Fakultät für Medizin,
Lehrstuhl für Neurogenomik

Restless legs syndrome-associated variants contribute to *MEIS1* regulation during neurodevelopment

Ana Antić Nikolić

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Angela Krackhardt

Betreuerin: Prof. Dr. Juliane Winkelmann

Prüfer der Dissertation:

1. Prof. Dr. Thomas Meitinger
2. Prof. Dr. Bernhard Hemmer

Die Dissertation wurde am 19.03.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 20.07.2021 angenommen.

To my parents Svetlana and Slavoljub

ABSTRACT

Restless legs syndrome (RLS) is a common neurological disorder with a lifetime prevalence affecting up to 10% of the population > 65 years. It is characterized by unpleasant sensations and an urge to move the legs, primarily at rest. The symptoms are typically relieved by movement. RLS often goes undiagnosed and can strongly affect the patient's quality of life. Genome-wide study association (GWAS) identified 19 genomic risk loci for RLS. Among them, the strongest risk factor was identified within an intronic region of the *MEIS1* gene. The lead SNP, rs113851554 tags a highly evolutionarily conserved non-coding region (HNCR) designated 602 based on genomic coordinates. Another strongly associated SNP in the LD block, rs12469063, tags a distinctly highly conserved non-coding region designated as 617. The latter was shown to exhibit allele-specific enhancer activity in the developing mouse forebrain. Besides, an independent RLS risk locus is identified 1.3 Mb downstream of *MEIS1* and can potentially be involved in the *MEIS1* regulatory network. This region is tagged by rs1820987. The aim of this thesis was to profile the regulatory landscape of *MEIS1* in neurodevelopment, to identify how RLS-associated risk variants could exert their regulatory function, and then functionally validate their regulatory effect on *MEIS1*.

To address the regulatory potential of RLS-associated risk variants, we profiled DNA accessibility by the ATAC-seq method using *in vitro* generated neural cell types at different stages of maturation. We analysed publicly available datasets of relevant tissues and cell types derived from humans and developing mice. To elucidate the spatial organization of the *MEIS1* locus, circular chromosomal conformation capture followed by sequencing (4C-seq) was performed in three different cell types using the *MEIS1* promoter as a viewpoint. To investigate the direction of the enhancer effect, we employed CRISPR-Cas9 assisted deletion of HCNR 617 and differentiation of pluripotent stem cell into ganglionic eminences-like progenitors to explore stage-specific regulatory activity.

ATAC-seq results showed a distinct accessibility pattern at the *MEIS1* locus during *in vitro* neurodevelopment. HNCR 602 was only accessible in inhibitory neurons and public datasets confirmed activity of this element specifically in neurons. Moreover, human lateral ganglionic eminences and putamen were enriched for open chromatin at HCNR 602. HCNR 617 exhibited regulatory features in neural cell types including progenitor stage and more

mature neural stages. In selected public datasets derived from the developing mouse, both enhancers were active predominately during forebrain development. Besides, we explored the chromatin signature at the secondary RLS risk locus in the vicinity of the *MEIS1* and identified regions of accessibility that could potentially carry a causal risk variant. Furthermore, chromosome conformation capture results showed a direct interaction between *MEIS1* promoter and HCNR 617 in human neural stem cells. Moreover, in glioblastoma cell line there was a long-range interaction between the *MEIS1* promoter, and the RLS risk locus located 1.3 Mb downstream of *MEIS1* which harbours the risk variant rs1820987. Finally, targeted deletion of regulatory element 617 confirmed an enhancer effect on *MEIS1*.

In summary, we prioritized cell types where RLS-associated risk variants exhibit features of regulation. We have found cell-specific enhancer-promoter contacts linked to RLS risk loci and confirmed an enhancer effect by genome editing. These experiments provide further evidence for *MEIS1* as a causal RLS gene, and further implicate the developing nervous system as an important region for the *MEIS1* effect on RLS risk, furthering our understanding of RLS etiology.

ZUSAMMENFASSUNG

Das Restless-Legs-Syndrom (RLS) ist eine häufige neurologische Erkrankung, mit einer altersabhängigen Prävalenz bis zu 10 % der älteren Bevölkerung. Es ist durch unangenehme Mißempfindungen und einen Bewegunsdrang in den Beinen, vor allem nachts, und in Ruhe gekennzeichnet. Die Symptome werden typischerweise durch Bewegung gelindert. In genomweiten Assoziationsstudien (GWAS) wurden 19 genomische Risikoloci für das RLS identifiziert. Unter ihnen wurde der stärkste Risikofaktor innerhalb einer intronischen Region des MEIS1-Gens identifiziert. Der SNP mit dem stärksten Effekt, rs113851554, markiert eine hochgradig evolutionär konservierte nicht-kodierende Region (HNCR 602). Ein weiterer stark assoziierter SNP im LD-Block, rs12469063, markiert ebenfalls eine hoch konservierte nicht-kodierende Region (HCNR 617). In vorherigen Untersuchungen konnte gezeigt werden, dass diese Region eine allelspezifische Enhancer-Aktivität im sich entwickelnden Vorderhirn der Maus hat. Außerdem wurde ein unabhängiger RLS-Risikolocus 1,3 Mb stromabwärts des *MEIS1*-Gens (rs1820987) identifiziert, der möglicherweise im *MEIS1*-Regulationsnetzwerk eingebunden ist. Das Ziel dieser Arbeit war es, die Regulation von *MEIS1* in der neuronalen Entwicklung zu analysieren und den Effekt der RLS-assoziierte Risikovarianten auf die regulatorische Funktion von *MEIS1* funktionell zu validieren.

Um das regulatorische Potenzial von RLS-assoziierten Risikovarianten zu untersuchen, haben wir die DNA-Zugänglichkeit mit der ATAC-seq-Methode unter Verwendung von in vitro generierten neuronalen Zelltypen in verschiedenen Reifungsstadien profiliert. Darüber hinaus analysierten wir öffentlich verfügbare Datensätze relevanter Gewebe und Zelltypen, die von Menschen und Mausembryonen stammen. Um die räumliche Organisation des *MEIS1*-Lokus aufzuklären, wurde zirkuläres Chromosomen-Konformations-Capture gefolgt von Sequenzierung (4C-seq) in drei verschiedenen Zelltypen durchgeführt, wobei der *MEIS1*-Promotor als Ausgangspunkt diente. Um schließlich die Richtung des Enhancer-Effekts zu untersuchen, setzten wir die CRISPR-Cas9-assistierte Deletion von HCNR 617 und die Differenzierung von pluripotenten Stammzellen in Ganglion-ähnliche Progenitoren ein, um die stadienspezifische regulatorische Aktivität zu untersuchen.

ATAC-seq-Ergebnisse zeigten ein ausgeprägtes Muster der DNA-Zugänglichkeit am *MEIS1*-Lokus während der neuronalen Entwicklung *in vitro*. HNCR 602 war nur in inhibitorischen Neuronen zugänglich, öffentliche Datensätze bestätigten die Aktivität dieses Elements speziell in Neuronen. Darüber hinaus konnte in humanen lateralen ganglionären Emminenzen und Putamen bei HCNR 602 offenes Chromatin detektiert werden. HCNR 617 zeigte regulatorische Merkmale in neuralen Zelltypen, einschließlich des Vorläuferstadiums und reiferer neuronaler Stadien. In ausgewählten öffentlichen Datensätzen, die von der sich entwickelnden Maus stammen, waren beide Enhancer überwiegend während der Entwicklung des Vorderhirns aktiv. Schließlich untersuchten wir die Chromatinsignatur am sekundären RLS-*MEIS1*-Risikolokus und fanden Regionen der Zugänglichkeit, die möglicherweise eine kausale Risikovariante tragen könnten. Darüber hinaus zeigten die Ergebnisse der Chromosomenkonformationserfassung eine direkte Interaktion zwischen dem *MEIS1*-Promotor und HCNR 617 in menschlichen neuralen Stammzellen. Außerdem wurde in der Glioblastom-Zelllinie eine Interaktion zwischen dem *MEIS1*-Promotor und dem RLS-Risikolokus, der sich 1,3 Mb stromabwärts von *MEIS1* befindet, festgestellt. Schließlich bestätigte die Deletion des regulatorischen Elements 617 einen Enhancer-Effekt auf *MEIS1*.

Zusammenfassend haben wir Zelltypen priorisiert, in denen RLS-assoziierte Risikovarianten regulatorisch wirken. Wir haben zellspezifische Enhancer-Promotor-Kontakte identifiziert, die mit RLS-Risikoloci verbunden sind, und den Enhancer-Effekt durch Genom-Editierung bestätigt. Diese Experimente bestätigen *MEIS1* als kausale RLS-Gen und implizieren dass das sich entwickelnde Nervensystem eine wichtige Region für den ″*MEIS1*-Effekt″ darstellt.

ABBREVIATIONS

4C-seq	Circular chromosome conformation capture, coupled to high-throughput sequencing
5'UTR	5' untranslated region
AD	Alzheimer's disease
ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput sequencing
ATP	Adenosine 5'-triphosphate
BBB	Blood-brain-barrier
CAD	Coronary artery disease
Cas9	CRISPR associated protein 9
ChIP-seq	Chromatin immunoprecipitation sequencing
CSF	Cerebrospinal fluid
CRBN	Cereblon
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
D2R	Dopamine receptor D2
DMEM/F12	Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12
DNase-seq	DNase I hypersensitive site sequencing
DPBS	Dulbecco's phosphate-buffered saline
DTT	Dithiothreitol
DZ	Dizygotic
E	Embryonic day
EC	Embryonic carcinoma
EC	Endothelial cell
EDTA	Ethylenediamine tetraacetic acid
EMSA	Electrophoretic mobility shift assay
EP300	E1A-associated protein p300
ESRD	End-stage renal disease
eQTL	Expression quantitative trait locus
FTL	Light chain ferritin

FTH	Heavy chain ferritin
GE	Ganglionic eminences
GE-like progenitors	Ganglionic eminences-like progenitors
GWAS	Genome wide association study
H3K27ac	Acetylation of the lysine residue at N-terminal position 27 of the histone H3 protein
H9 NSC	H9 derived neural stem cell
HCNR	Highly conserved non-coding region
hESC	human embryonic stem cell
iPSC	Induced pluripotent stem cell
IRLSSG	International Restless legs syndrome study group
KO	Knockout
LD	Linkage disequlibrium
LCL	Lymphoblastoid cell line
LGE	Lateral ganglionic eminence
M	Mole (Unit)
MAF	Minor allele frequency
$MgCl_2$	Magnesium chloride
MGE	Medial ganglionic eminence
mRNA	Messenger ribonucleic acid
MZ	Monozygotic
NaCl	Sodium chloride
NIM	Neural induction media
NSC	Neural stem cell
OCR	Open chromatin region
PCR	Polymerase chain reaction
PLMS	Periodic limb movements in sleep
qPCR	quantitative polymerase chain reaction
RLS	Restless leg syndrome
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROCK	Rho-associated protein kinase inhibitor
SFM	Serum- free media

SNP	Single nucleotide polymorphism
TF	Transcription factor
TH	tyrosine hydroxylase
Tris	Tris(hydroxymethyl)aminomethane
Tris-HCl	Tris-Hydrochlorid
TSS	Transcriptional start site
VP	Viewpoint
VSMC	Vascular smooth muscle cells

Table of Contents

INTRODUCTION

1.1 Restless legs syndrome

Restless legs syndrome (RLS) is a neurological sensorimotor disorder characterized by patient complaints of an urge to move the legs (or arms), usually accompanied by uncomfortable sensations in the affected limbs (Bogan and Cheray 2013). The earliest clinical description of RLS was given by a physician, Sir Thomas Willis in the 17th century (Willis and Eugenius Philiatros. 1977). Later on, a Swedish neurologist Karl-Axel Ekbom provided a detailed description and also termed the disorder RLS (Ekbom 1945).

The main clinical symptom of RLS is a compelling urge to move the legs, which occurs either by itself or accompanied by uncomfortable paraesthesia of the legs. Patients feel sensations such as crawling, itching, and burning, mainly in the upper calves (Karroum, Leu-Semenescu, and Arnulf 2012). Moving the legs or walking improves the urge. Symptoms generally develop in the evening or at night and might progressively worsen during the time but resolve by the early hours of the morning. The symptoms occur after a duration of relative inactivity (Wijemanne and Ondo 2017).

At the moment, there are no biomarkers for RLS, and clinical diagnosis is based on a detailed interview with a patient. Diagnostic criteria are initially defined by the International Restless Legs Syndrome Study Group (IRLSSG) in 1995 (Walters 1995) which were further updated in 2003 (Allen et al. 2003) and in 2014 (Allen et al. 2014). These diagnostic criteria include:

1. An urge to move the legs usually accompanied by uncomfortable and unpleasant sensations in the legs.

2. The urge to move the legs and any accompanying unpleasant sensations begin or worsen during periods of rest or inactivity for example resting or sitting.

3. The urge to move the legs and any accompanying unpleasant sensations are partially or totally relieved by movement, such as walking or stretching, at least as long as the activity continues.

4. The urge to move the legs and any accompanying unpleasant sensations during rest or inactivity only occur or are worse in the evening or night than during the night.

5. The existence of the above-mentioned symptoms is not merely accounted for as symptoms primary to another medical or behavioural condition (e.g. myalgia, venous stasis, leg oedema, arthritis, leg cramps, positional discomfort, habitual foot tapping).

Supportive diagnostic criteria are a positive family history of RLS, positive therapeutic response to dopaminergic drugs, and periodic limb movements in sleep (PLMS) (Allen et al. 2014). The fifth criterion was included in 2014 and denotes the importance of a detailed clinical assessment in order to exclude other possible conditions. Certain muscular abnormalities can be initially mistaken for RLS such as essential myoclonus, myokymia, and orthostatic tremor. Furthermore, diseases characterized by nocturnal leg discomfort and/or pain such as radiculopathies, neuropathies, vascular disorders, myalgias, arthralgias can be misidentified as RLS (Allen et al. 2014).

Several comprehensive reviews on the epidemiology of RLS showed that the disorder received significant attention in the past 25 years. There are more than 50 epidemiological studies conducted in different populations. Overall, RLS prevalence is higher in North America and Europe compared to Asia (Koo et al. 2015). RLS is two times more common in women compared to men across all populations and ages (Ohayon et al. 2012).

Two forms of the disease have been discussed, based on the age of onset. The early-onset form is considered when symptoms start at age 45 or earlier. This form is more common within families, and it has a slower progression of symptoms. The late-onset type starts after the age of 45 with rapid symptom progression, and it occurs less commonly in families (Allen and Earley 2000).

Formerly, RLS was considered primary or idiopathic (when no apparent cause is seen) and secondary or symptomatic (associated with some other disease). Numerous studies done over time observed the association of RLS with iron deficiency, low serum ferritin values, pregnancy, end-stage renal disease (ESRD), and with neurological disorders such as polyneuropathy (Garcia-Borreguero et al. 2011). However, a study from 2016 discovered an increased prevalence of RLS only in iron deficiency and kidney disease. Moreover, the authors reported that there is insufficient evidence for conditions such as anaemia (without iron deficiency), chronic obstructive pulmonary disease, multiple sclerosis, headache, stroke,

narcolepsy, and ataxias (Trenkwalder et al. 2016). Pathophysiological mechanisms involved are not entirely understood, however, there is evidence for the role of iron and dopamine. It is suggested that the role of iron in the RLS pathology is more relevant in the nervous system than in blood and that an impaired brain iron acquisition plays a role in RLS (Allen et al. 2015). A reduced concentration of ferritin and increased concentration of transferrin were detected in cerebrospinal fluid (CSF) of RLS patients (Earley et al. 2000; Mizuno et al. 2005). Furthermore, neuroimaging studies reported an iron deficiency in RLS patients (Allen et al. 2001; Godau et al. 2008). Brain iron deficiency in RLS patients was also demonstrated by neuropathological examination and immunohistochemistry of post-mortem brain tissue (Connor et al. 2003; 2004). Taken together, studies using different approaches suggested an impaired brain iron acquisition in RLS.

Furthermore, dopamine dysfunction was also recognized. The neuropathological studies showed a decrease in D2R (Dopamine receptor D_2) expression in the putamen in RLS patients, correlated with increased severity of RLS symptoms. Additionally, an increase in tyrosine hydroxylase (TH) activity in the substantia nigra and putamen was found (Connor et al. 2009). On that note, it is important to mention that iron acts as a cofactor to enzyme tyrosine hydroxylase (TH) in the conversion of tyrosine to L-Dopa. In the next step, L-dopa is converted to dopamine. This demonstrates that iron affects the biosynthesis of dopamine and highlights the interplay between iron and dopamine in RLS. One of the evidences for the role of dopaminergic pathology in RLS comes from positive pharmacological response to the treatment with dopaminergic agonists (Allen et al. 2004). Even though there is evidence for the existence of a hyperdopaminergic state in RLS, dopaminergic agents are used for treatment. Medication with dopaminergic agonist relieves the symptoms. However, effectiveness does not last very long. Since dopaminergic system is continuously supplemented with agonist, receptors become downregulated meaning that the number of receptors on the cell surface decreases. This downregulation of dopamine receptors relates to probable malfunctioning dopaminergic activity at night and thus gives the RLS symptoms in the night-time (Khan et al. 2017). This dopamine dysfunction could contribute to augmentation, a frequently seen phenomenon as worsening of the symptoms, even occurrence earlier in the daytime (Khan et al. 2017).

There are evidence-based published guidelines for the management of RLS. For each RLS patient, the treatment should be carefully tailored. For the first-line treatment, current

guidelines recommend non-ergoline dopamine agonists such as pramipexole, ropinirole, and rotigotine. It is important to carefully adjust the dosage to avoid the undesired effect of augmentation (Garcia-Borreguero et al. 2016). Levodopa was the first dopaminergic agonist used for RLS treatment, but now it is recommended only for intermittent usage, and not daily due to the high risk of augmentation (Garcia-Borreguero et al. 2016). One study reported that 70 % of the patients treated with levodopa, experienced augmentation (Allen and Earley 1996). Benzodiazepines are occasionally used for RLS, but they rather improve sleep quality than treating the RLS symptoms (Silber et al. 2013). For the treatment of more severe cases, treatment guidelines recommend medication with alpha-2-delta ligands such as pregabalin, gabapentin, and a prodrug gabapentin enacarbil. This medication is the treatment of choice if the sleep disturbance and insomnia develop or persist (Garcia-Borreguero et al. 2013). Finally, a low dose opioid such as methadone, oxycodone plus naloxone, tramadol, codeine should be reserved for the treatment of RLS in case of refractory symptoms who do not respond well to the first line medication or in case of development of augmentation. Opioid therapy, however, should be undertaken with precaution due to the risks of overdose and addiction (Silber et al. 2018). In all patients diagnosed with RLS, iron status should be examined. Evidence-based guidelines from 2018 provide recommendations for iron supplementation. If the serum ferritin level is less than 75 μg/L, oral iron supplementation such as ferrous sulphate plus vitamin C should be prescribed. In case of intolerance of oral iron or the presence of malabsorption, iron treatment should be administered intravenously (Silber et al. 2013; Allen et al. 2018). The genetic study from 2017 found an association between RLS and gene *CRBN* that encodes cereblon, a protein with a role in proteolysis and potassium channel regulation (Fischer et al. 2014; Schormair et al. 2017). An immunomodulatory drug thalidomide binds cereblon and therefore can be considered as a treatment option for RLS in specific patients (Salminen et al. 2018).

