


# SCIENTIFIC REPORTS



OPEN

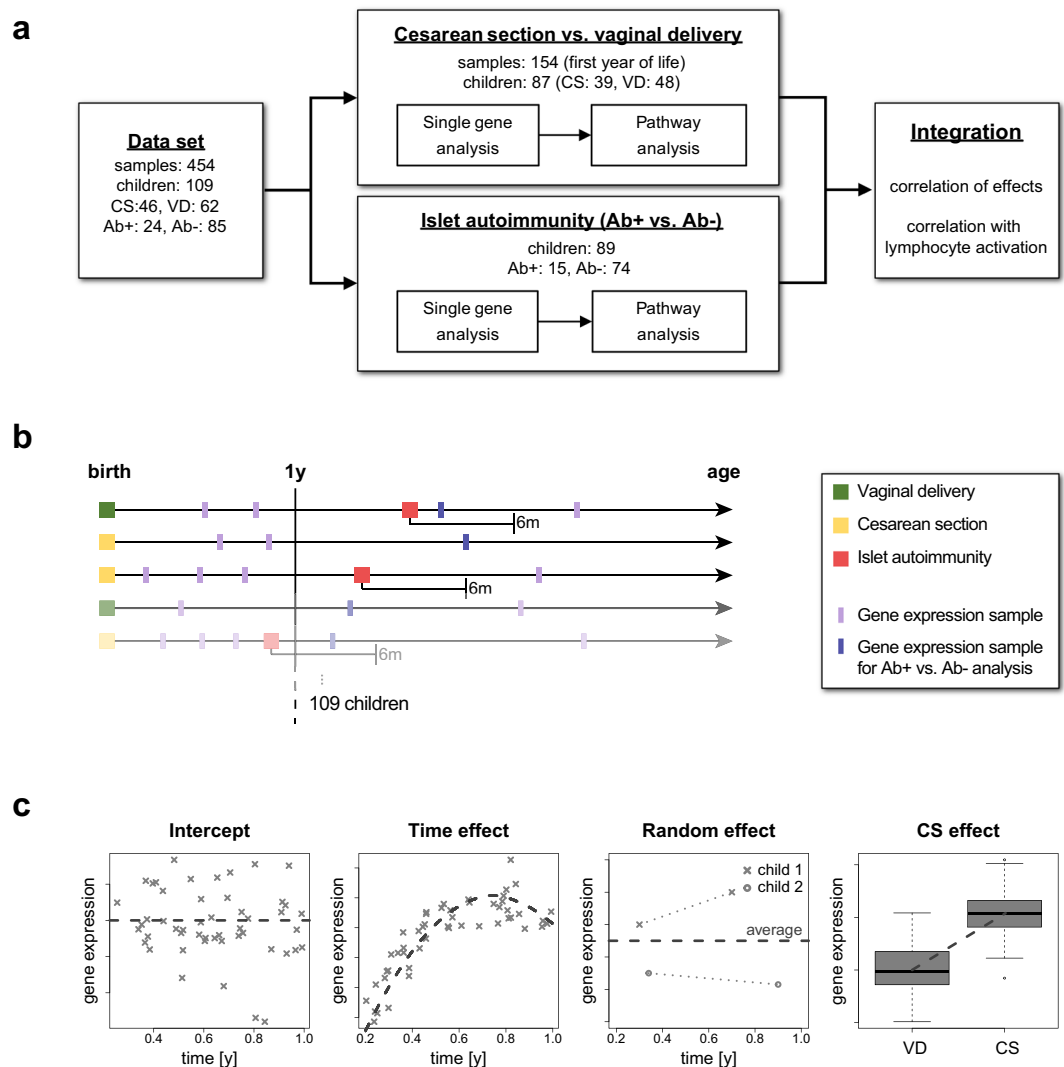
## Common patterns of gene regulation associated with Cesarean section and the development of islet autoimmunity – indications of immune cell activation

M. Laimighofer<sup>1,2</sup>, R. Lickert<sup>3</sup>, R. Fuerst<sup>3</sup>, F. J. Theis <sup>1,2</sup>, C. Winkler<sup>3</sup>, E. Bonifacio<sup>4,5</sup>, A.-G. Ziegler<sup>3</sup> & J. Krumsiek<sup>1,6,7</sup>

Birth by Cesarean section increases the risk of developing type 1 diabetes later in life. We aimed to elucidate common regulatory processes observed after Cesarean section and the development of islet autoimmunity, which precedes type 1 diabetes, by investigating the transcriptome of blood cells in the developing immune system. To investigate Cesarean section effects, we analyzed longitudinal gene expression profiles from peripheral blood mononuclear cells taken at several time points from children with increased familial and genetic risk for type 1 diabetes. For islet autoimmunity, we compared gene expression differences between children after initiation of islet autoimmunity and age-matched children who did not develop islet autoantibodies. Finally, we compared both results to identify common regulatory patterns. We identified the pentose phosphate pathway and pyrimidine metabolism - both involved in nucleotide synthesis and cell proliferation - to be differentially expressed in children born by Cesarean section and after islet autoimmunity. Comparison of global gene expression signatures showed that transcriptomic changes were systematically and significantly correlated between Cesarean section and islet autoimmunity. Moreover, signatures of both Cesarean section and islet autoimmunity correlated with transcriptional changes observed during activation of isolated CD4<sup>+</sup> T lymphocytes. In conclusion, we identified shared molecular changes relating to immune cell activation in children born by Cesarean section and children who developed autoimmunity. Our results serve as a starting point for further investigations on how a type 1 diabetes risk factor impacts the young immune system at a molecular level.

