine the genetic architecture of MS in a more homogenous population. With a total of 4,888 cases and 10,395 controls, this study constitutes the largest GWAS conducted for MS in a single population to date.

Patients recruited from German MS centers and diagnosed with relapsing-remitting, secondary progressive, or primary progressive MS, or with clinically isolated syndrome (CIS; often a precursor of relapsing-remitting MS) were genotyped on Illumina HumanOmniExpress-24 BeadChips to form the dataset DE1. In addition, MS patients previously genotyped on Illumina 660W-Quad BeadChips were used for a second dataset, DE2. The MS/CIS patients were combined with healthy controls from several German population-based or depression cohorts (KORA, HNR, SHIP, DOGS, FoCus, popgen, MARS, and GSK). After imputation, analyses of HLA alleles and GWAS were conducted in both datasets separately, followed by fixed-effects meta-analyses.

As expected, a variant within the MHC region, in LD with *HLA-DRB1*15:01*, showed the strongest association in the meta-analysis (rs3104373: odds ratio (OR) 2.90, 95% confidence interval (CI) 2.72-3.09, $p=1.3\times10^{-234}$, Figure 5A). A step-wise conditional logistic regression analysis of imputed HLA alleles identified seven independent HLA alleles to be associated with MS at genome-wide significance (Table 1). The well-established MS risk allele *HLA-DRB1*15:01* reached an OR of 2.85 (CI=2.66-3.06), $p=1.0\times10^{-191}$. The allele *HLA-DRB1*15:01* also reduced the age at

disease onset, yet this association did not reach genome-wide significance (n=1,196, β=-0.21, p=7.6×10^{-6}). A variant in LD with this HLA allele, rs4959027, showed the overall lowest p-value in the age at onset analyses (n=1,519, β=-0.20, p=1.5×10^{-7}, Figure 5B).

Outside the MHC region, fifteen loci reached genome-wide significance, of which five had not been identified in GWAS at genome-wide significance before (Figure 5A). All loci showed support for an association in both samples, DE1 and DE2. Interestingly, all significant or suggestive (p<1×10^{-6}) variants from the largest IMSGC GWAS available at the time of publication (IMSGC et al., 2013) and present in our dataset showed the same direction of effect. We calculated the probability of such an overlap occurring by coincidence using a binomial sign test and observed a p=5×10^{-32}. The five genome-wide significant loci were also analyzed in an independent Sardinian cohort (2,903 cases, 3,323 controls), where two of the variants (rs2812197 (*DLEU1*) and rs4925166 (*SHMT1*)) replicated.

Among the novel variants, rs4925166 on chromosome 17 showed the strongest association (OR=0.85, CI=0.81-0.90, p=2.7×10^{-9}). To fine-map this signal, we conducted *cis* expression quantitative trait locus (eQTL) analyses on n=242 DE1 patients, for whom whole-blood gene expression had been assessed using Illumina HT-12 BeadChips. We found that rs4925166 was part of an eQTL for the gene *SHMT1* (false discovery rate (FDR) 2.99×10^{-10}). Using other, non-MS datasets for which DNA methylation had been quantified using the Illumina HumanMethylation450 BeadChip, we identified that rs4925166 also affected DNA methylation levels at three CpG sites mapping to *SHMT1,* with CpG cg26763362 exhibiting the most robust support for an association. A causal mediation analysis of the association rs4925166 → CpG cg26763362 → *SHMT1 expression* revealed partial mediation of the effect of rs4925166 on *SHMT1* expression by DNA methylation (Figure 5C).

SHMT1 is a serine hydroxymethyltransferase acting in the folate cycle. It catalyzes the transfer of a carbon unit subsequently used for the synthesis of both nucleotides and methionine. SHMT1 thus plays a role in the metabolism of the methyl group donor substrate S-adenosylmethionine (SAM). Thereby, it maintains methylation homeostasis in the cell, shifting epigenetic regulation via methylation into the focus of MS susceptibility.

The putatively MS-associated genes at all five loci that reached, in our GWAS, genome-wide significance for the first time (*L3MBTL3*, *DLEU1*, *MAZ*, *ERG*, and *SHMT1*) are associated with regulatory mechanisms in immune cells.

HLA allele	**AF**	**OR (95 % CI)**	***p*-value**	***HLA* alleles in LD (r^2 > 0.9)**
*DRB1*15:01*	14.8	2.85 (2.66-3.06)	1.03×10^{-191}	*DQB1*06:02*
*A*02:01*	28.6	0.68 (0.64-0.73)	3.68×10^{-29}	
*B*38:01*	2.0	0.36 (0.27-0.49)	2.09×10^{-11}	
*DRB1*13:03*	1.5	1.96 (1.60-2.40)	6.42×10^{-11}	
*DPB1*03:01*	10.3	1.33 (1.22-1.46)	4.35×10^{-10}	
*DRB1*03:01*	12.2	1.29 (1.18-1.40)	1.85×10^{-08}	*DQA1*05:01, DQB1*02:01*
*DRB1*08:01*	3.0	1.63 (1.39-1.91)	2.36×10^{-09}	*DQA1*04:01, DQB1*04:02*

Table 1: Genome-wide significant HLA alleles.

The analysis was conducted using step-wise logistic regression. For each row, alleles from the rows above were used as covariates. The first allele was associated with MS independently of the others. Results are from fixed-effects meta-analyses of DE1 and DE2. AF, allele frequency of controls in %; OR, odds ratio; CI, confidence interval; LD, linkage disequilibrium.

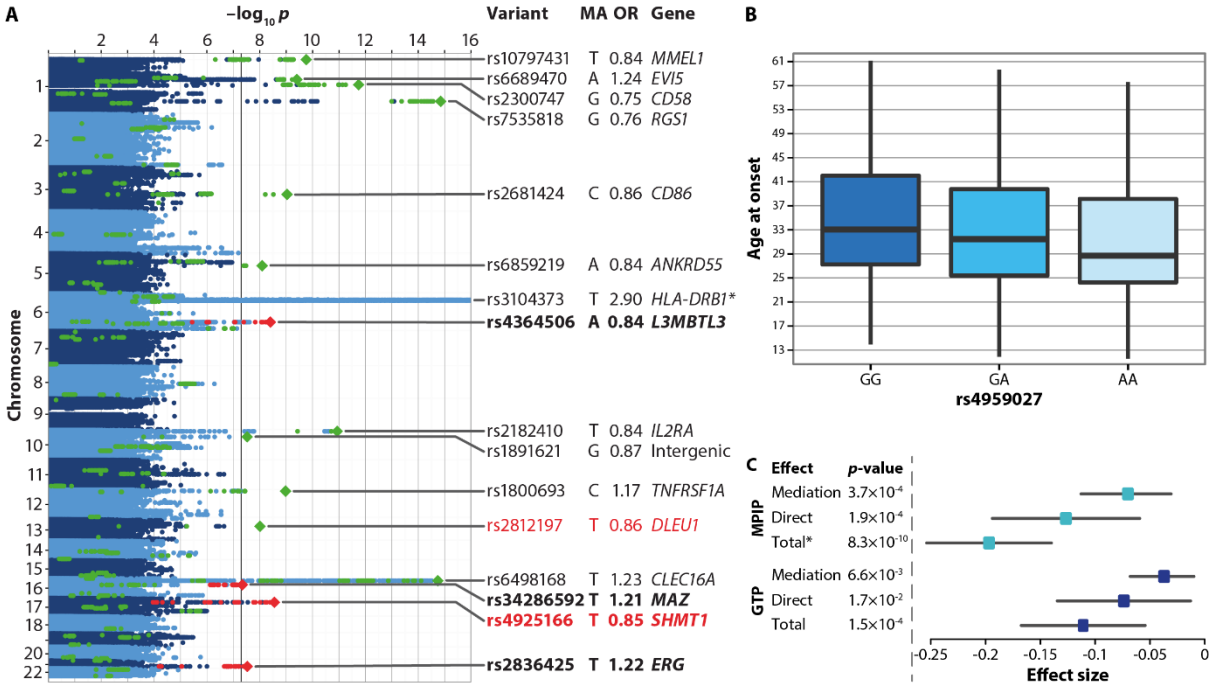


Figure 5: Results from the German MS GWAS.

A: Manhattan plot showing the strength of evidence for an association with MS in the German GWAS. Green dots represent established MS-associated variants and their proxies, as listed by Sawcer *et al.* (Sawcer et al., 2014). Top variants at the 15 non-MHC loci associated at the genome-wide significance threshold in our study are shown as diamonds. Novel variants showing genome-wide significance are plotted as red diamonds; their names are shown in bold font. Variants in high linkage disequilibrium ($r^2 \geq 0.7$) with these novel variants are shown as red dots. Variants replicated in the Sardinian cohort are shown in red font. MA, minor allele, OR, odds ratio (relative to the MA). The plot is truncated at $-\log_{10}(p) = 16$ for better visibility. The p-value of rs3104373 is 1.3×10^{-234}.
B: The genotype of SNP rs4959027 *vs.* the untransformed age at onset, two outliers were removed for better visibility (n=1,519). **C:** Causal mediation analysis results. Tested mediation effect: rs4925166 → CpG cg26763362 → *SHMT1* expression. Direct effect: rs4925166 → *SHMT1* expression.

Indirect evidence suggests that they could all be linked either directly or indirectly to epigenetic regulatory mechanisms. For example, the locus coding for the regulatory, long non-coding RNA *DLEU1* regulates the expression of NF-κB (Garding et al., 2013), a critical MS-associated transcription factor playing a central role in the regulation of inflammation (Housley et al., 2015). In turn, *DLEU1* is strictly regulated by DNA methylation at its promoter region (Garding et al., 2013).

Three variants at the *DLEU1* locus on chromosome 13 had previously shown suggestive associations (rs806321, rs9596270, and rs806349) (IMSGC et al., 2011, 2013; Lill et al., 2015; Patsopoulos et al., 2011). In our GWAS, a fourth SNP, rs2812197, was the top-associated signal at the locus. The associations of both rs806349 and rs806321 were dependent on rs2812197 in conditional analyses. In addition, a secondary signal might exist at the locus, represented by the variant rs9591325 (in LD with the previously published rs9596270). SNP rs9591325 maps to a functional region, indicating that this variant could either be the actual causal or a second causal variant at the *DLEU1* locus.

With the identification of several MS risk genes potentially affecting or affected by epigenetic regulation, the importance of epigenetic mechanisms for developing this multifactorial disorder is underlined. Importantly, environmental risk factors like smoking, viral infections, childhood obesity, or low sun exposure can influence the expression of disease-associated genes via epigenetic mechanisms (Waubant et al., 2019; Zhou et al., 2014). Our assessment of DNA methylation

being relevant for the development of MS was later confirmed in a study by Kular and colleagues, which showed that DNA methylation at *HLA-DRB1*15:01* is causally involved in the development of MS (Kular et al., 2018). The German KKNMS MS case/control GWAS was followed up in 2019 by a large IMSGC study on, in total, 47,479 cases and 68,374 controls (IMSGC, 2019a). This study identified 32 independent associations in the MHC region, 200 autosomal loci outside the MHC region, and one signal on chromosome X (Figure 6). Using different fine-mapping approaches, 551 potentially MS-associated genes were prioritized. Enrichment of the expression of these MS-specific genes was observed not only in cells of the adaptive immune system, e.g., in $CD4^+$ and $CD8^+$ T and B cells but also in the innate immune system, including natural killer and dendritic cells. For the first time, enrichment was demonstrated in immune cells of the brain, *i.e.*, microglia – but not in neurons or astrocytes (Figure 6). Genome-wide significant effects now explain 39% of the SNP-based MS heritability, and the suggestive effects potentially add another 9%.

Notably, these studies focused on common variants with a MAF of ≥1%. An additional share of the heritability can likely be explained by the effects of low-frequency or rare variants (Wainschtein et al., 2019). We analyzed such variants genotyped on Illumina HumanExome BeadChips both in German and international IMSGC samples (Dankowski et al., 2015; IMSGC et al., 2018) and found low-frequency coding variants with a MAF of ≤5% to explain 5% of the heritability. So far, only four associations of low-frequency variants could be established robustly (IMSGC et al., 2018). Reliable single variant-level analyses of rare variants still require larger sample sizes.

A vital challenge in GWAS is mapping variant-level associations to genes and functional pathways in specific cell types and developmental stages. The IMSGC has successfully conducted gene set and pathway analyses to identify biological networks influencing MS susceptibility (IMSGC, 2019b, 2019a; IMSGC et al., 2011). However, traditional functional analyses rely on a *post-hoc*, position-based annotation of variants for possible functional effects. As an innovative alternative, we have developed a novel approach called deepWAS, which analyzes functional units of SNPs in the context of cell type-specific chromatin features (Arloth et al., 2020). This method allows for the direct identification of functionally relevant SNPs together with their regulatory effects. We applied this method to the KKNMS GWAS dataset, describing 61 such regulatory SNP units (Arloth et al., 2020).

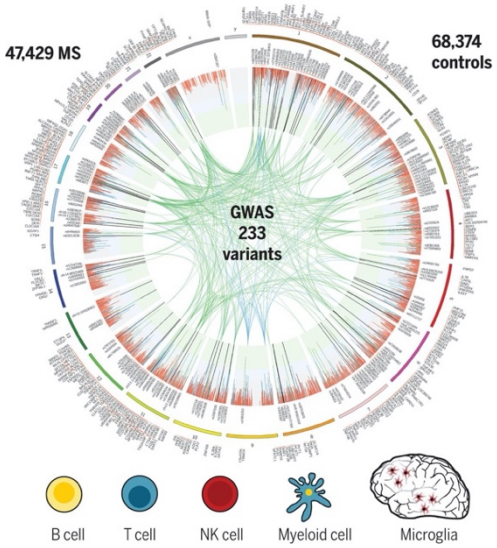


Figure 6: Results from the IMSGC MS GWAS.
The MS genomic map displayed as a circos plot of 201 autosomal non-MHC and X-chromosomal effects, 551 prioritized genes, and potential protein-protein interactions between them (center). Based on enrichment analyses of putative MS susceptibility genes, the study suggests that several peripheral immune cell populations and microglia are implicated in the development of MS.
With permission from The American Association for the Advancement of Science, the image was taken from (IMSGC, 2019a).

Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis is Linked to Human Multiple Sclerosis Risk Genes

Faber H, Kurtoic D, Krishnamoorthy G, Weber P, Pütz P, Müller-Myhsok B, Weber F, **Andlauer TFM** (2020):
Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis Is Linked to Human Multiple Sclerosis Risk Genes.
Frontiers in Immunology 11:2165

Despite the identification of more than 230 MS risk variants, we still do not functionally understand the biological underpinnings of MS. One issue in unraveling the biology of MS is the difficulty of studying the human CNS *in vivo*. Thus, MS research partly relies on animal models like experimental autoimmune encephalomyelitis (EAE). Although widely used, it is unclear to which extent EAE reflects human MS, and all available EAE models are, to some degree, artificial.

Among others, two major EAE models exist: first, MOG EAE, actively induced by injection of a myelin oligodendrocyte glycoprotein (MOG) peptide into *C57BL/6* mice together with Freud's adjuvant. Second, the transgenic model opticospinal EAE (OSE), which develops an autoimmune disease spontaneously. OSE mice express both a T cell receptor (TCR) recognizing a MOG peptide and B cells carrying MOG-specific receptors (Krishnamoorthy et al., 2006).

We analyzed the involvement of MS risk genes in gene expression changes occurring in diseased EAE mice to study to which degree these EAE models reflect the biology of MS. To this end, we compared gene expression profiles of total spinal cord preparations of MOG EAE and OSE mice, assessed using Illumina MouseWG-6 BeadChips. Six mouse types were examined (with four mice each): wildtype (WT); healthy OSE controls (OSE_0); OSE with disease score 1 (OSE_1); OSE score 4 (OSE_4); as a MOG EAE control, healthy control mice injected with complete Freund's adjuvant but not with a MOG peptide (CFA); as MOG EAE, *C57BL/6* wildtype mice injected with adjuvant and a MOG_{35-55} peptide, rated score 4 (MOG_4). We rated the mice for clinical signs of EAE using a 5-point scale. 0: healthy animal; 1: animal with a flaccid tail; 2: animal with impaired righting reflex and/or gait; 3: animal with one paralyzed hind leg; 4: animal with both hind legs paralyzed; 5: moribund animal or death of the animal after preceding clinical disease.

We assessed gene expression differences for five contrasts: OSE_1-OSE_0, OSE_4-OSE_0, MOG_4-CFA, and the two control contrasts CFA-WT and OSE_0-WT. In both MOG EAE and OSE models, gene expression differed between healthy and diseased mice (Figure 7A). More transcripts were differentially expressed in the OSE_4-OSE_0 (n=5,555) than in the MOG_4-CFA contrast (n=3,182). Of these, 4.88× more transcripts were differentially expressed in OSE_4-OSE_0 only. The fold changes observed in OSE_4-OSE_0 were higher than in OSE_1-OSE_0 (binomial test: $p=1.4\times10^{-65}$) or MOG_4-CFA ($p=5.8\times10^{-3}$; Figure 7C-D).

Next, we conducted over-representation analyses (ORA) using WebGestalt v2019 in *R* (Liao et al., 2019), based on the gene ontology (GO) biological process database, for three groups (Table 2): *Common disease transcripts* (CDT), *OSE_4-specific transcripts* (OSE_4sp), and *MOG_4-specific transcripts* (MOG_4sp). In the pathway analyses of genes differentially expressed in the CDT group, 1,379 redundant GO biological processes were significantly overrepresented after correcting for multiple testing. Among the top-associated GO terms were many immune-related gene sets, e.g., *immune response*, *regulation of immune system process*, and *T cell activation*. The same and additional immune-associated processes were also significant in the OSE_4sp group – but no immune system-specific process was significant for MOG_4sp. More immune-related expression changes were thus triggered in OSE than in MOG EAE.

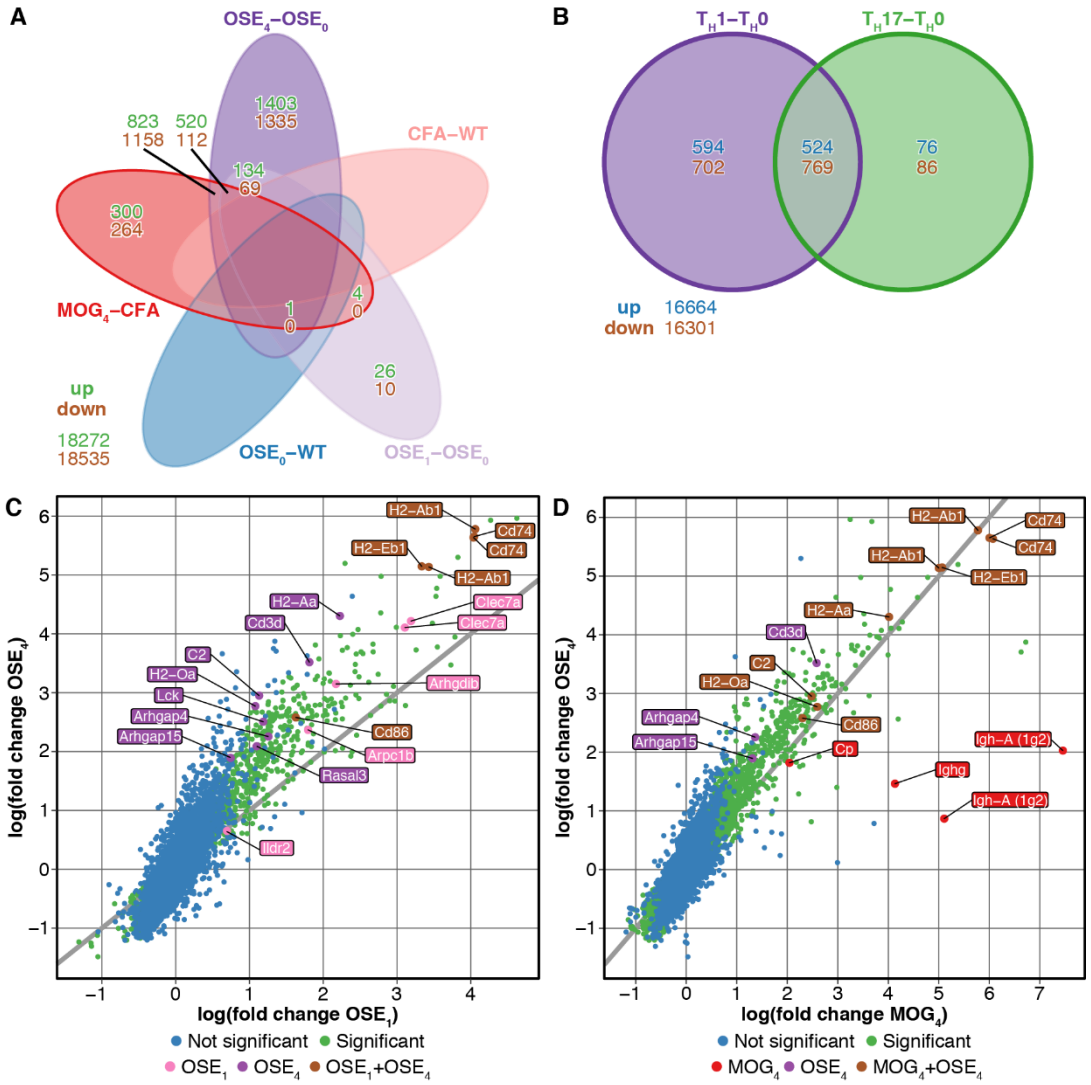


Figure 7: Results from differential expression analyses.
A-B: Venn diagrams of differentially expressed transcripts. **A:** Analyses of MOG EAE and OSE models. **B:** Analyses of T_H cell types. **C-D:** OSE_4 mice showed greater fold expression changes than OSE_1 (**C**) or MOG_4 mice (**D**). The ten most differentially expressed genes per group are labeled (magenta, OSE_4 only; brown, both groups).

Differentially expressed (DE) transcript group	**DE in contrasts**	**Not DE in contrasts**
Common disease transcripts (CDT)	OSE_4-OSE_0, MOG_4-MOG_0	OSE_0-WT, MOG_0-WT
OSE_4-specific transcripts (OSE_4sp)	OSE_4-OSE_0	MOG_4-MOG_0, OSE_0-WT, MOG_0-WT
MOG_4-specific (MOG_4sp) transcripts	MOG_4-MOG_0	OSE_4-OSE_0, OSE_0-WT, MOG_0-WT
OSE_1-specific transcripts (OSE_1sp)	OSE_1-OSE_0	MOG_4-MOG_0, OSE_4-OSE_0, OSE_0-WT, MOG_0-WT
OSE_1-expressed transcripts (OSE_1ex)	OSE_1-OSE_0	OSE_0-WT, MOG_0-WT

Table 2: Groups of differentially expressed (DE) transcripts.
These groups were used in over-representation analyses. WT, wildtype.

We used OSE, which shows variable disease development, to analyze mildly affected mice (OSE_1). Although 34 transcripts were differentially expressed only in OSE_1 animals (*OSE_1-specific transcripts* (OSE_1sp)), no GO terms were significantly overrepresented among these genes. However, for the 805 transcripts differentially expressed in OSE_1-OSE_0 but not in control contrasts (*OSE_1-expressed transcripts* (OSE_1ex); Table 2), the previously mentioned immune-related GO terms, as well as *B cell-mediated immunity* and *antigen processing and presentation*, were significantly overrepresented.

Based on the 200 MS-associated non-MHC autosomal loci, the IMSGC GWAS prioritized 551 MS candidate MS risk genes (IMSGC, 2019a). Of these, 499 transcripts of 265 genes were present in our gene expression dataset. The first component from a principal component analysis (PCA) of these transcripts was significantly higher in all diseased mice than in controls (Figure 8A). Thus, MS risk genes show higher average expression levels in EAE mice, especially in OSE_4, including *H2-Ab1* and *H2-Eb1* (Figure 8B), which are homologous to *HLA-DQB1* and *HLA-DRB5*. Note that the HLA alleles *HLA-DRB1*15:01*, *HLA-DQB1*06:02*, and *HLA-DRB5*01:01* are part of the extended MS risk haplotype *DR15-DQ6* (Hollenbach and Oksenberg, 2015). We analyzed whether our analysis groups (Table 2) were enriched for MS risk genes and found CDT, OSE_4sp, and OSE_1ex to be significantly enriched, but not the MOG_4sp group (Table 3).

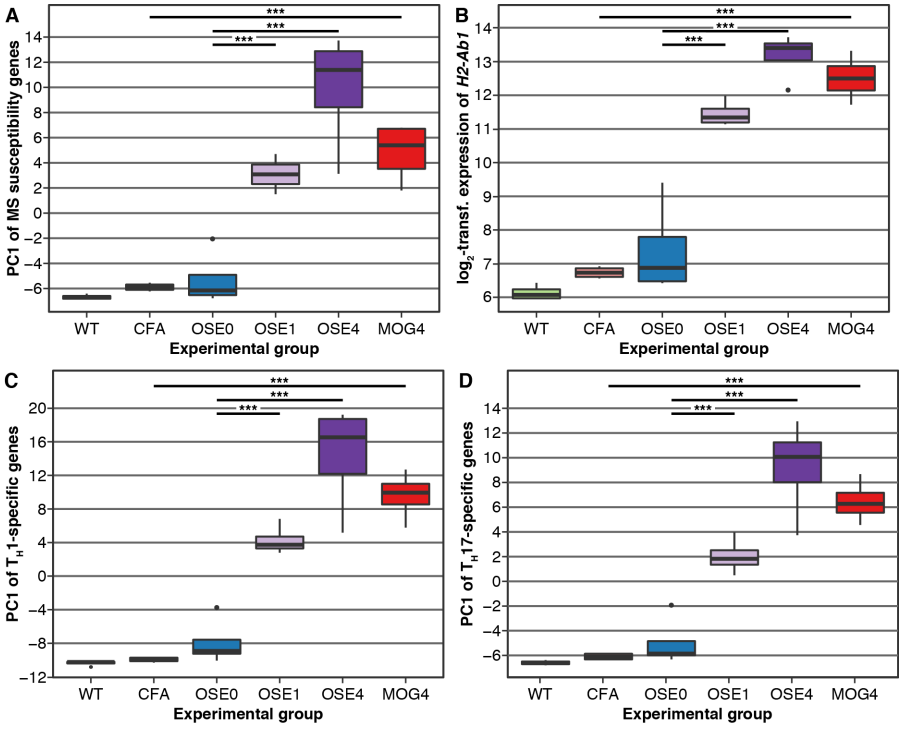


Figure 8: Comparison of gene expression levels.

A: PCA of gene expression profiles of MS risk genes.

B: *H2-Ab1* expression levels.

C-D: PCA of gene expression profiles of transcripts differentially expressed in (C) T_H1 and (D) T_H17 cells.

Significance levels:
* adjusted $p<0.05$,
** adjusted $p<0.01$,
*** adjusted $p<0.001$.

Differentially expressed (DE) transcript group	DE genes	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	2014	68	**$<1\times10^{-5}$**	**$<4\times10^{-5}$**
OSE_4sp	2362	68	**4.4×10^{-4}**	**8.8×10^{-4}**
MOG_4sp	469	11	3.2×10^{-1}	3.2×10^{-1}
OSE_1ex	693	34	**1.0×10^{-5}**	**4.0×10^{-5}**

Table 3: Enrichment of MS susceptibility genes.
We analyzed 265 genes. Enrichments significant after correction for multiple testing are highlighted in bold font.

T helper (T_H) cell differentiation is a crucial pathway in the etiology of MS (Figure 9). T_H1 cells have long been viewed as the main drivers of both EAE and MS (Glatigny and Bettelli, 2018). However, either T_H cell type can induce EAE (Jäger et al., 2009), supporting a critical function also of T_H17 cells in the development of MS. To analyze expression differences of T_H cell-specific transcripts in both EAE models, we generated *in vitro* polarized T_H1 and T_H17 cells from spleen cells of OSE mice. Next, we analyzed differential gene expression in T_H1 and T_H17 cells compared to undifferentiated T_H0 cells. In T_H1 cells, 8× more transcripts were differentially expressed than in T_H17 cells (Figure 7B). Interestingly, a high T_H1/T_H17 ratio is indicative of a lesion distribution pattern characterized by prominent spinal cord involvement, as is the case for both OSE and MOG EAE (Stromnes et al., 2008).

We conducted a PCA on transcripts differentially expressed in T_H1 and T_H17 cells. The first component for both cell types was higher in diseased mice than controls, especially in OSE_4 (Figure 8C-D). The T_H1 signature molecule *Tbx21* (*T-bet*) was significantly upregulated in all diseased mice, while *Ifng* was only upregulated in OSE_4. The T_H17-specific marker *Il17f* was also upregulated in OSE_4, but neither *Rorc* nor *Il17a* were. The analysis groups CDT, OSE_4sp, and OSE_1ex (Table 2) were significantly enriched for genes differentially expressed in either T_H1 or T_H17 cells (Table 4). MOG_4sp transcripts were only significant for T_H1.

Finally, we generated a list of transcripts differentially expressed in both the EAE models and in T_H1 or T_H17 cells. Immune-related biological processes, including the three terms mentioned above as well as *positive regulation of T cell proliferation*, were overrepresented among the genes present in the CDT, OSE_4sp, and OSE_1ex groups and also differentially expressed in T_H1 cells. By contrast, no GO terms at all were overrepresented for MOG_4sp or T_H17-specific genes.

Interestingly, CDT and OSE_1ex genes differentially expressed in T_H1 cells were significantly enriched for IMSGC MS risk genes (Table 5). OSE_4sp genes were enriched for MS risk genes in T_H1 cells at nominal significance (unadjusted p=0.0097). By contrast, T_H1-expressed MOG_4sp transcripts showed no trend for the enrichment of MS risk genes at all (unadjusted p=0.51).

In the context of T_H1-driven immune responses, the spontaneous OSE model might thus be linked to human MS risk genes more closely than induced MOG EAE is. Both EAE types constitute valuable MS models and appear to recapitulate key MS pathways. However, given the enrichment of both MS risk genes and T_H cell-specific transcripts in OSE_4, the OSE model may be more apt than MOG EAE for studying the function of MS risk genes and their associated pathways.

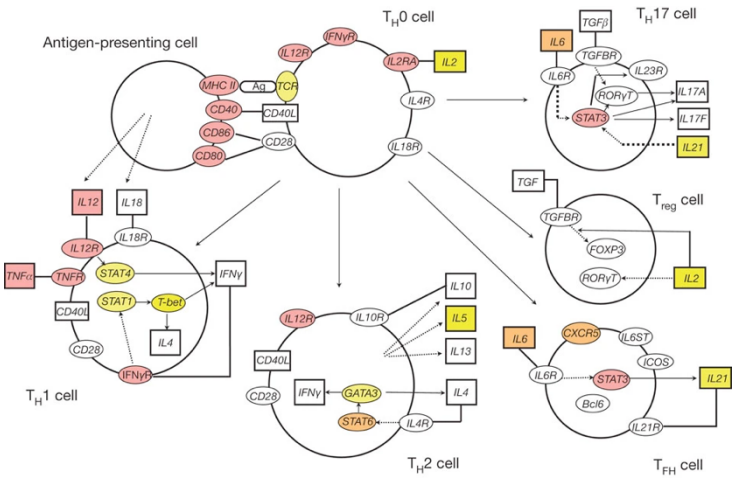


Figure 9: MS risk genes in the T helper cell differentiation pathway.

Colors: red, genome-wide significant; orange, strongly suggestive loci; yellow, suggestive loci.

With permission from Springer Nature, the image was taken from (IMSGC et al., 2011).

Differentially expressed (DE) transcript group	DE genes	Cell type	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	2014	T_H1	150	**$<1\times10^{-5}$**	**$<8\times10^{-5}$**
		T_H17	28	**2.0×10^{-2}**	**4.0×10^{-2}**
OSE_4sp	2362	T_H1	195	**$<1\times10^{-5}$**	**$<8\times10^{-5}$**
		T_H17	36	**2.0×10^{-3}**	**8.0×10^{-3}**
MOG_4sp	469	T_H1	35	**1.1×10^{-2}**	**3.3×10^{-2}**
		T_H17	7	9.8×10^{-2}	9.8×10^{-2}
OSE_1ex	693	T_H1	61	**2.0×10^{-5}**	**1.2×10^{-4}**
		T_H17	16	**1.0×10^{-3}**	**5.0×10^{-3}**

Table 4: Enrichment of TH-specific transcripts.
Analysis of 1,080 T_H1- and 145 T_H17-specific transcripts. Enrichments significant after correction for multiple testing are highlighted in bold font.

Differentially expressed (DE) transcript group	Cell type	EAE T_H cell list size	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	T_H1	150	10	**6.5×10^{-4}**	**5.2×10^{-3}**
	T_H17	30	3	2.1×10^{-2}	1.1×10^{-1}
OSE_4sp	T_H1	215	10	9.7×10^{-3}	5.8×10^{-2}
	T_H17	41	3	4.7×10^{-2}	1.6×10^{-1}
MOG_4sp	T_H1	37	1	5.1×10^{-1}	5.1×10^{-1}
	T_H17	7	1	1.3×10^{-1}	2.6×10^{-1}
OSE_1ex	T_H1	60	6	**1.1×10^{-3}**	**7.7×10^{-3}**
	T_H17	16	2	3.9×10^{-2}	1.6×10^{-1}

Table 5: Enrichment of MS risk genes among T_H-specific transcripts.
Enrichments significant after correction for multiple testing are highlighted in bold font.

As supported by the identification of putative epigenetic risk factors (Andlauer et al., 2016; Kular et al., 2018), genetic and environmental MS risk factors likely interact during MS disease development. EAE models like OSE that spontaneously develop disease might recapitulate such gene-by-environment interactions better than induced models such as MOG EAE can.

For successful treatment of MS without long-term disability, timely initiation of the appropriate therapy is vital. Our analyses of OSE_1 mice demonstrated that many gene expression changes observed in severely affected mice are already present in mildly affected animals. Interestingly, the GO term *response to interferon-beta* was highly overrepresented in both the OSE_1ex and CDT analysis groups. Interferon β is a well-established first-line treatment for MS. Other treatment-associated genes were also differentially expressed in all diseased mice, e.g., *Cd74*, involved in MHC class-II antigen presentation, the target of Milatuzumab and regulated by Fingolimod (Alinari et al., 2011), and *Cd52*, expressed on lymphocytes and the target of Alemtuzumab (Tintore et al., 2019). OSE thus provides the opportunity to study such therapy-associated genes throughout different disease stages, facilitating the evaluation of potential therapies.

Neither EAE model fully reflects a heterogeneous human disease like MS. Nevertheless, our results indicate that OSE, with its closer link to MS risk genes and T_H cell biology, may be better suited for studying the etiology of MS and for defining specific therapeutic targets than MOG-induced EAE is.

Genetic Risk Factors for Anti-Drug Antibodies Against Interferon β

Andlauer TFM, Link J, Martin D, . . ., Deisenhammer F, Fogdell-Hahn A, Hemmer B & on behalf of the ABIRISK consortium (2020): Treatment- and population-specific genetic risk factors for anti-drug antibodies against interferon-beta: a GWAS. **BMC Medicine** 18:298

Further studies relevant for this section:

Buck D*, **Andlauer TFM*** (*shared first*), . . ., Müller-Myhsok B, Hemmer B, The BEYOND and BENEFIT Study Groups (2018): Effect of HLA-DRB1 alleles and genetic variants on the development of neutralizing antibodies to interferon beta in the BEYOND and BENEFIT trials. **Multiple Sclerosis Journal** 25(4):565-573

A recombinant form of the cytokine interferon β (IFNβ) has been the first successful disease-modifying treatment for MS (Lublin, 2005). IFNβ-1b, approved in 1993, is raised in *Escherichia coli* bacteria and injected subcutaneously (*s.c.*). It was initially sold under the brand names Betaseron and Betaferon. Later, the same product was also marketed as Extavia. In 1996, IFNβ-1a was approved, which is produced in Chinese hamster ovary cells. The IFNβ-1a formulation sold under the brand name Avonex is administered via intramuscular (*i.m.*), the later-approved Rebif by *s.c.* injection. While more effective treatments for MS have been developed since, IFNβ remains a widely used first-line treatment for MS, not least because of its well-established long-term safety profile. Recently, IFNβ has gained renewed attention as a potential treatment of severe COVID-19 (Hadjadj et al., 2020; Lee and Shin, 2020; Monk et al., 2020).

All biopharmaceutical treatments, *i.e.*, therapeutically active polypeptides produced via biotechnological methods, are immunogenic. This means that the substances can provoke an unwanted immune response. Over time, the immune system of some patients forms a neutralizing response against the compound, including the development of anti-drug antibodies (ADA). Accordingly, up to 40% of patients treated with IFNβ develop ADA binding IFNβ (bADA) (Deisenhammer, 2014). Binding ADA neutralizing the biological activity of IFNβ, e.g., by inhibiting the interaction of the drug with its receptor (Figure 10), are called neutralizing ADA (nADA) (Kappos et al., 2005). Such antibodies are not only medically relevant for biopharmaceuticals and vaccination. For example, it was suggested that pre-existing neutralizing auto-antibodies against type I interferons may cause severe COVID-19 after SARS-CoV-2 infection (Bastard et al., 2020).

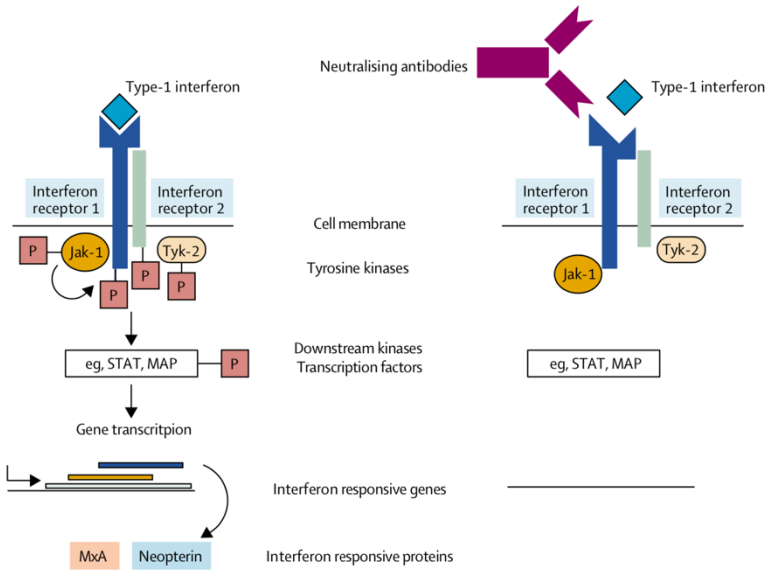


Figure 10: Interference of neutralizing ADA with type-I interferon signaling.
IFNβ signaling leads to the activation of transcription factors and the expression of IFNβ-responsive genes. Neutralizing ADA prevent the formation of IFNβ-receptor complexes and downstream transcriptional activation.
With permission from Elsevier, the image was taken from (Hemmer et al., 2005).

Previous research has reported variation in the immunogenicity of IFNβ preparations (Bertolotto et al., 2004). Besides the species in which they are produced, IFNβ-1a and -1b formulations are characterized by several differences: While the amino acid sequence of IFNβ-1a corresponds to the sequence of the mammalian protein IFNB1, IFNβ-1b differs from it at two positions (Bertolotto et al., 2004). Moreover, IFNβ-1b is not glycosylated, possibly affecting its immunogenicity by promoting the formation of protein aggregates (Bertolotto et al., 2004). Furthermore, post-translational modifications like deamidation, oxidation, and glycation can take place spontaneously. Their occurrence might depend on the manufacturing and processing of biopharmaceuticals (Jawa et al., 2013). Avonex does not contain serum albumin, a major driver of protein aggregation (Barnard et al., 2013). In consequence, also sequence-identical products like Avonex and Rebif may differ in their immunogenicity. Notably, the formulation of Rebif was changed in 2007 to decrease its immunogenicity (Jaber et al., 2007; Stefano et al., 2010).

Different factors modify the risk of developing ADA, including the use of immunosuppressants and antibiotics, infections, tobacco smoking (Hässler et al., 2020), possibly obesity (Callahan et al., 2014), and genetic variants. Several studies from the United States, Germany, Spain, and Sweden have proposed various genetic risk and protective factors for ADA. However, results differed between these studies, and they could not establish a consensus on the human leukocyte antigen (*HLA*) alleles affecting ADA formation (Barbosa et al., 2006; Buck et al., 2011; Hoffmann et al., 2008; Link et al., 2014; Núñez et al., 2014; Stickler et al., 2004; Weber et al., 2012). Only one previous study assessed the effect of genetic variants outside the MHC region (Weber et al., 2012).

We conducted two studies to further assess the genetic risk landscape for IFNβ-induced ADA. First, we published an analysis of 941 patients treated with IFNβ-1b in two phase III trials (Buck et al., 2018), BENEFIT (Kappos et al., 2006) and BEYOND (O'Connor et al., 2009). This longitudinal study focused on examining previously suggested MS-associated variants. After correction for multiple testing, *HLA-DRB1*04:01* and *HLA-DRB1*07:01* were significantly associated with the presence of nADA, mean nADA titer, maximum nADA titer, and the area under the nADA receiver-operating characteristic (ROC) curve (AUC). Moreover, our GWAS identified highly correlated SNPs in the MHC region to be associated with the different nADA endpoints. The previously proposed non-MHC SNP on chromosome 8 (Weber et al., 2012) did not replicate.

Subsequently, we conducted a retrospective cross-sectional study to establish a consensus on the heterogeneous findings from previous publications. All previous studies had comparably small sample sizes, only analyzed a single population, and mostly did not assess ADA with validated methods. We analyzed 2,757 MS patients, treated with three different IFNβ preparations, from the Karolinska Institutet Stockholm, Sweden (KI) and the Technical University of Munich, Germany (TUM). We determined both bADA levels and nADA titers in the same patients, allowing for systematic comparisons between both ADA types (Figure 11). We measured bADA levels by capture ELISA and nADA titers using a validated luciferase-based bioassay.

We randomized the patients into a discovery (n=2,000) and a replication (n=757) dataset. Binding and neutralizing ADA each target other parts of the IFNβ molecule. Nevertheless, we found bADA levels to be correlated with the presence of nADA (Spearman ρ=0.66) and with nADA titers (ρ=0.71). Compared to the presence of nADA determined via screening and titration, estimation of the nADA status from bADA levels had a sensitivity of 0.85 and a specificity of 0.84 (Figure 12). Thus, also patients with low bADA levels may produce nADA, indicating that the prediction of nADA from bADA titers may not be reliable.

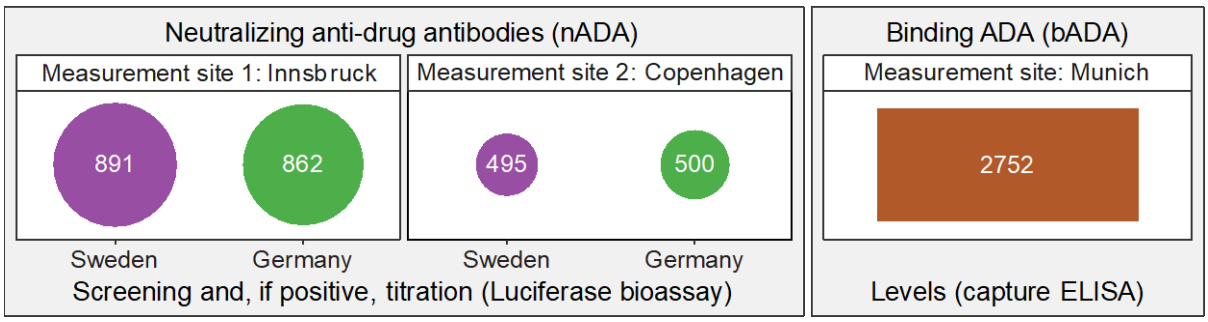


Figure 11: Distribution of samples for the ADA measurements.

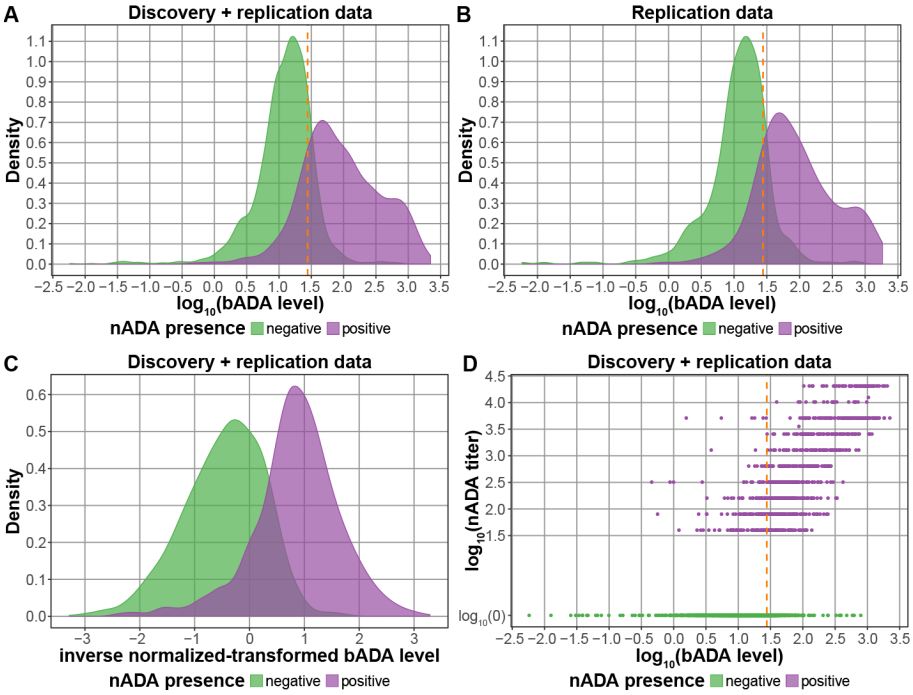


Figure 12: Comparison of bADA levels and nADA titers.
A-C: Density plots of $\log_{10}$ bADA levels stratified by nADA presence.
D: Comparison of bADA levels to nADA titers, colored by nADA presence.
The dashed line shows the cutoff optimizing the maximum sensitivity and specificity in the discovery data.

The SNP-based heritability, calculated using GCTA, was h^2_g=0.47 for bADA levels, h^2_g=0.50 for nADA titers, and h^2_{go}=0.48 on the observed scale for nADA presence (h^2_{gl}=0.79 on a liability scale). This result underlines the importance of genetic factors in the occurrence of IFNβ ADA. Binding ADA levels were genetically correlated with nADA titers (r_g=0.95) and presence (r_g=0.89). These findings indicate that the same genetic risk factors influence both bADA and nADA. It may constitute a stochastic process whether ADA are functionally neutralizing or merely bind IFNβ.

GWAS for nADA presence, nADA titers, and bADA levels identified the same or highly correlated variants mapping to the MHC region as associated with all three ADA measurements. We found no SNPs outside the MHC region to be robustly associated with ADA. None of the SNPs associated at genome-wide significance with ADA measurements in IFNβ-1a *s.c.*-treated patients showed significant statistical support for an association in IFNβ-1b *s.c.*-treated patients and vice versa.

Following the genome-wide GWAS, we analyzed HLA alleles in IFNβ-1a *s.c.*-treated patients. Here, allele *HLA-DQB1*06:02* and the ancestral haplotype *DR15-DQ6 (DRB1*15:01 + DQA1*01:02 + DQB1*06:02*), both smaller subsets of *B7-DQ6 (B*07:02 + DR15-DQ6*), showed the most robust support for conferring risk to all three ADA measurements (Figure 13A, Table 6). All other risk alleles were dependent on this extended haplotype.

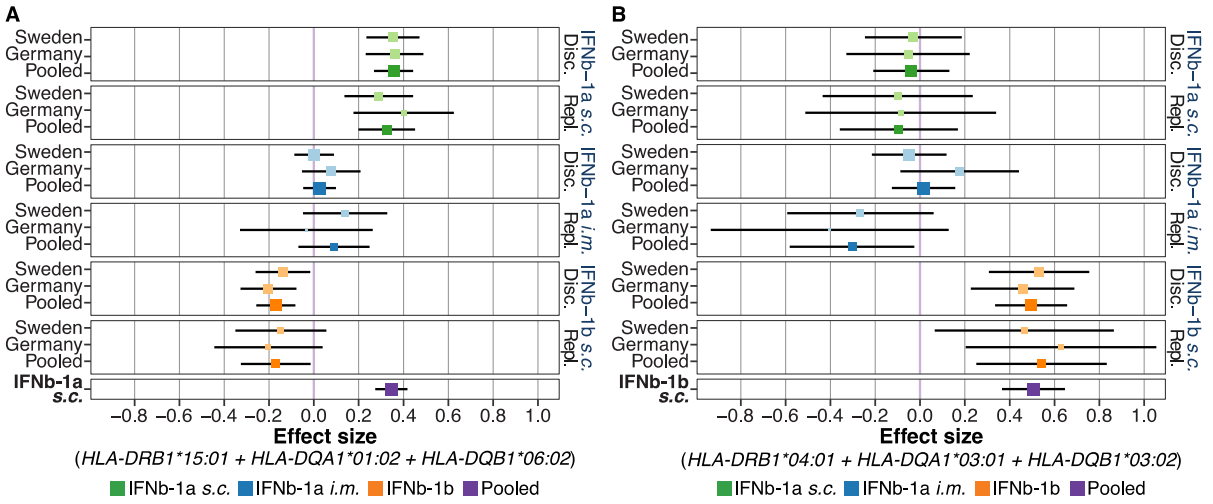


Figure 13: Treatment-specific HLA haplotypes.
A: The association of *DR15-DQ6* with nADA titers is specific for IFNβ-1a *s.c.* **B:** The association of *DR4-DQ3* with nADA titers is specific for IFNβ-1b *s.c.* Disc., discovery; Repl., replication.

The ancestral haplotype *DR3-DQ2* (*DRB1*03:01 + DQA1*05:01 + DQB1*02:01*) and its allele *HLA-DQB1*02:01* significantly protected against IFNβ-1a *s.c.*-induced ADA (Table 6). All other protective alleles were dependent on this extended haplotype. None of the protective alleles were significantly associated with ADA in patients treated with IFNβ-1b *s.c.*

By contrast, *HLA-DRB1*04:01* was the risk allele showing the most robust support for an association with all three ADA measurements in IFNβ-1b *s.c.*-treated patients. All associated alleles were part of the haplotype *DR4-DQ3* (*DRB1*04:01 + DQA1*03:01 + DQB1*03:02.*; Figure 13B, Table 6). These alleles were not significantly associated with ADA in patients treated with IFNβ-1a *s.c.*, and there were no significant protective alleles for patients receiving IFNβ-1b *s.c.*

Several factors likely contribute to these preparation-specific differences in genetic risk factors: The two diverse amino acids, diverging post-translational modifications, and, partly linked to that, the variable tendency of the IFNβ preparations to aggregate. Chemical alterations and spontaneously occurring protein modifications, caused by differences during the drugs' manufacturing, processing, and storage, can alter the proteins' surface and chemical properties.

All these differences may affect the protein stability and lead to altered processing of IFNβ in antigen-presenting cells, e.g., dendritic cells. Consequently, the diverse antigen peptides – produced depending on the state of IFNβ-1a and IFNβ-1b proteins – may be recognized and presented by different peptide-binding grooves and thus allelic variants of MHC class II molecules (Groot and Scott, 2007; Rombach-Riegraf et al., 2014).

When examining SNPs and HLA alleles previously reported to be associated with ADA, 15 of 20 candidate variants could be replicated. However, the previously published non-MHC SNP, MHC class I alleles, and *HLA-DRB1*11* alleles did not replicate.

Neutralizing ADA titers need to cross a threshold to become functionally relevant. Nevertheless, the same variants were associated with both nADA presence and titers. This finding suggests that most genetic risk factors might influence the likelihood of developing ADA instead of absolute titers. However, the two alleles *HLA-DRB1*04:08* and *HLA-DRB1*16:01* were associated only with bADA levels and/or with nADA titers but not with nADA presence. These two HLA alleles were also associated with nADA titers in nADA-positive patients, indicating that genetic factors exist that influence the amount of ADA.

ADA preparation	*Haplotype*	*AF KI*	*AF TUM*	*OR Disc.*	*p-value Disc.*	*OR Repl.*	$p_{one\text{-}sided}$ *Repl.*	*OR Pool.*	*p-value Pooled*
IFNβ-1a s.c.	DR15-DQ6	0.34	0.30	2.73	3.1×10^{-14}	3.41	1.5×10^{-07}	2.89	7.4×10^{-20}
IFNβ-1a s.c.	DR3-DQ2	0.13	0.12	0.40	9.1×10^{-07}	0.29	4.0×10^{-04}	0.37	3.7×10^{-09}
IFNβ-1b s.c.	DR4-DQ3	0.07	0.05	6.23	1.2×10^{-09}	14.7	6.1×10^{-06}	7.35	1.5×10^{-13}

Table 6: Selected significant HLA haplotypes.
Selected HLA haplotypes significantly associated ($p<3\times10^{-4}$) with the presence of nADA and replicated ($p_{one\text{-}sided}<1\times10^{-3}$). AF, allele frequency; KI, Karolinska Institutet, Sweden; TUM, Technical University of Munich, Germany; OR, odds ratio; Disc., discovery; Repl., replication; Pool., pooled.

Results from genetic analyses of ADA development can be translated to clinical care by generating personalized prediction models for ADA. To explore the feasibility of such prediction models, we derived PGS from our discovery-stage results and predicted the occurrence of nADA in the replication dataset. We did not observe evidence for a highly polygenic inheritance of nADA development (Table 7). Prediction models including only the top-associated variants performed better than models also incorporating signals selected by more liberal *p*-value thresholds. As expected, IFNβ-1a *s.c.*-specific models could not predict nADA in IFNβ-1b *s.c.*-treated patients and vice versa.

To generate risk estimates, we compared patients within the upper 30% of the top-associated PGS to patients within the lower 30% (Table 7). The sensitivity and specificity in this comparison (0.78 and 0.90, respectively, in Swedish IFNβ-1a *s.c.* patients and 0.80 and 0.76, respectively, in German patients) are not sufficient yet for a routine clinical test.

In the future, these prediction models could be optimized by the inclusion of environmental risk and protective factors like the use of immunosuppressants and antibiotics, infections, and tobacco smoking.

Nevertheless, our results support the use of genetic risk stratification as a personalized medicine tool, guiding IFNβ treatment preparation recommendations and the frequency of testing for the presence of nADA. Patients at high genetic risk for nADA should either switch to a different drug or be monitored more closely, as suggested for other conditions (Lewis and Vassos, 2020).

Preparation	*Model*	*Cohort*	*OR*	*95% CI*	*p-value*	*AUC*	R^2	*Sensit.*	*Specif.*
IFNβ-1a s.c.	PGS threshold $p\leq5\times10^{-08}$	KI	3.89	2.35-6.45	1.44×10^{-07}	0.85	0.42	0.78	0.78
IFNβ-1a s.c.	PGS threshold $p\leq5\times10^{-08}$	TUM	2.56	1.56-4.21	2.11×10^{-04}	0.76	0.24	0.68	0.65
IFNβ-1a s.c.	**PGS top *vs.* bottom 30%**	**KI**	**73.86**	**11.77-463.61**	$\mathbf{4.42\times10^{-06}}$	**0.91**	**0.59**	**0.78**	**0.90**
IFNβ-1a s.c.	**PGS top *vs.* bottom 30%**	**TUM**	**13.78**	**3.00-63.28**	$\mathbf{7.45\times10^{-04}}$	**0.83**	**0.38**	**0.80**	**0.76**
IFNβ-1a s.c.	rs77278603-A dominant	KI	9.16	2.48-33.79	8.79×10^{-04}	0.78	0.31	0.57	0.76
IFNβ-1a s.c.	rs77278603-A dominant	TUM	3.85	1.10-13.49	3.51×10^{-02}	0.72	0.20	0.56	0.68
IFNβ-1b s.c.	PGS threshold $p\leq1\times10^{-06}$	KI	2.40	1.45-3.97	6.46×10^{-04}	0.78	0.33	0.57	0.85
IFNβ-1b s.c.	PGS threshold $p\leq1\times10^{-06}$	TUM	2.15	1.43-3.23	2.28×10^{-04}	0.73	0.22	0.73	0.58
IFNβ-1b s.c.	PGS top vs. bottom 30%	KI	10.16	2.30-44.95	2.25×10^{-03}	0.83	0.46	0.58	0.87
IFNβ-1b s.c.	PGS top vs. bottom 30%	TUM	5.97	2.03-17.52	1.14×10^{-03}	0.78	0.33	0.69	0.75
IFNβ-1b s.c.	**rs28366299-A dominant**	**KI**	**9.78**	**2.68-35.74**	$\mathbf{5.62\times10^{-04}}$	**0.83**	**0.40**	**0.62**	**0.83**
IFNβ-1b s.c.	**rs28366299-A dominant**	**TUM**	**7.56**	**3.01-19.02**	$\mathbf{1.71\times10^{-05}}$	**0.80**	**0.33**	**0.77**	**0.57**

Table 7: Treatment-specific prediction of the presence of nADA.
Covariates used in all models: sex, age, treatment duration, titration site, and ancestry components. The top models per treatment preparation are indicated in bold font. OR, odds ratio; CI, confidence interval; *p*-value of the genetic component; AUC, area under the receiver operating characteristic curve; R^2, Nagelkerke's pseudo-R^2; Sensit., sensitivity; Specif., specificity.

Bipolar Multiplex Families Have an Increased Burden of Common Risk Variants for Psychiatric Disorders

Andlauer TFM, Guzman-Parra J, Streit F, …, Nöthen MM, Rietschel M (2019): Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. **Molecular Psychiatry**, doi:10.1038/s41380-019-0558-2

Further studies relevant for this section:

Guzman-Parra J, Streit F, …, **Andlauer TFM***, Rietschel M* (*shared last*) (2021): Clinical and genetic differences between bipolar disorder type 1 and 2 in multiplex families. **Translational Psychiatry**, 11:31

Andlauer TFM, Nöthen MM (2020): Polygenic scores for psychiatric disease: from research tool to clinical application. **medizinische genetik** 32(1) 39-45

Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, **Andlauer TFM**, …, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; MDD Working Group of the PGC (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. **Nature Genetics** 50(5):668-681

Polygenic disorders affecting the CNS – e.g., MS, MDD, and BD – can accumulate in families. Several factors may contribute to such an increased incidence of disorders in single families: First, exposure to shared environmental risk factors; second, a high burden of common (*i.e.*, frequently occurring) genetic risk variants; third, rare variants conferring risk. The three factors could act both separately and in combination.

As is the case for all polygenic disorders, many variants with small individual effect sizes contribute to genetic risk for BD (Stahl et al., 2019). Some risk variants confer risk to several psychiatric disorders like BD, MDD, and SCZ (PGC-CDG et al., 2019).

For BD, the phenotypic variance explained by genotyped common variants (*i.e.*, the SNP-based heritability) is estimated to be 0.17–0.23 on a liability scale, while BD's overall heritability is above 70% (Stahl et al., 2019). As for other polygenic disorders, this discrepancy cannot be easily explained, with rare variants constituting a prime candidate as the source of the missing heritability (Manolio et al., 2009; Wainschtein et al., 2019).

Researchers have often used multiplex families showing a high incidence of a disorder to screen for rare (causal) variants. However, multiplex families may also accumulate a high common-variant risk burden. For example, assortative mating or a limited available genetic pool can contribute to such a risk load. We explored this hypothesis that common variants contribute significantly to disorder incidence within multiplex families. To this end, we analyzed whether families with a high incidence of BD and MDD, belonging to the Andalusian ABiF study (Guzman-Parra et al., 2017), show a high polygenic risk load for psychiatric disorders.

We examined 33 families with 395 members, of which n=166 were diagnosed with BD (BD type I (BD-I), n=115; BD type II (BD-II), n=41; not otherwise specified BD, n=10), n=78 with MDD, and n=151 had no recorded history of an affective disorder. Each family contained at least two BD patients.

The family members were genotyped on Illumina Infinium PsychArray BeadChips. We combined n=384 family members with a population-based case/control dataset containing 161 BD patients (BD-I, n=156; BD-II, n=5) and 277 unscreened controls. Thirty-five unaffected and eight MDD-diagnosed married-in family members were excluded from all comparisons of families to other datasets. The joint dataset was imputed using the 1000 Genomes phase 3 reference panel.

We calculated PGS (Andlauer and Nöthen, 2020) for the psychiatric disorders BD, MDD, and SCZ using published GWAS as training data (PGC-SCZ, 2014; Stahl et al., 2019; Wray et al., 2018). As negative controls, we generated PGS for late-onset Alzheimer's disease (LOAD) (Lambert et al., 2013) and also simulated PGS. Furthermore, we computed *Shared* PGS from all variants with matching

effect directions and p<0.05 in the BD, MDD, and SCZ GWAS. The respective PGS weights were calculated by random-effects meta-analysis of the three GWAS. Disorder-specific PGS, e.g., of BD corrected for the correlation with MDD (BD-MDD), were calculated using genome-wide inferred statistics (GWIS) (Nieuwboer et al., 2016).

We conducted all case/control analyses using logistic mixed models with GMMAT. GenAbel's function *polygenic* was used for linear mixed models. In both cases, the genetic relationship matrix was modeled as a random effect to account for family structure.

In our study, we first compared family members diagnosed with BD to healthy controls from the general population. BD cases within the ABiF families had higher BD (p_{PGS}=0.1: OR=2.97, one-sided p=1.9×10^{-11}), SCZ, *Shared*, and GWIS BD-MDD PGS than population-based controls (Figure 14A). The association of the MDD PGS did not pass correction for multiple testing. An increased BD PGS in multiplex families compared to controls has also been observed in other studies (Fullerton et al., 2015). The significant increase of SCZ and *Shared* PGS might be a consequence of the genetic correlation between psychiatric disorders (PGC-CDG et al., 2019).

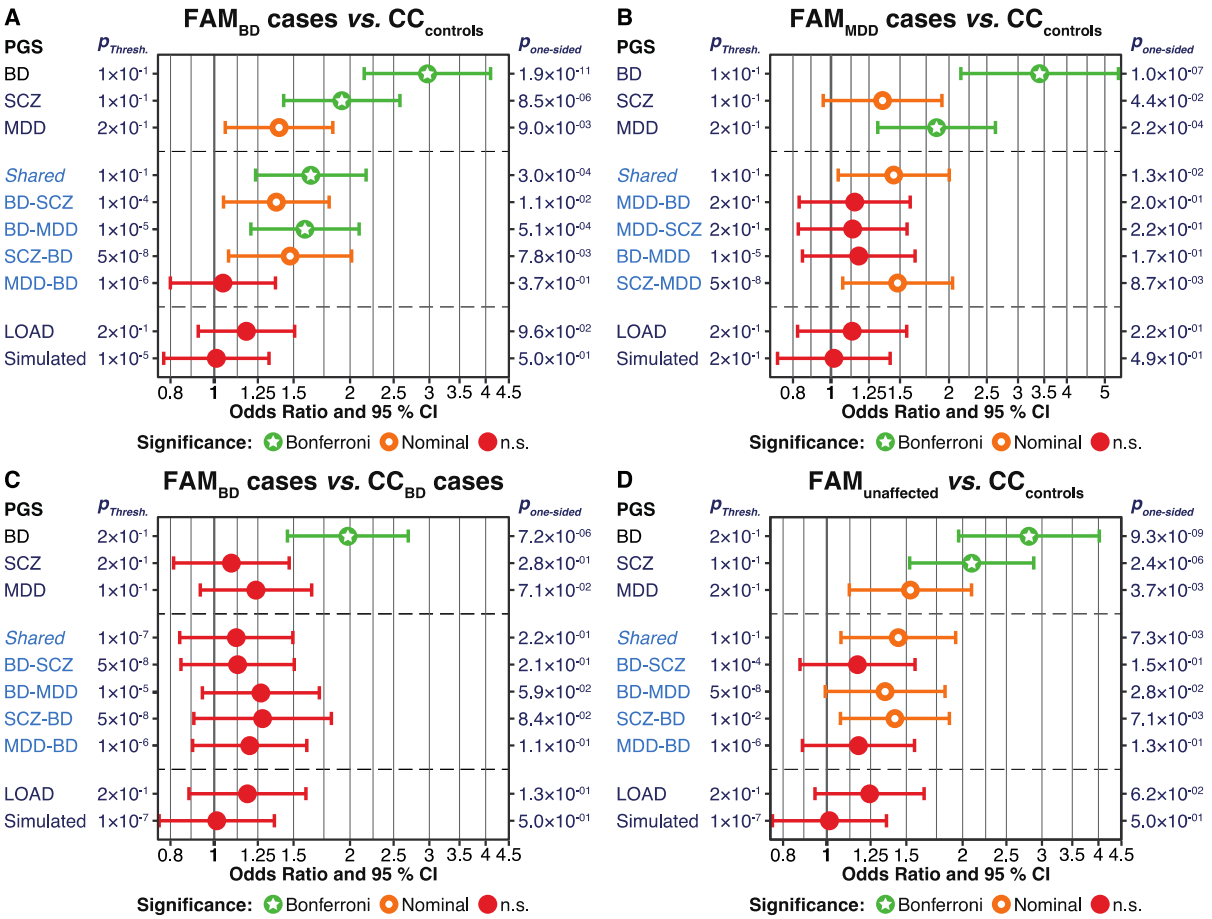


Figure 14: Comparison of polygenic risk scores between families and unrelated samples.
The plots show one-sided p-values, following the hypothesis that family members (FAM) have higher PGS than population-based cases and controls (CC). We plotted odds ratios (OR, filled circles) and 95% confidence intervals (CI). PGS were normalized by Z-score standardization. Simulated: Mean results from 10,000 simulated PGS. $p_{Thresh.}$, the top-associated p-value threshold used for calculating the PGS. Colors indicate significance: n.s., not significant; Nominal, p<0.05; Bonferroni, significant after Bonferroni correction for multiple testing (α=0.05/60=0.00083). A-D show four different comparisons between family members and BD cases or controls (see headings).

When comparing family members suffering from unipolar depression to controls, the BD (p_{PGS}=0.1: OR=3.41, one-sided p=1.0×10^{-7}) and MDD PGS were significantly increased, while SCZ and *Shared* PGS were nominally higher (Figure 14B). Note that the genetic correlation between MDD and SCZ is higher than between BD and SCZ (PGC-CDG et al., 2019).

Next, we compared familial BD cases to BD cases recruited from consecutive clinical admissions. Some of the latter BD cases may also belong to cryptic BD multiplex families, but were not ascertained as such. The familial BD cases showed significantly higher BD PGS than the sporadic BD cases (p_{PGS}=0.2: OR=1.98, one-sided p=7.2×10^{-6}, Figure 14C). There was no support for an increase of any other PGS in this comparison.

We also analyzed healthy family members relative to the population-based controls and observed significantly higher BD (p_{PGS}=0.2: OR=2.81, one-sided p=9.3×10^{-9}) and SCZ PGS for the family members (Figure 14D). The MDD and *Shared* PGS were nominally significant. No previously published family-based BD study had conducted a similar comparison.

Subsequently, we analyzed whether family members diagnosed with BD had higher psychiatric disorder PGS than unaffected family members. We observed significantly increased BD and BD-MDD GWIS PGS in BD patients, with the *Shared* and BD-SCZ PGS being nominally significant (Figure 15A). Two previous studies found only a nominally significant increase of BD PGS in different cohorts of affected *vs.* unaffected multiplex family members (de Jong et al., 2018; Szatkiewicz et al., 2019).

By contrast to the findings for BD patients, none of the PGS associations were significant after correction for multiple testing when analyzing family members with unipolar MDD (Figure 15B). Given the lower number of MDD cases in the ABiF families, the statistical power of MDD-based analyses was considerably lower than for BD. The increased BD PGS in familial MDD cases may indicate that, in some cases, the current MDD diagnosis constituted a prodromal stage of BD (Berk et al., 2007). Given the proposed severity continuum of affective psychiatric disorders, MDD may possibly be more strongly driven by BD risk variants in the ABiF families than in typical MDD cases from the general population.

We assessed whether assortative mating might have contributed to the high polygenic risk scores for BD in the ABiF families. Indeed, the unaffected married-in individuals had higher BD PGS than population-based controls (p=6.5×10^{-5}), but showed a similar risk load as other healthy family members (Figure 15C). Nevertheless, neither did the BD PGS increase significantly over generations nor did the age at onset decrease (Figure 15D). Thus, we did not observe support for anticipation. De Jong *et al.* made opposite observations in their multiplex families, *i.e.*, they reported anticipation but no evidence for assortative mating (de Jong et al., 2018).

That none of the married-in family members was diagnosed with BD does not argue against assortative mating. Undiagnosed individuals with high BD PGS may still display sub-threshold characteristics of BD, e.g., a broader range of emotions or creativity (Power et al., 2015). Assortative mating may have also occurred and led to an enrichment of risk variants already during previous generations. Assortative mating may then rather contribute to maintaining the high risk load than increasing it further, which would be observable as anticipation.

In this study, we strove to improve our understanding of the genetic risk landscape of psychiatric disorders by analyzing genetic risk in BD multiplex families. Our results indicate that a psychiatric cross-disorder risk burden increased the overall liability for affective disorders in the ABiF multiplex families, while specifically BD-associated genetic variants shaped the outcome towards BD and thus contributed to its high incidence.

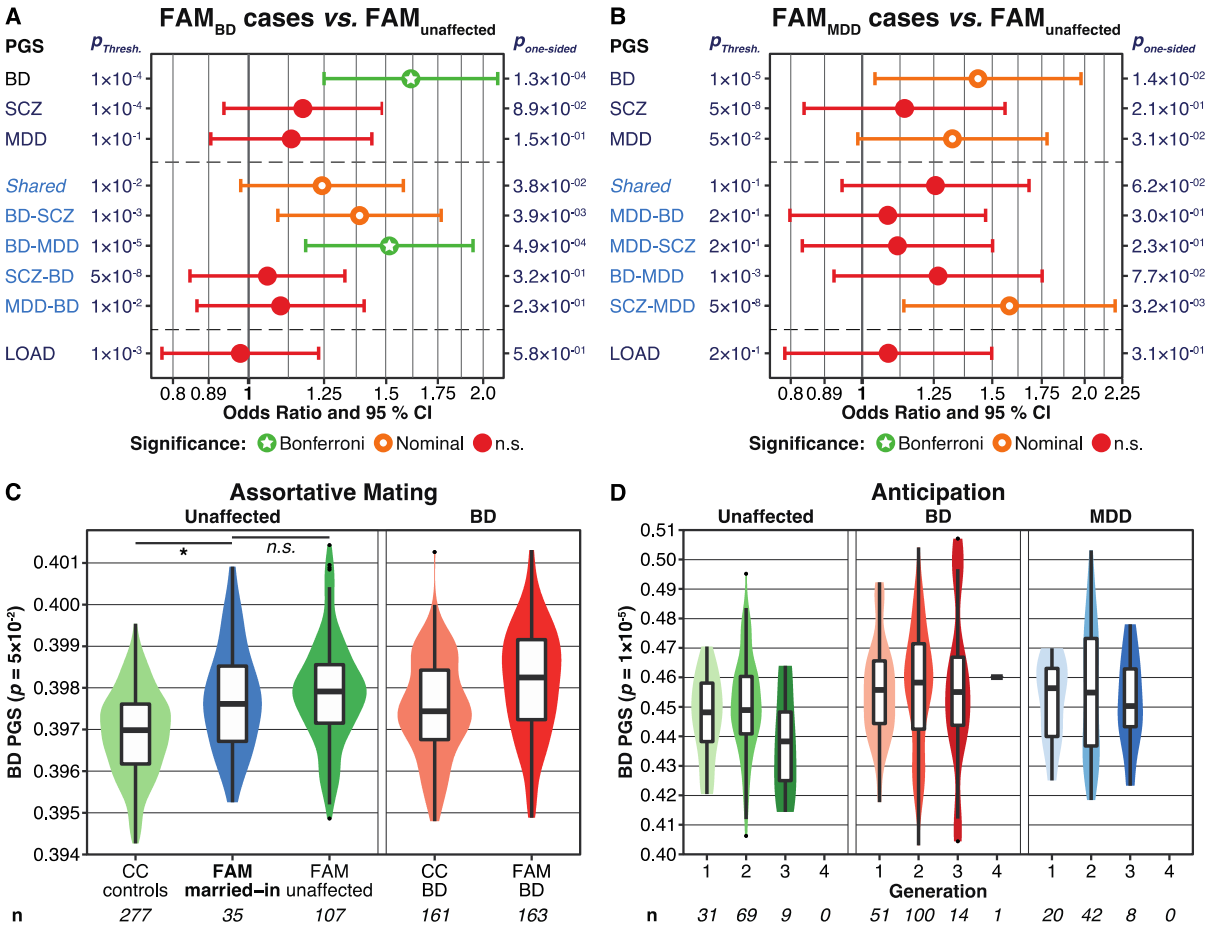


Figure 15: Analysis of polygenic risk scores within families.

A-B show one-sided *p*-values (hypothesis: affected family members (FAM) have higher PGS than unaffected ones). OR, odds ratio (filled circles); CI, confidence interval. $p_{Thresh.}$, top-associated *p*-value threshold. Significance: n.s., not significant; Nominal, $p<0.05$; Bonferroni, significant after Bonferroni correction for multiple testing.

C-D: n, sample size. **C:** Analysis of assortative mating for the BD PGS comparing unaffected, married-in individuals with no parent among the ABiF families to other study groups. Married-in family members had significantly higher BD PGS than population-based controls ($p=6.5\times10^{-5}$), but not compared to other unaffected family members ($p\geq0.167$). **D:** Analysis of anticipation, the BD PGS did not increase across generations ($p=0.45$).