As mentioned before, RLS is found to be associated with pregnancy. A meta-analysis from 2017 showed that overall RLS prevalence in pregnancy is 21% across all three trimesters (Chen et al. 2018).

RLS is also observed in the paediatric population, affecting 2–4% of school-aged children and adolescents. However, RLS remains frequently undiagnosed in children simply because children may describe the symptoms adequately (Picchietti et al. 2013).

1.2 RLS genetics studies

The familial occurrence, seen in patients, led to many genetic studies in this field and finally gave the evidence for RLS to be considered as a common complex genetic disease (Trenkwalder et al. 2018). A swedish neurologist Ekbom gave a comprehensive description of the disease and observed heredity since in a group of 33 studied patients, 11 of them stated to have close relatives having the same symptoms (Ekbom 1945). Subsequently, several studies reported a positive family history among patients. One study that involved 54 patients who fulfilled diagnostic criteria for the RLS, found positive family history for RLS in 23 of 25 idiopathic cases (92.0%) (Ondo and Jankovic 1996). In one research that focused on polysomnographic and genetic features of RLS, patients were surveyed for family history and 63 % reported at least one first-degree relative affected with RLS (Montplaisir et al. 1997). Another study involved 300 patients diagnosed according to the diagnostic criteria of IRLSSG. The familial cases are defined if at least one first degree relative was also examined and confirmed as affected by the authors of the study. It was found that 42.3 % of the idiopathic RLS cases had a definite positive history, whereas, only 11.7% of the secondary (uremic) RLS patients. This study also detected that patients with definite hereditary RLS were significantly younger at the age of onset than those with a negative family history (Winkelmann et al. 2000).

The genetic component in RLS is further supported by studies conducted in identical twins. Twin studies represent a very valuable tool to study the contribution of environmental and genetic components to different traits. The authors of one study investigated RLS in 12 monozygotic twin pairs and found that 10 pairs (83 %) were concordant for RLS (Ondo et al. 2000). Another study used twin design to study genetic aspects of RLS involved 933 monozygotic (MZ) and 1004 dizygotic (DZ) twin pairs. The concordance rate reported in MZ and DZ are 61 % and 45 % respectively and the heritability rate is 54 % (Desai et al. 2004). It should be noted that this study used only female probands and only two questions survey was used to detect the presence of RLS.

Initially, it was suggested that the disease follows a Mendelian inheritance, and linkage studies have been performed. One study explored a large French-Canadian family and mapped the first RLS locus (RLS1) on the short arm of chromosome 12 (Desautels et al. 2001). The second locus, (RLS2) on chromosome 14 was discovered in a large North Italian family (Bonati et al. 2003). Subsequently, additional RLS-loci were discovered on

chromosome 9, 2 20, 19 and 16 respectively (Liebetanz et al. 2006; Chen et al. 2004; Levchenko et al. 2006; Kemlink et al. 2008; Levchenko et al. 2009). With an exception to RLS1 that shows an autosomal recessive mode of heritability, other loci have an autosomal dominant inheritance pattern. However, these studies failed to pinpoint the causal gene due to the genetic heterogeneity of the disorder.

RLS is now being considered as a complex disease with an environmental and genetic component. Genome-wide association studies (GWAS) are a robust approach for unravelling the genetic factors contributing to complex diseases. The study design uses a hypothesis-free approach to study individuals with some specific phenotype and their matched controls. Genetic association studies search for a statistical association between some trait and genetic variation to pinpoint genomic regions that contribute to a specific disease (Lewis and Knight 2012). The genetic variation tested in GWAS is normally occurring DNA variation among individuals at the level of a single nucleotide named single nucleotide polymorphism (SNP). It has been estimated that these polymorphisms account for around 90% of genetic variation in the human genome (Collins et al. 1998). Thus, the SNPs serve as biomarkers for GWAS studies. Thanks to the development of high-throughput technology, hundreds of thousands of SNPs can be genotyped simultaneously. Precisely defined and accessed phenotype and adequate matched controls are some basic requirements for a successful GWA study. After stringent quality controls, by applying multidimensional complex statistical analysis, the allele frequency between cases and controls is tested for an association. It should be noted that these studies require a large sample size to have enough statistical power to detect a disease-associated variant. However, most detected disease-associated variants contribute with modest effect size to the phenotype. One of the reasons for that is that genotyping arrays used in GWAS screen the polymorphisms with minor allele frequency (MAF) above 1%. The common variants detected by GWAS explain only a small portion of heritability as shown in Figure 1.1 (Manolio et al. 2009). That means that only a small proportion of phenotypic variation can be explained due to genetic variance which contrasts with heritability estimates obtained from twin and family studies (Maher 2008). This missing heritability may be potentially explained with low-frequency ($0.5\% \leq MAF < 5\%$) and rare ($MAF < 0.5\%$) variants having additionally contributed to the risk phenotype (Lee et al. 2014). These low frequency and rare variants remain undetected by commercial arrays as they do not cover all variants in the genome. However, these rare variants can be imputed based on linkage disequilibrium (LD) and their role in complex diseases can be addressed. Of course, the costs,

large sample sizes to adequately access the rare variants, very complex statistical analysis, and interpretation make a GWAS very challenging task.

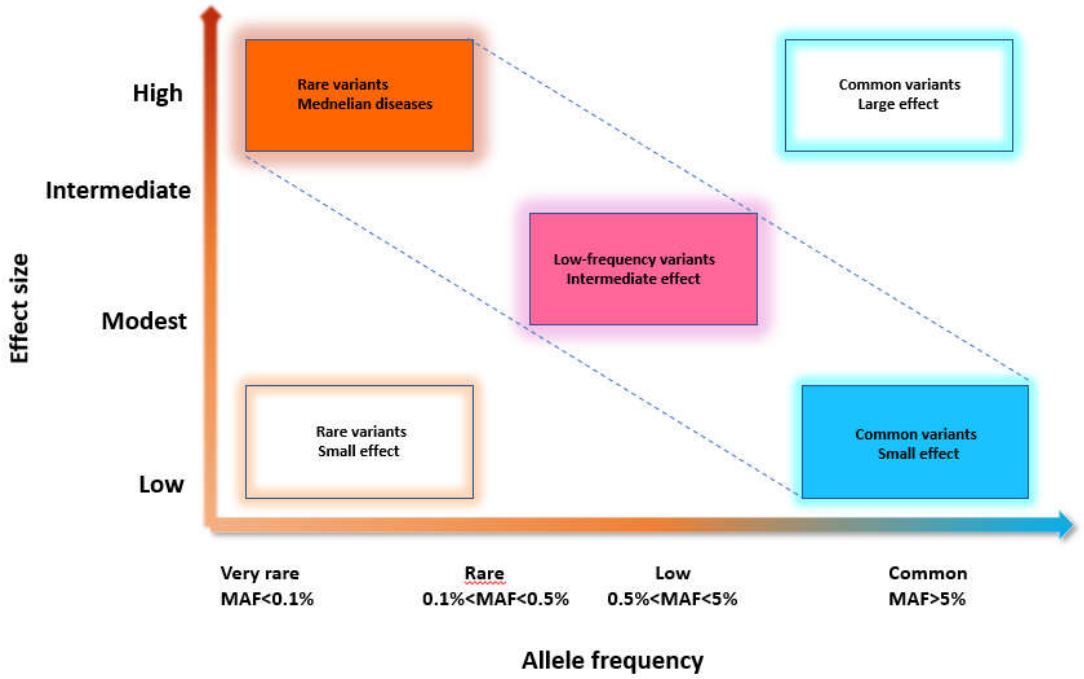


Figure 1.1: Genetic variants described based on allele frequency (x-axis) and effect size (y-axis) (adapted from Manolio et al. 2009)

Until today, there are around 4000 human GWA studies done on 2000 different traits (Buniello et al. 2019). Up to date, there have been three large GWA studies on RLS plus one GWAS meta-analysis. Also, one GWA study was conducted with RLS patients having periodic limb in sleep (PMLS) (Stefansson et al. 2007). The first study on RLS published in 2007 identified three RLS risk loci. The study involved 401 RLS German cases and 1644 controls in the exploratory phase and the same regions were confirmed in a replication phase that involved additional 1158 German and Canadian cases and 1178 controls (Winkelmann et al. 2007). The first candidate region spanned 32kb length in intron eight, exon nine, and intron nine of *MEIS1* located on chromosome 2. The risk haplotype was fully tagged by two SNPs, rs6710341 and rs12469063 who gave the strongest signal for RLS in a study from 2007 (Winkelmann et al. 2007). The second risk locus is within 113kb linkage disequilibrium block (LD) located within intron 5 of the *BTBD9* gene. The third genomic locus associated with RLS is mapped on chromosome 15q within a 48 kb LD block spanning the 3' end of the *MAP2K5* gene and the neighbouring *SKOR1* gene (Winkelmann et al. 2007).

These loci were again confirmed in a replication study from 2008 that involved 649 RLS patients and 1230 matched controls from the Czech Republic, Austria, and Finland (Kemlink et al. 2009). The same year a fourth risk locus was discovered. In this locus, there were two independent association signals found, in intron 8 and intron 10 of the *PTPRD* gene on chromosome 9. This locus was previously identified in the linkage studies as RLS3. The study involved 2,458 affected individuals and 4,749 controls from Germany, Austria, Czech Republic, and Canada. Also, the authors performed exome- sequencing in a subset of patients from RLS3- linked families, but no mutations were found (Schormair et al. 2008). This implicates the involvement of non-coding elements in the RLS pathology. Another RLS GWAS from 2011, involved 954 German RLS cases and 1,814 controls and replicated previously reported four loci, but also discovered two novel loci: an intergenic region on chromosome 2p14 (1.3 Mb downstream of *MEIS1*) and 5' UTR of *TOX3* gene and the adjacent non-coding *RNA BC034767* (Winkelmann et al. 2011). An RLS meta-analysis of genome-wide association studies in 2017, encompassed 15 126 cases and 95 725 controls of European ancestry. The authors discovered 13 new risk loci and confirmed previously reported 6 loci (Figure 1.2). The *MEIS1* gene on chromosome 2 was confirmed as the strongest genetic risk factor for RLS. The lead SNP rs113851554, located within intron 8 of the *MEIS1*, is a low-frequency variant, with odds ratio (OR) 1·82–2·16 (Schormair et al. 2017). Furthermore, the *MEIS1* locus was identified in GWA studies of insomnia disorder (Lane et al. 2017; Hammerschlag et al. 2017).

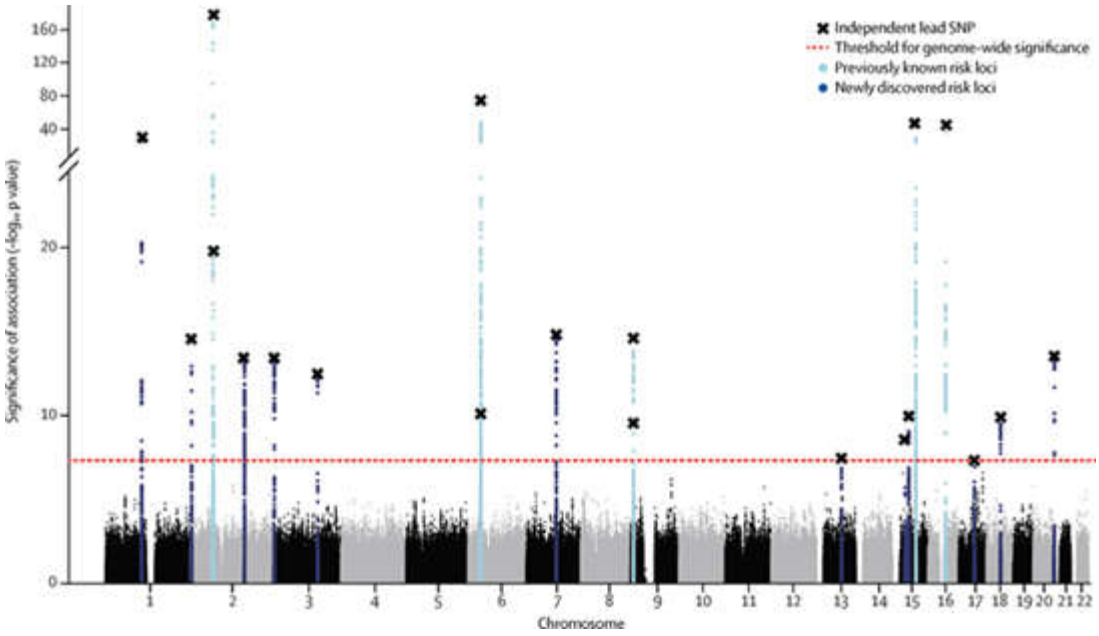


Figure 1.2: Manhattan plot of meta GWAS for RLS (Schormair 2017)

The authors performed an extensive analysis of candidate genes to understand and explain the complex phenotype displayed in RLS. A pathway enrichment analysis implies that candidate genes are related to neurodevelopment, specifically to neurogenesis, axon guidance, synaptogenesis, etc. Moreover, the results underline the importance of locomotion and DNA repairing mechanisms. By applying the LD regression score, the authors calculated that 19.6% heritability can be explained with SNPs from this dataset (Schormair et al. 2017). GWA studies are a powerful instrument to explore the role of the genes in complex diseases and to predict the risk allowing thus prevention and adequate treatment for the patients.

1.3 Non-coding sequences and gene regulation

Up to date, genome-wide association studies (GWASs) mapped thousands of loci associated with different phenotypes. Causal variants within identified regions are difficult to detect due to complex LD structure (Freedman et al. 2011; Lin et al. 2018). The variants are mostly found in non-coding regions and only a small portion falls within protein-coding sequences. Nearly 90% of the genetic variants identified by GWAS reside within non-coding sequences such as promoters, enhancers, silencers, insulators, long coding RNA, miRNAs (Mirza et al. 2014; Giral et al. 2018). Furthermore, the GWAS association signal also does not specify which gene or genes are affected by the causal variant (Gallagher and Chen-Plotkin 2018). Therefore, the functional interpretation of non-coding genetic variants poses a great challenge. There is emerging evidence that variants within non-coding regions most probably contribute to gene expression regulation. Gene expression is precisely controlled in a spatio-temporal manner. These changes in gene expression can affect the cellular structure and/or function and thereby contribute to a specific trait. The regulatory element is a stretch of DNA sequence located in an intronic or intergenic region that contains multiple binding sites for transcription factors (TF) and can modulate gene expression from the transcriptional start site (TSS). Such domains show certain features of functionality: high degree of conservation through species; epigenetic histone modifications like acetylation and methylation; DNA methylation. These elements can upregulate gene expression and then are called enhancers or downregulate gene expression and, in that case, they are termed silencers. A single base pair substitution can affect affinity for transcription factor (TF) binding and consequently change gene expression as depicted in Figure 1.3. For example, a risk allele G of variant rs10757278 associated with cardiovascular disease, located within the enhancer element on chromosome 9, disrupts the binding site for transcription factor STAT1. In return, STAT1 fails to recruit

silencing machinery, and a long coding RNA named *CDKN2B-AS* becomes upregulated (Harismendy et al. 2011). The functional consequence of this upregulation is the promotion of atherosclerosis through plaque formation (Ou et al. 2020). Another study extensively explored the functional role of risk variants associated with type 1 diabetes (T1D). By employing epigenetic profiling in T cells derived from patients and healthy controls, the authors unravelled T1D risk variants within putative regulatory elements modulating the expression of immune genes. A variant rs883868 located on chromosome 21 in high LD with the lead T1D GWAS SNP rs11203203, was found to allele specifically modulate the binding YY1. Yin and Yang 1 (YY1) is a transcriptional factor known as a structural regulator of three-dimensional DNA loops. The loss of YY1 binding subsequently leads to a loss of enhancer-promoter contact and lower expression of the *UBASH3A* gene (Gao et al. 2019). Another striking example of the regulatory function of a single SNP comes from GWAS follow-up study in prostate cancer (PCa). A single nucleotide polymorphism rs11672691 at 19q13 is associated with poor prognosis and survival (Bradley et al. 2018). In a series of experiments, the authors discovered that rs11672691 possesses a feature of bifunctionality, acting as a promoter and enhancer. This PCa associated SNP regulates the expression of two isoforms of lncRNA *PCAT19*. Specifically, rs11672691 is located in the promoter region of short isoform *PCAT19* and within the enhancer of the long isoform PCAT19. It is proposed that risk allele, by disrupting the binding of NKX3 in the promoter of the short isoform, switches the enhancer activity, and upregulates the expression of the long variant lncRNA *PCAT19*, which in returns promotes cancer growth (Hua et al. 2018). Moreover, variants discovered in GWAS for Parkinson's disease were examined for their functionality in disease pathogenesis. More precisely, a common variant rs356168 located within the regulatory element modulates the expression of the α-synuclein (*SNCA)* gene. The electrophoretic mobility shift assay analysis (EMSA) demonstrated allele-specific binding EMX2 and NKX6-1 with a preference for the non-risk allele. Next, by using the CRISPR-Cas9 editing system the authors engineered the induced pluripotent stem cells carrying the risk allele and differentiated the cells into neural precursors and neurons. In both differentiated cell types, the expression level of the *SNCA* gene was significantly higher compared to control lines. The role of *SNCA* in Parkinson's disease is well established- this gene encodes for a protein found in Lewy bodies which are the primary pathophysiological substrate in Parkinson's disease (Soldner et al. 2016). Overall, the interpretation of non-coding variants remains challenging, but substantial progress has been made.

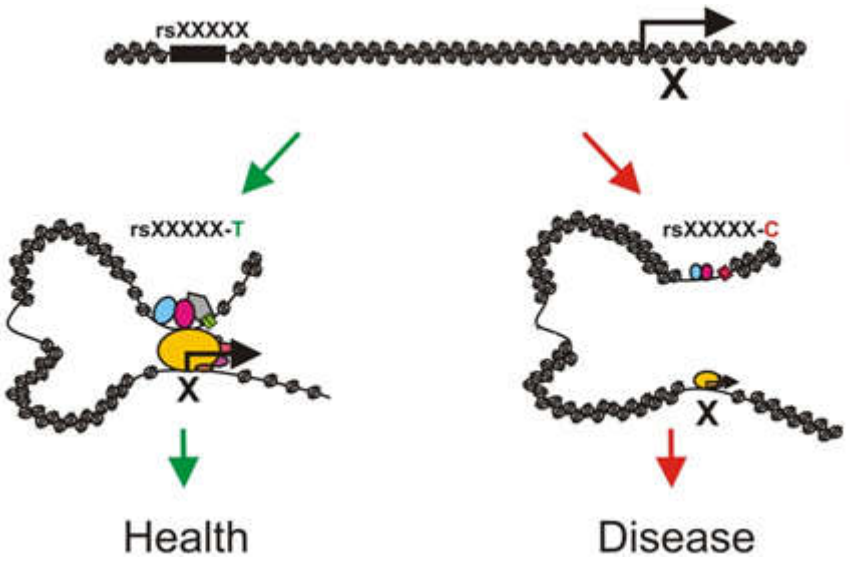


Figure 1.3: Model depicting how sequence variation in distal regulatory elements might influence phenotypes or disease states (Palstra et al., 2012)

1.4 *MEIS1* and its implication in neural development and RLS association

As mentioned above, *MEIS1* was one of the first RLS risk loci discovered and replicated in three subsequent studies. Today it is considered the strongest genetic risk factor for RLS. This gene is the only candidate within the locus, surrounded by two large intergenic regions. Thus, substantial effort has been made in the past years to understand the role of the *MEIS1* in RLS pathology. *MEIS1* (myeloid ectopic viral insertion site 1) is a highly evolutionarily conserved gene encoding a transcription factor. It is located on chromosome 2 in humans and chromosome 11 in mice. This gene was discovered and characterized as a common site for viral integration in a mouse model for leukaemia (Moskow et al. 1995). *MEIS1* exhibits highly complex expression patterns regulated by equally complex regulatory elements. A comparative genomic analysis showed an enrichment of highly conserved noncoding regions (HCNR) within the *MEIS1* locus (Royo et al. 2012). In humans and mice, there are three *MEIS* homologs *MEIS1, MEIS2,* and *MEIS3*, which belong to the TALE-superfamily (Three Amino-acid Loop Extension) of homeobox (HOX) genes. The *MEIS* gene family shares a high level of DNA and protein sequence identity (Schulte and Geerts 2019). Alternative splicing has been identified in the *MEIS* gene family, further extending its functional repertoire. The two main isoforms of the *MEIS1* are the *MEIS1A* isoform that contains exon 12 and the *MEIS1B* isoform without exon 12 which encodes an alternative C-terminus

(Sánchez-Guardado et al. 2011). MEIS1 is highly pleiotropic, with central roles in limb, vascular, cardiac, and neural development. *MEIS1* is essential for the maintenance of normal haematopoiesis in the bone marrow. *Meis1* knockout experiments demonstrated impaired haematopoiesis through a loss of hematopoietic stem cells, disrupted cell cycle, and self-renewal (Ariki et al. 2014). *Meis1* homozygous knockout mice exhibit abnormal vascular formation, heart anomalies, and die between embryonic days 11.5 and 14.5 due to a general failure of haematopoiesis (Azcoitia et al. 2005). Moreover, the oncogenicity of the *MEIS1* is well established. *MEIS1* is upregulated in acute myeloid leukaemia (Lawrence et al. 1999). In patients diagnosed with acute myeloid leukaemia, high MEIS1 expression levels were associated with poor response to chemotherapy (Liu et al. 2017). One study identified three regulatory regions of *MEIS1* that could potentially have a role in leukemogenesis (Xiang et al. 2014).