Type 1 diabetes is an autoimmune disease in which immune cells destroy insulin-producing beta cells in the pancreas. Loss of beta-cell mass leads to insulin deficiency, impaired glucose tolerance, and ultimately the onset of type 1 diabetes<sup>1</sup>. This process is marked by the appearance of islet autoantibodies to beta-cell-related antigens<sup>2</sup>. Autoantibody seroconversion usually occurs during childhood and early adolescence, with a peak conversion rate

<sup>1</sup>Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany. <sup>2</sup>Department of Mathematics, Technische Universität München, Garching, Germany. <sup>3</sup>Institute of Diabetes Research, Helmholtz Zentrum München, and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universität München, München, Germany. <sup>4</sup>DFG Center for Regenerative Therapies Dresden, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany. <sup>5</sup>Paul Langerhans Institute Dresden, German Center for Diabetes Research (DZD), Technische Universität Dresden, Dresden, Germany. <sup>6</sup>German Center for Diabetes Research (DZD), Neuherberg, Germany. <sup>7</sup>Institute for Computational Biomedicine, Englander Institute for Precision Medicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, USA. Correspondence and requests for materials should be addressed to J.K. (email: [jak2043@med.cornell.edu](mailto:jak2043@med.cornell.edu))

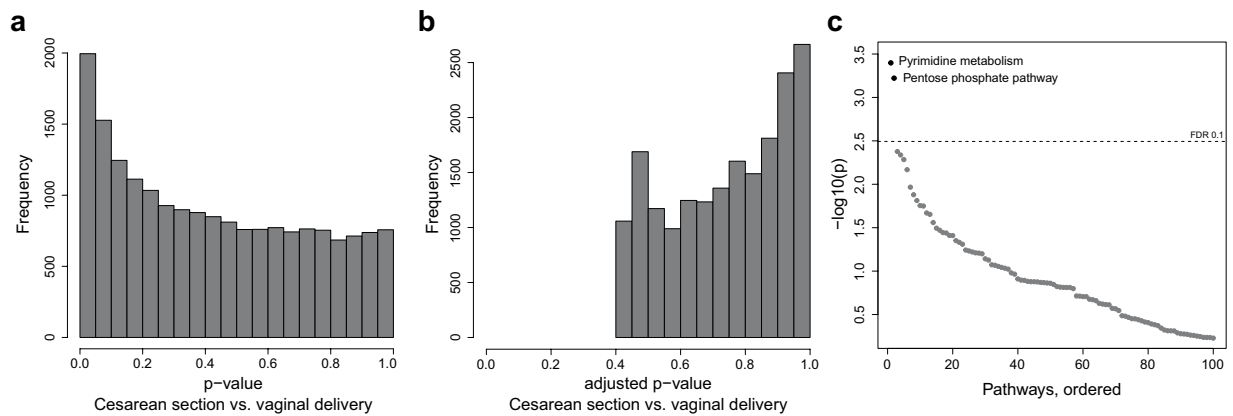


**Figure 1.** Overview. **(a)** Study workflow: Parallel analyses were performed to detect differential gene expression and pathway enrichment for Cesarean section and islet autoimmunity. The results were then compared in a combined analysis and related to expression patterns of lymphocyte activation. **(b)** Schematic overview of the longitudinal study design for Cesarean section and islet autoimmunity analyses. **(c)** Schematic illustration of the generalized additive mixed effect model (GAMM) to analyze the longitudinal dataset, including intercept, a time effect, a random effect for multiple measurements, and the investigated Cesarean section vs. vaginal delivery effect. Abbreviations: CS = Cesarean section, VD = Vaginal delivery, Ab+ = Islet autoantibody-positive, Ab- = Islet autoantibody-negative.

between 9 months and two years<sup>3</sup>. Notably, an increasing incidence of type 1 diabetes has been observed within the last few decades, especially in children<sup>4</sup>. Several previous studies indicated transcriptional changes in immune cells caused by the development of islet autoantibodies<sup>5,6</sup>, indicating substantial molecular changes long before the onset of type 1 diabetes.

Children born by Cesarean section have an odds ratio of 1.23 for the development of type 1 diabetes compared to vaginally delivered children<sup>7</sup>. Moreover, Cesarean section has been reported to be associated with a faster progression to type 1 diabetes, but not with an increase the risk for islet autoimmunity<sup>8</sup>. The underlying mechanisms of these associations are not yet fully understood. Some reports indicate that the mode of delivery affects colonization of microbiota in the intestinal tract<sup>9</sup>, which in turn affects the developing immune system of infants<sup>10</sup>. Such differences in the gut microbiome and its interaction with the immune system may lead to increased risk of asthma, childhood allergies<sup>11</sup>, and autoimmune diseases, such as type 1 diabetes.

Here we investigated the impact of Cesarean section and islet autoimmunity on the immune system by analyzing gene expression data in peripheral blood mononuclear cells (PBMCs) as a readout (Fig. 1a). The overall goal of the analysis was to elucidate common immune modulation patterns between an early risk factor, Cesarean section, and the development of autoimmunity. We first investigated the effects of Cesarean section on gene expression profiles of children in the first year of life (Fig. 1b). To this end, we analyzed data from several time points in this early period in children with an increased familial risk for type 1 diabetes. A generalized additive



**Figure 2.** Cesarean section analysis. **(a)** Histogram of unadjusted p-values for CS association per gene. **(b)** Histogram of p-values after multiple testing adjustment by controlling the false discovery rate. **(c)** Sorted  $-\log_{10}(p)$ -values of pathway enrichment. Dashed line indicates the multiple testing threshold at an FDR of 0.1. Abbreviations: CS = Cesarean section, VD = Vaginal delivery.

mixed model was used to account for the longitudinal information when extracting gene expression differences between children born by Cesarean section and vaginal delivery (Fig. 1c). In a second analysis, we compared gene expression levels of children with increased risk of familial type 1 diabetes after initiation of islet autoimmunity with age-matched children who did not develop islet autoantibodies. We then combined both results to identify genes and pathways that were differentially regulated in both analyses, to link Cesarean section and islet autoimmunity at the level of gene expression. Finally, we compared the patterns identified for Cesarean section and islet autoantibodies with gene expression data from activated human CD4+ T lymphocytes to identify candidates for common molecular mechanisms of immune activation.