Here, we did not differentiate between BD subtypes. Both BD-I and -II patients experience depressive episodes. Yet while a hypomanic episode is sufficient for a BD-II diagnosis, BD-I patients also experience at least one manic episode, often with psychotic symptoms,. This clinical variation is partly driven by genetic differences: BD-I patients show higher SCZ PGS than BD-II, while BD-II cases have higher MDD PGS (Stahl et al., 2019). Together with clinical findings, this result indicates that a severity spectrum of psychiatric disorders exists (MDD < BD-II < BD-I < SCZ). Thus, we analyzed whether a correlation between genetic risk for psychiatric disorders and disease severity exists in the ABiF study (Guzman-Parra et al., 2021). Indeed, BD-I patients suffered from more severe symptoms during manic and more frequently showed incapacity in depressive episodes. Compared to BD-II, BD-I cases also had lower MDD PGS, arguing against a simplistic genetic severity continuum. Furthermore, a higher BD PGS was significantly associated with suicidal ideation in ABiF BD patients. Disorder-specific risk load (MDD *vs.* SCZ) may thus shape disorder presentation and subtype development even in families. The relative BD risk load may influence BD severity and thereby aggravate symptoms, supporting a BD severity continuum.

Genetic Factors Influencing a Neurobiological Substrate for Psychiatric Disorders

Andlauer TFM, Mühleisen T, . . ., Müller-Myhsok B, Cichon S (2021): Genetic factors influencing a neurobiological substrate for psychiatric disorders. **Translational Psychiatry**, 11(1):192

Further studies relevant for this section:

Lee PH, Anttila V, Won H, Feng YA, Rosenthal J, Zhu Z, . . ., **Andlauer TFM**, . . ., Geschwind DH, Neale BM, Kendler, KS, Smoller JW (2019): Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. **Cell** 179(7):P1469-1482.e11

Andlauer TFM, Müller-Myhsok B, Ripke S (2018): Statistical genetics: Genome-wide studies. In: **Psychiatric Genetics – A Primer for Clinical and Basic Scientists**. Edts: T. Schulze and F. McMahon; Oxford University Press, doi:10.1093/med/9780190221973.003.0004

Patients suffering from CNS disorders show differences compared to healthy controls in CNS structural and functional magnetic resonance imaging (MRI) (Erp et al., 2018; Hibar et al., 2018; Schmaal et al., 2016). Genetic risk factors for the disorders might contribute to such observable differences in brain function and atrophy.

While most MS genetic risk loci indicate a molecular etiology of the disorder predominantly in the peripheral immune system, outside the CNS (IMSGC, 2019a), almost all genetic risk variants for psychiatric disorders can be linked to brain-specific functions, e.g., in neuronal development and differentiation or synaptic function and plasticity (PGC-SCZ, 2014; Stahl et al., 2019; Wray et al., 2018).

If a symptomatic severity continuum were to exist between psychiatric disorders and since the same genetic risk loci confer risk to several different psychiatric disorders, structural and functional brain-level alterations are expected to partly overlap across disorders.

Goodkind and colleagues conducted a meta-analysis on 193 case/control voxel-based morphometry (VBM) structural MRI studies including N=7,381 psychiatric patients from six diagnostic groups (SCZ, BD, MDD, addiction, obsessive-compulsive disorder, and anxiety) and N=8,511 healthy controls (Goodkind et al., 2015). The study described gray matter atrophy across affective and psychotic diagnoses in the left and right anterior insular cortices (AIC) and the dorsal anterior cingulate cortex (dACC, Figure 16).

The authors described gray matter loss in these three functionally connected regions as a common neurobiological substrate for psychiatric disorders. Follow-up analyses confirmed that the three regions represent the hub nodes of the salience network.

The salience network is a major functional brain network implicated in psychiatric disorders (Menon, 2015; Uddin, 2014). It is essential for evaluating behaviorally relevant signals, focusing attention, and executing goal-directed behaviors.

It is unclear whether gray matter loss in the salience network precedes (and thus constitutes a risk factor for) or is a consequence of psychiatric disorders (Seeley et al., 2009; Tozzi et al., 2020). Furthermore, the observed atrophy might also be a consequence of accelerated brain aging in psychiatric patients (Koutsouleris et al., 2014).

Therefore, we conducted a population-based GWAS of the combined regional common substrate volume proposed by Goodkind *et al.* to assess whether genetic risk factors might predispose to common substrate-specific gray matter atrophy. To this end, we extracted the common substrate from five population-based cohorts and combined all three regions by PCA (Figure 16). We termed the first principal component the Component of the Common Substrate (CCS).

We used N=2,271 individuals from four cohorts in the discovery-stage GWAS and further N=865 from one cohort for replication. All MRI data were processed using the same protocols.

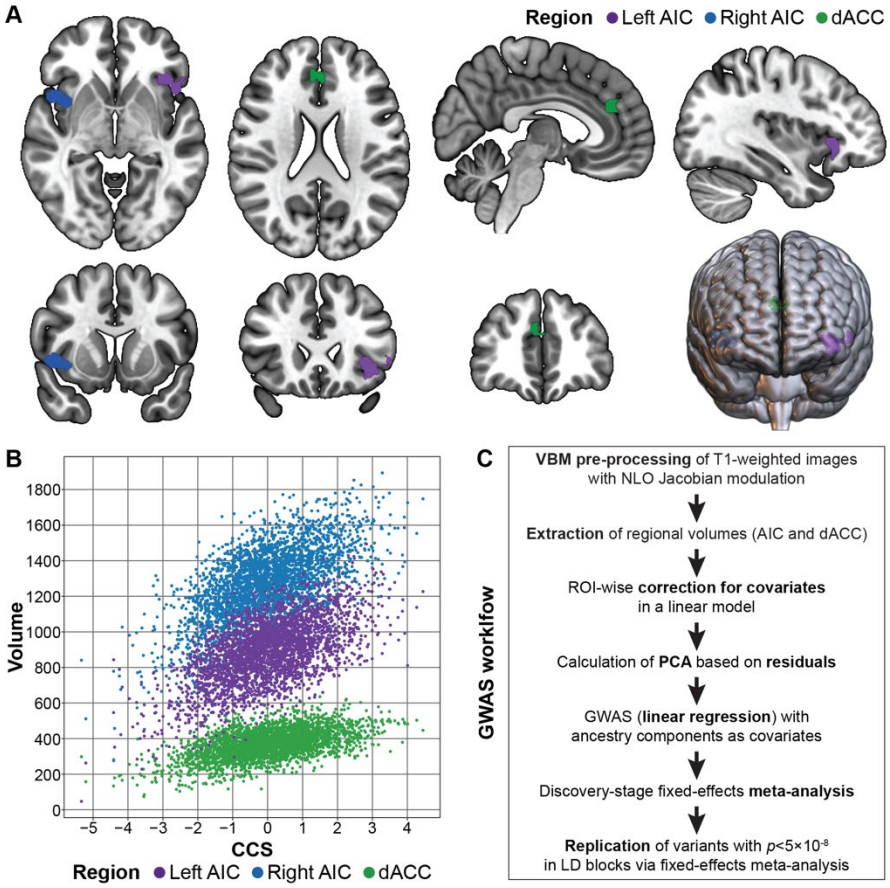


Figure 16: Calculation of the CCS and GWAS workflow.
A: The three regions forming the common neurobiological substrate of mental illness mapping to the left (magenta) and right (blue) anterior insular cortex (AIC) and the dorsal anterior cingulate cortex (dACC; green).
B: The first PCA component of the common substrate (CCS; x-axis) compared to the three regional volumes. For the calculation of the CCS, the volumes were corrected for age, age^2, and sex; handedness was used for three of the five GWAS cohorts. Correlations: left and right AIC: r=0.65; left AIC and dACC: r=0.52; right AIC and dACC: r=0.46.
C: GWAS workflow.

The five cohorts were imputed to the 1000 Genomes Phase I reference panel in batches by genotyping microarray type.
The CCS explained 66.5% of the variance of the three common substrate regional volumes, which reduced to 55.4% after correction for the covariates age, age^2, and sex (plus handedness in three of the cohorts). The SNP-based heritability of the CCS, calculated using GCTA, was estimated at h^2_g=0.50. This finding demonstrates that genetic factors play a significant role in shaping the common substrate volumes in unscreened individuals from the general population.
We conducted GWAS to identify specific genetic loci associated with a smaller CCS. One locus on chromosome 5 was significantly associated with the CCS in the four discovery cohorts (Figure 17). Among the twelve highly correlated, significantly associated variants, rs17076061 replicated (Figure 17; MAF=0.36; discovery: p=1.51×10^{-8}; replication: one-sided p=9.91×10^{-3}; meta-analysis: p=1.46×10^{-9}). The support for an association of this variant with the three individual common substrate regional volumes and the total gray matter volume was lower than for the CCS (pooled dataset; left AIC: p=7.00×10^{-8}; right AIC: p=2.63×10^{-6}; dACC: p=1.77×10^{-3}; total gray matter: p=1.85×10^{-4}).
Several predicted, uncharacterized long intergenic non-coding RNAs and two protein-coding genes, *BOD1* and *STC2*, map to the identified locus (Figure 17C). The SNP rs17076061 is significantly associated with the expression of *STC2* in pancreatic tissue (p=3.6×10^{-8}) and, possibly, in the anterior cingulate cortex (p=0.06). The gene *STC2* codes for a secreted glycoprotein putatively acting in an auto- or paracrine manner. It might have a neuroprotective function (Byun et al., 2010). BOD1 might be relevant for psychiatric disorders (Esmaeeli-Nieh et al., 2016; Kim et al., 2014).
In two independent gene-set analyses, the pathway "SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion" was robustly associated with the CCS.

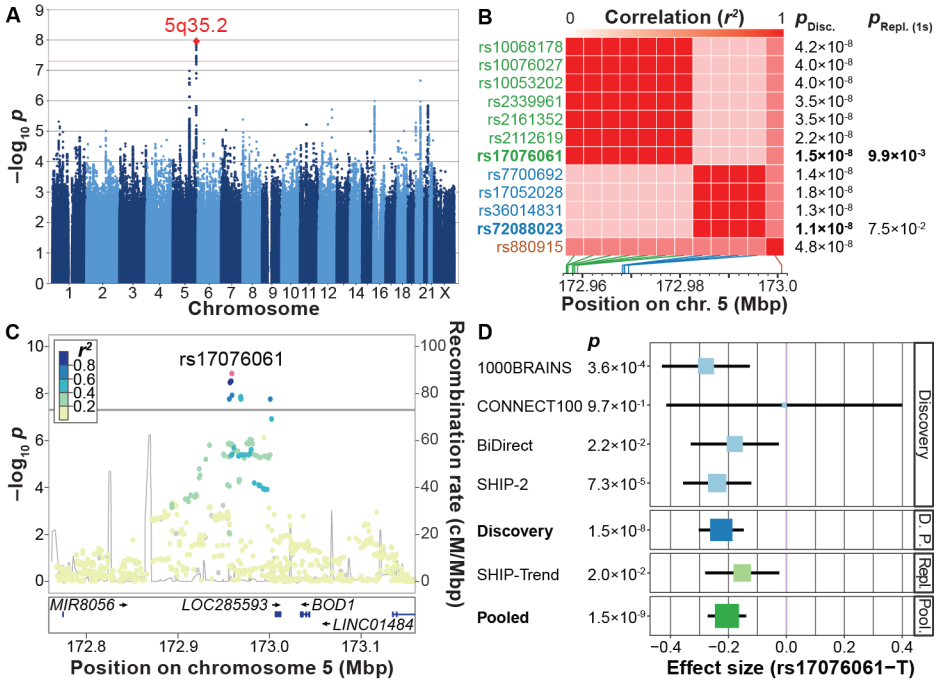


Figure 17: CCS GWAS results.
A: Manhattan plot of the discovery-stage GWAS. The red line indicates significance. **B:** Pairwise LD (1000 Genomes population CEU) between the 12 significant variants. **C:** Regional association plot of the GWAS meta-analysis. **D:** Forest plot for the SNP rs17076061 (T). $P_{Disc.}$, discovery-stage p-value; $p_{Repl.(1s)}$, replication-stage one-sided p-value; Mbp, mega base pair; cM, centiMorgan. D.P., pooled discovery cohorts; Repl., replication cohort; Pool., pooled analysis of discovery and replication cohorts.

SEMA3A codes for Semaphorin-3A, a chemorepellent receptor that mediates axon guidance and acts as a chemoattractant for dendrite growth. Plexins are its signal-transducing subunits. Semaphorin-3A and Plexin-A2 are associated with psychiatric disorders like SCZ and their synaptic pathology (Eastwood et al., 2003; Wray et al., 2007). Semaphorin-3A is expressed in MS lesions (Costa et al., 2015) and may contribute to neurodegeneration in Alzheimer's disease (Good et al., 2004) and neuronal regeneration after brain trauma (Mecollari et al., 2014).

Next, we examined whether genetic variants associated with the CCS might directly confer risk for psychiatric disorders. To this end, we first looked up the association of rs17076061 with psychiatric disorders in published GWAS. The SNP was not associated with MDD, BD, or SCZ (PGC-SCZ, 2014; Stahl et al., 2019; Wray et al., 2018) but showed very weak support for an association in a psychiatric cross-disorder GWAS (OR=1.035, one-sided p=0.048 (PGC-CDG et al., 2019)).

Next, we extended these analyses to the genome-wide level (Franke et al., 2016). Using LD score regression, we found no evidence for a genetic correlation between the CCS GWAS and the four psychiatric GWAS (p≥0.08). Similarly, rank-rank hypergeometric overlap tests showed no significant overlap of SNPs ranked by their association strength between the studies (p≥0.06; Figure 18). In binomial sign tests, CCS-associated variants did not show the opposite effect direction in the psychiatric disorder GWAS more often than expected by chance (p≥0.50). We also calculated PGS using the four psychiatric GWAS as training data. None of them were significantly associated with the CCS after correction for multiple testing (p≥0.05).

Finally, we generated PGS with the CCS GWAS as the training data in psychiatric case/control cohorts. We analyzed three cohorts, BiDirect (n=582 MDD patients; n=311 healthy controls), MPIP (n=385 MDD patients; n=197 healthy controls), and FOR2107 (n=769 MDD, n=127 BD, n=72 SCZ, and n=43 schizoaffective (SZA) patients; n=867 healthy controls). After correction for multiple testing, the CCS PGS was not significantly lower in any case-control comparison (unadjusted p≥0.02).

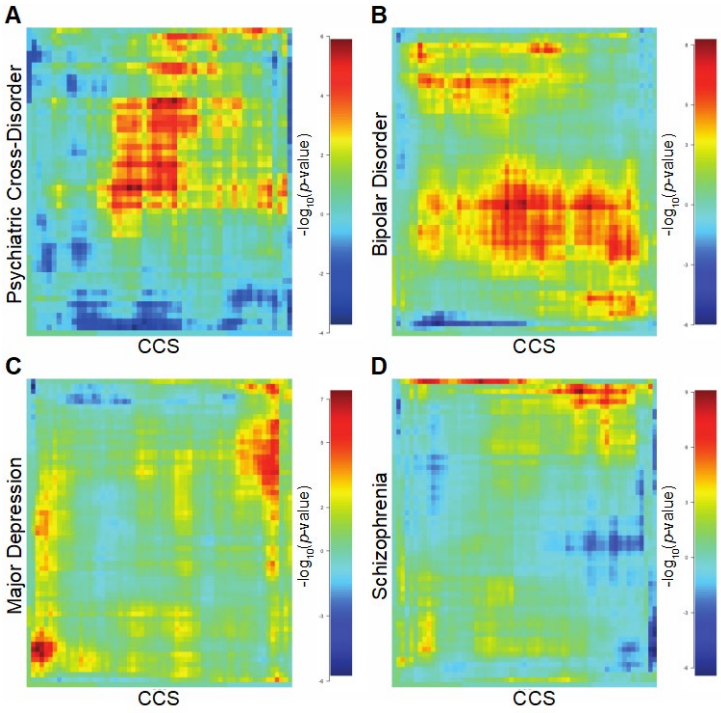


Figure 18: Rank-rank hypergeometric overlap maps.
Comparison of CCS-associated and psychiatric disorder GWAS-associated variants. We analyzed whether the order of genetic variants ranked by their association strength was random between studies using rank-rank hypergeometric overlap (RRHO) tests. For this analysis, variants were LD-pruned using the 1000 Genomes phase 3 EUR subset.
The respective RRHO overlaps with the CCS were:
A: Psych. cross-disorder: 0.29 (p=0.53)
B: Bipolar disorder: 0.21 (p=0.06)
C: Major depression: 0.04 (p=0.18)
D: Schizophrenia: 0.02 (p=0.15)

Given these negative results, we used the same case/control cohorts to assess whether the CCS, an artificial construct derived from Goodkind's meta-analysis, is smaller in psychiatric patients. We observed striking differences, with the CCS being significantly smaller in a meta-analysis of our three MDD cohorts (p=1.3×10^{-7}; Figure 19A). Moreover, the median CCS showed a stepwise decrease along the affective-psychosis axis in the transdiagnostic FOR2107 cohort (Figure 19; MDD: p=3.9×10^{-3}; BD: p=2.8×10^{-5}; SZA: p=2.6×10^{-8}; SCZ: p=6.6×10^{-10}), clearly replicating the existence of a common neurobiological substrate of psychiatric disorders.
Nevertheless, we found no genetic overlap between the CCS and psychiatric disorders. Several reasons may clarify this negative result. SNP rs17076061 explained only a small fraction of the CCS variance (R^2=1.2%). Moreover, diagnostic status explained only a small share of the CCS variance in our analyses (MDD: R^2=1.0%; SCZ: R^2=4.2%). Similarly, individual genetic variants explain only a small part of the observed disorder prevalence. We thus should expect any psychiatric risk mediated by this SNP to be very low.
Also the genome-wide comparisons have inherent problems: If only some variants were associated with both brain volume and disorder risk or if different loci exhibited mixed effect directions, genome-wide methods would likely fail to identify the overlap. Furthermore, none of the genome-wide methods used by us could adequately account for gene-by-gene or gene-by-environment effects. In particular gene-by-environment effects likely influence brain atrophy in the context of psychiatric disorders, which might have contributed to our negative results. Also environmental factors not documented in our datasets might have led to a decrease of common substrate regional volumes. For example, salience network dysfunction and cortical gray matter loss have been documented in individuals who have experienced childhood maltreatment (Harmelen et al., 2010; Tozzi et al., 2020; Werff et al., 2013).
Interestingly, other studies also found only either a weak or no significant relationship between the genetic architecture of MDD and SCZ and regional brain volumes (Franke et al., 2016; Reus et al., 2017; Satizabal et al., 2019; Smeland et al., 2017; Wigmore et al., 2017). So far, a robust correlation was merely identified between the brainstem volume and ADHD but not for SCZ or BD (Satizabal et al., 2019).

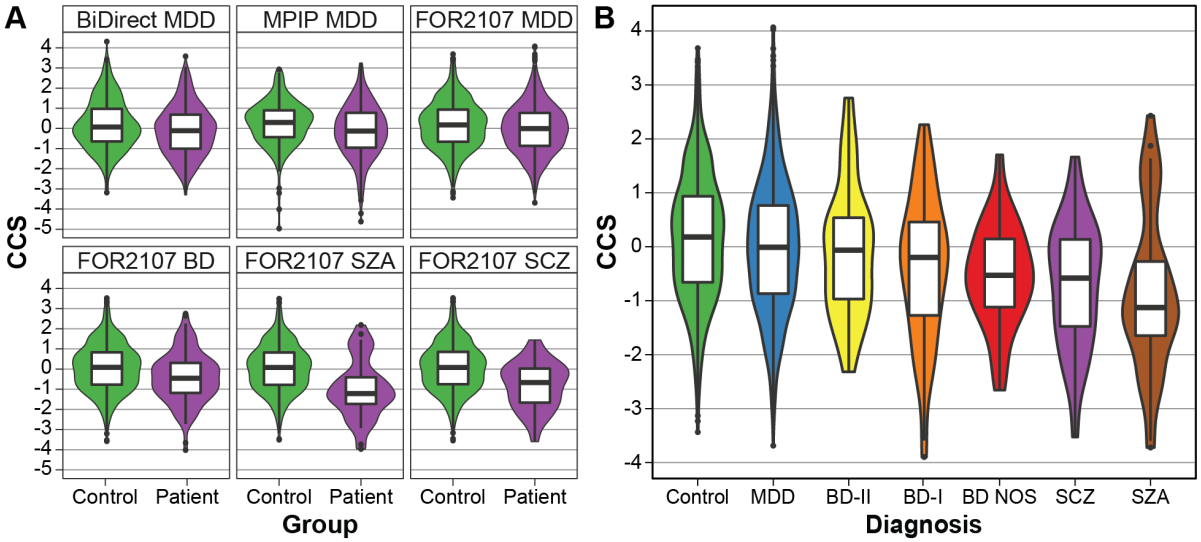


Figure 19: Analysis of the CCS in psychiatric case/control cohorts.
A: We observed a significantly smaller CCS in psychiatric patients than healthy controls in three different psychiatric case/control samples. **B:** The CCS across different diagnostic groups in the FOR2107 cohort.
Abbreviations: BD NOS, not otherwise specified BD; SZA, schizoaffective disorder; SCZ, schizophrenia.

Given this failure by us and others to robustly link genetic variants influencing brain volumes to psychiatric risk leaves three possibilities: First, given the complex and polygenic architecture of the studied traits and disorders, all studies were underpowered to detect the small effects. Second, gray matter atrophy is predominantly a consequence of psychiatric disorders and not a risk factor. Third, the connection between both is indirect: Smaller brain volumes, e.g., in hub regions of the salience network, decrease the robustness of brain networks, moving them closer to the boundaries of their normal functional range. In combination with other genetic and, especially, environmental risk factors, this vulnerability may increase the risk of developing a psychiatric disorder. High-powered longitudinal studies in well-characterized cohorts would be required to examine these options further. Also functional studies might be necessary to uncover a link to disorder risk (Liu et al., 2021). In the absence of such studies, we hypothesized that smaller common substrate regions might be a consequence of accelerated aging in psychiatric patients. If genetic variants influencing the CCS do not directly confer risk to psychiatric disorders, they might instead contribute to aging-related processes relevant for psychiatric disorders (Han et al., 2020). To examine this hypothesis, we compared our CCS GWAS results with GWAS summary statistics for epigenetic accelerated aging (Lu et al., 2017) and longevity (Deelen et al., 2014), using the battery of comparisons applied previously in the comparison to psychiatric comparisons. After correction for multiple testing, neither an analysis of SNP rs17076061 nor any genome-wide tests supported a significant overlap. The salience network is particularly involved in accelerated cognitive decline during aging (Corte et al., 2016). Therefore, we finally examined a proxy of salience network dysfunction. This network exhibits neurodegeneration in behavioral frontotemporal dementia (bvFTD) (Seeley et al., 2009), a disorder leading to severe executive disturbances and personality changes. However, we found no significant genetic overlap between the CCS and a bvFTD GWAS (Ferrari et al., 2014).
Triple interaction analyses between genetic variants, disease status, and the CCS in large case/control cohorts would be required to reliably examine the relationship between genetic variants associated with the CCS and accelerated aging in the salience network hubs in the context of psychiatric disorders.

Discussion and outlook

Extraordinary progress has been made in characterizing the genetic risk landscape of various disorders affecting the CNS, including MS, dyslexia, MDD, BD, and SCZ (Andlauer et al., 2016, 2019; Arloth et al., 2020; Dankowski et al., 2015; Gialluisi et al., 2019, 2020; IMSGC, 2019a; Kendall et al., 2021; Kular et al., 2018; Levey et al., 2020; Mullins et al., 2020; PGC-CDG et al., 2019; PGC-SCZ et al., 2020; Wray et al., 2018).

These studies also demonstrated the utility of polygenic scores (PGS). The development of PGS holds the promise that risk predictions may eventually become feasible even for complex disorders (Andlauer and Nöthen, 2020; Khera et al., 2018; Muse et al., 2017). Next to predicting disorder risk, PGS can also be employed to predict treatment efficacy and thus guide personalized therapy. For example, we have applied PGS to predict patients' risk of developing neutralizing antidrug antibodies, an important treatment complication (Andlauer et al., 2020).

However, PGS are not only a promising instrument for risk stratification but also a powerful research tool. During the last years, we have used PGS to characterize the psychiatric risk burden of BD and MDD cases in multiplex families (Andlauer et al., 2019), assessed how genetic variants influence psychiatric symptoms like suicidal ideation (Guzman-Parra et al., 2021), and examined whether genetic risk for psychiatric disorders correlates with a transdiagnostic CNS imaging marker (Andlauer et al., 2021).

PGS could be highly useful in characterizing known and latent disease subtypes and support the identification of more homogenous patient subgroups. Such applications of PGS can increase our understanding of disease etiology and support targeting therapies to the optimal patient groups. We have recently applied unsupervised clustering of various baseline variables to identify five psychiatric patient groups clustered by symptom severity across traditional diagnostic boundaries and characterized them using PGS (Pelin et al., 2021). The age at disease onset (AAO) may also delineate patients suffering from more severe disorder subtypes (Hagenaars et al., 2020; Kalman et al., 2018; Power et al., 2017). We have characterized the AAO in 12,977 BD patients and found it to be significantly lower in BD patients with high polygenic scores for autism spectrum disorder (β=-0.34 years, SE=0.08), MDD (β=-0.34 years, SE=0.08), SCZ (β=-0.39 years, SE=0.08), and educational attainment (β=-0.31 years, SE=0.08), but not BD (Kalman et al., 2021). This finding supports the hypothesis that BD patients with a younger AAO show a characteristic genetic risk fingerprint.

Despite all progress in understanding the genetic risk landscape of complex disorders, it is currently unclear whether patients suffering from these disorders will directly benefit from this improved knowledge in the short term. Here, the first challenge is to translate correlated genetic risk variants into functionally relevant information on causal disease pathways and mechanisms.

Different approaches have been developed to make such connections, many integrating GWAS findings with eQTL and other *omics* data (Jacobs et al., 2020; PGC SCZ et al., 2020). We applied some of these innovative methods to functionally annotate GWAS results. For example, we used causal mediation analyses on expression and methylation data to map MS-associated variants to genes (Andlauer et al., 2016). Subsequently, we used a deep learning-based framework to identify functionally relevant units associated with MS, *i.e.*, SNPs in the context of specific regulatory chromatin motifs, cell types, and treatments (Arloth et al., 2020).

In a recent, still unpublished study, we conducted a GWAS on circulating levels of the protein α-Klotho (Ellidag et al., 2016) and linked genetically-mediated changes in Klotho levels to Crohn's disease using Mendelian randomization (Gergei *et al.*,

submitted for publication). Note that Mendelian randomization and colocalization are highly useful methods for integrating genetic with tissue-specific gene expression and proteomics data (Zheng et al., 2020).

Moreover, multifactorial disorders are shaped by an intricate interplay of genetic and environmental influences. Analyzing genetics alone will not be sufficient to unravel the etiology of complex disorders in full. Therefore, future studies will also have to address how putative environmental risk factors, e.g., low sunlight exposure and vitamin D levels in the case of MS (Olsson et al., 2016; Ostkamp et al., 2021), interact with genetic risk factors. Likely, the integration of environmental with genetic variation involves epigenetic mechanisms like DNA methylation (Andlauer et al., 2016; Kular et al., 2018).

Another challenge is that an understanding of the molecular risk factors for a disorder does not directly lead to the development of improved treatments: Patients suffer from the clinical manifestations and symptoms emerging after the underlying disorders have developed. However, genetic risk factors contributing to the development of a disease do not necessarily shape these symptoms.

For example, MS likely develops primarily in the peripheral immune system, involving, among others, genes regulating lymphocyte differentiation (IMSGC et al., 2011). Subsequently, inflammation causes demyelination and neurodegeneration in the CNS, the extent of which shapes disease severity. Additional pathways not involving MS risk factors may contribute to these processes. For example, the permeability of the blood-brain barrier may influence disease severity, as indicated by successful drugs like natalizumab or fingolimod. Moreover, processes regulating myelin repair and neuron survival affect MS disease severity and progression (Dutta and Trapp, 2011; Lubetzki et al., 2020).

These processes are presumably also affected by genetic variants that are not associated with MS susceptibility. It will thus likely be necessary to study genetic factors influencing the clinical progression and disability of CNS disorders to identify new treatment options. To this end, we started international efforts to discover genetic variants associated with MS disease severity (e.g., using the MS severity score MSSS), the frequency of MS relapses (Vandebergh et al., 2021), MRI parameters like white-matter lesion volume (Mühlau et al., 2016), and severe symptoms like suicidal ideation in BD patients. We have also characterized the genetics of other disease-associated traits like immunoglobulin antibody levels produced in the brain, which is an immunological hallmark of MS (Gasperi et al., 2020). Over the next years, such projects will hopefully improve our knowledge on how genetic factors influence the disease course.

Furthermore, many complex disorders constitute heterogeneous syndromic concepts, defined by symptoms rather than objectively quantifiable biomarkers. Thus, it is necessary to conduct both transdiagnostic studies searching for genetic factors commonly underlying related disorders (Andlauer et al., 2021; PGC-CDG et al., 2019) and to cluster heterogeneous disorders into subtypes with possibly more homogeneous molecular etiologies (Pelin et al., 2021). For example, depression is a symptom-based construct consisting of different subtypes with likely different molecular pathomechanisms (Schwabe et al., 2019). Studying these subtypes together reduces statistical power.

Moreover, the generation of ever-larger GWAS dataset comes at a price: Instead of focusing on homogeneous subtypes, we integrate heterogeneous datasets based on shallow phenotypic data, including diagnostic codes from electronic health records (EHR) and self-reports from direct-to-consumer testing companies (Cai et al., 2020; Kendall et al., 2021; Levey et al., 2020; Wray et al., 2018). Data assessed using different phenotype definitions (e.g., structured interviews, ICD codes, or self-reports) does not necessarily reflect the same, clearly circumscribed disease entity (Cai et al., 2020). Accordingly, uncertainties in disease-related phenotypes and

cross-country differences in health systems may impede large, multicentric studies. In a recent study (Kalman et al., 2021), we found that when combining the age at disease onset of BD patients across continents or phenotypic definitions (e.g., assessed using structured interviews or defined by subjective impairment), the estimated heritability approached zero. Given that the heritability is the proportion of the phenotypic variance explained by genetic variance, the heritability decreases when the phenotypic variance increases (Fischbach and Niggeschmidt, 2019). Lack of internationally standardized phenotype definitions thus impairs the chances of successful genetic discoveries for severity-associated phenotypes in ever-larger studies.

However, large biobanks and EHR datasets can also be employed to apply machine learning methods for the identification of stable patient groups that exhibit similar disease characteristics or treatment responses (Paul et al., 2019). Accordingly, the analysis of EHRs holds great promise for identifying risk factors for disease (Hapfelmeier et al., 2019).

Further heterogeneity is caused by psychiatric diagnoses having diffuse boundaries. Possibly, subtypes of related but distinct diagnoses might overlap in their molecular etiology (Andlauer et al., 2019; Coleman et al., 2019; PGC-CDG et al., 2019). For example, BD type II is genetically more closely correlated to depression than type-I BD is (Guzman-Parra et al., 2021; Stahl et al., 2019).

About 8-9% of unipolar depression patients convert to a bipolar diagnosis later in life, with a BD family history being the most important predictor of a conversion (Musliner and Østergaard, 2018). Such diagnostic uncertainties hamper the optimal and timely treatment of patients. Most psychiatric treatments have been discovered by serendipity, and their molecular mode of action is often still unclear. For example, lithium is a well-established mood stabilizer used in BD, but it shows no or only a partial effect for 60% of patients (Hou et al., 2016). The reasons for this variance in response are unclear, but genetic factors likely play an important role. In the future, genetic studies will help to identify patients showing differential disease courses and treatment responses and will drive the characterization of molecular factors responsible for these differences.

Here, animal models will also remain essential. The first identified MDD risk gene codes for the presynaptic active zone scaffolding protein Piccolo (Sullivan et al., 2009; Wray et al., 2018). We have studied Piccolo's role in synaptic plasticity in mouse models (Waites et al., 2011). Notably, even superficially simple-appearing model systems like *D. melanogaster* are very well suited for characterizing the fundamental function of synaptic proteins in learning and memory (Andlauer and Sigrist, 2012; Andlauer et al., 2014; Christiansen et al., 2011; Fulterer et al., 2018).

Although modeling psychiatric and neurological disorders in animal models is challenging, useful mouse models have been generated for studying MS (Ben-Nun et al., 2014; Glatigny and Bettelli, 2018). We have shown that such models can be used to analyze human genetic MS risk factors and, presumably, their interactions with the environment (Faber et al., 2020). Animal and *in vitro* models like induced pluripotent stem cells will remain crucial for assessing the safety of drug candidates. Future drug development pipelines will have to firmly incorporate genomic data not only to prioritize drug targets but also to identify patients likely to profit from a specific therapy or at risk for side effects and treatment failure (Andlauer et al., 2020).

Currently, one of the biggest unaddressed issues for both efficient disorder risk models and the identification of safer drug targets is the insufficient inclusion of non-European ethnicities in genetic studies (Martin et al., 2019; Wojcik et al., 2019). When including more diverse ancestries, harnessing massive datasets while harmonizing phenotype definitions, and applying the rapidly improving advanced statistical methods, the next wave of genetic discoveries may finally deliver a noticeable benefit to patients worldwide.

Es ist uns aufgetragen, am Werke zu arbeiten,
aber es ist uns nicht gegeben, es zu vollenden.

Lion Feuchtwanger, Die Geschwister Oppermann (Talmud)

References

Alinari, L., Mahoney, E., Patton, J., Zhang, X., Huynh, L., Earl, C.T., Mani, R., Mao, Y., Yu, B., Quinion, C., et al. (2011). FTY720 increases CD74 expression and sensitizes mantle cell lymphoma cells to milatuzumab-mediated cell death. Blood *118*, 6893–6903.

Andlauer, T.F.M., and Nöthen, M.M. (2020). Polygenic scores for psychiatric disease: from research tool to clinical application. Medizinische Genetik *32*, 39–45.

Andlauer, T.F.M., and Sigrist, S.J. (2012). In vivo imaging of the Drosophila larval neuromuscular junction. Cold Spring Harbor Protocols *2012*, 481–489.

Andlauer, T.F.M., Scholz-Kornehl, S., Tian, R., Kirchner, M., Babikir, H.A., Depner, H., Loll, B., Quentin, C., Gupta, V.K., Holt, M.G., et al. (2014). Drep-2 is a novel synaptic protein important for learning and memory. eLife *3*, e03895.

Andlauer, T.F.M., Buck, D., Antony, G., Bayas, A., Bechmann, L., Berthele, A., Chan, A., Gasperi, C., Gold, R., Graetz, C., et al. (2016). Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. Science Advances *2*, e1501678.

Andlauer, T.F.M., Müller-Myhsok, B., and Ripke, S. (2018). Oxford Medicine Online. Undefined.

Andlauer, T.F.M., Guzman-Parra, J., Streit, F., Strohmaier, J., González, M.J., Flores, S.G., Fabeiro, F.J.C., Noriega, F. del R., Perez, F.P., González, J.H., et al. (2019). Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. Mol Psychiatr *26*, 1286–1298.

Andlauer, T.F.M., Link, J., Martin, D., Ryner, M., Hermanrud, C., Grummel, V., Auer, M., Hegen, H., Aly, L., Gasperi, C., et al. (2020). Treatment- and population-specific genetic risk factors for anti-drug antibodies against interferon-beta: a GWAS. BMC Med *18*, 298.

Andlauer, T.F.M., Mühleisen, T.W., Hoffstaedter, F., Teumer, A., Wittfeld, K., Teuber, A., Reinbold, C.S., Grotegerd, D., Bülow, R., Caspers, S., et al. (2021). Genetic factors influencing a neurobiological substrate for psychiatric disorders. Transl Psych. *11(1)*, 192

Arloth, J., Eraslan, G., Andlauer, T.F.M., Martins, J., Iurato, S., Kühnel, B., Waldenberger, M., Frank, J., Gold, R., Hemmer, B., et al. (2020). DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. Plos Comput Biol *16*, e1007616.

Barbosa, M.D.F.S., Vielmetter, J., Chu, S., Smith, D.D., and Jacinto, J. (2006). Clinical link between MHC class II haplotype and interferon-beta (IFN-beta) immunogenicity. Clinical Immunology (Orlando, Fla.) *118*, 42–50.

Barnard, J.G., Babcock, K., and Carpenter, J.F. (2013). Characterization and quantitation of aggregates and particles in interferon-β products: potential links between product quality attributes and immunogenicity. J Pharm Sci *102*, 915–928.

Bastard, P., Rosen, L.B., Zhang, Q., Michailidis, E., Hoffmann, H.-H., Zhang, Y., Dorgham, K., Philippot, Q., Rosain, J., Béziat, V., et al. (2020). Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science *370*, eabd4585.

Beck, K., Ehmann, N., Andlauer, T.F.M., Ljaschenko, D., Strecker, K., Fischer, M., Kittel, R.J., and Raabe, T. (2015). Loss of the Coffin-Lowry syndrome-associated gene RSK2 alters ERK activity, synaptic function and axonal transport in Drosophila motoneurons. Disease Models & Mechanisms *8*, 1389–1400.

Ben-Nun, A., Kaushansky, N., Kawakami, N., Krishnamoorthy, G., Berer, K., Liblau, R., Hohlfeld, R., and Wekerle, H. (2014). From classic to spontaneous and humanized models of multiple sclerosis: Impact on understanding pathogenesis and drug development. J Autoimmun *54*, 33–50.

Berk, M., Dodd, S., Callaly, P., Berk, L., Fitzgerald, P., Castella, A.R. de, Filia, S., Filia, K., Tahtalian, S., Biffin, F., et al. (2007). History of illness prior to a diagnosis of bipolar disorder or schizoaffective disorder. J Affect Disorders *103*, 181–186.

Bertolotto, A., Deisenhammer, F., Gallo, P., and Sørensen, P.S. (2004). Immunogenicity of interferon beta: differences among products. Journal of Neurology *251 Suppl 2*, II15–II24.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.

Brainstorm, Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., et al. (2018). Analysis of shared heritability in common disorders of the brain. Science (New York, NY) *360*, eaap8757.

Buck, D., Cepok, S., Hoffmann, S., Grummel, V., Jochim, A., Berthele, A., Hartung, H.-P., Wassmuth, R., and Hemmer, B. (2011). Influence of the HLA-DRB1 genotype on antibody development to interferon beta in multiple sclerosis. Arch Neurol. *68*, 480–487.

Buck, D., Andlauer, T.F.M., Igl, W., Wicklein, E.-M., Mühlau, M., Weber, F., Köchert, K., Pohl, C., Arnason, B., Comi, G., et al. (2018). Effect of HLA-DRB1 alleles and genetic variants on the development of neutralizing antibodies to interferon beta in the BEYOND and BENEFIT trials. Mult Scler J 135245851876308.

Byun, J.-S., Lee, J.-W., Kim, S., Kwon, K., Sohn, J.-H., Lee, K., Oh, D., Kim, S.-S., Chun, W., and Lee, H. (2010). Neuroprotective effects of stanniocalcin 2 following kainic acid-induced hippocampal degeneration in ICR mice. Peptides *31*, 2094–2099.

Cai, N., Revez, J.A., Adams, M.J., Andlauer, T.F.M., Breen, G., Byrne, E.M., Clarke, T.-K., Forstner, A.J., Grabe, H.J., Hamilton, S.P., et al. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 1–11.

Callahan, S.T., Wolff, M., Hill, H.R., Edwards, K.M., Keitel, W., Atmar, R., Patel, S., Sahly, H.E., Munoz, F., Glezen, W.P., et al. (2014). Impact of Body Mass Index on Immunogenicity of Pandemic H1N1 Vaccine in Children and Adults. J Infect Dis *210*, 1270–1274.

Christiansen, F., Zube, C., Andlauer, T.F., Wichmann, C., Fouquet, W., Owald, D., Mertel, S., Leiss, F., Tavosanis, G., Luna, A.J., et al. (2011). Presynapses in Kenyon Cell Dendrites in the Mushroom Body Calyx of Drosophila. The Journal of Neuroscience *31*, 9696–9707.

Coleman, J.R.I., Gaspar, H.A., Bryois, J., Breen, G., Byrne, E.M., Forstner, A.J., Holmans, P.A., Leeuw, C.A. de, Mattheisen, M., McQuillin, A., et al. (2019). The Genetics of the Mood Disorder Spectrum: Genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. Biol Psychiat.

Corte, V., Sperduti, M., Malherbe, C., Vialatte, F., Lion, S., Gallarda, T., Oppenheim, C., and Piolino, P. (2016). Cognitive Decline and Reorganization of Functional Connectivity in Healthy Aging: The Pivotal Role of the Salience Network in the Prediction of Age and Cognitive Performances. Frontiers in Aging Neuroscience *8*, 204.

Costa, C., Martínez-Sáez, E., Gutiérrez-Franco, A., Eixarch, H., Castro, Z., Ortega-Aznar, A., Cajal, R.S. y, Montalban, X., and Espejo, C. (2015). Expression of semaphorin 3A, semaphorin 7A and their receptors in multiple sclerosis lesions. Multiple Sclerosis Journal *21*, 1632–1643.

Daltrozzo, T., Hapfelmeier, A., Donnachie, E., Schneider, A., and Hemmer, B. (2018). A Systematic Assessment of Prevalence, Incidence and Regional Distribution of Multiple Sclerosis in Bavaria From 2006 to 2015. Front Neurol *9*, 871.

Dankowski, T., Buck, D., Andlauer, T.F.M., Antony, G., Bayas, A., Bechmann, L., Berthele, A., Bettecken, T., Chan, A., Franke, A., et al. (2015). Successful Replication of GWAS Hits for Multiple Sclerosis in 10,000 Germans Using the Exome Array. Genetic Epidemiology *39*, 601–608.

Deelen, J., Beekman, M., Uh, H.-W., Broer, L., Ayers, K.L., Tan, Q., Kamatani, Y., Bennet, A.M., Tamm, R., Trompet, S., et al. (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Human Molecular Genetics *23*, 4420–4432.

Deisenhammer, F. (2014). Interferon-Beta: Neutralizing Antibodies, Binding Antibodies, Pharmacokinetics and Pharmacodynamics, and Clinical Outcomes. Journal of Interferon & Cytokine Research : The Official Journal of the International Society for Interferon and Cytokine Research *34*, 938–945.

de Jong, S., Diniz, M.J.A., Saloma, A., Gadelha, A., Santoro, M.L., Ota, V.K., Noto, C., Consortium, M.D.D. and B.D.W.G. of the P.G., Curtis, C., Newhouse, S.J., et al. (2018). Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. Commun Biology *1*, 163.

Dendrou, C.A., Petersen, J., Rossjohn, J., and Fugger, L. (2018). HLA variation and disease. Nat Rev Immunol *18*, 325–339.

Dutta, R., and Trapp, B.D. (2011). Mechanisms of neuronal dysfunction and degeneration in multiple sclerosis. Prog Neurobiol *93*, 1–12.

Eastwood, S., Law, A., Everall, I., and Harrison, P. (2003). The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology. Molecular Psychiatry *8*, 148–155.

Ellidag, H.Y., Yilmaz, N., Kurtulus, F., Aydin, O., Eren, E., Inci, A., Dolu, S., Ince, F.D.A., Giray, Ö., and Yaman, A. (2016). The Three Sisters of Fate in Multiple Sclerosis: Klotho (Clotho), Fibroblast Growth Factor-23 (Lachesis), and Vitamin D (Atropos). Ann Neurosci *23*, 155–161.

Erp, T. van, Walton, E., Hibar, D.P., Schmaal, L., Jiang, W., Glahn, D.C., Pearlson, G.D., Yao, N., Fukunaga, M., Hashimoto, R., et al. (2018). Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. Biological Psychiatry *84*, 644–654.

Esmaeeli-Nieh, S., Fenckova, M., Porter, I.M., Motazacker, M.M., Nijhof, B., Castells-Nobau, A., Asztalos, Z., Weißmann, R., Behjati, F., Tzschach, A., et al. (2016). BOD1 Is Required for Cognitive Function in Humans and Drosophila. PLoS Genetics *12*, e1006022.

Faber, H., Kurtoic, D., Krishnamoorthy, G., Weber, P., Pütz, B., Müller-Myhsok, B., Weber, F., and Andlauer, T.F.M. (2020). Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis Is Linked to Human Multiple Sclerosis Risk Genes. Front Immunol *11*, 2165.

Feinstein, A., Magalhaes, S., Richard, J.-F., Audet, B., and Moore, C. (2014). The link between multiple sclerosis and depression. Nat Rev Neurol *10*, 507–517.

Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M., et al. (2014). Frontotemporal dementia and its subtypes: a genome-wide association study. The Lancet. Neurology *13*, 686–699.

Fischbach, K.-F., and Niggeschmidt, M. (2019). Erblichkeit der Intelligenz, Eine Klarstellung aus biologischer Sicht (Springer).

Franke, B., Stein, J.L., Ripke, S., Anttila, V., Hibar, D.P., Hulzen, K.J.E. van, Arias-Vasquez, A., Smoller, J.W., Nichols, T.E., Neale, M.C., et al. (2016). Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. Nat Neurosci. *19*, 420–431.

Fullerton, J.M., Koller, D.L., Edenberg, H.J., Foroud, T., Liu, H., Glowinski, A.L., McInnis, M.G., Wilcox, H.C., Frankland, A., Roberts, G., et al. (2015). Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young At-Risk Individuals. Am J Medical Genetics Part B Neuropsychiatric Genetics *168*, 617–629.

Fulterer, A., Andlauer, T.F.M., Ender, A., Maglione, M., Eyring, K., Woitkuhn, J., Lehmann, M., Matkovic-Rachid, T., Geiger, J.R.P., Walter, A.M., et al. (2018). Active Zone Scaffold Protein Ratios Tune Functional Diversity across Brain Synapses. Cell Reports *23*, 1259–1274.

Garding, A., Bhattacharya, N., Claus, R., Ruppel, M., Tschuch, C., Filarsky, K., Idler, I., Zucknick, M., Caudron-Herger, M., Oakes, C., et al. (2013). Epigenetic upregulation of lncRNAs at 13q14.3 in leukemia is linked to the In Cis downregulation of a gene cluster that targets NF-kB. Plos Genet *9*, e1003373.

Gasperi, C., Andlauer, T.F.M., Keating, A., Knier, B., Klein, A., Pernpeintner, V., Lichtner, P., Gold, R., Zipp, F., Bergh, F.T., et al. (2020). Genetic determinants of the humoral immune response in MS. Neurology - Neuroimmunol Neuroinflammation *7*, e827.

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun *10*, 1776.

Gialluisi, A., Andlauer, T.F.M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K.U., Czamara, D., Pourcain, B.S., Brandler, W., et al. (2019). Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Transl Psychiat *9*, 77.

Gialluisi, A., Andlauer, T.F.M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K.U., Czamara, D., Pourcain, B.S., Honbolygó, F., et al. (2020). Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. Mol Psychiatr 1–14.

Glatigny, S., and Bettelli, E. (2018). Experimental Autoimmune Encephalomyelitis (EAE) as Animal Models of Multiple Sclerosis (MS). Csh Perspect Med.

Good, P.F., Alapat, D., Hsu, A., Chu, C., Perl, D., Wen, X., Burstein, D.E., and Kohtz, S.D. (2004). A role for semaphorin 3A signaling in the degeneration of hippocampal neurons during Alzheimer's disease. Journal of Neurochemistry *91*, 716–736.

Goodkind, M., Eickhoff, S.B., Oathes, D.J., Jiang, Y., Chang, A., Jones-Hagata, L.B., Ortega, B.N., Zaiko, Y.V., Roach, E.L., Korgaonkar, M.S., et al. (2015). Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry *72*, 305–315.

Groot, A.S.D., and Scott, D.W. (2007). Immunogenicity of protein therapeutics. Trends Immunol *28*, 482–490.

Guzman-Parra, J., Rivas, F., Strohmaier, J., Forstner, A., Streit, F., Auburger, G., Propping, P., Diaz, G.O., González, M.J., Gil-Flores, S., et al. (2017). The Andalusian Bipolar Family (ABiF) Study: Protocol and sample description. Revista De Psiquiatría Y Salud Ment.

Guzman-Parra, J., Streit, F., Forstner, A.J., Strohmaier, J., González, M.J., Flores, S.G., Fabeiro, F.J.C., Noriega, F. del R., Perez, F.P., González, J.H., et al. (2021). Clinical and genetic differences between bipolar disorder type 1 and 2 in multiplex families. Transl Psychiat *11*, 31.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science *369*, 718–724.

Hagenaars, S.P., Coleman, J.R.I., Choi, S.W., Gaspar, H., Adams, M.J., Howard, D.M., Hodgson, K., Traylor, M., Air, T.M., Andlauer, T.F.M., et al. (2020). Genetic comorbidity between major depression and cardio-metabolic traits, stratified by age at onset of major depression. Am J Medical Genetics Part B Neuropsychiatric Genetics *183*, 309–330.

Han, L.K.M., Dinga, R., Hahn, T., Ching, C.R.K., Eyler, L.T., Aftanas, L., Aghajani, M., Aleman, A., Baune, B.T., Berger, K., et al. (2020). Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. Mol Psychiatr 1–16.

Hapfelmeier, A., Gasperi, C., Donnachie, E., and Hemmer, B. (2019). A large case-control study on vaccination as risk factor for multiple sclerosis. Neurology *93*, e908–e916.

Harmelen, A.-L. van, Tol, M.-J. van, Wee, N.J.A. van der, Veltman, D.J., Aleman, A., Spinhoven, P., Buchem, M.A. van, Zitman, F.G., Penninx, B.W.J.H., and Elzinga, B.M. (2010). Reduced Medial Prefrontal Cortex Volume in Adults Reporting Childhood Emotional Maltreatment. Biol Psychiat *68*, 832–838.

Harroud, A., Marrie, R.A., Fitzgerald, K.C., Salter, A., Lu, Y., Patel, M., and Kowalec, K. (2021). Mendelian randomization provides no evidence for a causal role in the bidirectional relationship between depression and multiple sclerosis. Mult Scler J 135245852199307.

Hässler, S., Bachelet, D., Duhaze, J., Szely, N., Gleizes, A., Abina, S.H.-B., Aktas, O., Auer, M., Avouac, J., Birchler, M., et al. (2020). Clinicogenomic factors of biotherapy immunogenicity in autoimmune disease: A prospective multicohort study of the ABIRISK consortium. Plos Med *17*, e1003348.

Heitmann, H., Andlauer, T.F.M., Korn, T., Mühlau, M., Henningsen, P., Hemmer, B., and Ploner, M. (2020). Fatigue, depression, and pain in multiple sclerosis: How neuroinflammation translates into dysfunctional reward processing and anhedonic symptoms. Mult Scler J 135245852097227.

Hemmer, B., Stüve, O., Kieseier, B., Schellekens, H., and Hartung, H.-P. (2005). Immune response to immunotherapy: the role of neutralising antibodies to interferon beta in the treatment of multiple sclerosis. Lancet Neurology *4*, 403–412.

Hibar, D.P., Westlye, L.T., Doan, N.T., Jahanshad, N., Cheung, J.W., Ching, C.R.K., Versace, A., Bilderbeck, A.C., Uhlmann, A., Mwangi, B., et al. (2018). Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. Mol Psychiatr *23*, 932–942.

Hillert, J. (1994). Human leukocyte antigen studies in multiple sclerosis. Ann Neurol *36*, S15–S17.

Hoffmann, S., Cepok, S., Grummel, V., Lehmann-Horn, K., Hackermüller, J., Hackermueller, J., Stadler, P.F., Hartung, H.-P., Berthele, A., Deisenhammer, F., et al. (2008). HLA-DRB1*0401 and HLA-DRB1*0408 are strongly associated with the development of antibodies against interferon-beta therapy in multiple sclerosis. American Journal of Human Genetics *83*, 219–227.

Hollenbach, J.A., and Oksenberg, J.R. (2015). The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun *64*, 13–25.

Holstiege, J., Steffen, A., Goffrier, B., and Bätzing, J. (2017). Epidemiology of multiple sclerosis – a population-based, Germany-wide study. Central Research Institute of Ambulatory Health Care in Germany (Zi). Versorgungsatlas Report *17/09*.

Hou, L., Heilbronner, U., Degenhardt, F., Adli, M., Akiyama, K., Akula, N., Ardau, R., Arias, B., Backlund, L., Banzato, C.E.M., et al. (2016). Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. Lancet *387*, 1085–1093.

Housley, W.J., Fernandez, S.D., Vera, K., Murikinati, S.R., Grutzendler, J., Cuerdon, N., Glick, L., Jager, P.L.D., Mitrovic, M., Cotsapas, C., et al. (2015). Genetic variants associated with autoimmunity drive NFκB signaling and responses to inflammatory stimuli. Sci Transl Med *7*, 291ra93.

Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. *22*, 343–352.

IMSGC (2019a). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science *365*, eaav7188.

IMSGC (2019b). A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis. Nature Communications *10*, 2236.

IMSGC, WTCCC2, Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C.A., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature *476*, 214–219.

IMSGC, Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kemppinen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet *45*, 1353–1360.

IMSGC, Mitrovic, M., Patsopoulos, N.A., Beecham, A.H., Dankowski, T., Goris, A., Dubois, B., D'hooghe, M.B., Lemmens, R., Damme, P.V., et al. (2018). Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. Cell 1–17.

Jaber, A., Driebergen, R., Giovannoni, G., Schellekens, H., Simsarian, J., and Antonelli, M. (2007). The Rebif® New Formulation Story. Drugs R D *8*, 335–348.

Jacobs, B.M., Taylor, T., Awad, A., Baker, D., Giovanonni, G., Noyce, A.J., and Dobson, R. (2020). Summary-data-based Mendelian randomization prioritizes potential druggable targets for multiple sclerosis. Brain Commun *2*, fcaa119.

Jäger, A., Dardalhon, V., Sobel, R.A., Bettelli, E., and Kuchroo, V.K. (2009). Th1, Th17, and Th9 Effector Cells Induce Experimental Autoimmune Encephalomyelitis with Different Pathological Phenotypes. J Immunol *183*, 7169–7177.

Jawa, V., Cousens, L.P., Awwad, M., Wakshull, E., Kropshofer, H., and Groot, A.S.D. (2013). T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. Clin Immunol *149*, 534–555.

Jersild, C., Hansen, G., Svejgaard, A., Fog, T., Thomsen, M., and Dupont, B. (1973). Histocompatibility Determinants in Multiple Sclerosis, With Special Reference to Clinical Course. Lancet *302*, 1221–1225.

Kalman, J.L., Papiol, S., Forstner, A.J., Heilbronner, U., Degenhardt, F., Strohmaier, J., Adli, M., Adorjan, K., Akula, N., Alda, M., et al. (2018). Investigating polygenic burden in age at disease onset in bipolar disorder: Findings from an international multicentric study. Bipolar Disord *21*, 68–75.

Kalman, J.L., Loohuis, L.M.O., Vreeker, A., McQuillin, A., Stahl, E.A., Ruderfer, D., Grigoroiu-Serbanescu, M., Panagiotaropoulou, G., Ripke, S., Bigdeli, T.B., et al. (2021). Characterization of Age and Polarity at Onset in Bipolar Disorder. Medrxiv 2021.04.16.21251163.

Kappos, L., Clanet, M., Sandberg-Wollheim, M., Radue, E.W., Hartung, H.P., Hohlfeld, R., Xu, J., Bennett, D., Sandrock, A., Goelz, S., et al. (2005). Neutralizing antibodies and efficacy of interferon beta-1a: a 4-year controlled study. Neurology *65*, 40–47.

Kappos, L., Polman, C.H., Freedman, M.S., Edan, G., Hartung, H.P., Miller, D.H., Montalban, X., Barkhof, F., Bauer, L., Jakobs, P., et al. (2006). Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. Neurology *67*, 1242–1249.

Kendall, K.M., van Assche, E., Andlauer, T.F.M., Choi, K.W., Luykx, J.J., Schulte, E.C., and Lu, Y. (2021). The genetic basis of major depression. Psychological Medicine doi: 10.1017/S0033291721000441.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.-H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet *50*, 1219–1224.

Kim, J., Shin, J.-Y., Kim, J.-I., Seo, J.-S., Webster, M.J., Lee, D., and Kim, S. (2014). Somatic deletions implicated in functional diversity of brain cells of individuals with schizophrenia and unaffected controls. Scientific Reports *4*, 3807.

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., Falkai, P., Riecher-Rössler, A., Möller, H.-J., Reiser, M., et al. (2014). Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. Schizophrenia Bulletin *40*, 1140–1153.

Krishnamoorthy, G., Lassmann, H., Wekerle, H., and Holz, A. (2006). Spontaneous opticospinal encephalomyelitis in a double-transgenic mouse model of autoimmune T cell/B cell cooperation. Journal of Clinical Investigation *116*, 2385–2392.

Kular, L., Liu, Y., Ruhrmann, S., Zheleznyakova, G., Marabita, F., Gomez-Cabrero, D., James, T., Ewing, E., Lindén, M., Górnikiewicz, B., et al. (2018). DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nature Communications *9*, 2397.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet *45*, 1452–1458.

Lee, J.S., and Shin, E.-C. (2020). The type I interferon response in COVID-19: implications for treatment. Nat Rev Immunol *20*, 585–586.

Levey, D.F., Stein, M.B., Wendt, F.R., Pathak, G.A., Zhou, H., Aslan, M., Quaden, R., Harrington, K.M., Sanacora, G., McIntosh, A.M., et al. (2020). GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in More than 1.2 Million Participants Yields 178 Independent Risk Loci. Medrxiv 2020.05.18.20100685.

Levinson, D.F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N.R., and Sullivan, P.F. (2014). Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? Biological Psychiatry *76*, 510–512.

Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. Genome Med *12*, 44.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199–W205.

Lill, C.M., Luessi, F., Alcina, A., Sokolova, E.A., Ugidos, N., Hera, B. de la, Guillot-Noël, L., Malhotra, S., Reinthaler, E., Schjeide, B.-M.M., et al. (2015). Genome-wide significant association with seven novel multiple sclerosis risk loci. J Med Genet *52*, 848–855.

Link, J., Ryner, M.L., Fink, K., Hermanrud, C., Lima, I., Brynedal, B., Kockum, I., Hillert, J., and Fogdell-Hahn, A. (2014). Human leukocyte antigen genes and interferon beta preparations influence risk of developing neutralizing anti-drug antibodies in multiple sclerosis. PLoS ONE *9*, e90479.

Liu, S., Smit, D.J.A., Abdellaoui, A., Wingen, G. van, and Verweij, K.J.H. (2021). Brain structure and function show distinct relations with genetic predispositions to mental health and cognition. MedRxiv doi: https://doi.org/10.1101/2021.03.07.21252728.

Lu, A.T., Hannon, E., Levine, M.E., Crimmins, E.M., Lunnon, K., Mill, J., Geschwind, D.H., and Horvath, S. (2017). Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. Nature Communications *8*, 15353.

Lubetzki, C., Zalc, B., Williams, A., Stadelmann, C., and Stankoff, B. (2020). Remyelination in multiple sclerosis: from basic science to clinical translation. Lancet Neurology *19*, 678–688.

Lublin, F. (2005). History of modern multiple sclerosis therapy. J Neurol *252*, iii3–iii9.

Maaser, A., Forstner, A.J., Strohmaier, J., Hecker, J., Ludwig, K.U., Sivalingam, S., Streit, F., Degenhardt, F., Witt, S.H., Reinbold, C.S., et al. (2018). Exome sequencing in large, multiplex bipolar disorder families from Cuba. PLoS ONE *13*, e0205895.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics *51*, 584–591.

Mecollari, V., Nieuwenhuis, B., and Verhaagen, J. (2014). A perspective on the role of class III semaphorin signaling in central nervous system trauma. Frontiers in Cellular Neuroscience *8*, 328.

Menon, V. (2015). Salience Network (Academic Press: Elsevier).

Monk, P.D., Marsden, R.J., Tear, V.J., Brookes, J., Batten, T.N., Mankowski, M., Gabbay, F.J., Davies, D.E., Holgate, S.T., Ho, L.-P., et al. (2020). Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Respir Medicine.

Mühlau, M., Andlauer, T.F.M., and Hemmer, B. (2016). HLA Genetic Risk Burden in Multiple Sclerosis. JAMA Neurology *73*, 1500–1501.

Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2020). Genome-wide association study of over 40,000 bipolar disorder cases provides novel biological insights. MedRxiv.

Muse, E.D., Wineinger, N.E., Schrader, B., Molparia, B., Spencer, E.G., Bodian, D.L., Torkamani, A., and Topol, E.J. (2017). Moving Beyond Clinical Risk Scores with a Mobile App for the Genomic Risk of Coronary Artery Disease. Biorxiv 101519.

Musliner, K.L., and Østergaard, S.D. (2018). Patterns and predictors of conversion to bipolar disorder in 91 587 individuals diagnosed with unipolar depression. Acta Psychiat Scand *137*, 422–432.

Ni, G., Zeng, J., Revez, J.R., Wang, Y., Ge, T., Restaudi, R., Kiewa, J., Nyholt, D.R., Coleman, J.R.I., Smoller, J.W., et al. (2020). A comprehensive evaluation of polygenic score methods across cohorts in psychiatric disorders. BioRxiv.

Nieuwboer, H.A., Pool, R., Dolan, C.V., Boomsma, D.I., and Nivard, M.G. (2016). GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. American Journal of Human Genetics *99*, 917–927.

Núñez, C., Cénit, M.C., Alvarez-Lafuente, R., Río, J., Fernández-Arquero, M., Arroyo, R., Montalban, X., Fernández, O., Oliver-Martos, B., Leyva, L., et al. (2014). HLA alleles as biomarkers of high-titre neutralising antibodies to interferon-β therapy in multiple sclerosis. J Med Genet *51*, 395–400.

O'Connor, P., Filippi, M., Arnason, B., Comi, G., Cook, S., Goodin, D., Hartung, H.-P., Jeffery, D., Kappos, L., Boateng, F., et al. (2009). 250 μg or 500 μg interferon beta-1b versus 20 mg glatiramer acetate in relapsing-remitting multiple sclerosis: a prospective, randomised, multicentre study. Lancet Neurology *8*, 889–897.

Olsson, T., Barcellos, L.F., and Alfredsson, L. (2016). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol.

Ostkamp, P., Salmen, A., Pignolet, B., Görlich, D., Andlauer, T.F.M., Schulte-Mecklenbeck, A., Gonzalez-Escamilla, G., Bucciarelli, F., Gennero, I., Breuer, J., et al. (2021). Sunlight exposure exerts immunomodulatory effects to reduce multiple sclerosis severity. Proc National Acad Sci *118*, e2018457118.

Patsopoulos, N.A. (2018). Genetics of Multiple Sclerosis: An Overview and New Directions. Csh Perspect Med.

Patsopoulos, N.A., Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist, ANZgene, GeneMSA, IMSGC, Esposito, F., Reischl, J., Lehr,

S., Bauer, D., et al. (2011). Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol *70*, 897–912.

Patsopoulos, N.A., Barcellos, L.F., Hintzen, R.Q., Schaefer, C., Duijn, C.M. van, Noble, J.A., Raj, T., IMSGC, ANZgene, Gourraud, P.-A., et al. (2013). Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genetics *9*, e1003926.

Paul, R., Andlauer, T.F.M., Czamara, D., Hoehn, D., Lucae, S., Pütz, B., Lewis, C.M., Uher, R., Müller-Myhsok, B., Ising, M., et al. (2019). Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. Transl Psychiat *9*, 1–15.

Pelin, H., Ising, M., Stein, F., Meinert, S., Meller, T., Brosch, K., Winter, N.R., Krug, A., Leenings, R., Lemke, H., et al. (2021). Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning. MedRxiv 2021.02.04.21251083.

PGC-CDG, Lee, P.H., Anttila, V., Won, H., Feng, Y.-C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., et al. (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell *179*, 1469-1482.e11.

PGC-SCZ (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

PGC-SCZ, Ripke, S., Walters, J.T., and O'Donovan, M.C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. MedRxiv.

Power, R.A., Steinberg, S., Bjornsdottir, G., Rietveld, C.A., Abdellaoui, A., Nivard, M.M., Johannesson, M., Galesloot, T.E., Hottenga, J.J., Willemsen, G., et al. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci *18*, 953–955.

Power, R.A., Tansey, K.E., Buttenschøn, H.N., Cohen-Woods, S., Bigdeli, T., Hall, L.S., Kutalik, Z., Lee, S.H., Ripke, S., Steinberg, S., et al. (2017). Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Biol Psychiat *81*, 325–335.

Reus, L.M., Shen, X., Gibson, J., Wigmore, E., Ligthart, L., Adams, M.J., Davies, G., Cox, S.R., Hagenaars, S.P., Bastin, M.E., et al. (2017). Association of polygenic risk for major psychiatric illness with subcortical volumes and white matter integrity in UK Biobank. Sci Rep *7*, 42140.

Rombach-Riegraf, V., Karle, A.C., Wolf, B., Sordé, L., Koepke, S., Gottlieb, S., Krieg, J., Djidja, M.-C., Baban, A., Spindeldreher, S., et al. (2014). Aggregation of Human Recombinant Monoclonal Antibodies Influences the Capacity of Dendritic Cells to Stimulate Adaptive T-Cell Responses In Vitro. Plos One *9*, e86322.

Satizabal, C.L., Adams, H.H.H., Hibar, D.P., White, C.C., Knol, M.J., Stein, J.L., Scholz, M., Sargurupremraj, M., Jahanshad, N., Roshchupkin, G.V., et al. (2019). Genetic architecture of subcortical brain structures in 38,851 individuals. Nat Genet *51*, 1624–1636.

Sawcer, S., Franklin, R.J.M., and Ban, M. (2014). Multiple sclerosis genetics. Lancet Neurology *13*, 700–709.

Schmaal, L., Hibar, D.P., Sämann, P.G., Hall, G.B., Baune, B.T., Jahanshad, N., Cheung, J.W., Erp, T.G.M. van, Bos, D., Ikram, M.A., et al. (2016). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Molecular Psychiatry.

Schwabe, I., Milaneschi, Y., Gerring, Z., Sullivan, P.F., Schulte, E., Suppli, N.P., Thorp, J.G., Derks, E.M., and Middeldorp, C.M. (2019). Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies. Psychol Med *49*, 2646–2656.

Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., and Greicius, M.D. (2009). Neurodegenerative diseases target large-scale human brain networks. Neuron *62*, 42–52.

Smeland, O.B., Wang, Y., Frei, O., Li, W., Hibar, D.P., Franke, B., Bettella, F., Witoelar, A., Djurovic, S., Chen, C.-H., et al. (2017). Genetic Overlap Between Schizophrenia and Volumes of Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic Mechanisms. Schizophrenia Bull *44*, 854–864.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H. x000E9 lamp x000E9 na A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature Genetics *51*, 793–803.

Stefano, N.D., Curtin, F., Stubinski, B., Blevins, G., Drulovic, J., Issard, D., Shotekov, P., Gasperini, C., and investigators, I. study (2010). Rapid benefits of a new formulation of subcutaneous interferon beta-1a in relapsing—remitting multiple sclerosis. Mult Scler J *16*, 888–892.

Stickler, M., Valdes, A.M., Gebel, W., Razo, O.J., Faravashi, N., Chin, R., Rochanayon, N., and Harding, F.A. (2004). The HLA-DR2 haplotype is associated with an increased proliferative response to the immunodominant CD4+ T-cell epitope in human interferon-β. Genes Immun *5*, 1–7.

Stromnes, I.M., Cerretti, L.M., Liggitt, D., Harris, R.A., and Goverman, J.M. (2008). Differential regulation of central nervous system autoimmunity by TH1 and TH17 cells. Nat Med *14*, 337–342.

Sullivan, P.F., Geus, E.J.C. de, Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T., Arolt, V., Baune, B.T., Blackwood, D., Cichon, S., et al. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol Psychiatr *14*, 359–375.

Szatkiewicz, J., Crowley, J.J., Adolfsson, A.N., Åberg, K.A., Alaerts, M., Genovese, G., McCarroll, S., Del-Favero, J., Adolfsson, R., and Sullivan, P.F. (2019). The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. Transl Psychiat *9*, 60.

Tintore, M., Vidal-Jordana, A., and Sastre-Garriga, J. (2019). Treatment of multiple sclerosis — success from bench to bedside. Nat Rev Neurol *15*, 53–58.

Tozzi, L., Garczarek, L., Janowitz, D., Stein, D.J., Wittfeld, K., Dobrowolny, H., Lagopoulos, J., Hatton, S.N., Hickie, I.B., Carballedo, A., et al. (2020). Interactive impact of childhood maltreatment, depression, and age on cortical brain structure: mega-analytic findings from a large multi-site cohort. Psychol Med *50*, 1020–1031.

Uddin, L.Q. (2014). Salience processing and insular cortical function and dysfunction. Nat Rev Neurosci *16*, 55–61.

Vandebergh, M., Andlauer, T.F.M., Zhou, Y., Mallants, K., Held, F., Aly, L., Taylor, B.V., Hemmer, B., Dubois, B., and Goris, A. (2021). Genetic variation in WNT9B increases relapse hazard in multiple sclerosis. Ann Neurol. doi: 10.1002/ana.26061

Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., Group, Topm.A.W., Consortium, T.-O. for P.M., Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., et al. (2019). Recovery of trait heritability from whole genome sequence data. Biorxiv 588020.

Waites, C.L., Leal-Ortiz, S.A., Andlauer, T.F., Sigrist, S.J., and Garner, C.C. (2011). Piccolo Regulates the Dynamic Assembly of Presynaptic F-Actin. The Journal of Neuroscience *31*, 14250–14263.

Waubant, E., Lucas, R., Mowry, E., Graves, J., Olsson, T., Alfredsson, L., and Langer-Gould, A. (2019). Environmental and genetic risk factors for MS: an integrated review. Ann Clin Transl Neur *6*, 1905–1922.

Weber, F., Cepok, S., Wolf, C., Berthele, A., Uhr, M., Bettecken, T., Buck, D., Hartung, H.P., Holsboer, F., Müller-Myhsok, B., et al. (2012). Single-nucleotide polymorphisms in HLA- and non-HLA genes associated with the development of antibodies to interferon-β therapy in multiple sclerosis patients. The Pharmacogenomics Journal *12*, 238–245.

Werff, S.J.A. van der, Pannekoek, J.N., Veer, I.M., Tol, M.-J. van, Aleman, A., Veltman, D.J., Zitman, F.G., Rombouts, S.A.R.B., Elzinga, B.M., and Wee, N.J.A. van der (2013). Resting-state functional connectivity in adults with childhood emotional maltreatment. Psychol Med *43*, 1825–1836.

Wigmore, E., Clarke, T.-K., Howard, D., Adams, M., Hall, L., Zeng, Y., Gibson, J., Davies, G., Fernandez-Pujals, A., Thomson, P., et al. (2017). Do regional brain volumes and major depressive disorder share genetic architecture? A study of Generation Scotland (n=19 762), UK Biobank (n=24 048) and the English Longitudinal Study of Ageing (n=5766). Translational Psychiatry *7*, e1205.

Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518.

Wray, N.R., James, M.R., Mah, S.P., Nelson, M., Andrews, G., Sullivan, P.F., Montgomery, G.W., Birley, A.J., Braun, A., and Martin, N.G. (2007). Anxiety and Comorbid Measures Associated With PLXNA2. Archives of General Psychiatry *64*, 318–326.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.F.M., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature Genetics *50*, 668–681.

Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet *27*, 3641–3649.

Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P.C., Hurle, M.R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., et al. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet *52*, 1122–1131.

Zhou, Y., Simpson, S., Holloway, A.F., Charlesworth, J., Mei, I. van der, and Taylor, B.V. (2014). The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler J *20*, 135–140.