Moreover, it has been shown that Meis1 has a role in correct patterning of the proximodistal limb axis during development in mouse, chicken, and *Drosophila* (Mercader et al. 1999). Meis1 overexpression in developing mouse altered proximodistal limb patterning by disrupting distal limb formation (Mercader et al. 2009). On the other hand, *Meis1* and *Meis2* deficiency in developing limbs led to phocomelia, a condition characterized by absent or severely hypoplastic proximal limbs (Delgado et al. 2020).

A role for MEIS1 in neural development has been indicated in several species. The *Xenopus* homolog of MEIS1 (Xmeis1) is involved in neural crest development during embryogenesis (Maeda et al. 2001). In the zebrafish visual system, *meis1* has a role in retinotectal mapping establishment (Erickson et al. 2010). Furthermore, intact meis function is necessary for proper hindbrain development (Waskiewicz et al. 2001). There is evidence that Meis1 is implicated in striatal development and sympathetic neuron differentiation in mice (Rataj-Baniowska et al. 2015; Bouilloux et al. 2016). It has been suggested that *MEIS* genes participate in neural differentiation through chromatin accessibility remodelling (Hau et al. 2017). An *in vitro* study using the pluripotent P19 embryonal carcinoma (EC) line showed that direct communication of Meis1 and Oct4 is necessary for neural differentiation (Yamada et al. 2013).

Two studies investigated MEIS1 expression in RLS patients compared to controls. In the first, the authors observed a significant decrease in *MEIS1* mRNA as well as protein level in lymphoblastoid cell lines (LCLs) and brain tissues (pons and thalamus) from RLS patients

homozygous for the intronic RLS risk haplotype (GG for rs1246906 and GG for rs2300478), compared with those homozygous for the non-risk haplotype (Xiong et al. 2009). These findings suggested that reduced *MEIS1* expression could contribute to RLS symptomatology. In a subsequent study, the same group searched for a connection between MEIS1 and iron metabolism in humans by measuring the expression of ferritin in lymphoblastoid lines and brain tissues (pons and thalamus) of RLS patients. In the thalamus and lymphoblastoid cell line the light chain ferritin (FTL) and heavy chain ferritin (FTH) mRNA and protein were significantly increased in homozygous risk haplotype carriers who also had reduced expression of *MEIS1*. Moreover, knockdown of the *MEIS1* orthologue *Unc-62* in *Caenorhabditis elegans* significantly increased ferritin expression, which implies a role of *Unc-62* in iron metabolism regulation. These results indicate that low *MEIS1* expression leads to increased expression of ferritin, suggesting dysfunction of iron transportation (Catoire et al. 2011). A more recent study discovered reduced expression of the RLS-associated *SKOR1* gene in brain specimens of patients carrying the *MEIS1* risk haplotype. In fact, electrophoretic mobility shift assays (EMSA) indicated that this regulation was achieved through two MEIS1 binding sites at the *SKOR1* promoter (Catoire et al. 2018).

Rare coding variants within *MEIS1* have also been screened for RLS causal effects. The nonsynonymous coding variant p.R272H substitutes arginine for histidine within the highly conserved TALE homeobox domain was detected in two studies involving familial cases (Vilariño-Güell et al. 2009; Schulte et al. 2011). One large-scale sequencing study screened more than 3000 cases and 3000 controls and observed an accumulation of rare variants in RLS cases compared to control (Schulte et al. 2014). This study investigated these variants further by employing an in vivo zebrafish model to study the impact of detected *MEIS1* variants on neurogenesis. A total of 17 detected nonsynonymous variants (13 in isoform 1 and 4 in isoform 2) were tested in a zebrafish complementation assay. Six of these variants are found to be deleterious for optic tectum development including the previously reported variant p.R272H. This approach strongly implicated rare variants in RLS (Schulte et al. 2014).

Mouse models have been used extensively to study RLS, but this is a rather challenging task due to the complex phenotype with motor and sensory components. As mentioned before, Meis1 homozygous knockouts are not viable after E14.5. Heterozygous *Meis1*-knockout mice showed signs of motor hyperactivity when young and middle-aged (Spieler et al. 2014; Salminen et al. 2017). Furthermore, in the middle-aged cohort the increase in locomotor

activity was observed at the beginning of the rest phase, suggesting an impaired circadian rhythm which resembles certain symptoms in RLS patients (Salminen et al. 2017).

As discussed, the RLS association signal within the *MEIS1* spans a 32 kb LD block. This region harbours seven highly conserved non-coding regions (HCNR) (Spieler et al. 2014). The lead RLS SNP, rs113851554, is located in intron eight of *MEIS1* within *HCNR* designated 602, based on genomic coordinates. This risk variant is a low-frequency variant (MAF = 0.053) with a significance level of $P = 2x10^{-280}$ and an odds ratio of 1.92 (Schormair et al. 2017). In addition, an ongoing meta-analysis reports a P-value of 10^{-1105} for rs113851554 (unpublished data). Another strongly associated SNP in the LD block, rs12469063, tags a distinctly HCNR designated as 617. This element was assayed for putative enhancer activity in zebrafish, exhibiting a reproducible neural expression pattern which was genotype dependent. This was confirmed using a transgenic beta-galactosidase (lacZ) enhancer, which demonstrated enhancer activity in the embryonic ganglionic eminences (GE). The enhancer activity was reduced with the risk allele (Spieler et al. 2014). In addition to the *MEIS1* risk locus, there is a noteworthy distinct RLS risk locus on chromosome 2, in an intergenic region located 1.3 Mb downstream of *MEIS1* tagged with SNP rs1280987. The RLS signal is detected in a 120 KB LD block, tagged with rs1820987 ($p=8.22\times10^{-147}$, unpublished metaGWAS). The SNP is located within a conserved genomic region, but its functional relevance has not yet been investigated. The genes *C1D, ETAA1 and MEIS1* are in the vicinity. It is possible that this region engages in long-range interaction with *MEIS1* and thereby regulates it. Figure 1.4 depicts the genomic positions of both RLS associated loci on chromosome 2.

In summary, the above-mentioned findings strongly implicate *MEIS1* in RLS pathogenesis, although the mechanism remains to be identified.

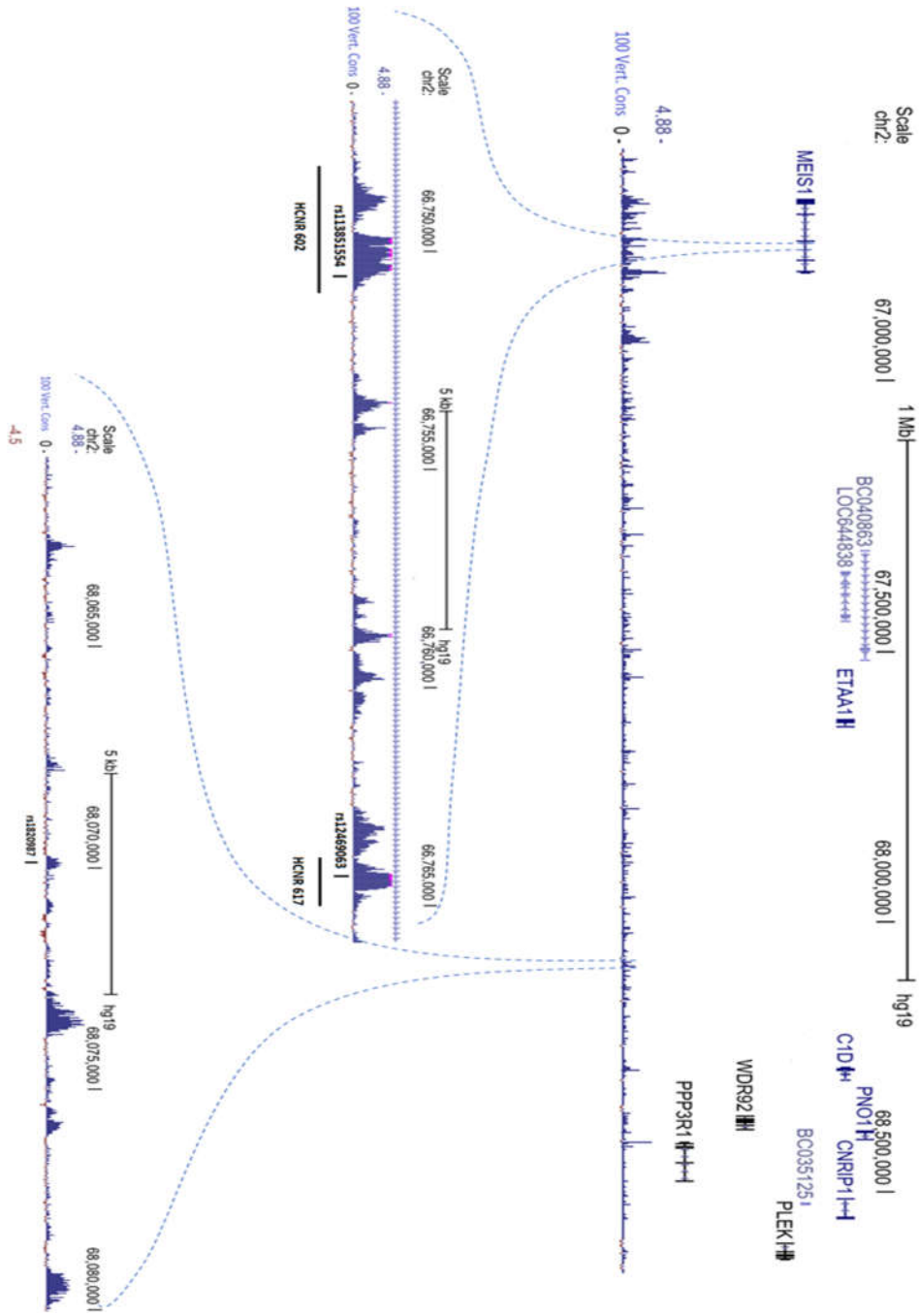


Figure 1.4: Genomic position of two RLS-associated risk loci: Upper part of the figure- *MEIS1* and an intergenic region on chromosome 2p14, marked with dashed blue lines. Lower part of the figure: zoomed MEIS1 risk locus with marked SNP positions of rs113851554 and rs12469063; zoomed intergenic risk locus with marked SNP position of rs1820987. 100 Vert. Conc corresponds to base wise conservation across 100 vertebrates (Pollard et al. 2010). Generated using UCSC Genome Browser, hg19 http://genome.ucsc.edu (Kent et al. 2002). The session URL: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2%3A66609087%2D68634033&hgsid=1135399453_l0PZM2YAWLSecgeJT5j1dhHZRsG1

1.5 Genomic approaches for GWAS follow-up studies

Genome-wide association studies (GWASs) discovered common genetic variants contributing to normal and pathological traits but finding the true causal variants poses a great challenge (Edwards et al. 2013). As mentioned before, the vast majority of GWAS variants reside in the non-coding regions so several approaches and strategies emerged in the post GWAS era focusing on functional characterization.

One approach is to investigate regulatory elements in the genome is to explore chromatin accessibility. Open chromatin regions (OCRs) are nucleosome-depleted regions that can be bound by transcription factors (TF) and further regulate gene expression (Krude 1995). Two widely used methods nowadays to map accessible chromatin and to infer regulatory elements are DNase I hypersensitive site sequencing (DNase-seq) (Song and Crawford 2010) and an assay of transposase accessible chromatin with whole-genome sequencing (ATAC-seq) (Buenrostro et al. 2013). Moreover, the methods can capture dynamic changes in chromatin remodelling and subsequently in regulatory landscapes during cellular differentiation. DNase-seq utilizes the endonuclease DNase I enzyme to preferentially digest nucleosome-depleted DNA regions followed by sequencing. Similarly, in ATAC-seq, a genetically engineered enzyme called transposase Tn5 is used to simultaneously cleave DNA and insert sequencing adapters in accessible DNA. Even though both techniques can comparably infer regulatory regions, DNase-seq requires substantially larger input material compared to the simple, low input material and time-efficient ATAC-seq protocol. Furthermore, a single cell ATAC-seq method is developed (Buenrostro et al. 2015).

To study human neurogenesis, de la Torre – Ubieta et al. assayed chromatin accessibility by employing ATAC-seq in human developing cortical plate and germinal zone. They identified enhancers regulating the key genes which drive human corticogenesis. Furthermore, GWAS variants associated with educational attainment, the risk for neuropsychiatric disease were enriched within these regulatory elements (de la Torre-Ubieta et al. 2018). Another study carried out ATAC-seq using an *in vitro* model of neuronal differentiation to study schizophrenia (SZ). It was shown that neural OCRs flanking TF-binding footprints can help prioritize putative functional noncoding schizophrenia GWAS risk variants (Forrest et al. 2017). It is important to mention that regulatory elements can regulate multiple genes and one gene can be regulated by multiple enhancers (Meddens 2016). Enhancers and silencers could be located within the gene body or distally from the genes they regulate and can be found

even up to several megabases away from the transcription start site. In order to regulate transcription, regulatory elements engage in physical proximity with gene promoters, mediated through the formation of chromatin loops. There are several techniques to study spatial genome organization. Chromosome conformation capture (3C) technique can detect a physical interaction between two loci. This assay can be used for two known regions which makes it disadvantageous. The 3C method was successfully used to confirm physical contact between risk variant rs6983267 identified in prostate and colorectal carcinoma and its target *MYC,* a proto-oncogene (Pomerantz et al. 2009). The next technique derived from 3C is called circular chromosome conformation capture 4C coupled with next-generation sequencing (Splinter et al. 2011). The method can be used to explore the interaction profile of a single locus of interest such as gene promoter or regulatory element. The advantage of the method is the ability to capture unknown interacting regions. By employing 4C-seq one study investigated obesity-associated variant rs9930506 located within intron 1 of the *FTO* gene. They found this variant in long-range interaction with an *IRX3* gene promoter located 500kb downstream. Together with the mouse model and human expression quantitative trait locus (eQTL), the authors implied the role of the *IRX3* gene in obesity (Smemo et al. 2014). Further, improving 3C methods led to the development of the Hi-C method allowing analysing the spatial organization of the entire genome in an unbiased way (Lieberman-Aiden et al. 2009). A high-resolution chromosome conformation capture detects promoter-enhancer interaction and can be a valuable tool to assign risk variants to their target genes. This approach was used to identify gene regulatory networks during neurogenesis. The authors of the study generated Hi-C interacting profiles using tissue from postmortem fetal cortex and integrated them with schizophrenia-associated risk variants. The study highlighted candidate genes and pathways involved in disease pathogenesis. Furthermore, they identified an interaction between risk SNP rs1191551 and the *FOXG1* gene located 760 Kb away from the risk locus. This contact was found to be functional as upon deletion of the rs1191551 flanking region, the *FOXG1* expression level was reduced in the neural progenitor cell line (Won et al. 2016).

However, proving an open chromatin region and/or physical proximity of a regulatory element to a promoter cannot tell us in which direction the effect goes.

By using CRISPR Cas9 gene-editing system one can successfully introduce a targeted deletion of a regulatory element containing a candidate causal SNP in a genome. The next

step is to measure the mRNA expression of a candidate target gene. In one such study, the authors inspected the functional relevance of coronary artery disease (CAD) risk SNP rs9349379, a common SNP in the third intron of the *PHACTR1* gene. They introduced an 88-bp deletion including the risk variant in an induced pluripotent stem cell (iPSC) and differentiated these cells into two vascular cell types: endothelial cells (ECs) and vascular smooth muscle cells (VSMCs). In both iPSC-derived ECs and VSMCs, loss of the 88 bp flanking rs9349379 resulted in the upregulation of the endothelin-1 (EDN1), a gene located 600 kilobases upstream from *PHACTR1*. Interestingly, the deletion of this regulatory element did not affect the expression level of *PHACTR1* or other genes within 1 megabase distance. Furthermore, the specificity of the risk variant was confirmed in vascular cell types as the deletion of the regulatory element did not affect gene expression in the neural crest progenitor cells (Gupta et al. 2017). Similarly, a risk variant for Alzheimer's disease (AD) rs6733839 was examined. Firstly, the authors performed extensive epigenetic profiling to identify cell types where the risk variants exhibit features of activity. They have found that AD risk variants are enriched in microglia regulatory elements. Next, they used the CRISPR-Cas9 editing system to delete 363 base pairs microglia specific enhancer harbouring the risk variant rs6733839. After performing deletion in induced pluripotent stem cells, this line was differentiated into neurons, microglia, and astrocytes. Deletion of the enhancer element reduced *BIN1* expression in microglia but not in neurons and astrocytes (Nott et al. 2019). These findings underline the relevance of cell specificity for investigating the functional role of variants contributing to complex diseases.

Finally, the above-mentioned methods are frequently combined with ChIP-seq, RNA-seq, and proteomics. With an integrated multi-omics approach all layers of gene regulation can be inspected and this finding can contribute to a better understanding of complex diseases.

1.6 Aim of the study

The aim of this work was to dissect the role of putative RLS risk variants within the *MEIS1* locus and to target cell lines where these variants exhibit their role. More precisely, the aim was to explore the regulatory features of two genomic regions within the *MEIS1*, HCNR 602 and HCNR 617, which harbour RLS-associated variants. In addition, I investigated another RLS risk locus, located 1.3 Mb downstream of *MEIS1* which could potentially be involved in the *MEIS1* regulatory network. To achieve these aims, I employed complementary genetic and epigenetic methods to assess the function of these regulatory elements.