## Results

**Effects of Cesarean section on the transcriptome in the first year of life.** To focus on early post-partum effects on genetic regulation after delivery by Cesarean section, we restricted our analysis to samples from children in their first year of life (Supplementary Fig. S1). This resulted in a total of 154 PBMC gene expression samples (71 Cesarean section, 83 vaginal delivery) from 87 children (39 Cesarean section, 48 vaginal delivery), Fig. 1a,b. The number of samples per child ranged between 1 and 4 (Fig. S1). Details on the dataset and preprocessing steps are provided in the Methods section.

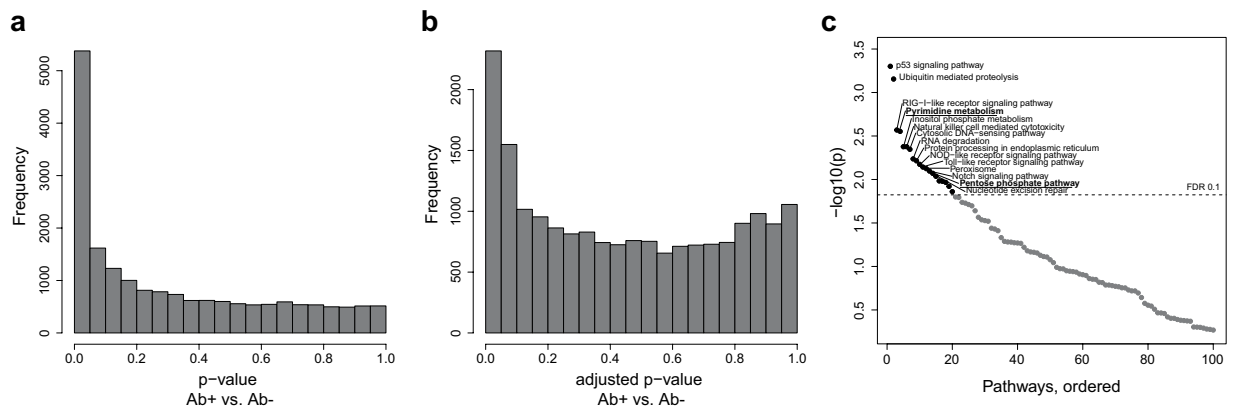
We used a generalized additive mixed effect model (GAMM) to account for the longitudinal nature of the dataset. The model consisted of three major parts (Fig. 1c): (1) the change of expression profiles over time was modeled using a spline-based function, (2) the overall trend of expression values per child was captured by a random mixed effect, and (3) a term representing the Cesarean section effect extracted the association in which we were primarily interested. Applying this model to the data set, we observed an enrichment of low p-values before adjusting for multiple testing (Fig. 2a and Supplementary Table S1). However, after multiple testing correction by controlling the false discovery rate (FDR) at 0.1, no significant single genes could be identified in this first step (Fig. 2b).

To improve statistical power and facilitate biological interpretability, we performed a pathway enrichment analysis based on the KEGG pathway database<sup>12</sup>. From this analysis, we obtained two significantly differentially expressed pathways after multiple testing correction ( $FDR \leq 0.1$ ): “Pentose phosphate pathway” and “Pyrimidine metabolism” (Fig. 2c and Supplementary Table S1). The “Pentose phosphate pathway” included 27 genes (20 up- and 7 down-regulated genes in children born by Cesarean section) and the “Pyrimidine metabolism” pathway contained 92 genes (57 up-regulated and 35 down-regulated).

Both pathways are well known to be involved in the synthesis of nucleotides during cell proliferation<sup>13,14</sup>. This indicated that there is a change in expression profiles of proliferation-related nucleotide synthesis pathways as an early consequence of delivery by Cesarean section.

**Effects of islet autoimmunity on the transcriptome.** To identify gene expression signatures associated with seroconversion, we compared the earliest sample of children after the development of islet autoantibodies (up to 6 months post-seroconversion; 15 children) with all available age-matched samples of children who did not develop islet autoantibodies (74 children). We applied linear regression to explain gene expression differences induced by islet autoantibody status, corrected for age (see Methods). We detected a strong accumulation of differentially expressed genes (Fig. 3a), of which 3,867 were significant after multiple testing correction ( $FDR \leq 0.1$ , Fig. 3b). The majority of these genes were found to be up-regulated in children with islet autoimmunity (64% up-regulated vs. 36% down-regulated, Supplementary Table 1).

Pathway enrichment analysis on KEGG pathways identified 20 as significantly regulated ( $FDR \leq 0.1$ , Fig. 3c and Supplementary Table 1). The top-ranked pathways were “p53 signaling pathway”, “Ubiquitin mediated proteolysis”, and “RIG-I-like receptor signaling pathway.” The two significant pathways from the Cesarean section analysis, “Pyrimidine metabolism” and “Pentose phosphate pathway”, also appeared as significant in the islet



**Figure 3.** Ab+ analysis. **(a)** Histogram of unadjusted p-values from single gene analysis of islet autoimmunity positive vs. age-matched children who did not develop islet autoimmunity. **(b)** Histogram of p-values after multiple testing, controlling the false discovery rate. **(c)** Sorted  $-\log_{10}(p)$  values of pathway enrichment for islet autoimmunity status. Dashed line indicates the multiple testing threshold at an FDR of 0.1. Abbreviations: Ab+ = Islet autoantibody-positive, Ab- = Islet autoantibody negative.

autoantibody analysis. Notably, comparing gene expression samples of children before seroconversion (up to 6 months before) and age-matched children without islet autoantibodies yielded no significant genes or pathways (Supplementary Fig. S2). In addition, we investigated children that had gene expression samples both before and after seroconversion (up to 6 months before and after) in a paired analysis, leaving only seven children. No genes or pathways were found to be differentially expressed (Supplementary Fig. S2).

Taken together, we observed several immune system-related and nucleotide synthesis pathways associated with islet autoimmunity, which included the pathways found in our Cesarean section analysis.