List of publications

Last author publications (including author contributions)

1. Guzman-Parra J, Streit F, Forstner AJ, Strohmaier J, González MJ, Gil Flores S, Cabaleiro Fabeiro FJ, del Río Noriega F, Perez Perez F, Haro González J, Orozco Diaz G, de Diego-Otero Y, Moreno-Kuestner B, Auburger G, Degenhardt F, Heilmann-Heimbach S, Herms S, Hoffmann P, Frank J, Foo JC, Sirignano L, Witt SH, Cichon C, Rivas F, Mayoral F, Nöthen MM, **Andlauer TFM***, Rietschel M* (*shared last*):
Clinical and genetic differences between bipolar disorder type 1 and 2 in multiplex families.
Translational Psychiatry 2021, 11:31

   ***Contributions**: Together with José Guzman-Parra, Fabian Streit, and Marcella Rietschel, I have designed the study, interpreted the results, and prepared the manuscript. I have processed the genetic data, analyzed both genetic and clinical data, and generated the data visualizations. I am the corresponding author for this article.*

2. Faber H, Kurtoic D, Krishnamoorthy G, Weber P, Pütz P, Müller-Myhsok B, Weber F*, **Andlauer TFM*** (*shared last*):
Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis Is Linked to Human Multiple Sclerosis Risk Genes.
Frontiers in Immunology 2020, 11:2165

   ***Contributions**: Together with Dunja Kurtoic and Bertram Müller-Myhsok, I have devised the statistical analyses for the study. I have supervised all and conducted selected statistical analyses, and have generated the data visualizations. Together with Hans Faber, Dunja Kurtoic, and Frank Weber I have interpreted the results. Together with Hans Faber, I have prepared the manuscript draft, which we revised primarily together with Dunja Kurtoic, Gurumoorthy Krishnamoorthy, and Frank Weber. Together with Bertram Müller-Myhsok, I am the corresponding author for this article.*

First author publications (including author contributions)

3. **Andlauer TFM***, Mühleisen T* (*shared first*), Hoffstaedter H, Teumer A, Wittfeld K, Teuber A, Reinbold CS, Grotegerd D, Bülow R, Caspers S, Dannlowski U, Herms S, Hoffmann P, Kircher T, Minnerup H, Moebus S, Nenadić I, Teismann H, Völker U, Etkin A, Berger K, Grabe HJ, Nöthen MM, Amunts K, Eickhoff SB, Sämann PG, Müller-Myhsok B, Cichon S:
Genetic factors influencing a neurobiological substrate for psychiatric disorders.
Translational Psychiatry 2021, 11(1):192

   ***Contributions**: Together with Thomas Mühleisen, Felix Hoffstaedter, Philipp Sämann, Bertram-Müller-Myhsok, and Sven Cichon, I have designed the study. I have processed all genetic data and conducted all statistical analyses except for the extraction of the MRI volumes, and have generated the data visualizations. Together with Thomas Mühleisen and Philipp Sämann, I have prepared the manuscript draft, which was revised by all authors.*

4. **Andlauer TFM**, Link J, Martin D, Ryner M, Hermanrud C, Grummel V, Auer M, Hegen H, Aly L, Gasperi C, Knier B, Müller-Myhsok B, Jensen PEH, Sellebjerg F, Kockum I, Olsson T, Pallardy M, Spindeldreher S, Deisenhammer F, Fogdell-Hahn A, Hemmer B & on behalf of the ABIRISK consortium:
Treatment- and population-specific genetic risk factors for anti-drug antibodies against interferon-beta: a GWAS.
BMC Medicine 2020, 18:298

   ***Contributions**: Together with Florian Deisenhammer, Anna Fogdell-Hahn, and Bernhard*

Hemmer, I have designed and administered the study and interpreted the results. I have processed all data, conducted all statistical analyses, and generated the data visualizations. I have prepared the manuscript draft, which was revised primarily together with Florian Deisenhammer, Anna Fogdell-Hahn, and Bernhard Hemmer, with input from all authors. Together with Bernhard Hemmer, I am the corresponding author for this article.

5. **Andlauer TFM***, Guzman-Parra J*, Streit F* *(shared first)*, Strohmaier J, González MJ, Gil Flores S, Cabaleiro Fabeiro FJ, del Río Noriega F, Perez Perez F, Haro González J, Orozco Diaz G, de Diego-Otero Y, Moreno-Kuestner B, Auburger G, Degenhardt F, Heilmann-Heimbach S, Herms S, Hoffmann P, Frank J, Foo JC, Treutlein J, Witt SH, Cichon C, Kogevinas M, Bipolar Disorder Working Group of the PGC, Major Depressive Disorder Working Group of the PGC, Rivas F, Mayoral F, Müller-Myhsok B, Forstner AJ, Nöthen MM, Rietschel M:
Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders.
Molecular Psychiatry 2019, 26:1286–1298 (2021)

   ***Contributions**: Together with José Guzman-Parra, Fabian Streit, and Marcella Rietschel, I have designed the study and interpreted the results. I have processed the genetic data, conducted all statistical analyses, and generated the data visualizations. Together with José Guzman-Parra, Fabian Streit, and Marcella Rietschel, I have prepared the manuscript, which was revised by all authors.*

6. Fulterer A*, **Andlauer TFM***, Ender A*, Maglione M* *(shared first)*, Eyring K, Woitkuhn J, Lehmann M, Matkovic-Rachid T, Geiger JRP, Walter AM, Nagel KI, Sigrist SJ:
Active Zone Scaffold Protein Ratios Tune Functional Diversity across Brain Synapses.
Cell Reports 2018, 23(5):1259-74

   ***Contributions**: Together with Andreas Fulterer, Katherine Nagel, and Stephan Sigrist, I have designed the study, interpreted the results, and prepared the manuscript draft. I have performed part of the experiments (confocal microscopy imaging), conducted statistical analyses (Figs. 1-4, 7), especially of the ratiometric data (Fig. 1-3), and generated the respective data visualizations.*

7. Buck D*, **Andlauer TFM*** *(shared first)*, Igl W, Wicklein E-M, Mühlau M, Weber F, Köchert K, Pohl C, Arnason B, Comi G, Cook S, Filippi M, Hartung H-P, Jeffery D, Kappos L, Barkhof F, Edan G, Freedman MS, Montalban X, Müller-Myhsok B, Hemmer B, The BEYOND and BENEFIT Study Groups:
Effect of HLA-DRB1 alleles and genetic variants on the development of neutralizing antibodies to interferon beta in the BEYOND and BENEFIT trials.
Multiple Sclerosis Journal 2018, 25(4):565-573

   ***Contributions**: I have processed the genetic data, conducted all statistical analyses, and generated the data visualizations. Together with Dorothea Buck, Eva-Maria Wicklein, Mark Mühlau, Karl Köchert, Christoph Pohl, Bertram Müller-Myhsok, and Bernhard Hemmer, I have interpreted the results. I have prepared the manuscript draft, which was revised by all authors.*

8. **Andlauer TFM***, Buck D* *(shared first)*, Antony G, Bayas A, Bechmann L, Berthele A, Chan A, Gasperi C, Gold R, Graetz C, Haas J, Hecker M, Infante-Duarte C, Knop M, Kümpfel T, Limmroth V, Linker RA, Loleit V, Luessi F, Meuth SG, Mühlau M, Nischwitz S, Paul F, Pütz M, Ruck T, Salmen A, Stangel M, Stellmann JP, Stürner KH, Tackenberg B, Then Bergh F, Tumani H, Warnke C, Weber F, Wiendl H, Wildemann B, Zettl UK, Ziemann U, Zipp F, Arloth J, Weber P, Radivojkov-Blagojevic M, Scheinhardt MO, Dankowski T, Bettecken T, Lichtner P, Czamara D, Carrillo-Roa T, Binder EB, Berger K, Bertram L, Franke A, Gieger C, Herms S, Homuth G, Ising M, Jöckel KH, Kacprowski T, Kloiber S, Laudes M, Lieb W, Lill CM, Lucae S, Meitinger T, Moebus S, Müller-Nurasyid M, Nöthen MM, Petersmann A, Rawal R, Schminke U, Strauch K, Völzke H, Waldenberger M, Wellmann J, Porcu E, Mulas A, Pitzalis M, Sidore C, Zara I, Cucca F,

Zoledziewska M, Ziegler A, Hemmer B, Müller-Myhsok B:
Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation.
Science Advances 2016, 2(6) e1501678

***Contributions**: Together with Dorothea Buck, Andreas Ziegler Bernhard Hemmer, and Bertram MülIer-Myhsok, I have designed the study and interpreted the results. I have processed all data, conducted all statistical analyses, and generated the data visualizations. I have prepared the manuscript draft, which was revised by all authors.*

9. **Andlauer TFM**, Scholz-Kornehl S, Tian R, Kirchner M, Babikir HA, Depner H, Loll B, Quentin C, Gupta VK, Holt MG, Dipt S, Cressy M, Wahl MC, Fiala A, Selbach M, Schwärzel M, Sigrist SJ:
Drep-2 is a novel synaptic protein important for learning and memory.
eLife 2014, 3.e03895

***Contributions**: Together with Matthew Holt, Markus Wahl, André Fiala, Matthias Selbach, Martin Schwärzel, and Stephan Sigrist, I have designed the study and interpreted the results. I have performed part of the experiments (especially confocal microscopy imaging, Figs. 2, 4, 6), conducted statistical analyses, and generated the data visualizations. Together with Stephan Sigrist, I have prepared the manuscript draft, which was revised by all authors.*

Co-author publications

10. Zeng L, Moser S, Mirza-Schreiber N, Lamina C, Coassin S, Nelson CP, Annilo T, Franzén O, Kleber ME, Mack S, **Andlauer TFM**, Jiang B, Stiller B, Li L, Willenborg C, Munz M, Kessler T, Kastrati A, Laugwitz KL, Erdmann J, Moebus S, Nöthen MM, Peters A, Strauch K, Müller-Nurasyid M, Gieger C, Meitinger T, Steinhagen-Thiessen E, März W, Metspalu A, Björkegren JLM, Samani NJ, Kronenberg F, Müller-Myhsok B, Schunkert H:
Cis-epistasis at the LPA locus and risk of cardiovascular diseases.
Cardiovascular Research doi:10.1093/cvr/cvab136

11. Vandebergh M, **Andlauer TFM**, Zhou Y, Mallants K, Held F, Aly L, Taylor BV, Hemmer B, Dubois B, Goris A:
Genetic variation in WNT9B increases relapse hazard in multiple sclerosis.
Annals of Neurology doi:10.1002/ana.26061

12. Heilbronner U, Papiol S, Budde M, **Andlauer TFM**, Strohmaier J, Streit F, Frank J, Degenhardt F, Heilmann-Heimbach S, Witt SH, Forstner AJ, Loerbroks A, Amelang M, Stürmer T, Müller-Myhsok B, Nöthen MM, Rietschel M, Schulze TG:
"The Heidelberg Five" personality dimensions: Genome-wide associations, polygenic risk for neuroticism, and psychopathology 20 years after assessment.
Am. J. Medical Genetics Part B: Neuropsychiatric Genetics doi:10.1002/ajmg.b.32837

13. Ostkamp P, Salmen A, Pignolet B, Görlich D, **Andlauer TFM**, Schulte-Mecklenbeck A, Gonzalez-Escamilla G, Bucciarelli F, Gennero I, Breuer J, Antony G, Schneider-Hohendorf T, Mykicki N, Bayas A, Then Bergh F, Bittner S, Hartung HP, Friese MA, Linker RA, Luessi F, Lehmann-Horn K, Mühlau M, Paul F, Stangel M, Tackenberg B, Tumani H, Warnke C, Weber F, Wildemann B, Zettl UK, Ziemann U, Müller-Myhsok B, Kümpfel T, Klotz L, Meuth SG, Zipp F, Hemmer B, Hohlfeld R, Brassat D, Gold R, Gross CC, Lukas C, Groppa S, Loser K, Wiendl H, Schwab N; German Competence Network Multiple Sclerosis (KKNMS) and the BIONAT Network:
Sunlight exposure exerts immunomodulatory effects to reduce multiple sclerosis severity.
PNAS 2021, 118(1):e2018457118

14. Gialluisi A, **Andlauer TFM**, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, Ludwig KU, Czamara D, St Pourcain B, Honbolygó F, Tóth D, Csépe V, Huguet G, Chaix Y, Iannuzzi S, Demonet J, Morris AP, Hulslander J, Willcutt EJ, DeFries JC, Olson RK, Smith SD, Pennington BF, Vaessen A, Maurer U, Lyytinen H, Peyrard-Janvid M, Leppänen PHT, Brandeis D, Bonte M, Stein JF, Talcott JB, Fauchereau F, Wilcke A, Kirsten H, Müller B, Francks C, Bourgeron T, Monaco AP, Ramus F, Landerl K, Kere J, Scerri TS, Paracchini S, Fisher SE, Schumacher J, Nöthen MM, Müller-Myhsok B & Schulte-Körne B:
Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia.
Molecular Psychiatry 2020, doi:10.1038/s41380-020-00898-x

15. Goltermann J, Redlich R, Grotegerd D, Dohm K, Leehr EJ, Böhnlein J, Förster K, Meinert S, Enneking V, Richter M, Repple J, DeVillers I, Kloecker M, Jansen A, Krug A, Nenadić I, Brosch K, Meller T, Stein F, Schmitt S, Rietschel M, Streit F, Witt SH, Forstner AJ, Nöthen MM, Baune BT, **Andlauer TFM**, Kircher T, Opel N, Dannlowski U:
Childhood maltreatment and cognitive functioning: the role of depression, parental education, and polygenic predisposition.
Neuropsychopharmacology 2020, doi:10.1038/s41386-020-00794-6

16. Nenadić I, Meller T, Schmitt S, Stein F, Brosch K, Mosebach J, Ettinger U, Grant P, Meinert S, Opel N, Lemke H, Fingas S, Förster K, Hahn T, Jansen A, **Andlauer TFM**, Forstner AJ, Heilmann-Heimbach S, Hall ASM, Awasthi S, Ripke S, Witt SH, Rietschel M, Müller-Myhsok B, Nöthen MM, Dannlowski U, Krug A, Streit F, Kircher T:
Polygenic risk for schizophrenia and schizotypal traits in non-clinical subjects.
Psychological Medicine 2020, doi:10.1017/S0033291720002822

17. Gasperi C, **Andlauer TFM**, Keating A, Knier B, Klein A, Pernpeintner V, Lichtner P, Gold R, Zipp F, Then Bergh F, Stangel M, Tumani H, Wildemann B, Wiendl H, Bayas A, Kümpfel T, Zettl UK, Linker RA, Ziemann U, Knop M, Warnke C, Friese MA, Paul F, Tackenberg B, Berthele A, Hemmer B:
Genetic determinants of the humoral immune response in MS.
Neurology Neuroimmunology & Neuroinflammation 2020, 7(5):e827

18. Hagenaars SP, Coleman JRI, Choi SW, Gaspar H, Adams MJ, Howard DM, Hodgson K, Traylor M, Air TM, **Andlauer TFM**, Arolt V, Baune BT, Binder EB, …, …, McIntosh AM, Deary IJ, Breen G, Lewis CM (*author list abbreviated*):
Genetic comorbidity between major depression and cardio-metabolic traits, stratified by age at onset of major depression.
American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 2020, doi:10.1002/ajmg.b.32807

19. Aly L, Havla J, Lepennetier G, **Andlauer TFM**, Sie C, Strauß EM, Hoshi MM, Kümpfel T, Hiltensperger M, Mitsdoerffer M, Mühlau M, Zimmer C, Hemmer B, Korn T, Knier B:
Inner retinal layer thinning in radiologically isolated syndrome predicts conversion to multiple sclerosis.
European Journal of Neurology 2020, doi:10.1111/ene.14416

20. Krug A, Wöhr M, Seffer D, Rippberger H, Sungur AÖ, Dietsche B, Stein F, Sivalingam S, Forstner AJ, Witt SH, Dukal H, Streit F, Maaser A, Heilmann-Heimbach S, **Andlauer TFM**, Herms S, Hoffmann P, Rietschel M, Nöthen MM, Lackinger M, Schratt G, Koch M, Schwarting RKW, Kircher T:

Advanced paternal age as a risk factor for neurodevelopmental disorders: a transl. study.
Molecular Autism 2020, 11(1):54

21. Cai N, Revez JA, Adams MJ, **Andlauer TFM**, Breen G, Byrne EM, Clarke TK, Forstner AJ, Grabe HJ, Hamilton SP, Levinson DF, Lewis CM, Lewis G, Martin NG, Milaneschi Y, Mors O, Müller-Myhsok B, Penninx BWJH, Perlis RH, Pistis G, Potash JB, Preisig M, Shi J, Smoller JW, Streit F, Tiemeier H, Uher R, Van der Auwera S, Viktorin A, Weissman MM, MDD Working Group of the PGC, Kendler KS, Flint J:
Minimal phenotyping yields genome-wide association signals of low specificity for major depression.
Nature Genetics 2020, 52(4):437-447

22. Dwyer DB, Kalman JL, Budde M, Kambeitz J, Ruef A, Antonucci LA, Kambeitz-Ilankovic L, Hasan A, Kondofersky I, Anderson-Schmidt H, Gade K, Reich-Erkelenz D, Adorjan K, Senner F, Schaupp S, **Andlauer TFM**, Comes AL, Schulte EC, …, …, Papiol S, Heilbronner U, Falkai P, Schulze TG, Koutsouleris N (*author list abbreviated*):
An Investigation of Psychosis Subgroups With Prognostic Validation and Exploration of Genetic Underpinnings: The PsyCourse Study.
JAMA Psychiatry 2020, 77(5):523-533

23. Comes AL, Czamara D, Adorjan K, Anderson-Schmidt H, **Andlauer TFM**, Budde M, Gade K, Hake M, Kalman JL, Papiol S, Reich-Erkelenz D, Klöhn-Saghatolislam F, Schaupp SK, Schulte EC, Senner F, Juckel G, Schmauß M, Zimmermann J, Reimer J, Reininghaus E, Anghelescu IG, Konrad C, Thiel A, Figge C, von Hagen M, Koller M, Dietrich DE, Stierl S, Scherk H, Witt SH, Sivalingam S, Degenhardt F, Forstner AJ, Rietschel M, Nöthen MM, Wiltfang J, Falkai P, Schulze TG, Heilbronner U:
The Role of Environmental Stress and DNA Methylation in the Longitudinal Course of Bipolar Disorder
International Journal of Bipolar Disorders 2020, 8(1):9

24. Arloth J, Eraslan G, **Andlauer TFM**, Martins J, Iurato S, Kühnel B, Waldenberger M, Frank J, Gold R, Hemmer B, Luessi F, Nischwitz S, Paul F, Wiendl H, Gieger C, Heilmann-Heimbach S, Kacprowski T, Laudes M, Meitinger T, Peters A, Rawal S, Strauch K, Lucae S, Müller-Myhsok B, Rietschel M, Theis FJ, Binder EB, Mueller NS:
DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning.
PLOS Computational Biology 2020, doi:10.1371/journal.pcbi.1007616

25. *Cross-Disorder Group of the Psychiatric Genomics Consortium:*
Lee PH, Anttila V, Won H, Feng YA, Rosenthal J, Zhu Z, Tucker-Drob EM, Nivard MG, Grotzinger AD, Posthuma D, Wang MM, Yu D, Stahl EA, …, …, **Andlauer TFM**, …, …, Smoller JW (*author list abbreviated*):
Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders.
Cell 2019, 179(7):P1469-1482.e11

26. *International Multiple Sclerosis Genetics Consortium:*
Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, Beecham AH, James T, Replogle J, Vlachos IS, McCabe C, Pers TH, Brandes A, White C, Keenan B, Cimpean M, Winn P, Panteliadis IP, Robbins A, **Andlauer TFM**, …, …, De Jager PL (*author list abbreviated*):
Multiple sclerosis genomic map implicates peripheral immune cells and microglia in

susceptibility.
Science 2019, 365(6460):eaav7188

27. Adorjan K, Mekonnen Z, Tessema F, Ayana M, Degenhardt F, Hoffmann P, Fricker N, Widmann M, Riedke H, Toennes S, Soboka M, Suleman S, **Andlauer TFM**, Tesfaye M, Rietschel M, Susser E, Odenwald M, Schulze TG, Mattheisen M:
Genotype-phenotype feasibility studies on khat abuse, traumatic experiences and psychosis in Ethiopia.
Psychiatric Genetics 2019, 30(1):34-38

28. Comes AL, Senner F, Budde M, Adorjan K, Anderson-Schmidt H, **Andlauer TFM**, Gade K, Hake M, Heilbronner U, Kalman JL, Reich-Erkelenz D, Klöhn-Saghatolislam F, Schaupp SK, Schulte EC, Juckel G, Dannlowski U, Schmauß M, Zimmermann J, Reimer J, Reininghaus E, Anghelescu IG, Arolt V, Baune BT, Konrad C, Thiel A, Fallgatter AJ, Nieratschker V, Figge C, von Hagen M, Koller M, Becker T, Wigand ME, Jäger M, Dietrich DE, Stierl S, Scherk H, Spitzer C, Folkerts H, Witt SH, Degenhardt F, Forstner AJ, Rietschel M, Nöthen MM, Wiltfang J, Falkai P, Schulze TG, Papiol S:
The genetic relationship between educational attainment and cognitive performance in major psychiatric disorders.
Translational Psychiatry 2019, 9(1):210

29. Glanville KP, Coleman JRI, Hanscombe KB, Euesden J, Choi SW, Purves KL, Breen G, Air TM, **Andlauer TFM**, Baune BT, Binder EB, Blackwood DHR, Boomsma DI, Buttenschøn HN, Colodro-Conde L, Dannlowski U, …, …, Major Depressive Disorder Working Group of the PGC, O'Reilly PF, Lewis CM (*author list abbreviated*):
Classical HLA alleles and C4 haplotypes are not significantly associated with depression.
Biological Psychiatry 2019, 87(5):419–430

30. Paul R, **Andlauer TFM**, Czamara D, Hoehn D, Lucae S, Pütz B, Lewis CM, Uher R, Müller-Myhsok B, Ising M, Sämann PG:
Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models.
Translational Psychiatry 2019, 9:187

31. van der Lee SJ, Conway OJ, Jansen I, Carrasquillo MM, Kleineidam L, van den Akker E, Hernández I, van Eijk KR, Stringa N, Chen JA, Zettergren A, **Andlauer TFM**, …, …, Ramirez A, Ruiz A, Slagboom E, van der Flier WM, Holstege H (*author list abbreviated*):
A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity.
Acta Neuropathologica 2019, 138(2):237-250

32. *International Multiple Sclerosis Genetics Consortium:*
Madireddy L, Patsopoulos NA, Cotsapas C, Bos SD, Beecham A, McCauley J, Kim K, Jia X, Santaniello A, Caillier SJ, **Andlauer TFM**, Barcellos LF, Berge T, Bernardinelli L, Martinelli-Boneschi F, Booth DR, Briggs F, Celius EG, Comabella M, Comi G, Cree BAC, …, …, De Jager P, Sawcer S, Oksenberg JR, Baranzini SE (*author list abbreviated*):
A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis.
Nature Communications 2019, 10:2236

33. Meller T, Schmitt S, Stein F, Brosch K, Mosebach J, Yüksel D, Zaremba D, Grotegerd D, Dohm K, Meinert S, Förster K, Redlich R, Opel N, Repple J, Hahn T, Jansen A, **Andlauer**

TFM, Forstner AJ, Heilmann-Heimbach S, Streit F, Witt SH, Rietschel M, Müller-Myhsok B, Nöthen MM, Dannlowski U, Krug A, Kircher T, Nenadić I:
Associations of schizophrenia risk genes ZNF804A and CACNA1C with schizotypy and modulation of attention in healthy subjects.
Schizophrenia Research 2019, 208:67-75

34. Gialluisi A, **Andlauer TFM**, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, Ludwig KU, Czamara D, St Pourcain B, Brandler W, Honbolygó F, Tóth D, Csépe V, Huguet G, Morris AP, Hulslander J, Willcutt EG, DeFries JC, Olson RK, Smith SD, Pennington BF, Vaessen A, Maurer U, Lyytinen H, Peyrard-Janvid M, Leppänen PHT, Brandeis D, Bonte M, Stein JF, Talcott JB, Fauchereau F, Wilcke A, Francks C, Bourgeron T, Monaco AP, Ramus F, Landerl K, Kere J, Scerri TS, Paracchini S, Fisher SE, Schumacher J, Nöthen MM, Müller-Myhsok B, Schulte-Körne G:
Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia.
Translational Psychiatry 2019, 9(1):77

35. Anderson-Schmidt H, Gade K, Malzahn D, Papiol S, Budde M, Heilbronner U, Reich-Erkelenz D, Adorjan K, Kalman JL, Senner F, Comes AL, Flatau L, Gryaznova A, Hake M, Reitt M, Schmauß M, Juckel G, Reimer J, Zimmermann J, Figge C, Reininghaus E, Anghelescu IG, Konrad C, Thiel A, von Hagen M, Koller M, Stierl S, Scherk H, Spitzer C, Folkerts H, Becker T, Dietrich DE, **Andlauer TFM**, Degenhardt F, Nöthen MM, Witt SH, Rietschel M, Wiltfang J, Falkai P, Schulze TG:
The influence of religious activity and polygenic schizophrenia risk on religious delusions in schizophrenia.
Schizophrenia Research 2019, 210, 255-261

36. *International Multiple Sclerosis Genetics Consortium:*
Mitrovič M, Patsopoulos NA, Beecham AH, Dankowski T, Goris A, Dubois B, D'hooghe MB, Lemmens R, Van Damme P, Søndergaard HB, Sellebjerg F, Sorensen PS, Ullum H, Thørner LW, Werge T, Saarela J, Cournu-Rebeix I, Damotte V, Fontaine B, Guillot-Noel L, Lathrop M, Vukusik S, Gourraud PA, **Andlauer TFM**, Pongratz V, Buck D, Gasperi C, Bayas A, Heesen C, Kümpfel T, Linker R, Paul F, Stangel M, …, …, Sawcer SJ, Oksenberg JR, De Jager PL, Kockum I, Hafler DA, Cotsapas C (*author list abbreviated*):
Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk
Cell 2018, 175(6):1679-87

37. Maaser A, Forstner AJ, Strohmaier J, Hecker J, Ludwig KU, Sivalingam S, Streit F, Degenhardt F, Witt SH, Reinbold CS, Koller AC, Raff R, Heilmann-Heimbach S, Fischer SB, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Herms S, Hoffmann P, Thiele H, Nürnberg P, Löhlein Fier H, Orozco-Díaz G, Carmenate-Naranjo D, Proenza-Barzaga N, Auburger GWJ, **Andlauer TFM**, Cichon S, Marcheco-Teruel B, Mors O, Rietschel M, Nöthen MM:
Exome sequencing in large, multiplex bipolar disorder families from Cuba.
PLOS ONE 2018, 13(10):e0205895

38. Budde M, Anderson-Schmidt H, Gade K, Reich-Erkelenz D, Adorjan K, Kalman JL, Senner F, Papiol S, **Andlauer TFM**, Comes AL, Schulte EC, …, …, Rietschel M, Nöthen MM, Falkai P, Schulze TG, Heilbronner U (*author list abbreviated*):
A longitudinal approach to biological psychiatric research: The PsyCourse study.

American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 2018, 180(2):89-102.

39. Kalman JL, Papiol S, Forstner AJ, Heilbronner U, Degenhardt F, Strohmaier J, Adli M, Adorjan K, Akula N, Alda M, Anderson-Schmidt H, **Andlauer TFM**, Anghelescu IG, Ardau R, Arias B, Arolt V, Aubry JM, Backlund L, Bartholdi K, Bauer M, Baune BT, Becker T, Bellivier F, Benabarre A, Bengesser S, Bhattacharjee AK, Biernacka JM, Birner A, …, …, Nöthen M, Rietschel M, Schulze TG (*author list abbreviated*):
Investigating polygenic burden in age at disease onset in bipolar disorder: Findings from an international multicentric study.
Bipolar Disorders 2018, 21(1):68-75

40. *Brainstorm Consortium:*
Anttila V, Bulik-Sullivan B, Finucane HK, …, …, **Andlauer TFM**, …, …, Grigoroiu-Serbanescu M, Edenberg HJ, Murray R (*author list abbreviated*):
Analysis of shared heritability in common disorders of the brain
Science 2018, 360(6395):eaap8757

41. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, James T, Ewing E, Lindén M, Górnikiewicz B, Aeinehband S, Stridh P, Link J, **Andlauer TFM**, Gasperi C, Wiendl H, Zipp F, Gold R, Tackenberg B, Weber F, Hemmer B, Strauch K, …, …, Kockum I, Feinberg AP, Jagodic M (*author list abbreviated*):
*DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis.*
Nature Communications 2018, 9(1):2397

42. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, **Andlauer TFM**, Bacanu SA, Bækvad-Hansen M, Beekman AFT, Bigdeli TB, Binder EB, Blackwood DRH, Bryois J, Buttenschøn HN, …, …, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (*author list abbreviated*):
Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression
Nature Genetics 2018, 50(5):668-681

43. Bigdeli TB, Ripke S, Peterson RE, Trzaskowski M, Bacanu SA, Abdellaoui A, **Andlauer TFM**, Beekman AT, Berger K, Blackwood DH, Boomsma DI, Breen G, Buttenschøn HN, Byrne EM, Cichon S, Clarke TK, Couvy-Duchesne B, Craddock N, …, …, Levinson DF, Lewis CM, Wray NR, Flint J, Sullivan PF, Kendler KS (*author list abbreviated*):
Genetic effects influencing risk for major depressive disorder in China and Europe
Translational Psychiatry 2017, 7(3):e1074

44. Gupta VK, Pech U, Bhukel A, Fulterer A, Ender A, Mauermann SF, **Andlauer TFM**, Antwi-Adjei E, Beuschel C, Thriene K, Maglione M, Quentin C, Bushow R, Schwärzel M, Mielke T, Madeo F, Dengjel J, Fiala A, Sigrist SJ:
Spermidine Suppresses Age-Associated Memory Impairment by Preventing Adverse Increase of Presynaptic Active Zone Size and Release
PLOS Biology 2016, 14(9):e1002563

45. Schirmer L, Worthington V, Solloch U, Loleit V, Grummel V, Lakdawala N, Grant D, Wassmuth R, Schmidt AH, Gebhardt F, **Andlauer TFM**, Sauter J, Berthele A, Lunn MP, Hemmer B:

*Higher frequencies of HLA DQB1*05:01 and anti-glycosphingolipid antibodies in a cluster of severe Guillain-Barré syndrome*
Journal of Neurology 2016, 263(10):2105-13

46. Dankowski T, Buck D, **Andlauer TFM**, Antony G, Bayas A, Bechmann L, Berthele A, Bettecken T, Chan A, Franke A, Gold R, Graetz C, Haas J, Hecker M, Herms S, Infante-Duarte C, Jöckel KH, Kieseier BC, Knier B, Knop M, …, …, Hemmer B, Ziegler A; German Competence Network for Multiple Sclerosis (KKNMS) (*author list abbreviated*):
Successful Replication of GWAS Hits for Multiple Sclerosis in 10,000 Germans Using the Exome Array
Genetic Epidemiology 2015, 39(8):601-8

47. Beck K, Ehmann N, **Andlauer TFM**, Ljaschenko D, Strecker K, Fischer M, Kittel RJ, Raabe T:
Loss of the Coffin-Lowry syndrome-associated gene RSK2 alters ERK activity, synaptic function and axonal transport in Drosophila motoneurons
Disease Models & Mechanisms 2015, 8(11):1389-400

48. Siebert M, Böhme MA, Driller JH, Babikir H, Mampell MM, Rey U, Ramesh N, Matkovic T, Holton N, Reddy-Alla S, Göttfert F, Kamin D, Quentin C, Klinedinst S, **Andlauer TFM**, Hell SW, Collins CA, Wahl MC, Loll B, Sigrist SJ:
A high affinity RIM-binding protein/Aplip1 interaction prevents the formation of ectopic axonal active zones
eLife 2015, 4.e06935

49. Nischwitz S, Wolf C, **Andlauer TFM**, Czamara D, Zettl UK, Rieckmann P, Buck D, Ising M, Bettecken T, Mueller-Myhsok B, Weber F:
MS susceptibility is not affected by single nucleotide polymorphisms in the MMP9 gene.
Journal of Neuroimmunology 2015, 279:46-9

50. Park B, Alves CH, Lundvig DM, Tanimoto N, Beck SC, Huber G, Richard F, Klooster J, **Andlauer TFM**, Swindell EC, Jamrich M, Le Bivic A, Seeliger MW, Wijnholds J:
PALS1 is essential for retinal pigment epithelium structure and neural retina stratification.
Journal of Neuroscience 2011, 31(47):17230-41

51. Waites CL, Leal-Ortiz SA, **Andlauer TFM**, Sigrist SJ, Garner CC:
Piccolo regulates the dynamic assembly of presynaptic f-actin.
Journal of Neuroscience 2011, 31(40):14250-63

52. Christiansen F, Zube C, **Andlauer TFM**, Wichmann C, Fouquet W, Owald D, Mertel S, Leiss F, Tavosanis G, Farca Luna AJ, Fiala A, Sigrist SJ:
Presynapses in Kenyon Cell Dendrites in the Mushroom Body Calyx of Drosophila.
Journal of Neuroscience 2011, 31(26):9696-9707

53. Kremer MC, Christiansen F, Leiss F, Paehler M, Knapek S, **Andlauer TFM**, Förstner F, Kloppenburg P, Sigrist SJ, Tavosanis G:
Structural Long-Term Changes at Mushroom Body Input Synapses.
Current Biology 2010, 20(21):1938-44

Peer-reviewed reviews

54. Kendall K, Assche EV, Andlauer TFM, Choi K, Luykx J, Schulte E, Lu Y:
The Genetic Basis of Major Depression.
Psychological Medicine 2021, doi: 10.1017/S0033291721000441

55. Heitmann H, **Andlauer TFM**, Korn T, Mühlau M, Henningsen P, Hemmer B, Ploner M:
Fatigue, depression, and pain in multiple sclerosis: How neuroinflammation translates into dysfunctional reward processing and anhedonic symptoms.
Multiple Sclerosis Journal 2020, doi:10.1177/1352458520972279

56. **Andlauer TFM**, Nöthen MM:
Polygenic scores for psychiatric disease: from research tool to clinical application.
medizinische genetik 2020, 32(1) 39-45

Non peer-reviewed reviews, book chapters, methods articles, and letters to the editor

57. **Andlauer TFM**, Müller-Myhsok B, Ripke S:
Statistical genetics: Genome-wide studies.
Chapter 4 in: **Psychiatric Genetics – A Primer for Clinical and Basic Scientists**
Editors: T. Schulze and F. McMahon
Oxford University Press 2018, doi:10.1093/med/9780190221973.003.0004

58. Mühlau M, **Andlauer TFM**, Hemmer B:
HLA Genetic Risk Burden in Multiple Sclerosis
JAMA Neurology 2016, 73(12):1500-1501

59. **Andlauer TFM**, Sigrist SJ:
In vivo imaging of Drosophila larval neuromuscular junctions to study synapse assembly.
Cold Spring Harbor Protocols 2012, 4:407-13

60. **Andlauer TFM**, Sigrist SJ:
Building an imaging chamber for in vivo imaging of Drosophila larvae.
Cold Spring Harbor Protocols 2012, 4:476-80

61. **Andlauer TFM**, Sigrist SJ:
In vivo imaging of the Drosophila larval neuromuscular junction.
Cold Spring Harbor Protocols 2012, 4:481-9

62. **Andlauer TFM**, Sigrist SJ:
Quantitative analysis of Drosophila larval neuromuscular junction morphology.
Cold Spring Harbor Protocols 2012, 4:490-3

63. Sigrist SJ, **Andlauer TFM**:
Fighting the famine with an amine: synaptic strategies for smart search.
Nature Neuroscience 2011, 14(2):124-6

64. **Andlauer TFM**, Sigrist SJ:
In Vivo Imaging of Drosophila Larval Neuromuscular Junctions to Study Synapse Assembly.
Chapter 21 in: **Drosophila Neurobiology: A Laboratory Manual**
Editors: B. Zhang, M.R. Freeman, and S. Waddell
Cold Spring Harbor Press 2010, ISBN 978-0-879699-05-5

65. Fischbach K, Linneweber GA, **Andlauer TFM**, Hertenstein A, Bonengel B, Chaudhary K:
The irre cell recognition module (IRM) proteins.
Journal of Neurogenetics 2009, 21(1-2):48-67

Curriculum vitae

PROFESSIONAL RESEARCH EXPERIENCE

Statistical Genetics and Drug Discovery 03/2021–
Principal Scientist for Human Genetics
Boehringer Ingelheim Pharma, Biberach an der Riß
Global Computational Biology & Digital Sciences

Statistical Genetics and Medical Informatics 01/2019–02/2021
Project leader for Genetics of Neurological Disorders
Klinikum rechts der Isar of the Technical University of Munich
Department of Neurology (Prof. Dr. Bernhard Hemmer)

Statistical Genetics 04/2014–12/2018
Max Planck Institute of Psychiatry, Munich
Department for Translational Research in Psychiatry
Group of Prof. Dr. Bertram Müller-Myhsok

Neurogenetics 10/2007–02/2014
Rudolf-Virchow-Center for Experimental Biomedicine, Würzburg and Freie Universität Berlin, *Institute for Biology / Genetics*
Lab of Prof. Dr. Stephan Sigrist

Neuromedical Genetics 07/2007–09/2007
Netherlands Institute for Neuroscience, Amsterdam
Lab of Dr. Jan Wijnholds

EDUCATIONAL RESEARCH EXPERIENCE

Neurogenetics 04/2006–04/2007
Albert Ludwigs University Freiburg i. Br., *Institute for Biology III*
Lab of Prof. Dr. Karl-Friedrich Fischbach

Neurophysiology 03/2004–06/2004
Karolinska Institute, Stockholm, *Institute for Neurophysiology*
Lab of Prof. Dr. Sten Grillner

Bacterial Genetics 01/2004–02/2004
University of Stockholm, *Department of Molecular Biosciences*
Lab of Prof. Dr. Elisabeth Haggård

Computational Biology 05/2003–07/2003
Albert Ludwigs University Freiburg i. Br., *Institute for Biology II*
Lab of Prof. Dr. Wolfgang Haehnel

EDUCATION

Doctoral Thesis in Biomedicine — 12/2013
Dr. rer. nat., grade: 1.0 (*summa cum laude*)
Julius Maximilians University Würzburg

Diploma in Biology — 04/2007
Diplom-Biologe, grade: 1.1 (*very good*)
Albert Ludwigs University Freiburg i. Br.
Major subject: *Genetics and Molecular Biology*
Minor subjects: *Neurobiology, Bioinformatics, Computer Science*

Erasmus Exchange — 08/2003–07/2004
Stockholm University

Abitur (A-level equivalent) — 06/1999
Goethe-Gymnasium Emmendingen, grade: 1.3 (*very good*)
Major subjects: *Mathematics, English*

MISCELLANEOUS EXPERIENCE

Head of the biobank laboratory — 11/2019–12/2020
Department of Neurology, Klinikum rechts der Isar, Munich

Certificate for university education (Bavaria) — 02/2021
Vertiefungsstufe, Technical University of Munich

Certificate for university education (Bavaria) — 06/2018
Aufbaustufe, Technical University of Munich

Certificate for medical didactics (Bavaria) — 08/2017
Grundstufe, Technical University of Munich

Mifek Kirschner Award — 07/2017
Max Planck Institute of Psychiatry, Munich

Lecturer — 05/2016–
Technical University of Munich, School of Medicine

PostDoc representative — 03/2016–09/2017
Max Planck Institute of Psychiatry, Munich

OpenAccess ambassador — 11/2014–08/2018
Max Planck Institute of Psychiatry, Munich

Teaching assistant at the summer course — 07/2008, 07/2009, and 07/2010
Neurobiology of Drosophila (Cold Spring Harbor Laboratory, NY)

Teaching assistant for imaging at the summer course — 07/2011
Neurobiology (Marine Biological Laboratory, Woods Hole, MA)

Elected member of the General students' committee (AStA) — 10/2006–06/2007
Albert Ludwigs University Freiburg im Breisgau

System administrator for students' computers — 04/2001–08/2003, 11/2004–06/2007
Faculty of Biology, Albert Ludwigs University Freiburg im Breisgau

Acknowledgments

This work would not have been possible without Maren's constant support, affection, and stimuli in all matters of life. This work is also your work and, to nail it down in Franz Schubert's words, dir "ist ja ohnehin alles gewidmet". Thank you for your patience over the last years during the many evenings and weekends you had to spend by yourself!
After already highlighting Scampolo's unique role in my doctoral thesis's acknowledgments, I had planned to put more emphasis on Puma this time. However, Puma was so courageous to explore new horizons already ahead of us and thus did not have the time to stay with us until I typed these letters. Therefore, it has once more been Scampi who pressed the keys when I was not looking and who jumped into video conferences and online presentations to steal the show.
Next to Puma, Dave Featherstone and Johannes Korthaus have been two inspiring individuals who have passed away while I was working on the projects presented here and whom I would otherwise have liked to send a copy of this booklet.
I am grateful to Cemsel Bafligil, Christiane Gasperi, and Janos Kalman for a critical reading of and helpful comments regarding this text. I thank Dominik Grotegerd for generating the images used for Figure 16A. Thanks to Christoph Fehre for his help with printing this booklet.
At the Max Planck Institute of Psychiatry, I would like to thank most and foremost Bertram Müller-Myhsok for his trust, guidance, and support over the years. Without these, at least my professional life would have taken a different turn.
I am also very grateful to my colleagues Benno Pütz, Nazanin Mirza-Schreiber, Alessandro Gialluisi, Christiane Wolf, Helena Pelin, Dunja Kurtoic, and Riya Paul, among others, for the joint projects, mutual learning, and friendship.
Outside the Statistical Genetics group, I would like to thank especially Darina Czamara, Alec Dick, Nadine Provençal, Janine Arloth, Philipp Sämann, Marcus Ising, Hans Faber, Sandra Nischwitz, Matthias Knop, Frank Weber, Alon Chen, and Elisabeth Binder for their support, guidance, and/or companionship.
At the Klinikum rechts der Isar, I am very grateful to Bernhard Hemmer for entrusting me with many stimulating responsibilities, always providing valuable advice, and for giving me the freedom to prioritize my many scientific projects flexibly. I would also like to thank Christiane Gasperi for the outstanding collaborations and joint passion for chocolate, as well as Mark Mühlau, Klaus Lehmann-Horn, Friederike Held, Henrik Heitmann, Matthias Bussas, Julia Hartberger, Isabell Cordts, and all lab members and students for the exciting joint projects, helpful discussions, and shared coffees.
I would also like to thank my colleagues at LMU for the excellent collaborations, especially Janos Kalman, Sergi Papiol, Urs Heilbronner, and Thomas Schulze (thank you for your political commitment around the world and for bringing me along to Ethiopia!).
Furthermore, I would like to thank Marcella Rietschel for our inspiring joint projects and for allowing me to travel to Jerusalem. Here, she stands for many motivating and instructive individuals that I had the pleasure to collaborate with – in the Psychiatric Genomics Consortium, the International MS Genetics Consortium, MultipleMS, ABIRISK, FOR2107, and others.
I thank Barbara Schormair for the great teamwork in teaching the course Genomic Medicine, Alexander Hapfelmeier for his support, and Thilo Rühle for our refreshing exchanges and adventures.
Moreover, I am very grateful to the entire company of the Bayerische Staatsoper at the Nationaltheater München, who carried me through my quotidian tasks in about 100 performances – before the pandemic abruptly separated us when there were still so many productions left to see.

Aber ich wünsche mir:
einmal, wenn die Stunde kommt
und es notwendig sein wird,
mich auch losmachen
und springen zu können,
bloß nicht zurück ins Geringere,
sondern vorwärts und ins Höhere.

Herrmann Hesse, Das Glasperlenspiel

NEUROLOGICAL DISORDERS

Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation

Till F. M. Andlauer,[1,2]* Dorothea Buck,[3]* Gisela Antony,[4] Antonios Bayas,[5] Lukas Bechmann,[6,7] Achim Berthele,[3] Andrew Chan,[8,56] Christiane Gasperi,[3] Ralf Gold,[8] Christiane Graetz,[9] Jürgen Haas,[10] Michael Hecker,[11] Carmen Infante-Duarte,[12] Matthias Knop,[1] Tania Kümpfel,[13] Volker Limmroth,[14] Ralf A. Linker,[15] Verena Loleit,[3] Felix Luessi,[9] Sven G. Meuth,[16] Mark Mühlau,[2,3] Sandra Nischwitz,[1] Friedemann Paul,[12] Michael Pütz,[17] Tobias Ruck,[16] Anke Salmen,[8,56] Martin Stangel,[18] Jan-Patrick Stellmann,[19] Klarissa H. Stürner,[19] Björn Tackenberg,[17] Florian Then Bergh,[20] Hayrettin Tumani,[21,22] Clemens Warnke,[23] Frank Weber,[1,24] Heinz Wiendl,[16] Brigitte Wildemann,[10] Uwe K. Zettl,[11] Ulf Ziemann,[25] Frauke Zipp,[9] Janine Arloth,[1,26] Peter Weber,[1] Milena Radivojkov-Blagojevic,[27] Markus O. Scheinhardt,[28] Theresa Dankowski,[28] Thomas Bettecken,[1] Peter Lichtner,[27,29] Darina Czamara,[1] Tania Carrillo-Roa,[1] Elisabeth B. Binder,[1,30] Klaus Berger,[31] Lars Bertram,[32,33] Andre Franke,[34] Christian Gieger,[35,36] Stefan Herms,[37,38] Georg Homuth,[39] Marcus Ising,[1] Karl-Heinz Jöckel,[40] Tim Kacprowski,[39] Stefan Kloiber,[1] Matthias Laudes,[41] Wolfgang Lieb,[42] Christina M. Lill,[9,32] Susanne Lucae,[1] Thomas Meitinger,[27,29] Susanne Moebus,[40] Martina Müller-Nurasyid,[43,44,45] Markus M. Nöthen,[37] Astrid Petersmann,[46] Rajesh Rawal,[35,36] Ulf Schminke,[47] Konstantin Strauch,[43,48] Henry Völzke,[49] Melanie Waldenberger,[35,36] Jürgen Wellmann,[31] Eleonora Porcu,[50] Antonella Mulas,[50,51] Maristella Pitzalis,[50] Carlo Sidore,[50] Ilenia Zara,[52] Francesco Cucca,[50,51] Magdalena Zoledziewska,[50,51] Andreas Ziegler,[28,53,54] Bernhard Hemmer,[2,3]*† Bertram Müller-Myhsok[1,2,55]*†

We conducted a genome-wide association study (GWAS) on multiple sclerosis (MS) susceptibility in German cohorts with 4888 cases and 10,395 controls. In addition to associations within the major histocompatibility complex (MHC) region, 15 non-MHC loci reached genome-wide significance. Four of these loci are novel MS susceptibility loci. They map to the genes *L3MBTL3*, *MAZ*, *ERG*, and *SHMT1*. The lead variant at *SHMT1* was replicated in an independent Sardinian cohort. Products of the genes *L3MBTL3*, *MAZ*, and *ERG* play important roles in immune cell regulation. *SHMT1* encodes a serine hydroxymethyltransferase catalyzing the transfer of a carbon unit to the folate cycle. This reaction is required for regulation of methylation homeostasis, which is important for establishment and maintenance of epigenetic signatures. Our GWAS approach in a defined population with limited genetic substructure detected associations not found in larger, more heterogeneous cohorts, thus providing new clues regarding MS pathogenesis.

INTRODUCTION

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system. Human leukocyte antigen (*HLA*) alleles, located within the major histocompatibility complex (MHC) region, have been identified as major genetic determinants for the disease (*1*, *2*). In addition, more than 100 non-MHC MS susceptibility variants have been described (*3*, *4*). Many of the genes carrying known susceptibility variants are involved in the regulation of either immune cell differentiation or signaling (*4*–*8*). However, because the heritability of MS is limited (*9*), environmental contributions to disease etiology are also important (*10*). Environmental influences can alter gene expression via epigenetic mechanisms (*11*). Epigenetic alterations, such as DNA methylation or histone modifications, have been observed in tissues and cells of MS patients (*8*, *12*–*14*). Nevertheless, the impact of epigenetic regulation in MS is not yet understood.The known genetic variants outside the MHC region have predominantly been established in large international collaborative studies. To achieve large sample sizes with the power to detect associations, these studies have combined sample sets from diverse ethnic populations (*4*–*6*). So far, the variants affecting MS susceptibility identified in these studies account for only 25% of disease heritability under an additive model of heritability (*3*), warranting for additional studies to fully unravel the genetic contribution to disease susceptibility. In contrast to the previously investigated large international cohorts, we have strived to examine the genetic contribution to MS susceptibility in a more homogeneous population, focusing entirely on German cases and controls. The genetic substructure among Germans is low (*15*). We therefore expected to have sufficient power to detect novel associations with moderate effect sizes in a data set showing little population stratification. Using a total of 4888 cases and 10,395 controls, we had 80% power to detect genome-wide significant associations with an odds ratio (OR) of 1.2 involving common variants with a minor allele frequency (MAF) of 21%. For rare single-nucleotide polymorphisms (SNPs) (MAF, 1%), the power surpassed 80% for an OR of 1.9.

RESULTS

Genome-wide association analyses

We recruited patients with either MS or clinically isolated syndrome (CIS) from MS centers throughout Germany and combined them with controls from several German population–based cohorts (Table 1). After quality control (QC), the data set DE1 consisted of 3934 cases and 8455 controls (control/case ratio, 2.15; table S1). We also compiled a second data set, called DE2, based on an independent group of

German cases previously used in the IMSGC/WTCCC2 (International Multiple Sclerosis Genetics Consortium/Wellcome Trust Case Control Consortium 2) MS study (Table 1) (*5*), and additional German controls, mostly from population-based cohorts. The data set DE2 contained 954 cases and 1940 controls after QC (control/case ratio, 2.03; table S2). We observed only moderate population substructure within these data sets (figs. S1 and S2), confirming previous genetic analyses of the German population (*15*).

Both data sets were imputed separately to the 1000 Genomes Phase 1 reference panel using SHAPEIT2 and IMPUTE2 (*16*–*18*). The resulting data sets contained more than 8 million high-quality variants with MAFs of at least 1% each. We separately conducted genome-wide association studies (GWAS) on both data sets using sex and the first eight multidimensional scaling (MDS) components of the genetic similarity matrix (GSM) as covariates, to control for any remaining population substructure. After assuring that the median genomic inflation of the two GWAS was in the expected range (table S3), results were combined using a fixed-effects pooled analysis. In this pooled analysis, the genomic inflation $\lambda_{1000,1000}$ outside the extended MHC region was 1.017 (table S3) (*19*).

Associations within the MHC region

The variant showing the strongest association in the pooled analysis of DE1 and DE2, rs3104373 [OR, 2.90; confidence interval (CI), 2.72 to 3.09; $P = 1.3 \times 10^{-234}$], lies within the MHC region between the genes *HLA-DRB1* and *HLA-DQA1*. This SNP is in strong linkage disequilibrium (LD) with the *HLA* allele *DRB1*15:01* ($r^2 = 0.99$) and thus corresponds to the established major MS risk locus (*1*, *2*). To confirm this finding, we imputed classical *HLA* alleles from our genotyping data (*20*). After QC, we obtained high-quality imputed alleles for a total of 3966 cases and 8329 controls from DE1 and DE2 (median accuracy, 96.1%; median call rate, 97.4%). Using stepwise conditional logistic regression (*1*, *2*, *5*), seven *HLA* alleles (Table 2) reached genome-wide significance (that is, $P < 5 \times 10^{-8}$). As expected, the most significantly associated allele was *DRB1*15:01* (OR, 2.85; CI, 2.66 to 3.06; $P = 1.0 \times 10^{-191}$). All seven alleles have been described as associated with MS in a recent detailed analysis of the MHC region (*2*).

Previous analyses of the MHC region have also identified associations between *HLA* alleles and age at onset of the disease, mainly with *DRB1*15:01* (*2*, *5*). We confirmed this finding in a subset of patients from our data set DE1. Age at onset was known for 1519 patients; for 1196 of them, imputed *HLA* alleles were available. Because the age at

Table 1. Clinical characteristics of German MS cases. PPMS, primary progressive MS (as opposed to bout-onset MS).

	Cohort DE1	Cohort DE2
Number of cases	3934	954
Age [mean (range)]	39 (13–79)	40 (17–82)
Female [*n* (%)]	2723 (69.2)	695 (72.9)
Male [*n* (%)]	1211 (30.8)	259 (27.1)
PPMS [*n* (%)]	105 (2.7)	63 (6.6)

[1]Max Planck Institute of Psychiatry, 80804 Munich, Germany. [2]Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany. [3]Department of Neurology, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany. [4]Central Information Office KKNMS, Philipps University Marburg, 35043 Marburg, Germany. [5]Department of Neurology, Klinikum Augsburg, 86156 Augsburg, Germany. [6]Department of Neurology, University of Leipzig, 04103 Leipzig, Germany. [7]Institute of Medical Microbiology, Otto-von-Guericke University, 39120 Magdeburg, Germany. [8]Department of Neurology, St. Josef Hospital, Ruhr-University Bochum, 44791 Bochum, Germany. [9]Department of Neurology, Focus Program Translational Neurosciences (FTN) and Research Center for Immunotherapy (FZI), Rhine-Main Neuroscience Network (rmn^2), University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany. [10]Department of Neurology, University Hospital Heidelberg, 69120 Heidelberg, Germany. [11]Department of Neurology, University of Rostock, 18147 Rostock, Germany. [12]NeuroCure Clinical Research Center, Department of Neurology, and Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine, and Charité University Medicine Berlin, 10117 Berlin, Germany. [13]Institute of Clinical Neuroimmunology, Ludwigs-Maximilians-Universität, 81377 Munich, Germany. [14]Department of Neurology, Hospital Köln-Merheim, 51109 Köln, Germany. [15]Department of Neurology, University Hospital Erlangen, 91054 Erlangen, Germany. [16]Department of Neurology, Klinik für Allgemeine Neurologie, University of Münster, 48149 Münster, Germany. [17]Clinical Neuroimmunology Group, Department of Neurology, Philipps-University of Marburg, 35043 Marburg, Germany. [18]Department of Neurology, Hannover Medical School, 30625 Hannover, Germany. [19]Institute of Neuroimmunology and Multiple Sclerosis and Department of Neurology, University Medical Centre Hamburg-Eppendorf, 20251 Hamburg, Germany. [20]Department of Neurology and Translational Center for Regenerative Medicine, University of Leipzig, 04103 Leipzig, Germany. [21]Department of Neurology, University of Ulm, 89081 Ulm, Germany. [22]Neurological Clinic Dietenbronn, 88477 Schwendi, Germany. [23]Department of Neurology, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. [24]Neurological Clinic, Medical Park, 65520 Bad Camberg, Germany. [25]Department of Neurology and Stroke, and Hertie Institute for Clinical Brain Research, Eberhard-Karls-Universität Tübingen, 72076 Tübingen, Germany. [26]Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [27]Institute of Human Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [28]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, 23562 Lübeck, Germany. [29]Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany. [30]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA 30329, USA. [31]Institut für Epidemiologie und Sozialmedizin der Universität Münster, 48149 Münster, Germany. [32]Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Integrative and Experimental Genomics, University of Lübeck, 23562 Lübeck, Germany. [33]School of Public Health, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK. [34]Institute of Clinical Molecular Biology, Kiel University, 24105 Kiel, Germany. [35]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [36]Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [37]Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany. [38]Department of Biomedicine, Division of Medical Genetics, University of Basel, 4031 Basel, Switzerland. [39]Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University and University Medicine Greifswald, 17475 Greifswald, Germany. [40]Institute of Medical Informatics, Biometry, and Epidemiology, University Hospital Essen, University Duisburg-Essen, 45122 Essen, Germany. [41]Department I of Internal Medicine, Kiel University, 24105 Kiel, Germany. [42]Institute of Epidemiology and Biobank popgen, Kiel University, 24105 Kiel, Germany. [43]Institute of Genetic Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [44]Department of Medicine I, Ludwig-Maximilians-Universität, 81377 Munich, Germany. [45]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, 80802 Munich, Germany. [46]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany. [47]Department of Neurology, University Medicine Greifswald, 17475 Greifswald, Germany. [48]Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany. [49]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany. [50]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, 09042 Cagliari, Italy. [51]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, 07100 Sassari, Italy. [52]Center for Advanced Studies, Research and Development in Sardinia (CRS4), Pula, 09010 Cagliari, Italy. [53]Zentrum für Klinische Studien, Universität zu Lübeck, 23562 Lübeck, Germany. [54]School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, Scottsville 3209, South Africa. [55]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK. [56]Department of Neurology, University Hospital Bern and University of Bern, 3010 Bern, Switzerland.
*These authors contributed equally to this work.
†Corresponding author. Email: hemmer@tum.de (B.H.); bmm@psych.mpg.de (B.M.-M.)

Table 2. Genome-wide significant *HLA* alleles. Alleles are in order of stepwise logistic regression. For each row, alleles from the rows above have been used as covariates in the model. AF (allele frequency of controls in %) is calculated from a joint set of DE1 and DE2. ORs and *P* values are from a fixed-effects pooled analysis of DE1 and DE2.

HLA allele	AF	OR (95% CI)	*P*	*HLA* alleles in LD ($r^2 > 0.9$)
DRB1*15:01	14.8	2.85 (2.66–3.06)	1.03×10^{-191}	DQB1*06:02
A*02:01	28.6	0.68 (0.64–0.73)	3.68×10^{-29}	
B*38:01	2.0	0.36 (0.27–0.49)	2.09×10^{-11}	
DRB1*13:03	1.5	1.96 (1.60–2.40)	6.42×10^{-11}	
DPB1*03:01	10.3	1.33 (1.22–1.46)	4.35×10^{-10}	
DRB1*03:01	12.2	1.29 (1.18–1.40)	1.85×10^{-8}	DQA1*05:01, DQB1*02:01
DRB1*08:01	3.0	1.63 (1.39–1.91)	2.36×10^{-9}	DQA1*04:01, DQB1*04:02

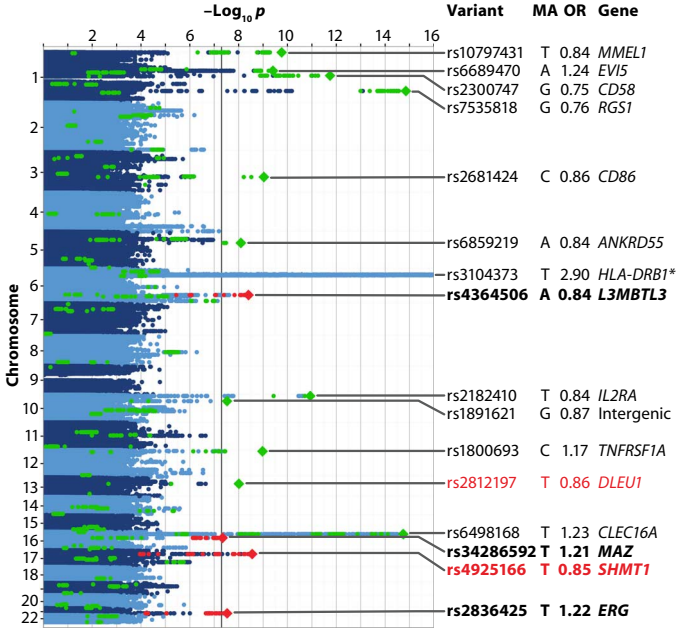


Fig. 1. Genome-wide representation of MS associations in the pooled analysis of German data sets. Manhattan plot showing strength of evidence for association (*P* value). Each variant is shown as a dot, with alternating shades of blue according to chromosome. Green dots represent established MS-associated variants and their proxies, as listed by Sawcer *et al.* (*3*) (except for rs2812197, which was not covered by that review). Top variants at the 15 non-MHC loci associated at the genome-wide significance threshold in our study are shown as diamonds. Novel variants showing genome-wide significance are plotted as red diamonds; their names are shown in bold font. Variants in high LD ($r^2 \geq 0.7$) with these novel variants are shown as red dots. Variants replicating in the Sardinian cohort are shown in red font. MA, minor allele. The OR is relative to the MA. Gene names for known loci are indicated as listed by Sawcer *et al.* (*3*). The plot is truncated at $-\log_{10}p = 16$ for better visibility; all truncated variants map to the MHC region. The lowest *P* value (rs3104373, *) was 1.3×10^{-234}.

onset was not normally distributed, rank-based inverse normal transformation was applied. The *HLA* allele most strongly associated with transformed age at onset was *DRB1*15:01* (effect size, −0.21; $P = 7.6 \times 10^{-6}$). When conducting a genome-wide analysis of transformed age at onset in all 1519 patients, no variant passed the threshold for genome-wide significance (fig. S3, A and B). The most strongly associated SNP was rs4959027 (effect size, −0.20; $P = 1.5 \times 10^{-7}$; fig. S3, C and D), which is in LD with *DRB1*15:01* ($r^2 = 0.72$). After conditioning for *DRB1*15:01* in the subset of cases with both age at onset and imputed *HLA* alleles available, the *P* value of rs4959027 was increased from 1.1×10^{-6} to 4.8×10^{-2}. We conclude that our findings for the MHC region are very well in line with previous studies and concentrated further analyses on associations with case/control status outside this region.

Associations outside the MHC region

Variants at 15 loci outside the MHC region showed genome-wide significance (Fig. 1, figs. S4 and S5, Table 3, and table S4). Ten of these loci have already been established in previous large MS GWAS (*3*, *4*, *6*). One more locus, *DLEU1* (*deleted in lymphocytic leukemia 1*), was only recently confirmed to be associated with MS in a candidate gene study (*21*). The remaining four signals are thus novel candidates for MS susceptibility loci. The lead variants at all 15 non-MHC loci showed $P < 5 \times 10^{-6}$ in DE1 and lower $P < 5 \times 10^{-8}$ in the pooled analysis of DE1 and DE2 and have thus replicated in DE2. We could not detect any significant interaction among the 15 top non-MHC variants or between them and SNP rs3104373 within the MHC region.

For validation of our findings, we compared our results to the largest study on MS genetic susceptibility published to date (Fig. 2) (*4*). Of the 108 non-MHC variants showing genome-wide significant or suggestive associations with MS in the published study, 104 variants were present in our data and could be analyzed. All of them showed the same direction of effect ($P = 5 \times 10^{-32}$, binomial sign test; CI, 0.97 to 1.00), 84 with nominal ($P < 0.05$) and 10 with genome-wide significance ($P < 5 \times 10^{-8}$). Fifty-eight of the variants had lower ORs and 35 had higher ORs in our data than in the published data set (*4*). It was expected to observe more signals with lower ORs than previously reported due to regression toward the mean.

Table 3. Genome-wide significant loci outside the MHC region and the top variant within the MHC region. Bold font in the left half of the table indicates novel loci, whereas bold font in the right half indicates variants that replicated in Sardinians. All *P* values shown are two-sided. Gene names of known loci are as listed by Sawcer *et al.* (*3*). C, chromosome. For additional details, see table S4.

Variant	C	MA	Gene	MAF DE	OR (CI) DE1 + DE2	*P* DE1 + DE2	*P* Sardinia	OR (CI) DE + Sardinia	*P* DE + Sardinia
rs10797431	1	T	*MMEL1*	34.1	0.84 (0.80–0.89)	1.81×10^{-10}			
rs6689470	1	A	*EVI5*	14.2	1.24 (1.16–1.33)	3.93×10^{-10}			
rs2300747	1	G	*CD58*	12.4	0.75 (0.69–0.81)	1.74×10^{-12}			
rs7535818	1	G	*RGS1*	19.2	0.76 (0.71–0.82)	1.51×10^{-15}			
rs2681424	3	C	*CD86*	49.7	0.86 (0.82–0.90)	9.51×10^{-10}			
rs6859219	5	A	*ANKRD55*	22.2	0.84 (0.79–0.89)	8.06×10^{-9}			
rs3104373	6	T	*HLA-DRB1*	13.6	2.90 (2.72–3.09)	1.34×10^{-234}			
rs4364506	**6**	**A**	***L3MBTL3***	**26.4**	**0.84 (0.80–0.89)**	$\mathbf{4.06 \times 10^{-9}}$	0.83	0.89 (0.85–0.93)	1.99×10^{-6}
rs2182410	10	T	*IL2RA*	38.1	0.84 (0.79–0.88)	1.15×10^{-11}			
rs1891621	10	G	Intergenic	46.7	0.87 (0.83–0.91)	2.94×10^{-8}			
rs1800693	12	C	*TNFRSF1A*	42.1	1.17 (1.11–1.23)	1.06×10^{-9}			
rs2812197	13	T	*DLEU1*	38.4	0.86 (0.82–0.91)	9.95×10^{-9}	$\mathbf{6.86 \times 10^{-3}}$	**0.87 (0.83–0.91)**	$\mathbf{2.83 \times 10^{-10}}$
rs6498168	16	T	*CLEC16A*	35.5	1.23 (1.17–1.29)	1.98×10^{-15}			
rs34286592	**16**	**T**	***MAZ***	**14.2**	**1.21 (1.13–1.30)**	$\mathbf{4.58 \times 10^{-8}}$	0.44	1.16 (1.09–1.23)	4.79×10^{-7}
rs4925166	**17**	**T**	***SHMT1***	**34.5**	**0.85 (0.81–0.90)**	$\mathbf{2.69 \times 10^{-9}}$	$\mathbf{5.63 \times 10^{-4}}$	**0.86 (0.82–0.90)**	$\mathbf{7.40 \times 10^{-12}}$
rs2836425	**21**	**T**	***ERG***	**12.7**	**1.22 (1.14–1.31)**	$\mathbf{2.84 \times 10^{-8}}$	0.35	1.18 (1.11–1.25)	1.54×10^{-7}

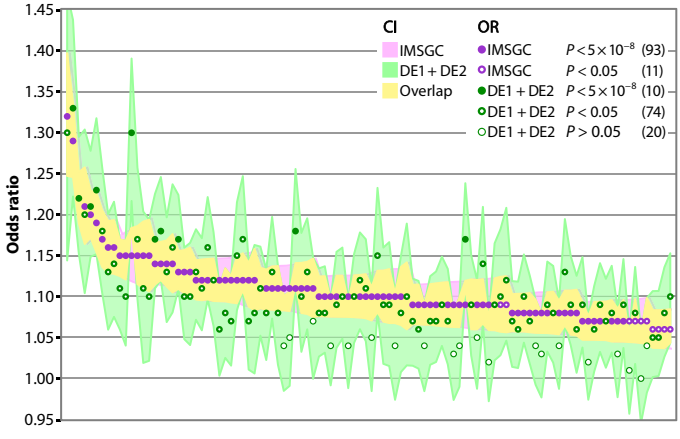


Fig. 2. Comparison of results from the pooled analysis of Germans to associations found in an IMSGC study. One hundred and four of the 108 variants showing genome-wide significant or suggestive associations with MS in the study published by the IMSGC in 2013 (*4*) were present in the pooled results of DE1 and DE2. All 104 variants showed the same direction of effect ($P = 5 \times 10^{-32}$, binomial sign test). Fifty-eight variants had lower ORs and 35 higher ORs compared to the published data set. *P* value–based categories labeled with different dots represent exclusive bins that add up to 104.

Next, we examined the four novel loci and *DLEU1* not found at genome-wide significance in a GWAS before in more detail. We investigated whether the five lead variants at these loci are significantly associated with MS in our German cohort only or whether they replicate in Sardinians, a genetically distinct population with low genetic heterogeneity. This independent Sardinian cohort consisted of 2903 cases (69.2% female, 1.2% PPMS) and 3323 controls (control/case ratio, 1.15) (*22*–*24*). Two of the variants (rs2812197 and rs4925166) replicated with $P < 0.01$ in the Sardinian data set; two more (rs34286592 and rs2836425) showed the same direction of effect but did not reach nominal significance (Table 3, table S4, and fig. S6).

SHMT1 as a novel MS susceptibility gene

The association of rs4925166 constituted the strongest signal among the novel variants. It showed an OR of 0.85 (CI, 0.81 to 0.90) and a *P* value of 2.7×10^{-9} in the pooled analysis of German data sets (Table 3). This variant replicated in the Sardinian cohort with a joint *P* value of 7.4×10^{-12} (fig. S6D). SNP rs4925166 is located on chromosome 17 in an intron of the gene *TOP3A*, coding for the DNA topoisomerase IIIα. However, strongly associated SNPs in this genomic region spread over several neighboring genes (Fig. 3A). We therefore conducted an expression quantitative trait locus (eQTL) analysis using a subset of 242 patients from data set DE1 to functionally link variants to nearby genes. We examined transcripts within a cis window of 1 million base pairs upstream and downstream of the lead variant for an association of blood gene expression levels with allele configuration (table S5). The variant rs4925166 and proxy SNPs ($r^2 > 0.7$) were found to be part of

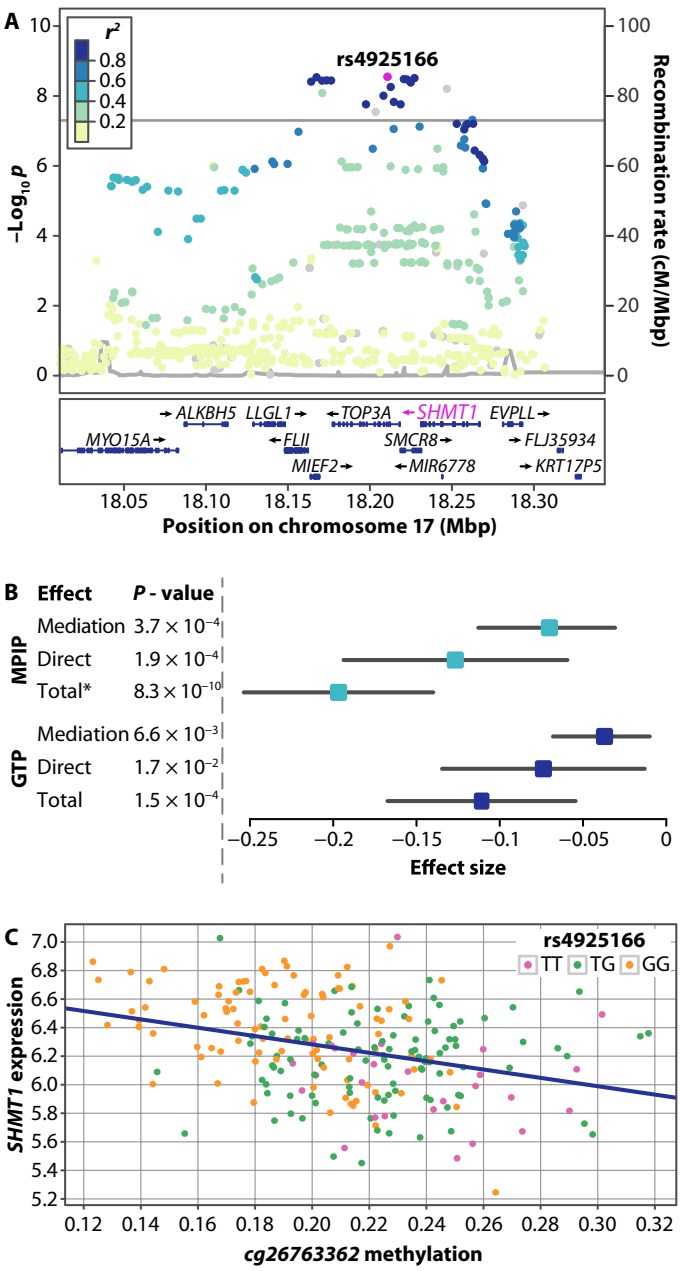


Fig. 3. Fine-mapping analysis results of locus rs4925166. (**A**) Regional plot for the rs4925166/*SHMT1* locus. Color of dots indicates LD with the lead variant (rs4925166; pink). Gray dots represent signals with missing r^2 values. cM, centimorgan. (**B**) Mediation analysis results in MPIP/GTP controls. Mediation effect: rs4925166→CpG cg26763362→*SHMT1* expression. Direct effect: rs4925166→*SHMT1* expression. Data have been calculated using the R package mediation (*30*), except for total effect (*), which was calculated by linear regression. Results were obtained using 1 million simulations. Effects and *P* values shown here differ from Table 5, as a lower number of samples contained both expression and methylation data than expression data alone. (**C**) Relationship between *cg26763362* methylation, *SHMT1* expression, and rs4925166 genotype in MPIP controls.

a strong eQTL with the gene *SHMT1* in DE1 samples [false discovery rate (FDR), 2.99×10^{-10}; Table 4 and fig. S7, A to C]. This eQTL was replicated in two independent control data sets [Max Planck Institute of Psychiatry (MPIP) data (*25*) and Grady Trauma Project (GTP) (*26*–*28*)] and in the publicly available GTEx eQTL database (*29*) (Table 4).

To investigate how rs4925166 influences the expression of *SHMT1*, we conducted an association analysis of the SNP with DNA methylation levels in blood. DNA methylation is an important epigenetic mechanism for regulation of gene expression. We tested the association between rs4925166 and DNA methylation levels at CpG sites in the two non-MS data sets MPIP and GTP. Methylation levels at 157 CpG sites that mapped to *SHMT1* were examined for an association with genotype. We observed eight significant (FDR <0.05) methylation QTLs (mQTLs) between rs4925166 and CpGs in *SHMT1* within the MPIP data set. Three of these associations could be replicated in the GTP data set (table S6 and fig. S7, D and E).

We wondered whether the CpG site showing the strongest association with rs4925166 (cg26763362) could fully explain the observed association between the SNP and *SHMT1* expression (causal direction: rs4925166→cg26763362→*SHMT1* expression) using mediation analysis (Table 4, tables S7 and S8, Fig. 3) (*30*). We observed partial mediation of the effect of rs4925166 on *SHMT1* expression by DNA methylation status of CpG site cg26763362. The association pattern indicates that an additional factor influences the relationship between the SNP, the CpG, and the gene expression (see the Supplementary Materials). Thus, we conclude that the genotype of rs4925166 affects the expression of *SHMT1* in a complex fashion, partially involving rs4925166-dependent DNA methylation.

Additional novel candidate loci associated with MS

Three loci showed genome-wide significance in the pooled analysis of German data sets DE1 and DE2 but not in Sardinians (Table 3). The strongest association, SNP rs4364506, was found on chromosome 6 and is located in an intron of the gene coding for the transcriptional regulator L3MBTL3 [Lethal(3)malignant brain tumor–like protein 3; fig. S5G). SNP rs2836425 on chromosome 21 constituted the second strongest signal identified in Germans only. This variant maps to an intron of the gene *ERG*, coding for a transcription factor (fig. S5P). The third SNP rs34286592 is located in an intron of the gene *MAZ* on chromosome 16, coding for the transcription factor MYC-associated zinc finger protein (fig. S5N). It maps to binding sites for transcription factors (fig. S8G).

When conditioning for the lead variants at the four newly identified MS-associated loci, no evidence for secondary signals was found. Thus, the lead variants also constitute the most likely causal variants. These variants all map to introns of genes. This makes a functional link between each variant and the gene it is located in probable. To further explore the functional connections between SNPs and genes, we conducted an eQTL analysis of the 15 loci showing genome-wide significant associations. We thereby identified four cis-eQTLs with FDR <0.05 in MS cases (table S5). In addition to the eQTL of rs4925166 and *SHMT1* already described above, three more significant eQTLs involved variants at two previously known MS susceptibility loci and three transcripts of the genes *MMEL1* and *ANKRD55*.