In order to gain insight into the regulatory elements of the *MEIS1*, I profiled open chromatin compartments in differentiated neural cell types and pluripotent stem cell. Moreover, I evaluated the *MEIS1* regulatory landscape in a subset of public datasets for both human and mice selected based on relevancy for RLS and MEIS1 function.

In search of regulatory elements of *MEIS1*, I also evaluated a three-dimensional (3D) organization of *MEIS1*. I investigated different cell types to search for cell type-specific interactions.

To validate the regulatory function of the RLS - associated regulatory element, I performed a targeted deletion with subsequent *in vitro* differentiation and evaluation of *MEIS1*.

In summary, I investigated *MEIS1* regulatory dynamics using an *in vitro* neural differentiation system as well as other cell types reflecting different stages of neural maturity to capture the spatiotemporal landscape of *MEIS1* regulation. The overall aim was to evaluate how RLS-associated variants regulate *MEIS1* in neurodevelopment and to prioritize cell types for further functional studies.

MATERIAL AND METHODS

2.1 Cell Lines

2.1.1 HMGU1

HMGU1 line is an induced pluripotent stem cell line (iPSC) obtained from Helmholtz Zentrum München iPSC Core Facility. The line is generated from a new-born male healthy donor from human foreskin fibroblasts by transfection of five mRNA reprogramming factors: Oct4, Sox2, Klf4, Lin28, and c.Myc. The source of the cells is BJ (ATCC CRL-2522). The line is maintained in mTeSR™1 culture media (Stemcell Technologies # 85850) on cell culture dishes Corning® Costar® (Sigma, # CLS3516-50EA) coated with Matrigel Matrix (Corning, # 354277) or Geltrex Matrix (Thermo Fisher Scientific, # A1413202).

2.1.2 HMGU12

HMGU12 line is an induced pluripotent stem cell line (iPSC) obtained from Helmholtz Zentrum München iPSC Core Facility. The line is generated from a new-born male healthy donor from human foreskin fibroblasts by transfection of six mRNA reprogramming factors: Oct4, Sox2, Klf4, Lin28, c.Myc, and Nanog. The source of the cells is BJ (ATCC CRL-2522). The line is maintained in mTeSR™1 culture media (Stemcell Technologies # 85850,) on cell culture dishes Corning® Costar® (Sigma, # CLS3516-50EA) coated with Matrigel Matrix (Corning, #354277) or Geltrex Matrix (Thermo Fisher Scientific, # A1413202).

2.1.3 GIBCO® Human Neural Stem Cells (H9-Derived)

H9 NSC line (# N7800200) was purchased from Thermofisher Scientific. This line is derived from H9 (WA09) human embryonic stem cells (hESCs). The cells were cultured as a monolayer, maintained in complete StemPro® NSC SFM (Thermofisher Scientific #A1050901) consisted of KnockOut™ D-MEM/F-12 with StemPro® Neural Supplement, EGF, bFGF, and GlutaMAX (Thermofisher Scientific, #35050061) in cell culture dishes coated with Geltrex according to the manufacturer's protocol.

2.1.4 iCell® GABANeurons

Gabaergic neurons (iCell GABANeurons Kit, #01279) were purchased from Cellular Dynamics International (Madison, WI, USA). iCell GABANeurons are a highly pure population of human neurons, composed primarily of GABAergic neurons, derived from induced pluripotent stem (iPS) cells using CDI's proprietary differentiation and purification protocols. The cells were thawed, plated, and maintained according to the manufacturer's instructions (https://fujifilmcdi.com/assets/CDI_iCellNeuronsUsersGuide.pdf) and collected for downstream experiments five days post-plating.

2.1.5 iCell® GlutaNeurons

Glutamatergic neurons (iCell GlutaNeurons Kit, #01279) were purchased from Cellular Dynamics International (Madison, WI, USA). iCell GlutaNeurons are a highly pure population of human neurons, composed predominantly of cortical glutamatergic neurons, derived from the induced pluripotent stem (iPS) cells using CDI's proprietary differentiation and purification protocols. These cells were thawed, plated and maintained according to the manufacturer instructions (https://fujifilmcdi.com/assets/CDI_iCellGlutaNeurons_UG.pdf) and collected for downstream experiments five days post-plating.

2.1.6 SHSY-5Y

Neuroblastoma cell line (SHSY-5Y) originates from a metastatic tumor. It is subcloned from cell line SK-N-SH derived from bone marrow biopsy of a 4 years old female patient with neuroblastoma (Biedler et al. 1978). The line is maintained in DMEM medium (Thermofisher Scientific #41966-029) supplemented with 15% of fetal bovine serum (Thermofisher Scientific, #10270-106) and 0.5% of penicillin-streptomycin.

2.2 Cell culture maintenance

2.2.1 Cell thawing

The cell's vials stored in liquid nitrogen were thawed in a water bath at 37º C for no more than 3 minutes. Next, 1 ml of media was added dropwise, and the entire content was transferred into a 15 ml Falcon tube. The tube was centrifuged at 300g for 4 minutes. Next, the supernatant was aspirated, and the cells were resuspended in 2 ml pre-warmed media and

seeded on a previously coated 6 - well plate. The plate was placed in the incubator at 37° C and 5 % CO_2 atmosphere.

2.2.2 Cell splitting

For general subculturing of HMGU12 and HMGU1 lines, we followed recommendations obtained from iPSC Core Facility. Cells were washed once with 1 ml of Versene solution (Thermofisher Scientific, #15040066). Next, 1 ml of the same dissociation agent was added, and cells were incubated for no more than 7 minutes in the incubator at 37° C. At that time, the detachment process was monitored under the microscope. Following, the Versene was gently aspirated and 1 ml of pre-warmed media was added. Cells were carefully pipetted several times to detach the clumps using a 1ml pipette. The cell clusters were seeded into freshly coated 6- well plates in mTESR media at the split ratio between 1:4 and 1:10.

For the single cell splitting, cells are first washed with Dulbecco's phosphate-buffered saline (DPBS) (Thermofisher Scientific, #14190144), and then treated with TrypLE Select Enzyme (Thermofisher Scientific, #12604013) for 5 minutes at 37° C. Next, a prewarmed media was added to neutralize dissociation. The cell suspension was transferred into a 15 ml falcon tube and centrifuged at 300g for 4 minutes. The supernatant was removed, and cells are either counted prior to seeding or seeded on an adequate splitting ratio, depending on the downstream application. The same procedure applies to H9 NSC lines.

2.2.3 Cell freezing

Cells were dissociated with Versene solution or Trypsin-EDTA 0.25% (Thermofisher Scientific, #25200056). After dissociation, the reagent was neutralized with prewarmed media, the cells are transferred into 15 ml Falcon, centrifuged, and then resuspended in PSC Cryopreservation medium (Thermofisher Scientific, #A2644601), 1ml per one vial. Following, vials are frozen in cryogenic boxes and stored at -80° for no more than 48 hours. Subsequently, vials are moved into liquid nitrogen.

2.3 Methods

2.3.1 Alt-R CRISPR/Cas9 system

To delete a highly conserved non-coding region 617 harbouring RLS risk variant rs12469063, first, we designed two guides flanking conserved element 617. We used an online tool (http:// crispr.mit.edu/). Guides were positioned upstream and downstream of the conserved sequence and introduced 922 bp deletion encompassing the SNP rs12469063. To ensure efficient deletion, we employed Cas9-gRNA ribonucleoproteins. The advantage over plasmid delivery is pre-assembled complex of Cas9 protein and gRNA. Since Cas9 protein gets degraded fast, it reduces the off-target effect. All components were purchased from Integrated DNA Technologies IDT and the complexes were assembled according to IDT protocols. Firstly, crRNA (contains target-specific nucleotide sequences) and tracrRNA (transactivating CRISPR RNA) are resuspended in nuclease-free IDTE buffer at 200μM final concentration. To assemble guide RNA duplexes, crRNA, for each guide and trRNA were mixed in equimolar concentration in a sterile microcentrifuge tube and incubated at 95ºC for 5 minutes and subsequently cooled down to 20ºC at 0.1 ºC/ min rate. To assemble ribonucleoprotein complexes, crRNA:tracrRNA duplex was mixed with Alt-R HiFi recombinant S. pyogenes Cas9 nuclease and incubated at room temperature for 20 minutes.

GUIDE NAME	*SEQUENCE*
crRNA 617 1	TCAAGCTGGTATCTTTGGGA
crRNA 617 2	AACCATTAAGTAGAAAGACA

Table 2.1: Guide sequences for enhancer deletion

2.3.2 Transfection of the iPS cells

For transfection, we applied the nucleofection method using P3 Primary Cell 4D-Nucleofector™ X Kit (Lonza, #V4XP-3032) and 4D-NucleofectorTM X Unit. Briefly, HMGU12 iPSC cells were supplemented with RevitaCell™ Supplement (100X) (Thermofisher Scientific, #A2644501) at least one hour prior to nucleofection. Revitacell is a supplement containing a proprietary ROCK inhibitor that promotes single cell viability during dissociation and transfection. Subsequently, cells were dissociated into single cells using TrypLE Select Enzyme for 5 minutes at 37º C, and cells are counted on Countess II Automated Cell Counter (Thermofisher Scientific, #AMQAX1000). 150 000 cells were used

for nucleofection. First, 150 000 cells were spun down at 300g for 4 minutes and the supernatant was removed. Cell pellets are resuspended in a nucleofector solution composed of 16.4 µl of Nucleofector Solution + 3.6µl of Supplement. RNP complexed for each guide were conucleofected in the same reaction. Such a mixture was transferred into 20 µl Nucleocuvette™ strip (one well per one reaction). The strip was placed into 4D-NucleofectorTM X Unit and cells were nucleofected using CB-156 program. The nucleofected cells were transferred into 24-well plate in pre-warmed mTESR media supplemented with RevitaCell™ Supplement (100X). 24h post-nucleofection, the media was replaced. The media was replaced daily and after 3 days the cells were dissociated, counted, and 500 cells were seeded in 6 cm dish, coated with Geltrex supplied with mTESR media supplemented with Revita. After one week, colonies were manually picked and plated in 24-well plate. After five days, cells were split again into two 24-well plates, one for deletion screening and one for clone picking.

For deletion screening in 24-well plate, we used Kapa express kit for DNA isolation (Kapabiosystems, #07961618001). Cells in 24-well plate were dissociated with 500 µl per well of TrypLE Select Enzyme for 5 minutes, neutralized with the same amount of DMEM/F-12 (Thermofisher Scientific, #11320074), and the plate was centrifuged. After the centrifugation step, the supernatant was carefully aspirated. The cells are then resuspended in 100 µl of lysis solution (per well) composed of 88 µl of H_2O, 10 µl of 10X KAPA Express Extract Buffer and 2 µl of KAPA Express Extract Enzyme. The entire content of the plate was transferred to 96- well PCR plate (Thermofisher Scientific, #AB0700), and lysis was performed in a thermocycler using the following cycling parameters: 75ºC for 10 minutes and 95ºC for 5 minutes. Next, PCR was performed to screen for deletion. We used two primer pairs to correctly confirm the deletion of enhancer. The first primer pair is designed to flank the deletion and amplifies 1499 bp fragment. To distinguish between homozygous and heterozygous deletion clones, we performed PCR using primer pair amplifying the inner portion of deleted fragments. The primers are designed to amplify the 213 bp inner portion of the enhancer that is expected to be deleted. Primers are ordered from Metabion, desalted, and diluted to 10µl working stock concentration. For the PCR reaction Q5® High-Fidelity 2X Master Mix (NEB, #M0492L) was used in the set up shown in Table 2.2.

COMPONENT	*VOLUME FOR 25 µL REACTION*
Q5 High-Fidelity 2X Master Mix	12.5
10 µM Forward Primer	1.25
10 µM Reverse Primer	1.25
DNA	2
H_2O	8

Table 2.2: The reaction set up for PCR

The PCR was performed using cycling conditions from Table 2.3.

CYCLING STEP	*TIME (SEC)*	*TEMP (°C)*
Initial denaturation	60	98
30 x	10 30 30 sec/kb	98 58 72
Final extension	120	72
Hold	∞	4

Table 2.3: Cycling parameters for PCR

Finally, the amplification products were visualized on 1.5% agarose gel.

Clones found in initial screening were propagated and deletion was confirmed with Sanger sequencing. The primer sequences for deletion detection are shown in Table 2.4.

CYCLING STEP	*TIME (SEC)*	*TEMP (°C)*
Initial denaturation	60	98
30 x	10 30 30 sec/kb	98 58 72
Final extension	120	72
Hold	∞	4

Table 2.4: Primer sequences for deletion detection

2.3.3 Cell differentiation

HMGU12 cell line and HMGU1 cell (cell lines obtained in editing experiment) were cultured in mTESR media, in 6-well plates coated with Matrigel. We differentiated the cells into neural progenitors using previously published protocol from with slight changes (Close et al. 2017). The cells were grown to 90% confluency and then dissociated with TrypLE Select Enzyme for 5 minutes at 37°C. The dissociation was stopped by adding an equal amount of DMEM/F-12. Next, the cells are counted using Countess II Automated Cell Counter and 500 000 iPS cells were plated per well of 24 well plate in mTESR media, supplemented with RevitaCells. Following day (D1) the media is replaced with neural induction media (NIM) consisting of: DMEM/F12, 1X N-2 supplement (Thermofisher Scientific, #17502048) 1X B-27 supplement (Thermofisher Scientific, #17504044), 2mM Glutamax Supplement (Thermofisher Scientific, #35050061), 0.1 mM MEM Non-Essential Amino Acids Solution (Thermofisher Scientific, #11140035), 0.11 mM 2-mercaptoethanol (Thermofisher Scientific, #31350010), 0.05% (v/v) Bovine Albumin Fraction V Solution (Thermofisher Scientific, #15260037), Penicillin-Streptomycin (Thermofisher Scientific, #15140122), 100 nM LDN193189 (Biomol, #Cay11802-1) 10 µM SB431 (Cayman Chemical, #13031-5) and 2 µM XAV939 (R and D systems, #3748/10). This media was changed daily until day 5 when it was replaced by media containing 75% of NIM and 25 % of N2 media. N2 media composition as follows: DMEM/F12 (1:1), N-2 supplement, 0.15% (w/v) dextrose, 55 µM 2-Mercaptoethanol, Penicillin-Streptomycin. Next, on day 7 the cells are fed with media composed 50%: 50% NIM/N2. On day 9, the cells are fed with media 25%: 75% NIM/N2. On D10, the cells were dissociated into a single cell suspension with TrypLE Express Enzyme and plated onto Matrigel-coated 24-well plates in 25% NIM supplemented with Revita, at a density of 1.1 × 106/cm2. On the same day, one part of the cells was collected for RNA extraction and ATAC-seq. Following, on D11 the medium was replaced with N2/B27 medium (N2 medium (above) + B-27 supplement) containing 0.65M purmorphamine for HMGU12 line and 1M purmorphamine for HMGU1 line (Biomol, #Cay10009634-1). On D19-23, cells were fed daily with N2/B27 medium without purmorphamine. On day 24, the cells are collected for RNA extraction and ATAC-seq.

2.3.4 Gene expression

Total RNA from HMGU12, HMGU1 control and deletions lines, neural progenitor lines, iNeurons, and H9 NSC was extracted with RNeasy Mini Kit (Qiagen, #74104) according to

the instructions of the manufacturer. For each sample, the RNA concentration is measured on Nanodrop. RNA is transcribed into complementary DNA (cDNA) using High-Capacity cDNA Reverse Transcription Kit according to the provided protocol of the manufacturer (Thermo Fisher Scientific, #4368814). Up to 500 ng of RNA was used for reverse transcription. cDNA was diluted 1:2 with H_2O and used for quantitative PCR (qPCR).

Quantitative PCR was performed in duplicate for each sample using TaqMan™ Universal PCR Master Mix (Thermofisher Scientific, #4304437). The TaqMan gene expression assays used in this experiment are: *MEIS1*: Hs00180020_m1, the assay probe spans an exon junction and amplicon length is 88 bp; *MEIS2*: Hs00487743_m1 the assay probe spans an exon junction and amplicon length is 66 bp; *POU5F1* (*OCT4*): Hs00999632_g1 the assay probe spans an exon junction and amplicon length is 77 bp; *NANOG*: Hs02387400_g1 the assay probe spans an exon junction and amplicon length is 109 bp; *PAX6:* Hs00240871_m1 the assay probe spans an exon junction and amplicon length is 76 bp; *NESTIN:* Hs04187831_g1 the assay probe spans an exon junction and amplicon length is 58 bp; *FOXG1:* Hs01850784_s1; the assay probe spans within a single exon and amplicon length is 71 bp; *GSX2*: Hs00370195_m1 the assay probe spans an exon junction and amplicon length is 69 bp; *NKX2.1*: Hs00968940_m1; the assay probe spans an exon junction and amplicon length is 72 bp; *DLX2:* Hs00269993_m1 the assay probe spans an exon junction and amplicon length is 56 bp; *GAPDH*: Hs02758991_g, the assay probe spans an exon junction and amplicon length is 93 bp. The following experiment set up in Table 2.5 was used.

COMPONENT	*VOLUME FOR 10 µL REACTION VOLUME*
TaqMan® Universal PCR Master Mix 2X	5
Taqman gene expression assay	0.5
cDNA template	1
H_2O	3.5

Table 2.5: The reaction set up for qPCR

The qPCR was performed using 7900HT Fast Real-Time PCR System (Applied Biosystems using cycling conditions in Table 2.6.

CYCLING STEP	*TIME (MIN)*	*TEMP (°C)*
UNG incubation	2	50
Polymerase activation	10	95
40 x	15 sec 1	95 60
Hold	forever	4

Table 2.6: Cycling parameters for qPCR

For the relative quantification, we used the comparative ΔCт method. GAPDH was used as endogenous controls for expression normalization.

2.3.5 Staining

Cells are fixed in 4% Paraformaldehyde (Thermofisher Scientific, #28906) for 15 minutes and washed 3x with PBS. Next, cells are blocked in blocking solution (0.1% Triton X, 1x PBS; 10% serum) for 1h and then incubated with primary antibody overnight at 4° C. After overnight incubation cells are washed with PBS (3x 5 min) and incubated with secondary antibody diluted in blocking solution for 1h at room temperature. Samples are washed again and incubated in DAPI for 3 minutes. Primary antibodies used are: SOX2 (1:300 Abcam, 97959), OCT4 (1:300 Abcam, #18796), PAX6 (1:300 Biozol, #BDL901301), Nestin (1:300 Santa Cruz, #sc23927). Secondary antibodies used are Alexa Fluor® 488 Goat Anti-Mouse IgG antibody (1:500 Thermofisher Scientific, #A-11001) and Alexa Fluor 594 Goat anti-Rabbit IgG (H+L) Antibody (1:500 Thermofisher Scientific, #A-11012).

2.3.6 Circular chromosome conformation capture sequencing (4C-seq)

In order to evaluate the spatial organization of the *MEIS1* and identify promoter-enhancer interaction, I employed circular chromosome conformation capture 4C-seq in different cell types using the *MEIS1* promoter as a viewpoint. A graphical method summary and final library presentation are shown in Figure 2.1.