**Coherent gene expression changes between Cesarean section and islet autoimmunity.** We further investigated the similarities of transcriptional changes between the two risk factors. First, we found that the individual gene regulation of both “Pyrimidine metabolism” and the “Pentose phosphate pathway” pathway showed similar patterns of up- and down-regulation between Cesarean section and islet autoimmunity (Figs 4a and S3).

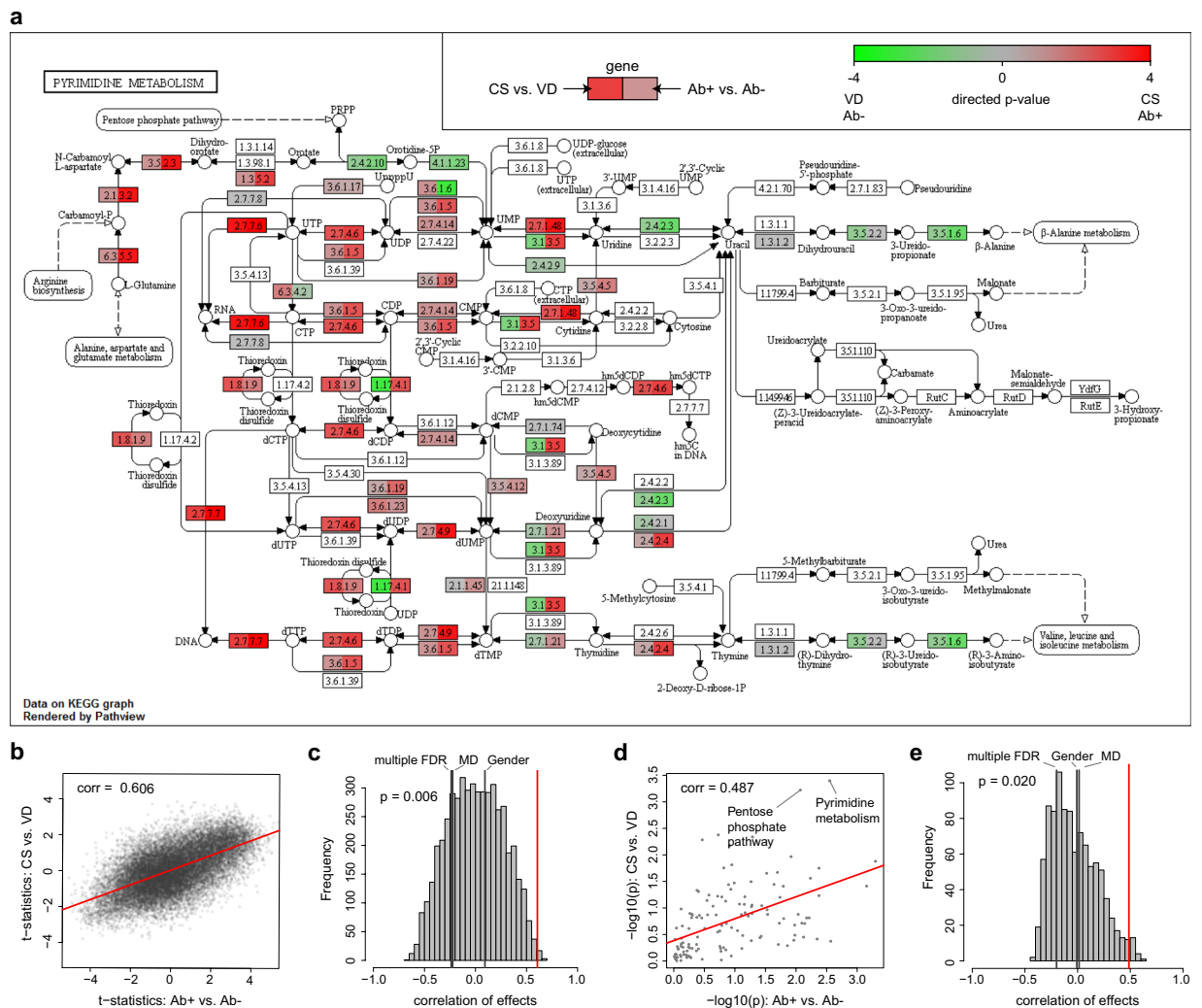
Extending this analysis, we quantified the relationship of gene regulation between the two factors at a systematic level. We correlated the standardized effects from Cesarean section and development of islet autoantibodies across all genes (Fig. 4b), revealing a striking correlation of 0.606 ( $p = 0.0062$ , Fig. 4c). In other words, genes regulated by Cesarean section, despite the rather weak signal strength in our study, were also regulated in the same direction by the initiation of islet autoimmunity. A similar correlation between the effects of Cesarean section and islet autoimmunity was observed at the pathway level, with a correlation of 0.49 ( $p = 0.02$ , Fig. 4d,e), supporting the functional agreement of transcriptional changes between the two risk factors. Importantly, we did not observe that children born by Cesarean section developed islet autoantibodies more frequently than children born by vaginal delivery, ruling out a confounding effect not related to gene expression (Supplementary Fig. S4).

In contrast to the profound Cesarean section to islet autoantibodies correlation at transcript level, we found the effects of gender, maternal diabetes and multiple first-degree relatives to be randomly correlated with Cesarean section at the single gene level (maternal diabetes,  $r = -0.22$ ,  $p = 0.53$ ; gender,  $r = 0.09$ ,  $p = 0.78$ ; multiple first-degree relatives,  $r = -0.24$ ,  $p = 0.49$ ) and at the pathway level (maternal diabetes,  $r = 0.01$ ,  $p = 0.97$ ; gender,  $r = -0.01$ ,  $p = 0.97$ ; multiple first-degree relatives,  $r = -0.23$ ,  $p = 0.34$ ), as shown in Figs 4c,e and S5. This demonstrates the specificity of the Cesarean section–islet autoantibody effect correlation.

In summary, this analysis showed that the gene expression changes in these two risk factors, Cesarean section and islet autoimmunity, are remarkably coherent.