Fine-mapping of *DLEU1*

Three variants located on chromosome 13 (rs806321, rs9596270, and rs806349), all intronic within the gene for the long noncoding RNA

DLEU1, have been described previously as associated with MS in three large studies (*4*–*6*), yet the variants did not show genome-wide significance in any of them. The association of rs806349 has recently been confirmed in a candidate-driven follow-up analysis of suggestive MS associations (*21*). However, the variant rs806349 reached a *P* value of only 2.7×10^{-4} in our analysis (Table 5). Instead, a different SNP (rs2812197) in weak LD with rs806349 ($r^2 = 0.4$) showed genome-wide significance in the pooled analysis of DE1 and DE2 and also replicated in Sardinians (Tables 3 and 5, and fig. S5K). The association of previously described rs806349 is completely dependent on the more strongly associated rs2812197 (Table 5). Thus, it is unlikely that rs806349 is the causal SNP at this locus. The same is true for rs806321 (*5*), which is not independent of rs2812197 either (Table 5).

The *DLEU1* locus contains evidence for a second signal, rs9591325 (Table 5 and fig. S5L), in poor LD with rs2812197, but in high LD with rs9596270, which was identified by Patsopoulos *et al.* (*6*) as a suggestive MS-associated variant. The two signals were partially independent of each other (Table 5). SNP rs9591325 is located in a clearly functional region with binding sites for many transcription factors, which is not the case for the other four variants (fig. S8, B to F). Although rs2812197 shows the overall strongest association at *DLEU1*, the functional data indicate that rs9591325 might be either the actual or a second causal variant. Additional studies with larger sample sizes are required to fully answer this question.

Table 4. eQTL and mQTL analysis for rs4925166. Direction of effect is relative to the minor allele T. Note that the effect sizes cannot be directly compared because normalization methods and covariates partly differ between studies. Additional eQTLs and mQTLs are described in the Supplementary Materials. Because only the single eQTL rs4925166/*SHMT1* was examined in GTEx data, no FDR is indicated here. NA, not applicable.

Expression				
Data set	**Transcript**	**Effect**	***P***	**FDR**
DE1	*SHMT1*	0.36	4.42×10^{-13}	2.99×10^{-10}
MPIP	*SHMT1*	0.19	4.26×10^{-12}	1.28×10^{-11}
GTP	*SHMT1*	0.11	3.12×10^{-4}	1.25×10^{-3}
GTEx	*SHMT1*	0.56	09.2×10^{-28}	NA
Methylation				
Data set	**CpG**	**Effect**	***P***	**FDR**
MPIP	*cg26763362*	−0.03	3.21×10^{-20}	5.04×10^{-18}
GTP	*cg26763362*	−0.03	1.98×10^{-14}	1.58×10^{-13}

DISCUSSION

The present study constitutes the largest GWAS on MS conducted in a single population to date. By pooled analysis of 3934 cases in data set DE1 and 954 cases in data set DE2, we identified strong associations in the MHC region with a *P* value of up to 1.3×10^{-234}. In addition, 15 loci outside the MHC region were associated at a genome-wide significant level (Fig. 1 and Table 3). Associations in the MHC region were examined using imputed *HLA* alleles. Stepwise conditional logistic regression identified *DRB1*15:01* and six more associated *HLA* alleles (Table 2), in line with results from previous studies (*2*). All genome-wide significant and suggestive non-MHC MS susceptibility variants published by the IMSGC in 2013 (*4*) and present in our data ($n = 104$) were replicated regarding direction of effects in our samples ($P = 5 \times 10^{-32}$; Fig. 2).

Four of the 15 non-MHC loci have not been found to be associated with MS in previous studies. One more locus, *DLEU1*, did not reach genome-wide significance in previous GWAS but has recently been confirmed as MS-associated in a candidate SNP study (*21*). The lead variants at *DLEU1* and at the novel locus *SHMT1* replicated in an independent Sardinian cohort containing 2903 cases (Table 3 and fig. S6). Variants at the other three novel loci did not reach nominal significance in Sardinians, yet two of them showed the same direction of effect. Because of their consistency and replication within the German cohorts, these three associations can nevertheless be considered as plausible. As the Sardinian population is genetically distinct from Germans, future studies are required to replicate these findings in other cohorts.

Previous genetic analyses of MS susceptibility have indicated immune system–related processes as relevant for the development of MS (*4*). Functions of known MS susceptibility genes have been mapped to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways Janus kinase/signal transducers and activators of transcription (JAK/STAT) signaling, acute myeloid leukemia (AML), and T cell receptor signaling (*7*). Accordingly, MS-associated genes are predominantly expressed in immune cells (*7*, *8*). The five genes examined in detail in our study (*L3MBTL3*, *DLEU1*, *MAZ*, *ERG*, and *SHMT1*) are associated with regulatory mechanisms in immune cells as well.

The gene *L3MBTL3* encodes a Polycomb group protein that maintains the transcriptionally repressive state of genes (*31*) and is frequently deleted in several forms of acute leukemia, including AML (*32*). Genes

Table 5. Fine-mapping of the *DLEU1* locus. MAF (controls in %) and r^2 (with rs2812197) are calculated from a joint set of DE1 and DE2, ORs and *P* values from the pooled analysis of DE1 and DE2. Second and third *P* value columns are from conditional analysis.

Variant	**MAF**	**OR (CI)**	***P***	***P* (rs2812197)**	***P* (rs9591325)**	r^2	**Reference**
rs2812197	38.4	0.86 (0.82–0.91)	9.95×10^{-9}		4.79×10^{-5}	1.00	
rs806321	48.5	0.89 (0.85–0.94)	6.36×10^{-6}	0.81	2.02×10^{-3}	0.66	(*5*)
rs806349	46.0	1.10 (1.04–1.15)	2.73×10^{-4}	0.99	0.019	0.41	(*4*, *21*)
rs9591325	8.1	0.78 (0.70–0.85)	2.26×10^{-7}	9.13×10^{-4}		0.14	
rs9596270	8.1	0.78 (0.71–0.86)	4.45×10^{-7}	1.49×10^{-3}	0.27	0.14 (0.99*)	(*6*)

*r^2 with rs9591325.

associated with AML constitute one of the most significant pathway categories linked to MS susceptibility variants (*7*). The murine ortholog of *L3MBTL3*, *MBT-1*, has been found to regulate maturation of myeloid progenitor cells (*33*). The regulatory long noncoding RNA *DLEU1* is often deleted in cases of B cell chronic lymphocytic leukemia and mantle cell leukemia (*34*). This locus regulates the expression of *NF-κB* (*35*), a transcription factor implicated in MS pathology (*4*, *36*, *37*). MAZ is an inflammation-responsive transcription factor (*38*) up-regulated during chronic myeloid leukemia (*39*). It binds to the promoter of the gene *MYC*, which is associated with MS (*5*). The transcription factor ERG is important for hematopoiesis (*40*), and the expression of this oncogene is associated with both AML and acute T cell lymphoblastic leukemia (*41*). ERG regulates the expression of MS-associated *NF-κB* (*42*), as *DLEU1* does. Finally, SHMT1 is a serine hydroxymethyltransferase acting in the folate cycle. It catalyzes the transfer of a carbon unit subsequently used for synthesis of both nucleotides and methionine. SHMT1 is thus an essential component in the metabolism of the substrate *S*-adenosylmethionine (SAM), the major methyl group donor during both protein and DNA methylation (*43*, *44*). By this effect on regulation of gene expression, one-carbon metabolism plays an important role in oncogenesis. Lack of *SHMT1* function is, among other effects, associated with acute lymphocytic leukemia (*44*–*46*). Thus, each of the five genes is involved in regulatory processes of the immune system.

Although a clearer picture has already emerged regarding the cell types and broad pathways relevant for the etiology of MS (*3*, *7*), little is still known about the mechanisms by which risk genes act. Analysis of the known functions of the five genes examined in this study revealed that four of them regulate transcription, especially of immune-related genes. Moreover, indirect evidence suggests that they could all be linked either directly or indirectly to epigenetic regulatory mechanisms: L3MBTL3 recognizes epigenetic histone lysine methylation (*31*) and ERG interacts with ESET, a histone H3–specific methyltransferase (*47*). The best known regulatory target of the transcription factor MAZ is *MYC* (*48*), a regulator of epigenetic chromatin state that is associated with MS (*5*, *49*). *DLEU1* is strictly regulated by DNA methylation at its promoter region (*35*). Finally, SHMT1 is essential for maintaining methylation homeostasis in the cell by catalyzing an important reaction in the generation of the methyl donor substrate SAM. Accordingly, the establishment of SHMT1 as an MS risk factor further puts epigenetic regulation by methylation in the focus of MS susceptibility.

In recent years, several studies have addressed the role of DNA methylation in the etiology and progression of MS. Methylation differences between MS cases and healthy controls have been analyzed in small, cross-sectional studies. Despite negative results in $CD4^+$ cells (*12*, *50*), Bos and colleagues (*12*) recently observed significant differences in overall DNA methylation levels in $CD8^+$ T cells. Another study demonstrated differentially methylated and expressed genes in brain tissue of MS patients compared to controls (*14*). Furthermore, differential methylation of the major risk locus *HLA-DRB1* was observed in MS patients (*51*). Several groups have found either hypermethylation or hypomethylation of specific genes to be associated with inflammation or demyelination in MS patients (*11*).

In summary, these studies argue in favor of DNA methylation being relevant for the development of MS. By finding novel risk genes with potential roles in epigenetic regulation, our study adds further indication that epigenetic mechanisms might be important for MS susceptibility. A disturbed homeostasis of methyl donors, caused by an altered expression of *SHMT1*, is likely to have an impact on the disease. As epigenetic mechanisms constitute a major route for environmental risk factors to influence expression of disease-associated genes (*11*), regulation of DNA and protein methylation is an interface where genetic and environmental risk factors for MS might intersect. Detailed analyses of DNA methylation patterns and their interaction with MS susceptibility genes in larger cohorts and among different cell populations and tissues are now required to better understand the role of epigenetic mechanisms in MS.

MATERIALS AND METHODS

Study samples

Two cohorts of cases, referred to as DE1 and DE2, were analyzed. Both data sets included patients with CIS, bout-onset MS, and PPMS. For cohort DE1, 4503 cases were recruited across multiple sites in Germany (for details, see the Supplementary Materials). For cohort DE2, 1002 cases were recruited across multiple sites in Germany (see the Supplementary Materials). The latter cohort was used in a previous publication (*5*). Controls for these cohorts were obtained from several population-based cohorts across Germany to match the different geographical regions where cases were recruited: KORA from the southeastern Germany region of Augsburg (*52*, *53*), HNR from central western Germany (*54*), SHIP from the northeastern region of West Pomerania (*55*), DOGS from Dortmund in central western Germany (*56*), and FoCus (*57*) and PopGen (*58*) from Kiel, northern Germany. In addition, controls from two studies on depression conducted in southeastern Germany were included (*59*, *60*). For a more detailed description of control cohorts, see the Supplementary Materials. All responsible ethics committees provided positive votes for the individual studies. All study participants gave written informed consent. In case of minors, parental informed consent was obtained.

Genotyping and QC

Samples of cohort DE1 were genotyped using the Illumina HumanOmniExpress-24 v1.0 or v1.1 BeadChips. Samples of cohort DE2 were genotyped using the Illumina Human 660-Quad platform. For both cohorts, identical, stringent QC was conducted on samples and variants. QC steps on samples included removal of individuals with genotyping rate <2%, cryptic relatives (relatedness ≥1/16), and genetic population outliers. QC steps on variants included removal of variants with call rate <2% and MAF <1%. For a full description of QC, see the Supplementary Materials. Each set of cohorts was combined with controls genotyped on similar arrays, producing case/control data sets DE1 and DE2. QC was repeated on the merged data sets, leading to final figures of 3934 cases and 8455 controls for DE1 (table S1), as well as 954 cases and 1940 controls for DE2 (table S2).

Imputation

Prephasing (haplotype estimation) of genotype data was conducted using SHAPEIT2, followed by imputation using IMPUTE2 in 5–megabase pair (Mbp) chunks (*16*–*18*). The 1000 Genomes Phase 1 June 2014 release was used as a reference panel. Imputed variants were filtered for MAF (≥1%), INFO metric (≥0.8), and HWE ($P \geq 10^{-6}$). For additional details, see the Supplementary Materials.

HLA alleles were separately imputed from genotyping data for DE1 and DE2 using HIBAG v1.6.0 (*20*). Alleles with a posterior probability

>0.5 were converted to hard calls. Results were validated using *HLA* typing of 442 patients from DE1 (see the Supplementary Materials).

Statistical analyses of genotype data

GWAS was conducted on data sets DE1 and DE2 using PLINK2 v1.90b3s (*61*). Sex and the first eight MDS components were used as covariates in logistic regression. Data sets were combined using a fixed-effects model in METASOFT (*62*). For maximum precision, logistic regression and meta-analysis of lead SNPs were repeated in R v3.2.3 using package meta v4.3.2. All follow-up analyses (for example, conditional and interaction analyses) were conducted in R. Locus-specific Manhattan plots were generated using LocusZoom with European samples of the 1000 Genomes March 2012 reference panel on the hg19 build (*63*). For analysis of *HLA* alleles, stepwise logistic regression was conducted in R as previously described (*1*, *2*, *5*).

Gene expression and methylation data

For a subset of 242, mostly treatment-naïve patients from data set DE1 (73 male and 169 female) whole-blood RNA was collected using Tempus Blood RNA Tubes (Applied Biosystems). RNA was hybridized to Illumina HT-12 v4 Expression BeadChips (Illumina) and further processed as described in the Supplementary Materials. In summary, QC was conducted in R 3.2.1 using the packages beadarray and lumi (*64*, *65*). Probes were transformed and normalized through variance stabilization and normalization (*66*). Probes, which showed a detection $P < 0.05$ in more than 10% of the samples, which could not be mapped to a known transcript, or which were identified as cross-hybridizing by the Re-Annotator pipeline (*67*), were removed. This left 20,302 transcripts from 242 samples. Technical batch effects were identified by inspecting the association of the first two principal components of expression levels with amplification round, amplification plate, and amplification plate column and row, as well as expression chip. The data were then adjusted using ComBat (*68*). Gene expression and methylation data of the two control cohorts MPIP and GTP were published and described previously and are summarized in the Supplementary Materials (*25*–*28*).

Statistical analysis of gene expression and methylation data

For each of the 15 genome-wide significant loci, all 429 transcripts beginning or ending within 1 Mbp upstream or downstream of a lead variant were determined. Associations between genotype and expression levels were examined in data set DE1 by linear regression using sex, age, and three MDS components as covariates. To account for multiple testing, *P* values were first corrected for the number of transcripts per cis window, followed by calculation of the FDR for the total number of variants tested. Replication of eQTLs with an FDR <0.05 in data set DE1 was conducted in control cohorts MPIP and GTP. For MPIP, the covariates sex, age, body mass index (BMI), disease status, and three MDS components were used in linear regression. For GTP, covariates were sex, age, and four MDS components. eQTLs were also looked up in the GTEx database (*29*). Here, only associations in whole blood were considered.

For analysis of the association of rs4925166 with DNA methylation at *SHMT1*, 210 CpG probes were identified in data set MPIP that mapped to *SHMT1*. After removing the quartile of probes showing the lowest variation in methylation status, 157 CpGs remained. Association of DNA methylation with imputed genotype was assessed by linear regression using sex, age, BMI, disease status, three MDS components, and estimated cell counts as covariates. The eight CpG probes showing an FDR <0.05 were replicated in data set GTP, using sex, age, four MDS components, and estimated cell counts as covariates. Mediation analysis was conducted as outlined in the Supplementary Materials, including nonparametric bootstrap for estimation of CIs and *P* values (*30*).

Replication of the results in a Sardinian cohort

The replication case group consisted of 2903 unrelated Sardinian MS patients that were diagnosed and selected using the McDonald criteria (*22*–*24*). Only 35 of these patients were diagnosed with PPMS (1.2%). Two thousand ten (69.2%) cases were female, and 893 (30.8%) were male; the average age at onset was 32 years. The matching control group of healthy individuals is composed of 2880 unrelated adult volunteer blood donors from the same locations where the cases were collected, as well as 443 Affected Family BAsed pseudo-Controls (AFBACs) derived from 242 MS and 201 type 1 diabetes family trios (*23*). AFBAC allele and haplotype frequencies were constructed using the two alleles in each trio that are not transmitted from the parents to the affected child. These familial pseudo-controls are matched to the cases for ethnic origin and are thus robust to population stratification.

All individuals were genotyped using the Illumina ImmunoChip array. In addition, 2040 (962 cases and 1078 controls) were genotyped with the Illumina HumanOmniExpress array and 3917 (2111 cases and 1806 controls) with the Affymetrix 6.0 array. One hundred seventy-four individuals (170 case and 4 controls) were genotyped using both HumanOmniExpress and Affymetrix 6.0 (*22*). After QC, we used 883,557 SNPs as baseline for imputation (*17*) of 20.1 million untyped SNPs using a Sardinian-specific reference panel, including 3514 Sardinian individuals sequenced to an average coverage of 4.16-fold (*69*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/6/1501678/DC1

Supplementary Results

Supplementary Materials and Methods

table S1. QC of data set DE1.

table S2. QC of data set DE2.

table S3. Genomic inflation.

table S4. Genome-wide significant loci.

table S5. eQTLs with FDR <0.05 in data set DE1.

table S6. Replicated mQTLs of rs4925166 and CpG sites in *SHMT1*.

table S7. Mediation analysis.

table S8. Causal mediation analysis.

fig. S1. Substructure analysis results in DE1.

fig. S2. Substructure analysis results in DE2.

fig. S3. GWAS with age at onset.

fig. S4. Forest plots of all non-MHC top genome-wide significant variants.

fig. S5. Locus-specific Manhattan plots.

fig. S6. Forest plots of novel variants replicated in a Sardinian cohort.

fig. S7. eQTL and mQTL analysis for rs4925166.

fig. S8. Transcription factor binding sites.

References (*70*–*75*)

REFERENCES AND NOTES

1. N. A. Patsopoulos, L. F. Barcellos, R. Q. Hintzen, C. Schaefer, C. M. van Duijn, J. A. Noble, T. Raj, IMSGC, ANZgene, P.-A. Gourraud, B. E. Stranger, J. Oksenberg, T. Olsson, B. V. Taylor, S. Sawcer, D. A. Hafler, M. Carrington, P. L. De Jager, P. I. W. de Bakker, Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLOS Genet.* **9**, e1003926 (2013).
2. L. Moutsianas, L. Jostins, A. H. Beecham, A. T. Dilthey, D. K. Xifara, M. Ban, T. S. Shah, N. A. Patsopoulos, L. Alfredsson, C. A. Anderson, K. E. Attfield, S. E. Baranzini, J. Barrett,

T. M. C. Binder, D. Booth, D. Buck, E. G. Celius, C. Cotsapas, S. D'Alfonso, C. A. Dendrou, P. Donnelly, B. Dubois, B. Fontaine, L. Fugger, A. Goris, P.-A. Gourraud, C. Graetz, B. Hemmer, J. Hillert; International IBD Genetics Consortium (IIBDGC), I. Kockum, S. Leslie, C. M. Lill, F. Martinelli-Boneschi, J. R. Oksenberg, T. Olsson, A. Oturai, J. Saarela, H. B. Søndergaard, A. Spurkland, B. Taylor, J. Winkelmann, F. Zipp, J. L. Haines, M. A. Pericak-Vance, C. C. A. Spencer, G. Stewart, D. A. Hafler, A. J. Ivinson, H. F. Harbo, S. L. Hauser, P. L. De Jager, A. Compston, J. L. McCauley, S. Sawcer, G. McVean; International Multiple Sclerosis Genetics Consortium, International IBD Genetics Consortium (IIBDGC), Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* **47**, 1107–1113 (2015).

3. S. Sawcer, R. J. M. Franklin, M. Ban, Multiple sclerosis genetics. *Lancet Neurol.* **13**, 700–709 (2014).
4. International Multiple Sclerosis Genetics Consortium (IMSGC), A. H. Beecham, N. A. Patsopoulos, D. K. Xifara, M. F. Davis, A. Kemppinen, C. Cotsapas, T. S. Shah, C. Spencer, D. Booth, A. Goris, A. Oturai, J. Saarela, B. Fontaine, B. Hemmer, C. Martin, F. Zipp, S. D'Alfonso, F. Martinelli-Boneschi, B. Taylor, H. F. Harbo, I. Kockum, J. Hillert, T. Olsson, M. Ban, J. R. Oksenberg, R. Hintzen, L. F. Barcellos; Wellcome Trust Case Control Consortium 2 (WTCCC2), International IBD Genetics Consortium (IIBDGC), C. Agliardi, L. Alfredsson, M. Alizadeh, C. Anderson, R. Andrews, H. Bach Søndergaard, A. Baker, G. Band, S. E. Baranzini, N. Barizzone, J. Barrett, C. Bellenguez, L. Bergamaschi, L. Bernardinelli, A. Berthele, V. Biberacher, T. M. C. Binder, H. Blackburn, I. L. Bomfim, P. Brambilla, S. Broadley, B. Brochet, L. Brundin, D. Buck, H. Butzkueven, S. J. Caillier, W. Camu, W. Carpentier, P. Cavalla, E. G. Celius, I. Coman, G. Comi, L. Corrado, L. Cosemans, I. Cournu-Rebeix, B. A. C. Cree, D. Cusi, V. Damotte, G. Defer, S. R. Delgado, P. Deloukas, A. di Sapio, A. T. Dilthey, P. Donnelly, B. Dubois, M. Duddy, S. Edkins, I. Elovaara, F. Esposito, N. Evangelou, B. Fiddes, J. Field, A. Franke, C. Freeman, I. Y. Frohlich, D. Galimberti, C. Gieger, P.-A. Gourraud, C. Graetz, A. Graham, V. Grummel, C. Guaschino, A. Hadjixenofontos, H. Hakonarson, C. Halfpenny, G. Hall, P. Hall, A. Hamsten, J. Harley, T. Harrower, C. Hawkins, G. Hellenthal, C. Hillier, J. Hobart, M. Hoshi, S. E. Hunt, M. Jagodic, I. Jelčić, A. Jochim, B. Kendall, A. Kermode, T. Kilpatrick, K. Koivisto, I. Konidari, T. Korn, H. Kronsbein, C. Langford, M. Larsson, M. Lathrop, C. Lebrun-Frenay, J. Lechner-Scott, M. H. Lee, M. A. Leone, V. Leppä, G. Liberatore, B. A. Lie, C. M. Lill, M. Lindén, J. Link, F. Luessi, J. Lycke, F. Macciardi, S. Männistö, C. P. Manrique, R. Martin, V. Martinelli, D. Mason, G. Mazibrada, C. McCabe, I.-L. Mero, J. Mescheriakova, L. Moutsianas, K.-M. Myhr, G. Nagels, R. Nicholas, P. Nilsson, F. Piehl, M. Pirinen, S. E. Price, H. Quach, M. Reunanen, W. Robberecht, N. P. Robertson, M. Rodegher, D. Rog, M. Salvetti, N. C. Schnetz-Boutaud, F. Sellebjerg, R. C. Selter, C. Schaefer, S. Shaunak, L. Shen, S. Shields, V. Siffrin, M. Slee, P. Soelberg Sorensen, M. Sorosina, M. Sospedra, A. Spurkland, A. Strange, E. Sundqvist, V. Thijs, J. Thorpe, A. Ticca, P. Tienari, C. van Duijn, E. M. Visser, S. Vucic, H. Westerlind, J. S. Wiley, A. Wilkins, J. F. Wilson, J. Winkelmann, J. Zajicek, E. Zindler, J. L. Haines, M. A. Pericak-Vance, A. J. Ivinson, G. Stewart, D. Hafler, S. L. Hauser, A. Compston, G. McVean, P. De Jager, S. J. Sawcer, J. L. McCauley, Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013).
5. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, S. Sawcer, G. Hellenthal, M. Pirinen, C. C. A. Spencer, N. A. Patsopoulos, L. Moutsianas, A. Dilthey, Z. Su, C. Freeman, S. E. Hunt, S. Edkins, E. Gray, D. R. Booth, S. C. Potter, A. Goris, G. Band, A. B. Oturai, A. Strange, J. Saarela, C. Bellenguez, B. Fontaine, M. Gillman, B. Hemmer, R. Gwilliam, F. Zipp, A. Jayakumar, R. Martin, S. Leslie, S. Hawkins, E. Giannoulatou, S. D'alfonso, H. Blackburn, F. Martinelli Boneschi, J. Liddle, H. F. Harbo, M. L. Perez, A. Spurkland, M. J. Waller, M. P. Mycko, M. Ricketts, M. Comabella, N. Hammond, I. Kockum, O. T. McCann, M. Ban, P. Whittaker, A. Kemppinen, P. Weston, C. Hawkins, S. Widaa, J. Zajicek, S. Dronov, N. Robertson, S. J. Bumpstead, L. F. Barcellos, R. Ravindrarajah, R. Abraham, L. Alfredsson, K. Ardlie, C. Aubin, A. Baker, K. Baker, S. E. Baranzini, L. Bergamaschi, R. Bergamaschi, A. Bernstein, A. Berthele, M. Boggild, J. P. Bradfield, D. Brassat, S. A. Broadley, D. Buck, H. Butzkueven, R. Capra, W. M. Carroll, P. Cavalla, E. G. Celius, S. Cepok, R. Chiavacci, F. Clerget-Darpoux, K. Clysters, G. Comi, M. Cossburn, I. Cournu-Rebeix, M. B. Cox, W. Cozen, B. A. C. Cree, A. H. Cross, D. Cusi, M. J. Daly, E. Davis, P. I. W. de Bakker, M. Debouverie, M. B. D'hooghe, K. Dixon, R. Dobosi, B. Dubois, D. Ellinghaus, I. Elovaara, F. Esposito, C. Fontenille, S. Foote, A. Franke, D. Galimberti, A. Ghezzi, J. Glessner, R. Gomez, O. Gout, C. Graham, S. F. A. Grant, F. R. Guerini, H. Hakonarson, P. Hall, A. Hamsten, H.-P. Hartung, R. N. Heard, S. Heath, J. Hobart, M. Hoshi, C. Infante-Duarte, G. Ingram, W. Ingram, T. Islam, M. Jagodic, M. Kabesch, A. G. Kermode, T. J. Kilpatrick, C. Kim, N. Klopp, K. Koivisto, M. Larsson, M. Lathrop, J. S. Lechner-Scott, M. A. Leone, V. Leppä, U. Liljedahl, I. Lima Bomfim, R. R. Lincoln, J. Link, J. Liu, Å. R. Lorentzen, S. Lupoli, F. Macciardi, T. Mack, M. Marriott, V. Martinelli, D. Mason, J. L. McCauley, F. Mentch, I.-L. Mero, T. Mihalova, X. Montalban, J. Mottershead, K.-M. Myhr, P. Naldi, W. Ollier, A. Page, A. Palotie, J. Pelletier, L. Piccio, T. Pickersgill, F. Piehl, S. Pobywajlo, H. L. Quach, P. P. Ramsay, M. Reunanen, R. Reynolds, J. D. Rioux, M. Rodegher, S. Roesner, J. P. Rubio, I.-M. Rückert, M. Salvetti, E. Salvi, A. Santaniello, C. A. Schaefer, S. Schreiber, C. Schulze, R. J. Scott, F. Sellebjerg, K. W. Selmaj, D. Sexton, L. Shen, B. Simms-Acuna, S. Skidmore, P. M. A. Sleiman, C. Smestad, P. Soelberg Sørensen, H. Bach Søndergaard, J. Stankovich, R. C. Strange, A.-M. Sulonen, E. Sundqvist, A.-C. Syvänen, F. Taddeo, B. Taylor, J. M. Blackwell, P. Tienari, E. Bramon, A. Tourbah, M. A. Brown, E. Tronczynska, J. P. Casas, N. Tubridy, A. Corvin, J. Vickery, J. Jankowski, P. Villoslada, H. S. Markus, K. Wang, C. G. Mathew, J. Wason, C. N. A. Palmer, H.-E. Wichmann, R. Plomin, E. Willoughby, A. Rautanen, J. Winkelmann, M. Wittig, R. C. Trembath, J. Yaouanq, A. C. Viswanathan, H. Zhang, N. W. Wood, R. Zuvich, P. Deloukas, C. Langford, A. Duncanson, J. R. Oksenberg, M. A. Pericak-Vance, J. L. Haines, T. Olsson, J. Hillert, A. J. Ivinson, P. L. De Jager, L. Peltonen, G. J. Stewart, D. A. Hafler, S. L. Hauser, G. McVean, P. Donnelly, A. Compston, Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
6. N. A. Patsopoulos; Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium, F. Esposito, J. Reischl, S. Lehr, D. Bauer, J. Heubach, R. Sandbrink, C. Pohl, G. Edan, L. Kappos, D. Miller, J. Montalbán, C. H. Polman, M. S. Freedman, H. P. Hartung, B. G. Arnason, G. Comi, S. Cook, M. Filippi, D. S. Goodin, D. Jeffery, P. O'Connor, G. C. Ebers, D. Langdon, A. T. Reder, A. Traboulsee, F. Zipp, S. Schimrigk, J. Hillert, M. Bahlo, D. R. Booth, S. Broadley, M. A. Brown, B. L. Browning, S. R. Browning, H. Butzkueven, W. M. Carroll, C. Chapman, S. J. Foote, L. Griffiths, A. G. Kermode, T. J. Kilpatrick, J. Lechner-Scott, M. Marriott, D. Mason, P. Moscato, R. N. Heard, M. P. Pender, V. M. Perreau, D. Perera, J. P. Rubio, R. J. Scott, M. Slee, J. Stankovich, G. J. Stewart, B. V. Taylor, N. Tubridy, E. Willoughby, J. Wiley, P. Matthews, F. M. Boneschi, A. Compston, J. Haines, S. L. Hauser, J. McCauley, A. Ivinson, J. R. Oksenberg, M. Pericak-Vance, S. J. Sawcer, P. L. De Jager, D. A. Hafler, P. I. de Bakker, Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* **70**, 897–912 (2011).
7. International Multiple Sclerosis Genetics Consortium, Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am. J. Hum. Genet.* **92**, 854–865 (2013).
8. K. K.-H. Farh A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. H. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. John Luckey, N. A. Patsopoulos, P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, B. E. Bernstein, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
9. C. H. Hawkes, A. J. Macgregor, Twin studies and the heritability of MS: A conclusion. *Mult. Scler.* **15**, 661–667 (2009).
10. A. K. Hedström, T. Olsson, L. Alfredsson, The role of environment and lifestyle in determining the risk of multiple sclerosis. *Curr. Top. Behav. Neurosci.* **26**, 87–104 (2015).
11. Y. Zhou, S. Simpson Jr., A. F. Holloway, J. Charlesworth, I. van der Mei, B. V. Taylor, The potential role of epigenetic modifications in the heritability of multiple sclerosis. *Mult. Scler.* **20**, 135–140 (2014).
12. S. D. Bos, C. M. Page, B. K. Andreassen, E. Elboudwarej, M. W. Gustavsen, F. Briggs, H. Quach, I. S. Leikfoss, A. Bjølgerud, T. Berge, H. F. Harbo, L. F. Barcellos, Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. *PLOS One* **10**, e0117403 (2015).
13. J. L. Huynh, P. Casaccia, Epigenetic mechanisms in multiple sclerosis: Implications for pathogenesis and treatment. *Lancet Neurol.* **12**, 195–206 (2013).
14. J. L. Huynh, P. Garg, T. Htwe Thin, S. Yoo, R. Dutta, B. D. Trapp, V. Haroutunian, J. Zhu, M. J. Donovan, A. J. Sharp, P. Casaccia, Epigenome-wide differences in pathology-free regions of multiple sclerosis–affected brains. *Nat. Neurosci.* **17**, 121–130 (2014).
15. M. Steffens M. Steffens, C. Lamina, T. Illig, T. Bettecken, R. Vogler, P. Entz, E.-K. Suk, M. Reza Toliat, N. Klopp, A. Caliebe, I. R. König, K. Köhler, J. Ludemann, A. Diaz Lacava, R. Fimmers, P. Lichtner, A. Ziegler, A. Wolf, M. Krawczak, P. Nürnberg, J. Hampe, S. Schreiber, T. Meitinger, H.-E. Wichmann, K. Roeder, T. F. Wienker, M. P. Baur, SNP-based analysis of genetic substructure in the German population. *Hum. Hered.* **62**, 20–29 (2006).
16. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009).
17. B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G. R. Abecasis, Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
18. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
19. M. L. Freedman M. L. Freedman, D. Reich, K. L. Penney, G. J. McDonald, A. A. Mignault, N. Patterson, S. B. Gabriel, E. J. Topol, J. W. Smoller, C. N. Pato, M. T. Pato, T. L. Petryshen, L. N. Kolonel, E. S. Lander, P. Sklar, B. Henderson, J. N. Hirschhorn, D. Altshuler, Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36**, 388–393 (2004).
20. X. Zheng, J. Shen, C. Cox, J. C. Wakefield, M. G. Ehm, M. R. Nelson, B. S. Weir, HIBAG—HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* **14**, 192–200 (2014).

21. C. M. Lill, F. Luessi, A. Alcina, E. A. Sokolova, N. Ugidos, B. de la Hera, L. Guillot-Noël, S. Malhotra, E. Reinthaler, B.-M. M. Schjeide, J. Y. Mescheriakova, A. Mashychev, I. Wohlers, D. A. Akkad, O. Aktas, I. Alloza, A. Antigüedad, R. Arroyo, I. Astobiza, P. Blaschke, A. N. Boyko, M. Buttmann, A. Chan, T. Dörner, J. T. Epplen, O. O. Favorova, M. Fedetz, O. Fernández, A. García-Martínez, L.-A. Gerdes, C. Graetz, H.-P. Hartung, S. Hoffjan, G. Izquierdo, D. S. Korobko, A. Kroner, C. Kubisch, T. Kümpfel, L. Leyva, P. Lohse, N. A. Malkova, X. Montalban, E. V. Popova, P. Rieckmann, A. S. Rozhdestvenskii, C. Schmied, I. V. Smagina, E. Y. Tsareva, A. Winkelmann, U. K. Zettl, H. Binder, I. Cournu-Rebeix, R. Hintzen, A. Zimprich, M. Comabella, B. Fontaine, E. Urcelay, K. Vandenbroeck, M. Filipenko, F. Matesanz, F. Zipp, L. Bertram, Genome-wide significant association with seven novel multiple sclerosis risk loci. *J. Med. Genet.* **52**, 848–855 (2015).
22. S. Sanna, M. Pitzalis, M. Zoledziewska, I. Zara, C. Sidore, R. Murru, M. B. Whalen, F. Busonero, A. Maschio, G. Costa, M. Cristina Melis, F. Deidda, F. Poddie, L. Morelli, G. Farina, Y. Li, M. Dei, S. Lai, A. Mulas, G. Cuccuru, E. Porcu, L. Liang, P. Zavattari, L. Moi, E. Deriu, M. Francesca Urru, M. Bajorek, M. Anna Satta, E. Cocco, P. Ferrigno, S. Sotgiu, M. Pugliatti, S. Traccis, A. Angius, M. Melis, G. Rosati, G. R. Abecasis, M. Uda, M. Giovanna Marrosu, D. Schlessinger, F. Cucca, Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis. *Nat. Genet.* **42**, 495–497 (2010).
23. M. Zoledziewska, G. Costa, M. Pitzalis, E. Cocco, C. Melis, L. Moi, P. Zavattari, R. Murru, R. Lampis, L. Morelli, F. Poddie, P. Frongia, P. Pusceddu, M. Bajorek, A. Marras, A. M. Satta, A. Chessa, M. Pugliatti, S. Sotgiu, M. B. Whalen, G. Rosati, F. Cucca, M. G. Marrosu, Variation within the *CLEC16A* gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia. *Genes Immun.* **10**, 15–17 (2009).
24. N. Barizzone, I. Zara, M. Sorosina, S. Lupoli, E. Porcu, M. Pitzalis, M. Zoledziewska, F. Esposito, M. Leone, A. Mulas, P. Ferrigno, F. R. Guerini, P. Brambilla, G. Farina, R. Murru, F. Deidda, S. Sanna, A. Loi, C. Barlassina, D. Vecchio, A. Zauli, F. Clarelli, D. Braga, F. Poddie, R. Cantello, V. Martinelli, G. Comi, J. Frau, L. Lorefice, M. Pugliatti, G. Rosati; PROGEMUS (PROgnostic GEnetic factors in MUltiple Sclerosis) Consortium PROGRESSO (Italian network of Primary Progressive Multiple Sclerosis) Consortium, M. Melis, M. G. Marrosu, D. Cusi, F. Cucca, F. Martinelli Boneschi, S. Sanna, S. D'Alfonso, The burden of multiple sclerosis variants in continental Italians and Sardinians. *Mult. Scler.* **21**, 1385–1395 (2015).
25. J. Arloth, R. Bogdan, P. Weber, G. Frishman, A. Menke, K. V. Wagner, G. Balsevich, M. V. Schmidt, N. Karbalai, D. Czamara, A. Altmann, D. Trümbach, W. Wurst, D. Mehta, M. Uhr, T. Klengel, A. Erdhart, C. E. Carey, E. D. Conley; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC), A. Ruepp, B. Müller-Myhsok, A. R. Ahmad, E. B. Binder, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC), Genetic differences in the immediate transcriptome response to stress predict risk-related brain function and psychiatric disorders. *Neuron* **86**, 1189–1202 (2015).
26. E. B. Binder, R. G. Bradley, W. Liu, M. P. Epstein, T. C. Deveau, K. B. Mercer, Y. Tang, C. F. Gillespie, C. M. Heim, C. B. Nemeroff, A. C. Schwartz, J. F. Cubells, K. J. Ressler, Association of *FKBP5* polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. *JAMA* **299**, 1291–1305 (2008).
27. D. Mehta, T. Klengel, K. N. Conneely, A. K. Smith, A. Altmann, T. W. Pace, M. Rex-Haffner, A. Loeschner, M. Gonik, K. B. Mercer, B. Bradley, B. Müller-Myhsok, K. J. Ressler, E. B. Binder, Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8302–8307 (2013).
28. A. S. Zannas, J. Arloth, T. Carrillo-Roa, S. Iurato, S. Röh, K. J. Ressler, C. B. Nemeroff, A. K. Smith, B. Bradley, C. Heim, A. Menke, J. F. Lange, T. Brückl, M. Ising, N. R. Wray, A. Erhardt, E. B. Binder, D. Mehta, Lifetime stress accelerates epigenetic aging in an urban, African American cohort: Relevance of glucocorticoid signaling. *Genome Biol.* **16**, 266 (2015).
29. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou; The GTEx Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
30. D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, Mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).
31. N. Nady, L. Krichevsky, N. Zhong, S. Duan, W. Tempel, M. F. Amaya, M. Ravichandran, C. H. Arrowsmith, Histone recognition by human malignant brain tumor domains. *J. Mol. Biol.* **423**, 702–718 (2012).
32. M. Merup, T. Calero Moreno, M. Heyman, K. Rönnberg, D. Grandér, R. Detlofsson, O. Rasool, Y. Liu, S. Söderhäll, G. Juliusson, G. Gahrton, S. Einhorn, 6q deletions in acute lymphoblastic leukemia and non-Hodgkin's lymphomas. *Blood* **91**, 3397–3400 (1998).
33. S. Arai, T. Miyazaki, Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein MBT-1. *EMBO J.* **24**, 1863–1873 (2005).
34. S. Stilgenbauer, J. Nickolenko, J. Wilhelm, S. Wolf, S. Weitz, K. Döhner, T. Boehm, H. Döhner, P. Lichter, Expressed sequences as candidates for a novel tumor suppressor gene at band 13q14 in B-cell chronic lymphocytic leukemia and mantle cell lymphoma. *Oncogene* **16**, 1891–1897 (1998).
35. A. Garding, N. Bhattacharya, R. Claus, M. Ruppel, C. Tschuch, K. Filarsky, I. Idler, M. Zucknick, M. Caudron-Herger, C. Oakes, V. Fleig, I. Keklikoglou, D. Allegra, L. Serra, S. Thakurela, V. Tiwari, D. Weichenhan, A. Benner, B. Radlwimmer, H. Zentgraf, S. Wiemann, K. Rippe, C. Plass, H. Döhner, P. Lichter, S. Stilgenbauer, D. Mertens, Epigenetic upregulation of lncRNAs at 13q14.3 in leukemia is linked to the *In Cis* downregulation of a gene cluster that targets NF-kB. *PLOS Genet.* **9**, e1003373 (2013).
36. J. Yan, J. M. Greer, NF-κB, a potential therapeutic target for the treatment of multiple sclerosis. *CNS Neurol. Disord. Drug Targets* **7**, 536–557 (2008).
37. W. J. Housley, S. D. Fernandez, K. Vera, S. R. Murikinati, J. Grutzendler, N. Cuerdon, L. Glick, P. L. De Jager, M. Mitrovic, C. Cotsapas, D. A. Hafler, Genetic variants associated with autoimmunity drive NFκB signaling and responses to inflammatory stimuli. *Sci. Transl. Med.* **7**, 291ra93 (2015).
38. A. Ray, S. Dhar, A. Shakya, P. Ray, Y. Okada, B. K. Ray, SAF-3, a novel splice variant of the SAF-1/MAZ/Pur-1 family, is expressed during inflammation. *FEBS J.* **276**, 4276–4286 (2009).
39. L. Dahéron, S. Salmeron, S. Patri, A. Brizard, F. Guilhot, J.-C. Chomel, A. Kitzis, Identification of several genes differentially expressed during progression of chronic myelogenous leukemia. *Leukemia* **12**, 326–332 (1998).
40. S. J. Loughran, E. A. Kruse, D. F. Hacking, C. A. de Graaf, C. D. Hyland, T. A. Willson, K. J. Henley, S. Ellis, A. K. Voss, D. Metcalf, D. J. Hilton, W. S. Alexander, B. T. Kile, The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* **9**, 810–819 (2008).
41. C. D. Baldus, T. Burmeister, P. Martus, S. Schwartz, N. Gökbuget, C. D. Bloomfield, D. Hoelzer, E. Thiel, W. K. Hofmann, High expression of the ETS transcription factor *ERG* predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J. Clin. Oncol.* **24**, 4714–4720 (2006).
42. B. Hoesel, J. A. Schmid, The complexity of NF- κB signaling in inflammation and cancer. *Mol. Cancer* **12**, 86 (2013).
43. K. S. Crider, T. P. Yang, R. J. Berry, L. B. Bailey, Folate and DNA methylation: A review of molecular mechanisms and the evidence for folate's role. *Adv. Nutr.* **3**, 21–38 (2012).
44. J. W. Locasale, Serine, glycine and one-carbon units: Cancer metabolism in full circle. *Nat. Rev. Cancer* **13**, 572–583 (2013).
45. R. de Jonge, W. J. E. Tissing, J. Hendrik Hooijberg, G. Jansen, G. J. L. Kaspers, J. Lindemans, G. J. Peters, R. Pieters, Polymorphisms in folate-related genes and risk of pediatric acute lymphoblastic leukemia. *Blood* **113**, 2284–2289 (2009).
46. C. F. Skibola, M. T. Smith, A. Hubbard, B. Shane, A. C. Roberts, G. R. Law, S. Rollinson, E. Roman, R. A. Cartwright, G. J. Morgan, Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. *Blood* **99**, 3786–3791 (2002).
47. L. Yang, L. Xia, D. Y. Wu, H. Wang, H. A. Chansky, W. H. Schubach, D. D. Hickstein, Y. Zhang, Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene* **21**, 148–152 (2002).
48. S. A. Bossone, C. Asselin, A. J. Patel, K. B. Marcu, MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7452–7456 (1992).
49. N. V. Varlakhanova, P. S. Knoepfler, Acting locally and globally: Myc's ever-expanding roles on chromatin. *Cancer Res.* **69**, 7487–7490 (2009).
50. S. E. Baranzini, J. Mudge, J. C. van Velkinburgh, P. Khankhanian, I. Khrebtukova, N. A. Miller, L. Zhang, A. D. Farmer, C. J. Bell, R. W. Kim, G. D. May, J. E. Woodward, S. J. Caillier, J. P. McElroy, R. Gomez, M. J. Pando, L. E. Clendenen, E. E. Ganusova, F. D. Schilkey, T. Ramaraj, O. A. Khan, J. J. Huntley, S. Luo, P.-y. Kwok, T. D. Wu, G. P. Schroth, J. R. Oksenberg, S. L. Hauser, S. F. Kingsmore, Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351–1356 (2010).
51. M. C. Graves, M. Benton, R. A. Lea, M. Boyle, L. Tajouri, D. Macartney-Coxson, R. J. Scott, J. Lechner-Scott, Methylation differences at the *HLA-DRB1* locus in CD4+ T-Cells are associated with multiple sclerosis. *Mult. Scler.* **20**, 1033–1041 (2013).
52. R. Holle, M. Happich, H. Löwel, H. E. Wichmann; MONICA/KORA Study Group, KORA-a research platform for population based health research. *Gesundheitswesen* **67** (Suppl. 1), S19–S25 (2005).
53. H.-E. Wichmann, C. Gieger, T. Illig; MONICA/KORA Study Group, KORA-gen–Resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67** (Suppl. 1), S26–S30 (2005).
54. A. Schmermund, S. Möhlenkamp, A. Stang, D. Grönemeyer, R. Seibel, H. Hirche, K. Mann, W. Siffert, K. Lauterbach, J. Siegrist, K.-H. Jöckel, R. Erbel, Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *Am. Heart J.* **144**, 212–218 (2002).
55. H. Völzke, D. Alte, C. O. Schmidt, D. Radke, R. Lorbeer, N. Friedrich, N. Aumann, K. Lau, M. Piontek, G. Born, C. Havemann, T. Ittermann, S. Schipf, R. Haring, S. E. Baumeister, H. Wallaschofski, M. Nauck, S. Frick, A. Arnold, M. Jünger, J. Mayerle, M. Kraft, M. M. Lerch, M. Dörr, T. Reffelmann, K. Empen, S. B. Felix, A. Obst, B. Koch, S. Gläser, R. Ewert, I. Fietze, T. Penzel, M. Dören, W. Rathmann, J. Haerting, M. Hannemann,

J. Röpcke, U. Schminke, C. Jürgens, F. Tost, R. Rettig, J. A. Kors, S. Ungerer, K. Hegenscheid, J.-P. Kühn, J. Kühn, N. Hosten, R. Puls, J. Henke, O. Gloger, A. Teumer, G. Homuth, U. Völker, C. Schwahn, B. Holtfreter, I. Polzer, T. Kohlmann, H. J. Grabe, D. Rosskopf, H. K. Kroemer, T. Kocher, R. Biffar, U. John, W. Hoffmann, Cohort profile: The study of health in Pomerania. *Int. J. Epidemiol.* **40**, 294–307 (2011).
56. K. Berger, DOGS. *Bundesgesundheitsbla.* **55**, 816–821 (2012).
57. N. Müller, D. M. Schulte, K. Türk, S. Freitag-Wolf, J. Hampe, R. Zeuner, J. O. Schröder, I. Gouni-Berthold, H. K. Berthold, W. Krone, S. Rose-John, S. Schreiber, M. Laudes, IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.* **56**, 1034–1042 (2015).
58. M. Krawczak, S. Nikolaus, H. von Eberstein, P. J. P. Croucher, N. E. El Mokhtari, S. Schreiber, PopGen: Population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **9**, 55–61 (2006).
59. P. Muglia, F. Tozzi, N. W. Galwey, C. Francks, R. Upmanyu, X. Q. Kong, A. Antoniades, E. Domenici, J. Perry, S. Rothen, C. L. Vandeleur, V. Mooser, G. Waeber, P. Vollenweider, M. Preisig, S. Lucae, B. Müller-Myhsok, F. Holsboer, L. T. Middleton, A. D. Roses, Genome-wide association study of recurrent major depressive disorder in two European case–control cohorts. *Mol. Psychiatry* **15**, 589–601 (2010).
60. M. A. Kohli, S. Lucae, P. G. Saemann, M. V. Schmidt, A. Demirkan, K. Hek, D. Czamara, M. Alexander, D. Salyakina, S. Ripke, D. Hoehn, M. Specht, A. Menke, J. Hennings, A. Heck, C. Wolf, M. Ising, S. Schreiber, M. Czisch, M. B. Müller, M. Uhr, T. Bettecken, A. Becker, J. Schramm, M. Rietschel, W. Maier, B. Bradley, K. J. Ressler, M. M. Nöthen, S. Cichon, I. W. Craig, G. Breen, C. M. Lewis, A. Hofman, H. Tiemeier, C. M. van Duijn, F. Holsboer, B. Müller-Myhsok, E. B. Binder, The neuronal transporter gene *SLC6A15* confers risk to major depression. *Neuron* **70**, 252–265 (2011).
61. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
62. B. Han, E. Eskin, Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
63. R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, C. J. Willer, LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
64. M. J. Dunning, M. L. Smith, M. E. Ritchie, S. Tavaré, Beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183–2184 (2007).
65. P. Du, W. A. Kibbe, S. M. Lin, *Lumi*: A pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
66. W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (Suppl. 1), S96–S104 (2002).
67. J. Arloth, D. M. Bader, S. Röh, A. Altmann, Re-Annotator: Annotation pipeline for microarray probe sequences. *PLOS One* **10**, e0139516 (2015).
68. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
69. C. Sidore, F. Busonero, A. Maschio, E. Porcu, S. Naitza, M. Zoledziewska, A. Mulas, G. Pistis, M. Steri, F. Danjou, A. Kwong, V. D. O. del Vecchyo, C. W. K. Chiang, J. Bragg-Gresham, M. Pitzalis, R. Nagaraja, B. Tarrier, C. Brennan, S. Uzzau, C. Fuchsberger, R. Atzeni, F. Reinier, R. Berutti, J. Huang, N. J. Timpson, D. Toniolo, P. Gasparini, G. Malerba, G. Dedoussis, E. Zeggini, N. Soranzo, C. Jones, R. Lyons, A. Angius, H. M. Kang, J. Novembre, S. Sanna, D. Schlessinger, F. Cucca, G. R. Abecasis, Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47**, 1272–1281 (2015).
70. M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. Christopher Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, M. Snyder, Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
71. J. Wang, J. Zhuang, S. Iyer, X.-Y. Lin, M. C. Greven, B.-H. Kim, J. Moore, B. G. Pierce, X. Dong, D. Virgil, E. Birney, J.-H. Hung, Z. Weng, Factorbook.org: A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res* **41**, D171–D176 (2013).
72. J. Wang, J. Zhuang, S. Iyer, X. Lin, T. W. Whitfield, M. C. Greven, B. G. Pierce, X. Dong, A. Kundaje, Y. Cheng, O. J. Rando, E. Birney, R. M. Myers, W. S. Noble, M. Snyder, Z. Weng, Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812 (2012).
73. M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, R. A. Irizarry, Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
74. J.-P. Fortin, A. Labbe, M. Lemire, B. W. Zanke, T. J. Hudson, E. J. Fertig, C. M. T. Greenwood, K. D. Hansen, Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
75. E. A. Houseman, W. P. Accomando, D. C. Koestler, B. C. Christensen, C. J. Marsit, H. H. Nelson, J. K. Wiencke, K. T. Kelsey, DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).

Acknowledgments: We thank V. Grummel, J. Hornung, and N. Miksch for technical assistance as well as N. Provençal and S. Ripke for scientific discussions. This study makes use of data generated by the WTCCC. A full list of the investigators who contributed to the generation of the data is available at www.wtccc.org.uk. **Funding:** This work was supported by the German Ministry for Education and Research (BMBF) as part of the "German Competence Network Multiple Sclerosis" (KKNMS) (grant nos. 01GI0916 and 01GI0917) and the Munich Cluster for Systems Neurology (SyNergy). D.B., B.H., F.Z., and H.W. were supported by the German Research Foundation (DFG) (grant no. CRC128). Funding for the WTCCC project was provided by the Wellcome Trust under awards 076113 and 085475. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine and Headache Society (DMKG) and unrestricted grants of equal share from Almirall, AstraZeneca, Berlin-Chemie, Boehringer, Boots Healthcare, GlaxoSmithKline (GSK), Janssen-Cilag, McNeil Pharma, Merck Sharp & Dohme (MSD), and Pfizer to the University of Münster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Münster; genotyping was supported by the BMBF (grant no. 01ER0816). The FoCus study was supported by the BMBF (grant no. 0315540A). The Heinz Nixdorf Recall study was supported by the Heinz Nixdorf Foundation Germany, the BMBF, and the DFG (ER 155/6-1 and ER 155/6-2). The KORA study was initiated and financed by the Helmholtz Zentrum München–German Research Center for Environmental Health, which is funded by the BMBF and the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The PopGen 2.0 network is supported by a grant from the BMBF (grant no. 01EY1103). SHIP is part of the Community Medicine Research Network of the University Medicine Greifswald (www.community-medicine.de), which was initiated and funded by the BMBF and the State of Mecklenburg-Pomerania; genome-wide data have been supported by the BMBF (grant no. 03ZIK012). **Author contributions:** Conception and design: T.F.M.A., D.B., A.Z., B.H., and B.M.-M.; recruitment of German cases: D.B., G.A., A. Bayas, L. Bechmann, A. Berthele, A.C., C. Gasperi, R.G., C. Graetz, J.H., M.H., C.I.-D., M.K., T. Kümpfel, V. Limmroth, R.A.L., V. Loleit, F.L., S.G. Meuth, M.M., S.N., F.P., M. Pütz, T.R., A.S., M. Stangel, J.-P.S., K.H. Stürner, B.T., F.T.B., H.T., C.W., F.W., H.W., B.W., U.K. Zettl, U. Ziemann, F.Z., and B.H.; acquisition of genotype or expression data of German cases (DE1): P.W., M.R.-B., P.L., T.B., and F.W.; German control cohorts: K.B., L. Bertram, A.F., C. Gieger, S.H., G.H., M.I., K.-H.J., S.K., T. Kacprowski, M.L., W.L., C.L., S.L., T.M., S. Moebus, M.M.-N, M.N., A.P., R.R., K. Strauch, U.S., H.V., M.W., and J.W.; provided replication data (Sardinian cohort): A.M., M. Pitzalis, E.P., C.S., I.Z., F.C., and M.Z.; analysis and interpretation of data: T.F.M.A., D.B., J.A., M.O.S., T.D., A.Z., D.C., T.C.-R., E.B., B.H., and B.M.-M.; drafting or revising the manuscript: T.F.M.A., D.B., F.C., M.Z., B.H., and B.M.-M. **Competing interests:** A.Bayas received honoraria for consultancy and/or as speaker from Merck Serono, Biogen, Bayer Vital, Novartis, Teva, Roche, and Sanofi/Genzyme and for trial activities from Biogen, Merck Serono, and Novartis; he received grants for congress trips and participation from Biogen, Novartis, Sanofi/Genzyme, and Merck Serono. F.T.B. received funding from the DFG and BMBF; he received, through his institution, research support for investigator-initiated studies from Actelion, Bayer Schering, Novartis, and Teva; he has served on scientific advisory boards for Novartis, Sanofi/Genzyme, and Teva; he has received support to attend scientific meetings from Actelion, Bayer Schering, Biogen Idec, Sanofi/Genzyme, Merck-Serono, Novartis, and Teva; he has received personal honoraria for speaking from Bayer Schering, Biogen Idec, CSL Behring, Sanofi/Genzyme, Merck-Serono, and Teva. A.Berthele received travel grants, research grants, and speaker honoraria from Bayer Healthcare, Biogen, Merck Serono, Novartis, Teva, and Sanofi. D.B. received compensation for activities with Bayer Healthcare, Biogen, Merck Serono, and Novartis; she is supported by the Abirisk Consortium. A.C. received compensation for activities with Almirall Hermal GmbH, Bayer Schering, Biogen Idec, Merck Serono, Novartis, and Teva Neuroscience; research support from Bayer Schering, Biogen Idec, Merck Serono, and Novartis; and research grants from the BMBF (KKNMS; CONTROL MS, 01GI0914). R.G. received compensation for activities with Bayer Healthcare, Biogen Idec, and Teva Neuroscience; an editorial capacity from Therapeutic Advances in Neurological Disorders; patent payments from Biogen Idec; and research support from Bayer Healthcare, Biogen Idec, Merck Serono, Teva Neuroscience, Novartis, and BMBF (KKNMS; CONTROL MS, 01GI0914). M.H. received speaker honoraria and travel expenses from Bayer Healthcare, Biogen Idec, Novartis, and Teva. B.H. served on scientific advisory boards for Roche, Novartis, Bayer Schering, Merck Serono, Biogen Idec, GSK, Chugai Pharmaceutical, Genentech, and Genzyme Corporation; he serves on the international advisory board of Archives of Neurology and Experimental Neurology; he received speaker honoraria from Bayer Schering, Novartis, Biogen Idec, Merck Serono, Roche, and Teva Pharmaceutical Industries; he received research support from Biogen Idec, Bayer Schering, Merck Serono, Five Prime, Metanomics, Chugai Pharmaceutical, and Novartis. He has filed a patent for the detection of antibodies and T cells against KIR4.1 in a subpopulation of MS patients and genetic determinants of neutralizing antibodies to

interferon-β. M.K. received travel grants for attention of scientific meetings from Genzyme and grant support from Merck Serono. F.L. received travel grants from Teva Pharmaceutical and Merck Serono. S. Meuth received honoraria for lecturing and travel expenses for attending meetings and has received financial research support from Bayer, Bayer Schering, Biogen Idec, Genzyme, Merck Serono, MSD, Novartis, Novo Nordisk, Sanofi-Aventis, and Teva. M.M. received research support from BMBF, DFG, Hertie Foundation, Merck Serono, and Novartis and travel expenses for attending meetings from Bayer and Merck Serono; he received honoraria for lecturing from Merck Serono, and investigator fees for a phase 3 clinical study from Biogen Idec. S.N. received travel grants for attention of scientific meetings from Merck Serono, Biogen Idec, and Teva and grant support from Bayer Schering and Novartis. T.R. received travel expenses and financial research support from Genzyme and honoraria for lecturing from Teva and Genzyme. A.S. received personal compensation for activities with Novartis, Sanofi, and Almirall Hermal GmbH. U.S.'s research activities were funded by the National Institute of Neurological Disorders and Stroke (subaward no. A08580 M10A10647), the BMBF (grant no. 03IS2061A), the German Federal State of Mecklenburg–West Pomerania, and the Siemens Healthcare. He received travel expenses and/or honoraria for lectures or educational activities not funded by industry. J.-P.S. and K. Stürner received research grants and speaker honoraria from Bayer Healthcare, Biogen, Merck Serono, Novartis, and Sanofi-Aventis. M. Stangel received honoraria for scientific lectures or consultancy from Bayer Healthcare, Biogen Idec, Baxter, CSL Behring, Grifols, Merck Serono, Novartis, Sanofi-Aventis, and Teva. His institution received research support from Bayer Healthcare, Biogen Idec, Merck Serono, Novartis, and Teva. His laboratory has grant support from the DFG, the Ministry of Science and Culture of Lower Saxony [Niedersachsen-Research Network on Neuroinfectiology (N-RENNT)], the BMBF, and the Röver Foundation. B.T. received consultancy and speaker honoraria and/or research grants from Bayer Healthcare, Biogen, CSL Behring, Genzyme, Grifols, Merck Serono, Novartis, Octapharma, Roche, Sanofi-Aventis, and Teva. H.T. received honoraria for speaking/consultation and travel grants from Bayer Healthcare, Biogen Idec, Merck Serono, Genzyme, Novartis Pharma, Siemens Health Products, and Teva Pharmaceutical and research grants from Biogen Idec, Merck Serono, Novartis Pharmaceuticals, Siemens Health Products, Teva Pharmaceutical, and the BMBF. C.W. received honoraria for participation to advisory boards and/or research funding from Novartis, Bayer, Biogen, and Teva. F.W. received honoraria from Genzyme and Novartis for serving on a scientific advisory board and a travel grant for the attention of a scientific meeting from Merck Serono, Novartis, and Biogen. He received grant support from Merck Serono, Novartis, and the BMBF (projects Biobanking and Omics in CONTROL MS; KKNMS). H.W. received compensation for serving on scientific advisory boards/steering committees for Bayer Healthcare, Biogen Idec, Genzyme, Merck Serono, Novartis, and Sanofi-Aventis. He received speaker honoraria and travel support from Bayer Vital GmbH, Bayer Schering AG, Biogen Idec, CSL Behring, EMD Serono, Fresenius Medical Care, Genzyme, Merck Serono, OmniaMed, Novartis, and Sanofi-Aventis and compensation as a consultant from Biogen Idec, Merck Serono, Novartis, and Sanofi-Aventis. He received research support from Bayer Vital, Biogen Idec, Genzyme, Merck Serono, Novartis, Sanofi-Aventis Germany, and Sanofi U.S. as well as grants and research support from Bayer Healthcare, Biogen Idec, BMBF, DFG, Else Kröner-Fresenius Foundation, Fresenius Foundation, Hertie Foundation, Merck Serono, Novartis, NRW Ministry of Education and Research, Interdisciplinary Center for Clinical Studies (IZKF) Münster, RE Children's Foundation, Sanofi-Aventis/Genzyme, and Teva Pharmaceutical. B.W. received honoraria for speaking/consultation and travel grants from Bayer Healthcare, Biogen Idec, Merck Serono, Genzyme, a Sanofi Company, Novartis Pharmaceuticals, and Teva Pharmaceutical GmbH and research grants from Biogen Idec, Biotest, Merck Serono, Novartis Pharmaceuticals, Teva Pharmaceutical, the BMBF, and the Dietmar Hopp Foundation. U. Ziemann received honoraria for speaking/consultation from Bayer Healthcare, Biogen Idec, Bristol-Myers Squibb, Cortec, Medtronic GmbH, Servier; and research funding from Biogen Idec. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 20 November 2015
Accepted 27 April 2016
Published 17 June 2016
10.1126/sciadv.1501678

Citation: T. F. M. Andlauer, D. Buck, G. Antony, A. Bayas, L. Bechmann, A. Berthele, A. Chan, C. Gasperi, R. Gold, C. Graetz, J. Haas, M. Hecker, C. Infante-Duarte, M. Knop, T. Kümpfel, V. Limmroth, R. A. Linker, V. Loleit, F. Luessi, S. G. Meuth, M. Mühlau, S. Nischwitz, F. Paul, M. Pütz, T. Ruck, A. Salmen, M. Stangel, J.-P. Stellmann, K. H. Stürner, B. Tackenberg, F. Then Bergh, H. Tumani, C. Warnke, F. Weber, H. Wiendl, B. Wildemann, U. K. Zettl, U. Ziemann, F. Zipp, J. Arloth, P. Weber, M. Radivojkov-Blagojevic, M. O. Scheinhardt, T. Dankowski, T. Bettecken, P. Lichtner, D. Czamara, T. Carrillo-Roa, E. B. Binder, K. Berger, L. Bertram, A. Franke, C. Gieger, S. Herms, G. Homuth, M. Ising, K.-H. Jöckel, T. Kacprowski, S. Kloiber, M. Laudes, W. Lieb, C. M. Lill, S. Lucae, T. Meitinger, S. Moebus, M. Müller-Nurasyid, M. M. Nöthen, A. Petersmann, R. Rawal, U. Schminke, K. Strauch, H. Völzke, M. Waldenberger, J. Wellmann, E. Porcu, A. Mulas, M. Pitzalis, C. Sidore, I. Zara, F. Cucca, M. Zoledziewska, A. Ziegler, B. Hemmer, B. Müller-Myhsok, Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci. Adv.* **2**, e1501678 (2016).

frontiers in Immunology

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/fimmu.2020.02165

Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis Is Linked to Human Multiple Sclerosis Risk Genes

Hans Faber[1†‡], Dunja Kurtoic[1‡], Gurumoorthy Krishnamoorthy[2], Peter Weber[1], Benno Pütz[1], Bertram Müller-Myhsok[1,3], Frank Weber[1†‡] and Till F. M. Andlauer[1,4*‡]*

[1] Max Planck Institute of Psychiatry, Munich, Germany, [2] Max Planck Institute of Biochemistry, Martinsried, Germany, [3] Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom, [4] Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

Recent genome-wide association studies have identified over 230 genetic risk loci for multiple sclerosis. Current experimental autoimmune encephalomyelitis (EAE) models requiring active induction of disease may not be optimally suited for the characterization of the function of these genes. We have thus used gene expression profiling to study whether spontaneous opticospinal EAE (OSE) or MOG-induced EAE mirrors the genetic contribution to the pathogenesis of multiple sclerosis more faithfully. To this end, we compared gene expression in OSE and MOG EAE models and analyzed the relationship of both models to human multiple sclerosis risk genes and T helper cell biology. We observed stronger gene expression changes and an involvement of more pathways of the adaptive immune system in OSE than MOG EAE. Furthermore, we demonstrated a more extensive enrichment of human MS risk genes among transcripts differentially expressed in OSE than was the case for MOG EAE. Transcripts differentially expressed only in diseased OSE mice but not in MOG EAE were significantly enriched for T helper cell-specific transcripts. These transcripts are part of immune-regulatory pathways. The activation of the adaptive immune system and the enrichment of both human multiple sclerosis risk genes and T helper cell-specific transcripts were also observed in OSE mice showing only mild disease signs. These expression changes may, therefore, be indicative of processes at disease onset. In summary, more human multiple sclerosis risk genes were differentially expressed in OSE than was observed for MOG EAE, especially in T_H1 cells. When studying the functional role of multiple sclerosis risk genes and pathways during disease onset and their interactions with the environment, spontaneous OSE may thus show advantages over MOG-induced EAE.

Keywords: experimental autoimmune encephalomyelitis (EAE), myelin oligodendrocyte glycoprotein (MOG), T helper cell (Th), multiple sclerosis, risk genes, gene expression

OPEN ACCESS

Edited by:
Marcello Moccia,
University of Naples Federico II, Italy

Reviewed by:
Georgina Xanthou,
Biomedical Research Foundation of the Academy of Athens (BRFAA), Greece
Richard Milner,
San Diego Biomedical Research Institute, United States

****Correspondence:***
Bertram Müller-Myhsok
bmm@psych.mpg.de
Till F. M. Andlauer
till.andlauer@tum.de

†Present address:
Hans Faber,
Schön Klinik Bad Aibling Harthausen, Bad Aibling, Germany
Frank Weber,
Neurologische Klinik, Sana Kliniken des Landkreises Cham GmbH, Cham, Germany

‡These authors have contributed equally to this work

Specialty section:
This article was submitted to Multiple Sclerosis and Neuroimmunology, a section of the journal Frontiers in Immunology

Received: *21 May 2020*
Accepted: *10 August 2020*
Published: *18 September 2020*

Citation:
Faber H, Kurtoic D, Krishnamoorthy G, Weber P, Pütz B, Müller-Myhsok B, Weber F and Andlauer TFM (2020) Gene Expression in Spontaneous Experimental Autoimmune Encephalomyelitis Is Linked to Human Multiple Sclerosis Risk Genes. Front. Immunol. 11:2165. doi: 10.3389/fimmu.2020.02165

INTRODUCTION

Although animal models are widely used in human research, it is still discussed whether they can adequately mirror diseases like multiple sclerosis (MS) that only exist in humans. MS is a chronic inflammatory disease of the central nervous system (CNS), with both environmental and genetic risk factors contributing to disease susceptibility. The recent identification of more than 230

genetic risk loci for MS (1, 2) requires a reassessment of the widely used experimental autoimmune encephalomyelitis (EAE) animal models. To support analyses of the primary cause and etiology of MS, animal models should ideally replicate mechanisms taking place during MS disease induction.

Most EAE models are actively induced by injection of myelin-derived antigens in conjunction with potent adjuvants (3). One such antigen is myelin oligodendrocyte glycoprotein (MOG), a component of the outer surface of myelin (4). Injection of the MOG_{35-55} peptide into *C57BL/6* mice leads to chronic EAE (5) and thus serves as a popular animal model to date. A related model, passively-transferred EAE, is caused by bulk transfer of *in vitro*-activated myelin-specific T cells (6).

By contrast, transgenic models such as opticospinal EAE (OSE) spontaneously develop autoimmune disease and may, therefore, be better suited to study disease onset than induced EAE is. Spontaneous models can be used for identifying environmental triggers of MS (7, 8) and might support analyses of genetic risk factors for human MS. They circumvent problems specific to induced ones, such as adjuvant inoculation, with its partially unknown effects. In OSE, ∼50% of the animals develop a spontaneous inflammatory demyelinating CNS disease, predominantly affecting optic nerves and the lumbar part of the spinal cord (9). These mice carry two transgenic modifications: they express a T cell receptor (TCR) recognizing the MOG_{35-55} peptide and B cells with MOG-specific receptors. In OSE, MOG-specific B cells function as antigen-presenting cells to trigger disease onset by activating MOG-specific T cells (10). Notably, B cell-depleting treatments for MS appear to target primarily cellular and not humoral B cell responses, and, thus, result in a reduced T cell activation (11).

For a long time, T_H1 cells were considered as the predominant drivers of EAE and MS (4). This hypothesis was challenged by emerging evidence for a substantial role of T_H17 cells in the disease etiology, including the discovery that the transfer of T_H17 cells can induce EAE. In fact, both T_H cell types can induce EAE, albeit with distinct pathologies (12). In humans, genome-wide association studies (GWAS) have identified many MS risk loci that support a central role of T_H cells and T_H cell differentiation in the pathophysiology of MS (1, 2, 13).

Despite their valuable contributions to our understanding of MS pathophysiology and drug development, the relationship of EAE to human MS remains controversial (14). All available EAE models are, to some degree, artificial. Therefore, knowledge of whether gene expression changes in diseased mice involve MS risk genes can support the choice of an EAE model for specific research projects. The present study had three aims: First, to characterize gene expression differences in diseased OSE and MOG_{35-55} EAE mice, two widely used EAE models with markedly different forms of induction. Second, to explore which of OSE or MOG-induced EAE resembles human MS more closely. To this end, we examined to which degree genes differentially expressed in spinal cord samples of OSE and MOG EAE showed significant enrichment of human MS risk genes. Third, to analyze expression differences of T_H cell-specific transcripts in both EAE models.

MATERIALS AND METHODS

Mice, Animal Handling, and Scoring

All mice used in this study had a *C57BL/6* background and were bred in the animal facilities of the Max Planck Institute of Biochemistry and Neurobiology, Martinsried, Germany. For the OSE model, double-transgenic *2D2* (TCR_{MOG}) × IgH_{MOG} (OSE) mice were used (9). For MOG EAE, wildtype *C57BL/6* mice were immunized subcutaneously with 200 μg of a MOG peptide consisting of the amino acids 35–55, emulsified in complete Freund's adjuvant supplemented with 5 mg/ml *Mycobacterium tuberculosis* (strain H37Ra, Thermo Fischer Scientific BD Difco), as described previously (9). Pertussis toxin (400 ng, List Biological Laboratories) was injected intraperitoneally on the day of immunization and 48 h later. Control mice (CFA) received the same treatment but without the MOG peptide. For the analysis of EAE models, only female mice were used. For the T_H cell analyses, OSE mice of mixed gender were used (15).

Scores for clinical signs of EAE were assessed daily according to the standard 5-point scale (9, 16): 0: healthy animal; 1: animal with a flaccid tail; 2: animal with impaired righting reflex and/or gait; 3: animal with one paralyzed hind leg; 4: animal with both hind legs paralyzed; 5: moribund animal or death of the animal after preceding clinical disease. See **Supplementary Figure 1** for the disease course of MOG EAE compared to control mice. Following our ethically approved protocol, the mice were sacrificed when they reached a score of 4. The animal welfare committee of the government of Upper Bavaria (Tierschutzkommission der Regierung von Oberbayern, Munich, Germany) approved the protocol. The animal procedures were in strict accordance with the guidelines set down by the animal welfare committee of the government of Upper Bavaria.

In vitro $CD4^+$ T Cell Differentiation

T cells derived from the spleen of a mixed-gender pool of four OSE mice were used to polarize pathogenic effector T_H1 and T_H17 cells, as described previously (15). In brief, four separate batches of four mice each were used for this experiment. To generate T_H1 cells, total erythrocyte-lysed spleen cells from OSE mice were cultured in the presence of a MOG peptide (amino acids 1–125), IL-12, IL-18, and anti-IL-4. After 3 days, IL-2 was added to the culture. To generate T_H17 cells, total erythrocyte-lysed spleen cells from OSE mice were cultured in the presence of a MOG peptide (amino acids 1–125), TGF-β1, IL-6, IL-23, anti-IL-4, and anti-IFN-γ. After 3 days, IL-23 was added to the culture. In both cases, cells were re-stimulated after 6 days and harvested after 9 and, once more, after 12 days. Naïve T_H0 cells were harvested on day 0. The success of polarization was evaluated by flow cytometry, ELISA, and quantitative real-time PCR (**Supplementary Figure 2** and **Supplementary Methods**).