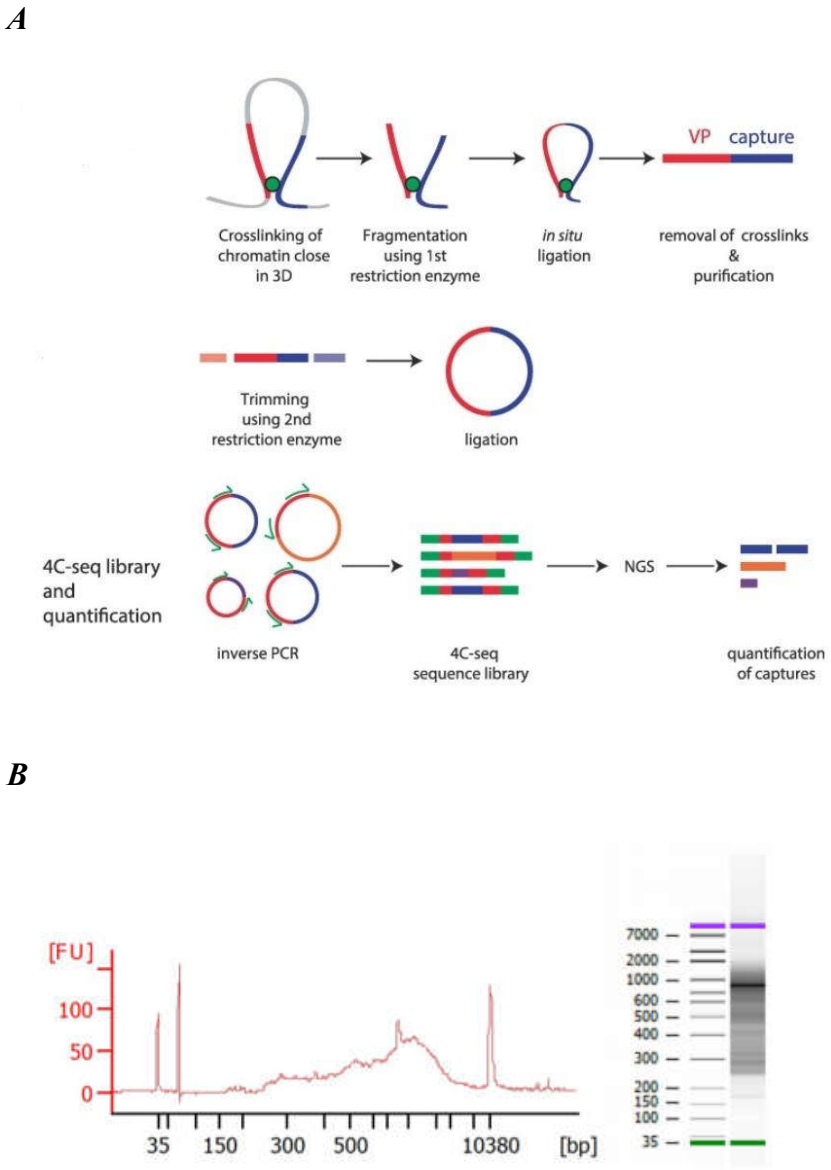


Figure 2.1: 4C-seq method workflow (A) adapted from (Krijger et al. 2020) and fragment distribution of 4C-seq library on Bioanalyzer (B)

Cell lines used for circular chromosome conformation capture are HMGU1, H9NSC, and SH-SY5Y neuroblastoma cell line. We used the protocol previously described (van de Werken et al. 2012). For each fixation, fresh 4% paraformaldehyde was prepared. For this, 16% Formaldehyde (Thermofisher Scientific, #28906) was diluted with sterile water to 4 %.

The cells were dissociated into single cells and resuspended in 5 ml of DPBS in a 15 ml Falcon tube. Next, we added 5 ml of 4% formaldehyde to achieve a final concentration of 2%. Cells are incubated in fixation solution for 10 minutes while rotating. Next, 1.425 ml of 1 M glycine was added to quench the reaction, and Falcons are placed on ice. Subsequently, the cells were centrifuged for 10 minutes at 400g at 4°C. The supernatant was discarded, and the

cells were resuspended in 5ml of cold lysis buffer (50 mM Tris pH 7.5,150 mM NaCl, 5 mM EDTA, 0.5% NP-40 (Sigma #I8896), 1% TX-100 (Sigma #T8532), 1 × protease inhibitors (Roche #11245200)) and incubated on ice for 30 minutes. To make sure that the lysis was complete, cells were inspected under the microscope using Methyl green pyronin (Dianova, #MGP125). In the case of efficient lysis, nuclei are stained blue, and cytoplasm was stained pink. Next, the lysed pellets are centrifuged, and washed once with DPBS, and resuspended in 440 µl of water, in a 1.5 ml Eppendorf tube. 60µl of 1X NEBuffer™ DpnII was added, and samples are placed in thermomixer at 37ºC. Then, 15 µl of 10 % SDS was added and incubated for one hour while shaking at 900 rpm with occasional pipetting to break the cell aggregates. After, 75µl of 20 % of Triton X (Biomol, #600217100) was added and again incubated for one hour while shaking. A 10µl aliquot of undigested control was taken from each sample before starting the digestion, and then the first round of digestion was done with 400 U of DpnII (New England Biolabs, R0543M). Following the overnight digestion, 10 µl of digested control from each sample is taken to evaluate the digestion efficiency. In each 10 µl aliquot, 90 µl of 10 mM of Tris-HCl and 2.5 µl of 30 mg/ml proteinase K is added, mixed and samples are incubated for one hour at 65 ºC. Afterward, samples are purified using phenol-chloroform and loaded on 0.6% agarose gel. Efficient digestion is presented as a smear whereas undigested samples are seen as one high molecular weight band. After successful digestion is verified, the DpnII enzyme is inactivated by incubating samples for 20 minutes at 65ºC. The samples are transferred in 50 ml Falcon tubes and 700 µl of 10x ligation buffer added (400 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, and 5 mM ATP), plus 5.7 ml of water and 150 U of T4 Ligase (Thermofisher Scientific, #EL0012). The ligation step was carried out by overnight incubation at 16 ºC. The following day, 100 µl aliquots are taken to assess proper ligation. The control aliquots are incubated with proteinase K at 65 ºC for one hour. Subsequently, the samples are loaded on 0.6 % gel next to undigested and digested control. In the case of proper ligation, a band of high molecular weight is seen. Upon successful ligation, the samples are decrosslinked using 15 µl of Proteinase K at 65˚C overnight. DNA was purified using phenol-chloroform and the second round of digestion was performed with 60 U per sample of Csp6I (Thermofisher Scientific, #ER0211). Digestion was followed by overnight ligation at 16ºC. To do so, the samples are moved to 50 ml Falcon tubes, mixed with 12.1 ml of water, 1.4 ml of 10x ligation buffer, and 250 U of T4 ligase and incubated overnight. The following day probes are purified using Amicon® Ultra Centrifugal Filters (Millipore, #UFC900324) and DNA is quantified using

Qubit dsDNA HS Assay Kit (Thermofisher Scientific, #Q32851). To generate a 4C library, I used 1 µg of DNA for PCR amplification divided into 10 PCR reactions, using Expand™ Long Template PCR System (Sigma Aldrich, #11681834001). The experiment setup is shown in Table 2.7.

COMPONENT	*VOLUME FOR 25 µL REACTION*
4C DNA template	100 ng
Expand Long Template buffer 1	2.5
Primer F (100µM)	0.25
Primer R (100µM)	0.25
dNTP (10mM)	0.5
Expand Long Template Polymerase	0.35
H_2O	Up to 25

Table 2.7: The reaction set up for PCR

The PCR reaction was performed using cycling conditions in Table 2.8

CYCLING STEP	*TIME (MIN)*	*TEMP (°C)*
Initial denaturation	02:00	94
30x	00:15 01:00 03:00	94 55 68
Final elongation	05:00	68
Store	∞	4

Table 2.8: Cycling parameters for PCR

After pooling, the libraries are purified using SPRI beads (Beckam Coulter, #A63880). Purification was done at 1:1.8 ratio sample to the beads according to manufacturer instructions and fragment size was checked using Agilent High Sensitivity kit (Agilent, #5067-4626). The primers for chromosome conformation capture 4C-seq are designed using primer3 (Untergasser et al. 2012). They are located upstream from the transcription start site (TSS) of the *MEIS1* promoter. The forward primer is directed towards GATC sequence, the cut site for DpnII, and the reverse primer towards the secondary enzyme, the cut site of

Csp6I. They were synthesized with Truseq adapters on top. The adapter sequence is identical for the forward primer for each sample (Truseq universal adapter) whereas adapters on the reverse primer contain 6 base pairs index sequences (Truseq index adapter). The viewpoint specific primer sequences (bold) with Truseq adapters containing indexes (red) are shown in Table 9.

NAME	*PRIMER SEQUENCE*
F1_pro	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGC AACACACACTTTACACAC**
R1pro iPSC_1	CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**
R2pro iPSC_2	CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**
R3pro hNSC_1	CAAGCAGAAGACGGCATACGAGAT**GCCTAA**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**
R4pro hNSC_1	CAAGCAGAAGACGGCATACGAGAT**TGGTCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**
R5pro SH-SY5Y _1	CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**
R6pro SH-SY5Y _2	CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATC**GACTTTCCTCGAAATTATTGG**

Table 2.9: Primer sequences for 4C-seq

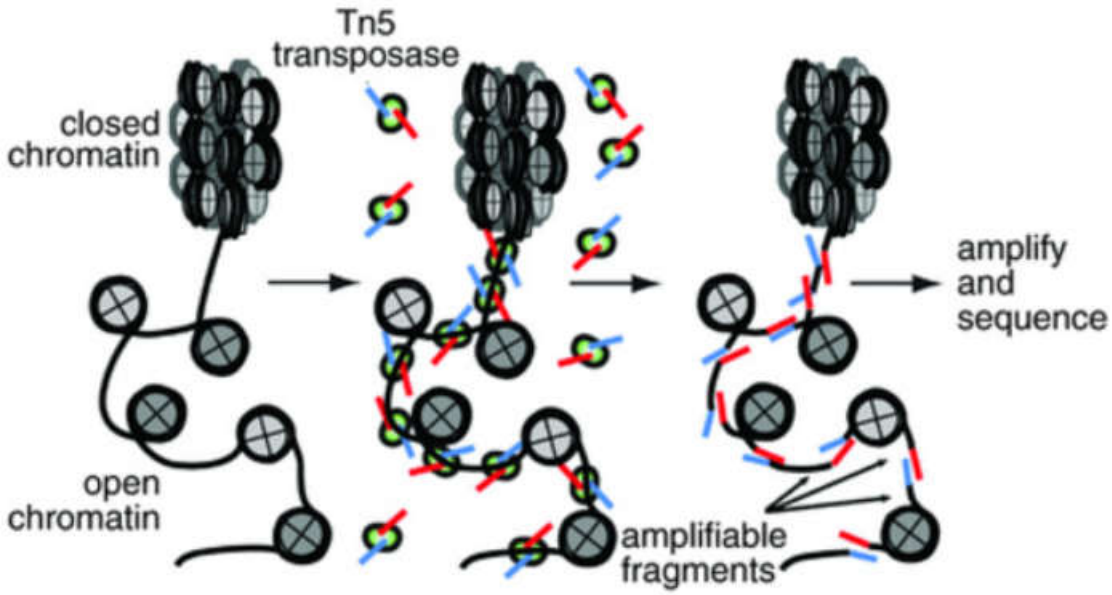
Analysis

The samples were sequenced on the Illumina MiSeq sequencing platform, spiked with 30% PhiX to increase library complexity. Data processing and analysis were done by Dr. Chen Zhao from Institute of Neurogenomics, Helmholtz Zentrum München. The reads were trimmed with Trimmomatic v0.36 (Bolger, Lohse, and Usadel 2014), aligned with bowtie2 (Langmead and Salzberg 2012) and subsequently processed, filtered, analysed, and visualized according to Basic4Cseq in R v 4.0.2 (Walter et al. 2014).

2.3.7 ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing)

To evaluate chromatin accessibility and to infer potential regulatory elements, I used an assay for transposase-accessible chromatin with high throughput sequencing in different cell lines. A graphical method summary and final library presentation are shown in Figure 2.2.

A



B

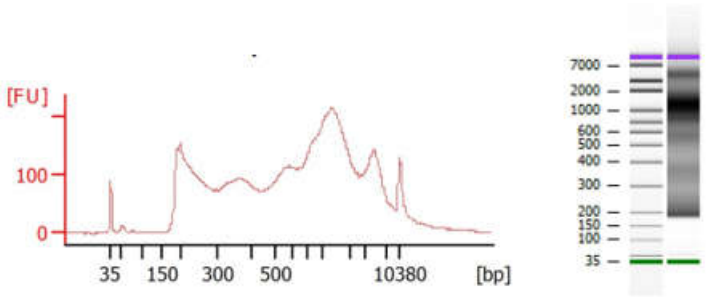


Figure 2.2: ATAC-seq -method workflow (A) adapted from (Buenrostro et al. 2013) and fragment distribution of ATAC-seq library on Bioanalyzer (B)

For ATAC-seq, I used HMGU12, HMGU12-derived neural progenitor, HMGU12-derived ganglionic progenitors, H9-derived neural stem cells, and gabaergic and glutamatergic neurons. I followed a previously published protocol (Buenrostro et al. 2015). For each replicate, 50 000 cells were harvested. Cells were washed once with ice-cold DPBS and centrifuged at 500g for 5 minutes at 4°C. Next, the cells were resuspended in ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630 (Sigma #3021)) and centrifuged for 20 minutes at 500g and 4°C. Following, the pellets were resuspended into 50µl of transposition reaction mix and incubated for 30 min at 37℃. Transposition reaction mix contains 25 µl of 2x reaction buffer, 2.5 µl of Nextera Tn5

Transposase (Illumina, #FC-121-1030), and 22.5 µl of Nuclease-free water (Thermofisher Scientific, #AM9938). Then, samples were purified using Qiagen MinElute Kit (Qiagen, #28004) and eluted in 10 µl of elution buffer. Following purification, samples were amplified using up to a maximum of 12 cycles. To determine an adequate number of cycles for each replicate, the initial PCR amplification of 5 cycles was performed using the following set up shown in Table 2.10.

COMPONENT	*VOLUME FOR 50 µL REACTION*
NEBNext High-Fidelity 2x PCR Master Mix	25
i5 Index primer	5
i7 Index primer	5
Transposed DNA	10
H_2O	5

Table 2.10: Reaction setup for PCR

The PCR reaction was performed using cycling conditions in Table 2.11.

CYCLING STEP	*TIME (MIN)*	*TEMP (°C)*
Extension	05:00	72
Initial denaturation	00:30	98
5x	00:10 00:30 01:00	98 63 72
Store	∞	4

Table 2.11: Cycling parameters for PCR

Next, 5µl of each PCR reaction was taken to perform qPCR. Sybr Green I (Thermofisher Scientific, #S7567) was used to assay an additional number of cycles needed. The following setup, shown in the Table 2.12 was used.

COMPONENT	*VOLUME FOR 15 µL REACTION*
NEBNext High-Fidelity 2x PCR Master Mix	5
i5 Index primer	0.5
i7 Index primer	0.5
Previously PCR-amplified DNA	5
100× SYBR Green I	0.09
H_2O	3.91

Table 2.12: Reaction setup for qPCR

The qPCR was performed using LightCycler 480 Instrument II (Roche Life Science) using cycling conditions in the Table 2.13.

CYCLING STEP	*TIME (MIN)*	*TEMP (°C)*
Initial denaturation	00:30	98
25x	00:10 00:30 01:00	98 63 72
Store	∞	4

Table 2.13: The cycling parameters for qPCR

After qPCR was finished, an additional number of cycles for each sample was assessed by plotting the fluorescence versus cycle. The number of cycles corresponded to ¼ of the maximal fluorescence intensity. The remaining 45 µl of each sample is subsequently amplified for additional 5-7 cycles.

The barcoded primers from Nextera Index Kit (Illumina, #FC-121-1011) were used. The samples are purified with SPRI beads (Beckam Coulter, #A63880) at 1:1.8 ratio sample to the beads, according to manufacturer instructions and the fragment size was checked using Agilent High Sensitivity kit (Agilent, #5067-4626). The libraries were sequenced on HiSeq 4000 sequencing platform, two samples per lane, 100 bp paired end reads.

Analysis

Data processing and analysis were done by Dr. Chen Zhao from Institute of Neurogenomics, Helmholtz Zentrum München. Briefly, reads were trimmed and aligned with bowtie2

(Langmead and Salzberg 2012), with standard parameters and a maximum fragment length of 2,000. Duplicate reads were removed with Picard http://broadinstitute.github.io/picard/. De-duplicated reads were filtered for high quality (samtools (H. Li et al. 2009), MAPQ ≥ 30), nonmitochondrial chromosome, non-Y chromosome, and proper pairing (samtools flag 0 × 2). Peaks were called with macs2 (Zhang et al. 2008) and filtered out with IDR threshold of 0.1 (Q. Li et al. 2011) and blacklist of artefactual regions in hg19 (Amemiya, Kundaje, and Boyle 2019). The libraries were quality controlled by downsampling to 5 million reads and evaluating transcription start site enrichment (RefSeq) (O'Leary et al. 2016).

RESULTS

3.1 Neural differentiation of human induced pluripotent stem cells toward ganglionic eminences progenitors

I sought to confirm previously reported findings and explore *MEIS1* expression pattern using human *in vitro* model of differentiation of induced pluripotent stem cell line (iPSC) into ganglionic eminences-like neural progenitors using a previously published protocol (Close et al. 2017). Figure 3.1 summarizes the protocol. As previously described in the method section, iPSCs (HMGU12) are differentiated towards GE-like neural progenitors in two key steps. Firstly, a telencephalic neural induction was initiated using a combination of small molecules SB431542, LDN1933189, and XAV939. The first two components, SB431542 and LDN1933189, act synergistically as inhibitors of SMAD signalling and thereby promote neural induction (Chambers et al. 2009). Together with XAV939 which inhibits WTN signalling, this combination of small molecules enables efficient and fast neural induction of human embryonic stem cells and induced pluripotent stem cells (Maroof et al. 2013). In the second step, to achieve more specific cell population towards ganglionic eminences, the cells were replated after ten days of neural induction. On day 11 the ventralization by 0.65µM of puromorphamine was initiated and lasted for 8 days. On day 19, purmorphamine treatment was stopped and the cells were maintained in N2B27 medium for additional 5 days. A small molecule purmorphamine is an agonist of protein Smoothened which is a key component of the Sonic hedgehog signalling pathway (Shh) (Sinha and Chen 2006). Shh promotes cellular specification towards interneurons (Close et al. 2017).

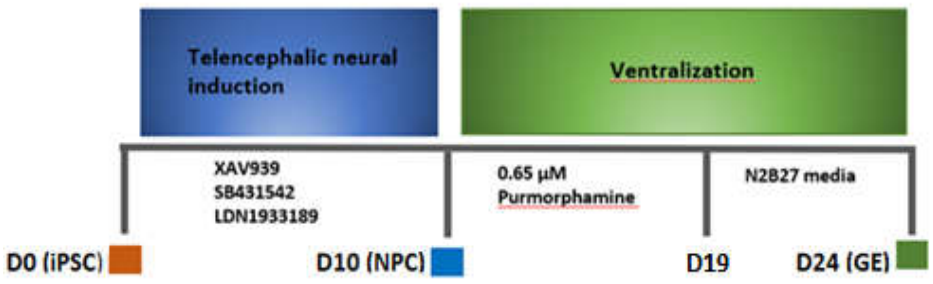


Figure 3.1: Neural differentiation of induced pluripotent stem cells towards GE-like neural progenitors (Adapted from Close et al. 2017).