Finally, we investigated whether the two activated pathways are specific to the development of islet autoantibodies and therefore type 1 diabetes. To this end, we performed the same pathway enrichment analysis for gene expression data from lupus disease (GSE81622), juvenile idiopathic arthritis and juvenile dermatomyositis (GSE11083, includes samples for both diseases), comparing cases against controls. In contrast to seroconversion, we did not observe the pentose phosphate pathway and pyrimidine metabolism to be activated, providing evidence that these two pathways are specific to islet autoimmunity of type 1 diabetes (Supplementary Table S2).

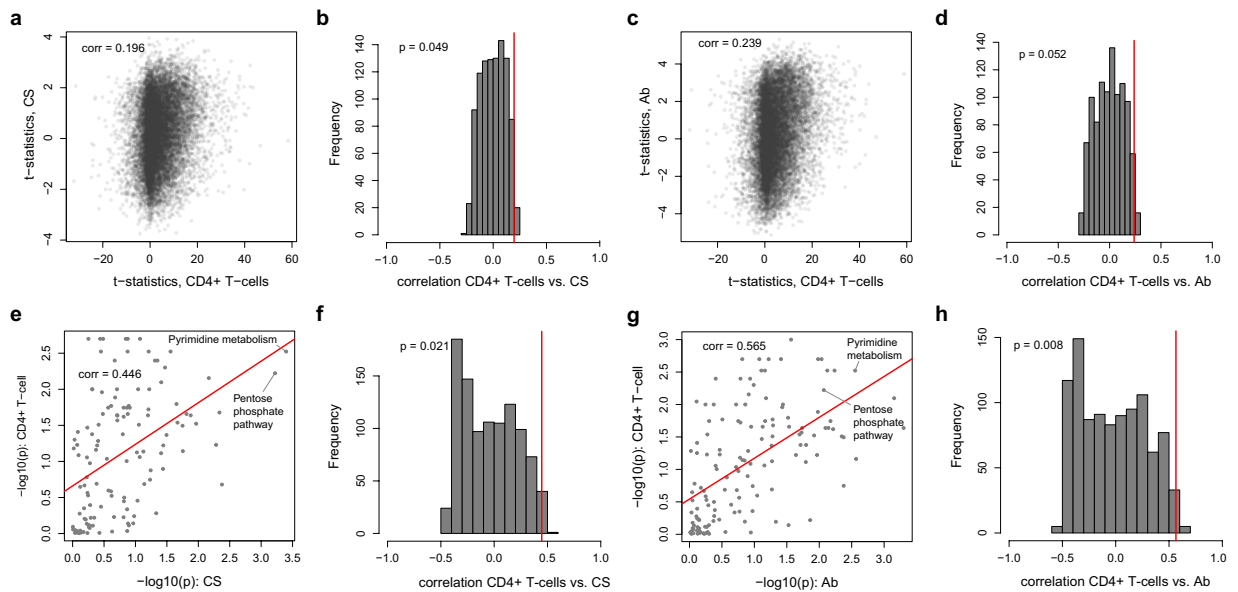
**Signatures of immune cell activation.** The pentose phosphate pathway is a universal, central metabolic pathway in the cytosol, which supports cell proliferation and survival<sup>15</sup>. The non-oxidative branch of the pentose phosphate pathway branches off glycolysis and generates ribose 5-phosphate as a precursor for the synthesis of nucleotides and amino acids necessary for cell growth and division. Moreover, the pyrimidine metabolism pathway is directly related to the pentose phosphate pathway in its role in nucleotide synthesis. Since we investigated PMBCs, these differentially regulated pathways in Cesarean section and islet autoimmunity point toward a general activation of immune cells. A direct proof of this hypothesis in the same children was unfeasible in the context of this study. Instead, we collected evidence from several secondary analysis steps.



**Figure 4.** Comparison of Cesarean section and islet autoimmunity signatures. **(a)** The pyrimidine metabolism pathway is shown as an example, with two-sided node coloring according to directed p-values ( $-\log_{10}(p) * \text{sign}(t\text{-statistic})$ ). Pathway image from KEGG<sup>12</sup>. **(b)** T-statistics per gene from both analyses; the x-axis indicates results from islet autoimmunity status and the y-axis the results from Cesarean section vs. vaginal delivery. **(c)** Empirical distribution of correlation coefficients between Cesarean section and permuted class labels of islet autoimmunity status for 5,000 permutations. The red line indicates the ‘true’ correlation between the results from the analysis of Cesarean section and islet autoimmunity status. Black lines indicate the correlation between Cesarean section and multiple first-degree relatives (multiple FDR), maternal diabetes (MD), and gender. **(d)** Correlation of pathway directed p-values in Cesarean section analysis and islet autoimmunity status analysis. **(e)** Empirical distribution of correlation coefficients between Cesarean section and permuted class labels of islet autoimmunity status at pathway level for 1,000 permutations. The red line indicates the ‘true’ correlation between Cesarean section pathways and islet autoimmunity status pathways. Abbreviations: CS = Cesarean section, VD = Vaginal delivery, Ab+ = Islet autoantibody-positive, Ab- = Islet autoantibody negative.

First, we investigated whether regulated genes from both analyses enriched immune genes annotated in innateDB<sup>16</sup>. Indeed, there was a significant accumulation of higher standardized effects for immune genes compared to non-immune genes, for both Cesarean section (Wilcoxon rank sum test:  $p = 0.0002$ ) and autoimmunity ( $p = 2.6 \times 10^{-15}$ ); see Supplementary Table S3.