Microarrays

Two separate microarray experiments were performed on the Illumina gene expression profiling platform: The first comprised RNA isolated from total spinal cord preparations of healthy and diseased EAE mice. The second experiment analyzed gene

expression profiles of naïve T_H0 cells and *in vitro* polarized T_H1 and T_H17 cells. For the analysis of EAE models, the Sentrix BeadChip Array MouseWG-6 v2 (Illumina, San Diego, USA) was used; for the T_H cell microarray, the Sentrix BeadChip Array MouseWG-6 v1.1 (Illumina, San Diego, USA). Four chips (24 samples, four per experimental group) were hybridized in the EAE experiment, three chips (18 samples from four separate experiments: $4 \times T_H0$, $7 \times T_H1$, $7 \times T_H17$) were used for the T_H cell analysis. In each experiment, all samples and chips were processed in parallel. RNA processing, array hybridization, and quantification followed the same protocols in both experiments: First, concentration and purity of total RNA were assessed by 260 nm UV absorption and by 260/280 ratios, respectively (Nanophotometer, Implen, Munich, Germany). Second, RNA integrity was evaluated using a chip-based electrophoretic assay (Agilent RNA 6000 Nano Kit used in conjunction with the Agilent 2100 bioanalyzer, Agilent Technologies, Waldbronn, Germany). Mean RNA integrity numbers were 8.4 (SD = 0.5) for the EAE and 9.0 (SD = 0.5) for the T_H cell experiment. Third, RNA was amplified and labeled using the Illumina TotalPrep RNA Amplification Kit (Ambion, Houston, TX, USA) and hybridized onto Illumina gene expression arrays following the manufacturer's instructions. Fourth, fluorescence signals were scanned on an Illumina BeadStation and analyzed by in-house software routines. The manufacturer's built-in controls were analyzed, including hybridization controls and sample-dependent parameters. All microarrays fulfilled Illumina's recommendations for quality control (QC).

Quality Control of Microarrays

Raw probe intensities were exported as summary data using Illumina's GenomeStudio, and further statistical processing was carried out using *R* v3.3.2 (17). For the analysis of EAE models, summary data was loaded using the Bioconductor package *beadarray* (18), and QC was conducted with *lumi* (19) and *vsn* (20). Each probe was transformed and normalized through variance stabilization and normalization. Probes were removed if they showed a detection p-value < 0.05 in >10% of the samples or had a "no match" or "bad" probe quality in the *illuminaMousev2.db* package. This procedure left 21,483 transcripts from 24 samples. For the T_H cell experiment, summary data was loaded using *limma* (21) and QC was conducted with *limma* and *vsn*. Probes were transformed, normalized, and filtered as described above, based on the *illuminaMousev1p1.db* package. This pipeline left 17,858 transcripts. Technical batch effects were examined by inspecting the association of the first ten principal components of expression levels with expression chip and position on the chip.

Analysis of Differential Expression

Principal component analysis (PCA) was conducted in *R* using the function *prcomp* without scaling of variables; PCs were scaled for display. K-means clustering was performed using *kmeans* with $k = 4$; the analysis was repeated 100 times and the most stable clustering solution was selected. Differential expression was assessed with *limma*. For the analysis of differential expression across the EAE models, six mouse types were examined (with four mice each): wild-type (WT); healthy OSE controls (OSE_0); OSE with disease score 1 (OSE_1); OSE score 4 (OSE_4); as a MOG EAE control, healthy control mice injected with complete Freund's adjuvant but not with a MOG peptide (CFA); as MOG_{35-55} EAE, *C57BL/6* wildtype mice injected with adjuvant and MOG_{35-55} peptide, rated score 4 (MOG_4). The design matrix was constructed from the six mouse types. Each expression chip contained one sample per mouse type. The four chips were added to the model as random effects via the *duplicateCorrelation* function. The five contrasts MOG_4-CFA, CFA-WT, OSE_4-OSE_0, OSE_1-OSE_0, and OSE_4-WT were computed on the fitted linear model and moderated t-tests were calculated using the *eBayes* function. For the T_H cell experiment, the design matrix was constructed from the three cell types (naïve T_H0, T_H1, T_H17), with the four mouse pools treated as random effects. Only T_H1 and T_H17 cells harvested on day 9 were analyzed. The two contrasts T_H1-T_H0 and T_H17-T_H0 were examined.

Overrepresentation Analyses

Overrepresentation analyses (ORA) were conducted using WebGestalt v2019 (22) in *R*, based on the gene ontology (GO) biological process database. Genes were submitted as unique Entrez IDs, and the reference was *genome protein-coding*. The significance level was determined using a hypergeometric test, followed by calculation of the Benjamini-Hochberg false discovery rate (FDR) (23).

Enrichment Tests

The enrichment of genes was calculated using a permutation test in *R*. For this test, the unique Entrez IDs of genes were used. First, the amount of unique differentially expressed genes was determined, and the same number of random genes was selected. Second, the number of these random genes overlapping with the test set of genes (e.g., MS susceptibility genes) was determined. These two steps were repeated 100,000 times. Third, to calculate a p-value, the number of observations where the overlap between random genes and test genes was equal to or larger than the overlap between differentially expressed genes and test genes was counted and divided by the number of permutations.

For the enrichment analysis with MS susceptibility genes, the 558 genes outside the major histocompatibility complex (MHC) region listed in Supplementary Table 18 of the MS genomic map published in 2019 by the IMSGC were used (2). From this list, *CTB-50L17.10*, *RP11-345J4.5*, *JAZF1-AS1*, *ZEB1-AS1*, *GATA3-AS1*, *SSTR5-AS1*, and *RPL34-AS1* were excluded to generate the list of 551 prioritized putative MS susceptibility genes described in the IMSGC publication.

RESULTS

We compared gene expression profiles of total spinal cord preparations derived from two EAE models, OSE and MOG_{35-55} EAE. Double-transgenic OSE mice developed CNS autoimmunity spontaneously, predominantly affecting the lumbar part of the spinal cord. In the MOG_{35-55} EAE model, the disease was induced in *C57BL/6* wildtype (WT)

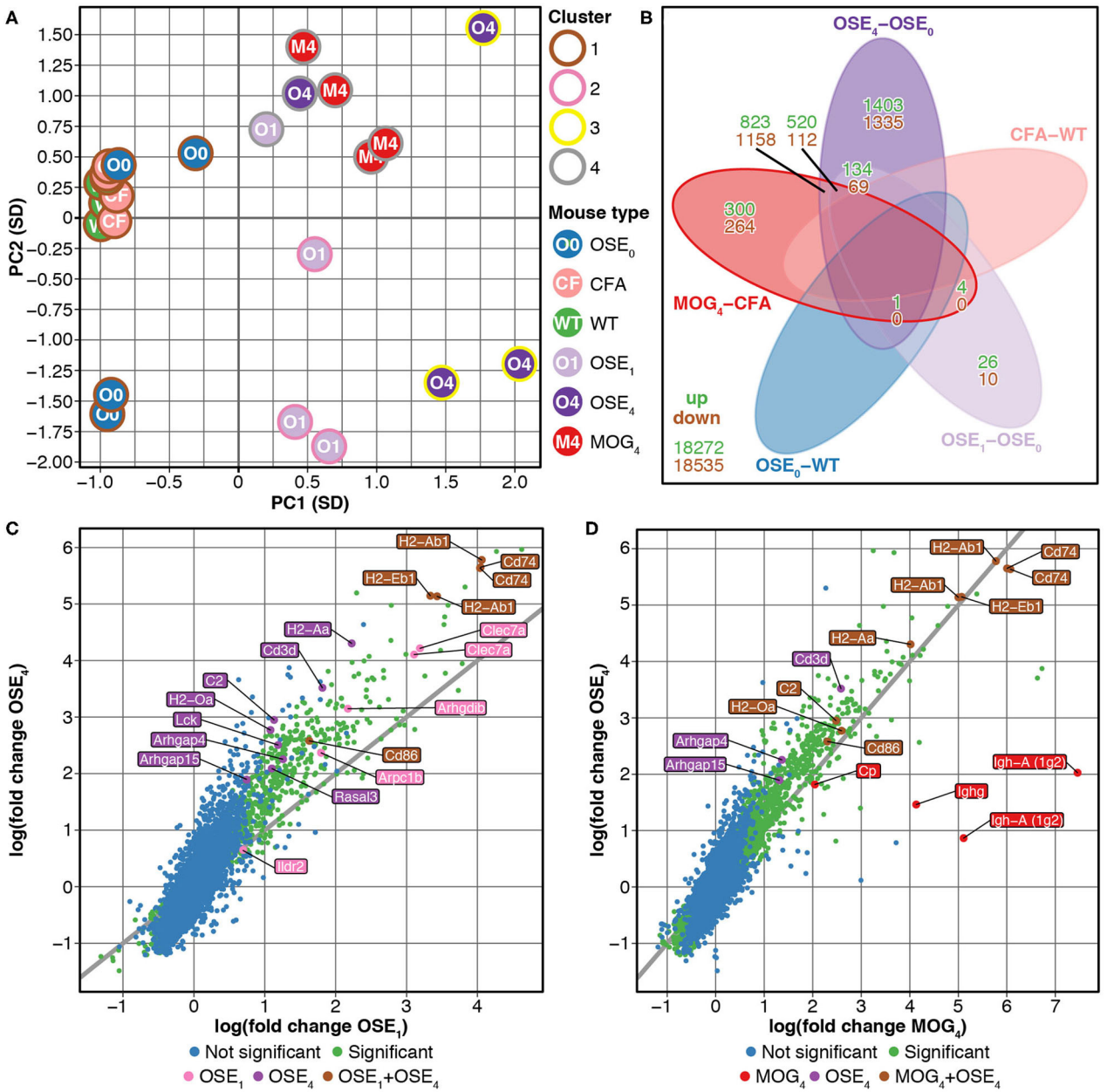


FIGURE 1 | Differential expression analysis of OSE and MOG EAE models. **(A)** Principal component analysis (PCA) of gene expression profiles grouped by k-means clustering. The cluster solution displayed here was the most frequent (34 of 100). In 97 of 100 analyses, all WT, OSE_0, and CFA mice were placed into separate clusters than OSE_1, OSE_4, and MOG_4 mice (**Supplementary Table 1**). Because of the spontaneous nature of disease development in OSE mice, gene expression in diseased OSE_4 animals showed more variance than was, e.g., observed in MOG_4 mice, which exhibit a more stereotypic disease course (9). PC, principal component; SD, standard deviations. **(B)** Venn diagram highlighting the number of transcripts differentially expressed in the analyzed contrasts. For this plot, up- and downregulated transcripts were analyzed separately, and transcripts differentially expressed in opposing directions are therefore included in the counts. **(C,D)** OSE_4 mice showed greater fold changes of expression levels than **(C)** OSE_1 and **(D)** MOG_4 mice, each compared to its control condition (OSE_0 and CFA, respectively) (**Supplementary Table 3**). For each group, the ten most differentially expressed genes (with Entrez IDs) are labeled. If two probes per gene were present among the top differentially expressed transcripts, the gene was counted only once, but both probes are plotted. The groups are: differentially expressed in OSE_1 only (light magenta), differentially expressed in OSE_4 only (dark magenta), differentially expressed in MOG_4 only (red), differentially expressed in **(C)** both OSE_1 and OSE_4 or **(D)** both MOG_4 and OSE_4 (brown; with higher expression levels observed for OSE_4).

mice by immunization with a MOG_{35-55} peptide. PCA of gene expression profiles separated healthy (OSE_0, CFA, and WT) from diseased [OSE score 1 (OSE_1), OSE score 4 (OSE_4), and MOG score 4 (MOG_4)] animals along the first component (**Figure 1A**). Most variance in gene expression was thus observed between healthy and diseased mice and not between EAE models. Because

of the spontaneous nature of disease development in OSE mice, gene expression in diseased OSE_4 animals showed more variance than was, e.g., observed in MOG_4 mice, which exhibit a more stereotypic disease course (9).

Unsupervised k-means clustering on PCs further supported this finding, which consistently (in 97 of 100 replications) placed healthy and diseased animals into separate clusters (**Figure 1A**, **Supplementary Table 1**). The most frequent cluster solution (34/100) placed all healthy mice together in cluster 1; additional clusters were OSE_1 only (cluster 2), OSE_4 only (cluster 3), and a mixed cluster of the remaining diseased animals (cluster 4). We could thus successfully detect disease-relevant gene expression changes in the animals.

Stronger Gene Expression Changes in the OSE Model

Next, we analyzed gene expression changes in OSE and MOG EAE mice. We examined differential expression for five contrasts: OSE_1-OSE_0, OSE_4-OSE_0, MOG_4-CFA, and the two control contrasts CFA-WT and OSE_0-WT (**Figure 1B**, **Supplementary Table 2**). In the control contrast CFA-WT, no transcript was differentially regulated. A single transcript was upregulated in OSE_0-WT, *T cell receptor alpha chain* (*Tcra*), which was also upregulated in all other contrasts except CFA-WT. The number of significantly up- and downregulated transcripts was higher for OSE_4-OSE_0 ($n = 5{,}555$) than for MOG_4-CFA ($n = 3{,}182$). In total, the expression of 864 transcripts differed significantly between MOG_4 and OSE_4 mice (**Supplementary Table 2**). Interestingly, 4.88× more transcripts were differentially expressed specifically only in OSE_4-OSE_0 than only in MOG_4-CFA (**Figure 1B**). Moreover, fold changes were higher in OSE_4-OSE_0 than in either OSE_1-OSE_0 (binomial test: $p = 1.4 \times 10^{-65}$ for all transcripts, $p = 9.9 \times 10^{-119}$ for transcripts differentially expressed in both contrasts, **Figure 1C**) or MOG_4-CFA ($p = 5.8 \times 10^{-3}$ for all, $p = 2.7 \times 10^{-221}$ for differentially expressed transcripts, **Figure 1D**; **Supplementary Table 3**). Stronger global gene expression changes were thus triggered in OSE than in MOG EAE.

Overrepresentation of Immune System Processes Especially for OSE

To characterize the expression changes in the different EAE models further, we conducted ORA analyses of the analyzed contrasts (**Supplementary Table 4**, **Supplementary Figure 3**) and of differentially expressed transcripts for three groups (**Supplementary Figure 4**): First, *common disease transcripts* (CDT), differentially expressed for both contrasts OSE_4-OSE_0 and MOG_4-CFA but not in the two control contrasts OSE_0-WT or CFA-WT. Second, *OSE_4-specific transcripts* (OSE_4sp), differentially expressed for the contrast OSE_4-OSE_0 but not for MOG_4-CFA or the control contrasts. Third, *MOG_4-specific transcripts* (MOG_4sp), differentially expressed for MOG_4-CFA but not for OSE_4-OSE_0 or the control contrasts. When examining CDT, 1,379 redundant GO biological processes remained significant after correction for multiple testing (**Supplementary Table 4**). Together with other immune-related gene sets, *immune response*, *regulation of immune system process*, and *T cell activation* were among the top-associated terms (adjusted $p < 2 \times 10^{-16}$). These and other immune-associated processes remained significant in OSE_4sp (adjusted $p \leq 3.5 \times 10^{-2}$, **Supplementary Figure 4**). By contrast, no immune system-specific process was significant for MOG_4sp. More expression changes in the immune system were, therefore, triggered in OSE than in MOG EAE.

Activation of the Adaptive Immune System in OSE_1 Mice

While MOG EAE develops rapidly in a highly stereotypical manner, the clinical course of OSE is usually slower and shows more inter-individual variability (24). OSE thus allows for studying disease at different stages, and we analyzed mice showing a mild disease score of 1 (OSE_1). Compared to OSE_0, 34 transcripts were differentially expressed specifically in OSE_1 animals and not in any other contrast [*OSE_1-specific transcripts* (OSE_1sp), **Supplementary Table 5**]. These transcripts are potentially indicative of changes during mild or early disease. However, no significant GO biological processes were identified for them. Transcripts differentially regulated in both OSE_1 and OSE_4 consistently showed the same direction of regulation compared to OSE_0 [binomial test $p = 4.36 \times 10^{-252}$, 95% confidence interval (CI) 0.995–1.0, **Supplementary Table 3**]. When analyzing all transcripts differentially expressed in OSE_1-OSE_0 but not in control contrasts [*OSE_1-expressed transcripts* (OSE_1ex), **Supplementary Table 5**], 805 processes were significant after correction for multiple gene sets. Among them were the three previously highlighted GO terms (adjusted $p < 2 \times 10^{-16}$, **Supplementary Table 4**, **Supplementary Figure 4**). Furthermore, the gene sets *B cell mediated immunity* and *antigen processing and presentation* were significantly overrepresented not only in the analysis of CDT but also for the OSE_1ex transcripts, indicating a potential role of B cells also in mildly affected OSE mice.

Enrichment of MS Susceptibility Genes Among Transcripts Expressed in OSE

Over 230 independent genetic loci associated with MS susceptibility in humans have been identified (1, 2). Based on these GWAS loci, 551 human MS susceptibility candidate genes have been proposed (2), for which expression data of 499 transcripts were available in our dataset. We conducted a PCA on these transcripts (265 genes) to analyze whether the expression of MS risk genes was increased in the EAE models. The first component, explaining 75.7% of the variance in expression of these transcripts, was significantly higher in all disease groups than in controls, indicating high expression levels of MS-associated genes in EAE, with the highest levels observed for OSE_4 (**Figure 2A**, **Supplementary Table 6**). Also individual MS risk genes, e.g., *H2-Ab1*, *Cd52*, and *Cd86* (1, 2), as well as further putative MS-associated genes like *Cd74*, were among the transcripts showing the lowest differential expression

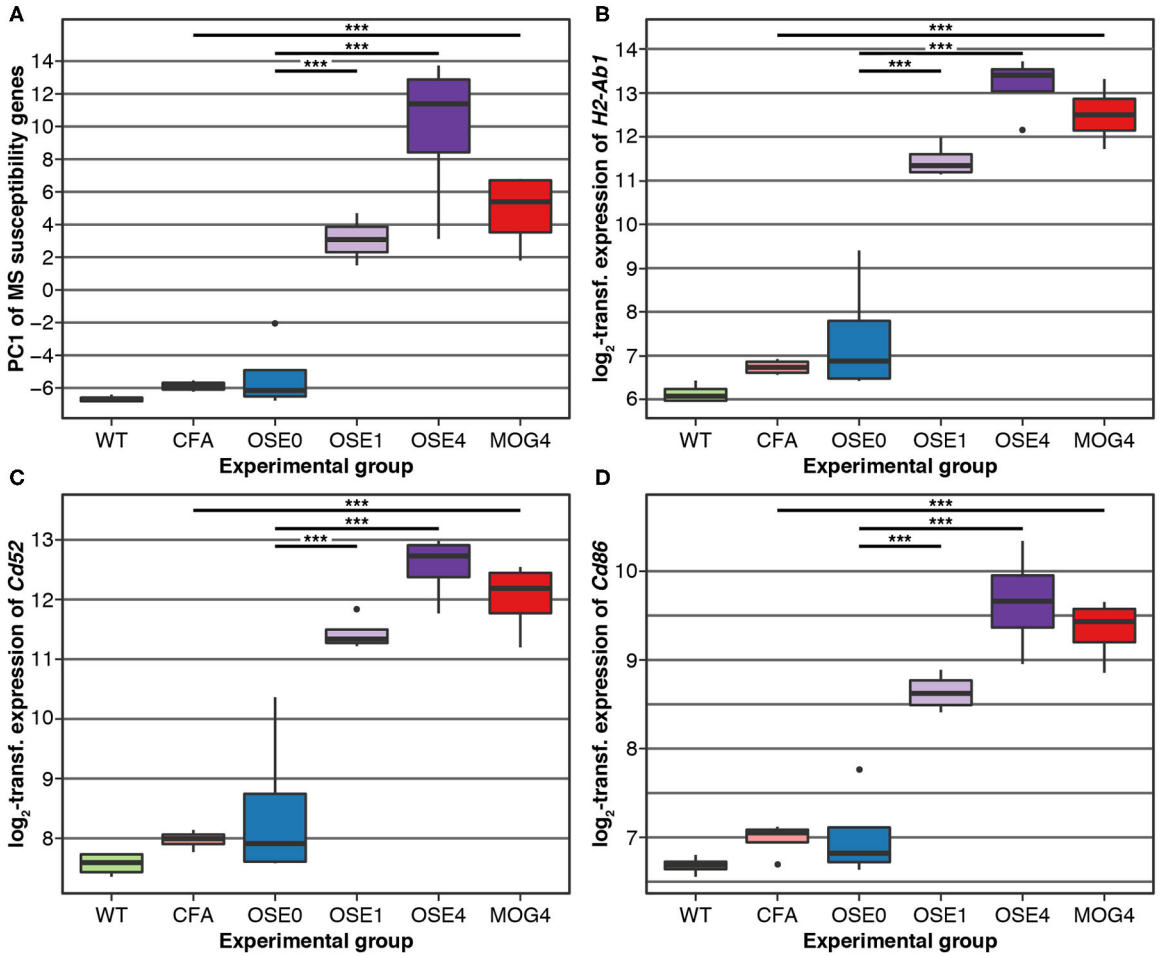


FIGURE 2 | MS risk genes show a higher expression in diseased EAE animals. **(A)** Principal component analysis (PCA) of gene expression profiles of putative MS risk genes. Diseased mice showed higher MS risk gene expression levels (**Supplementary Table 6**). PC, principal component; y-axis unit, standard deviations. **(B–D)** Examples of expression levels of three putative MS risk genes, *H2-Ab1*, *Cd52*, and *Cd86* (**Supplementary Table 2**). In all three cases, diseased mice showed significantly higher expression levels, with the highest expression observed for OSE_4 mice. Significance levels: *** adjusted $p < 0.001$.

TABLE 1 | Enrichment of MS susceptibility genes.

DE transcript group	DE genes	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	2,014	68	**$<1 \times 10^{-5}$**	**$<4 \times 10^{-5}$**
OSE_4sp	2,362	68	**4.4×10^{-4}**	**8.8×10^{-4}**
MOG_4sp	469	11	3.2×10^{-1}	3.2×10^{-1}
OSE_1ex	693	34	**1.0×10^{-5}**	**4.0×10^{-5}**

Of 551 genes considered, 265 were present in our data. P-values were computed using 100,000 permutations. Enrichments significant after Holm-Bonferroni correction for multiple testing (four tests) are highlighted in bold font (adjusted p-value < 0.05). DE, differentially expressed; WT, wildtype; CDT, common disease transcripts (differentially expressed for both contrasts OSE_4-OSE_0 and MOG_4-CFA but not in the two control contrasts OSE_0-WT or CFA-WT); OSE_4sp, OSE_4-specific transcripts (differentially expressed for the contrast OSE_4-OSE_0 but not in MOG_4-CFA, OSE_0-WT, or CFA-WT); MOG_4sp, MOG_4-specific transcripts (differentially expressed for the contrast MOG_4-CFA but not in OSE_4-OSE_0, OSE_0-WT, or CFA-WT); OSE_1ex, OSE_1-expressed transcripts (differentially expressed in OSE_1-OSE_0 but not in OSE_0-WT or CFA-WT).

p-values. They were significantly upregulated in all three diseased mouse types (**Figures 1C,D**, **2B–D**, **Supplementary Figure 5**, **Supplementary Table 2**). Furthermore, differentially expressed genes from the analysis sets CDT, OSE_4sp, and OSE_1ex were significantly enriched for MS risk genes, while the MOG_4sp genes were not (**Table 1**). OSE might thus be more closely connected to the etiology of human MS than MOG_{35-55} EAE is.

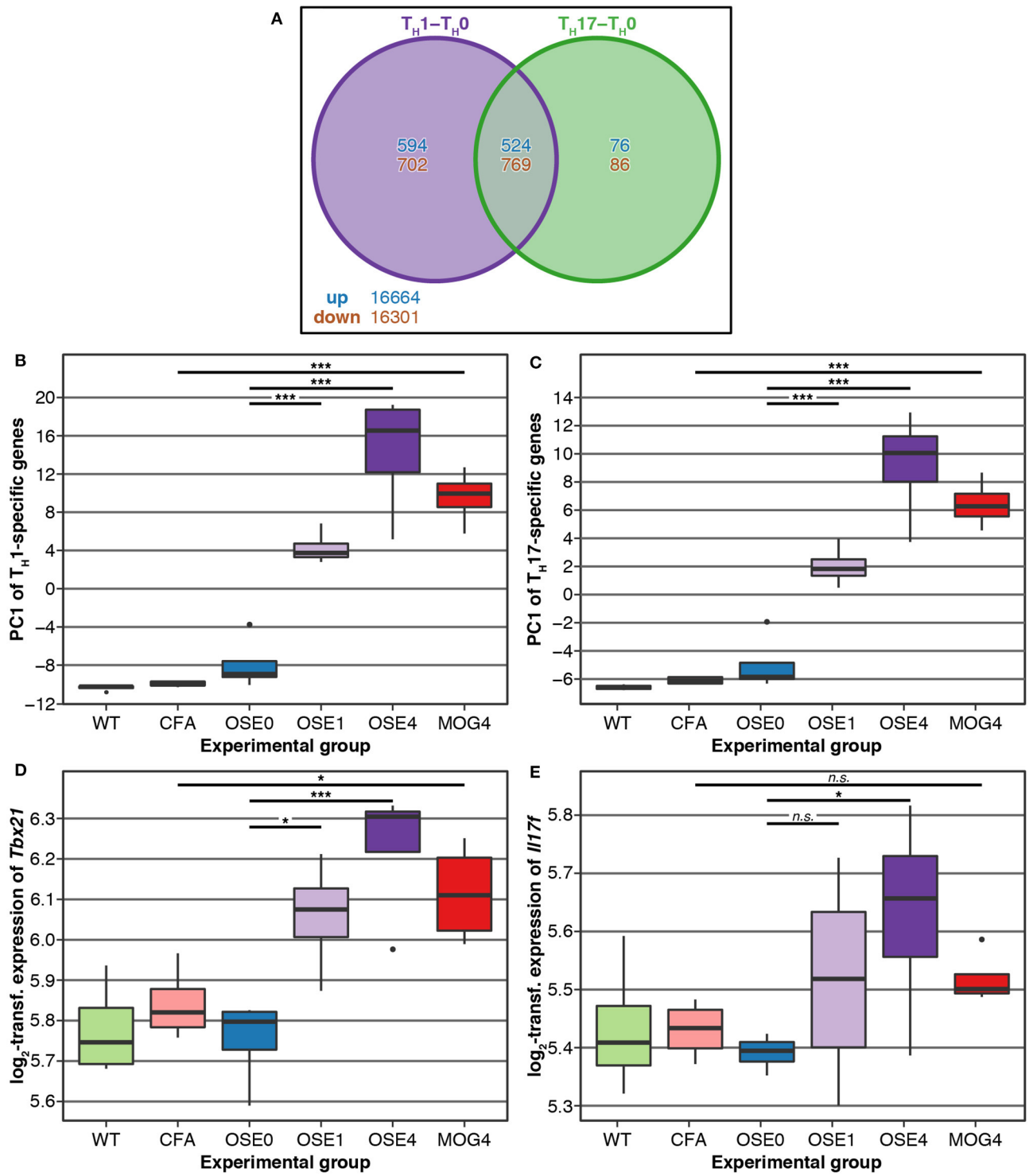


FIGURE 3 | Genes differentially expressed in T_H cells show a higher expression in diseased EAE animals, especially in OSE_4. **(A)** Venn diagram highlighting the number of transcripts differentially expressed in T_H cells. For this plot, up- and downregulated transcripts were analyzed separately, and transcripts differentially expressed in opposing directions are therefore included in the counts. **(B,C)** Principal component analysis (PCA) of gene expression profiles of transcripts differentially expressed in **(B)** T_H1 and **(C)** T_H17 cells. Diseased mice showed higher T_H cell-specific expression levels (**Supplementary Table 7**). PC, principal component; y-axis unit: standard deviations. **(D,E)** Examples of expression levels of the T_H1 signature molecule **(D)** *Tbx21* and the T_H17 signature molecule **(E)** *Il17f*. Diseased mice showed significantly higher expression levels of *Tbx21*, yet only OSE_4 mice showed an increased expression of *Il17f*. Furthermore, the T_H1 signature molecule *Ifng* was expressed significantly higher only in OSE_4; the T_H17 signature molecules *Rorc* and *Il17a* were not differentially expressed in any contrast (**Supplementary Table 8**). Significance levels: n.s.: $p \geq 0.05$, * adjusted $p < 0.05$, *** adjusted $p < 0.001$.

Gene Expression in OSE Overlaps With T_H Cell-Specific Transcripts

T_H cell differentiation was identified as a key pathway in the etiology of MS (13). We, therefore, analyzed whether gene expression changes in EAE models were related to T_H cell differentiation. To this end, gene expression profiling of *in vitro* polarized T_H1 and T_H17 cells was conducted, derived from OSE mice. Compared to naïve T_H0 cells, 8× more transcripts were differentially expressed specifically in T_H1 than in T_H17 cells (**Figure 3A**). None of the transcripts differentially expressed in both T_H1 and T_H17 were regulated in opposite directions.

We examined via PCA whether the expression of T_H1- and T_H17-specific, differentially expressed probes was higher in EAE models than controls. The first component of T_H1-and T_H17-specific gene expression explained 49.6 and 68.6% of the variance, respectively. For both T_H cell types, the first component of cell-specific transcripts was significantly higher in all disease groups than in controls, with the highest levels for OSE_4 (**Figures 3B,C**, **Supplementary Table 7**). Among signature molecules for T_H1 cells, *Tbx21* (*T-bet*) was significantly upregulated in all diseased mice, *Ifng* only in OSE_4 (**Figure 3D**, **Supplementary Table 8**). Of the examined T_H17 markers, only *Il17f* was upregulated in OSE_4, neither *Rorc* nor *Il17a* were differentially expressed (**Figure 3E**, **Supplementary Table 8**).

After correction for multiple testing, the CDT, OSE_4sp, and OSE_1ex analysis sets were significantly enriched for T_H1- and T_H17-specific transcripts (**Table 2**). In the case of MOG_4sp transcripts, the overlap was lower and only significant for T_H1-specific probes. These experiments indicate a stronger overlap of known MS-associated immune responses involving T_H cells with OSE than with MOG EAE.

Finally, we analyzed whether EAE-associated genes differentially expressed in T_H1 or T_H17 cells were more closely connected to human MS. To this end, we intersected the lists of EAE-specific and T_H-specific transcripts. Immune-related biological processes were overrepresented for CDT, OSE_4sp, and OSE_1ex genes intersected with T_H1-specific genes. (**Supplementary Table 9**, **Supplementary Figure 6**). No terms were significantly overrepresented for any T_H17-specific or MOG_4sp genes.

CDT and OSE_1ex genes differentially expressed in T_H1 cells were significantly enriched for the IMSGC MS risk genes ($p < 7 \times 10^{-4}$, **Table 3**). The enrichment for OSE4sp did not withstand correction for multiple testing. Neither any of the T_H17-specific gene sets nor the genes from the MOG_4sp group were enriched for these risk genes. Thus, we conclude that OSE entails gene expression changes involving human MS gene risk genes, especially in T_H1 cells, which were not observed to the same degree for MOG EAE.

DISCUSSION

With the identification of over 230 MS risk loci in recent GWAS, we move closer to understanding the etiology of MS. Further research relies on adequate animal models that have to be reassessed in the context of GWAS data. Given the interplay of genetics and environment in human MS, spontaneous EAE models like OSE might be more apt for studying the genetic risk component of MS than induced EAE models that require active experimental manipulation. In the present study, we performed spinal cord gene expression profiling to, first, characterize differences between spontaneous OSE and MOG-induced EAE and, second, to analyze the relationship of both models to human MS risk genes and T_H cell biology.

OSE May Reflect the Etiology of MS Better Than MOG EAE Does

In comparison to MOG EAE, gene expression changes in OSE were stronger and more closely linked to immune pathways. This might reflect a more complex mode of disease induction in OSE than is the case for MOG EAE. OSE features active B and T cell cooperation, a mechanism highly relevant for the pathophysiology of human MS, as demonstrated by the effectivity of B cell-depleting treatments (10, 11). More than MOG EAE, OSE-specific transcripts were enriched for both human MS risk genes and T_H cell-specific transcripts and showed

TABLE 2 | Enrichment of T_H-specific transcripts.

DE transcript group	DE genes	Cell type	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	2,014	T_H1	150	**$<1 \times 10^{-5}$**	**$<8 \times 10^{-5}$**
		T_H17	28	**2.0×10^{-2}**	**4.0×10^{-2}**
OSE_4sp	2,362	T_H1	195	**$<1 \times 10^{-5}$**	**$<8 \times 10^{-5}$**
		T_H17	36	**2.0×10^{-3}**	**8.0×10^{-3}**
MOG_4sp	469	T_H1	35	**1.1×10^{-2}**	**3.3×10^{-2}**
		T_H17	7	9.8×10^{-2}	9.8×10^{-2}
OSE_1ex	693	T_H1	61	**2.0×10^{-5}**	**1.2×10^{-4}**
		T_H17	16	**1.0×10^{-3}**	**5.0×10^{-3}**

1,080 T_H1- and 145 T_H17-specific transcripts were examined. P-values were computed using 100,000 permutations. Enrichments significant after Holm-Bonferroni correction for multiple testing (eight tests) are highlighted in bold font (adjusted p-value < 0.05). DE, differentially expressed; WT, wildtype; CDT, common disease transcripts (differentially expressed for both contrasts OSE_4-OSE_0 and MOG_4-CFA but not in the two control contrasts OSE_0-WT or CFA-WT); OSE_4sp, OSE_4-specific transcripts (differentially expressed for the contrast OSE_4-OSE_0 but not in MOG_4-CFA, OSE_0-WT, or CFA-WT); MOG_4sp, MOG_4-specific transcripts (differentially expressed for the contrast MOG_4-CFA but not in OSE_4-OSE_0, OSE_0-WT, or CFA-WT); OSE_1ex, OSE_1-expressed transcripts (differentially expressed in OSE_1-OSE_0 but not in OSE_0-WT or CFA-WT).

TABLE 3 | Enrichment of MS susceptibility genes among T_H-specific transcripts.

DE transcript group	Cell type	EAE T_H cell list size	Overlapping genes	*p*-value	Adjusted *p*-value
CDT	T_H1	150	10	**6.5×10^{-4}**	**5.2×10^{-3}**
	T_H17	30	3	2.1×10^{-2}	1.1×10^{-1}
OSE_4sp	T_H1	215	10	9.7×10^{-3}	5.8×10^{-2}
	T_H17	41	3	4.7×10^{-2}	1.6×10^{-1}
MOG_4sp	T_H1	37	1	5.1×10^{-1}	5.1×10^{-1}
	T_H17	7	1	1.3×10^{-1}	2.6×10^{-1}
OSE_1ex	T_H1	60	6	**1.1×10^{-3}**	**7.7×10^{-3}**
	T_H17	16	2	3.9×10^{-2}	1.6×10^{-1}

The p-values were computed using 100,000 permutations. Enrichments significant after Holm-Bonferroni correction for multiple testing (eight tests) are highlighted in bold font (adjusted p-value <0.05). DE, differentially expressed; WT, wildtype; CDT, common disease transcripts (differentially expressed for both contrasts OSE_4-OSE_0 and MOG_4-CFA but not in the two control contrasts OSE_0-WT or CFA-WT); OSE_4sp, OSE_4-specific transcripts (differentially expressed for the contrast OSE_4-OSE_0 but not in MOG_4-CFA, OSE_0-WT, or CFA-WT); MOG_4sp, MOG_4 -specific transcripts (differentially expressed for the contrast MOG_4-CFA but not in OSE_4-OSE_0, OSE_0-WT, or CFA-WT); OSE_1ex, OSE_1 -specific transcripts (differentially expressed in OSE_1-OSE_0 but not in OSE_0-WT or CFA-WT).

an overrepresentation of immune-specific gene sets. We thus hypothesize that OSE shows advantages over MOG EAE in studying the functional role of human MS risk genes and their associated immune pathways.

Nevertheless, many of the differentially expressed genes indicate that both EAE models faithfully recapitulate critical functional pathways of MS, especially regarding the role of antigen presentation and $CD4^+$ T cells in MS immunopathogenesis (25, 26). Transcripts for the *HLA* genes *H2-Eb1* and *H2-Ab1*, homologous to *HLA-DRB5* and *HLA-DQB1*, were among the most differentially expressed probes. The alleles *HLA-DRB5*01:01* and *HLA-DQB1*06:02* are part of the *DR15-DQ6* haplotype and are, most likely because of linkage disequilibrium with *HLA-DRB1*15:01*, strongly associated with MS risk (27). In MS, memory B cells mediate autoproliferation of brain-homing T_H1 cells in a *HLA-DRB1*15:01*-dependent manner (28). Interestingly, the antigen-presenting function of MOG-specific B cells is, in cooperation with T cells, important for the development of OSE (10). Among putative non-MHC MS risk genes (1, 2), *Cd86*, *Cd52*, and *Cd74* showed very robust support for differential expression.

We could thus show that EAE, and in particular OSE, constitutes a valuable model for studying the role of human MS risk genes. Several previous studies support this finding: First, humanized EAE models successfully replicated *HLA*-related risk variants, including *HLA-DRB*15:01* (29). Second, knockout mice lacking the MS-associated *Il7r* are resistant to EAE (30). Third, shared human and EAE risk loci exist that are linked to T_H cell differentiation (31). Fourth, an overlap of upregulated genes between myelin-reactive T cells from MS patients and encephalitogenic $CD4^+$ T cells isolated from EAE was described (32). Fifth, in a passive-transfer EAE study, several MS risk genes were suggested to be implicated in the transition from *in vitro*-generated MOG-specific T_H17 cells to encephalitogenic $CD4^+$ T cells (33).

Functional pathways involving MS risk genes interact with environmental factors to trigger an autoimmune response, as demonstrated by the role of epigenetic factors for MS risk (1, 34). Spontaneous EAE models might resemble gene-environment interactions more faithfully than MOG EAE does. For instance, in a spontaneous EAE model, disease onset could be prevented in mice kept under germ-free conditions (7). In this model, a higher incidence of EAE was observed following the transfer of the human gut microbiome from MS patients than when transferring the microbiome from the patient's healthy twin (8).

T_H1-Specific Transcripts Are Enriched for MS Risk Genes

Our gene set analyses point at a central role of lymphocyte activation in EAE induction and shed light on the ongoing controversy regarding the relative importance of T_H1 and T_H17 cells in mediating CNS autoimmunity (35). In accordance with previous studies (9), we observed a higher differential expression of selected T_H1- than of T_H17-specific transcripts in diseased mice. Interestingly, a high T_H1/T_H17 ratio is indicative of a lesion distribution pattern characterized by prominent spinal cord involvement, as is the case for both EAE models investigated in our study (12, 15, 36).

CDT and OSE_1ex transcripts differentially expressed in T_H1 cells were significantly enriched for MS risk genes (**Table 3**). We did not observe such an enrichment for transcripts differentially expressed in T_H17 cells. Albeit also OSE_4sp genes were only enriched for risk genes in T_H1 cells at nominal significance (unadjusted $p = 0.0097$), T_H1-expressed MOG_4sp transcripts showed no trend for the enrichment of MS risk genes at all (unadjusted $p = 0.51$). In GO overrepresentation analyses, immune-related biological processes like *positive regulation of T cell proliferation* were significant for OSE_4sp-genes differentially expressed in T_H1 cells, but no GO gene sets at all were overrepresented in T_H1-specific MOG_4sp genes. In the context of T_H1-driven immune responses, the OSE model might thus be linked more closely to human MS risk genes than MOG_{35-55} EAE is. However, T_H17 cells can shift toward a T_H1 phenotype in EAE (37, 38). The T_H1 markers analyzed in the EAE models may, accordingly, reflect expression in a significant proportion of former T_H17 cells. Therefore, our findings do not argue against a relevant impact of T_H17 cells in either EAE model.

Expression Patterns Across Different Disease Stages Can Be Studied Using OSE

Most genes differentially expressed in OSE_1 mice were also recapitulated in severely affected OSE_4 mice and showed the same direction of regulation in both disease stages. Many factors active in severe EAE thus also influence EAE during a mild or, potentially, early disease course. Effective immunotherapy is facilitated if the same biological pathways are continuously active throughout the entire disease. For example, the gene set *response to interferon-beta* was highly overrepresented in both OSE_1ex and CDT and *Cd52* was differentially expressed in all diseased mice. Studying mild OSE cases might, therefore, constitute an interesting model for defining the initial triggers of MS and the identification of novel therapeutic options.

LIMITATIONS

Our gene expression analysis of two EAE models had several limitations: First, the microarrays used covered only part of the murine transcriptome and thus, some MS risk genes could not be analyzed. Second, the statistical power of our analyses was restricted by the sample size. Third, the initial phases of EAE are hard to define since the disease develops over a short period. We thus analyzed mild OSE cases as a proxy for early disease. It is, however, unknown whether these animals would have developed more severe EAE later.

CONCLUSIONS

Although hundreds of genetic MS risk loci have been identified, their functional role in the etiology of the disease still has to be resolved. Ideally, suitable animal models recapitulate molecular and functional pathways involving these genes. They may thus move research closer to the primary cause and etiology of MS, thereby supporting the identification of effective immunotherapies. No animal model fully reflects a heterogeneous human disease like MS and each EAE model available today only replicates a part of the human disease. Researchers will thus continue to study different aspects of MS using a variety of EAE models. Our results indicate that OSE, with its closer link to MS risk genes and T_H cell biology, may be better suited for studying the etiology of MS and for defining specific therapeutic targets than MOG-induced EAE is. Future studies will show whether OSE can fulfill this promise to model the human MS genetic risk landscape faithfully.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession number(s) can be found at: https://www.ebi.ac.uk/arrayexpress/, E-MTAB-9132; https://www.ebi.ac.uk/arrayexpress/, E-MTAB-9133.

ETHICS STATEMENT

The animal study was reviewed and approved by Tierschutzkommission der Regierung von Oberbayern, Munich, Germany.

AUTHOR CONTRIBUTIONS

HF, GK, PW, and FW contributed to the original conception and design of the study. HF, GK, and PW conducted experiments. DK and TA devised the statistical analyses. DK, BP, and TA conducted statistical analyses. BM-M and FW supervised the study. HF and TA drafted the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

TA was supported by the German Federal Ministry of Education and Research (BMBF) through the DIFUTURE consortium of the Medical Informatics Initiative Germany (grant 01ZZ1804A) and by the European Union's Horizon 2020 Research and Innovation Programme (grant MultipleMS, EU RIA 733161). TA and BM-M were supported by the BMBF through the Integrated Network IntegraMent, under the auspices of the e:Med Programme (grant 01ZX1614J). GK was supported by the European Research Council starting grant (GAMES; 635617), German research foundation (DFG) SFB TR-128 (Project A1), and by the Max Planck Society.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.02165/full#supplementary-material

Supplementary Figure 1 | Disease course of MOG EAE in *C57BL/6* mice. MOG EAE was induced in *C57BL/6* mice as described in the Methods. The plot shows mean clinical scores ($n = 4$) and the standard error of the mean. The disease score of MOG EAE mice began to increase on day ten. The mice were sacrificed when they reached a score of 4. Control mice (CFA) consistently remained at a disease score of 0.

Supplementary Figure 2 | Efficiency of T_H1 and T_H17 differentiation. **(A)** T cells from OSE mice were activated under T_H1- and T_H17-polarizing conditions and analyzed for intracellular IL-17 and IFN-γ cytokine expression by flow cytometry. The data represent the percentage of cytokine-producing cells in the gated $CD4^+$ populations of naïve (T_H0), T_H1, and T_H17 cells. **(B)** Quantification of IL-17 and IFN-γ in the culture supernatants of T_H1 and T_H17 cells by ELISA. The plots show the mean and the standard error of the mean (SEM). **(C)** IL-17 and IFN-γ mRNA expression of naïve (T_H0), T_H1-, and T_H17-polarized cells quantified by real-time PCR. The data are representative of three independent experiments. The plots show the mean and SEM.

Supplementary Figure 3 | Overrepresented immune system pathways in OSE and MOG EAE contrasts. The plots show overrepresented GO terms that are descendants of the term *Immune System Process* (**Supplementary Table 4**) for the contrasts **(A)** OSE_1-OSE_0, **(B)** OSE_4-OSE_0, **(C)** MOG_4-CFA, and **(D)** MOG_4-OSE_4. The -$\log_{10}$(FDR) from hypergeometric tests is shown on the x-axis and used for coloring the plots (darker colors represent lower FDRs).

Supplementary Figure 4 | The top 40 overrepresented immune system pathways in the differentially expressed transcripts groups. The plots show the top

40 overrepresented GO terms that are descendants of the term *Immune System Process* (**Supplementary Table 4**) for the transcript groups **(A)** CDT, common disease transcripts (differentially expressed for both contrasts OSE_4-OSE_0 and MOG_4-CFA but not in the two control contrasts OSE_0-WT or CFA-WT), **(B)** OSE_4sp, OSE_4-specific transcripts (differentially expressed for the contrast OSE_4-OSE_0 but not in MOG_4-CFA, OSE_0-WT, or CFA-WT), **(C)** OSE_1ex, OSE_1-expressed transcripts (differentially expressed in OSE_1-OSE_0 but not in OSE_0-WT or CFA-WT). Note that no GO terms that are descendants of the term *Immune System Process* were significantly overrepresented for the group MOG_4sp, MOG_4-specific transcripts (differentially expressed for the contrast MOG_4-CFA but not in OSE_4-OSE_0, OSE_0-WT, or CFA-WT). The -$\log_{10}$(FDR) from hypergeometric tests is shown on the x-axis and used for coloring the plots (darker colors represent lower FDRs).

Supplementary Figure 5 | Expression levels of *Cd74* and *Icam1*. Diseased mice showed an increased expression of **(A)** *Cd74* and **(B)** *Icam1* (**Supplementary Table 2**). Significance levels: * adjusted $p < 0.05$, ** adjusted $p < 0.01$, *** adjusted $p < 0.001$.

Supplementary Figure 6 | The top 40 overrepresented immune system pathways in the differentially expressed transcripts groups intersected with T_H1-specific genes. The plots show the top 40 overrepresented GO terms that are descendants of the term *Immune System Process* (**Supplementary Table 9**) for the transcript groups **(A)** CDT (common disease transcripts) intersected with T_H1-specific genes, **(B)** OSE_4sp (OSE_4-specific transcripts) intersected with T_H1-specific genes, **(C)** OSE_1ex (OSE_1-expressed transcripts) intersected with T_H1-specific genes. Note that no GO terms were significantly overrepresented for any T_H17-specific or MOG_4sp genes. The -$\log_{10}$(FDR) from hypergeometric tests is shown on the x-axis and used for coloring the plots (darker colors represent lower FDRs).

Supplementary Table 1 | Results from k-means clustering (**Figure 1A**).

Supplementary Table 2 | Differential expression results for all analyzed contrasts. FC, fold change; CI, 95% confidence interval.

Supplementary Table 3 | Comparison of fold changes in expression levels between contrasts (**Figures 1C,D**). CI, 95% confidence interval.

Supplementary Table 4 | Results from WebGestalt gene ontology (GO) over-representation analyses (ORA) for the analysis groups CDT, OSE_4sp, MOG_4sp, and OSE_1ex (see the legend of **Table 1** and **Supplementary Figure 5**) as well as the contrasts OSE_1-OSE_0, OSE_4-OSE_0, MOG_4-CFA, and MOG_4-OSE_4 (**Supplementary Figure 4**). FDR, 5% false discovery rate. Gene sets that are labeled in bold if they are descendants of the GO term "Immune System Process".

Supplementary Table 5 | Differential expression results for the analysis groups (see the legend of **Table 1**) CDT, CDT, OSE_4sp, MOG_4sp, OSE_1ex, and OSE_1sp. OSE_1sp transcripts were differentially expressed in OSE_1-OSE_0 but not in any other contrast. FC, fold change, CI, 95% confidence interval. For CDT transcripts, separate coefficients are provided for the two contrasts MOG_4-CFA and OSE_4-OSE_0.

Supplementary Table 6 | Results from the principal component analysis (PCA) of gene expression profiles of putative MS risk genes (**Figure 2A**). FC, fold change, CI, 95% confidence interval. *P*-values were adjusted for multiple testing using Holm-Bonferroni correction.

Supplementary Table 7 | Results from the principal component analysis (PCA) of gene expression profiles of T_H cell-specific transcripts (**Figures 3B,C**). FC, fold change; CI, 95% confidence interval. *P*-values were adjusted for multiple testing using Holm-Bonferroni correction.

Supplementary Table 8 | Differential expression results for selected T_H1 and T_H17 signature molecules (**Figures 3D,E**). DE, differentially expressed.

Supplementary Table 9 | Results from Webgestalt gene ontology (GO) over-representation analyses (ORA) for the analysis groups CDT, OSE_4sp, and OSE_1ex (see the legend of **Table 1**) among T_H1-specific transcripts (**Supplementary Figure 6**). For MOG_4sp or T_H17-specific transcripts, no significant gene sets were found. FDR, 5% false discovery rate. Gene sets that are labeled in bold if they are descendants of the GO term "Immune System Process".

REFERENCES

1. Andlauer TF, Buck D, Antony G, Bayas A, Bechmann L, Berthele A, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv.* (2016) 2:e1501678. doi: 10.1126/sciadv.1501678
2. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* (2019) 365:eaav7188. doi: 10.1126/science.aav7188
3. Hurst WE. The effects of the injection of normal brain emulsion into rabbits, with special reference to the aetiology of the paralytic accidents of antirabic treatment. *J Hyg.* (1932) 32:33–44. doi: 10.1017/S0022172400017800
4. Glatigny S, Bettelli E. Experimental autoimmune encephalomyelitis (EAE) as animal models of multiple sclerosis (MS). *Csh Perspect Med.* (2018) 8:a028977. doi: 10.1101/cshperspect.a028977
5. Mendel I, Rosbo NK de, Ben-Nun A. A myelin oligodendrocyte glycoprotein peptide induces typical chronic experimental autoimmune encephalomyelitis in H-2b mice: fine specificity and T cell receptor Vβ expression of encephalitogenic T cells. *Eur J Immunol.* (1995) 25:1951–9. doi: 10.1002/eji.1830250723
6. Ben-Nun A, Wekerle H, Cohen IR. The rapid isolation of clonable antigen-specific T lymphocyte lines capable of mediating autoimmune encephalomyelitis. *Eur J Immunol.* (1981) 11:195–9. doi: 10.1002/eji.1830110307
7. Berer K, Mues M, Koutrolos M, Rasbi Z, Boziki M, Johner C, et al. Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature.* (2011) 479:538. doi: 10.1038/nature10554
8. Berer K, Gerdes L, Cekanaviciute E, Jia X, Xiao L, Xia Z, et al. Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proc Natl Acad Sci USA.* (2017) 114:10719–24. doi: 10.1073/pnas.1711233114
9. Krishnamoorthy G, Lassmann H, Wekerle H, Holz A. Spontaneous opticospinal encephalomyelitis in a double-transgenic mouse model of autoimmune T cell/B cell cooperation. *J Clin Investig.* (2006) 116:2385–92. doi: 10.1172/JCI28330
10. Molnarfi N, Schulze-Topphoff U, Weber MS, Patarroyo JC, Prod'homme T, Varrin-Doyer M, et al. MHC class II–dependent B cell APC function is required for induction of CNS autoimmunity independent of myelin-specific antibodies. *J Exp Med.* (2013) 210:2921–37. doi: 10.1084/jem.20130699
11. Lehmann-Horn K, Kronsbein HC, Weber MS. Targeting B cells in the treatment of multiple sclerosis: recent advances and remaining challenges. *Ther Adv Neurol Diso.* (2013) 6:161–73. doi: 10.1177/1756285612474333
12. Jäger A, Dardalhon V, Sobel RA, Bettelli E, Kuchroo VK. Th1, Th17, and Th9 effector cells induce experimental autoimmune encephalomyelitis with different pathological phenotypes. *J Immunol.* (2009) 183:7169–77. doi: 10.4049/jimmunol.0901906
13. The International Multiple Sclerosis Genetics Consortium, The Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* (2011) 476:214–9. doi: 10.1038/nature10251
14. Ben-Nun A, Kaushansky N, Kawakami N, Krishnamoorthy G, Berer K, Liblau R, et al. From classic to spontaneous and humanized models of multiple sclerosis: impact on understanding pathogenesis and drug development. *J Autoimmun.* (2014) 54:33–50. doi: 10.1016/j.jaut.2014.06.004
15. Domingues HS, Mues M, Lassmann H, Wekerle H, Krishnamoorthy G. Functional and pathogenic differences of Th1 and Th17 cells in experimental autoimmune encephalomyelitis. *PLoS ONE.* (2010) 5:e15531. doi: 10.1371/journal.pone.0015531
16. Miller SD, Karpus WJ, Davidson TS. Experimental autoimmune encephalomyelitis in the mouse. *Curr Protoc Immunol.* (2010) 88:15.1.1–20. doi: 10.1002/0471142735.im1501s88
17. R Core Team *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. (2020). Available online at: https://www.R-project.org/ (accessed August 1, 2020).

18. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics.* (2007) 23:2183–4. doi: 10.1093/bioinformatics/btm311
19. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* (2008) 24:1547–48. doi: 10.1093/bioinformatics/btn224
20. Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics.* (2002) 18:S96–104. doi: 10.1093/bioinformatics/18.suppl_1.S96
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
22. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* (2019) 47:W199–205. doi: 10.1093/nar/gkz401
23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B Methodol.* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
24. Krishnamoorthy G, Holz A, Wekerle H. Experimental models of spontaneous autoimmune disease in the central nervous system. *J Mol Med.* (2007) 85:1161–73. doi: 10.1007/s00109-007-0218-x
25. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* (2013) 9:e1003926. doi: 10.1371/journal.pgen.1003926
26. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet.* (2015) 47:1107–13. doi: 10.1038/ng.3395
27. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: a comprehensive review. *J Autoimmun.* (2015) 64:13–25. doi: 10.1016/j.jaut.2015.06.010
28. Jelcic I, Nimer FA, Wang J, Lentsch V, Planas R, Jelcic I, et al. Memory B cells activate brain-homing, autoreactive CD4+ T cells in multiple sclerosis. *Cell.* (2018) 175:85–100.e23. doi: 10.1016/j.cell.2018.08.011
29. Gregersen JW, Kranc KR, Ke X, Svendsen P, Madsen LS, Thomsen AR, et al. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature.* (2006) 443:574–7. doi: 10.1038/nature05133
30. Ashbaugh JJ, Brambilla R, Karmally SA, Cabello C, Malek TR, Bethea JR. IL7Rα contributes to experimental autoimmune encephalomyelitis through altered T cell responses and nonhematopoietic cell lineages. *J Immunol.* (2013) 190:4525–34. doi: 10.4049/jimmunol.1203214
31. Blankenhorn EP, Butterfield R, Case LK, Wall EH, Rio R del, Diehl SA, et al. Genetics of experimental allergic encephalomyelitis supports the role of T helper cells in multiple sclerosis pathogenesis. *Ann Neurol.* (2011) 70:887–96. doi: 10.1002/ana.22642
32. Cao Y, Goods BA, Raddassi K, Nepom GT, Kwok WW, Love JC, et al. Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. *Sci Transl Med.* (2015) 7:287ra74. doi: 10.1126/scitranslmed.aaa8038
33. Hoppmann N, Graetz C, Paterka M, Poisa-Beiro L, Larochelle C, Hasan M, et al. New candidates for CD4 T cell pathogenicity in experimental neuroinflammation and multiple sclerosis. *Brain.* (2015) 138:902–17. doi: 10.1093/brain/awu408
34. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun.* (2018) 9:2397. doi: 10.1038/s41467-018-04732-5
35. Hiltensperger M, Korn T. The interleukin (IL)-23/T helper (Th)17 axis in experimental autoimmune encephalomyelitis and multiple sclerosis. *Csh Perspect Med.* (2017) 8:a029637. doi: 10.1101/cshperspect.a029637
36. Stromnes IM, Cerretti LM, Liggitt D, Harris RA, Goverman JM. Differential regulation of central nervous system autoimmunity by TH1 and TH17 cells. *Nat Med.* (2008) 14:337–42. doi: 10.1038/nm1715
37. Hirota K, Duarte JH, Veldhoen M, Hornsby E, Li Y, Cua DJ, et al. Fate mapping of IL-17-producing T cells in inflammatory responses. *Nat Immunol.* (2011) 12:255–63. doi: 10.1038/ni.1993
38. Korn T, Kallies A. T cell responses in the central nervous system. *Nat Rev Immunol.* (2017) 17:179–94. doi: 10.1038/nri.2016.144

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Andlauer et al. BMC Medicine (2020) 18:298
https://doi.org/10.1186/s12916-020-01769-6

BMC Medicine

RESEARCH ARTICLE Open Access

Treatment- and population-specific genetic risk factors for anti-drug antibodies against interferon-beta: a GWAS

Till F. M. Andlauer[1,2*], Jenny Link[3], Dorothea Martin[1], Malin Ryner[3], Christina Hermanrud[3], Verena Grummel[1], Michael Auer[4], Harald Hegen[4], Lilian Aly[1,5], Christiane Gasperi[1], Benjamin Knier[1,5], Bertram Müller-Myhsok[2,6,7], Poul Erik Hyldgaard Jensen[8], Finn Sellebjerg[8], Ingrid Kockum[3], Tomas Olsson[3], Marc Pallardy[9], Sebastian Spindeldreher[10,11], Florian Deisenhammer[4], Anna Fogdell-Hahn[3], Bernhard Hemmer[1,7*] and on behalf of the ABIRISK consortium

Abstract

Background: Upon treatment with biopharmaceuticals, the immune system may produce anti-drug antibodies (ADA) that inhibit the therapy. Up to 40% of multiple sclerosis patients treated with interferon β (IFNβ) develop ADA, for which a genetic predisposition exists. Here, we present a genome-wide association study on ADA and predict the occurrence of antibodies in multiple sclerosis patients treated with different interferon β preparations.

Methods: We analyzed a large sample of 2757 genotyped and imputed patients from two cohorts (Sweden and Germany), split between a discovery and a replication dataset. Binding ADA (bADA) levels were measured by capture-ELISA, neutralizing ADA (nADA) titers using a bioassay. Genome-wide association analyses were conducted stratified by cohort and treatment preparation, followed by fixed-effects meta-analysis.

Results: Binding ADA levels and nADA titers were correlated and showed a significant heritability (47% and 50%, respectively). The risk factors differed strongly by treatment preparation: The top-associated and replicated variants for nADA presence were the *HLA*-associated variants rs77278603 in IFNβ-1a *s.c.*- (odds ratio (OR) = 3.55 (95% confidence interval = 2.81–4.48), $p = 2.1 \times 10^{-26}$) and rs28366299 in IFNβ-1b *s.c.*-treated patients (OR = 3.56 (2.69–4.72), $p = 6.6 \times 10^{-19}$). The rs77278603-correlated *HLA* haplotype *DR15-DQ6* conferred risk specifically for IFNβ-1a *s.c.* (OR = 2.88 (2.29–3.61), $p = 7.4 \times 10^{-20}$) while *DR3-DQ2* was protective (OR = 0.37 (0.27–0.52), $p = 3.7 \times 10^{-09}$). The haplotype *DR4-DQ3* was the major risk haplotype for IFNβ-1b *s.c.* (OR = 7.35 (4.33–12.47), $p = 1.5 \times 10^{-13}$). These haplotypes exhibit large population-specific frequency differences. The best prediction models were achieved for ADA in IFNβ-1a *s.c.*-treated patients. Here, the prediction in the Swedish cohort showed AUC = 0.91 (0.85–0.95), sensitivity = 0.78, and specificity = 0.90; patients with the top 30% of genetic risk had, compared to patients in the bottom 30%, an OR = 73.9 (11.8–463.6, $p = 4.4 \times 10^{-6}$) of developing nADA. In the German cohort, the AUC of the same model was 0.83 (0.71–0.92), sensitivity = 0.80, specificity = 0.76, with an OR = 13.8 (3.0–63.3, $p = 7.5 \times 10^{-4}$).

(Continued on next page)

* Correspondence: till.andlauer@tum.de; hemmer@tum.de
[1]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Ismaninger Str 22, 81675 Munich, Germany
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusions: We identified several *HLA*-associated genetic risk factors for ADA against interferon β, which were specific for treatment preparations and population backgrounds. Genetic prediction models could robustly identify patients at risk for developing ADA and might be used for personalized therapy recommendations and stratified ADA screening in clinical practice. These analyses serve as a roadmap for genetic characterizations of ADA against other biopharmaceutical compounds.

Keywords: Multiple sclerosis, Interferon beta, Anti-drug antibodies, Human leukocyte antigen (HLA) system, Genetics, Genome-wide association study, Prediction

Background

Interferon β (IFNβ) preparations are a treatment option for multiple sclerosis (MS). IFNβ-1a is produced in Chinese hamster ovary cells and administered either via intramuscular (*i.m.*) or subcutaneous (*s.c.*) injection. IFNβ-1b is raised using *Escherichia coli* and injected subcutaneously. The amino acid sequence of IFNβ-1b differs at two positions from the mammalian protein [1]. Moreover, IFNβ-1b is not glycosylated, which may affect its immunogenicity, e.g., by promoting the formation of protein aggregates [1, 2]. Posttranslational modifications like deamidation, oxidation, and glycation can also occur spontaneously, depending on the manufacturing and processing of biopharmaceuticals [3]. Therefore, also sequence-identical compounds like IFNβ-1a *s.c.* and *i.m.* can differ in their immunogenicity [1, 4].

Up to 40% of patients treated with IFNβ develop anti-drug antibodies (ADA) that bind IFNβ (binding ADA, bADA) [1, 5–7]. A subset of bADA inhibits the interaction of IFNβ with its receptor and thus neutralizes the drug's biological activity (neutralizing ADA, nADA) [8, 9]. Previous studies have already identified genetic factors influencing the development of ADA but could not establish a consensus on the human leukocyte antigen (*HLA*) alleles [10–17] and single nucleotide polymorphisms (SNPs) [14, 15] contributing to ADA development.

The primary aim of the present, retrospective study was to characterize the contribution of genetic risk to ADA development by analyzing a large, cross-sectional sample from two different sites: the Karolinska Institutet Stockholm, Sweden (KI), and the Technical University of Munich, Germany (TUM). In these analyses, it was an objective to establish a consensus on the heterogeneous findings from previous studies, especially regarding the associations of *HLA* alleles. Both bADA levels and nADA titers were determined in the same patients, allowing for systematic comparisons between the two antibody types. Genome-wide association studies (GWAS) on bADA levels, nADA titers, and nADA presence, as well as analyses of the association of imputed *HLA* alleles with ADA, were conducted. As primary analyses, results were pooled across treatments; as secondary analyses, treatment-specific results were evaluated. The secondary aim of the study was to use these genetic factors for the prediction of ADA development.

Methods

Sample inclusion criteria

Patient inclusion criteria of this retrospective study were as follows: diagnosis of either clinically isolated syndrome (CIS) or multiple sclerosis (MS), age at first treatment with IFNβ ≥ 18 years, availability of genotype data, and a serum sample fulfilling the sample inclusion criteria. Patients were diagnosed using the current McDonald criteria at the time of diagnosis. The sample inclusion criteria for bADA-/nADA-negative samples were as follows: ≥ 12 months of treatment with IFNβ; if more than one sample was eligible, the first sample available at least 12 months after initiation of treatment with IFNβ was selected; and no previous positive screening for bADA or nADA. The sample inclusion criteria for previously bADA-/nADA-positive samples were as follows: ≥ 6 months of treatment with IFNβ; if previously treated with an IFNβ product, not having been ADA-positive during a previous IFNβ treatment period; and if more than one sample was eligible, the first sample available at least 6 months after initiation of treatment with IFNβ was selected. Based on these criteria, 1810 patients were eligible at KI and 1488 at TUM. The respective local ethics committees approved the study, and all participants provided written informed consent.

Power calculation

In a previous ADA GWAS, Weber et al. identified a genome-wide significant SNP explaining 2.5% of the variance of bADA levels [14]. To have sufficient power for identification of additional associated variants, 2000 patients were assigned to the discovery-stage analyses. In this dataset, 80% of statistical power can be reached for a variant explaining 1.96% of the variance at a p value of 5×10^{-8} (calculated using the R package *pwr*). Effect sizes in the replication stage are expected to be smaller than in the discovery stage [18]. We thus estimated that at least 682 patients are necessary for replicating up to ten linkage disequilibrium (LD)-independent signals with 80% power, explaining 1.7% of the variance using a one-

sided hypothesis. Because of an expected reduction in power due to heterogeneity and an expected decrease in the number of available samples after titration and quality control (QC), we initially selected 800 patients for the replication stage.

Selection of patients

To select approximately 2800 patients for ADA screening and titration, all available previously bADA-/nADA-positive samples ($n = 984$) were combined with previously ADA-negative samples ($n = 2314$) best-matching ADA-positive ones (Additional file 1). Propensity score matching was conducted using the *R* package *optmatch* [19], based on recruitment site, gender, the age at the blood draw, the IFNβ treatment preparation, the total duration of IFNβ treatment, and eight multi-dimensional scaling (MDS) ancestry components of the genetic identity-by-state (IBS) matrix, calculated from the genotype data to account for population stratification (Additional file 2). From the selected patients, new bADA levels and nADA titers (see below) could be determined for 938 previously bADA-/nADA-positive and 1819 previously ADA-negative samples (Table 1 and Additional file 3). These patients were randomized into a discovery ($n = 2000$), and a replication ($n = 757$) set, using adaptive randomization to minimize differences regarding recruitment site, nADA measurement site (Innsbruck or Copenhagen, see below), gender, the age at the blood draw, the IFNβ treatment preparation, and the total duration of IFNβ treatment.

ADA screening and titration

Binding ADA levels were measured by capture ELISA [20] at a single site (Munich) and were calculated from optical densities using a standard curve (Additional file 2). For the assessment of nADA titers, measured as the inverse of serum dilutions using a luciferase-based bioassay [21], samples were first screened, and titration was only conducted for samples positive during screening [22]. Assessment of nADA titers was conducted at two separate sites (Innsbruck and Copenhagen), to which samples were assigned using adaptive randomization to minimize differences regarding the recruitment site, gender, the age at the blood draw, the IFNβ treatment preparation, and the total duration of IFNβ treatment. We obtained 2748 valid measurements for nADA screening and titers as well as 2752 bADA levels; for 2743 patients; both nADA titers and bADA levels were available (1990 in the discovery and 753 in the replication set). The presence of nADA was defined as samples positive in the screening for nADA and showing a nADA titer ≥ 40 tenfold reduction units per milliliter. Correlations of bADA and nADA were calculated in a combined dataset of all samples. For the estimation of the nADA status from bADA levels, the cutoff was established using nested cross-validation in the discovery dataset (Additional file 2). Sensitivity and specificity were calculated by the application of this cutoff to the replication data.

Genotyping and imputation

SNPs were genotyped on Illumina microarrays, and QC was conducted separately for KI and TUM data in PLINK v1.90b3.44 or higher [23], as described before [24]. Genotype data were imputed to the 1000 Genomes Phase 3 reference panel using SHAPEIT2 and IMPUTE2 [25–27]. The resulting datasets contained 9,096,778 and 8,550,834 high-quality variants with a MAF ≥ 1% for KI and TUM, respectively. *HLA* allele imputation was performed using SNP2HLA v1.0.3/Beagle v3.04 and the Type 1 Diabetes Genetics Consortium imputation panel,

Table 1 Sample characteristics

Treatment preparation	IFNβ-1a *i.m.*		IFNβ-1a *s.c.*		IFNβ-1b *s.c.*	
Cohort	KI Sweden	TUM Germany	KI Sweden	TUM Germany	KI Sweden	TUM Germany
N (%)	345 (24.7)	251 (18.4)	590 (42.3)	558 (40.9)	459 (32.9)	554 (40.6)
Mean age (SD)	46.7 (9.9)	40.1 (9.6)	44.1 (9.9)	38.9 (9.6)	45.4 (10.4)	41.4 (10.4)
Female sex (%)	216 (62.6)	191 (76.1)	440 (74.6)	406 (72.8)	334 (72.8)	390 (70.4)
Median treatment duration in months (MAD)	21.0 (8.1)	40.0 (20.4)	30.0 (15.0)	55.2 (23.1)	24.9 (12.9)	46.9 (23.4)
Progressive MS (%)	60 (17.4)	34 (13.5)	120 (20.3)	90 (16.1)	130 (28.3)	128 (23.1)
nADA positive (%)	45 (13.0)	41 (16.3)	204 (34.6)	188 (33.7)	245 (53.4)	255 (46.0)
Median nADA titer (MAD) *nADA-positive samples*	320 (280)	320 (280)	640 (600)	1280 (1240)	320 (280)	320 (280)
Median bADA level (MAD) *all samples*	13.9 (6.5)	9.6 (5.7)	23.2 (13.9)	16.3 (10.5)	35.8 (19.9)	29.4 (20.0)
Median bADA level (MAD) *nADA-positive samples*	63.8 (42.6)	25.8 (22.3)	109.0 (84.8)	115.0 (104.0)	69.0 (37.0)	73.8 (50.7)

N (%) refers to the entire cohort, the other percentages to the respective column. The nADA and bADA measurements shown here were obtained within the present study. Non-parametric summary statistics are provided for variables that were not normally distributed. Progressive MS = patients with a primary or secondary progressive disease course, as opposed to clinically isolated syndrome and relapsing-remitting MS. The dataset contained 1.6% primary progressive, 0.6% progressive-relapsing, and 18.2% secondary progressive MS patients. The frequency of nADA did not differ between progressive (35.2%) and other (35.5%) MS patients. Patients were diagnosed using the current McDonald criteria at the time of diagnosis. *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany; *SD* standard deviation; *MAD* median absolute deviation

as previously described [28–30]. The extended haplotypes were determined based on the haplotype phasing estimated in Beagle. An additional file provides further details on QC and imputation (Additional file 2).

Estimation of heritability and GWAS

ADA titers/levels were transformed by rank-based inverse normal transformation before analyses. Sex, age, treatment preparation, treatment duration, titration site, and eight ancestry components were used as covariates in all analyses. The covariate treatment preparation was also used in preparation-specific analyses and controlled, beyond the three preparation types, for (a) whether treatment with IFNβ-1a *s.c.* had begun before 2008 (change of the formulation [31]); (b) in the TUM cohort, the dose of IFNβ-1a *s.c.* (22 vs. 44 μg); and (c) in the KI cohort, the IFNβ-1b *s.c.* brand used.

The SNP heritability and genetic correlations were estimated with GCTA GREML on a combined dataset of KI and TUM genotypes [32–35], using the covariates mentioned above plus treatment preparation and the recruitment site.

GWAS were conducted separately for the presence of nADA, nADA titers, and bADA levels. ADA titers/levels were analyzed by linear regression models, the presence of nADA by logistic regression. GWAS were run stratified by cohort (KI Sweden and TUM Germany) and by treatment preparation (IFNβ-1a *i.m.*, IFNβ-1a *s.c.*, IFNβ-1b *s.c.*). For each treatment preparation, samples from Sweden and Germany were analyzed separately in PLINK; GWAS results were pooled per cohort using fixed-effects meta-analysis in METAL [36]. For plots of the ancestry components in both cohorts, see Additional file 4. In the primary analysis (GWAS across treatment preparations), the three treatment groups were subsequently pooled by fixed-effects meta-analysis. The threshold for genome-wide significance was $\alpha = 5 \times 10^{-8}$. For replication, the significance threshold α was corrected for the total number of variants analyzed across all SNP-based analyses in the replication phase ($n = 16$) using Bonferroni's method, i.e., $\alpha = 0.05/16 = 3 \times 10^{-3}$. SNPs prioritized for replication had to fulfill the following criteria: (I) genome-wide significance ($p < 5 \times 10^{-8}$) in the discovery-stage GWAS; (II) within each window of 100,000 bp, only the SNP with the lowest p value was selected; and (III) LD with SNPs showing lower p values had to be $r^2 < 0.2$ in each cohort. Although we used a study design involving discovery and replication, for completeness, association results in the pooled complete dataset are reported as secondary results. More details are provided in Additional file 2.

Permutation analyses

All replicated associations from hypothesis-free linear regression analyses were validated using permutation analyses. In these analyses, the null distribution of test statistics was empirically determined by repeating regression analyses either 200 million or 1 million times with random sampling of phenotype data. To calculate a p value, the number of tests was counted where a model with a random genotype-phenotype association showed the same or a more extreme p value than the correct, non-randomized model; this number was divided by the total number of tests (200 or 1 million). Permutation-based p values were pooled per cohort and treatment using Stouffer's Z-score method [37]. For GWAS variants, 200 million permutations per dataset (discovery/replication), cohort, and treatment preparation were carried out (allowing for p values down to 1×10^{-8}); for stepwise conditional models and *HLA* alleles, the default was 1 million permutations per group (for p values down to 1×10^{-6}). If these permutation p values were $< 1 \times 10^{-6}$, 200 million permutations were conducted. If the permutation p values were $< 1 \times 10^{-8}$, they were set to 1×10^{-8}.

EQTL analyses

The significant *cis*-expression quantitative trait loci (eQTLs) in whole blood were looked up in the GTEx v8 database (https://gtexportal.org/) downloaded on April 1, 2020 (dbGaP accession number phs000424.v8.p2) [38].

Gene-set analyses

Gene-set analyses were conducted with MAGMA v1.07b [39]. First, SNPs within gene boundaries were annotated to RefSeq genes (0 bp window). Second, gene analysis was performed on the pooled GWAS summary statistics, based on LD information from the 1000 Genomes EUR reference panel, using both mean- and top-SNP gene models. Third, gene-level analyses used a combination of the curated 186 KEGG and 1499 Reactome pathways from the MSigDB 7.0 database gene sets [40].

HLA and stepwise conditional analyses

The association of *HLA* alleles was analyzed in *R* v3.3 or higher. As in the GWAS, sex, age, treatment preparation and duration, titration site, and eight ancestry components were used as covariates. Separate regression models were run per cohort and treatment preparation, followed by a two-level meta-analysis: results were combined using fixed-effects meta-analysis first by cohort and then by preparation. For assessment of significance, we applied Bonferroni correction for testing 131 alleles and extended haplotypes [41] (rounded down to $\alpha = 3 \times 10^{-4}$). In the replication phase, we corrected for multiple testing of 41 *HLA* alleles and haplotypes prioritized

across all analyses (rounded down to $\alpha = 1 \times 10^{-3}$). Note that the associations of all *HLA* alleles and haplotypes presented in this study also reached genome-wide significance ($p < 5 \times 10^{8}$) in the pooled analyses of discovery and replication samples, except for the super-extended haplotypes *C7-DQ6* and *A3-DQ6*, which reached a $p < 10 \times 10^{-8}$.