During the differentiation process, I monitored the expression pattern of the marker genes at three different time points: day 0 which represent undifferentiated, induced pluripotent stem cell line (iPSC), at day 10 which defines neural progenitors (NPC), and at day 24 which represent the stage of ganglionic eminences-like neural progenitors (GE NPC). Furthermore, efficient telencephalic neural induction was confirmed through immunostaining for PAX6 and NESTIN and day 10 of differentiation as demonstrated in Figure 3.2.

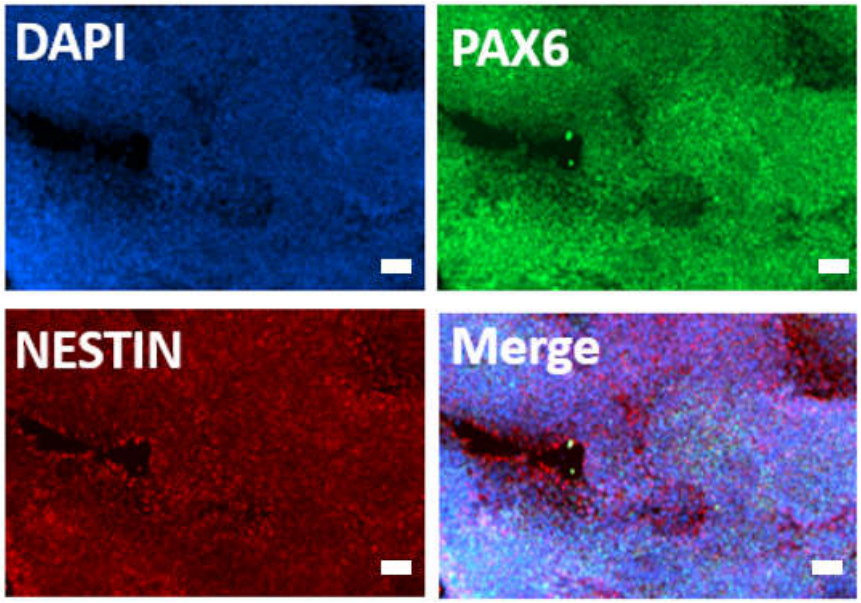


Figure 3.2: Immunofluorescence staining for PAX6 and NESTIN at day 10 of differentiation. Scale bar, 50 µm.

On day 10 of neural induction, qPCR results showed significant upregulation (around 2500-fold change) of the *PAX6* gene. This finding demonstrates efficient neural inductions as PAX6 is a marker of neuroepithelial cells (Chambers et al. 2009; Callaerts et al. 1997). On day 24, the expression level of *PAX6* was significantly downregulated compared to the NPC stage (Figure 3.3A) as described by (Zhu et al. 2019). This suggests that cell phenotype corresponds to ventral progenitors. It is important to mention that *Pax6* is expressed in the pallial ventricular zone of lateral ganglionic eminences (Flames et al. 2007). Next, the expression levels of *FOXG1*, a forebrain fate marker (Yuan et al. 2015) were evaluated. As shown in Figure 3.3B, *FOXG1* is extensively upregulated upon neural induction at day 10 of neural differentiation (218-fold increase). On day 24, the upregulation became more prominent as FOXG1 is required for patterning of the ventral telencephalon (Manuel et al. 2010).

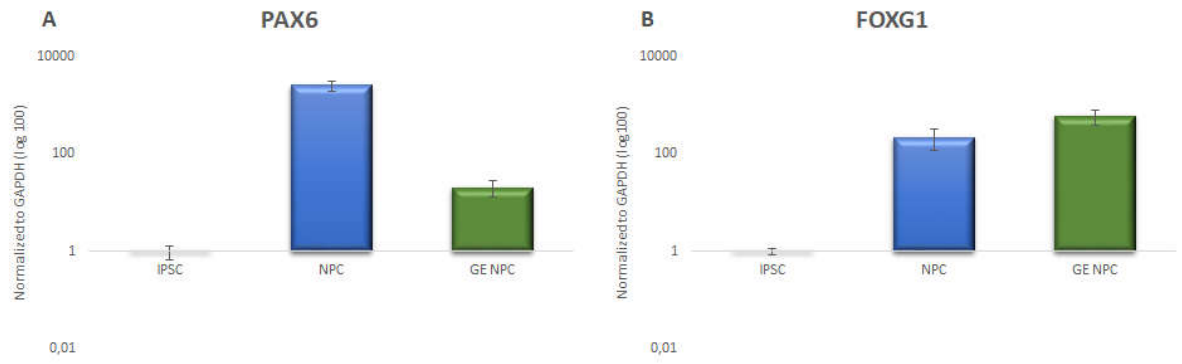


Figure 3.3: The expression dynamic of *PAX6* (A) and *FOXG1* (B) evaluated by qPCR during differentiation, at three different time points: IPSC - HMGU12 induced pluripotent stem cell line; NPC - HMGU12 derived neural progenitor cell line; GE NPC - HMGU12 derived ganglionic eminences-like progenitor cell line. The values are relative to the induced pluripotent stem cell line (log scale). N=6 biological replicates per cell clone.

Next, to evaluate the effect of purmorphamine and successful ventralization of the cells towards ganglionic eminences fate, I measured mRNA expression level of two marker genes of ganglionic eminences- *GSX2* and *NKX2-1*. More precisely, *GSX2* encodes for transcriptional factor required for specification of lateral ganglionic eminence progenitor cells (Pei et al. 2011). The upregulation was significant at day 24 (900-fold increase) compared to the other two stages which confirms that cells underwent a transition to ventralized precursors (Figure 3.4A). NKX2-1 is a transcriptional factor highly expressed in medial ganglionic eminences (MGE), involved in the migration of MGE-derived interneurons (Nóbrega-Pereira et al. 2008; Butt et al. 2008). Being almost undetectable in iPSC and NPC, *NKX2-1* became notably expressed at day 24 (24000-fold change) as the cells adopted ganglionic eminence identity (Figure 3.4B).

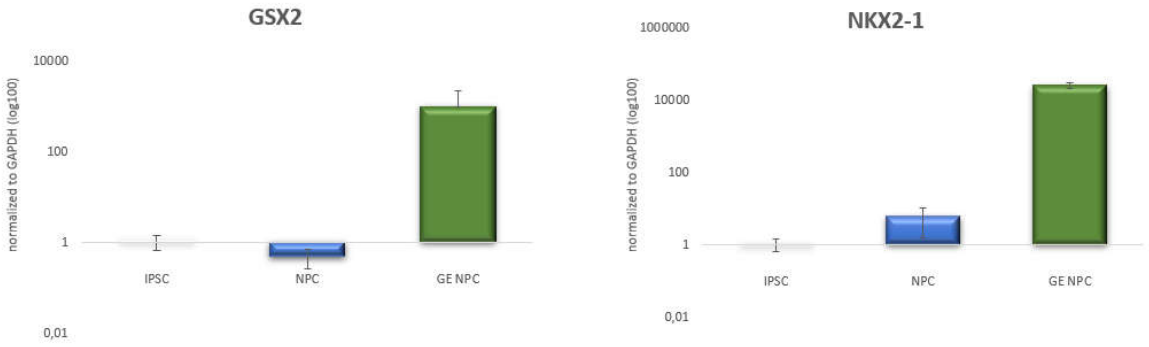


Figure 3.4: The expression dynamic of *GSX2* (A) and *NKX2-1* (B) evaluated by qPCR during differentiation, at three different time points: IPSC - HMGU12 induced pluripotent stem cell line; NPC - HMGU12 derived neural progenitor cell line; GE NPC - HMGU12 derived ganglionic eminences-like progenitor cell line. The values are relative to the induced pluripotent stem cell line (log scale). N=6 biological replicates per cell clone.

Finally, the expression levels of *MEIS1* and *MEIS2* genes at the respective differentiation stages were investigated. As previously discussed, *MEIS1* encodes a transcriptional factor with a pleiotropic role in embryonic development, especially cardiogenesis, haematopoiesis, neurogenesis, and limb patterning (Schulte and Geerts 2019). The expression level of *MEIS1* was significantly increased upon neural induction (192-fold change). This observation is in line with previously published results of MEIS1/Meis1 expression in human and mouse neural stem cells (Maisel et al. 2007; Barber et al. 2013). Upon ventralisation, the *MEIS1* mRNA level dropped compared to the previous stage but still significantly upregulated compared to the pluripotent stage (Figure 3.5A). It is important to mention that Meis1 is expressed in the mantle zone of ganglionic eminences (Toresson, Parmar, and Campbell 2000). Furthermore, RLS-associated variant rs12469063 was found to alter the expression activity of the HCNR 617 in ganglionic eminences of the mouse (Spieler et al. 2014). *MEIS2* gene is a homolog of *MEIS1* having to some extent an overlapping expression pattern and presumably similar function (Schulte and Geerts 2019). The *MEIS2* expression was notably elevated in neural progenitor compared to iPSC. Upon purmorphamine treatment, mRNA level became upregulated resulting in a 1500-fold change compared to iPSC stage and a 7-fold change compared to neural progenitors (Figure 3.5B). One study reports high Meis2 expression in the subventricular zone of lateral ganglionic eminences of the developing mouse embryo (Toresson, Parmar, and Campbell 2000). Moreover, studies that used *in vitro* generation of medium spiny neurons, originating from the LGE structure, confirmed high levels of *MEIS2* in LGE -like neural progenitors (Ma et al. 2012; Yuan et al. 2015).

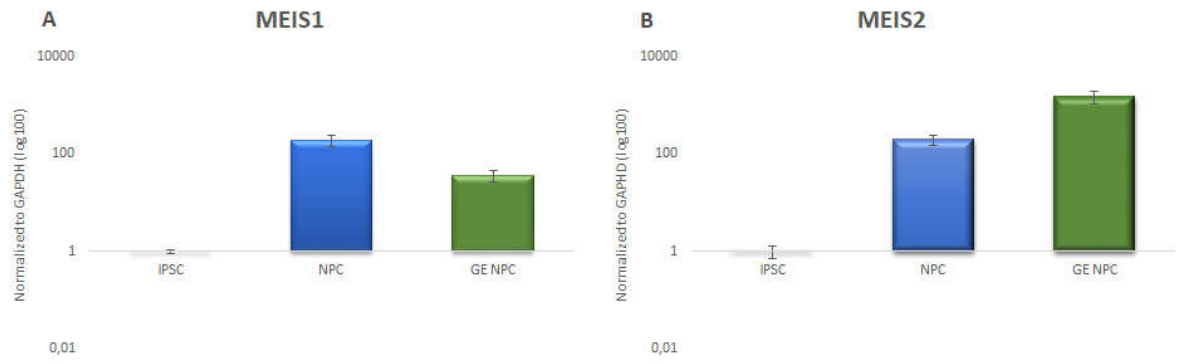


Figure 3.5: The expression dynamic of *MEIS1* (A) and *MEIS2* (B) evaluated by qPCR during differentiation, at three different time points: IPSC - HMGU12 induced pluripotent stem cell line; NPC - HMGU12 derived neural progenitor cell line; GE NPC - HMGU12 derived ganglionic eminences -like progenitor cell line The values are relative to the induced pluripotent stem cell line (log scale). N=6 biological replicates per cell clone.

3.2 Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) reveals cell-specific regulatory features of RLS-associated regulatory elements

To examine the regulatory landscape of the *MEIS1* locus and how RLS risk variants contribute to its regulation, I looked at chromatin accessibility by employing assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq). This technology was developed in 2013 (Buenrostro et al. 2013) and promptly emerged as one of the most frequently used methods for chromatin investigation. The advantage of this method is simplicity, low material input, and time efficiency. It is frequently combined with other high-throughput technologies such as ChIP -seq, RNA -seq, and Hi-C providing a possibility for an integrative multi-omics approach to study regulatory networks in a given cell type or tissue.

In order to map the chromatin state of the *MEIS1* locus during neural differentiation, we generated data sets for 6 different cell types, for each cell line two replicates: induced pluripotent stem cell line HMGU12, neural progenitors derived from HMGU12 (NPC), and ganglionic eminences progenitors derived from HMGU12 (GE NPC), neural stem cell GIBCO® Human Neural Stem Cells (H9-derived), gabaergic (iCell® GABANeurons) and glutamatergic neurons (iCell® GlutaNeurons). Samples were sequenced on a Illumina HiSeq 4000 sequencing platform, with 2x100 paired-end sequencing, two samples per lane.

In Table 3.1 the basic quality control metrics for each sample are shown, such as the number of total reads, percentage of properly mapped and paired reads. The library complexity is evaluated by non-redundant Fraction (NRF), PCR Bottlenecking Coefficients 1 (PBC1), and PCR Bottlenecking Coefficients 2 (PBC2). One of the parameters in the table is the Transcription Start Site (TSS) enrichment score which represents a signal-to-noise ratio (Dunham et al. 2012). With the exception of one sample (HMGU12 iPSC, replicate 1 where TSS score was 4.59 which is just below the ENCODE threshold for acceptable value), all other samples were in line with ENCODE ATAC-seq data quality standards (https://www.encodeproject.org/atac-seq/).

ATAC-seq QC-items	*HMGU12_ips_1*	*HMGU12_ips_2*	*HMGU12_NPC_1*	*HMGU12_NPC_2*	*HMGU12_LGE_1*	*HMGU12_LGE_2*	*hNSC_1*	*hNSC_2*	*GLUTA_1*	*GLUTA_2*	*GABA_1*	*GABA_2*
Total	64041194	61663377	52908974	50400648	50178947	61662729	5572344	60638564	341994144	543962066	38354170	47724534
	1	3	4	3	8	1	25	7		9	9	9
% Mapped	99.72	99.77	99.66	99.68	99.61	99.64	99.84	99.66	99.24	99.27	99.47	99.49
% Properly Paired	64.24	59.14	75.05	74.62	68.12	69.81	99.39	99.15	70.07	64.55	69.84	76.54
NRF = Distinct/Tot	0.835331	0.710346	0.918846	0.923788	0.882257	0.890947	0.79791	0.859098	0.734347	0.857991	0.885839	0.909263
						5						
al PBC1 = OnePair/D	0.939979	0.920122	0.949911	0.952257	0.948781	0.942548	0.82148	0.890133	0.775446	0.913224	0.920438	0.929978
						8						
istinct PBC2 = OnePair/T	17.14908	12.776316	20.711555	21.662732	20.489201	18.085948	5.50710	9.155275	4.627575	12.072885	13.397365	14.906046
	1					7						
woPair TSS_enric hment	4.594541	8.5420742	7.9860283	7.0728600	13.259717	12.149027	23.0920	13.14637	23.997256	22.6708715	23.034537	18.173549
	671	3	85	73	57	19	6836	304	8	3	69	28

Table 3.1 Basic quality parameters for ATAC-seq

The correlation of replicates by calculating the Cohen's Kappa coefficient for each cell line was evaluated as shown in Table 3.2.

SAMPLE	*KAPPA VALUE*
HMGU12 induced pluripotent stem cell line	0.596367172
HMGU12-derived neural progenitor cell line	0.804902649
HMGU12-derived ganglionic eminences-like progenitor cell line	0.79511461
NSC-H9	0.81638432
GABA neurons	0.777675885
GLUTA neurons	0.753464388

Table 3.2: Cohen's Kappa correlation between two replicates for each cell line

We profiled chromatin accessibility at different stages of neural development using a cellular model of differentiation. For the purpose of this project, I focused on the *MEIS1* locus although with this method we can evaluate all RLS candidate genes. The first three cell lines in Figure 3.6 are models for the generation of ganglionic eminences (GE) -like progenitors. As previously stated, we used induced pluripotent stem cell line HMGU12 and differentiated it into neural progenitors and further into ganglionic eminences progenitors.

A promoter accessibility has a very dynamic pattern during differentiation. At the pluripotent state, the promoter has very low accessibility compared to neural progenitors and GE-like progenitors. This might be due to the fact that the *MEIS1* is weakly expressed at the pluripotent cell stage. Moreover, intronic regions showed no features of activity in iPSC cells whereas the other two stages of neural progenitors have a distinct pattern of accessibility. Furthermore, we observed a progressive increase in chromatin accessibility at RLS-associated regulatory element 617 in neural progenitors and GE-like progenitors. In addition, two potentially regulatory elements, located in intron 7, displayed prominent ATAC-seq peaks - one in neural progenitor cell line and the other one in GE-like progenitors. These elements are however, not associated with RLS but might strongly contribute to *MEIS1* expression. Furthermore, I assayed commercially available cell lines: H9-derived neural stem cell line and two types of neurons, glutamatergic and GABAergic. As seen in Figure 3.6, these cell lines had a distinct accessibility pattern. In the H9-derived neural stem cell line, open chromatin was detected at RLS associated HCNR 617 as well as in intron 7, which was observed in iPSC-derived neural progenitors and GE-like progenitors. This similarity in

accessible chromatin suggests the resemblance in the phenotype of these cell lines. Furthermore, two types of neurons have differential accessibility pattern at the *MEIS1* locus, which suggests that the *MEIS1* regulatory network is highly cell specific. Both neuron types were highly accessible in the promoter region, whereas substantial differences are observed in intronic regions. Excitatory neurons had a weak, yet detectable signal at HCNR 617 as well as in intron 7, previously described in neural progenitors and neural stem cells. The gabaergic neurons displayed one unique feature of regulation compared to all other assayed cell types. This feature is open chromatin at HCNR 602 which harbours the strongest RLS associated signal, a risk variant rs11385554. Besides, prominent signals were found at HCNR 617 and at the element in intron 7. Another accessible region in inhibitory neurons was detected in intron 6. Finally, I concluded that RLS-associated conserved elements 602 and 617 were highly active in gabaergic neurons which strongly supports their role in *MEIS1* regulation in the respective cell types. This finding also prioritizes inhibitory neurons as cell type where *MEIS1* dysregulation could contribute to the molecular mechanism of RLS.