Second, we compared results from the mixture of PBMC cells with published data on activation in isolated immune cells. For the analysis, we used transcriptomics data from isolated naïve and activated human CD4+ T cells<sup>17</sup>. We calculated gene expression differences before and after activation, and applied enrichment analysis to identify differentially expressed pathways. In particular, the ‘‘Pentose phosphate pathway’’ ( $p = 0.049$ ) and ‘‘Pyrimidine metabolism’’ ( $p = 0.035$ ) pathways were significantly differentially expressed in activated CD4+ T cells (Supplementary Table S1). To compare the effects of CD4+ T cell activation with effects from the Cesarean section and islet autoantibody analyses, we calculated correlations of the standardized effects. We observed borderline significant correlations between changes in activated CD4+ T cell and Cesarean section ( $r = 0.20$ ,  $p = 0.049$ , Fig. 5a,b) and islet autoimmunity ( $r = 0.24$ ,  $p = 0.052$ , Fig. 5c,d). Remarkably, the association was



**Figure 5.** Signatures of immune cell activation. (a) Correlation between single gene effects in Cesarean section (CS) compared to the association between naïve and activated CD4+ cells. (b) Histogram of correlation of permuted class labels of CD4 T cells and Cesarean section at the single gene level and the “true” correlation effect. (c) Correlation between single gene effects in islet autoimmunity status (Ab) compared to the association between naïve and activated CD4+ cells. (d) Histogram of correlation of permuted class labels of CD4 T cells and islet autoimmunity status at the single gene level and the “true” effect. (e) Correlation between pathway effects in Cesarean section compared to the association between naïve and activated pathways for CD4+ cells. (f) Histogram of correlation of permuted class labels of CD4 T cells and Cesarean section at the pathway level and the “true” effect. (g) Correlation between pathway effects in islet autoimmunity status compared to the association between naïve and activated pathways for CD4+ cells. (h) Histogram of correlation of permuted class labels of CD4 T cells and islet autoimmunity status at the pathway level and the “true” effect.

substantially stronger at pathway level for both Cesarean section ( $r = 0.45$ ,  $p = 0.021$ , Fig. 5e,f) and islet autoimmunity ( $r = 0.57$ ,  $p = 0.008$ , Fig. 5g,h). The pathway association was replicated in a second transcriptomics dataset from naïve and activated CD4+ T cells, monocytes, and natural killer cells, published in<sup>18</sup>. We found that the pathway association was present in CD4+ T cells and monocytes; however, we did not observe any association for natural killer cells. Detailed results are discussed in Supplementary Note S1.

In summary, we observed a significant correlation between the functional effects of human lymphocyte activation and the effects of Cesarean section and islet autoimmunity on gene expression in PBMCs.

## Discussion

We identified coherent gene expression signatures for Cesarean section, an early risk factor for type 1 diabetes, and islet autoantibody positivity, an obligatory stage of autoimmune response prior to the development of autoimmune type 1 diabetes. Specifically, at the transcriptome level, we identified two pathways involved in nucleotide synthesis and cell proliferation that were regulated in PBMCs of children born by Cesarean section. This analysis required an extended statistical model to incorporate the complex time information in our present dataset. Islet autoantibody analysis revealed various pathways generally involved in immune processes, including the aforementioned nucleotide synthesis pathways. Comparing global gene expression signatures, we found that transcriptomic changes were systematically and significantly correlated between the Cesarean section and islet autoantibody-positive analyses. Importantly, transcriptional signatures of Cesarean section did not correlate with gender, maternal diabetes, or multiple first-degree relatives, demonstrating that the correlation is specific to Cesarean section and islet autoimmunity. In a functional follow-up analysis, both Cesarean section and islet autoimmunity signatures were significantly correlated with gene expression changes observed during activation of isolated CD4+ T lymphocytes. At pathway level, this correlation was also observed for monocytes and natural killer cells in a second dataset.

We can speculate on the biological basis of our statistical observations. The coherent regulation of proliferation pathways in blood PBMCs may indicate a general activation of the immune system and immune cells for both Cesarean section and seroconversion. In the case of Cesarean section, this activation might indirectly reflect different microbial exposures during birth. This idea is indirectly supported by findings that the microbiome is affected by the mode of delivery, and in turn affects the developing immune system<sup>9,10</sup>. For the autoimmunity results, the precise interplay and timing of the occurrence of environmental stimuli, immune response and the development of islet autoantibodies still need to be elucidated. Regarding a hypothetical disease trajectory leading to type 1 diabetes, it is conceivable that the observed gene expression signatures may reflect a transient deflection

of the immune system and that this primes a child for subsequent progression to type 1 diabetes in the presence of other risk factors.

An interesting general observation in our analysis was the increased correlation of gene expression signatures when performing pathway analysis instead of single gene analysis. This pattern was observed both for the comparison of Cesarean section and islet autoantibody-positive signatures, and for comparisons of these two signatures with the isolated immune cells. These findings indicate that pathway analysis substantially reduced the noise compared to single gene analysis, which allowed us to identify common patterns at a functional level.

Our study could be extended and improved in several directions. (1) The present dataset has rather low statistical power for the islet autoantibody analysis, with only 15 positive samples available for at most 6 months post-seroconversion. While differential expression changes were remarkably significant, the analysis should be validated with a larger sample size. (2) The hypothesis of coherent immune system activation should ideally be confirmed in an isolated primary T cell population from children after Cesarean section or after seroconversion. We took an indirect route using PBMCs rather than a more general activation experiment with isolated immune cells. (3) The investigated samples stem from a study with increased familial risk for type 1 diabetes. Ideally, future studies would include gene expression measurements of children without familial risk. (4) The incorporation of further environmental factors, such as nutrition and medical parameters, in combination with the children's genetic background, is expected to give a more complete picture of the parameters leading to autoimmunity and type 1 diabetes.

In summary, we found a transcriptional link between Cesarean section and islet autoimmunity, pointing toward a transiently altered immune system in the susceptible period of islet autoimmunity generated by Cesarean section, which was remarkably coherent with the changes observed after islet autoimmunity.