Stepwise conditional regression was conducted, as previously described [42, 43], first only for *HLA* alleles and then for a joint dataset of *HLA* alleles and SNPs mapping to the extended MHC region. In brief, the association of all alleles/SNPs was first tested in separate regression models. The top-associated allele/SNP was then added as a covariate to the regression model, and the analysis was repeated for all remaining alleles/SNPs. This addition of top-associated alleles/SNPs as covariates was repeated until no allele/SNP was significant anymore after correction for multiple testing.

Polygenic risk scores (PRS) and prediction of nADA

PRS were calculated in *R* v3.33 using imputed genetic data, as described previously [44, 45]. For each PRS, the effect sizes of variants from the discovery-stage analyses (training data), below a selected discovery-stage *p* value threshold, were multiplied by the imputed SNP dosage in the replication-stage test data and then summed to produce a single PRS per threshold. For each analysis group, eight PRS based on different GWAS *p* value thresholds were calculated on the discovery data. More details are provided in Additional file 2.

For the prediction of the presence of nADA in the replication dataset, logistic regression of the eight PRS, the top single GWAS variant, and the top *HLA* allele from the discovery stage was conducted using the GWAS models. The area under the receiver operating characteristic curve (AUC) was calculated using the *R* package *pROC*, its 95% confidence interval (CI) with the function *ci.auc* (2000 stratified bootstrap replicates). At this stage, we adapted the significance threshold for ten tests using the Bonferroni correction (Fig. 3a, b). For each treatment preparation, the model with the highest AUC was selected. The performance of all top models was subsequently compared, with a significance threshold adapted for 160 comparisons ($\alpha = 3.13 \times 10^{-4}$, Fig. 3c).

The sensitivity and specificity of the predictions were calculated using the package *OptimalCutpoints*, maximizing both measures (*MaxSpSe*). To avoid overfitting, the cutpoint was selected using nested cross-validation with three outer and four inner folds. In each outer cross-validation instance, the cutoff producing the maximum sensitivity across three inner cross-validation folds was tested on the remaining fold. Nested cross-validation was repeated 100 times, and the mean cutoff of the 100 repetitions was used as the final cutoff. Nagelkerke's pseudo-R^2 was calculated using the package *fmsb*. For a comparison of patients either at low or high genetic risk, patients within the lower 30% of genetic risk were compared to the patients in the upper 30%. We initially selected a 10% cutoff for this contrast and increased it in 10% steps until the sample size in the replication dataset sufficed for the stable convergence of regression models.

Results

From 2757 MS patients recruited in Sweden and Germany and treated with three different IFNβ preparations (Table 1), bADA levels were measured by capture ELISA [20] and nADA titers using a luciferase-based bioassay [21, 22] (Additional files 1-3). The bADA levels were correlated with nADA presence (Spearman $\rho = 0.66$) and nADA titers ($\rho = 0.71$). Compared to the presence of nADA determined via screening and titration, estimation of the nADA status from bADA levels had a sensitivity = 0.85 and a specificity = 0.84 (Fig. 1).

SNP heritability and genetic correlations

The SNP-based heritability estimated from the genotype data was $h^2_g = 0.47$ (standard error (SE) = 0.15, $p = 1.4 \times 10^{-4}$) for the inverse-normal transformed bADA levels and $h^2_g = 0.50$ (SE = 0.15, $p = 2.9 \times 10^{-4}$) for the transformed nADA titers. The SNP heritability of the presence of nADA was $h^2_{go} = 0.48$ on the observed scale (SE = 0.15, $p = 4.9 \times 10^{-4}$) and, assuming an incidence of 0.35 for ADA, $h^2_{gl} = 0.79$ (SE = 0.25) on a liability scale. Genetic correlations of bADA levels with nADA presence ($r_g = 0.89$, SE = 0.14, $p = 1.2 \times 10^{-3}$) and titers ($r_g = 0.95$, SE = 0.11, $p = 7.0 \times 10^{-4}$) were very high.

Outline of the genetic association analyses

To improve control for type I errors, patients were randomized a priori into a discovery ($n = 2000$) and a replication ($n = 757$) set for all genetic association analyses (Additional file 3). The sizes of both datasets were guided using a power calculation (see the "Methods" section). We conducted three separate analyses: first, a pooled analysis of all three treatment preparation groups; second, an analysis of patients treated with IFNβ-1a *s.c.*; and third, an analysis of patients treated with IFNβ-1b *s.c.* For each of these three analysis levels, we conducted separate GWAS of nADA presence, nADA titers, and bADA levels. Because of the small number of IFNβ-1a *i.m.*-treated patients with ADA in the present study, we did not analyze this treatment preparation on its own. In addition to the GWAS, we analyzed imputed *HLA* alleles in the same manner. To estimate the number of independent association signals in the major histocompatibility complex (MHC) region,

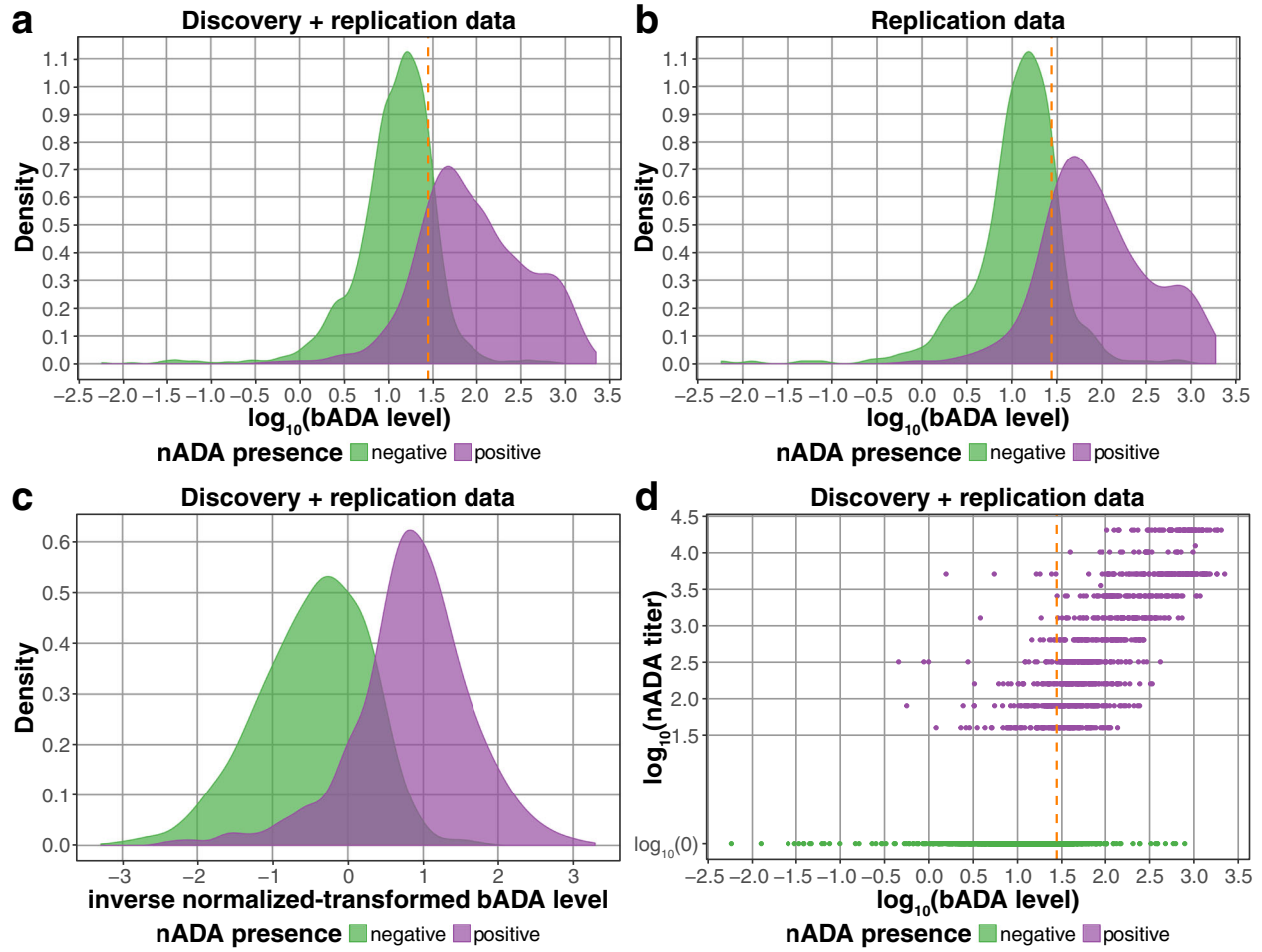


Fig. 1 Comparison of bADA levels and nADA titers. The orange dashed line indicates the cutoff at a $\log_{10}$ bADA level of 1.442256, which optimized the maximum sensitivity and specificity in the discovery data. This cutoff had a sensitivity = 0.83 and a specificity = 0.82 in the discovery dataset; a sensitivity = 0.85 and a specificity = 0.84 in the replication dataset; and a sensitivity = 0.84 and a specificity = 0.82 in the combined dataset. **a** Density plot showing $\log_{10}$ bADA levels in the combined discovery and replication dataset stratified by nADA presence. **b** Density plot showing $\log_{10}$ bADA levels in the replication data stratified by nADA presence. **c** Density plot showing rank-based inverse-normal transformed bADA levels in the combined dataset stratified by nADA presence. **d** Comparison of $\log_{10}$ bADA levels to $\log_{10}$ nADA titers in the combined dataset, colored by nADA presence

we carried out conditional analyses in a combined dataset of GWAS variants and *HLA* alleles.

In all GWAS, only variants within the MHC region were significant on a genome-wide scale ($p < 5 \times 10^{-8}$) in the discovery-stage analyses and replicated (Additional files 5-10). There was no indication for systematic inflation of test statistics; all genomic inflation factors λ were in the expected range (Additional file 11).

GWAS across IFNβ preparations

In the discovery GWAS of nADA presence across all three treatment preparation groups, the strongest association was observed for the insertion TTTTTTT of the variant rs9281971, which was associated with decreased risk for nADA (Table 2 and Additional files 12 and 13). This insertion had a frequency of 36.0% in Swedish and 38.6% in German patients. No other genome-wide significant variant with linkage disequilibrium (LD) $r^2 < 0.2$ with the top signal was identified. The insertion replicated at genome-wide significance and was also the top association signal when pooling discovery and replication data (discovery: odds ratio (OR) = 0.59 (95% confidence interval (CI) = 0.50–0.69), $p = 1.9 \times 10^{-11}$; replication: OR = 0.44 (0.32–0.59), $p_{one\text{-}sided} = 2.4 \times 10^{-08}$; discovery + replication: OR = 0.55 (0.48–0.63), $p = 2.3 \times 10^{-17}$). Inversely, for each copy lacking the insertion TTTTTTT at this site, the OR associated with risk for nADA was thus 1.82 (1.58–2.08). This association was supported in all three treatment groups but was strongest in IFNβ-1a *s.c.*-treated patients (Additional file 14). Because the MHC region shows long-range LD patterns, we conducted stepwise conditional regression analyses in the pooled dataset (discovery +

Table 2 Genome-wide significant variants from GWAS across IFNβ preparations

ADA	Variant	Chr.	Pos. (bp)	MA	MAF (KI)	MAF (TUM)	OR/β (Disc.)	*p* (Disc.)	OR/β (Repl.)	$p_{(one\text{-}sided)}$ (Repl.)	OR/β (Pool.)	*p* (Pool.)
nADA pres.	rs9281971	6	32,596,722	$(T)_7$	0.36	0.39	0.59^{OR}	1.9×10^{-11}	0.44^{OR}	2.4×10^{-08}	0.55^{OR}	2.3×10^{-17}
nADA titer	rs9271377	6	32,587,165	G	0.30	0.33	-0.15^{β}	4.0×10^{-11}	-0.17^{β}	4.0×10^{-05}	-0.16^{β}	1.5×10^{-14}
nADA titer	rs9281971	6	32,596,722	$(T)_7$	0.36	0.39	-0.14^{β}	1.4×10^{-10}	-0.23^{β}	4.6×10^{-09}	-0.16^{β}	2.5×10^{-17}
bADA level	rs9271377	6	32,587,165	G	0.30	0.33	-0.23^{β}	1.3×10^{-13}	-0.20^{β}	8.5×10^{-05}	-0.22^{β}	1.2×10^{-16}
bADA level	rs9272071	6	32,599,487	C	0.32	0.38	-0.21^{β}	1.8×10^{-11}	-0.28^{β}	2.0×10^{-08}	-0.23^{β}	4.5×10^{-18}

The top GWAS association signals that showed genome-wide significance in the discovery-stage analysis ($\alpha = 5 \times 10^{-8}$) and replicated ($\alpha = 3 \times 10^{-3}$) in the analysis across all three treatment preparations. For nADA presence, odds ratios are provided (marked by ^{OR}), and for quantitative ADA measures effect sizes (marked by $^{\beta}$). For detailed association statistics, including conditional analyses, correlated *HLA* alleles, nearby genes, eQTL results, permutation *p* values, and preparation-specific association results, see Additional file 12. For locus-specific Manhattan plots of each locus, see Additional file 13. For forest plots of each association, including treatment preparation-specific effects, see Additional file 14. *Abbreviations: Chr.* chromosome; *Pos.* position in base pairs (build *hg19*); *MA* minor and effect allele; *MAF* minor allele frequency; *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany; *OR* odds ratio; *β* effect size; *Disc* discovery; *Repl.* replication; *Pool.* pooled; *pres.* presence; $(T)_7$ TTTTTTT

replication) to estimate the number of independently associated signals. No variant except rs9281971-TTTTTTT was significantly associated with nADA presence on a genome-wide scale (Additional file 15).

In addition to the analysis of the dichotomous nADA presence, we also conducted GWAS for the quantitative measures nADA titers and bADA levels. In both cases, the top-associated and replicated SNP was rs9271377 (Table 2). As a secondary analysis, we pooled the discovery and replication GWAS. Here, variant rs9281971-TTTTTTT was the strongest association for nADA titers (9558 base pairs (bp) downstream from rs9271377; LD $r^2 = 0.39$ in Swedish and $r^2 = 0.51$ in German patients). For bADA levels, SNP rs9272071 (2759 bp downstream from rs9271377; $r^2 = 0.77$ in Swedish and $r^2 = 0.81$ in German patients) was the top pooled association. These associations were supported in all three treatment groups, but to a lesser degree in IFNβ-1a *i.m.*-treated patients (Additional file 14). Because of possible deviations from normality, the associations of all replicated ADA variants were confirmed using nonparametric permutation analyses (Additional file 12). In stepwise conditional regression analyses in the pooled dataset, rs9281971-TTTTTTT was the only significant variant for nADA titers, while four variants reached significance for bADA levels (rs9272071, rs28746882, rs1265086, and *HLA-DRB1*04:04* (Additional file 13)).

The three top-associated variants map directly upstream of the gene *HLA-DQA1* (rs9271377 18.0 kbp, rs9281971 8.5 kbp, and rs9272071 5.7 kbp upstream (Additional file 13)). The variants were all in weak to moderate LD with the *HLA* allele *DQA1*05:01* (LD range: $0.31 \geq r^2 \leq 0.50$) and part of *cis*-expression quantitative trait loci (eQTLs) with *HLA-DRB5* (GTEx v8, see Additional file 16) [38]. In gene-set analyses using KEGG and Reactome gene sets [39, 40], several immune-related pathways were significant after correction for multiple testing, e.g., "antigen processing and presentation," "Translocation of ZAP-70 to Immunological synapse," and "PD-1 signaling" (Additional file 17).

Treatment-specific GWAS: IFNβ-1a *s.c.*

Forest plots of effect sizes in treatment preparation subgroups suggested that preparation-specific genetic risk factors may exist (Additional file 10). Therefore, we also conducted analyses separately for the two main treatment preparations used in our cohorts (IFNβ-1a *s.c.* and IFNβ-1b *s.c.*). We did not conduct hypothesis-free preparation-specific analyses for IFNβ-1a *i.m.*, due to its much lower number of ADA-positive patients (Table 1).

In GWAS of nADA presence and nADA titers in IFNβ-1a *s.c.*-treated patients, variant rs77278603 was genome-wide significant in the discovery stage and confirmed in the replication dataset (OR = 3.55 (2.81–4.48), $p = 2.1 \times 10^{-26}$; Table 3 and Additional file 12). The variant maps downstream of *HLA-DRB5* (Additional file 18). In the secondary meta-analysis of discovery and replication GWAS, two different but correlated SNPs upstream of *HLA-DQA1* were the most strongly associated signals, rs9271700 for nADA presence and rs9271673 for nADA titers (Table 3 and Additional file 12). Both variants were significant eQTLs with an *HLA-DRB5* transcript (Additional file 16). All three SNPs associated with nADA in IFNβ-1a *s.c.*-treated patients were in LD with the *HLA* allele *HLA-DRB1*15:01* ($r^2 \geq 0.71$).

For bADA levels in IFNβ-1a *s.c.*-treated patients, rs9281962 was the top-associated variant in both the discovery and pooled analysis (Table 3 and Additional file 12), which was in very high LD with the nADA-associated SNPs rs9271700 and rs9271673 ($r^2 \geq 0.93$). In stepwise conditional regression analyses in the pooled dataset, none but the respective top-associated variants were significantly associated with IFNβ-1a *s.c.*-induced ADA (Additional file 15).

Notably, none of the variants associated at genome-wide significance with ADA measurements in IFNβ-1a *s.c.*-treated patients showed statistical support for an association in IFNβ-1b *s.c.*-treated patients with $p < 0.001$ in the discovery stage (Additional files 12 and 19). When analyzing both IFNβ-1a preparations together, results were highly similar to when analyzing IFNβ-1a *s.c.*-treated patients alone (Additional file 12).

Treatment-specific GWAS: IFNβ-1b *s.c.*

In IFNβ-1b *s.c.*-treated patients, SNP rs28366299 was significantly associated with nADA presence in both the discovery and pooled analysis (OR = 3.56 (2.69–4.72), $p = 6.6 \times 10^{-19}$; Table 4 and Additional file 12). It maps upstream of *HLA-DRB1*, which is correlated with *HLA-DQA1*03:01* ($r^2 \geq 0.46$) and an eQTL with an *HLA-DQA2* transcript (Additional files 16 and 20). We confirmed the association of this SNP with nADA presence in a published, independent study on 941 IFNβ-1b *s.c.*-treated patients [15], where it replicated robustly (OR 2.37 (1.81–3.08), one-sided $p = 9.88 \times 10^{-11}$; meta-analysis with the present study: OR 2.87 (2.37–3.48), $p = 7.74 \times 10^{-27}$). The same variant was also associated with nADA titers (Table 4 and Additional file 12). SNP rs9272775, intronic in *HLA-DQA1* and correlated with rs28366299 ($r^2 \geq 0.79$), was the top variant in the pooled analysis of nADA titers. In the independent study [15], variant rs9272775 replicated with one-sided $p = 6.05 \times 10^{-17}$ (meta-analysis $p = 7.62 \times 10^{-40}$).

The analysis of bADA levels in IFNβ-1b *s.c.*-treated patients produced similar results (Table 4 and Additional file 12). In the discovery stage, rs78279385 (LD with rs9272775 $r^2 \geq 0.88$) was the variant showing the most robust support for an association. In the meta-

Table 3 Genome-wide significant variants from treatment-specific GWAS for IFNβ-1a *s.c*

ADA	Variant	Chr.	Pos. (bp)	MA	MAF (KI)	MAF (TUM)	OR/β (Disc.)	*p* (Disc.)	OR/β (Repl.)	$p_{(one\text{-}sided)}$ (Repl.)	OR/β (Pool.)	*p* (Pool.)
nADA pres.	rs77278603	6	32,469,421	A	0.43	0.40	3.33^{OR}	5.3×10^{-19}	4.43^{OR}	1.9×10^{-09}	3.55^{OR}	2.1×10^{-26}
nADA pres.	rs9271700	6	32,593,198	G	0.42	0.39	3.16^{OR}	8.6×10^{-19}	5.08^{OR}	1.2×10^{-10}	3.48^{OR}	5.4×10^{-27}
nADA titer	rs77278603	6	32,469,421	A	0.43	0.40	0.38^{β}	2.2×10^{-19}	0.36^{β}	8.1×10^{-10}	0.37^{β}	2.4×10^{-27}
nADA titer	rs9271673	6	32,592,833	C	0.41	0.39	0.37^{β}	3.0×10^{-19}	0.38^{β}	5.2×10^{-11}	0.37^{β}	2.1×10^{-28}
bADA level	rs9281962	6	32,594,597	T	0.44	0.43	0.51^{β}	6.0×10^{-22}	0.51^{β}	5.5×10^{-12}	0.51^{β}	4.6×10^{-32}

The top GWAS association signals that showed genome-wide significance in the discovery-stage analysis ($\alpha = 5 \times 10^{-8}$) and replicated ($\alpha = 3 \times 10^{-3}$) in the analysis of IFNβ-1a *s.c.*-treated patients. For nADA presence, odds ratios are provided (marked by ^{OR}), and for quantitative ADA measures effect sizes (marked by $^{\beta}$). For detailed association statistics, including conditional analyses, correlated *HLA* alleles, nearby genes, eQTL results, permutation *p* values, and preparation-specific association results, see Additional file 12. For locus-specific Manhattan plots of each locus, see Additional file 18. For forest plots of each association, including treatment preparation-specific effects, see Additional file 19. *Abbreviations: Chr.* chromosome; *Pos.* position in base pairs (build *hg19*); *MA* minor and effect allele; *MAF* minor allele frequency; *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany; *OR* odds ratio; *β* effect size; *Disc* discovery; *Repl.* replication; *Pool.* pooled; *pres.* presence

Table 4 Genome-wide significant variants from treatment-specific GWAS for IFNβ-1b *s.c*

ADA	Variant	Chr.	Pos. (bp)	MA	MAF (KI)	MAF (TUM)	OR/β (Disc.)	*p* (Disc.)	OR/β (Repl.)	$p_{(one\text{-}sided)}$ (Repl.)	OR/β (Pool.)	*p* (Pool.)
nADA pres.	rs28366299	6	32,560,870	A	0.20	0.19	3.11^{OR}	2.1×10^{-12}	5.84^{OR}	5.2×10^{-09}	3.56^{OR}	6.6×10^{-19}
nADA titer	rs28366299	6	32,560,870	A	0.20	0.19	0.40^{β}	1.8×10^{-16}	0.52^{β}	1.1×10^{-09}	0.43^{β}	5.3×10^{-24}
nADA titer	rs9272775	6	32,610,257	C	0.22	0.23	0.38^{β}	2.5×10^{-16}	0.51^{β}	3.2×10^{-10}	0.41^{β}	2.5×10^{-24}
bADA level	rs78279385	6	32,451,758	A	0.23	0.25	0.39^{β}	2.7×10^{-14}	0.39^{β}	2.0×10^{-06}	0.39^{β}	5.5×10^{-19}
bADA level	rs9272775	6	32,610,257	C	0.22	0.23	0.39^{β}	1.0×10^{-13}	0.46^{β}	1.1×10^{-07}	0.41^{β}	1.6×10^{-19}

The top GWAS association signals that showed genome-wide significance in the discovery-stage analysis ($\alpha = 5 \times 10^{-8}$) and replicated ($\alpha = 3 \times 10^{-3}$) in the analysis of IFNβ-1b *s.c.*-treated patients. For nADA presence, odds ratios are provided (marked by ^{OR}), and for quantitative ADA measures effect sizes (marked by $^{\beta}$). For detailed association statistics, including conditional analyses, correlated *HLA* alleles, nearby genes, eQTL results, permutation *p* values, and preparation-specific association results, see Additional file 12. For locus-specific Manhattan plots of each locus, see Additional file 20. For forest plots of each association, including treatment preparation-specific effects, see Additional file 21. *Abbreviations: Chr.* chromosome; *Pos.* position in base pairs (build *hg19*); *MA* minor and effect allele; *MAF* minor allele frequency; *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany; *OR* odds ratio; *β* effect size; *Disc* discovery; *Repl.* replication; *Pool.* pooled; *pres.* presence

analysis of discovery and replication GWAS, rs9272775 was the top-associated SNP. None of the variants associated at genome-wide significance with any ADA measurement in IFNβ-1b *s.c.*-treated patients showed statistical support for an association in IFNβ-1a *s.c.*-treated patients with $p < 0.009$ in the discovery stage (Additional files 12 and 21).

In stepwise conditional regression analyses, rs559242105, in LD with *HLA-DPB1*03:01* ($r^2 \geq 0.78$), was identified as a secondary signal for nADA presence and titers when conditioning for the respective top SNP (Additional file 15). In stepwise conditional analyses of bADA levels, the allele *HLA-DRB1*04:01* reached a lower p value ($\beta = 0.61$, $p = 8.3 \times 10^{-20}$) than the top GWAS variant rs9272775 ($\beta = 0.41$, $p = 1.6 \times 10^{-19}$). Note that this constitutes the only analysis where an *HLA* allele reached a lower p value than the best available SNP. In this analysis, variant rs17205731 was identified as a significant secondary signal. The two secondary variants rs559242105 (nADA presence/titers) and rs17205731 (bADA levels) are not in LD with each other; both map to *HLA*-associated loci further downstream than the primary association signal (Additional file 20). The variant rs559242105 was independent of *HLA-DRB1*04:01* (nADA titers without conditioning for *HLA-DRB1*04:01*: $p = 5.2 \times 10^{-10}$, conditioned for *HLA-DRB1*04:01*: 1.1×10^{-08}). Possibly, the secondary signal of rs559242105 corresponds to the protective association observed for *HLA-DPB1*03:01* (conditional model of *HLA-DPB1*03:01* OR = 0.49 (0.35–0.69), $p = 4.7 \times 10^{-05}$) (Additional file 15).

Analysis of *HLA* variants across IFNβ preparations

Most previous studies have not conducted GWAS but instead analyzed the association of *HLA* alleles with ADA. Therefore, we also conducted a dedicated association analysis of imputed *HLA* alleles with ADA (Table 5). In this secondary analysis, the discovery-stage significance threshold was set to $\alpha = 3 \times 10^{-4}$, corresponding to the Bonferroni correction for 131 analyzed alleles and extended haplotypes. The full list of significantly associated *HLA* alleles and haplotypes is shown in (Additional file 22). In analyses of nADA presence and nADA titers across all treatment preparations, no *HLA*

Table 5 Selected significant *HLA* alleles and haplotypes

ADA	Allele/HT	AF (KI)	AF (TUM)	OR/β (Disc.)	*p* (Disc.)	OR/β (Repl.)	$p_{(one\text{-}sided)}$ (Repl.)	OR/β (Pool.)	*p* (Pool.)
All preparations: risk alleles									
bADA levels	HLA-DQA1*01:02	0.42	0.36	0.18^{β}	1.6×10^{-08}	0.17^{β}	3.7×10^{-04}	0.17^{β}	4.8×10^{-11}
bADA levels	B7-DQ6	0.18	0.17	0.17^{β}	8.6×10^{-06}	0.21^{β}	3.8×10^{-04}	0.18^{β}	2.7×10^{-08}
All preparations: protective alleles									
bADA levels	DR3-DQ2	0.13	0.12	-0.23^{β}	9.6×10^{-08}	-0.21^{β}	6.7×10^{-04}	-0.23^{β}	5.0×10^{-10}
IFNβ-1a *s.c.*: risk alleles									
nADA pres.	DR15-DQ6	0.34	0.30	2.73^{OR}	3.1×10^{-14}	3.41^{OR}	1.5×10^{-07}	2.89^{OR}	7.4×10^{-20}
nADA titer	DR15-DQ6	0.34	0.30	0.36^{β}	1.3×10^{-15}	0.32^{β}	2.3×10^{-07}	0.35^{β}	3.6×10^{-21}
bADA levels	DR15-DQ6	0.34	0.30	0.44^{β}	6.8×10^{-15}	0.45^{β}	4.4×10^{-08}	0.44^{β}	3.5×10^{-21}
IFNβ-1a *s.c.*: protective alleles									
nADA pres.	DR3-DQ2	0.13	0.12	0.40^{OR}	9.1×10^{-07}	0.29^{OR}	4.0×10^{-04}	0.37^{OR}	3.7×10^{-09}
nADA titer	DR3-DQ2	0.13	0.12	-0.31^{β}	8.6×10^{-08}	-0.28^{β}	5.0×10^{-04}	-0.30^{β}	3.4×10^{-10}
bADA levels	DR3-DQ2	0.13	0.12	-0.41^{β}	2.1×10^{-08}	-0.40^{β}	1.5×10^{-04}	-0.41^{β}	2.5×10^{-11}
IFNβ-1b *s.c.*: risk alleles									
nADA pres.	HLA-DRB1*04:01	0.09	0.07	6.82^{OR}	3.4×10^{-13}	14.7^{OR}	1.7×10^{-07}	7.95^{OR}	1.4×10^{-18}
nADA pres.	DR4-DQ3	0.07	0.05	6.23^{OR}	1.2×10^{-09}	14.7^{OR}	6.1×10^{-06}	7.35^{OR}	1.5×10^{-13}
nADA titer	HLA-DRB1*04:01	0.09	0.07	0.56^{β}	4.1×10^{-15}	0.56^{β}	9.2×10^{-06}	0.56^{β}	3.7×10^{-19}
nADA titer	DR4-DQ3	0.07	0.05	0.50^{β}	1.6×10^{-09}	0.54^{β}	1.3×10^{-04}	0.51^{β}	1.9×10^{-12}
bADA levels	HLA-DRB1*04:01	0.09	0.07	0.62^{β}	4.6×10^{-15}	0.61^{β}	1.8×10^{-06}	0.62^{β}	8.7×10^{-20}
bADA levels	DR4-DQ3	0.07	0.05	0.56^{β}	4.9×10^{-10}	0.53^{β}	2.2×10^{-04}	0.56^{β}	9.0×10^{-13}

Selected four-digit *HLA* alleles and extended haplotypes that were significantly associated ($p < 3 \times 10^{-4}$) with an ADA measurement and replicated ($p_{one\text{-}sided} < 1 \times 10^{-3}$). Alleles that are part of one of the listed extended haplotypes and showed a similar or weaker association than the haplotypes and which did not remain significant when conditioning for the haplotypes are not displayed separately. For nADA presence, odds ratios are provided (marked by ^{OR}), and for quantitative ADA measures effect sizes (marked by $^{\beta}$). For a detailed table of all results, see Additional file 22. For forest plots of each association, including treatment preparation-specific effects, see Additional files 14, 19, and 21. *HT* haplotype; *AF* allele frequency; *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany; *β* effect size; *OR* odds ratio; *Disc* discovery; *Repl.* replication; *Pool.* pooled; *pres.*presence. Abbreviations of the haplotypes: *B7-DQ6*, *HLA-B*07:02 + HLA-DRB1*15:01 + HLA-DQA1*01:02 + HLA-DQB1*06:02*; *DR15-DQ6*, *HLA-DRB1*15:01 + HLA-DQA1*01:02 + HLA-DQB1*06:02*; *DR3-DQ2*, *HLA-DRB1*03:01 + HLA-DQA1*05:01 + HLA-DQB1*02:01*; *DR4-DQ3*, *HLA-DRB1*04:01 + HLA-DQA1*03:01 + HLA-DQB1*03:02*

allele was significant after correction for multiple testing and replicated.

For bADA levels, the top-associated and replicated *HLA* risk allele across preparations was *HLA-DQA1*01:02* ($p = 4.79 \times 10^{-11}$). This allele is part of the extended ancestral haplotype *B7-DQ6* (the combined presence of *HLA-B*07:02*, *HLA-DRB1*15:01*, *HLA-DQA1*01:02*, and *HLA-DQB1*06:02* on the same chromosome), which was also associated ($p = 2.70 \times 10^{-8}$). However, conditional analyses indicated an independent effect of *HLA-DQA1*01:02* from the extended haplotype (Additional file 22). The extended ancestral haplotype *DR3-DQ2* (the combined presence of *HLA-DRB1*03:01*, *HLA-DQA1*05:01*, and *HLA-DQB1*02:01* on the same chromosome) was the top protective association for bADA levels across preparations ($p = 4.97 \times 10^{-10}$; Table 5 and Additional file 22). In conditional analyses, none of these *HLA* associations was independent of the top bADA SNP rs9271377 (Additional file 22).

Treatment-specific *HLA* analyses

For many identified *HLA* alleles, support for an association was predominantly observed in patients treated with either IFNβ-1a *s.c.* or -1b *s.c.* but not in both groups simultaneously (Additional file 14).

When analyzing IFNβ-1a *s.c.*-treated patients, allele *HLA-DQB1*06:02* and the ancestral haplotype *DR15-DQ6*, both smaller subsets of *B7-DQ6*, were the *HLA* risk variants showing the most robust support for an association in each of the three ADA measurements (*DR15-DQ6* nADA presence: OR = 2.88 (2.29–3.61), $p = 7.4 \times 10^{-20}$; Fig. 2a, Table 5 and Additional file 22). All other risk alleles associated with IFNβ-1a *s.c.*-induced ADA were part of this extended haplotype and did not remain significant when conditioning for *DR15-DQ6*. None of these variants passed the discovery-stage significance threshold in patients receiving IFNβ-1b *s.c.* (Additional file 19). When conditioning the IFNβ-1a *s.c.*-associated risk *HLA* haplotype and alleles for the top GWAS SNP rs77278603, none of the HLA signals remained significant (Additional file 22). The *DR15-DQ6* risk association from the *HLA* analysis thus likely corresponds to the GWAS association of rs77278603 and correlated SNPs.

The ancestral haplotype *DR3-DQ2* and its allele *HLA-DQB1*02:01* were the protective alleles showing the lowest *p* values in IFNβ-1a *s.c.*-treated patients (*DR3-DQ2* nADA presence: OR = 0.37 (0.27–0.52), $p = 3.7 \times 10^{-9}$; Table 5 and Additional file 22), with all other protective alleles being part of this extended haplotype. No other allele remained significant when conditioning for *DR3-DQ2*. None of these variants were significantly associated in patients treated with IFNβ-1b *s.c.* (e.g., *DR3-DQ2*, nADA presence $p = 0.27$) or IFNβ-1a *i.m.* after correction for multiple testing (Additional files 19 and 22).

In IFNβ-1b *s.c.*-treated patients, *HLA-DRB1*04:01* was the risk allele that showed the most robust support for an association for all three ADA measurements, and all associated alleles were part of the haplotype *DR4-DQ3* (the combined presence of *HLA-DRB1*04:01*, *HLA-DQA1*03:01*, and *HLA-DQB1*03:02* on the same chromosome). The pooled association strength of *DR4-*

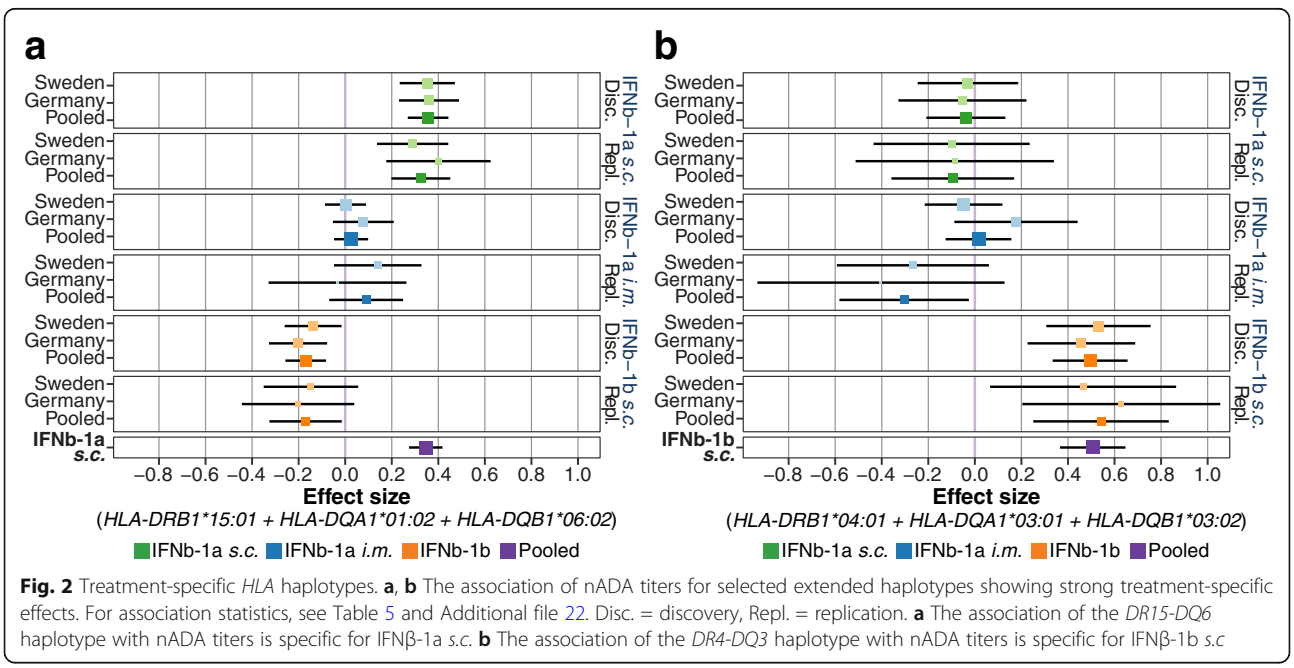


Fig. 2 Treatment-specific *HLA* haplotypes. **a**, **b** The association of nADA titers for selected extended haplotypes showing strong treatment-specific effects. For association statistics, see Table 5 and Additional file 22. Disc. = discovery, Repl. = replication. **a** The association of the *DR15-DQ6* haplotype with nADA titers is specific for IFNβ-1a *s.c.* **b** The association of the *DR4-DQ3* haplotype with nADA titers is specific for IFNβ-1b *s.c*

Table 6 Association of candidate variants

Measurement	SNP/*HLA* allele	First publication	Published effect direction	OR/β (pooled)	$p_{(one\text{-}sided)}$ (all prep.)	$p_{(one\text{-}sided)}$ (IFNβ-1a *s.c.*)	$p_{(one\text{-}sided)}$ (IFNβ-1b *s.c.*)
nADA presence	rs2454138-A	[15]	risk	1.40^{OR}	$\mathbf{1.59 \times 10^{-05}}$	9.65×10^{-01}	$\mathbf{2.41 \times 10^{-16}}$
nADA presence	rs522308-T	[15]	risk	1.39^{OR}	$\mathbf{2.06 \times 10^{-05}}$	9.80×10^{-01}	$\mathbf{8.76 \times 10^{-17}}$
nADA presence	rs9272105-A	[14]	protective	0.62^{OR}	$\mathbf{1.21 \times 10^{-13}}$	$\mathbf{2.76 \times 10^{-12}}$	$\mathbf{1.14 \times 10^{-05}}$
nADA presence	HLA-DQA1*02:01	[16]	risk	1.27^{OR}	2.12×10^{-02}	5.74×10^{-01}	$\mathbf{2.99 \times 10^{-04}}$
nADA presence	HLA-DRB1*04:01	[12]	risk	1.76^{OR}	$\mathbf{9.46 \times 10^{-06}}$	8.71×10^{-01}	$\mathbf{1.62 \times 10^{-19}}$
nADA titer	HLA-DRB1*04:08	[12]	risk	1.14^{β}	$\mathbf{7.50 \times 10^{-09}}$	$\mathbf{1.56 \times 10^{-05}}$	$\mathbf{6.21 \times 10^{-05}}$
nADA titer in nADA-positive	HLA-DRB1*04:08	[12]	risk	0.48^{β}	$\mathbf{2.32 \times 10^{-03}}$	2.45×10^{-02}	$\mathbf{3.59 \times 10^{-04}}$
nADA presence	HLA-DRB1*07:01	[10]	risk	1.28^{OR}	1.90×10^{-02}	5.47×10^{-01}	$\mathbf{3.03 \times 10^{-04}}$
nADA presence	HLA-DRB1*08:01	[13]	risk	1.38^{OR}	2.52×10^{-02}	$\mathbf{1.66 \times 10^{-05}}$	9.14×10^{-01}
nADA presence	HLA-DRB1*15:01	[13]	risk	1.32^{OR}	$\mathbf{6.27 \times 10^{-05}}$	$\mathbf{4.24 \times 10^{-19}}$	1.00×10^{0}
bADA levels	HLA-DRB1*16:01	[11]	risk	0.54^{β}	$\mathbf{7.10 \times 10^{-06}}$	1.12×10^{-02}	$\mathbf{3.34 \times 10^{-05}}$
nADA titer in nADA-positive	HLA-DRB1*16:01	[11]	risk	0.62^{β}	$\mathbf{2.11 \times 10^{-04}}$	2.93×10^{-01}	$\mathbf{5.07 \times 10^{-05}}$
nADA presence	HLA-DQA1*05:01	[13] / [16]	protective	0.68^{OR}	$\mathbf{3.09 \times 10^{-07}}$	$\mathbf{3.15 \times 10^{-09}}$	4.59×10^{-03}
nADA presence	HLA-DQB1*02:01	[16]	protective	0.65^{OR}	$\mathbf{6.24 \times 10^{-06}}$	$\mathbf{1.46 \times 10^{-09}}$	1.94×10^{-01}
nADA presence	HLA-DRB1*03:01	[11] / [16]	protective	0.65^{OR}	$\mathbf{5.43 \times 10^{-06}}$	$\mathbf{1.55 \times 10^{-09}}$	1.83×10^{-01}
nADA presence	HLA-DRB1*04:04	[11]	protective	0.56^{OR}	$\mathbf{2.29 \times 10^{-03}}$	6.91×10^{-03}	1.03×10^{-01}

The table shows previously published SNPs and *HLA* alleles that showed a one-sided $p < 2.5 \times 10^{-3}$ (Bonferroni correction for 20 tests) either in the pooled analysis across treatment preparations or in the analysis of IFNβ-1a *s.c.*-treated or IFNβ-1b *s.c.*- treated patients. *P* values below this threshold are labeled in bold font. For nADA presence, odds ratios are provided (marked by ^{OR}), and for quantitative ADA measures effect sizes (marked by $^{\beta}$), both refer to the pooled analysis across treatment preparations. For a detailed table of all results, see Additional file 23. *β* effect size; *OR* odds ratio; *prep* treatment preparations. "nADA titer in nADA-positive" refers to an analysis of nADA titers restricted to nADA-positive patients

DQ3 for nADA presence was OR = 7.35 (4.32–12.47), $p = 1.5 \times 10^{-13}$ in patients treated with IFNβ-1b *s.c.* (Fig. 2b, Table 5, and Additional file 22). Of note, when conditioning for *DR4-DQ3*, the association of *HLA-DRB1*04:01* remained significant, suggesting it to constitute the primary signal (Additional file 22). However, statistical power for fine-mapping this signal was limited because of the low allele frequencies of the allele and the haplotype (Table 3). These alleles and the haplotype were not significantly associated in IFNβ-1a *s.c.*-treated patients (e.g., *DR4-DQ3*, nADA presence for IFNβ-1a *s.c.*, $p = 0.22$ (Additional files 21 and 22)). In conditional analyses of the IFNβ-1b *s.c.*-associated risk *HLA* haplotype and alleles for the top GWAS SNP rs28366299, *HLA-DRB1*04:01* and *DR4-DQ3* remained significant (Additional files 15 and 22) and vice versa (Additional file 8), indicating that rs28366299 represents an independent signal from the *HLA* association. There were no significant protective alleles for patients receiving IFNβ-1b *s.c.*

Analysis of candidate variants

As a secondary analysis, we analyzed the association of SNPs and *HLA* alleles previously published to be associated with ADA (Table 6 and Additional file 23). Here, we used the Bonferroni correction for 20 SNPs and alleles, corresponding to a significance threshold of $\alpha = 2.5 \times 10^{-3}$. We found support for an association of 15 of the 20 candidate variants (Table 6), but not for SNP rs4961252 (which does not map to the MHC region), the MHC class I alleles, or the *HLA-DRB1*11* alleles. Notably, *HLA-DRB1*04:08* was only associated with bADA levels and nADA titers but not with nADA presence, *HLA-DRB1*16:01* only with bADA levels. In follow-up analyses, we observed that these two *HLA* alleles were also associated with nADA titers in nADA-positive patients (Table 6 and Additional file 23).

Prediction of ADA

We predicted the occurrence of nADA in the replication dataset using the genetic models derived in the discovery dataset. For each treatment preparation, we analyzed eight polygenic risk scores (PRS), the top single GWAS variant, and the top *HLA* allele from the discovery stage (Fig. 3a, b and Additional files 24 and 25). Based on the AUC, the best predictions were achieved in models either featuring only the top variant or by PRS consisting of variants showing strong support for an association (Fig. 3). We thus did not observe evidence for a highly polygenic inheritance of nADA development.

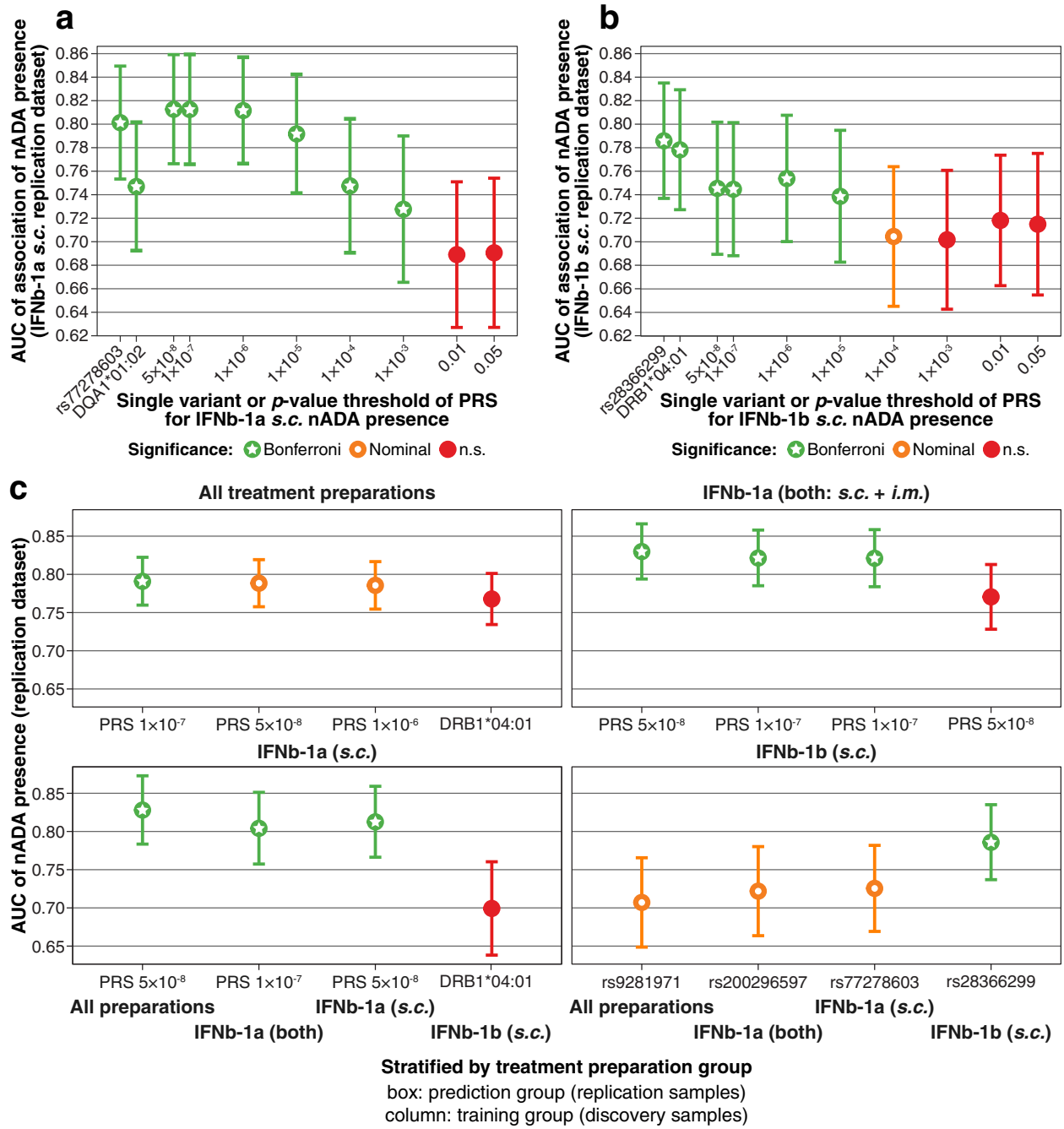


Fig. 3 Prediction of nADA. Treatment-specific prediction of the presence of nADA in the replication data. Eight PRS, the top single GWAS variant, and the top *HLA* allele from the discovery stage were analyzed, with sex, age, treatment preparation and duration, titration site, and ancestry components as covariates. The plots show the area under the receiver operating characteristic curve (AUC) and its 95% confidence interval (CI) calculated using bootstrapping. Bonferroni = significant after Bonferroni correction for multiple testing; nominal = nominally significant ($p < 0.05$); n.s. = not significant. **a**, **b** AUC of all ten prediction models for **a** IFNβ-1a *s.c.* and **b** IFNβ-1b *s.c.*; Bonferroni correction for ten tests. **c** The model with the highest AUC for each treatment preparation, Bonferroni correction for 160 tests ($\alpha = 3.13 \times 10^{-4}$). Boxes show the prediction groups (replication data) and columns within each box the training data groups (discovery data)

Interestingly, only treatment-specific predictions were significant; IFNβ-1a *s.c.*-specific models could not predict nADA in IFNβ-1b *s.c.*-treated patients and vice versa (Fig. 3c and Additional file 26). Prediction models containing either the top PRS or SNP showed distinctly increased AUCs, Nagelkerke's pseudo-R^2, sensitivities, and

Table 7 Treatment-specific prediction of the presence of nADA in the replication data

Preparation	Model	Cohort	OR	95% CI	*p*	AUC	R^2	Sensitivity	Specificity
IFNβ-1a *s.c.*	Without genetics	KI				0.65	0.08	0.66	0.52
IFNβ-1a *s.c.*	Without genetics	TUM				0.60	0.06	0.71	0.42
IFNβ-1a *s.c.*	PRS nADA presence 5×10^{-08}	KI	3.89	2.35–6.45	1.44×10^{-07}	0.85	0.42	0.78	0.78
IFNβ-1a *s.c.*	PRS nADA presence 5×10^{-08}	TUM	2.56	1.56–4.21	2.11×10^{-04}	0.76	0.24	0.68	0.65
IFNβ-1a *s.c.*	**PRS top 30% vs. bottom 30%**	**KI**	**73.86**	**11.77–463.61**	$\mathbf{4.42 \times 10^{-06}}$	**0.91**	**0.59**	**0.78**	**0.90**
IFNβ-1a *s.c.*	**PRS top 30% vs. bottom 30%**	**TUM**	**13.78**	**3.00–63.28**	$\mathbf{7.45 \times 10^{-04}}$	**0.83**	**0.38**	**0.80**	**0.76**
IFNβ-1a *s.c.*	SNP rs77278603 additive coding	KI	4.49	2.41–8.36	2.14×10^{-06}	0.82	0.36	0.74	0.74
IFNβ-1a *s.c.*	SNP rs77278603 additive coding	TUM	3.88	1.78–8.47	6.67×10^{-04}	0.73	0.21	0.63	0.74
IFNβ-1a *s.c.*	SNP rs77278603-A dominant coding	KI	9.16	2.48–33.79	8.79×10^{-04}	0.78	0.31	0.57	0.76
IFNβ-1a *s.c.*	SNP rs77278603-A dominant coding	TUM	3.85	1.10–13.49	3.51×10^{-02}	0.72	0.20	0.56	0.68
IFNβ-1b *s.c.*	Without genetics	KI				0.70	0.15	0.48	0.82
IFNβ-1b *s.c.*	Without genetics	TUM				0.58	0.02	0.82	0.20
IFNβ-1b *s.c.*	PRS nADA presence 1×10^{-06}	KI	2.40	1.45–3.97	6.46×10^{-04}	0.78	0.33	0.57	0.85
IFNβ-1b *s.c.*	PRS nADA presence 1×10^{-06}	TUM	2.15	1.43–3.23	2.28×10^{-04}	0.73	0.22	0.73	0.58
IFNβ-1b *s.c.*	PRS top 30% vs. bottom 30%	KI	10.16	2.30–44.95	2.25×10^{-03}	0.83	0.46	0.58	0.87
IFNβ-1b *s.c.*	PRS top 30% vs. bottom 30%	TUM	5.97	2.03–17.52	1.14×10^{-03}	0.78	0.33	0.69	0.75
IFNβ-1b *s.c.*	SNP rs28366299 additive coding	KI	4.51	1.72–11.80	2.14×10^{-03}	0.77	0.31	0.57	0.82
IFNβ-1b *s.c.*	SNP rs28366299 additive coding	TUM	6.91	3.18–15.03	1.07×10^{-06}	0.79	0.32	0.74	0.61
IFNβ-1b *s.c.*	**SNP rs28366299-A dominant coding**	**KI**	**9.78**	**2.68–35.74**	$\mathbf{5.62 \times 10^{-04}}$	**0.83**	**0.40**	**0.62**	**0.83**
IFNβ-1b *s.c.*	**SNP rs28366299-A dominant coding**	**TUM**	**7.56**	**3.01–19.02**	$\mathbf{1.71 \times 10^{-05}}$	**0.80**	**0.33**	**0.77**	**0.57**

Predictors in the model without genetics: sex, age, treatment duration, and titration site. The genetic models contained the same base model plus the indicated genetic factors and ancestry components. The top models are indicated in bold font. *OR* odds ratio; *CI* 95% confidence interval; *p* *p* value of the genetic component; *AUC* area under the receiver operating characteristic curve; R^2 Nagelkerke's pseudo-R^2; *KI* Karolinska Institutet, Sweden; *TUM* Technical University of Munich, Germany

specificities over models containing only the covariates (Table 7).

Finally, patients with a high and low genetic risk burden were contrasted [46]. To this end, patients within the upper 30% of the top-associated PRS were compared to the patients within the lower 30%. This specific threshold was set to allow for a large enough sample size in the replication dataset for the stable convergence of regression models when conducting cross-validation. In addition, the respective top SNP was analyzed using dominant coding, thereby comparing no copy of the risk allele to any copy. Here, the best prediction was achieved for patients treated with IFNβ-1a *s.c.* (Table 7). In the Swedish cohort, it had an AUC = 0.91 (CI = 0.85–0.95), pseudo-R^2 = 0.59, sensitivity = 0.78, and specificity = 0.90. Patients with the top 30% of genetic risk had, compared to patients in the bottom 30%, an OR = 73.9 (CI = 11.8–463.6, $p = 4.4 \times 10^{-06}$) of developing nADA. In the German cohort, the same model had an AUC = 0.83 (0.71–0.92), pseudo-R^2 = 0.38, sensitivity = 0.80, and specificity = 0.76; the OR of patients with the top 30% of genetic risk was 13.8 (3.0–63.3, $p = 7.5 \times 10^{-4}$).

Discussion

Several studies have previously assessed genetic risk factors for ADA. They were limited by much smaller sample sizes, only analyzed a single population, focused mostly on *HLA* alleles, or did not consistently assess ADA with sensitive and validated methods. The latter is also reflected in the increased number of nADA-positive samples in the measurements conducted for the present study, compared to previous results (Additional files 1 and 3). Most importantly, the existing studies neither reached a consensus on genetic risk factors nor could they delineate a robust prediction model for ADA. To our knowledge, the present study constitutes the most extensive genetic characterization of ADA risk to date, is the first systematic comparison of the genetics of different ADA types, and includes the first genetic prediction model for ADA against IFNβ.

Previously reported genetic risk factors

All genetic variants robustly associated with ADA in the present study map to the MHC region and are linked to the expression of *HLA* genes or amino acid

changes in the peptide-binding groove of HLA molecules. SNP rs9272105, mapping to the MHC region and previously identified in a study conducted on a subset of the patients analyzed here [14], was significantly associated across treatment preparations in the present study. However, we found no support for an association of variant rs4961252 on chromosome 8, identified in the same study [14], which confirms a previously failed replication attempt [15]. Both variants already identified in an independent study of IFNβ-1b *s.c.*-treated patients [15] replicated only in individuals treated with IFNβ-1b *s.c.*

Previous studies prioritized sixteen different *HLA* alleles as potentially associated with nADA presence, nADA titers, or bADA levels [10–13, 15, 16]. Of these, eleven were significantly associated with an ADA measurement for any treatment preparation in our study (Table 6), and five were not (Additional file 23). Importantly, the present study does not constitute a formal replication for many of the candidate *HLA* alleles because of the extensive sample overlap with previous Swedish and German studies. The *HLA* alleles that did not replicate had low frequencies, with a maximum AF = 0.06, and showed only weak support in the original studies. Notably, in some previous studies, highly correlated alleles were analyzed as if they were independent variants, and some studies failed to correct them appropriately for multiple testing. Both factors may have led to an overestimation of the number of associated *HLA* alleles in previous studies.

One study analyzed Spanish patients [16], a population whose allele frequencies and linkage patterns differ from the individuals studied here, and whose results may thus not be fully comparable to the present study. Three MHC class I HLA alleles reported to be associated with ADA by Núñez et al. [16] did not replicate in the present study (Additional file 23). All three alleles are more frequent in Spain than in Germany and Sweden, with, e.g., *HLA-B*14:02* showing a frequency of 1% in Sweden [47], 2% in Germany, and 4% in Spain [48]. To reliably assess whether the associations of these alleles are specific to Spanish populations or whether the lack of correction for multiple testing led to type I errors in the original study, independent replication studies on Spanish patients are required. Next to population-specific effects, joint analyses of patients receiving different proportions of IFNβ treatment preparations constituted a source of heterogeneity in previous studies. In our comprehensive analyses, we could now consolidate several *HLA* alleles published in previous studies into few treatment-specific haplotypes.

Treatment preparation-specific risk

Before our study, it was unclear whether treatment preparation-independent or preparation-specific genetic risk factors dominate ADA risk. The associations of the top GWAS SNPs identified in the analyses across all treatment preparations were mostly supported in all treatment preparations (Additional file 14). Nevertheless, we observed lower p values and larger effect sizes in the preparation-specific analyses than in the models combining patients across treatments. The combination of patients receiving different treatment preparations thus created heterogeneity that decreased statistical power. This hypothesis was further supported in stepwise conditional analyses. Here, we observed more evidence for the existence of independent risk loci in analyses across preparations than was the case in treatment-specific analyses. Likely, such presumably independent loci in the combined analysis reflect treatment preparation-specific effects. These findings thus argue in favor of conducting treatment-specific rather than cross-treatment analyses. In future studies of ADA against biopharmaceuticals, analyses of preparation-specific risk factors should, therefore, be prioritized.

Although differences in the antigenic potential of the various IFNβ preparations are known [1], the extent of preparation-specific genetic risk observed in the present study is striking (Fig. 2). There are several plausible explanations for why the preparations might be processed differently by the immune system. While the amino acid sequence of IFNβ-1a is identical to natural human IFNβ, IFNβ-1b diverges at two positions: IFNβ-1b lacks the N-terminal methionine, and a cysteine at position 17 is substituted by serine. Furthermore, the products are raised in different cell types, prokaryotic *E. coli* and eukaryotic Chinese hamster ovary cells, leading to different post-translational modifications, especially glycosylation [1].

Lack of glycosylation facilitates the formation of protein aggregates, increasing the immunogenicity of IFNβ-1b [1, 2]. Previous research demonstrated that, among the three preparations analyzed in the present study, IFNβ-1b shows the highest tendency to aggregate [49]. IFNβ-1a *i.m.*, which does not contain human serum albumin, forms the fewest aggregates and shows the lowest rate of ADA. Furthermore, aggregates observed with IFNβ-1a *s.c.* preparations are mainly formed by human serum albumin [49]. Differences in IFNβ protein aggregation might, in addition to increased presentation of peptides by dendritic cells and, thus, increased T cell activation [50, 51], also contribute to the diversification of genetic risk factors. When taken up by antigen-presenting cells, e.g., dendritic cells, IFNβ oligomers are likely degraded differently from monomers. Such differences in processing could produce diverse peptides, which may be presented by different MHC class II molecules [50].

Post-translational modifications not only affect aggregate formation but, together with differences in the amino acid sequences, also alter the biochemical

properties of IFNβ-1a and IFNβ-1b. Thereby, both post-translational modifications and differences in the amino acid sequence may contribute to the preparation-specific associations with *HLA* alleles [4]. For example, altering epitopes by glycosylation strongly affects antigen recognition [52]. Possibly, glycosylated IFNβ-1a peptides are thus preferentially recognized by different peptide-binding grooves of MHC molecules than IFNβ-1b-derived epitopes are. Similarly, also the amino acid changes may alter the binding of IFNβ peptides to MHC molecules and T cell recognition [53].

Additional factors in the processing of treatment preparations can influence how the immune system recognizes them. Spontaneously occurring modifications like deamidation, oxidation, and glycation alter the surface and chemical properties of proteins. These modifications even diverge between preparations sharing the identical amino acid sequence, e.g., IFNβ-1a *s.c.* and *i.m.*, by differential production, processing, or storing of the biopharmaceuticals [3]. Other chemical alterations of amino acids like phosphorylation, PEGylation, methylation, or acetylation can be applied during the manufacturing of drugs, e.g., to alter their stability, and also change epitopes, leading to differential binding to allelic variants of HLA heterodimers [54, 55]. Importantly, these modifications also happen after administration of the product in vivo, and glycosylation may well affect the likelihood of them taking place.

In summary, diverging post-translational modifications may contribute to the observed differences in preparation-specific genetic risk factors. Notably, the MHC class II peptide-binding groove is formed by heterodimers of two HLA proteins, likely contributing to the association of haplotypes spanning *HLA* α and β chain genes, like *HLA-DQA1* and *HLA-DQB1*, with IFNβ ADA. However, it is unlikely that preparation-specific risk can entirely be attributed to genetic factors. For example, the dosage and injection frequency of preparations may affect the likelihood of developing ADA, independently of genetic risk [8, 56, 57]. Nevertheless, most patients develop ADA within the first months of IFNβ treatment [58], arguing against pronounced long-term dosage-specific effects and underlining the importance of genetic risk.

Next to having to rely on imputed *HLA* alleles, the low number of available patients that developed ADA under treatment with IFNβ-1a *i.m.*, rendering IFNβ-1a *i.m.*-specific analyses unfeasible, constitutes a limitation of the present study. We expect genetic risk factors for IFNβ-1a *i.m.*-induced ADA to exist, but whether these are independent of IFNβ-1a *s.c.*-associated risk remains to be shown.

The complexity of the genetic risk landscape

Using conditional analyses, we did not find evidence for more than one genetic risk locus for IFNβ-1a *s.c.*-induced ADA. Results from previous studies can thus, at least for Swedish and German patients, be consolidated to the extended haplotype *DR15-DQ6*. In the present dataset, it is impossible to assess whether the combined *DR15-DQ6* haplotype constitutes the real risk factor for IFNβ-1a-*s.c.* or whether any of the single alleles *HLA-DQB1*06:02* or *HLA-DQA1*01:02* convey this risk, with the haplotype showing an association merely because of LD. *DR15-DQ6* ($MAF_{KI} = 0.34$, $MAF_{TUM} = 0.29$) is less common than the two single alleles, especially compared to *HLA-DQA1*01:02* ($MAF_{KI} = 0.42$, $MAF_{TUM} = 0.36$). Because statistical power is dependent on the AF, the slightly lower statistical support for the association of *DR15-DQ6*, compared to the single alleles, likely reflects these differences in AF and power. We thus hypothesize that the combined haplotype *DR15-DQ6* constitutes the primary signal. Nevertheless, such fine-mapping and the differentiation between the correlated alleles is irrelevant for risk predictions. Because of the strong correlation of alleles observed within the extended haplotype, any of these alleles can reliably be used as a proxy for the others in prediction models.

Similarly, conditional analyses support the association of the extended haplotype *DR3-DQ2* as the primary protective genetic signal for IFNβ-1a *s.c.*, without evidence for secondary signals. However, in the present sample, the association of this haplotype cannot be separated from *HLA-DQB1*02:01*. By contrast, genetic risk for IFNβ-1b *s.c.*-induced ADA appears to be more complicated. The association of the haplotype *DR4-DQ3* could not fully explain the signal of its allele *HLA-DRB1*04:01*. Moreover, we found evidence for a secondary signal in stepwise conditional regression analyses. Notably, the prediction models for IFNβ-1b *s.c.* did not perform as well as the prediction for IFNβ-1a *s.c.*-induced ADA did, which possibly reflects this more complex risk landscape. To truly unravel an additional potential polygenic contribution to ADA risk, the current study still lacked the sample size necessary for reliably detecting polygenic variants with their expected small effect sizes [59, 60].

Population-specific risk differences

Alleles from all three associated haplotypes, *HLA-DRB1*15:01*, *HLA-DRB1*04:01*, and the protective *HLA-DRB1*03:01*, are concurrent risk factors for MS [42]. The unfortunate coincidence that *HLA-DRB1*15:01* and *HLA-DRB1*04:01* are enriched among MS patients and also constitute ADA risk factors likely contributes to the high incidence of IFNβ ADA among MS patients.

Interestingly, these alleles also show substantial population-specific differences [47, 48, 61]: The IFNβ-1a *s.c.* risk allele *HLA-DRB1*15:01* is less frequent in

Southern Europe and for Ashkenazi, Southern Hispanic, and African ancestries (e.g., Italy 5.6–6.4%, Southwestern Spain 5.2–8.6%). At the same time, it is especially frequent in individuals with ancestry from other parts of Europe (e.g., Northern Spain 16.7–32.1%, Germany 12.9–17.2%, Sweden 16.1%). Note that being the most important MS risk variant, allele *HLA-DRB1*15:01* is more frequent among MS patients than in the respective general population. However, population-specific frequency differences exist on top: In the present study, the frequency of *HLA-DRB1*15:01* was markedly higher in Swedish (36.1%) than in German MS patients (30.9%). On average, Swedish patients may thus, in comparison to German patients, be at higher risk of developing nADA against IFNβ-1a *s.c.* The IFNβ-1b *s.c.* risk allele *HLA-DRB1*04:01* is more frequent in parts of Northwestern, Northern, and Central Europe (e.g., England 12.4–13.5%, Denmark 17.6%, Sweden 13.7%) than in Southern Europe and most other ancestries (e.g., Italy 1.7–4.1%, Spain 2.0–3.8%). Germany lies in between with frequencies of 6.8–9.4%.

While the frequencies of both risk alleles for IFNβ-1a *s.c.*- and IFNβ-1b *s.c.*-induced ADA thus roughly decrease along a North-South gradient within Europe, their relative frequencies differ sharply in some ancestries (Additional file 27). For example, in Northern Spain, the major genetic risk factor for IFNβ-1a *s.c.*-specific ADA occurs > 8 times more often than the one for IFNβ-1b *s.c.*-induced ADA. Such substantial, population-specific differences in risk allele frequencies likely exist for ADA against any biopharmaceutical. If genetic risk factors for a biopharmaceutical are known, and personalized genotyping data for patients are not available, recommendations for the choice of a specific treatment preparation could thus be made on a population level. Where the availability of genetic testing is limited, patients from populations with higher frequencies of risk alleles could be prioritized for genetic testing, as already practiced for other treatments [62].

Genetic factors contributing to nADA titers

The same or highly correlated risk factors contributed to the presence of nADA and the magnitude of nADA titers and bADA levels. The heritability of nADA, explained by common variants, after correction for confounders like treatment preparation and duration, sex, and age, was very high—$h^2_{gl} = 0.79$ on a liability scale. This result underlines the importance of genetic factors in the occurrence of IFNβ ADA. Although nADA titers need to cross a threshold to become functionally relevant, the associations of genetic risk factors with both nADA presence and titers may indicate that most genetic risk factors mainly influence the likelihood of developing ADA and less the absolute titers. Interestingly, the association of the candidate variants *HLA-DRB1*04:08* and *HLA-DRB1*16:01* was only significant in analyses of quantitative nADA titers or bADA levels but not for nADA presence. This finding indicates that genetic factors influencing the amount of ADA likely exist. In fact, follow-up analyses found that these two alleles were also associated with nADA titers in nADA-positive patients treated with IFNβ-1b *s.c.* In the present study, we did not conduct hypothesis-free GWAS of nADA titers in the smaller subsample restricted to nADA-positive patients. To reliably distinguish between influences of genetic variants on either the likelihood of nADA development or the amount of nADA, larger patient samples than analyzed in the present study should be collected for future studies.

Comparison to other MS- and ADA-related analyses

The allele *HLA-DRB1*15:01* is associated with the risk of MS [42], earlier age at MS disease onset [24], and developing nADA against IFNβ-1a *s.c.* Moreover, the same allele and associated haplotypes are also associated with intrathecal immunoglobulin G levels [63] and Epstein Barr viral loads and titers in MS patients [64, 65]. By contrast, a strong negative association between *HLA-DRB1*15:01* and JC polyomavirus antibody status was reported [66]. The *HLA* allele *HLA-DRB1*15:01*, therefore, constitutes the key genetic risk factor for MS, which also differentially influences gene-by-environment interactions, disease severity, and treatment complications.

We identified *HLA-DQA1*05:01* to protect from nADA against IFNβ-1a *s.c.* Interestingly, the same allele is strongly associated with the risk of ADA against the widely used anti-tumor necrosis factor (TNF) treatments for Crohn's disease [67]. This association was consistent across the two anti-TNF biopharmaceutical drugs adalimumab and infliximab. The *HLA* allele *HLA-DRB1*03*, also protective against IFNβ-1a *s.c.*-induced nADA, is, together with *HLA-DQA1*05:01*, part of the haplotype *DR3-DQ2*. *HLA-DRB1*03* was published as a risk factor against adalimumab and infliximab in patients suffering from either inflammatory bowel disease or rheumatoid arthritis [68, 69]. Whether treatment-specific genetic risk factors also exist for anti-TNF biopharmaceuticals and whether the haplotype *DR3-DQ2* or one of the single *HLA* alleles confers the risk for anti-TNF ADA could be an interesting topic of future studies.

Prediction of ADA

The prediction models performed better for IFNβ-1a *s.c.*-induced than for IFNβ-1b *s.c.*-induced nADA, and they could predict nADA better in the Swedish KI than in the

German TUM cohort. Our results indicate that, compared to IFNβ-1b *s.c.*, genetic risk for IFNβ-1a *s.c.*-induced ADA is more dominated by a single locus. Overall, more patients receiving IFNβ-1a *s.c.* than IFNβ-1b *s.c.* were analyzed (1145 vs. 1010). Both factors likely contributed to better prediction models and performance in IFNβ-1a *s.c.*-treated patients. The top-associated ADA risk SNP was more frequent in Swedish patients than in German ones (43% vs. 40%), and the Swedish sample contained more patients treated with IFNβ-1a *s.c.* (Sweden 590, Germany 558), of whom more were nADA-positive (34.6% vs. 33.7%). Although these individual differences were small, they may have contributed to prediction models performing better in the Swedish dataset.

Contrasting the samples in the top and bottom percentiles of polygenic risk score distributions is a common practice to compare individuals carrying a high genetic risk burden to the ones not at risk [46]. The sensitivity and specificity reached in the comparison of individuals in the top 30% nADA risk group compared to the bottom 30% (0.78 and 0.90, respectively, in Swedish IFNβ-1a *s.c.* patients and 0.80 and 0.76, respectively, in German patients) may still not be sufficient for a routine clinical test. However, these prediction models could certainly be optimized by the inclusion of additional predictive factors, e.g., body mass index [70], not available in the present retrospective setting. The significant predictive improvement of the genetic risk model compared to a model containing only demographic and clinical variables (Table 7) underlines the importance of incorporating genetics in prediction models for ADA. The high odds of patients at genetic risk for nADA (Sweden: OR = 73.9, Germany: OR = 13.8) support the use of genetic stratification as a personalized medicine tool—patients at high genetic risk should either switch to a different drug or be monitored more closely, as suggested for other conditions [71].

Conclusions

We have conducted a comprehensive characterization of genetic risk for IFNβ-induced ADA, consolidating previous research. Next to treatment-specific risk factors, we described ancestry-specific effects relevant for treatment choice in specific populations. Our robust prediction models could be employed for personalized medicine, guiding treatment recommendations, and efficient nADA testing regimes. Importantly, our study can serve as a blueprint for the analysis of genetic factors influencing ADA against other biopharmaceuticals and in the context of further diseases.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12916-020-01769-6.

Additional file 1. Previous measurements and design of new ADA measurements. Previous ADA measurements in the Swedish KI and German TUM cohorts per treatment preparation and distribution of samples for the new ADA measurements. For part of the TUM patients, only previous bADA measurements were available.

Additional file 2. Additional details supporting the Methods section

Additional file 3. New ADA measurements and design of the datasets for analyses. New ADA measurements in the Swedish KI and German TUM cohorts per treatment preparation and assignments of samples into the discovery and replication datasets. In the discovery and replication datasets, the first number indicates nADA and the second number bADA measurements. The distinction into negative and positive patients was made using nADA measurements.

Additional file 4. Visualization and analysis of population stratification. For detailed figure legends, see the file.

Additional file 5. Manhattan plots of the GWAS across IFNβ preparations. Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance; the top genome-wide significant variant is labeled with a red diamond.