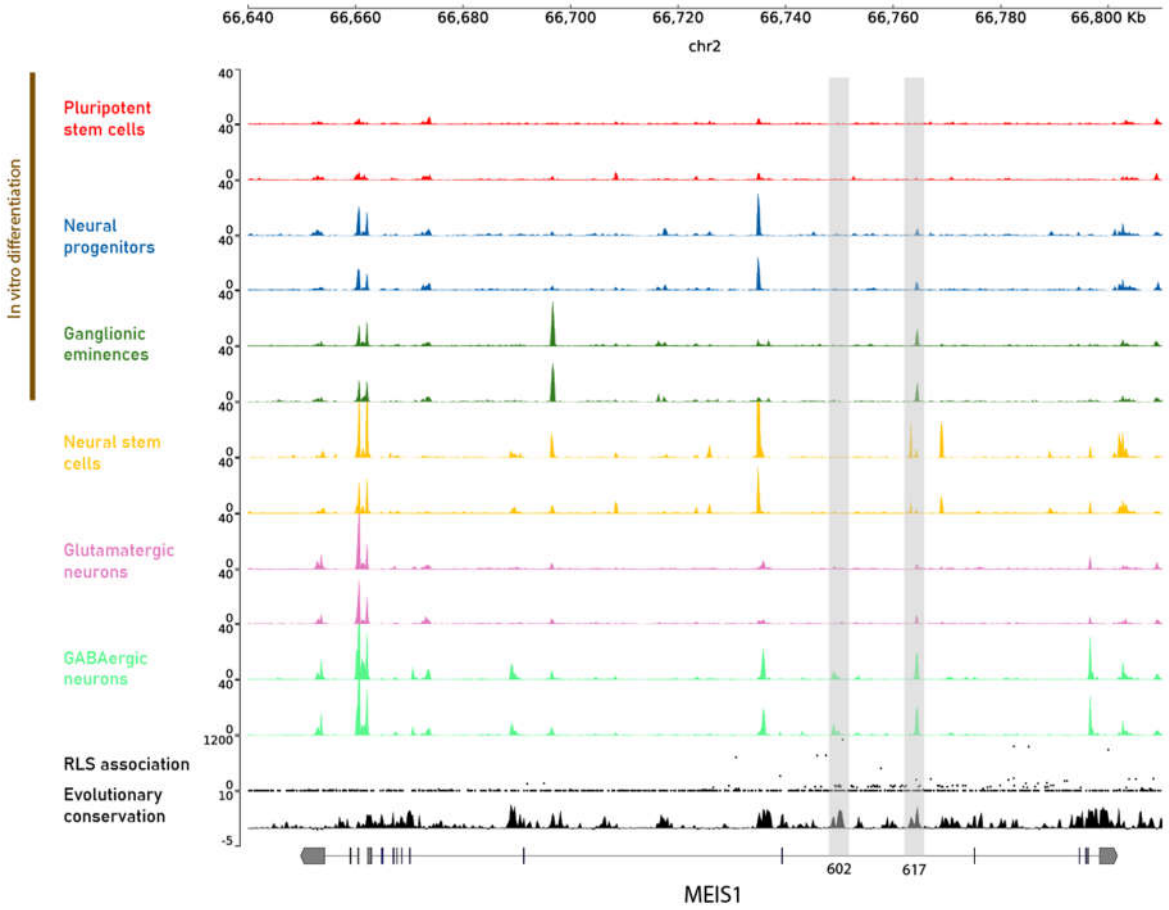


Figure 3.6: ATAC-seq signal (-log10 P-value) at *MEIS1* locus in six different cell types: Pluripotent stem cells - HMGU12 induced pluripotent stem cell line; Neural progenitors - HMGU12 derived neural progenitor cell line; Ganglionic eminence - HMGU12 derived ganglionic eminences-like progenitor cell line; Neural stem cells - H9-derived neural stem cells; Glutamatergic neurons - Excitatory neurons; GABAergic neurons - Inhibitory neurons. RLS association is from unpublished metaGWAS, the scale is -log10P. Evolutionary conservation is phyloP 100-way (Pollard et al. 2010). Both replicates are shown. Conserved elements 602 and 617 are highlighted in gray.

Overall, the observed accessibility dynamics at *MEIS1* in these cell lines correlates with expression levels. As seen in Figure 3.7, the highest mRNA expression of *MEIS1* is detected in inhibitory neurons which reflects complex regulation in this cell type.

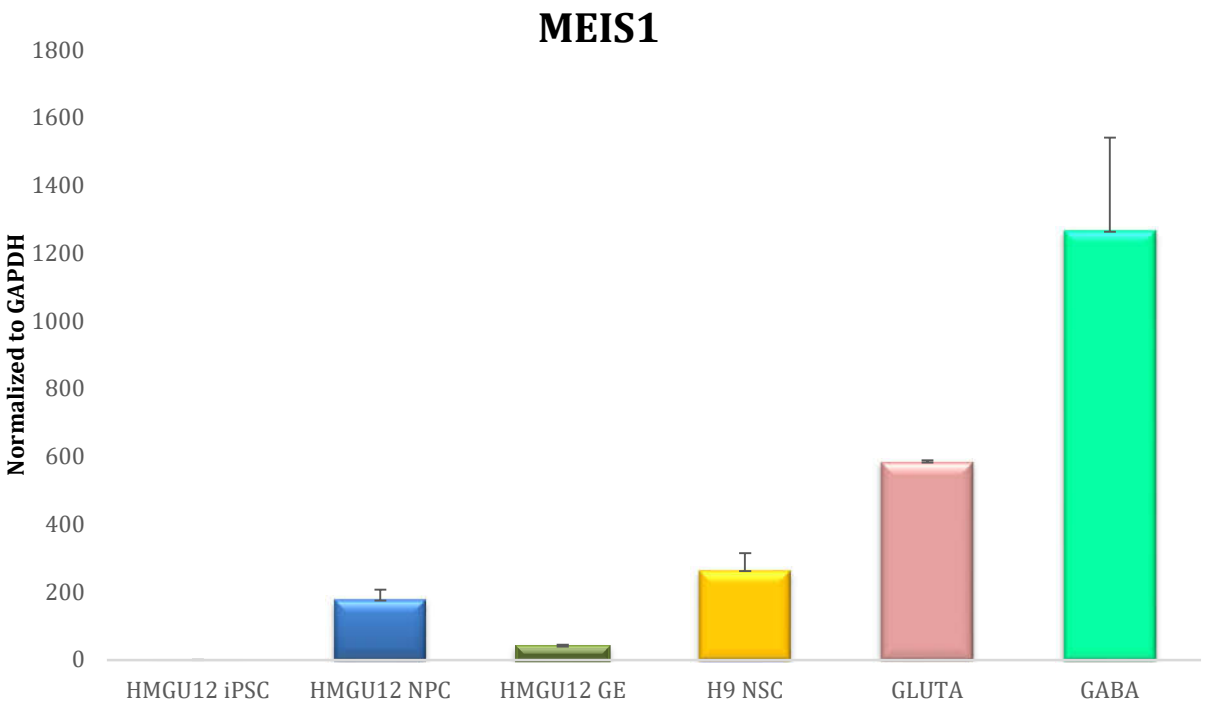


Figure 3.7: *MEIS1* mRNA expression evaluated by qPCR in six different cell types- HMGU12 iPSC - HMGU12 induced pluripotent stem cell line; HMGU12 NPC - HMGU12 derived neural progenitor cell line; HMGU12 GE - HMGU12 derived ganglionic eminences -like progenitor cell line; H9 NSC- H9-derived neural stem cells; GLUTA- Excitatory neurons; GABA-Inhibitory neurons. The values are relative to H9-derived neural stem cells. N=3-6 biological replicates per cell clone.

In order to further evaluate the complexity of the *MEIS1* regulatory network, and compare our findings with other datasets, I investigated open chromatin from publicly available datasets from different studies. Hereby, I show a collection of available datasets, carefully selected to reveal cell types and time points where RLS associated regulatory elements exhibit features of activity. In the section Appendix, the list of all publicly data sets with a link to the source can be found. As shown in Figure 3.8 we looked at *MEIS1* locus in different cell lines and tissues assayed for chromatin accessibility by two methods: ATAC-seq and DNAse-seq. We also explored ChIP-seq data sets. First track represents DNA-seq for H9-derived neural progenitors. As seen before in neural progenitors derived from HMGU12 and neural stem cell derived from H9, there was a signal in HCNR 617 and in intron 7. This consistent activity of HCNR 617 throughout a different dataset strongly supports the paradigm that this element is active at neural progenitor stage.

To evaluate the regulatory landscape revealed in profiled inhibitory and excitatory neurons, we have explored ENCODE ChIP-seq datasets (Dunham et al. 2012; Davis et al. 2018) for *in vitro* differentiated neurons. We visualized ChIP-seq data for H3K27ac, a marker of active

enhancers (Creyghton et al. 2010) and EP300 protein, a transcriptional activator, frequently bound on the enhancers (Visel et al. 2009). These datasets confirm the regulatory nature of the HCNR 602 harbouring the lead RLS SNP rs113851554, in neurons. Furthermore, an observed EP300 peak at HCNR 617 supports the enhancer activity at neuronal stage.

In addition, we explored two ATAC-seq datasets of human tissue from the developing brain - fetal LGE (Markenscoff-Papadimitriou et al. 2020) and post-mortem adult brain tissue - putamen (Fullard et al. 2018). The first data set is generated from dissected human lateral ganglionic eminences (LGE) at gestational week 19, a transient structure that yields striatal medium spiny neurons. The second dataset is generated from neurons dissected from putamen, a brain structure that is part of the striatum. The striatum is part of the motoric system, where most of the neurons (95%) are medium spiny neurons, a type of inhibitory neurons (Yager et al. 2015). The accessibility pattern of human lateral ganglionic eminences suggests the activity of HCNR 617 at that stage. Furthermore, open chromatin is observed in HCNR 602 in both human datasets, again supporting the notion that this element contributes to the *MEIS1* regulatory network in a highly specific temporospatial fashion, as a developmental and adult enhancer. Finally, we explored datasets in the non-neural cell lines where *MEIS1* is highly expressed. As mentioned, *MEIS1* plays an important role in cardiogenesis, so we explored the available DNAse-seq dataset of the fetal heart (Meuleman et al. 2020). We also included *in vitro* differentiated cardiac muscle cells. Since the role of *MEIS1* in haematopoiesis is very well established (Unnisa et al. 2012), we examined the chromatin accessibility of *MEIS1* locus in the hematopoietic progenitor cell line. In the first two datasets, there is an accessibility peak at HCNR 617. Whether this element and to what extent contributes to *MEIS1* regulation in cardiogenesis remains yet to be explained. On the other hand, HCNR 602 stayed inaccessible further strengthening the premise of cell-specific activity.

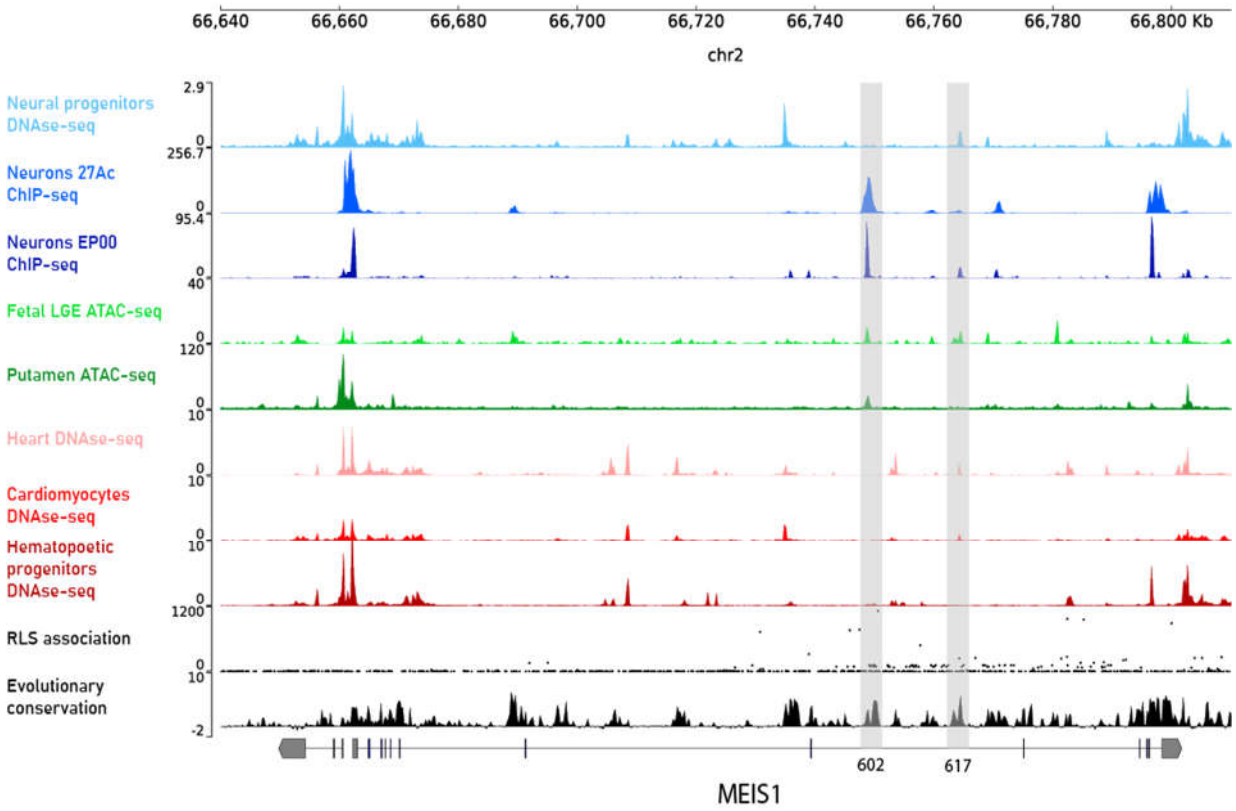


Figure 3.8: Publicly available datasets for DNAse-seq (read depth-normalized signal) ChIP-seq and ATAC-seq (-log10 P-value) at *MEIS1* locus form eight datasets. The source of datasets is listed in Appendix (Dunham et al. 2012; Davis et al. 2018; Markenscoff-Papadimitriou et al. 2020; Fullard et al. 2018). RLS association is from unpublished metaGWAS, the scale is -log10P. Evolutionary conservation is phyloP 100-way (Pollard et al. 2010). Conserved elements 602 and 617 are highlighted in gray.

We took advantage of genome-wide datasets and decided to search for regulatory features in another RLS GWAS locus on chromosome 2, located in the intergenic region 1.3 Mb downstream of the *MEIS1*. This locus was firstly identified in GWA study from 2011 (Winkelmann et al. 2011) and again confirmed in a meta-study (Schormair et al. 2017). The association is detected in 120 KB LD block, tagged with rs1820987 $p=8.22\times10^{-147}$ (unpublished metaGWAS). There are several genes in the surrounding area such as *C1D, ETAA1,* and *MEIS1*. It has been proposed that the RLS signal in this region could have a long-range regulation *MEIS1* (Winkelmann et al. 2011). Even though *C1D* and *ETAA1* reside linearly closer to the region encompassing RLS-associated risk variant, it is observed, in some studies, that regulatory elements form contacts and regulate more distant genes rather than closer ones (Smemo et al. 2014). We visualized ATAC-seq data (Figure 3.9) of the genomic region tagged by rs1820987 in the HMGU12 line and its derived cell lines and also in commercial neural stem cells and neurons. We highlighted the genomic region where the tag SNP resides. It is characterized by a high degree of conservation but in this data set, there

are no peaks of accessibility. However, upstream and downstream there are two conserved genomic elements displaying features of accessibility, in neural stem cells and neurons. Since these two conserved regions are part of the 120 kb LD block, they could potentially carry the causal RLS- risk variant. Whether this risk locus regulates *MEIS1* remains to be investigated.

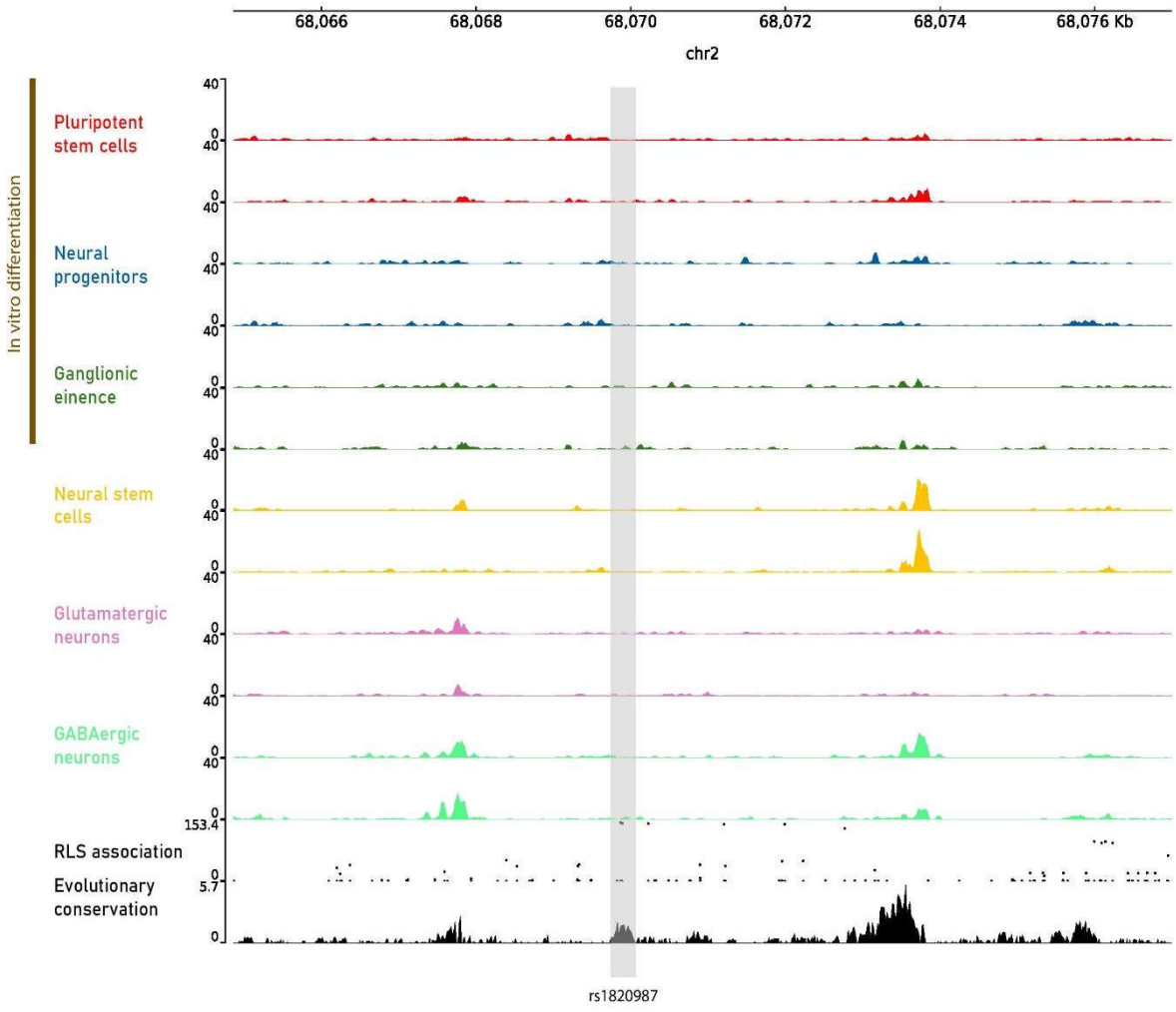


Figure 3.9: ATAC-seq signal (-log10 P-value) at intergenic RLS risk locus on chromosome 2, tagged with rs1820987 in six different cell types: Pluripotent stem cells - HMGU12 induced pluripotent stem cell line; Neural progenitors - HMGU12 derived neural progenitor cell line; Ganglionic eminence - HMGU12 derived ganglionic eminences -like progenitor cell line; Neural stem cells - H9-derived neural stem cells; Glutamatergic neurons - Excitatory neurons; GABAergic neurons - Inhibitory neurons. RLS association is from unpublished metaGWAS, the scale is -log10P. Evolutionary conservation is phyloP 100-way (Pollard et al. 2010) Both replicates are shown. The tag SNP rs1820987 is highlighted in gray.

MEIS1 is highly pleiotropic, with central roles in limb, vascular, cardiac, and neural development. Accordingly, the *MEIS1* locus contains many highly conserved non-coding regions (HCNRs), presumably acting as *cis*-regulatory elements to confer refined spatio-

temporal control of *MEIS1* expression in the respective developmental contexts. Among them, HCNR 602 and 617 showed high conservation in mice (Spieler et al. 2014). In order to precisely evaluate spatio-temporal dynamic of *Meis1* regulation and access the accessibility of HCNR 602 and 617, we collected ATAC-seq datasets of developing mice released on ENCODE. We grouped the available data into two categories – neural and non-neural tissues and evaluated the open chromatin at *Meis1* locus. Figure 3.10 represents a collection of the data in three different brain regions: forebrain, midbrain, and hindbrain. Each region was assayed at seven developmental time points starting from embryonic day E11.5 until E16.5 and on postnatal day 0. From these assays, we observed that *Meis1* is active throughout different brain regions and different stages as promoter exhibit prominent ATAC-seq peaks.