## Methods

**PBMC gene expression data.** We used the BABYDIET PBMC gene expression data deposited in ArrayExpress (Accession Number: E-MTAB-1724)<sup>6</sup>, in combination with non-public data on Cesarean section, islet autoimmunity, family history, gender and age. In this data set, 454 samples from 109 children were available. One individual of the original 109 children was removed in the Cesarean section analysis, since no information about the type of delivery was recorded. Raw gene expression data of 33,297 probes were normalized using the Robust Multi-Array Average (RMA) method<sup>19</sup>. Probes without annotation in the Affymetrix hugene11 data and duplicates were removed using the R package *genefilter* with default settings, leaving 18,720 genes for further analysis. Age at sampling ranged from 0.21 years to 9.15 years, with a median of 1.53 years.

**Generalized additive mixed model for time-resolved data.** Transcriptomics samples were available for multiple time points per child. To model gene expression from multiple, non-matching timepoints in relation to Cesarean section in a joint approach, we employed a generalized additive mixed model (GAMM)<sup>20</sup>. In this GAMM, the model structure is defined as

$$y_{i,j} = \beta_{0,i} + \beta_{CS,i} \mathbf{x}_{CS,j} + \sum_k b_{k,i}(\mathbf{t}_{i,j}) \beta_{k,i} + b_{ID,i} \mathbf{x}_{ID,j} + \epsilon_{i,j} \quad (1)$$

where  $y_{i,j}$  is the gene expression of sample  $j$  for gene  $i$ ,  $\beta_{0,i}$  is the intercept or gene-wise average expression,  $\beta_{CS,i}$  describes the effect of Cesarean section on gene expression with  $\mathbf{x}_{CS,j} = 1$  for Cesarean section and  $\mathbf{x}_{CS,j} = 0$  for vaginal delivery. Adjusting for the multiple measurements in time  $\mathbf{t}_{i,j}$ , a spline function was included as  $\sum_k b_{k,i}(\mathbf{t}_{i,j}) \beta_{k,i}$  per gene over all samples, where  $k$  is the estimated number of basis functions,  $b_{k,i}$  are the basis functions of the spline and  $\beta_{k,i}$  their regression coefficients. In addition, we included a random effect as  $b_{ID,i} \sim N(0, \sigma_i)$ , with  $\sigma_i$  estimated per gene, correcting for the dependence of several measurements per child, denoted as  $\mathbf{x}_{ID,j}$ , and an error term  $\epsilon_{i,j}$  as i.i.d. normally distributed noise. The noise  $\epsilon_{i,j}$  and the noise of the subject specific random effect  $\sigma_i$  were assumed to be independent. Inference of the GAMM was performed using a restricted maximum likelihood approach<sup>21</sup> using the R package *mgcv*.

**Linear model for seroconversion analysis.** To investigate differences in gene expression between children after development of islet autoantibodies and age-matched children who did not develop islet autoantibodies, we selected the first sample up to six months after development of islet autoimmunity for each child, when such a sample was available (see Fig. 1b). These children were compared to all available age-matched children who did not develop autoantibodies, by applying a linear regression model per gene:

$$y_{i,j} = \beta_{0,i} + \beta_{AB,i} \mathbf{x}_{AB,j} + \beta_{age,i} \mathbf{x}_{age,j} + \epsilon_{i,j} \quad (2)$$

where  $y_{i,j}$  is the gene expression of child  $j$  for gene  $i$ ,  $\beta_{0,i}$  the intercept,  $\beta_{AB,i}$  describes the effect on gene expression of islet autoimmunity, with  $\mathbf{x}_{AB,j} = 1$  for islet autoantibody-positives and  $\mathbf{x}_{AB,j} = 0$  for islet autoantibody-negatives, and  $\beta_{age,i}$  the effect of age  $\mathbf{x}_{age,j}$  on gene expression. The residual term  $\epsilon_{i,j}$  was defined as i.i.d. normally distributed noise.

**Pathway enrichment analysis.** We performed pathway enrichment analysis based on the KEGG database<sup>12</sup>. We identified differentially regulated pathways using the "Significance Analysis of Function and Expression" (SAFE) algorithm<sup>22</sup>, R package *safe*. In this approach, p-values from local gene tests are computed and Wilcoxon rank sum statistics assess whether local statistics are systematically increased in the pathway, compared to the background (global test). Local gene tests were calculated on the residuals of models from equations (1) and (2), but excluding the term for the factor of interest (Cesarean section or islet autoimmunity). SAFE uses a

sample permutation-based approach for p-value calculation, which avoids false positive results due to correlating transcripts. The number of permutations was set to 5,000.

For node coloring of pathways (Fig. 4a,b), we used a “directed p-value”, defined as  $(-\log_{10}(\text{p-value}) * \text{sign}(\text{regression coefficient}))$ . For nodes in the pathway that had multiple genes annotated, the highest effect was shown.

**Comparing Cesarean section and islet autoimmunity results.** To compare the results obtained from the Cesarean section and islet autoimmunity (Ab+) analyses, we calculated the Pearson correlation between the standardized effect estimates of both analyses. Standardized effect estimates were represented by the t-statistic in the single gene analysis, and a “directed p-value” (analogously to previous section) in the pathway analysis. To assess the statistical significance of the correlation, we calculated

$$p_{perm} = \frac{1}{B} \sum_{i=1}^B I(\text{abs}(\rho_{i(\text{CS}_{Ab+})}) > \rho_{\text{obs}(\text{CS}_{Ab+})}) \quad (3)$$

with B being the number of permutations,  $\rho_{\text{obs}(\text{CS}_{Ab+})}$  the observed “true” correlation of effects between both analyses,  $I()$  the indicator function and  $\rho_{i(\text{CS}_{Ab+})}$  the correlation between the effects of Cesarean section and permuted islet autoimmunity status. The number of permutations was fixed at B = 5,000 for the single gene analysis and 1,000 for the pathway analyses. The same analysis with 1,000 permutations was repeated for the factors maternal diabetes, gender, and multiple first-degree relatives.