Additional file 6. Manhattan plots of the GWAS on patients treated with IFNβ-1a *s.c.* Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance; the top genome-wide significant variant is labeled with a red diamond.

Additional file 7. Manhattan plots of the GWAS on patients treated with IFNβ-1b *s.c.* Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance; the top genome-wide significant variant is labeled with a red diamond.

Additional file 8. Manhattan plots of the MHC region of the GWAS across IFNβ preparations. Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS, showing only the MHC region. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance. For (A-C) discovery-stage plots, the prioritized variants are labeled with red diamonds for (D-F) replication-stage plots, the top genome-wide significant variant is labeled with a red diamond, and for (G-I) pooled discovery + replication plots, the replicated variants are labeled with red diamonds, and the top pooled variant is labeled in magenta.

Additional file 9. Manhattan plots of the MHC region of the GWAS on patients treated with IFNβ-1a *s.c.* Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS, showing only the MHC region. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance. For (A-C) discovery-stage plots, the prioritized variants are labeled with red diamonds, for (D-F) replication-stage plots, the top genome-wide significant variant is labeled with a red diamond, and for (G-I) pooled discovery + replication plots, the replicated variants are labeled with red diamonds, and the top variant from the pooled analysis is labeled in magenta.

Additional file 10. Manhattan plots of the MHC region of the GWAS on patients treated with IFNβ-1b *s.c.* Manhattan plots of the (A-C) discovery-stage, (D-F) replication-stage, and (G-I) pooled discovery + replication GWAS, showing only the MHC region. The red line between $-\log_{10}p = 7$ and $-\log_{10}p = 8$ indicates genome-wide significance. For (A-C) discovery-stage plots, the prioritized variants are labeled with red diamonds, for (D-F) replication-stage plots, the top genome-wide significant variant is labeled with a red diamond, and for (G-I) pooled discovery + replication plots, the replicated variants are labeled with red diamonds, and the top variant from the pooled analysis is labeled in magenta.

Additional file 11. Genomic inflation factors for all GWAS. Lambda = Median genomic inflation factor.

Additional file 12. Table of the top GWAS associations. Variants prioritized in the discovery GWAS (bold font if replicated) and top

variants from the pooled analysis of discovery + replication data. All effect sizes are relative to the minor allele. Bp = base pairs, MAF = minor allele frequency, beta = regression effect size, SE = standard error, P = *p*-value, cond. = conditional analysis, R2 = linkage disequilibrium r^2.

Additional file 13. Regional association plots of the top GWAS variants in the analysis across IFNβ preparations. Regional association plots of variants from the GWAS generated using LocusZoom v1.4 and the 1000 Genomes 1000G_Nov2014 EUR reference panel [72]. The color of dots indicates LD with the lead variant (pink). Gray dots represent signals with missing LD r^2 values. If no LD information was present in the database on the top variant, LD with the variant showing the second-lowest *p*-value is indicated. The gray line indicates genome-wide significance. cM: centimorgan, chr: chromosome, Mb: mega base pairs.

Additional file 14. Forest plots of the top GWAS variants and *HLA* alleles in the analysis across IFNβ preparations. Green: IFNβ-1a *s.c.*, blue: IFNβ-1a *i.m.*, orange: IFNβ-1b *s.c.*, magenta: pooled discovery–/replication-stage analyses. D. = discovery, R. = replication, P. = pooled discovery + replication.

Additional file 15. Results from stepwise conditional analyses. Results from stepwise conditional analyses on the pooled discovery and replication data. Genome-wide significant *p*-values are labeled in bold font. All effect sizes are relative to the minor allele. Bp = base pairs, MAF = minor allele frequency, beta = regression effect size, SE = standard error, P = *p*-value, R2 = linkage disequilibrium r^2.

Additional file 16. Results from eQTL analyses. Summary statistics as downloaded from GTEx v8 (https://gtexportal.org/). Significance thresholds are shown in the column *Gene-level P threshold*.

Additional file 17. Results from MAGMA gene set analyses. FDR = 5% false discovery rate.

Additional file 18. Regional association plots of the top GWAS variants in the analysis of IFNβ-1a s.c.-treated patients. Regional association plots of variants from the GWAS generated using LocusZoom v1.4 and the 1000 Genomes 1000G_Nov2014 EUR reference panel [72]. The color of dots indicates LD with the lead variant (pink). Gray dots represent signals with missing LD r^2 values. If no LD information was present in the database on the top variant, LD with the variant showing the second-lowest *p*-value is indicated. The gray line indicates genome-wide significance. cM: centimorgan, chr: chromosome, Mb: mega base pairs.

Additional file 19. Forest plots of the top GWAS variants and *HLA* alleles in the analysis of IFNβ-1a s.c.-treated patients. Green: IFNβ-1a *s.c.*, blue: IFNβ-1a *i.m.*, orange: IFNβ-1b *s.c.*, magenta: pooled discovery–/replication-stage analyses. D. = discovery, R. = replication, P. = pooled discovery + replication.

Additional file 20. Regional association plots of the top GWAS variants in the analysis of IFNβ-1b s.c.-treated patients. Regional association plots of variants from the GWAS generated using LocusZoom v1.4 and the 1000 Genomes 1000G_Nov2014 EUR reference panel [72]. The color of dots indicates LD with the lead variant (pink). Gray dots represent signals with missing LD r^2 values. If no LD information was present in the database on the top variant, LD with the variant showing the second-lowest *p*-value is indicated. The gray line indicates genome-wide significance. cM: centimorgan, chr: chromosome, Mb: mega base pairs.

Additional file 21. Forest plots of the top GWAS variants and *HLA* alleles in the analysis of IFNβ-1b s.c.-treated patients. Green: IFNβ-1a *s.c.*, blue: IFNβ-1a *i.m.*, orange: IFNβ-1b *s.c.*, magenta: pooled discovery–/replication-stage analyses. D. = discovery, R. = replication, P. = pooled discovery + replication.

Additional file 22. Association statistics of all replicated *HLA* alleles. AF = allele frequency, beta = regression effect size, SE = standard error, P = *p*-value, cond. = conditional analysis.

Additional file 23. Results from analyses of previously published candidate SNPs and *HLA* alleles. Variants, alleles, and the respective *p*-values are labeled in bold font if they showed a one-sided $p < 2.5 \times 10^{-3}$ (Bonferroni correction for 20 tests) either in the pooled analysis across treatment preparations or in the analysis of IFNβ-1a *s.c.*-treated or IFNβ-1b *s.c.*- treated patients. "nADA titer in nADA-positive" refers to an analysis of nADA titers restricted to nADA-positive patients. All effect sizes are relative to the minor allele. Bp = base pairs, MAF = minor allele frequency, AF = allele frequency, beta = regression effect size, SE = standard error, P = *p*-value, cond. = conditional analysis.

Additional file 24. Treatment preparation-specific prediction of the presence of nADA in the replication data. Eight PRS, the top single GWAS variant, and the top *HLA* allele from the discovery stage were analyzed in the replication data using the covariates sex, age, treatment preparation, treatment duration, titration site, and eight ancestry components. Upper table: Prediction of the presence of nADA in IFNβ-1a *s.c.*-treated patients from the replication data using all ten prediction models based on analyses for IFNβ-1a *s.c.* in the discovery data. Lower table: Prediction of the presence of nADA in IFNβ-1b *s.c.*-treated patients from the replication data using all ten prediction models based on analyses for IFNβ-1b *s.c.* in the discovery data. Beta = regression effect size, SE = standard error, P = *p*-value.

Additional file 25. Treatment preparation-specific prediction of the presence of nADA in the replication data: performance of single models. Eight PRS, the top single GWAS variant, and the top HLA allele from the discovery stage. Covariates: sex, age, treatment preparation, treatment duration, titration site, and ancestry components. The plots show the area under the receiver operating characteristic curve (AUC) and its 95% confidence interval (CI). Bonferroni = significant after Bonferroni correction for multiple testing; nominal = nominally significant ($p < 0.05$); n.s. = not significant

Additional file 26. Treatment preparation-specific prediction of the presence of nADA in the replication data: comparison of top models. For each top model, the plots show either the AUC and its 95% CI or Nagelkerke's pseudo-R^2 and its 95% CI. Boxes show the prediction groups (replication data) and columns within each box the training data groups (discovery data). Bonferroni = significant after Bonferroni correction for multiple testing; nominal = nominally significant ($p < 0.05$); n.s. = not significant.

Additional file 27. Comparison of allele frequencies for *HLA-DRB1*15:01* and *HLA-DRB1*04:01*. The allele frequencies (AF) were queried from allelefrequencies.net on July 27th 2020 [48]. All four-digit European Silver and Gold populations with data on both *HLA-DRB1*15:01* and *HLA-DRB1*04:01* were used and populations with relative differences for both alleles are shown (i.e., with an AF above or below the average for one allele without the other allele being in the same group). Populations with an AF below the average for *HLA-DRB1*15:01* and above the average for *HLA-DRB1*04:01* are colored in green. Populations with an AF above the average for *HLA-DRB1*15:01* and below the average for *HLA-DRB1*04:01* are colored in magenta. In addition, the European populations with the highest or lowest AF for the respective allele (if not already present) as well as the largest German population and the Swedish SweHLA sample [47] are shown in gray.

Abbreviations

ADA: Anti-drug antibodies; AF: Allele frequency; AUC: Area under the receiver operating characteristic curve; bp: Base pairs; bADA: Binding ADA; CI: 95% confidence interval; eQTL: Expression quantitative trait locus; GWAS: Genome-wide association study; IFNβ: Interferon β; *i.m.*: Intramuscular; HLA: Human leukocyte antigen; kbp: Kilobase pairs; KI: Karolinska Institutet Stockholm, Sweden; LD: Linkage disequilibrium; MAD: Median absolute deviation; MHC: Major histocompatibility complex; MAF: Minor allele frequency; MDS: Multidimensional scaling; MS: Multiple sclerosis; nADA: Neutralizing ADA; OR: Odds ratio; *s.c.*: Subcutaneous; PRS: Polygenic risk scores; SD: Standard deviation; SE: Standard error; SNP: Single nucleotide polymorphism; TUM: Technical University of Munich, Germany

Acknowledgements

We are thankful for the excellent technical work preparing samples and measuring nADA by the technical staff, Birgit Kassow and Ulla Abildtrup, of DMSC, Department of Neurology, Rigshospitalet, University of Copenhagen, Denmark; Marlies Jank of the Department of Neurology, Medical University of Innsbruck, Austria; and Anna Mattsson, of Karolinska Institutet, Stockholm, Sweden. We thank Niek de Vries, Faculty of Medicine, University of Amsterdam, Netherlands, for helpful discussions on *HLA* haplotypes.

Authors' contributions

TFMA drafted the conceptual study design, administered the project, curated data, devised the methodology for data analysis, conducted and interpreted the formal statistical data analysis, generated the data visualizations, and has prepared and revised the original manuscript draft. JL drafted the conceptual study design, curated and provided data on and investigated the Swedish patients, and revised the manuscript. DM drafted the conceptual study design, acquired funding, recruited and investigated the German patients, curated and provided their data, and revised the manuscript. MR curated data on and investigated the Swedish patients and revised the manuscript. CH curated data on and investigated the Swedish patients and reviewed the manuscript. VG performed the ELISA measurements of bADA levels and acquired data. MA contributed to nADA screening and titration, acquired data, and revised the manuscript. HH contributed to nADA screening and titration, acquired data, and revised the manuscript. LA recruited and acquired data on German patients and revised the manuscript. CG recruited and acquired data on German patients and revised the manuscript. BK recruited and acquired data on German patients and reviewed the manuscript. BMM devised the methodology for data analysis, supervised statistical data analyses, and reviewed the manuscript. PEHJ conducted nADA screening and titration, acquired data, and reviewed the manuscript. FS contributed to nADA screening and titration and reviewed the manuscript. IK provided data on the Swedish patients. TO provided data on the Swedish patients and reviewed the manuscript. MP acquired funding, administered the project, and reviewed the manuscript. SeS drafted the conceptual study design, acquired funding, administered the project, and revised the manuscript. FD drafted the conceptual study design, acquired funding, administered the project, devised the methodology for ADA measurement, conducted nADA screening and titration, acquired data, and revised the manuscript. AFH drafted the conceptual study design, acquired funding, administered and supervised the project, devised the methodology for ADA measurement, provided data on Swedish patients, and revised the manuscript. BH drafted the conceptual study design, acquired funding, supervised the project, provided data on German patients, and revised the manuscript. The author(s) read and approved the final manuscript.

Funding

This study was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115303 (ABIRISK), resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007–2013) and EFPIA companies' in kind contribution.

TFMA, CG, and BH were supported by the German Federal Ministry of Education and Research (BMBF) through the DIFUTURE consortium of the Medical Informatics Initiative Germany (grant 01ZZ1804A) and by the European Union's Horizon 2020 Research and Innovation Programme (grant MultipleMS, EU RIA 733161). TFMA and BMM were supported by the BMBF through the Integrated Network IntegraMent, under the auspices of the e:Med Programme (grant 01ZX1614J). LA received funding from a clinician scientist program by the Munich Cluster for Systems Neurology (SyNergy). CG received a research stipend from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). TO has research funding from the Swedish Research Council, the Swedish Brain Foundation, and Margaretha af Ugglas Foundation.

This research was performed using a reference panel assembled by the Type 1 Diabetes Genetics Consortium (T1DGC), a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF). The data were supplied by the NIDDK Central Repositories. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding authors on reasonable request.

Ethics approval and consent to participate

The ethics committees of the *Regionala etikprövningsnämnden i Stockholm* for Karolinska Institutet, Stockholm, Sweden, and of the Klinikum rechts der Isar, Technical University of Munich, Germany, approved the study, and all participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

TFMA, JL, DM, MR, VG, LA, CG, PEHJ, IK, and MP have no competing interests to declare.

CH is an employee of Sanofi Genzyme.

MA has received speaker honoraria and/or travel grants from Biogen, Novartis, Merck, and Sanofi Genzyme.

HH has participated in meetings sponsored by and received speaker honoraria or travel funding from Bayer, Biogen, Merck, Novartis, Sanofi-Genzyme, Siemens, and Teva, and received honoraria for acting as consultant for Biogen and Teva.

BK received a research grant and travel compensations from Novartis outside the submitted work.

FS has served on scientific advisory boards, been on the steering committees of clinical trials, served as a consultant, received support for congress participation, received speaker honoraria, or received research support for his laboratory from Biogen, Merck, Novartis, Roche, Sanofi Genzyme, and Teva.

TO has received unrestricted MS research grants, and honoraria for advisory boards/lectures from Biogen, Novartis, Sanofi, Merck, and Roche.

SeS is a former employee and has stocks and/or stock options in Novartis.

FD has participated in meetings sponsored by or received honoraria for acting as an advisor/speaker for Almirall, Alexion, Biogen, Celgene, Genzyme-Sanofi, Merck, Novartis Pharma, Roche, and TEVA ratiopharm. His institution has received research grants from Biogen and Genzyme Sanofi. He is section editor of the MSARD Journal (Multiple Sclerosis and Related Disorders).

AFH has received unrestricted research grants from Merck-Serono and BiogenIdec, served as consultant for Johnson & Johnson, and received honoraria for lectures by BiogenIdec and Sanofi-Aventis.

During the last 2 years, BH has served on scientific advisory boards for Novartis; he has served as DMSC member for AllergyCare, Polpharma, and TG therapeutics; he or his institution have received speaker honoraria from Desitin; his institution received research grants from Regeneron for MS research; he holds part of a patent for the detection of antibodies against KIR4.1 in a subpopulation of patients with MS. None of these conflicts are relevant to the topic of the study.

BMM and BH hold parts of a patent for genetic determinants of neutralizing antibodies to interferon.

Author details

[1]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Ismaninger Str 22, 81675 Munich, Germany. [2]Max Planck Institute of Psychiatry, Kraepelinstr 2-10, 80804 Munich, Germany. [3]Department of Clinical Neuroscience, Karolinska Institutet, Visionsgatan 18, 17176 Stockholm, Sweden. [4]Department of Neurology, Medical University of Innsbruck, Anichstr 35, 6020 Innsbruck, Austria. [5]Institute of Experimental Neuroimmunology, Technical University of Munich, Trogerstr 9, 81675 Munich, Germany. [6]Institute of Translational Medicine, University of Liverpool, Crown Street, Liverpool L69 3BX, UK. [7]Munich Cluster for Systems Neurology (SyNergy), Feodor-Lynen-Str. 17, 81377 Munich, Germany. [8]DMSC, Department of Neurology, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark. [9]Inflammation, Microbiome and Immunosurveillance, Université Paris-Saclay, INSERM, Faculté de Pharmacie, rue JB Clément, 92290 Châtenay-Malabry, France. [10]Novartis Institutes for Biomedical Research, Novartis Pharma AG, 4056 Basel, Switzerland. [11]Integrated Biologix GmbH, Steinenvorstadt 33, 4051 Basel, Switzerland.

Received: 14 May 2020 Accepted: 28 August 2020
Published online: 04 November 2020

References

1. Bertolotto A, Deisenhammer F, Gallo P, Sørensen P. Immunogenicity of interferon beta: differences among products. J Neurol. 2004;251(Suppl 2): ii15–24.
2. Sethu S, Govindappa K, Alhaidari M, Pirmohamed M, Park K, Sathish J. Immunogenicity to biologics: mechanisms, prediction and reduction. Arch Immunol Ther Exp. 2012;60(5):331–44.
3. Jawa V, Cousens LP, Awwad M, Wakshull E, Kropshofer H, De Groot AS. T-cell dependent immunogenicity of protein therapeutics: preclinical assessment and mitigation. Clin Immunol. 2013;149(3):534–55.
4. Engelhard VH, Altrich-Vanlith M, Ostankovitch M, Zarling AL. Post-translational modifications of naturally processed MHC-binding epitopes. Curr Opin Immunol. 2006;18(1):92–7.
5. Deisenhammer F. Interferon-beta: neutralizing antibodies, binding antibodies, pharmacokinetics and pharmacodynamics, and clinical outcomes. J Interf Cytokine Res. 2014;34(12):938–45.
6. Bertolotto A, Gilli F, Sala A, Capobianco M, Malucchi S, Milano E, Melis F, Marnetto F, Lindberg RL, Bottero R, et al. Persistent neutralizing antibodies abolish the interferon beta bioavailability in MS patients. Neurology. 2003; 60(4):634–9.
7. Sorensen PS, Ross C, Clemmesen KM, Bendtzen K, Frederiksen JL, Jensen K, Kristensen O, Petersen T, Rasmussen S, Ravnborg M, et al. Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis. Lancet. 2003;362(9391):1184–91.
8. Kappos L, Clanet M, Sandberg-Wollheim M, Radue EW, Hartung HP, Hohlfeld R, Xu J, Bennett D, Sandrock A, Goelz S, et al. Neutralizing antibodies and efficacy of interferon beta-1a: a 4-year controlled study. Neurology. 2005; 65(1):40–7.
9. Pachner AR, Cadavid D, Wolansky L, Skurnick J. Effect of anti-IFNbeta; antibodies on MRI lesions of MS patients in the BECOME study. Neurology. 2009;73(18):1485–92.
10. Barbosa M, Vielmetter J, Chu S, Smith DD, Jacinto J. Clinical link between MHC class II haplotype and interferon-beta (IFN-β) immunogenicity. Clin Immunol. 2006;118(1):42–50.
11. Buck D, Cepok S, Hoffmann S, Grummel V, Jochim A, Berthele A, Hartung HP, Wassmuth R, Hemmer B. Influence of the HLA-DRB1 genotype on antibody development to interferon beta in multiple sclerosis. Arch Neurol. 2011;68(4):480–7.
12. Hoffmann S, Cepok S, Grummel V, Lehmann-Horn K, Hackermuller J, Stadler PF, Hartung HP, Berthele A, Deisenhammer F, Wassmuth R, et al. HLA-DRB1*0401 and HLA-DRB1*0408 are strongly associated with the development of antibodies against interferon-beta therapy in multiple sclerosis. Am J Hum Genet. 2008;83(2):219–27.
13. Link J, Lundkvist Ryner M, Fink K, Hermanrud C, Lima I, Brynedal B, Kockum I, Hillert J, Fogdell-Hahn A. Human leukocyte antigen genes and interferon beta preparations influence risk of developing neutralizing anti-drug antibodies in multiple sclerosis. PLoS One. 2014;9(3):e90479.
14. Weber F, Cepok S, Wolf C, Berthele A, Uhr M, Bettecken T, Buck D, Hartung HP, Holsboer F, Muller-Myhsok B, et al. Single-nucleotide polymorphisms in HLA- and non-HLA genes associated with the development of antibodies to interferon-beta therapy in multiple sclerosis patients. Pharmacogenomics J. 2012;12(3):238–45.
15. Buck D, Andlauer TF, Igl W, Wicklein EM, Muhlau M, Weber F, Kochert K, Pohl C, Arnason B, Comi G, et al. Effect of HLA-DRB1 alleles and genetic variants on the development of neutralizing antibodies to interferon beta in the BEYOND and BENEFIT trials. Mult Scler. 2019;25(4):565–73.
16. Núñez C, Cénit MC, Alvarez-Lafuente R, Río J, Fernández-Arquero M, Arroyo R, Montalbán X, Fernández O, Oliver-Martos B, Leyva L, et al. HLA alleles as biomarkers of high-titre neutralising antibodies to interferon-β therapy in multiple sclerosis. J Med Genet. 2014;51(6):395–400.
17. Stickler M, Valdes AM, Gebel W, Razo OJ, Faravashi N, Chin R, Rochanayon N, Harding FA. The HLA-DR2 haplotype is associated with an increased proliferative response to the immunodominant CD4+ T-cell epitope in human interferon-β. Genes Immun. 2004;5(1):1–7.
18. Kraft P. Curses—winner's and otherwise—in genetic epidemiology. Epidemiology. 2008;19(5):649–51.
19. Hansen BB, Klopfer SO. Optimal full matching and related designs via network flows. J Comput Graph Stat. 2006;15(3):609–27.
20. Pachner AR. An improved ELISA for screening for neutralizing anti-IFN-beta antibodies in MS patients. Neurology. 2003;61(10):1444–6.
21. Hermanrud C, Ryner M, Luft T, Jensen PE, Ingenhoven K, Rat D, Deisenhammer F, Sorensen PS, Pallardy M, Sikkema D, et al. Development and validation of cell-based luciferase reporter gene assays for measuring neutralizing anti-drug antibodies against interferon beta. J Immunol Methods. 2016;430:1–9.
22. Jensen PEH, Warnke C, Ingenhoven K, Piccoli L, Gasis M, Hermanrud C, Fernandez-Rodriguez BM, Ryner M, Kramer D, Link J, et al. Detection and kinetics of persistent neutralizing anti-interferon-beta antibodies in patients with multiple sclerosis. Results from the ABIRISK prospective cohort study. J Neuroimmunol. 2019;326:19–27.
23. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7.
24. Andlauer TF, Buck D, Antony G, Bayas A, Bechmann L, Berthele A, Chan A, Gasperi C, Gold R, Graetz C, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. Sci Adv. 2016;2(6):e1501678.
25. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5–6.
26. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955.
27. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.
28. International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, Bakker PIWd, Walker BD, Ripke S, Brumme CJ, Pulit SL , et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010, 330(6010):1551–1557.
29. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, Raychaudhuri S, de Bakker PI. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One. 2013;8(6):e64683.
30. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84(2):210–23.
31. Giovannoni G, Barbarash O, Casset-Semanaz F, Jaber A, King J, Metz L, Pardo G, Simsarian J, Sorensen PS, Stubinski B, et al. Immunogenicity and tolerability of an investigational formulation of interferon-beta1a: 24- and 48-week interim analyses of a 2-year, single-arm, historically controlled, phase IIIb study in adults with multiple sclerosis. Clin Ther. 2007;29(6):1128–45.
32. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012;28(19):2540–2.
33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
34. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7): 565–9.
35. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011; 88(3):294–305.
36. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–1.
37. Stouffer SA, Suchman EA, Devinney LC, Star SA, Jr RMW: The American soldier: adjustment during army life. Princeton: Princeton University Press; 1949.
38. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015; 348(6235):660–5.
39. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.
40. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25.
41. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, et al. Variation analysis and gene annotation

of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics. 2008;60(1):1–18.
42. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, Raj T, IMSGC, ANZgene, Gourraud PA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013;9(11):e1003926.
43. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, Shah TS, Patsopoulos NA, Alfredsson L, Anderson CA, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015;47(10):1107–13.
44. Andlauer TFM, Guzman-Parra J, Streit F, Strohmaier J, Gonzalez MJ, Gil Flores S, Cabaleiro Fabeiro FJ, Del Rio NF, Perez FP, Haro Gonzalez J, et al. Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. Mol Psychiatry. 2019. https://doi.org/10.1038/s41380-019-0558-2.
45. Andlauer TFM, Mühleisen TW, Hoffstaedter F, Teumer A, Wittfeld K, Teuber A, Reinbold CS, Bülow R, Caspers S, Herms S, et al. Genetic factors influencing a neurobiological substrate for psychiatric disorders. bioRxiv. 2019:774463. https://doi.org/10.1101/774463.
46. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–24.
47. Nordin J, Ameur A, Lindblad-Toh K, Gyllensten U, Meadows JRS. SweHLA: the high confidence HLA typing bio-resource drawn from 1000 Swedish genomes. Eur J Hum Genet. 2020;28(5):627–35.
48. Faviel, McCabe A, Eduardo, Jones J, Takeshita L, Nestor, Glenda, Ramsbottom K, Ghattaoraya G, Alfirevic A, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 2020;48(D1):D783–8.
49. Barnard JG, Babcock K, Carpenter JF. Characterization and quantitation of aggregates and particles in interferon-beta products: potential links between product quality attributes and immunogenicity. J Pharm Sci. 2013;102(3):915–28.
50. Rombach-Riegraf V, Karle AC, Wolf B, Sorde L, Koepke S, Gottlieb S, Krieg J, Djidja MC, Baban A, Spindeldreher S, et al. Aggregation of human recombinant monoclonal antibodies influences the capacity of dendritic cells to stimulate adaptive T-cell responses in vitro. PLoS One. 2014;9(1):e86322.
51. Gallais Y, Szely N, Legrand F-X, Leroy A, Pallardy M, Turbica I. Effect of growth hormone and IgG aggregates on dendritic cells activation and T-cells polarization. Immunol Cell Biol. 2017;95(3):306–15.
52. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, et al. Antibody neutralization and escape by HIV-1. Nature. 2003;422(6929):307–12.
53. Kalluri SR, Grummel V, Hracsko Z, Pongratz V, Pernpeintner V, Gasperi C, Buck D, Hemmer B, Consortium A. Interferon-beta specific T cells are associated with the development of neutralizing antibodies in interferon-beta treated multiple sclerosis patients. J Autoimmun. 2018;88:83–90.
54. Schultz HS, Ostergaard S, Sidney J, Lamberth K, Sette A. The effect of acylation with fatty acids and other modifications on HLA class II:peptide binding and T cell stimulation for three model peptides. PLoS One. 2018;13(5):e0197407.
55. De Groot AS, Scott DW. Immunogenicity of protein therapeutics. Trends Immunol. 2007;28(11):482–90.
56. Ross C, Clemmesen KM, Svenson M, Sorensen PS, Koch-Henriksen N, Skovgaard GL, Bendtzen K. Immunogenicity of interferon-beta in multiple sclerosis patients: influence of preparation, dosage, dose frequency, and route of administration. Danish Multiple Sclerosis Study Group. Ann Neurol. 2000;48(5):706–12.
57. Hemmer B, Stuve O, Kieseier B, Schellekens H, Hartung HP. Immune response to immunotherapy: the role of neutralising antibodies to interferon beta in the treatment of multiple sclerosis. Lancet Neurol. 2005;4(7):403–12.
58. Sorensen PS, Koch-Henriksen N, Ross C, Clemmesen KM, Bendtzen K, Danish Multiple Sclerosis Study G: Appearance and disappearance of neutralizing antibodies during interferon-beta therapy. Neurology 2005, 65(1):33–39.
59. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668–81.
60. International Multiple Sclerosis Genetics Consortium: Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science 2019, 365(6460):eaav7188.
61. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kahari AK, Lundin P, Che H, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 2017;25(11):1253–60.
62. Ferrell PB Jr, McLeod HL. Carbamazepine, HLA-B*1502 and risk of Stevens-Johnson syndrome and toxic epidermal necrolysis: US FDA recommendations. Pharmacogenomics. 2008;9(10):1543–6.
63. Gasperi C, Andlauer TFM, Keating A, Knier B, Klein A, Pernpeintner V, Lichtner P, Gold R, Zipp F, Then Bergh F, et al. Genetic determinants of the humoral immune response in MS. Neurol Neuroimmunol Neuroinflamm. 2020;7(5):e827.
64. Agostini S, Mancuso R, Guerini FR, D'Alfonso S, Agliardi C, Hernis A, Zanzottera M, Barizzone N, Leone MA, Caputo D, et al. HLA alleles modulate EBV viral load in multiple sclerosis. J Transl Med. 2018;16(1):80.
65. Jacobs BM, Giovannoni G, Cuzick J, Dobson R. Systematic review and meta-analysis of the association between Epstein-Barr virus, multiple sclerosis and other risk factors. Mult Scler. 2020. https://doi.org/10.1177/1352458520907901.
66. Sundqvist E, Buck D, Warnke C, Albrecht E, Gieger C, Khademi M, Lima Bomfim I, Fogdell-Hahn A, Link J, Alfredsson L, et al. JC polyomavirus infection is strongly controlled by human leucocyte antigen class II variants. PLoS Pathog. 2014;10(4):e1004084.
67. Sazonovs A, Kennedy NA, Moutsianas L, Heap GA, Rice DL, Reppell M, Bewshea CM, Chanchlani N, Walker GJ, Perry MH, et al. HLA-DQA1*05 carriage associated with development of anti-drug antibodies to infliximab and adalimumab in patients with Crohn's disease. Gastroenterology. 2020;158(1):189–99.
68. Billiet T, Vande Casteele N, Van Stappen T, Princen F, Singh S, Gils A, Ferrante M, Van Assche G, Cleynen I, Vermeire S. Immunogenicity to infliximab is associated with HLA-DRB1. Gut. 2015;64(8):1344–5.
69. Liu M, Degner J, Davis JW, Idler KB, Nader A, Mostafa NM, Waring JF. Identification of HLA-DRB1 association to adalimumab immunogenicity. PLoS One. 2018;13(4):e0195325.
70. Callahan ST, Wolff M, Hill HR, Edwards KM, Keitel W, Atmar R, Patel S, Sahly HE, Munoz F, Paul Glezen W, et al. Impact of body mass index on immunogenicity of pandemic H1N1 vaccine in children and adults. J Infect Dis. 2014;210(8):1270–4.
71. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5):djv036.
72. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336–7.

Publisher's Note

Molecular Psychiatry (2021) 26:1286–1298
https://doi.org/10.1038/s41380-019-0558-2

ARTICLE

Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders

Till F. M. Andlauer[1,2] · Jose Guzman-Parra[3] · Fabian Streit[4] · Jana Strohmaier[4] · Maria José González[5] · Susana Gil Flores[6] · Francisco J. Cabaleiro Fabeiro[7] · Francisco del Río Noriega[8] · Fermin Perez Perez[5] · Jesus Haro González[9] · Guillermo Orozco Diaz[10] · Yolanda de Diego-Otero[3] · Berta Moreno-Küstner[11] · Georg Auburger[12] · Franziska Degenhardt[13] · Stefanie Heilmann-Heimbach[13] · Stefan Herms[13,14] · Per Hoffmann[13,14,15] · Josef Frank[4] · Jerome C. Foo[4] · Jens Treutlein[4] · Stephanie H. Witt[4] · Sven Cichon[14,15] · Manolis Kogevinas[16] · Bipolar Disorder Working Group of the Psychiatric Genomics Consortium · Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium · Fabio Rivas[3] · Fermín Mayoral[3] · Bertram Müller-Myhsok[1,17,18] · Andreas J. Forstner[13,14,19,20] · Markus M. Nöthen[13] · Marcella Rietschel[4]

Received: 21 August 2018 / Revised: 17 July 2019 / Accepted: 24 July 2019 / Published online: 11 November 2019

Abstract

Multiplex families with a high prevalence of a psychiatric disorder are often examined to identify rare genetic variants with large effect sizes. In the present study, we analysed whether the risk for bipolar disorder (BD) in BD multiplex families is influenced by common genetic variants. Furthermore, we investigated whether this risk is conferred mainly by BD-specific risk variants or by variants also associated with the susceptibility to schizophrenia or major depression. In total, 395 individuals from 33 Andalusian BD multiplex families (166 BD, 78 major depressive disorder, 151 unaffected) as well as 438 subjects from an independent, BD case/control cohort (161 unrelated BD, 277 unrelated controls) were analysed. Polygenic risk scores (PRS) for BD, schizophrenia (SCZ), and major depression were calculated and compared between the cohorts. Both the familial BD cases and unaffected family members had higher PRS for all three psychiatric disorders than the independent controls, with BD and SCZ being significant after correction for multiple testing, suggesting a high baseline risk for several psychiatric disorders in the families. Moreover, familial BD cases showed significantly higher BD PRS than unaffected family members and unrelated BD cases. A plausible hypothesis is that, in multiplex families with a general increase in risk for psychiatric disease, BD development is attributable to a high burden of common variants that confer a specific risk for BD. The present analyses demonstrated that common genetic risk variants for psychiatric disorders are likely to contribute to the high incidence of affective psychiatric disorders in the multiplex families. However, the PRS explained only part of the observed phenotypic variance, and rare variants might have also contributed to disease development.

Introduction

Bipolar disorder (BD), characterised by alternating episodes of mania and depression, has a lifetime prevalence of ~1% and is a substantial contributor to disability throughout the world [1]. Nevertheless, reliable data concerning the aetiology of BD remain scarce. The heritability of BD is estimated to be above 70% [2–4], thus demonstrating an important genetic component in the development of the disorder. Genome-wide association studies (GWAS) in case/control samples have reported that single-nucleotide polymorphisms (SNP) with minor allele frequencies (MAF) of ≥ 1% explain a substantial proportion of the genetic risk for BD [5–12]: the heritability explained by such common

Joint last authors: Andreas J. Forstner, Markus M. Nöthen, Marcella Rietschel

Joint first authors: Till F. M. Andlauer, Jose Guzman-Parra, Fabian Streit

Members of Bipolar and Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium are listed in the Supplementary Material

Supplementary information The online version of this article (https://doi.org/10.1038/s41380-019-0558-2) contains supplementary material, which is available to authorised users.

✉ Marcella Rietschel
Marcella.Rietschel@zi-mannheim.de

Extended author information available on the last page of the article

variants (i.e., the SNP heritability) is estimated to be 0.17–0.23 on a liability scale [12]. Common variants also make a substantial contribution to the development of schizophrenia (SCZ) and major depressive disorder (MDD) [13, 14]. These three psychiatric disorders have a shared genetic component, whereby relatives of patients with BD have, in addition to BD, an increased risk for MDD and SCZ [15]. In fact, GWAS have shown that many genetic risk variants are associated with all three disorders [16–21].

Besides common variants with small individual effects, rare variants with larger effects may also contribute to BD development [22, 23]. In theory, such rare variants should be enriched in families with a high prevalence of illness, termed multiplex families, in comparison to unrelated BD cases. However, it remains unclear whether and to what extent disease incidence in multiplex families is caused by rare variants, a high load of common variants, or a combination of both.

To elucidate the molecular genetic causes of BD, we established the Andalusian Bipolar Family (ABiF) study in 1997, which recruited BD multiplex families [24–26]. In the present analyses, we first investigated whether common genetic variants make a significant contribution to the occurrence of BD in ABiF families. Next, we examined whether BD development was attributable to (a) BD-specific risk variants, (b) variants conferring risk to all three disorders BD, MDD, and SCZ, or, (c) a combination of both. To this end, polygenic risk scores (PRS) based on GWAS of BD, MDD, and SCZ were calculated for and compared between ABiF family members and unrelated BD cases and unrelated controls from the same population. Because of the strong genetic correlation between BD, SCZ, and MDD, standard PRS for BD cannot distinguish between BD-specific risks and factors shared between these disorders. To differentiate between genetic risk shared across and specific to any of the three disorders, we calculated PRS of disorder-specific risk variants using genome-wide inferred statistics (GWIS) and PRS of shared risk variants. To evaluate the possibility that population or technical differences between cohorts biased the results on psychiatric PRS, PRS for late-onset Alzheimer's disease (LOAD) and simulated PRS were analysed as negative controls. Assuming a polygenic model with a contribution of common risk variants, we expected increased psychiatric PRS in the ABiF family members compared to unrelated samples and increased psychiatric PRS in patients compared to controls.

Materials and methods

Sample description

The ABiF study recruited BD multiplex families in Andalusia, Spain [24–26]. The present analyses included 395 members of 33 ABiF families. Diagnoses were assigned by two trained clinicians according to the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria using the best estimate approach [24]. Diagnoses comprised (Table 1 and Supplementary Table S1): BD, $n = 166$ (families $(FAM)_{BD}$; BD type I (BD-I): $n = 115$; BD type II (BD-II): $n = 41$; not otherwise specified (NOS) BD: $n = 10$); MDD, $n = 78$ MDD (FAM_{MDD}); no history of an affective disorder, $n = 151$ ($FAM_{unaffected}$). Six unaffected individuals with a history of substance abuse were excluded from the analyses. Forty-four subjects married into the families and had no parent in the ABiF cohort (36 unaffected; 8 MDD). Furthermore, an independent, previously reported Spanish BD case/control (CC) sample was analysed. Here, BD cases (CC_{BD}) were recruited from consecutive clinical admissions and BD was diagnosed, as in the ABiF families, using DSM-IV [9]; unrelated control individuals ($CC_{controls}$) were recruited in the framework of the longitudinal European Community Respiratory Health Survey (ECRHS) study. Blood for genotyping was acquired at the ECRHS2 assessment in 2000–2001. After quality control (QC), the combined data set of both cohorts comprises data from 384 FAM (163 FAM_{BD}, 73 FAM_{MDD}, 142 $FAM_{unaffected}$, and 6 $FAM_{unaffected}$ with a history of substance abuse) and 438 CC subjects (161 unrelated BD cases; BD-I: $n = 156$; BD-II: $n = 5$) and 277 unrelated controls. Of the 161 CC_{BD} cases, 59 (36.6%) reported a family history of BD. However, in contrast to the data collection in the ABiF families, this information relied only on the self-report by the respective CC_{BD} patient, and was not validated via an interview of further family members. BD diagnoses were not available for the unrelated controls, but the self-reported prevalence of current depression in this cohort was 3.3% at the time of genotyping and the self-reported prevalence of lifetime depression was 14.4% at the follow-up 10 years after genotyping, indicating that the cohort is fairly representative of a typical population in regard to the prevalence of depression [27].

Note that, while all subjects passed QC in the family-only sample, 11 family members were excluded during QC of the joint sample because they showed significant differences in autosomal heterozygosity from the mean. Reported numbers of subjects thus differ slightly for different comparisons. The joint data set contained 35 unaffected, married-in family members who were excluded from analyses using the combined sample (unless specified otherwise). A detailed description of QC procedures is provided in the Supplementary Methods.

The study was approved by the respective local ethics committees (Comités de ética de la investigación provincial de Cádiz, Córdoba, Granada, Jaén and Málaga), and all participants provided written informed consent. For five

Table 1 Characteristics of 389 individuals from the 33 ABiF families, the 161 unrelated bipolar cases (CC_{BD}), and the 277 unrelated controls ($CC_{controls}$)

	FAM_{BD} ($n = 166$)	FAM_{MDD} ($n = 78$)	$FAM_{unaffected}$ ($n = 145$)	CC_{BD} ($n = 161$)	$CC_{controls}$ ($n = 277$)
Median age at interview (MAD)	40.5 (12.5)	44.5 (11.5)	44 (15)[d]	44 (11)[d]	43 (6)[b,c] *Missing = 3*
Median age at onset (MAD)	20 (5)[c] *Missing = 3*	26 (8)[a] *Missing = 1*		23 (6)[a]	
Female sex *n* (%)	103 (62.1)[b,d]	54 (69.2)[b,d]	51 (35.2)[a,c,d]	91 (56.5)[b]	132 (47.7)[a,b]
Married-in *n* (%)	0 (100.0)	8 (10.0)	36 (24.8)		
Educational level					
Primary school	118 (71.5)[c]	53 (68.0)	102 (70.3)[c]	88 (54.7)[a,b]	
Secondary school	39 (23.6)[c]	18 (23.1)[c]	36 (24.8)[c]	58 (36.0)[a,b]	
University degree	8 (4.9) *Missing = 1*	7 (9.0)	7 (4.8)	15 (9.3)	
Severe impairment during disorder	105 (65.6)[c] *Missing = 6*	4 (5.6)[a,c] *Missing = 6*		160 (99.4)[a]	
History of psychosis	110 (66.3)[c]	4 (5.1)[a,c]		159 (98.8)[a]	
History of suicide attempts	41 (24.7)[b]	2 (2.6)[a,c]	1 (0.7)[a,c] *Missing = 1*	40 (24.8)[b]	

Six unaffected individuals with a history of substance abuse were excluded from the analyses and are not shown in this table. Age and age at onset were analysed using Mann–Whitney *U*-tests; median and median absolute deviation (MAD) are shown. Categorical values were analysed using chi-squared (χ^2) tests with two degrees (education) or one degree (other) of freedom; number (*n*) and percentage (%) of subjects are shown. *Missing*: number of individuals with missing data. All subjects passed QC in the FAM sample (numbers as shown in the table), but 11 family members were excluded during QC of the joint sample, therefore reported numbers differ slightly between comparisons. Note that the unaffected, married-in family members were excluded from analyses of the combined data set (FAM + CC sample) unless specified otherwise. Differences between the following groups were at least nominally significant (for details and *p*-values adjusted for multiple testing see Supplementary Table 1)

[a]Different from FAM_{BD}

[b]Different from $FAM_{unaffected}$

[c]Different from CC_{BD}

[d]Different from $CC_{controls}$

adolescents (age 15–17 years), written informed consent was also obtained from the parents.

Genotyping and imputation

Genome-wide genotyping of the FAM sample was carried out using the Illumina Infinium PsychArray BeadChip (PsychChip). QC and population substructure analyses were performed in PLINK v1.9 [28], as described in the Supplementary Methods. Genotyping and basic QC of the CC sample were conducted previously and are described elsewhere [9]. The study used two genotype data sets: Analyses of family members by themselves used variants genotyped on the PsychChip. For analyses on the combined FAM + CC sample, the genotype data of the CC data set were, for the variants genotyped in both samples, merged with the genotype data of the FAM sample. Both genotype data sets (family-only and combined) were imputed independently to the 1000 Genomes phase 3 reference panel using SHAPEIT and IMPUTE2 [29–31]. After imputation and post-imputation QC, the combined data set of both cohorts contained 6,862,461 variants with an INFO metric of ≥ 0.8 and a MAF of ≥ 1%. The imputed FAM data set without the CC subjects contained 8,628,089 variants.

Calculation of polygenic risk scores

PRS were calculated in *R* v3.3 [32] using imputed genetic data. For each PRS, the effect sizes of variants below a selected *p*-value threshold, both obtained from large GWAS (training data), were multiplied by the imputed SNP dosage in the test data and then summed to produce a single PRS per threshold. Test statistics and alleles in the GWAS training data were flipped so that effect sizes were always positive. Thus, the PRS represent cumulative, additive risk. PRS were scaled to represent the relative risk load (minimum possible cumulative risk load = 0, maximum = 1). For each disorder, ten PRS based on different GWAS *p*-value thresholds ($<5 \times 10^{-8}$, $<1 \times 10^{-7}$, $<1 \times 10^{-6}$, $<1 \times 10^{-5}$, $<1 \times 10^{-4}$, <0.001, <0.01, <0.05, <0.1, <0.2) were calculated. The number of SNPs used for each PRS is shown in

Supplementary Table S2. For additional details, see the Supplementary Methods.

For BD, MDD, and SCZ diagnoses, summary statistics of GWAS by the Psychiatric Genomics Consortium (PGC) were used as training data. For BD, the data freeze contained 20,352 cases and 31,358 controls [12]. As selected index patients from the ABiF families and the unrelated Spanish BD case/control data set were part of this BD GWAS, we recalculated summary statistics for this PGC GWAS without these Spanish samples, to prevent false-positive results caused by sample overlap between training and test samples. For MDD and SCZ, published data sets were used. These contained 130,664 cases and 330,470 controls for MDD [14] and 33,640 cases and 43,456 controls for SCZ [13]. There was no overlap between the subjects included in those GWAS and the ABiF and Spanish case/control samples. Variants with an INFO metric of < 0.6 in the GWAS summary statistics were removed.

Shared psychiatric PRS were generated using all variants showing an association at $p < 0.05$ in the GWAS of BD, SCZ, and MDD and for which effect sizes pointed in the same direction across studies. For this shared set of variants, *p*-values and effect sizes, used as weights in the PRS, were obtained using random-effects meta-analysis. PRS were then calculated using the meta-analysis summary statistics. We generated disorder-specific summary statistics to assess genetic risk unique to each disorder. To this end, genome-wide inferred statistics (GWIS) were calculated as explained in detail elsewhere [33]. For example, we calculated BD GWAS summary statistics corrected for the MDD GWAS results (BD-MDD). These BD-MDD GWIS results are similar to results obtained from a conditional analysis for BD corrected for MDD. They represent a genetic unique BD liability, which is estimated based on the heritability of BD and the coheritability of BD and MDD, both estimated using LD score regression [34]. As recommended for this method, variants with an INFO metric of <0.9 or >1.1 were removed. Disorder-specific PRS, e.g., BD-MDD PRS, were then calculated based on the corresponding GWIS summary statistics.

To confirm whether family members and BD cases had an increased PRS specifically for the tested psychiatric disorders but not because of population or technical differences between cohorts, PRS for late-onset Alzheimer's disease (LOAD) were calculated as a negative control, based on a GWAS by the International Genomics of Alzheimer's Project (IGAP) with 17,008 cases and 37,154 controls [35]. For additional details, see the Supplementary Methods. Furthermore, 10,000 simulated PRS for each of the ten *p*-value thresholds were calculated as negative controls. To this end, random variants from across the genome were drawn, using the same number of variants as for the BD PRS at each threshold and random effect sizes from the pool of all available BD, SCZ, and MDD effects. The code for simulating PRS is available at: https://gitlab.com/tillandlauer/abif-prs-analyses/.

Statistical analysis

PRS analyses on binary variables (e.g., diagnoses and comparisons between cohorts) were conducted in *R* with the function *glmm.wald* of the package *GMMAT*, using a logistic mixed model, fitted by maximum likelihood using Nelder–Mead optimisation [36] to account for family structure. For logistic models, PRS underwent *Z*-score standardisation to generate comparable odds ratios (OR). Family structure was modelled as a random effect, with a genetic relationship matrix calculated on pruned genotype data in GEMMA [37].

Linear mixed models (LMMs) taking family structure into account were calculated using the function *polygenic* of the package *GenABEL* [38] for analyses of quantitative variables (anticipation and age at onset). In these analyses, test statistics, including 95% confidence intervals (CI), were calculated using bootstrapping (package *boot* [39, 40]) and *p*-values were validated using permutation analysis (10,000 permutations). In these permutation analyses, the null distribution of test statistics was empirically determined by repeating regression analyses 10,000 times with random sampling of phenotype data. To calculate a *p*-value, the number of tests were counted where a model with a random genotype-phenotype association showed the same or a more extreme *p*-value than the correct, non-randomised model and this number was divided by the total number of tests (10,000).

For each analysis of PRS, all ten PRS *p*-value thresholds were analysed. In analyses of the combined FAM and CC data set, sex was used as a covariate. In the analysis of FAM data alone, sex and age at the time of the interview were used as fixed effects covariates; whether an individual had married into the family was incorporated as a second random effect. Following the hypothesis that family members or subjects with a psychiatric diagnosis have increased PRS for psychiatric disorders, one-sided *p*-values were calculated for all PRS-based analyses. In all analyses, *p*-values below the significance threshold $\alpha = 0.05$ were considered as nominally significant. Unless otherwise stated, this threshold was corrected for $10 \times 6 = 60$ tests using the Bonferroni method ($\alpha = 0.05/60 = 8.33 \times 10^{-4}$). For further details, see the Supplementary Methods.

To determine whether population or technical differences might have influenced the observed effects independently of diagnosis groups, simulated PRS, generated as described above, were analysed. For each model, association statistics of the 10,000 simulated PRS were calculated for the ten *p*-value thresholds; the disorder PRS at the threshold showing

the lowest mean association p-value was analysed further: The number of simulated PRS at this threshold that showed the same or a stronger association was counted and compared to the association of the disorder PRS. This count was used as the number of successes in a binomial test to estimate the probability of success. For computational efficiency, models were fitted using restricted maximum likelihood estimation and the average information optimisation algorithm for this analysis.

Results

Figures 1 and 2 show the test statistics for the PRS with the training GWAS p-value threshold p_{PRS} that showed the strongest association per PRS type. Full results for all ten p_{PRS} per PRS type calculated using logistic mixed models are provided in Supplementary Figs. S1–S12 and in Supplementary Tables S3–S11.

FAM_{BD} cases had higher psychiatric PRS than controls from the general population

On average, familial FAM_{BD} cases had higher BD PRS than unrelated $CC_{controls}$ across the p_{PRS} thresholds (Fig. 1a, b; Supplementary Figs. S1 and S2 and Supplementary Table S3). The most substantial support for an increased BD PRS was found with the threshold $p_{PRS} = 0.1$ (OR = 2.97, one-sided $p = 1.9 \times 10^{-11}$). FAM_{BD} cases also had significantly higher SCZ PRS than $CC_{controls}$; the increase of the MDD PRS was nominally significant (Fig. 1b).

Shared PRS generated from the variants associated jointly with BD, SCZ, and MDD were significantly increased at $p_{PRS} \geq 0.01$ in FAM_{BD} cases compared to $CC_{controls}$. The GWIS BD-MDD PRS—the BD PRS corrected for associations shared with MDD—were significantly increased in FAM_{BD} cases compared to $CC_{controls}$. All other disorder-specific GWIS PRS were not significantly higher in FAM_{BD} cases after correction for multiple testing.

No significant increase was found for the negative-control PRS for late-onset Alzheimer's disease (LOAD) and associations of the PRS for BD and SCZ and of the *Shared* PRS were significantly stronger than simulated PRS in FAM_{BD} compared to $CC_{controls}$ (Table 2).

FAM_{BD} cases had higher BD PRS than unrelated CC_{BD} cases

The BD PRS was significantly higher in FAM_{BD} than in CC_{BD} cases at $p_{PRS} \geq 0.05$, but no other type of PRS was increased in FAM_{BD} compared to CC_{BD} cases (Fig. 1c, d). The association of the BD PRS was significantly stronger than simulated PRS (Table 2).

Unaffected family members showed higher psychiatric PRS than CC controls

In the comparison of $FAM_{unaffected}$ to $CC_{controls}$, PRS for BD and SCZ were significantly higher in unaffected family members (Fig. 1e, f). The increases of the MDD, *Shared*, BD-MDD, and SCZ-BD GWIS PRS were nominally significant. The associations of BD and SCZ PRS were significantly stronger than the associations of simulated PRS (Table 2).

FAM_{BD} cases had an increased PRS specifically for BD

In comparison to $FAM_{unaffected}$, the BD PRS and the BD-MDD disorder-specific PRS were significantly higher in FAM_{BD} (Fig. 2a, b). The *Shared* and the BD-SCZ PRS were increased at nominal significance.

Effects of assortative mating on BD PRS in family members

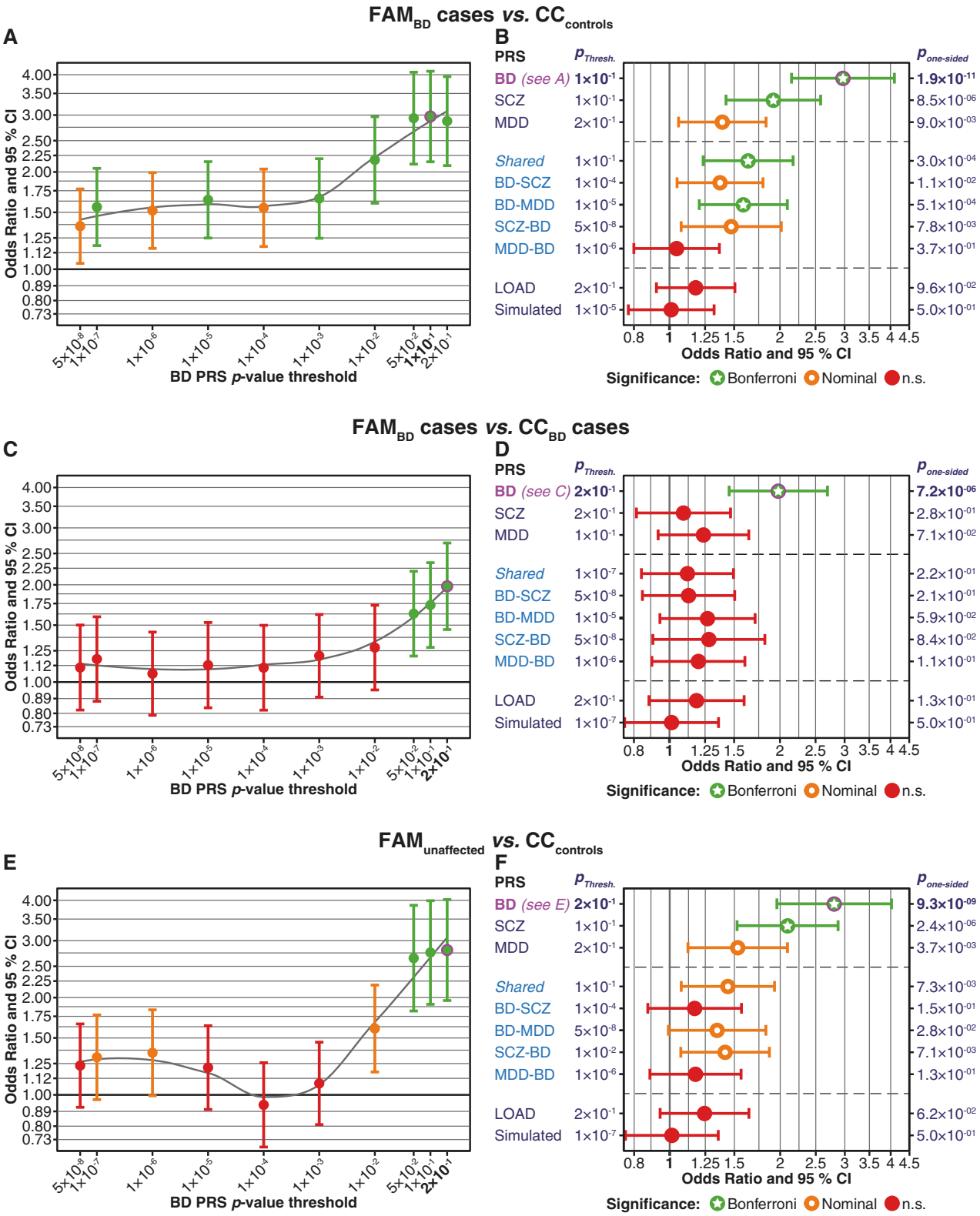
Eight of the 44 individuals who had married into the families had a diagnosis of MDD and none of BD (Table 1). While the unaffected married-in individuals had higher BD PRS than $CC_{controls}$ ($p = 6.5 \times 10^{-5}$), their BD PRS was not higher than the PRS of other $FAM_{unaffected}$ (Fig. 2c; Supplementary Fig. S6 and Supplementary Table S7). We also examined possible anticipation of BD in the families: neither did the BD PRS increase significantly over generations nor did the age at onset decrease over time (Fig. 2d; Supplementary Fig. S7 and Supplementary Table S8).

FAM_{MDD} cases had higher psychiatric PRS than $CC_{controls}$

In comparison to $CC_{controls}$, FAM_{MDD} cases had significantly higher BD and MDD PRS, increases of the *Shared* and SCZ PRS were nominally significant (Supplementary Figs. S8 and S9 and Supplementary Table S9). Both the BD and MDD PRS were increased at nominal significance when comparing FAM_{MDD} to $FAM_{unaffected}$ (Supplementary Fig. S10 and Supplementary Table S10). Notably, in both comparisons, FAM_{MDD} showed a nominal increase in SCZ-MDD PRS, but not in SCZ-BD PRS.

Discussion

Genome-wide association studies in large samples of unrelated patients and controls have unravelled the polygenic nature of BD, i.e., many common variants, each with a small effect size, contribute to BD. It has also been consistently shown that BD, MDD, and SCZ share many



risk-conferring variants. The aim of the present study was to investigate whether common variants also contribute to BD in families with a high density of the disorder and if so, whether these variants are specific to BD.

We found that, compared to CC$_{controls}$, unrelated subjects from the general population unscreened for BD, affected and unaffected ABiF family members had an elevated genetic risk for the tested psychiatric disorders, mainly for BD but

◀ **Fig. 1** Comparison of PRS between FAM and CC samples. Married-in family members were excluded from these analyses. The plots show one-sided *p*-values, following the hypothesis that family members have higher PRS than individuals from the CC samples. All PRS have been normalised using *Z*-score standardisation. **a**, **b** Comparison of FAM_{BD} cases to $CC_{controls}$. **a** FAM_{BD} cases had higher BD PRS across all ten p_{PRS} thresholds. The plot shows odds ratios (OR, *y*-axis, filled circles) and 95% confidence intervals (CI); p_{PRS} thresholds are shown on the *x*-axis. Results for each threshold are coloured by their degree of significance (one-sided *p*-values): red = not significant, orange = nominally significant, green = significant after Bonferroni correction for multiple testing ($\alpha = 0.05/60 = 0.00083$). The top-associated PRS ($p_{PRS} = 0.1$) is indicated in bold font and was marked by a magenta circle (also in **b**). **b** For ten different PRS, this plot shows association statistics for the top-associated p_{PRS} thresholds. The *x*-axis shows ORs. BD, SCZ, MDD: Standard PRS using the respective PGC GWAS summary statistics. *Shared*: Shared psychiatric PRS (SNPs with BD, MDD, SCZ $p < 0.05$, random-effects meta-analysis). BD-SCZ, BD-MDD: BD-specific GWIS PRS corrected for SCZ and MDD, respectively. SCZ-BD and MDD-BD: GWIS PRS for SCZ and MDD, each corrected for BD. LOAD: PRS for late-onset Alzheimer's disease. Simulated: Mean and CI of the 10,000 simulated PRS at the p_{PRS} with the lowest mean association *p*-value of all simulated PRS. The column to the left of the plot: p_{PRS} with the strongest association. Supplementary Fig. S2 shows plots for all p_{PRS}. Column to the right: $p_{one\text{-}sided}$ = one-sided *p*-value. For full association test statistics, see Supplementary Table S3. Bonferroni = significant after Bonferroni correction for multiple testing; nominal = nominally significant ($p < 0.05$); n.s. = not significant. **c**, **d** Comparison of FAM_{BD} cases and unrelated CC_{BD} cases. See Supplementary Fig. S3 and Table S4 for more detailed plots and full association test statistics. **e**, **f** Comparison of $FAM_{unaffected}$ and $CC_{controls}$. See Supplementary Fig. S4 and Table S5 for more detailed plots and full association test statistics

also for SCZ. FAM_{BD} cases were characterised by a particularly high load of BD-specific risk variants: The strongest association observed across all comparisons was the increase of the BD PRS in FAM_{BD} compared to $CC_{controls}$. In addition, the BD but not the SCZ and MDD PRS of FAM_{BD} were significantly higher than the PRS of unrelated CC_{BD} cases and unaffected family members. Together with the disorder-specific GWIS PRS, these results support the major contribution of BD-associated variants to the high density of the disorder in the investigated families.

An increased polygenic psychiatric risk has also been described in other studies of BD multiplex families [41–43]. However, the scope and results of these studies differed from the present study to some extent: Fullerton et al. [41] described an increased BD PRS in affected family members compared to unrelated controls and, when selecting families with a high polygenic BD risk load, also to unaffected family members. They constructed PRS only based on a small set of 32 SNPs from an older GWAS [10], and no other PRS were investigated. De Jong et al. [43] focused their analyses in a large Brazilian family with BD and MDD on assortative mating and anticipation and found BD and SCZ PRS to be increased at nominal significance in affected compared to unaffected members. In a large Swedish pedigree with mainly BD but also some SCZ cases, Szatkiewicz et al. [42] reported increased SCZ PRS in affected family members compared to family-level and population controls, as well as BD PRS increased at nominal significance in affected family members compared to family controls. However, no differences were observed between unaffected family members and population controls. Of note, none of these studies investigated differences in PRS between families and unrelated BD cases.

Compared to the CC_{BD} in our study, FAM_{BD} displayed, apart from an earlier age at onset, signs of a less severe clinical picture, i.e., less frequent impairment and less psychosis. This could be explained by the fact that CC_{BD} cases were almost all BD-I patients recruited from consecutive admissions to a hospital, while most of the FAM_{BD} cases were reached through other family members in the context of the study. Apart from this, the FAM_{BD} did not display any striking differences in clinical features compared to the CC_{BD}. Thus, we consider it likely that the increased PRS in the FAM_{BD} is linked to the familial aggregation and not to clinical characteristics.

It appears striking that none of the ABiF family members have been diagnosed with SCZ. However, this can most likely be attributed to ascertainment bias as the recruitment strategy focused on BD multiplex families. With respect to this lack of SCZ diagnoses in the ABiF families, it is of interest that the family members showed not only an increased BD PRS but also increased SCZ and *Shared* PRS compared to unrelated controls. This increase could be an indirect consequence of the genetic correlation between BD and SCZ [14, 16, 18–21]. Furthermore, affected family members also had higher *Shared* PRS than $CC_{controls}$. Of the psychiatric disorder GWAS data sets (i.e., SCZ, BD, and MDD) used in the present analysis, the SCZ GWAS both identified the largest amount of risk loci (108, 30, and 44, respectively) and the corresponding PRS explained the highest amount of case/control variance (7%, 4%, and 2% on a liability scale, respectively) [12–14]. Taking this and the genetic correlations between the disorders into account, the SCZ PRS might have included more cross-disorder signals with smaller effects than the PRS of BD and MDD. If family members had an increased *Shared* risk burden, this cross-disorder risk might have rendered them vulnerable to psychiatric disorders in general, with the high BD PRS then shaping the final BD diagnosis outcome. Of note, the analyses of FAM_{MDD} cases are discussed in the Supplementary Data.

Our study furthermore indicates that assortative mating may have contributed to the increased BD PRS in the ABiF families: in their study, de Jong et al. [43] found no increased PRS in married-in subjects, but an increase of polygenic risk and a decrease in age at onset over generations. We observed that individuals who married into the ABiF families had higher BD PRS than $CC_{controls}$, and

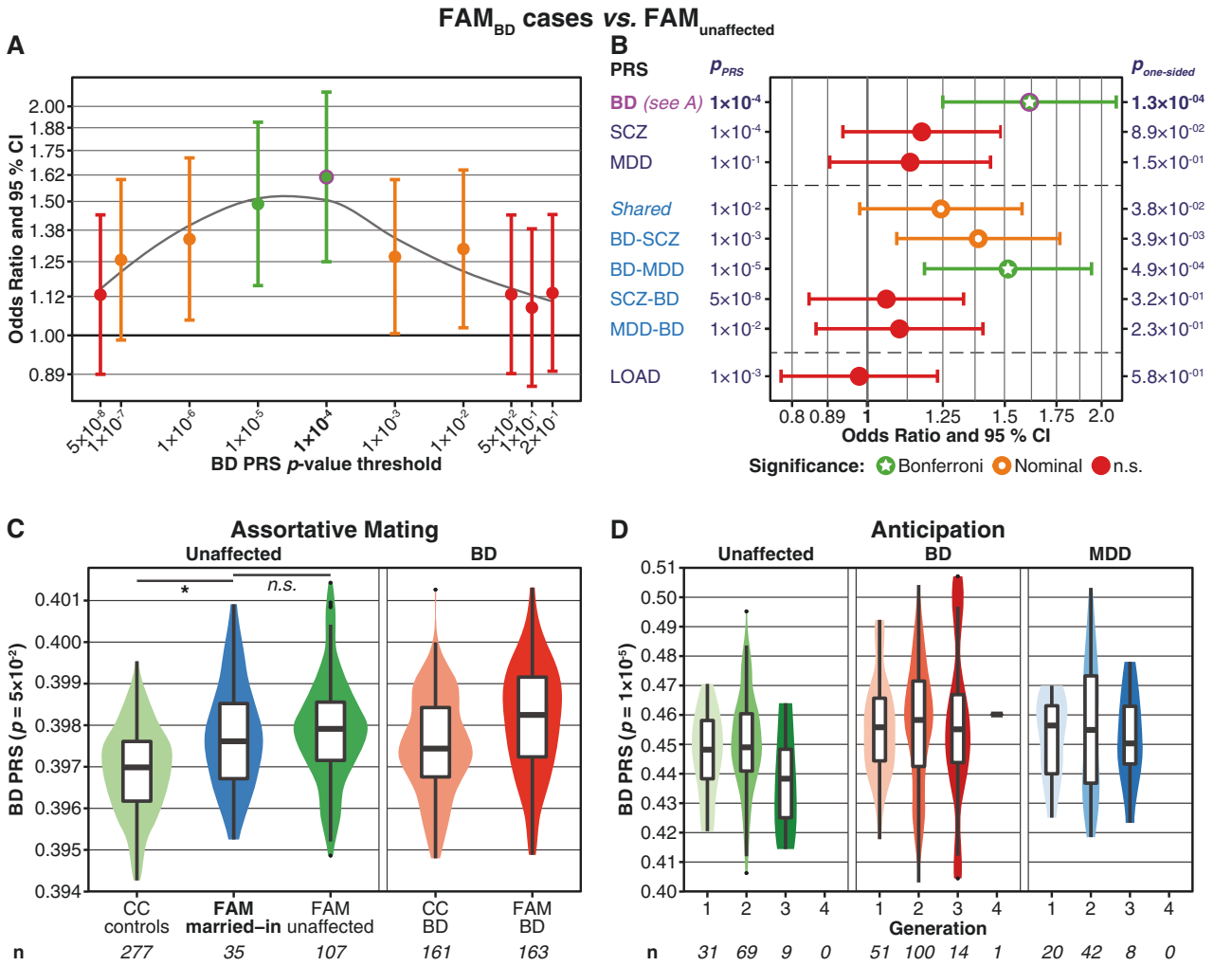


Fig. 2 **a**, **b** Comparison of PRS between FAM_{BD} cases and $FAM_{unaffected}$. The plots show one-sided *p*-values, following the hypothesis that BD cases have higher PRS than unaffected individuals. Further details of the plots are as described in the legend for Fig. 1. See Supplementary Fig. S5 and Table S6 for more detailed plots and full association test statistics. **c**, **d** Analyses of assortative mating (**c**) and anticipation (**d**). These plots were not adjusted for covariates; *n* = sample size. The *y*-axis shows the PRS values. **c:** Assortative mating. The plot shows violin- and boxplots of the BD PRS ($p_{PRS} = 0.05$), comparing unaffected, married-in individuals with no parent among the ABiF families to other FAM and CC subjects. At $p_{PRS} = 0.05$, married-in family members showed the highest BD PRS compared to $CC_{controls}$ ($p = 6.5 \times 10^{-5}$, Supplementary Fig. S6A and Table S7). The BD PRS of married-in individuals was not significantly higher than the PRS of $FAM_{unaffected}$ at any p_{PRS} ($p \geq 0.167$, Supplementary Fig. S6B and Table S7). Covariate used: sex. One-sided *p*-values were calculated, following the hypothesis that married-in individuals have higher PRS than other unaffected subjects. Note that, in the context of assortative mating, the boxplots of affected BD cases are displayed for reference only and have not been included in the analysis. **d** Anticipation: the BD PRS did not increase across generations. The plot shows violin- and boxplots of the BD PRS ($p_{PRS} = 1 \times 10^{-5}$) across different generations of the FAM sample for the three diagnosis groups. At $p_{PRS} = 1 \times 10^{-5}$, the association of the BD PRS with generation was strongest but not significant ($p = 0.45$; Supplementary Fig. S7A and Table S8). Married-in family members were excluded from this analysis. Covariates used: sex, age at the interview, diagnostic group. One-sided *p*-values were calculated, following the hypothesis that the PRS increase across generations

their BD risk load was similar to other $FAM_{unaffected}$. At the time of the interview, none of the married-in family members had a diagnosis of BD. Nevertheless, their increased BD PRS suggests that assortative mating may have occurred. Unaffected individuals with an above average BD PRS may display sub-threshold characteristics of BD, such as a broader range of emotions [44–46]. Consistent with the observation that married-in subjects did not have higher BD PRS than the other $FAM_{unaffected}$, no increase in BD PRS was found across generations. However, assortative mating may have contributed to the establishment and maintenance of a high genetic risk load for BD in these families. Furthermore, assortative mating may have already occurred in previous generations, for which no DNA was available. Of note, DNA was not available for all ABiF family members of the current

Table 2 The psychiatric disorder PRS can distinguish better between groups than simulated PRS

Group	Disorder PRS	Simulated PRS min. p_{PRS}	N simulated PRS with $p \leq p$ of disorder PRS	Prob. of success	95% CI
FAM_{BD} vs. $CC_{controls}$	BD	1×10^{-5}	0	**$<1 \times 10^{-4}$**	**0–0**
	SCZ	1×10^{-5}	0	**$<1 \times 10^{-4}$**	**0–0**
	MDD	1×10^{-5}	91	0.0091	0.007–0.011
	Shared	1×10^{-5}	2	**0.0002**	**0–0.001**
FAM_{BD} vs. CC_{BD}	BD	1×10^{-7}	0	**$<1 \times 10^{-4}$**	**0–0**
	SCZ	1×10^{-7}	2858	0.2858	0.277–0.295
	MDD	1×10^{-7}	744	0.0744	0.069–0.080
	Shared	1×10^{-7}	2229	0.2229	0.215–0.231
$FAM_{unaffected}$ vs. $CC_{controls}$	BD	1×10^{-7}	0	**$<1 \times 10^{-4}$**	**0–0**
	SCZ	1×10^{-7}	0	**$<1 \times 10^{-4}$**	**0–0**
	MDD	1×10^{-7}	37	0.0037	0.003–0.005
	Shared	1×10^{-7}	74	0.0074	0.006–0.009
FAM_{MDD} vs. $CC_{controls}$	BD	2×10^{-1}	0	**$<1 \times 10^{-4}$**	**0–0**
	SCZ	2×10^{-1}	409	0.0409	0.037–0.045
	MDD	2×10^{-1}	2	**0.0002**	**0–0.001**
	Shared	2×10^{-1}	125	0.0125	0.010–0.015

In binomial tests with 10,000 trials, the number of successes was the number of simulated PRS that showed the same or a stronger association than the disorder PRS (one-sided *p*-values). The 10,000 simulated PRS with ten *p*-value thresholds each were calculated by drawing random variants from across the genome, using the same number of variants as for the BD PRS at each threshold and random effect sizes from the pool of all available BD, SCZ, and MDD effects. For the present test, the p_{PRS} of the simulated PRS showing the lowest mean association *p*-value was chosen. *Prob.* = binomial test probability estimate of success; *CI* = confidence interval of the probability estimate, both calculated using the *R* package *binom* (method: *exact*). Significance threshold: 0.05/16 = 0.003125, comparisons surpassing this threshold are shown in bold font

generations, limiting the scope of the analysis of assortative mating.

Although both the FAM and CC samples were recruited in Spain [9], minor population differences may have influenced the present results. Even if such minor differences existed, it is unlikely that they caused the highly significant associations observed for the psychiatric PRS, given that the pairwise genetic relationship matrix was used as a random effect in the association analyses. Additionally, results from three further analyses support our assumption that systematic differences between the genotype data of FAM, $CC_{controls}$, and CC_{BD} samples did not distort our findings: First, we did not find significant differences between the cohorts in a population substructure analysis (see Supplementary Fig. S11 and Supplementary Methods). Second, PRS for LOAD were not significantly increased in family members in any analysis. Since LOAD shows no genetic correlation with BD, MDD, or SCZ [14, 47, 48], this result further supports the specificity of our analyses. Third, when a psychiatric disorder PRS was significantly increased in family members, this association was stronger than for simulated PRS. While these findings cannot entirely exclude the influence of unknown confounders on our results, we consider them as strong evidence that the high psychiatric PRS observed in family members compared to controls cannot be attributed to population or technical differences between the cohorts.

The lower a p_{PRS} threshold in the GWAS training data, the fewer SNPs were included in the calculation of the corresponding PRS. In most cases, significant differences between groups were not observed for these low p_{PRS} but the higher thresholds based on thousands of variants. This is commonly observed and in line with the polygenic nature of psychiatric disorders as complex disorders, with genome-wide significant SNPs only accounting for a small share of the polygenic signal. The training GWAS used for BD, SCZ, and MDD, the largest available for these phenotypes, differ in the number of included subjects, their statistical power, and the number of identified signals. Therefore, the derived PRS also differ in the number of SNPs used in the calculation of each threshold (see Supplementary Table S2). However, even though the BD GWAS was based on the smallest number of subjects and contained the lowest number of genome-wide-associated loci among the three GWAS, the BD PRS showed the strongest associations with BD case status or family membership, underlining the substantial contribution of BD risk variants to the development of BD in the ABiF families.

One limitation of the study is that the subjects of the unrelated control cohort were not systematically screened

for psychiatric disorders. The lifetime prevalence of unipolar depression in this cohort (up to 14.4% until the time of the interview) was in line with typically observed numbers [27], the prevalence of BD was not assessed. However, as BD has a lifetime prevalence of ~1%, we expect up to three BD cases among the 277 controls, a number we consider unlikely to have markedly influenced our results. Moreover, using controls unscreened for BD instead of "super-healthy" controls as a comparison to family members and unrelated BD cases represents a conservative approach and thereby strengthens the observed group differences in psychiatric PRS.

Similarly, around one third of the CC_{BD} reported a family history of BD. The CC_{BD} thus do not represent a sample of truly sporadic BD cases. However, the aim of our study was to investigate how members of multiplex BD families differ from typical BD cases regarding the polygenic contribution to their disorder. The observation that ABiF multiplex cases showed a higher polygenic psychiatric risk than CC_{BD}, despite part of the CC cases also reporting a family history for BD, thus rather strengthens the validity of our findings.

The present study generated substantial evidence that members of the ABiF families, including unaffected subjects, carried a higher risk burden of common genetic risk variants than an unrelated control sample mainly for the psychiatric disorders BD and SCZ and, at least the FAM_{MDD} cases, for MDD. In line with previous theoretical assumptions [49] and preliminary results from a pilot study in a single ABiF family [26], our results suggest that a high polygenic load of common risk variants is a major contributor to the increased risk for BD and MDD in families with a high density of BD. However, given the modest effect sizes of the PRS, they explained only a fraction of the phenotypic variance, and rare mutations such as copy number variants [50] or rare single-nucleotide variants likely also play an important role in each of the families. Sequencing studies carried out in multiplex families have suggested rare variants are involved in the aetiology of BD [51–53]. To date, however, it has proven difficult to identify replicable causal associations between rare variants and BD susceptibility. In a pilot study that analysed a single ABiF pedigree, we did not identify any rare causal variants for BD [26]. The analysis of rare variants in the remaining ABiF families using next-generation sequencing technologies is envisioned for the future, including integrative analyses in international consortia such as the Bipolar Sequencing Consortium [54]. Of note, the present analyses did not assess single families separately, but integrated PRS associations across all examined 33 ABiF families. Thus, the degree to which common and rare variants shaped the emergence of psychiatric disorders may vary between families.

Furthermore, PRS are commonly based on and applied to sets of unrelated individuals, and polygenic risk might act differently in the case of familial genetic background. Moreover, a broad range of environmental factors have been shown to influence the risk of psychiatric disorders and might act on top of the increased genetic risk in these families. However, environmental factors have not been systematically assessed in the present study. To further enhance our understanding regarding the aetiology of BD, integrated analyses of common and rare variants, as well as of environmental risk in the ABiF families are warranted in the future.

Acknowledgements The study was supported by the German Federal Ministry of Education and Research (BMBF), through the Integrated Network IntegraMent, under the auspices of the e:Med programme (grants 01ZX1314A to MMN and SC; 01ZX1314G to MR; 01ZX1614J to BMM) through grants 01EE1406C to MR and 01EE1409C to MR and SHW, and through ERA-NET NEURON, "SynSchiz—Linking synaptic dysfunction to disease mechanisms in schizophrenia—a multilevel investigation" (01EW1810 to MR) and BMBF grants 01EE1409C and 01EE1406C to MR and SHW; by the German Research Foundation (DFG grants FOR2107; RI908/11-2 to MR; NO246/10-2 to MMN; MU1315/8-2 to BMM; WI 3439/3-2 to SHW), by the Andalusian regional Health and Innovation Government (grants PI-0060-2017, RC-0006-2015 the Nicolas Monarde Programme for YDO and CTS-546) and by the Swiss National Science Foundation (SNSF grant 156791 to SC). MMN is a member of the DFG-funded cluster of excellence ImmunoSensation. The PGC has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320). We thank the research participants and employees of 23andMe, Inc. for their contribution to the MDD meta-analysis published in [14]. We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for the present analyses. See the Supplementary Data for extended Acknowledgements.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Merikangas KR, Jin R, He J-P, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. Arch Gen Psychiatry. 2011;68:241–51.
2. Craddock N, Sklar P. Genetics of bipolar disorder. Lancet. 2013;381:1654–62.
3. Edvardsen J, Torgersen S, Røysamb E, Lygren S, Skre I, Onstad S, et al. Heritability of bipolar spectrum disorders. Unity or heterogeneity? J Affect Disord. 2008;106:229–40.
4. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry. 2003;60:497–502.
5. Charney AW, Ruderfer DM, Stahl EA, Moran JL, Chambert K, Belliveau RA, et al. Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. Transl Psychiatry. 2017;7:e993.
6. Hou L, Bergen SE, Akula N, Song J, Hultman CM, Landén M, et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum Mol Genet. 2016;25:3383–94.
7. Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N, et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol Psychiatry. 2017;511:421.
8. Lee MTM, Chen C-H, Lee CS, Chen CC, Chong MY, Ouyang WC, et al. Genome-wide association study of bipolar I disorder in the Han Chinese population. Mol Psychiatry. 2011;16:548–56.
9. Mühleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat Commun. 2014;5:3339.
10. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 2011;43:977–83.
11. Cichon S, Mühleisen TW, Degenhardt FA, Mattheisen M, Miró X, Strohmaier J, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. Am J Hum Genet. 2011;88:372–81.
12. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019;51:793–803.
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
14. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50:668–81.
15. Song J, Bergen SE, Kuja-Halkola R, Larsson H, Landén M, Lichtenstein P. Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. Bipolar Disord. 2015;17:184–93.
16. Ruderfer DM, Fanous AH, Ripke S, McQuillin A, Amdur RL, Schizophrenia Working Group of Psychiatric Genomics Consortium, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol Psychiatry. 2014;19:1017–24.
17. Forstner AJ, Hecker J, Hofmann A, Maaser A, Reinbold CS, Mühleisen TW, et al. Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. PLoS ONE. 2017;12:e0171595.
18. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381:1371–9.
19. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984–94.
20. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu, Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell. 2018;173:1705–.e16.
21. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, et al. Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. bioRxiv. 2019:528117.
22. McClellan J, King M-C. Genomic analysis of mental illness: a changing landscape. JAMA. 2010;303:2523–4.
23. Kerner B. Toward a deeper understanding of the genetics of bipolar disorder. Front Psychiatry. 2015;6:105.
24. Guzman-Parra J, Rivas F, Strohmaier J, Forstner A, Streit F, Auburger G, et al. The Andalusian Bipolar Family (ABiF) Study: Protocol and sample description. Rev Psiquiatr Salud Ment. 2017. https://doi.org/10.1016/j.rpsm.2017.03.004.
25. Schumacher J, Kaneva R, Jamra RA, Díaz GO, Ohlraun S, Milanova V, et al. Genomewide scan and fine-mapping linkage studies in four European samples with bipolar affective disorder suggest a new susceptibility locus on chromosome 1p35-p36 and provides further evidence of loci on chromosome 4q31 and 6q24. Am J Hum Genet. 2005;77:1102–11.
26. Collins AL, Kim Y, Szatkiewicz JP, Bloom RJ, Hilliard CE, Quackenbush CR, et al. Identifying bipolar disorder susceptibility loci in a densely affected pedigree. Mol Psychiatry. 2013;18:1245–6.
27. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health. 2013;34:119–38.
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
29. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:e1000529.
30. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9.
31. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10:5–6.
32. R Core Team. R: A Language and Environment for Statistical Computing. R Core Team: Vienna, Austria; 2017.
33. Nieuwboer HA, Pool R, Dolan CV, Boomsma DI, Nivard MG. GWIS: genome-wide inferred statistics for functions of multiple phenotypes. Am J Hum Genet. 2016;99:917–27.
34. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
35. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452–8.
36. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet. 2016;98:653–66.

37. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014;11:407–9.
38. Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. PLoS ONE. 2013;8:e65395.
39. Canty A, Ripley BD. Boot: Bootstrap R (S-Plus) functions. comprehensive R archive network (CRAN), 2017. https://cran.r-project.org/web/packages/boot/.
40. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge University Press, Cambridge, 1997.
41. Fullerton JM, Koller DL, Edenberg HJ, Foroud T, Liu H, Glowinski AL, et al. Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young At-Risk Individuals. Am J Med Genet B Neuropsychiatr Genet. 2015;168:617–29.
42. Szatkiewicz J, Crowley JJ, Adolfsson AN, Åberg KA, Alaerts M, Genovese G, et al. The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. Transl Psychiatry. 2019;9:60.
43. de Jong S, Diniz MJA, Saloma A, Gadelha A, Santoro ML, Ota VK, et al. Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. Commun Biol. 2018;1:163.
44. Peyrot WJ, Robinson MR, Penninx BW, Wray NR. Exploring boundaries for the genetic consequences of assortative mating for psychiatric traits. JAMA Psychiatry. 2016;73:1189–95.
45. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of nonrandom mating within and across 11 major psychiatric disorders. JAMA Psychiatry. 2016;73:354–61.
46. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci. 2015;18:953–5.
47. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
48. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360:eaap8757.
49. Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. Eur J Hum Genet. 2010;18:1039–43.
50. Green EK, Rees E, Walters JTR, Smith K-G, Forty L, Grozeva D, et al. Copy number variation in bipolar disorder. Mol Psychiatry. 2016;21:89–93.
51. Goes FS, Pirooznia M, Parla JS, Kramer M, Ghiban E, Mavruk S, et al. Exome Sequencing of familial bipolar disorder. JAMA Psychiatry. 2016;73:590–7.
52. Cruceanu C, Schmouth J-F, Torres-Platas SG, Lopez JP, Ambalavanan A, Darcq E, et al. Rare susceptibility variants for bipolar disorder suggest a role for G protein-coupled receptors. Mol Psychiatry. 2017;3:577.
53. Ament SA, Szelinger S, Glusman G, Ashworth J, Hou L, Akula N, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. Proc Natl Acad Sci USA. 2015;112:3576–81.
54. Shinozaki G, Potash JB. New developments in the genetics of bipolar disorder. Curr Psychiatry Rep. 2014;16:493.

Affiliations

Till F. M. Andlauer[1,2] · Jose Guzman-Parra[3] · Fabian Streit[4] · Jana Strohmaier[4] · Maria José González[5] · Susana Gil Flores[6] · Francisco J. Cabaleiro Fabeiro[7] · Francisco del Río Noriega[8] · Fermin Perez Perez[5] · Jesus Haro González[9] · Guillermo Orozco Diaz[10] · Yolanda de Diego-Otero[3] · Berta Moreno-Küstner[11] · Georg Auburger[12] · Franziska Degenhardt[13] · Stefanie Heilmann-Heimbach[13] · Stefan Herms[13,14] · Per Hoffmann[13,14,15] · Josef Frank[4] · Jerome C. Foo[4] · Jens Treutlein[4] · Stephanie H. Witt[4] · Sven Cichon[14,15] · Manolis Kogevinas[16] · Bipolar Disorder Working Group of the Psychiatric Genomics Consortium · Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium · Fabio Rivas[3] · Fermín Mayoral[3] · Bertram Müller-Myhsok[1,17,18] · Andreas J. Forstner[13,14,19,20] · Markus M. Nöthen[13] · Marcella Rietschel[4]

1 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany

2 Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

3 Department of Mental Health, University Regional Hospital of Málaga, Institute of Biomedicine of Málaga (IBIMA), Málaga, Spain

4 Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

5 Department of Mental Health, Hospital of Puerto Real, Cádiz, Spain

6 Department of Mental Health, University Hospital of Reina Sofia, Córdoba, Spain

7 Department of Mental Health, Hospital of Jaén, Jaén, Spain

8 Department of Mental Health, Hospital of Jerez de la Frontera, Jerez de la Frontera, Spain

9 Department of Mental Health, Hospital Punta de Europa, Algeciras, Spain

10 Unidad de Gestión Clínica del Dispositivo de Cuidados Críticos y Urgencias del Distrito Sanitario Málaga—Coin-Guadalhorce, Málaga, Spain

11 Department of Personality, Assessment and Psychological Treatment, University of Malaga, Institute of Biomedicine of Málaga (IBIMA), Málaga, Spain

12 Department of Neurology, Goethe University Medical School, Frankfurt am Main, Germany

[13] Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany

[14] Department of Biomedicine, University of Basel, Basel, Switzerland

[15] Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany

[16] Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain

[17] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[18] Institute of Translational Medicine, University of Liverpool, Liverpool, UK

[19] Centre for Human Genetics, University of Marburg, Marburg, Germany

[20] Department of Psychiatry (UPK), University of Basel, Basel, Switzerland

Andlauer et al. *Translational Psychiatry* (2021)11:192
https://doi.org/10.1038/s41398-021-01317-7

Translational Psychiatry

ARTICLE Open Access

Genetic factors influencing a neurobiological substrate for psychiatric disorders

Till F. M. Andlauer[1,2], Thomas W. Mühleisen[3,4,5], Felix Hoffstaedter[3], Alexander Teumer[6], Katharina Wittfeld[7,8], Anja Teuber[9,10], Céline S. Reinbold[5,11], Dominik Grotegerd[12], Robin Bülow[13], Svenja Caspers[3,14], Udo Dannlowski[12], Stefan Herms[5,15], Per Hoffmann[3,5,15,16], Tilo Kircher[17,18,19], Heike Minnerup[9], Susanne Moebus[20], Igor Nenadić[17,18,19], Henning Teismann[9], Uwe Völker[21], Amit Etkin[22,23,24], Klaus Berger[9], Hans J. Grabe[25], Markus M. Nöthen[15,16], Katrin Amunts[3,4,26], Simon B. Eickhoff[3,27], Philipp G. Sämann[1], Bertram Müller-Myhsok[1,28] and Sven Cichon[3,5,29]

Abstract

A retrospective meta-analysis of magnetic resonance imaging voxel-based morphometry studies proposed that reduced gray matter volumes in the dorsal anterior cingulate and the left and right anterior insular cortex—areas that constitute hub nodes of the salience network—represent a common substrate for major psychiatric disorders. Here, we investigated the hypothesis that the common substrate serves as an intermediate phenotype to detect genetic risk variants relevant for psychiatric disease. To this end, after a data reduction step, we conducted genome-wide association studies of a combined common substrate measure in four population-based cohorts ($n = 2271$), followed by meta-analysis and replication in a fifth cohort ($n = 865$). After correction for covariates, the heritability of the common substrate was estimated at 0.50 (standard error 0.18). The top single-nucleotide polymorphism (SNP) rs17076061 was associated with the common substrate at genome-wide significance and replicated, explaining 1.2% of the common substrate variance. This SNP mapped to a locus on chromosome 5q35.2 harboring genes involved in neuronal development and regeneration. In follow-up analyses, rs17076061 was not robustly associated with psychiatric disease, and no overlap was found between the broader genetic architecture of the common substrate and genetic risk for major depressive disorder, bipolar disorder, or schizophrenia. In conclusion, our study identified that common genetic variation indeed influences the common substrate, but that these variants do not directly translate to increased disease risk. Future studies should investigate gene-by-environment interactions and employ functional imaging to understand how salience network structure translates to psychiatric disorder risk.

Introduction

Numerous studies have identified regional differences in the brain structure of psychiatric patients and described both transdiagnostic and disorder-specific processes of gray matter (GM) reduction in patients[1–8]. One of these reports was the large retrospective meta-analysis of 193 studies by Goodkind et al. that compared 7381 psychiatric patients from six diagnostic groups (schizophrenia, bipolar disorder (BD), major depressive disorder (MDD), addiction, obsessive-compulsive disorder, and anxiety) with 8511 psychiatrically healthy controls using voxel-based morphometry (VBM) from structural magnetic resonance imaging (MRI)[1]. Across all diagnoses, they found that GM volumes were lower in the left and right anterior insular cortices (AIC) and the dorsal anterior cingulate cortex (dACC). Subsequentially, they performed structural and functional connectivity analyses and confirmed that these three regions were tightly connected and represent hub nodes of the salience network:[1,9,10] This network serves stimulus selection,

Correspondence: Sven Cichon (s.cichon@fz-juelich.de)
[1]Max Planck Institute of Psychiatry, Munich, Germany
[2]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany
Full list of author information is available at the end of the article
These authors contributed equally: Till F. M. Andlauer, Thomas W. Mühleisen, Bertram Müller-Myhsok, Sven Cichon

controls the focus of attention, and is involved in the selection of goal-directed behavior and in saliency detection of exogenous or internal cues[9–11]. Independent studies indicate that functional differences in salience processing in these brain regions are associated with several neuropsychiatric disorders and their progression[11]. Eventually, Goodkind et al. hypothesized that lower GM of this network represents a common neurobiological substrate for psychiatric disorders[1].

However, the etiology of the common substrate reductions has not been investigated so far and remains unclear. One possible explanation involves the loss of GM at disease manifestation and during the further course of disease, implying a regionally specific vulnerability toward a degenerative process—similar to known neurodegenerative disease entities[12,13]. An alternative explanation implies that reduced GM exists before disease onset, shaped by genetic or early environmental influences such as childhood adversity:[14] Here, premorbid structural abnormalities of the salience network could increase a subject's vulnerability to psychiatric disease. More recently, structural salience network integrity was reported to mediate between polygenic risk for schizophrenia and auditory hallucinations[15]. A third explanation involves brain-aging processes that occur in a network-dependent way and often with a strong non-linear component:[16–18] Here, accelerated aging could increase the disease risk over the lifespan by genetic or environmental factors. All three explanation models might apply in parallel and lead to combined effects at the morphological level.

Many studies have analyzed genetic risk factors for psychiatric disorders such as schizophrenia, BD, and MDD[19]. These disorders show substantial heritability[20] and are genetically correlated with each other[21,22]. Genome-wide association studies (GWAS) identified single-nucleotide polymorphisms (SNPs) contributing risk for several psychiatric disorders, suggesting pleiotropy and partially overlapping etiologies[22,23]. Imaging genomics is a growing discipline that exploits imaging-based measures to explore the genetic basis of brain organization[24]. The clinical value of this concept to detect risk variants for psychiatric disease, however, depends on a detectable correlation between the intermediate phenotype and the clinical level. Following this line of thought, the common substrate suggested by Goodkind et al. is a promising intermediate phenotype, particularly due to its transdiagnostic effects.

The present study aimed to identify genetic variants influencing the substrate in the general population. As a conceptual decision, patient cohorts were not included in our genetic analyses to avoid any interference with secondary disease effects on the common substrate, such as treatment effects or other disease-related epiphenomena. Our imaging analyses involved a prospective, harmonized VBM preprocessing protocol applied to high-resolution structural MRI data of five population-based cohorts. To account for the network character of the three common substrate regions, we combined them into a single marker using principal component analysis (PCA). We analyzed the first principal *component of the common substrate* (CCS) of the population-based cohorts through GWAS, followed by meta-analysis. As our main result, we identified a novel genetic locus significantly associated with the CCS. In a series of secondary analyses, we characterized the genetic relationship between the CCS and risk for psychiatric disorders and investigated a potentially modulating role of age.

Methods and materials

Sample description

For the GWAS, 3136 individuals from five population-based cohorts were pooled. Four cohorts were used in the discovery (1000BRAINS[25], $n = 539$; CONNECT100[26], $n = 93$; BiDirect[27], $n = 589$; SHIP-2[28], $n = 1050$; total $n = 2271$) and the second-largest cohort available was used in the replication stage (SHIP-Trend[28], $n = 865$). For follow-up analyses, three psychiatric patient/control cohorts with 1978 patients and 1375 controls were used, BiDirect ($n = 582$ MDD patients; $n = 311$ healthy controls[29]), Max Planck Institute of Psychiatry (MPIP) ($n = 385$ MDD patients; $n = 197$ healthy controls[30,31]), and FOR2107 ($n = 769$ MDD, $n = 127$ BD, $n = 72$ schizophrenia, and $n = 43$ schizoaffective patients; $n = 867$ healthy controls[32,33]). The BiDirect cohort is a prospective observational study[27]. Probands were recruited in the area of Münster and underwent a structured clinical interview for DSM-IV axis I disorders and all MDD patients received treatment for acute depression[27]. The MPIP cohort represents subsamples of the Munich Antidepressant Response Signature study, an observational study on psychiatric in-patients treated for MDD[30], and the recurrent unipolar depression study, a cross-sectional case/control imaging genetics study[31] (see[5,6] for diagnostic instruments). FOR2107 is an ongoing multicenter study recruiting from the areas of Marburg and Münster in Germany[32]. All subjects underwent a structured clinical interview for DSM-IV axis I disorders, administered by trained clinical raters. Basic demographic characteristics of the cohorts can be found in Supplementary Tables S1 and S2. The studies were approved by the local ethics committees; all participants provided written informed consent.

VBM preprocessing and extraction of regional and total GM volumes

VBM-like preprocessing with MATLAB-based SPM (version 8, https://www.fil.ion.ucl.ac.uk/spm/software/

spm8/) and the VBM8 toolbox (version r445, http://dbm.neuro.uni-jena.de/vbm8/) were used to process all T1-weighted images ($n = 3136$ for the GWAS and $n = 3361$ for the patient/control analyses). Processing was performed locally by the participating sites and comprised the following steps: (i) spatial registration to a reference brain in MNI152 space, (ii) segmentation of T1-weighted images into GM, white matter, and cerebrospinal fluid by a three-step algorithm implemented in the VBM8 toolbox, (iii) bias correction of intensity non-uniformities, (iv) application of the diffeomorphic normalization algorithm DARTEL for iterative linear and non-linear spatial normalization of the GM and white matter maps to MNI space (IXI555 template)[34], (v) non-linear-only Jacobian modulation to correct for linear global scaling effects while preserving local GM volumes. The quality of processed GM segments in MNI space was assessed using the "Check sample homogeneity using covariance" function in VBM8. Three spatially disjunct regional GM volumes, based on binarized versions of the joint result areas from the study by Goodkind et al.[1], and total GM volume were extracted.

Extracted GM volumes were, separately for each cohort, corrected for age, age^2, and sex in multiple linear regression models. To allow for a valid interpretation of the individual coefficients, we conducted a Gram–Schmidt orthonormalization of age (first term) and age^2 (second term) in R v.3.5.2, using the function QR of the package matlib (see the Supplementary Material). Handedness was used as an additional covariate for 1000BRAINS, CONNECT100, and BiDirect, coil type for the MPIP sample, as well as body coil type and site for FOR2107. Residuals of these regional volume regression models were combined using PCA to create a single measure, which we named the CCS (Fig. 1).

Genotyping, quality control, and imputation

DNA extraction and genome-wide genotyping were conducted as described before[28,31,35–37]. Genotyping was carried out on different Illumina and Affymetrix microarrays (see the Supplementary Methods and Supplementary Table S3). Quality control and imputation were conducted separately for each genotyping batch, using the same protocols, in PLINK, R, and XWAS[38,39]. Genotype data were imputed to the 1000 Genomes phase 1 reference panel using SHAPEIT and IMPUTE2[40,41], as described in the Supplementary Methods and previously[42]. The population substructure of all five GWAS cohorts is shown in Supplementary Fig. S1.

Heritability estimation and GWAS

The SNP-based heritability of the CCS was estimated using genome-wide complex trait analysis on a combined sample of the imputed data from all five cohorts[43] (see the Supplementary Methods). GWAS was conducted separately per cohort using PLINK with ancestry components as covariates. Variants on the X chromosome were analyzed separately by sex, followed by p value-based meta-analysis to allow for different effect sizes per sex. A two-stage design was implemented for the GWAS, using four cohorts as the discovery sample and SHIP-Trend as an independent replication sample. The cohorts were combined with fixed-effects meta-analysis using METAL[44]. There was no indication for genomic inflation of the GWAS test statistics in the single cohorts or the meta-analysis ($\lambda_{1000} = 1.01$, see Supplementary Table S4 and Supplementary Fig. S2).

Linkage disequilibrium (LD) was analyzed using the European 1000 Genomes CEU population in LDmatrix[45]. The two SNPs that showed the most robust genome-wide support ($p < 5 \times 10^{-8}$) for an association in the discovery stage and were partially independent of each other (LD $r^2 < 0.5$ with more strongly associated variants) were carried forward to the replication stage. Here, a one-sided p value $< \alpha = 0.05/2$ (correcting for two LD-independent variants) was considered as a successful replication. See the Supplementary Methods for a detailed description of heritability and GWAS methods.

Gene-set analyses

Gene-set analyses were conducted on the meta-analysis of the discovery- and replication-stage GWAS, using 674 REACTOME gene sets containing 10–200 genes curated from MsigDB 6.2[46]. Only SNPs within gene boundaries were mapped to RefSeq genes (0 bp window). Analyses were conducted in MAGMA v1.07 using both mean and top SNP gene models[47] and in MAGENTA v2 using a top SNP approach[48]. Here, false discovery rates were calculated to correct for multiple testing.

Comparison to published GWAS of psychiatric disorders and polygenic score analyses

For genome-wide comparisons between our GWAS meta-analysis and published GWAS of psychiatric disorders, summary statistics from the following Psychiatric Genomics Consortium (PGC) GWAS were used: cross-disorder 2019[22], BD 2019[49], MDD 2018 (with 23andMe)[50], and schizophrenia 2014[51]. For additional comparisons, the following GWAS were used: IFGC behavioral frontotemporal dementia (bvFTD) 2014[52], longevity 85/90 2014[53], and three different GWAS from 2017 on epigenetic accelerated aging (EAA):[54] accelerated aging in all examined brain regions, accelerated aging in prefrontal cortex, and neuronal proportion in the prefrontal cortex.

To further characterize the relationship between the CCS and risk for psychiatric disorders, we ran four

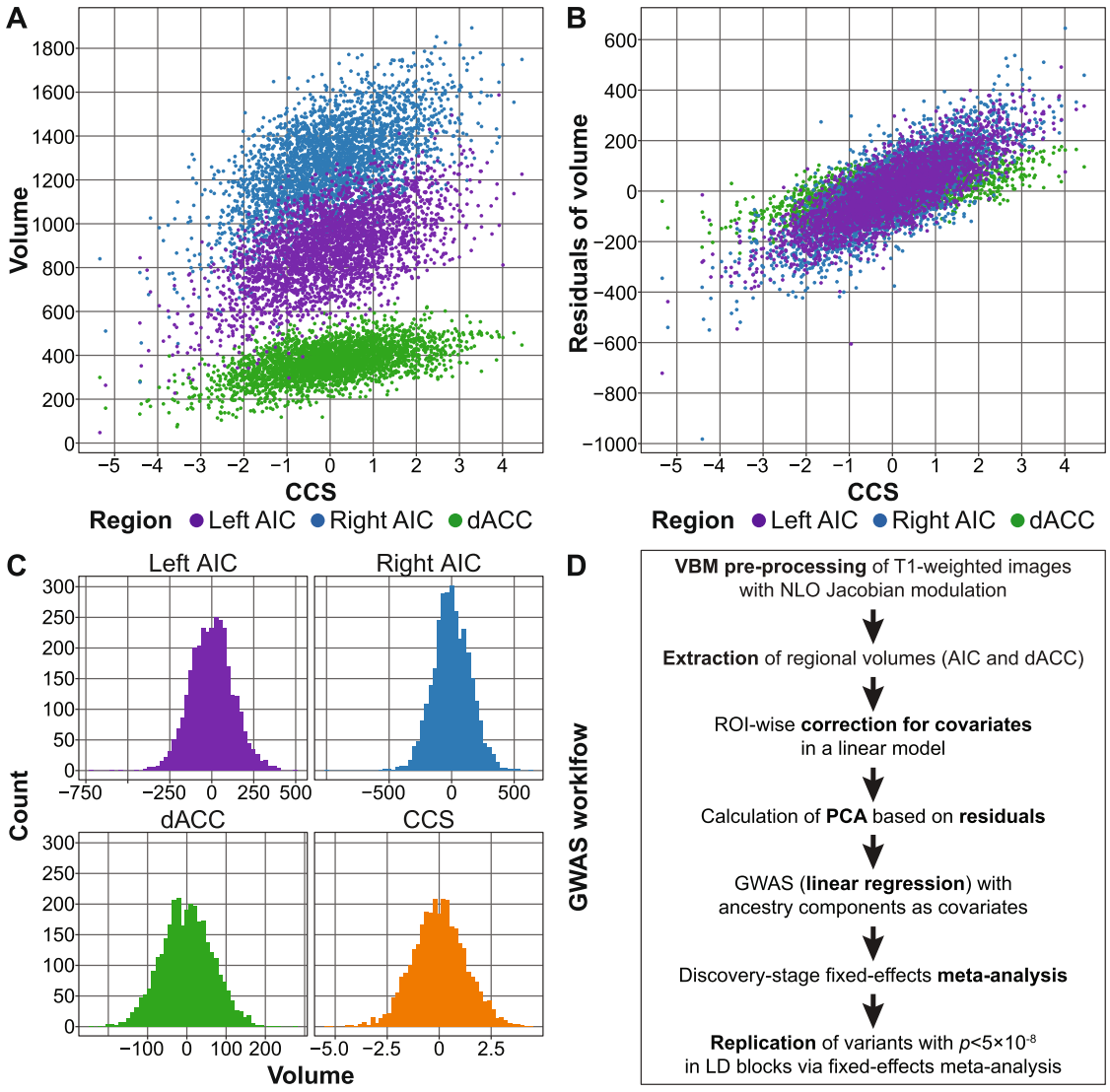


Fig. 1 Generation of the component of the common neurobiological substrate (CCS) and genome-wide association study (GWAS) analysis workflow. A, **B** Comparison between the CCS and the three individual volumes (**A**) and the residuals of the three volumes after correction for covariates (**B**). AIC anterior insula cortex, dACC dorsal anterior cingulate cortex. **C** Histograms of the three extracted volumes and the CCS. Note that **A**–**C** show combined data from all five GWAS cohorts. Correlations were left and right AIC: $r = 0.65$, left AIC and dACC: $r = 0.52$, right AIC and dACC: $r = 0.46$. **D** GWAS analysis workflow. All measures were extracted using NLO-based Jacobian modulation. All GM volumes were corrected for age, age^2, and sex as covariates; handedness was used as an additional covariate for the three samples 1000BRAINS, CONNECT100, and BiDirect. PCA: principal component analysis, LD: linkage disequilibrium.

analyses using GWAS summary statistics from published PGC studies, following a published, well-acknowledged workflow[55]. Polygenic scores (PGSs) were calculated and analyzed in R using imputed genetic data[56,57]. Here, we used the PGC GWAS as training and our population-based GWAS cohorts as test data. Furthermore, we also calculated PGS using the CCS GWAS summary statistics as training and the patient/control cohorts as test data. We ran LD score regression (LDSC) comparing the genetic correlation of published GWAS to the CCS GWAS summary statistics with standard settings[58,59]. We analyzed whether the order of SNPs ranked by their association strength was random between studies using rank–rank hypergeometric overlap (RRHO) tests[60]. For this analysis, variants were LD-pruned in the 1000 Genomes phase 3 EUR subset[61]. Binomial sign tests were conducted on LD-clumped variants in R (binom.test) to analyze whether SNPs associated with the CCS at either $p < 0.05$ or $p < 1 \times 10^{-5}$ showed the opposite direction of effects in other GWAS more often than expected by chance. For additional details on these analyses, see the Supplementary Methods.

Secondary analyses of age-interaction effects

We explored the possibility that the original VBM studies entering the meta-analysis of Goodkind et al.[1] picked up age-by-diagnosis effects by analyzing patient/control cohorts and by verifying that our main genetic association was not age-dependent. We performed secondary analyses that probed (a) the possibility of "accelerated aging" of the CCS phenotype in psychiatric disorders and (b) the possibility of heterogeneity of the SNP effect across different age ranges.

Results

Combination of the three brain regions

To analyze a combined measure of the published common neurobiological substrate for psychiatric disorders[1], we combined the volumes of the left AIC, right AIC, and dACC by PCA. The first principal component, which we refer to as the CCS, explained 66.5% of the phenotypic variance of the three volumes and 55.4% after correction of the volumes for covariates (Fig. 1 and Supplementary Methods).

Heritability of the CCS

After correction for covariates, the CCS showed a SNP heritability estimate of $h^2_g = 0.50$ (standard error (SE) = 0.18; p value = 0.0033).

GWAS of the CCS

In the discovery-stage GWAS (Supplementary Fig. S2A and Supplementary Table S4), 12 SNPs on chromosome *5q35.2* showed genome-wide significant associations with the CCS (significance threshold $p < 5 \times 10^{-8}$; Fig. 2A and Supplementary Table S5). Most of these variants were highly correlated with each other (Supplementary Table S6). The two partially LD-independent SNPs (pairwise LD $r^2 = 0.267$ in CEU samples) with the most robust support for an association were analyzed further (Fig. 2B). Of these, the minor allele T of the SNP rs17076061 (frequency in our GWAS cohorts: 0.36, Fig. 2C) was significantly associated in the replication cohort in the same direction (discovery: $\beta = -0.22$ standard deviations (SD) (SE = 0.04), $p = 1.51 \times 10^{-8}$; replication: $\beta = -0.15$ (SE = 0.07), one-sided $p = 9.91 \times 10^{-3}$, Supplementary Fig. S3) and was also the top-associated variant in the genome-wide meta-analysis of discovery and replication samples ($\beta = -0.21$ (SE = 0.03), $p = 1.46 \times 10^{-9}$; Fig. 2D, Supplementary Table S5, Supplementary Figs. S2B, S4, and S5). SNP rs17076061-T was associated with the CCS at genome-wide significance but not with the three single region volumes or the whole-brain GM volume (Table 1). After z-score transformation to allow effect size comparisons, the effect size was larger for the CCS (−0.159 SD) than for the total GM (−0.099 SD).

Gene-set analyses

In two separate gene-set analyses using GWAS meta-analysis results, four pathways were significantly associated with the CCS. The top-associated pathway in both analyses (MAGMA: adjusted $p = 2.2 \times 10^{-3}$; MAGENTA: false discovery rate $q = 2.4 \times 10^{-3}$) was "SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion" (https://www.reactome.org/content/detail/R-HSA-399955). Please see Supplementary Tables S7 and S8 for the full results of these analyses.

Comparison of the top GWAS SNP and the genetic architecture of the CCS with genetic risk for disease

To investigate whether rs17076061 is associated with risk for common psychiatric disorders, we looked up the SNP in published results from large GWAS of psychiatric disorders by the PGC (cross-disorder[22], BD[49], MDD[50], and schizophrenia[51]). Here, the cross-disorder GWAS showed the strongest effect, albeit not significant after correction for multiple testing (OR = 1.035, unadjusted one-sided $p = 0.048$; Supplementary Table S9). Next, we conducted genome-wide comparisons: using LDSC, we found no significant genetic correlation between the CCS GWAS and the four psychiatric GWAS (Table 2 and Supplementary Table S10). Furthermore, RRHO tests showed no significant overlap of SNPs ranked by their association strength (Table 2, Supplementary Table S11, and Supplementary Fig. S6). In binomial sign tests, CCS-associated variants did not show the opposite effect direction in the psychiatric disorder GWAS more often than expected by chance (Table 2 and Supplementary Table S12).

Analysis of polygenic scores

Next, we calculated PGSs based on the four PGC GWAS (psychiatric cross-disorder, MDD, BD, schizophrenia) as training data and analyzed associations of these disease-associated PGSs with the CCS in our population cohorts. None of the PGSs were associated with the CCS after correction for multiple testing (Table 2, Supplementary Table S13, and Supplementary Fig. S7).

Last, we inverted the direction of the approach and built a PGS based on our CCS GWAS as training data, using ten different p value thresholds, and compared it between patients and controls from four clinical diagnoses (MDD, BD, schizoaffective disorder, and schizophrenia) as available from three patient/control cohorts (BiDirect, MPIP, FOR2107). We expected the CCS PGS to be lower in psychiatric patients. No consistent results were observed regarding the expected direction of the patient/control comparisons and a specific threshold, and no single effect proved robust to multiple testing correction (Supplementary Table S14).

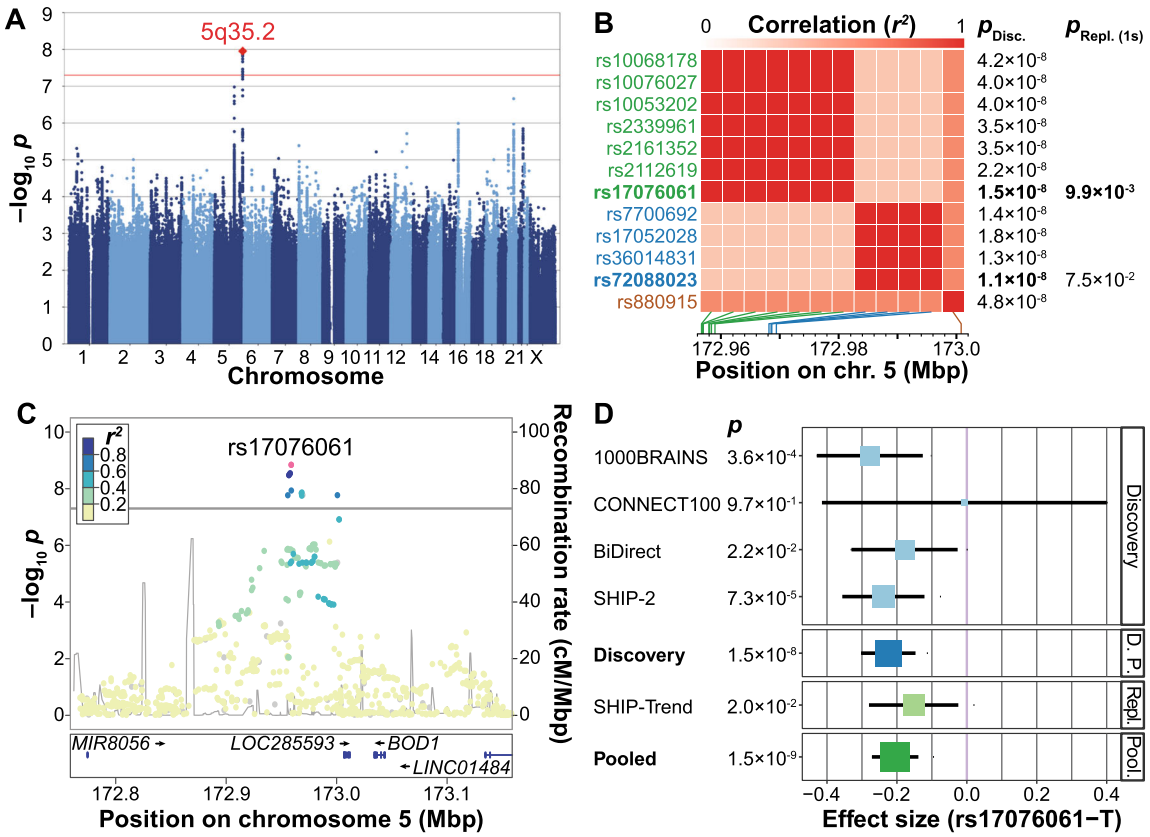


Fig. 2 Presentation of the genome-wide association study (GWAS) results. A Manhattan plot showing the strength of evidence for an association (*p* value) in the discovery stage component of the common neurobiological substrate (CCS) GWAS. Each variant is shown as a dot, with alternating shades of blue according to chromosome; the top-associated locus *5q32.2* is labeled with a red diamond. The red line marks the genome-wide significance level. **B** Matrix of the pairwise linkage disequilibrium (LD) pattern (1000 Genomes population CEU) between the 12 variants that reached genome-wide significance in the discovery GWAS. The two variants rs17076061 and rs72088023 ($r^2 = 0.267$) showed the strongest support for an association in their respective LD blocks and were analyzed in the replication stage. All other variants had pairwise LD > 0.5 with either of these two variants, their association strengths are provided for comparison only. $P_{Disc.}$ discovery stage GWAS *p* value, $p_{Repl.(1s)}$ one-sided *p* value in the replication cohort, Mbp mega base pair. **C** Regional association plot of the top-associated locus after pooled analysis of the discovery stage GWAS and the replication sample. The dot color indicates LD with the lead variant (rs17076061; pink). Gray dots represent signals with missing LD r^2 values. cM: centimorgan. **D** Forest plot of the pooled analysis of the replicated variant rs17076061 in discovery and replication cohorts. D. P.: pooled analysis of discovery stage cohorts, Repl.: replication, Pool.: pooled analysis of the discovery GWAS and the replication cohort SHIP-Trend.

Analyses of age-dependent effects

In an imaging meta-analysis of our three MDD/control cohorts (BiDirect, MPIP, FOR2107), we confirmed that the CCS was reduced in MDD patients compared to controls ($p = 1.3 \times 10^{-7}$; Fig. 3A and Supplementary Table S15). In the transdiagnostic FOR2107 cohort, the median CCS showed a stepwise decrease along the affective-psychosis axis (controls: median = 0.18; MDD: median = −0.010, comparison to controls: $p = 3.9 \times 10^{-3}$; BD: median = −0.35, $p = 2.8 \times 10^{-5}$; schizoaffective disorder: median = −1.13, $p = 2.6 \times 10^{-8}$; schizophrenia: median = −0.58, $p = 6.6 \times 10^{-10}$; combined analysis of all four diagnostic groups in FOR2107: $p = 1.5 \times 10^{-7}$; Fig. 3A, Supplementary Table S15, and Supplementary Fig. S8). This finding strongly affirmed the results of Goodkind et al.[1].

In these analyses, we noticed a possible influence of age on the association between the patient/control status and the CCS. When adding a linear and quadratic age-interaction term to the MDD regression models, the linear interaction term was not significant ($p = 0.72$). However, the age^2-by-diagnosis interaction term was significant ($p = 0.014$), pointing to a possible non-linear age dependency in MDD. No such effect was detected in the other diagnostic groups (Fig. 3B and Supplementary Table S15). To explore non-linear age dependencies in a complementary approach, we stratified all patient/control cohorts into five non-overlapping age groups (Fig. 3C). Heterogeneity in a meta-analysis of the CCS associations stratified by age would have indicated strong non-linear effects of age on the CCS. However, we detected no

Table 1 Association results from the genome-wide meta-analysis of discovery and replication samples in different gray matter (GM) regions.

rs17076061	Effect size	SE	*p* value
CCS	−0.159	0.026	1.41×10^{-9}
Left AIC	−0.142	0.026	7.00×10^{-8}
Right AIC	−0.124	0.026	2.63×10^{-6}
dACC	−0.083	0.026	1.77×10^{-3}
Total gray matter	−0.099	0.026	1.85×10^{-4}

For this comparison, all measures were centered and scaled using *z*-score transformation before the analysis to make the effect sizes of the different measures comparable. The unit of the effect sizes is thus standard deviation (SD). Accordingly, the CCS coefficients shown here differ from the ones presented in Fig. 2 and Supplementary Table S5. The effect size refers to the minor allele T. All measures were extracted using non-linear only (NLO)-based Jacobian modulation.
AIC anterior insula cortex, *dACC* dorsal anterior cingulate cortex, *SE* standard error.

significant heterogeneity between the age groups ($Q = 3.21$, $p = 0.52$; Fig. 3C and Supplementary Table S16).

When adding interaction terms to the model, neither the age-by-SNP ($p = 0.48$) nor the age^2-by-SNP ($p = 0.50$) interaction became significant in the meta-analysis of the five GWAS population cohorts, while the main SNP effect remained stable (Supplementary Table S17). Similarly, when stratifying the analysis by age groups, the SNP main effect size varied, yet without significant heterogeneity ($Q = 2.25$, $p = 0.69$; Fig. 3D and Supplementary Table S17).

To investigate whether our specific implementation of the global brain size correction influenced the association results, we switched from non-linear only Jacobian modulation of the GM probability maps to full Jacobian modulation, with the total intracranial volume entered as an explicit volumetric covariate. Our association results remained stable, independent of the correction method used (Supplementary Methods and Supplementary Table S15).

Comparison of the genetic architecture of the CCS with the genetics of aging traits

To further explore whether genetic variants associated with the CCS might influence aging-related processes, we compared our CCS GWAS results with GWAS for EAA[54] and longevity[53]. The common substrate regions represent the salience network, which is specifically prone to neurodegeneration in bvFTD[12,62], a subtype of frontotemporal dementia with severe executive disturbances and personality changes. Therefore, we also analyzed a possible overlap with GWAS results for bvFTD[52]. SNP rs17076061 showed no significant association in any of these GWAS (Supplementary Table S9). Single findings for longevity and EAA were nominally significant in PGS analyses and sign tests. However, overall, no significant genetic overlap with any of these GWAS was found with LDSC (Supplementary Table S10), RRHO tests

Table 2 Comparisons of the component of the common neurobiological substrate (CCS) and the CCS genetic architecture with psychiatric disorders.

LD score regression (LDSC)

GWAS comparison	r_g	*p* value
Psychiatric cross-disorder	0.0005	0.99
Bipolar disorder	0.17	0.08
Major depression	−0.03	0.75
Schizophrenia	0.08	0.38

Rank–rank hypergeometric overlap (RRHO)

GWAS comparison	Overlap	*p* value
Psychiatric cross-disorder	0.29	0.53
Bipolar disorder	0.21	0.06
Major depression	0.04	0.18
Schizophrenia	0.02	0.15

Binomial sign tests ($p < 0.05$)

GWAS comparison	Probability	*p* value
Psychiatric cross-disorder	0.50	0.77
Bipolar disorder	0.50	0.67
Major depression	0.50	0.82
Schizophrenia	0.50	0.64

Binomial sign tests ($p < 1 \times 10^{-5}$)

GWAS comparison	Probability	*p* value
Psychiatric cross-disorder	0.33	0.93
Bipolar disorder	0.54	0.50
Major depression	0.33	0.93
Schizophrenia	0.54	0.50

Polygenic scores (PGS)

Training GWAS	Effect size	*p* value	p_T
Psychiatric cross-disorder	−0.78	0.30	5×10^{-8}
Bipolar disorder	−0.64	0.05	1×10^{-7}
Major depression	−5.01	0.31	1×10^{-2}
Schizophrenia	−0.58	0.24	1×10^{-7}

Details on the four training genome-wide association studies (GWAS) datasets are provided in the Methods section. LDSC: linkage disequilibrium score regression using genome-wide summary statistics (Supplementary Table S10); r_g genetic correlation. RRHO: rank–rank hypergeometric overlap test showing the relative overlap of genome-wide summary statistics (Supplementary Table S11). *Sign tests* one-sided binomial sign tests for CCS GWAS *p* value thresholds $p < 0.05$ and $p < 1 \times 10^{-5}$ and the corresponding probability of success (Supplementary Table S12). *PGS* association of polygenic scores with the CCS; *pT* training GWAS data *p* value threshold; *effect size* linear regression effect size at the *pT* showing the strongest support for an association (see Supplementary Table S13 for results of all ten thresholds); *p* value: one-sided *p* value not corrected for multiple testing. The significance level adjusted for multiple testing was $\alpha = 0.05/(10 \times 4) = 0.00125$.

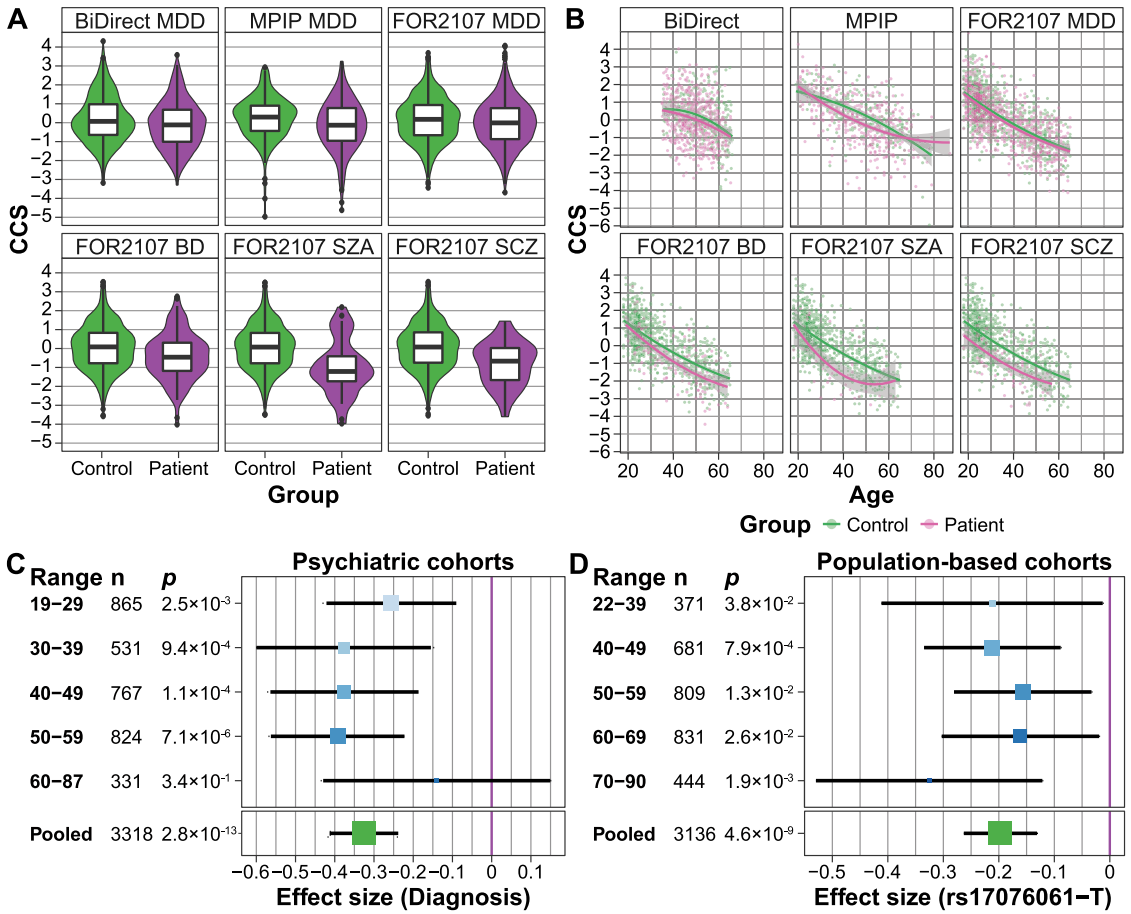


Fig. 3 Analyses of age-by-diagnosis and age-by-SNP effects on the component of the common neurobiological substrate (CCS). A A significant smaller CCS was observed in MDD (BiDirect, MPIP, FOR2107), BD (FOR2107), SZA (FOR2107), and SCZ (FOR2107) patients, affirming the transdiagnostic finding by Goodkind et al.[1] (Supplementary Table S15). **B** Age[2] trajectories of the patient/control groups plotted for each cohort. A non-linear, quadratic age dependency was observed in MDD (pooled MPIP, BiDirect, FOR2107), but no other diagnostic group. Data points represent the CCS after residualization against all covariates except for age and age[2] (separate fit for patients and controls (Supplementary Table S15). **C** Age-stratified analyses of the association between diagnosis and the CCS using five age groups in the combined patient/control cohorts (with cohort as a covariate). No significant heterogeneity was observed. Size and color of the effect sizes per bin are proportional to the sample size. **D** Age-stratified analyses of the association between the top SNP (rs17076061) and the CCS using five age groups in the combined five population-based GWAS cohorts (with cohort modeled as a covariate). No significant heterogeneity between the age groups was observed. Size and color of the effect sizes per bin are proportional to the sample size.

(Supplementary Fig. S9 and Supplementary Table S11), sign tests (Supplementary Table S12), or PGS analyses (Supplementary Fig. S7 and Supplementary Table S13) after correction for multiple testing.

Discussion

In the present study, we investigated the genetic architecture of an MRI-based volumetric marker that has previously been identified as a common neurobiological substrate for major psychiatric disorders[1], mapping to areas of the salience network. As the primary analysis, we conducted a population-based GWAS on this substrate that was calculated from the original three-region substrate using dimensional reduction by PCA. Thereby, we generated the CCS, a construct that simplified our genetic analyses while retaining a large fraction of the phenotypic variance. In secondary analyses, we studied the relationship between the CCS and risk for psychiatric disease as well as age-by-SNP and age-by-diagnosis effects on the CCS. Overall, our study produced three main findings.

First, the minor allele T of the intergenic SNP rs17076061 was associated with a decreased CCS at genome-wide significance and replicated. The association signal from the CCS was stronger than those from the three separate regions indicating that our approach stabilized the CCS association by reducing the statistical noise. The SNP maps directly to an evolutionarily constrained element in mammals[63], supporting a regulatory

role of the variant. The locus on chromosome *5q35.2* harbors several predicted, uncharacterized long intergenic non-coding RNAs and two protein-coding genes expressed in the brain with either psychiatric or neuroprotective functions[64–67]. The latter genes are biorientation of chromosomes in cell division 1 (*BOD1*) and stanniocalcin 2 (*STC2*), located 75 kbp downstream and 202 kbp upstream of rs17076061, respectively.

The SNP is part of a significant expression quantitative trait locus (eQTL) for *STC2* in pancreatic tissue ($p = 3.6 \times 10^{-8}$). However, this eQTL was not significant in the anterior cingulate cortex ($p = 0.06$), and the anterior insula was not available in GTEx v8[68]. Notably, the sample size for the ACC was half of that for the pancreas, decreasing the statistical power. In neurons, rs17076061 thus likely influences the expression of *STC2*, which expresses a secreted glycoprotein with a possible auto- or paracrine function. In the regulation of apoptosis, the unfolded protein response promotes the expression of the potentially neuroprotective STC2 in neuronal cells[66,67].

Our second main finding is that the neurodevelopmental pathway "SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion" was significantly associated with the CCS. Semaphorin-3A (SEMA3A) is a chemorepellent mediating axon guidance and a chemoattractant for dendrite growth, whereas plexins are the signal-transducing subunits of the Semaphorin-3A receptor. Semaphorin-3A and Plexin-A2 are associated with different psychiatric disorders:[69–72] Plexin-A2 is associated with schizophrenia, anxiety, and MDD[72,73], while Semaphorin-3A is upregulated in the brain of schizophrenia patients and has been suggested to contribute to the synaptic pathology of the disorder[70]. Furthermore, Semaphorin-3A may contribute to neurodegeneration in Alzheimer's disease[71], and the pathway is important for neuronal regeneration after brain trauma[74].

A third set of analyses focused on the question whether our approach—correlating a disease-associated structural brain phenotype with population-based genomic variation—would lead to the detection of genetic variants relevant for psychiatric disorders. Here, we found a discrepancy between detecting a genome-wide significant SNP (rs17076061) on the one hand, while not detecting an association between this SNP and major psychiatric diagnoses (MDD, BD, and schizophrenia) on the other hand. This finding obviously contradicts the latent expectation that the CCS could represent a "risk endophenotype" that exhibits a substantial heritability of 50% in the studied population. Although our top SNP explained only a small fraction of the CCS variance ($R^2 =$ 1.2%, sample size-weighted mean across three cohorts), there still remains a disconnection between this finding and the lack of an observed psychiatric risk conveyed by the SNP.

One explanation for this observation is the low correlation between the CCS and psychiatric diagnoses: Goodkind et al.[1] used the revised activation likelihood estimation (ALE) meta-analysis framework to test for a spatial convergence of morphometric patient/control differences and found the three-region substrate. However, ALE does not process effect sizes from the original studies, which impeded a comparison with our results. We thus analyzed patient/control cohorts of the affective-psychosis spectrum to assess the CCS variance explained by the diagnostic status, ranging from 1.0% for MDD to 4.2% for schizophrenia (R^2). Therefore, in a model that attributes disease risk to the presence of a smaller CCS (less GM), we expect the risk effect mediated by a single SNP to be very low. Compatible with this model, the association of rs1707601 with disease risk was only nominally significant in the large and most recent cross-disorder study by the PGC (26,432 patients and 49,926 controls[22]). Evidence from large consortium studies showed that psychiatric disease-specific PGSs explain only a small fraction of the disease phenotype[19]. This, along with the low disease/CCS correlation, may explain our observation that PGSs calculated from published GWAS were not associated with the CCS in our population-based cohorts.

The polygenic nature of both the CCS and risk for psychiatric disease demanded more detailed comparisons between association signals from the CCS GWAS and GWAS of major psychiatric disorders applying complementary statistical approaches (LDSC, RRHO, binominal sign tests). Our results suggest that no such genetic overlap exists, adding our study to a line of similar previous reports: Large studies on MDD and schizophrenia, for example, found only weak or no relationship between the genetic architecture of these diagnoses and regional brain volumes[2,55,75–77]. Similarly, a meta-analysis of genetic factors influencing subcortical volumes in about 40,000 individuals identified no robust correlations between subcortical volumes and BD or schizophrenia[75]. One may speculate that differences between disease-predisposing ("causal") and secondary ("epiphenomenological") brain changes (due to substance use or other comorbidities) could play a role for this heterogeneity. Methodologically, the analyses of genetic overlap, as conducted by us and others[61,75], investigated genome-wide similarities between GWAS. If only some variants showed a joint association or different loci exhibited mixed effect directions, these methods could fail to detect similarities. Similarly, our PGSs for a larger CCS were not lower in psychiatric patients diagnosed with MDD, BD, schizoaffective disorder, or schizophrenia. This finding supports the hypothesis that the standard approach of PGS, which only accounts for common additive effects, does not adequately capture epistatic gene-by-gene or

gene-by-environment effects that influence complex traits and, even more, disease risk. Future studies are warranted to explore such relationships based on models that allow for non-additive, particularly interactive effects[78].

Another possible explanation for the dissociation between our genetic findings and disease risk is that other premorbid environmental influences, such as the prenatal environment or early life adversity, were not addressed in our study. Such influences could aggravate a morphological risk pattern without being directly reflected in genetic associations. Well-documented examples for these influences are specific correlations between early childhood adversity and salience network dysfunction or GM loss[79–81]. In this line of thinking, undetected environmental factors may have shaped the CCS beyond genetic effects in our population cohorts. It is evident that only longitudinal studies of patients and controls can disentangle this challenging question, particularly as longitudinal brain changes themselves show a significant heritable component[82].

In our attempt to understand the function of the top SNP from our GWAS (rs17076061), we considered that aspects of pathological aging (accelerated aging) could play a role. In this regard, reports on different structural brain markers suggest that several major psychiatric diseases are associated with accelerated aging, with different effect sizes and different regional patterns[6,83]. The salience network, in particular, is involved in an accelerated cognitive decline during aging[84]. Beyond a cross-sectional replication of small but robust CCS differences between patients and controls, we recognized that the CCS could harbor non-linear age-by-diagnosis interactions in MDD. In fact, the SNP effect proved robust against the inclusion of age-interaction terms, without significant heterogeneity when analyzed in age-binned subgroups. These results suggest that rs17076061 may have a stable effect on the CCS over the adult lifespan. However, we could not entirely exclude the influence of higher-order non-linear deviations that we could not analyze in the present study. Concordant with this observation, we did not find genetic overlaps between our GWAS and GWAS of longevity (representing an extreme form of healthy aging), or bvFTD (representing an extreme form of salience network degeneration). To clarify the relationship between the CCS and a possibly accelerated salience network aging in psychiatric disease, larger patient/control cohorts are required that allow triple-interaction analyses (genetics, disease status, and CCS).

Our study has several limitations. First, more comprehensive investigations of age dependencies would have been possible from more homogeneous age distributions in the population cohorts. Still, our main goal demanded to assemble large samples, given the expected small effects of common variants. Second, environmental factors such as childhood adversity were either not available or acquired with heterogeneous instruments in the population cohorts, preventing an inclusion of this dimension as an important source of variance or interaction factor. Third, the operationalization of the CCS followed the specific result map of Goodkind et al.[1], which is a sparse representation of the salience network. Data-driven definitions, e.g., through structural covariance, as exemplified before[85], may capture a larger portion of the volumetric variance of the salience network[17,18].

In conclusion, we detected a replicable, genome-wide significant association of a common variant (rs17076061) with GM areas that represent hubs of the salience network in adult individuals from the general population. The genetic architecture of this network was not correlated with genetic risk for major psychiatric disorders. Future gene-by-environment interaction and functional imaging analyses may enable us to understand how salience network structure translates to psychiatric disease risk.

Acknowledgements

The study was supported by the German Federal Ministry of Education and Research (BMBF), through the Integrated Network IntegraMent, under the auspices of the e:Med programme (grants 01ZX1314A to M. M. N. and S. Cichon, 01ZX1614J to B. M.-M.), by the German Research Foundation (DFG; grant FOR2107: NO 246/10-1 to M. M. N., MU1315/8-2 to B. M.-M.), by the European Union's Horizon 2020 Research and Innovation Programme (grants 785907 (HBP SGA2) to S. Cichon, S. Caspers, and K. A. and 826421 (VirtualBrainCloud) to S. B. E.), and by the Swiss National Science Foundation (SNSF; grant 156791 to S. Cichon). T. F. M. A. was supported by the BMBF through the DIFUTURE consortium of the Medical Informatics Initiative Germany (grant 01ZZ1804A) and by the European Union's Horizon 2020 Research and Innovation Programme (grant MultipleMS, EU RIA 733161). S. B. E., K. A., and F. H. were supported by the Helmholtz Portfolio Theme Supercomputing and Modeling for the Human Brain, S. Caspers receives funding from the Initiative and Networking Fund of the Helmholtz Association, and H. M. from the BMBF (01ER1205); M. M. N. is a member of the DFG-funded cluster of excellence ImmunoSensation. The BiDirect study is supported by grants of the BMBF to the University of Münster (01ER0816 and 01ER1506). SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the BMBF (01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network "Greifswald Approach to Individualized Medicine (GANI_MED)," funded by the BMBF (03IS2061A). Whole-body MR imaging was supported by a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. Genome-wide data were supported by the BMBF (03ZIK012). The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. We thank the Heinz Nixdorf Foundation, Germany, for the generous support of the Heinz Nixdorf Recall Study. We thank the respective working groups of the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/) for making the summary statistics of their GWAS available. This work is part of the German multicenter consortium "Neurobiology of Affective Disorders. A translational perspective on brain structure and function," funded by the DFG (Forschungsgruppe/Research Unit FOR2107). The FOR2107 study was funded by the DFG: grants KI 588/14-1, KI 588/14-2 to T. K.; DA 1151/5-1, DA 1151/5-2 to U. D.; NE 2254/1-2 to I. N.; NO 246/10-1, NO 246/10-2 to M. M. N.; MU1315/8-2 to B. M.-M.; extended FOR2107 acknowledgments are available in the Supplement. Extended MPIP acknowledgments are available in the Supplement. We thank the International FTD-Genomics Consortium (IFGC; https://ifgcsite.wordpress.com/) for sharing summary statistics for the bvFTD

subgroup. The research group's members and their affiliations and funding sources can be found in the Supplementary Material. We thank the research participants and employees of 23andMe, Inc. for their contribution to the MDD meta-analysis published in[50]. The members of the 23andMe Research Team are listed in the Supplementary Material.

Author details
[1]Max Planck Institute of Psychiatry, Munich, Germany. [2]Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. [3]Institute of Neuroscience and Medicine (INM-1, INM-7), Research Centre Jülich, Jülich, Germany. [4]Cécile and Oskar Vogt Institute of Brain Research, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [5]Department of Biomedicine, University of Basel, Basel, Switzerland. [6]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [7]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Greifswald, Germany. [8]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [9]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [10]Institut für Energie- und Umwelttechnik e.V. (IUTA, Institute of Energy and Environmental Technology), Duisburg, Germany. [11]Department of Psychology, Center for Lifespan Changes in Brain and Cognition, University of Oslo, Oslo, Norway. [12]Department of Psychiatry and Psychotherapy, Westfälische Wilhelms-Universität Münster, Münster, Germany. [13]Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany. [14]Institute for Anatomy I, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [15]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. [16]Institute of Human Genetics, University of Bonn, Bonn, Germany. [17]Department of Psychiatry and Psychotherapy, Philipps-Universität Marburg, Marburg, Germany. [18]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Marburg, Germany. [19]Marburg University Hospital – UKGM, Marburg, Germany. [20]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Essen, Germany. [21]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [22]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94304, USA. [23]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94304, USA. [24]Veterans Affairs Palo Alto Healthcare System, and the Sierra Pacific Mental Illness, Research, Education, and Clinical Center (MIRECC), Palo Alto, CA 94394, USA. [25]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [26]JARA-Brain, Jülich-Aachen Research Alliance, Jülich, Germany. [27]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [28]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [29]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland

Funding
Open Access funding enabled and organized by Projekt DEAL

Conflict of interest
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-021-01317-7.

Received: 29 October 2019 Revised: 7 January 2021 Accepted: 20 January 2021
Published online: 29 March 2021

References

1. Goodkind, M. et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315 (2015).
2. Smeland, O. B. et al. Genetic overlap between schizophrenia and volumes of hippocampus, putamen, and intracranial volume indicates shared molecular genetic mechanisms. *Schizophr. Bull.* **44**, 854–864 (2017).
3. van Erp, T. G. M. et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* **21**, 547–553 (2015).
4. van Erp, T. G. M. et al. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol. Psychiatry* **84**, 644–654 (2018).
5. Schmaal, L. et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol. Psychiatry* **21**, 806–812 (2016).
6. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).
7. Hibar, D. P. et al. Subcortical volumetric abnormalities in bipolar disorder. *Mol. Psychiatry* **21**, 1710–1716 (2016).
8. Koutsouleris, N. et al. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. *Schizophr. Bull.* **40**, 1140–1153 (2014).
9. Menon V. Salience network. 597–611 (Elsevier, 2015).
10. Seeley, W. W. et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* **27**, 2349–2356 (2007).
11. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.* **16**, 55–61 (2015).
12. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62**, 42–52 (2009).
13. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).
14. Tozzi, L. et al. Interactive impact of childhood maltreatment, depression, and age on cortical brain structure: mega-analytic findings from a large multi-site cohort. *Psychol. Med.* **50**, 1020–1031 (2020).
15. Alloza, C. et al. Psychotic-like experiences, polygenic risk scores for schizophrenia and structural properties of the salience, default mode and central-executive networks in healthy participants from UK Biobank. *Transl. Psychiatry* **10**, 122 (2020).
16. DuPre, E. & Spreng, R. N. Structural covariance networks across the life span, from 6 to 94 years of age. *Netw. Neurosci.* **1**, 302–323 (2017).
17. Fjell, A. M. et al. Critical ages in the life course of the adult brain: nonlinear subcortical aging. *Neurobiol. Aging* **34**, 2239–2247 (2013).
18. Ziegler, G. et al. Brain structural trajectories over the adult lifespan. *Hum. Brain. Mapp.* **33**, 2377–2389 (2012).
19. Sullivan, P. F. et al. Psychiatric genomics: an update and an agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
20. Pettersson, E. et al. Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. *Psychol. Med.* **49**, 1166–1173 (2019).
21. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, 6395 (2018).
22. Lee, P. H. et al. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482 (2019).
23. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
24. Mühleisen, T. W., Forstner, A., Hoffmann, P. & Cichon, S. Brain imaging genomics: influences of genomic variability on the structure and function of the human brain. *Medizinische Genetik.* **32**, 47–56 (2020).
25. Caspers, S. et al. Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. *Front. Aging Neurosci.* **6**, 149 (2014).
26. Roski, C. et al. Activation shift in elderly subjects across functional systems: an fMRI study. *Brain Struct. Funct.* **219**, 707–718 (2014).
27. Teismann, H. et al. Establishing the bidirectional relationship between depression and subclinical arteriosclerosis—rationale, design, and characteristics of the BiDirect Study. *BMC Psychiatry* **14**, 174 (2014).
28. Volzke, H. et al. Cohort profile: the study of health in Pomerania. *Int. J. Epidemiol.* **40**, 294–307 (2011).

29. Hermesdorf, M. et al. Reduced fractional anisotropy in patients with major depressive disorder and associations with vascular stiffness. *Neuroimage Clin.* **14**, 151–155 (2017).
30. Hennings, J. M. et al. Clinical characteristics and treatment outcome in a representative sample of depressed inpatients—findings from the Munich Antidepressant Response Signature (MARS) project. *J. Psychiatr. Res.* **43**, 215–229 (2009).
31. Inkster, B. et al. Association of GSK3beta polymorphisms with brain structural changes in major depressive disorder. *Arch. Gen. Psychiatry* **66**, 721–728 (2009).
32. Kircher, T. et al. Neurobiology of the major psychoses: a translational perspective on brain structure and function-the FOR2107 consortium. *Eur. Arch. Psychiatry Clin. Neurosci.* **26**, 949–962 (2019).
33. Vogelbacher, C. et al. The Marburg-Munster Affective Disorders Cohort Study (MACS): a quality assurance protocol for MR neuroimaging data. *Neuroimage* **172**, 450–460 (2018).
34. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
35. Ising, M. et al. A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch. Gen. Psychiatry* **66**, 966–975 (2009).
36. Caspers, S. et al. Pathway-specific genetic risk for Alzheimer's disease differentiates regional patterns of cortical atrophy in older adults. *Cereb. Cortex* **30**, 801–811 (2019).
37. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
38. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015 (2015).
39. Gao, F. et al. XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J. Hered.* **106**, 666–671 (2015).
40. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955 (2012).
41. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
42. Andlauer, T. F. et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci. Adv.* **2**, e1501678 (2016).
43. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
44. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
45. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
46. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
47. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PloS Comput. Biol.* **11**, e1004219 (2015).
48. Segrè, A. V. et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PloS Genet.* **6**, e1001058 (2010).
49. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
50. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
51. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
52. Ferrari, R. et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).
53. Deelen, J. et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).
54. Lu, A. T. et al. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat. Commun.* **8**, 15353 (2017).
55. Franke, B. et al. Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat. Neurosci.* **19**, 420–431 (2016).
56. Andlauer, T. F. M. et al. Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. *Mol. Psychiatry.* **26**, 1286–1298 (2021).
57. Andlauer, T. F. M. & Nöthen, M. M. Polygenic scores for psychiatric disease: from research tool to clinical application. *Medizinische Genetik.* **32**, 39–45 (2020).
58. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
59. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
60. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).
61. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
62. Pievani, M. et al. Coordinate-based meta-analysis of the default mode and salience network for target identification in non-invasive brain stimulation of Alzheimer's disease and behavior variant frontotemporal dementia networks. *J. Alzheimer's Dis.* **57**, 825–843 (2017).
63. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
64. Esmaeeli-Nieh, S. et al. BOD1 is required for cognitive function in humans and drosophila. *PloS Genet.* **12**, e1006022 (2016).
65. Kim, J. et al. Somatic deletions implicated in functional diversity of brain cells of individuals with schizophrenia and unaffected controls. *Sci. Rep.* **4**, 3807 (2014).
66. Byun, J. S. et al. Neuroprotective effects of stanniocalcin 2 following kainic acid-induced hippocampal degeneration in ICR mice. *Peptides* **31**, 2094–2099 (2010).
67. Ito, D. et al. Characterization of stanniocalcin 2, a novel target of the mammalian unfolded protein response with cytoprotective properties. *Mol. Cell Biol.* **24**, 9456–9469 (2004).
68. Mele, M. et al. Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
69. Costa, C. et al. Expression of semaphorin 3A, semaphorin 7A and their receptors in multiple sclerosis lesions. *Mult. Scler. J.* **21**, 1632–1643 (2015).
70. Eastwood, S. L., Law, A. J., Everall, I. P. & Harrison, P. J. The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology. *Mol. Psychiatry* **8**, 148–155 (2003).
71. Good, P. F. et al. A role for semaphorin 3A signaling in the degeneration of hippocampal neurons during Alzheimer's disease. *J. Neurochem.* **91**, 716–736 (2004).
72. Wray, N. R. et al. Anxiety and comorbid measures associated with PLXNA2. *Arch. Gen. Psychiatry* **64**, 318–326 (2007).
73. Mah, S. et al. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol. Psychiatry* **11**, 471–478 (2006).
74. Mecollari, V., Nieuwenhuis, B. & Verhaagen, J. A perspective on the role of class III semaphorin signaling in central nervous system trauma. *Front. Cell Neurosci.* **8**, 328 (2014).
75. Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
76. Wigmore, E. M. et al. Do regional brain volumes and major depressive disorder share genetic architecture? A study of Generation Scotland (n=19 762), UK Biobank (n=24 048) and the English Longitudinal Study of Ageing (n=5766). *Transl. Psychiatry* **7**, e1205 (2017).
77. Reus, L. M. et al. Association of polygenic risk for major psychiatric illness with subcortical volumes and white matter integrity in UK Biobank. *Sci. Rep.* **7**, 42140 (2017).
78. Zwir, I. et al. Uncovering the complex genetics of human personality: response from authors on the PGMRA Model. *Mol. Psychiatry* **25**, 2210–2213 (2020).
79. Yu, M. et al. Childhood trauma history is linked to abnormal brain connectivity in major depression. *Proc. Natl Acad. Sci. USA* **116**, 8582–8590 (2019).
80. van der Werff, S. J. A. et al. Resting-state functional connectivity in adults with childhood emotional maltreatment. *Psychol. Med.* **43**, 1825–1836 (2013).
81. van Harmelen, A. L. et al. Reduced medial prefrontal cortex volume in adults reporting childhood emotional maltreatment. *Biol. Psychiatry* **68**, 832–838 (2010).
82. Brouwer, R. M. et al. Genetic influences on individual differences in longitudinal changes in global and subcortical brain volumes: results of the ENIGMA plasticity working group. *Hum. Brain Mapp.* **38**, 4444–4458 (2017).

83. Han L. K. M. et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol. Psychiatry* https://doi.org/10.1038/s41380-020-0754-0 (2020).
84. La Corte, V. et al. Cognitive decline and reorganization of functional connectivity in healthy aging: the pivotal role of the salience network in the prediction of age and cognitive performances. *Front. Aging Neurosci.* **8**, 204 (2016).
85. Gupta, C. N., Turner, J. A. & Calhoun, V. D. Source-based morphometry: a decade of covarying structural brain patterns. *Brain Struct. Funct.* **224**, 3031–3044 (2019).