**Immune cell activation analysis.** To investigate the enrichment of genes within annotated immune genes from innateDB<sup>16</sup>, we calculated Wilcoxon rank sum tests using the t-statistics from the single gene level analysis (see above) to identify differences between the distribution of immune genes and non-immune genes. For comparison of standardized effects, we used two published datasets, one from isolated CD4+ T cells before and after activation (GEO-33272, processed data downloaded) and one containing different subsets of human leukocytes before and after activation (GEO-22886, processed data downloaded). Single gene differential analysis was performed on log<sub>2</sub>-transformed expression values using linear regression, with gene expression as the response, and activation status as the explaining variables. For SAFE pathway enrichment, the number of permutations was set to 1,000. To calculate permutation-based p-values of Pearson correlations, we used 1,000 permutations.

All statistical analyses were performed using the computing environment R version 3.3.2.

**Ethics approval and consent to participate.** The BABYDIET study was approved by the ethics committee of Ludwig-Maximilians University (Protocol No. 329/00) and is registered at ClinicalTrials.gov NCT01115621. The parents or guardians of each child provided informed consent for participation in the BABYDIET study. All samples and information were collected after obtaining signed informed consent.

## Data Availability

This publication uses publicly available gene expression datasets, available via ArrayExpress (accession number E-MTAB-1724) and the Gene Expression Omnibus (Accession Numbers GSE33272 and GSE22886). The informed consent given by study participants does not cover data posting of the clinical covariates in public databases.

## References

- Donath, M. Y. & Halban, P. A. Decreased beta-cell mass in diabetes: significance, mechanisms and therapeutic implications. *Diabetologia* **47**, 581–589, <https://doi.org/10.1007/s00125-004-1336-4> (2004).
- Krischer, J. P. *et al.* The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* **58**, 980–987, <https://doi.org/10.1007/s00125-015-3514-y> (2015).
- Giannopoulou, E. Z. *et al.* Islet autoantibody phenotypes and incidence in children at increased risk for type 1 diabetes. *Diabetologia* **58**, 2317–2323, <https://doi.org/10.1007/s00125-015-3672-y> (2015).
- Lipman, T. H. *et al.* Increasing incidence of type 1 diabetes in youth: twenty years of the Philadelphia Pediatric Diabetes Registry. *Diabetes care* **36**, 1597–1603, <https://doi.org/10.2337/dc12-0767> (2013).
- Kallionpaa, H. *et al.* Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. *Diabetes* **63**, 2402–2414, <https://doi.org/10.2337/db13-1775> (2014).
- Ferreira, R. C. *et al.* A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes* **63**, 2538–2550, <https://doi.org/10.2337/db13-1777> (2014).
- Cardwell, C. R. *et al.* Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia* **51**, 726–735, <https://doi.org/10.1007/s00125-008-0941-z> (2008).
- Bonifacio, E., Warncke, K., Winkler, C., Wallner, M. & Ziegler, A. G. Cesarean section and interferon-induced helicase gene polymorphisms combine to increase childhood type 1 diabetes risk. *Diabetes* **60**, 3300–3306, <https://doi.org/10.2337/db11-0729> (2011).
- Biasucci, G., Benenati, B., Morelli, L., Bessi, E. & Boehm, G. Cesarean delivery may affect the early biodiversity of intestinal bacteria. *The Journal of nutrition* **138**, 1796S–1800S (2008).
- Caicedo, R. A., Schanler, R. J., Li, N. & Neu, J. The developing intestinal ecosystem: implications for the neonate. *Pediatric research* **58**, 625–628, <https://doi.org/10.1203/01.PDR.0000180533.09295.84> (2005).
- Neu, J. & Rushing, J. Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. *Clinics in perinatology* **38**, 321–331, <https://doi.org/10.1016/j.clp.2011.03.008> (2011).
- Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000).
- Jiang, P., Du, W. & Wu, M. Regulation of the pentose phosphate pathway in cancer. *Protein & cell* **5**, 592–602, <https://doi.org/10.1007/s13238-014-0082-8> (2014).
- Lane, A. N. & Fan, T. W. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic acids research* **43**, 2466–2485, <https://doi.org/10.1093/nar/gkv047> (2015).
- O’Neill, L. A., Kishton, R. J. & Rathmell, J. A guide to immunometabolism for immunologists. *Nature reviews. Immunology* **16**, 553–565, <https://doi.org/10.1038/nri.2016.70> (2016).



16. Breuer, K. *et al.* InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic acids research* **41**, D1228–1233, <https://doi.org/10.1093/nar/gks1147> (2013).
17. Maliga, Z. *et al.* A genomic toolkit to investigate kinesin and myosin motor function in cells. *Nature cell biology* **15**, 325–334, <https://doi.org/10.1038/ncb2689> (2013).
18. Abbas, A. R. *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes and immunity* **6**, 319–331, <https://doi.org/10.1038/sj.gene.6364173> (2005).
19. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264, <https://doi.org/10.1093/biostatistics/4.2.249> (2003).
20. Wood, S. N. *Generalized additive models: an introduction with R*. (CRC press, 2017).
21. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 3–36 (2011).
22. Barry, W. T., Nobel, A. B. & Wright, F. A. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics* **21**, 1943–1949, <https://doi.org/10.1093/bioinformatics/bti260> (2005).

## Acknowledgements

We thank Dr. Nikola Müller and Dr. Steffen Sass for sharing their expertise on transcriptomics data. This work was supported by grants from Deutsche Forschungsgemeinschaft ZI-310/14-1, ZI-310/14-2, ZI-310/14-3, and ZI-310/14-4 (BABYDIET), by a grant from the European Union's Seventh Framework Programme [FP7-Health-F5-2012] under grant agreement 305280 (MIMOmics), and by the Helmholtz Cross-Program Initiative Personalized Medicine 'iMED'.

## Author Contributions

Study concept and design: M.L., A.G.Z., J.K. Statistical analysis: M.L. Acquisition of data: R.L., R.F., C.W., A.G.Z. Analysis and interpretation of data: M.L., E.B., J.K. Drafting of the manuscript: M.L., E.B., A.G.Z., J.K. Critical revision of the manuscript for important intellectual content: M.L., R.L., R.F., F.J.T., C.W., E.B., A.G.Z., J.K. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42750-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019