In the forebrain region, *Meis1* is expressed from E10.5 and exhibits a highly dynamic pattern. High levels of *Meis1* are observed in the caudal ganglionic eminences and in the mantle zone of lateral and medial ganglionic eminences (Toresson et al. 2000). By looking at the forebrain development, we see that HCNR 617 is highly accessible in all time points from E11.5 until E16.5 and the accessibility reduces at postnatal day 0. This finding speaks in favour of reported enhancer activity of this element in mouse developing forebrain (Spieler et al. 2014). Furthermore, we observed open chromatin at HCNR 602 starting from E12.5 and persisting until postnatal day 0. It is possible that 602 and 617 enhancers cooperate during forebrain development. Next, we evaluated the regulatory landscape in the midbrain. The expression of *meis1* is detected in the midbrain region of zebrafish between 11 and 15 hours postfertilization. Moreover, *meis1* is detected in the dorsal midbrain which gives rise to optic tectum. On that note, *meis1* ablation disrupted proper retinotectal development (Erickson et al. 2010). One study evaluated *Meis1* expression in mouse developing telencephalon (E11.5) by whole mount *in situ* hybridization and found a weaker expression pattern of *Meis1* in midbrain compared to forebrain and hindbrain (Toresson, Parmar, and Campbell 2000). We observed a small yet detectable peak at 617 at stages E11.5, E12.5, and E13.5 which subsequently disappear which might be due to transient activity of this regulatory element. The element 602 remained silent in this brain region.

Next, we looked at open chromatin peaks in the hindbrain region of the developing mouse. In the zebrafish, *meis1* expression in the hindbrain region is detected first at 11 hpf and stays detectable even at 50 hpf (Erickson et al. 2010). In the developing mouse, *Meis1* is detected at a high level in the hindbrain (Toresson et al. 2000). The *Meis* genes are required for the proper hindbrain segmentation. An *in vivo meis1* knockdown in zebrafish disrupted proper hindbrain patterning (Waskiewicz et al. 2001). By evaluating chromatin accessibility of

hindbrain in developing mouse, we observed a transient peak at HCNR 617 around E12.5 and E13.5 which shows a strict temporal window of activity. At HCNR 602 there is a small, transient peak potentially involved in *Meis1* regulation during hindbrain development.

The role of *MEIS1* in telencephalon is complex and precisely controlled by a network of regulatory elements. Some of these regulatory elements are found to be associated with RLS. How these elements regulate *MEIS1* and how they contribute to RLS, remains to be elucidated.

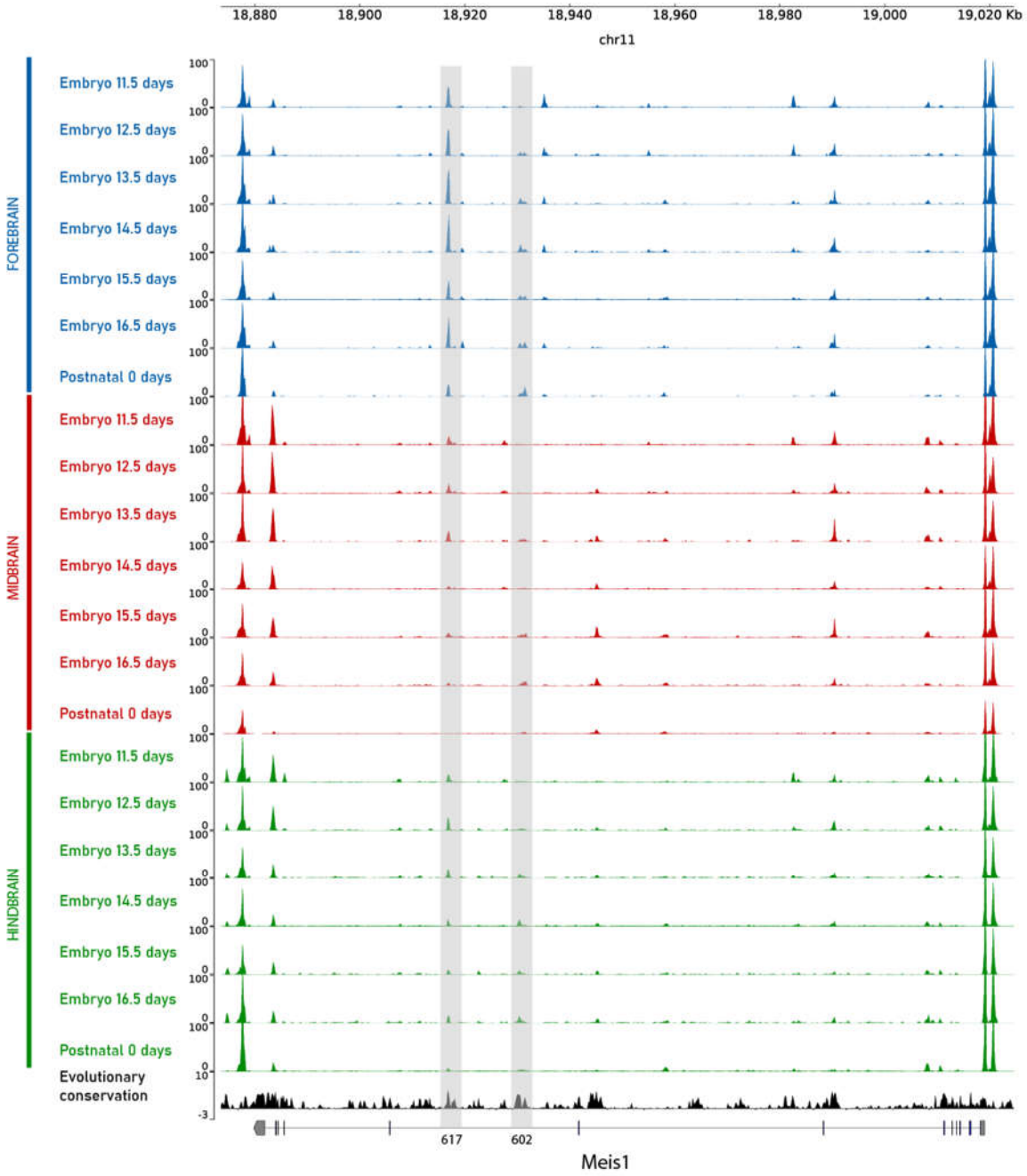


Figure 3.10: Chromatin accessibility assayed by ATAC-seq (-log10 P-value) in the murine *Meis1* locus. The source of datasets (ENCODE) is listed in Appendix (Dunham et al. 2012; Davis et al. 2018). Datasets are from the mouse forebrain, midbrain, and hindbrain at different developmental stages (E11.5-Postnatal day 0) for two replicates. Evolutionary conservation is phyloP60-way (Pollard et al. 2010). Note: HCNR 602 and 617 have inverted orientation in the mouse genome.

The role of *MEIS1* in embryogenesis and organogenesis is very complex and its expression is tightly regulated in a spatiotemporal manner. We sought to determine the accessibility profile and thereby potential regulatory network of *MEIS1* in different tissue such as the heart, limbs, and liver, where MEIS1 was shown to have a relevant role. Moreover, we wanted to see if

RLS associated regulatory element exhibit properties of activity. We extracted data from ENCODE (Dunham et al. 2012; Davis et al. 2018) for mouse heart, limb, and liver between E11.5 and 16.5 and postnatal day 0 and presented in Figure 3.11.

In heart development, the role of *MEIS1* is extensively investigated. In the mouse embryos, *Meis1* is expressed in the anterior heart field and outflow tract between E8.5 and E9.5 (Dupays et al. 2015). *Meis1*-null embryos displayed cardiac abnormalities, including defects in the outflow tract (Stankunas et al. 2008). Furthermore, a targeted ablation of *Meis1* in cardiomyocytes revealed a function of Meis1 in cardiac cell cycle arrest (Mahmoud et al. 2013). One study employed in vitro differentiation of human embryonic stem cells into cardiomyocytes and found the peak of *MEIS1* expression at day 5 of differentiation, suggesting involvement in early cardiac development (den Hartogh et al. 2016). In addition, GWAS identified an association of *MEIS1* with cardiographic PR interval (Pfeufer et al. 2010). Open chromatin at *Meis1* locus of developing mouse heart has a unique pattern compared to limb and liver but RLS associated regulatory elements 617 and 602 seemed to be inactive at these stages, as no accessibility peaks were detected. However, it is worth mentioning that HCNR 617 is accessible in human datasets of the fetal heart (day 96) and *in vitro* differentiated cardiomyocytes shown in Figure 3.8. This observed phenomenon could be explained by functional divergence of enhancers.

Furthermore, the Meis genes are implicated in limb development. *Meis1* expression is restricted to the proximal domain of developing limbs, a pattern necessary for correct proximodistal specification of the limbs. (Mercader et al. 1999). An ectopic expression of Meis1 eliminates the structure of distal limbs converting them into proximal identity (Mercader et al. 2009). A more recent study showed that the elimination of Meis genes during limb development leads to underdeveloped limbs, a phenotype resembling a human congenital condition called phocomelia (Delgado et al. 2020). As seen in Figure 3.10, the ATAC-seq peak is restricted at element 617 in two-time points, E11.5, and E12.5. This suggests that HCNR 617 is transiently active and potentially regulates *MEIS1* expression during limb development in strictly controlled time points. On the other hand, HCNR 602 had no accessibility peaks during limb development.

Finally, we explored datasets obtained from different stages of developing liver. It is known that *Meis1* has a crucial role in haematopoiesis and the liver is one of the sites of fetal haematopoiesis. Moreover, the expression of *Meis1* is confirmed in hematopoietic stem cells

in the fetal liver. In fact, *Meis1* homozygous knockout mice die due to failure in haematopoiesis (Azcoitia et al. 2005). As seen in Figure 3.11, the chromatin signature of *Meis1* in the fetal liver is very distinct compared to other tissues. Enrichment in the ATAC-seq signal is strong around the promoter site in all developmental stages. One prominent peak detected in the intron 6 with maximum accessibility at day E13.5 is possibly contributing to *Meis1* expression during a tightly controlled haematopoiesis establishment. On the other hand, there were no features of chromatin accessibility at RLS associated elements 602 and 617. This speaks in favour of differential regulation of *MEIS1* in different cell types and tissues.

In summary, two HCNR regions 602 and 617 associated with increased risk for RLS exhibit stage-specific and cell-specific accessibility in humans and mice. In humans, the enhancer 617 is active in the neural progenitor stage and in *in vitro* derived neurons. During mouse embryonic development, the enhancer is prominently accessible in the forebrain. Based on epigenetic assays, enhancer element 602 possesses accessible chromatin in mature neurons as well as in human developing lateral ganglionic eminences. In addition, the enhancer accessibility is seen in developing mouse forebrain. These findings support the view that HCNR 602 acts as a stage-specific developmental enhancer and adult enhancer. Taken together, the nature of RLS associated regulatory elements suggest that RLS is a neurodevelopmental disorder with adult-onset. In this view, subtle developmental effects could create a vulnerability to cumulative stress that can eventually trigger disease onset.

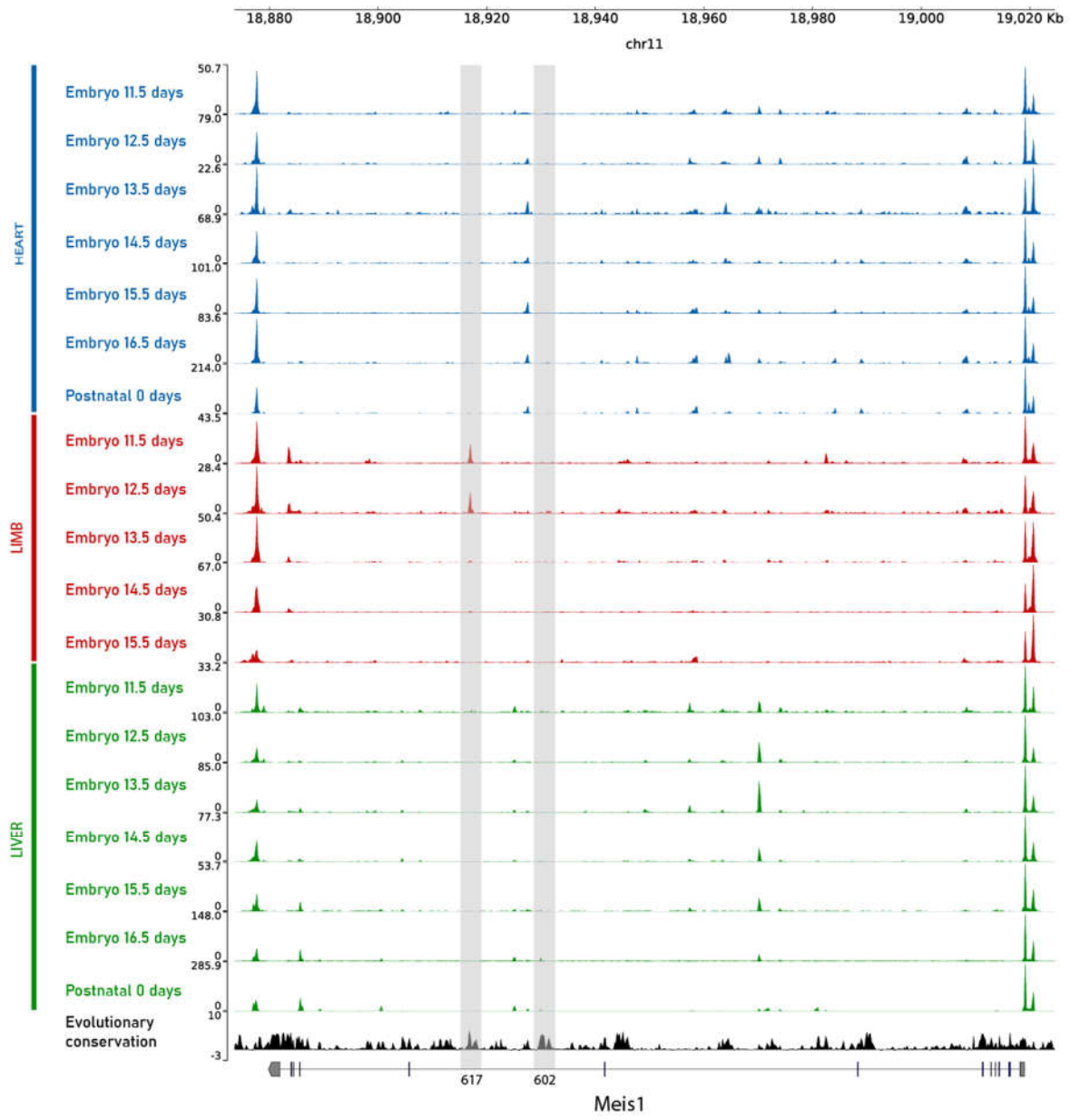


Figure 3.11: Chromatin accessibility assayed by ATAC-seq (-log10 P-value) in the murine *Meis1* locus. The source of datasets (ENCODE) is listed in Appendix (Dunham et al. 2012; Davis et al. 2018). Datasets are from mouse heart, limb, and liver at different developmental stages (E11.5-Postnatal day 0) for two replicates. Evolutionary conservation is phyloP60-way (Pollard et al. 2010). Note: HCNR 602 and 617 have inverted orientation in the mouse genome.

3.3 Chromosome conformation capture 4C-seq reveals cell-specific promoter-enhancer contacts

The newly emerged methods for the spatial organization of the genome such as 4C-seq and Hi-C allowed profiling interactions between genes and their regulators. The enhancers may

often reside at a great genomic distance from the genes they regulate, but when engaging in contact with their target genes, they are brought to spatial proximity. Furthermore, one regulatory element can contact multiple genes. Changes in enhancer sequence such as mutation or deletion could cause a change in gene expression and contribute to disease. Since many of the GWAS SNPs are located in noncoding regions they could have a role in gene expression (Schoenfelder and Fraser 2019).

To characterize the regulatory landscape of *MEIS1*, we performed 4C-seq using the *MEIS1* promoter as a viewpoint. Furthermore, we sought to determine if RLS associated variants within the *MEIS1* locus are in contact with the *MEIS1* promoter. For this experiment, we used three different cells line: HMGU1 induced pluripotent stem cell line, neural stem cell GIBCO® Human Neural Stem Cells (H9 -derived), and SH-SY5Y neuroblastoma cell line. By employing these cell lines, we wanted to examine *MEIS1* promoter interaction with potential regulatory elements at different stages of the neuronal specification. As mentioned in the methods section, two biological replicates were used for each cell line. The libraries were sequences on the Miseq platform, and the downstream analysis was performed by Basic4C-seq pipeline (Walter et al. 2014). We obtained between 1 and 2 million reads per replicate. The reads were distributed on chromosome 2 ranging from 63 % - 70%. Moreover, between 43% - 54% of the reads were mapped in a 500kb area surrounding the viewpoint as shown in Table 3.3.

CELL LINE	*VIEWPOINT*	*REPLICATE*	*READS TOTAL*	*MAPPED READS*	*READS AT CHR2*	*READS 500K CIS*
iPSC	MEIS1 pro	1	1193605	944428	605489	447495
iPSC	MEIS1 pro	2	1608705	1424289	901613	696435
hNSC	MEIS1 pro	1	1410868	1037570	664656	458235
hNSC	MEIS1 pro	2	1420318	1040808	689046	481922
SH-SY5Y	MEIS1 pro	1	2032582	1274993	930966	695932
SH-SY5Y	MEIS1 pro	2	1684094	1311410	900424	576214

Table 3.3: Basic sequencing parameters for circular chromosome conformation capture 4C-seq

The 4C-seq analysis showed a very distinct interaction pattern between the three cell types we used in this assay, as shown in Figure 3.12. At the pluripotent stage, the *MEIS1* promoter

interacted significantly only with the region at the 3'end region. No significant interactions were detected upstream or downstream of the gene body. This might be due to the fact that *MEIS1* is expressed at a very low level in induced pluripotent stem cells (based on qPCR) and according to ATAC-seq, has no accessible peaks surrounding the gene. This observed interaction spans from intron 9 until 3'UTR region which is marked by a high degree of conservation. Within this interacting region, there could reside regulatory elements that control the expression of *MEIS1* in a specific manner. On the other hand, in neural stem cells, the interaction pattern of the *MEIS1* promoter involved multiple genomic elements. Two significant interacting regions were detected within intron 7 which colocalize with ATAC-seq accessibility peaks in the respective cell line. Besides, the *MEIS1* promoter contacted RLS associated element 617 but there was no interaction with 602. These results are in line with chromatin accessibility data generated for this cell line. Another interaction region of the promoter was detected at the larger genomic portion, spanning from intron 8 to 3'UTR region. This region is characterized by the presence of conserved regions enriched for accessible chromatin which strongly suggests the involvement of this genomic locus in *MEIS1* regulation. Finally, we profiled an interaction pattern in the neuroblastoma cells line, and besides the contact at 3`end, there were no significant interactions with other intronic elements within *MEIS1*. Since there was a mutual contact with 3`end in all three cell lines, it is worth mentioning that at the 3' end there is a binding site of CTCF detected by ChIP-seq in different cell types (Dunham et al. 2012; Davis et al. 2018). This transcription factor is involved in the establishment and maintenance of chromatin architecture (Kentepozidou et al. 2020). This suggests that CTCF facilitates loop formation between the 5`and 3`end of the gene and brings in proximity enhancers and promoters (Kubo et al. 2021).

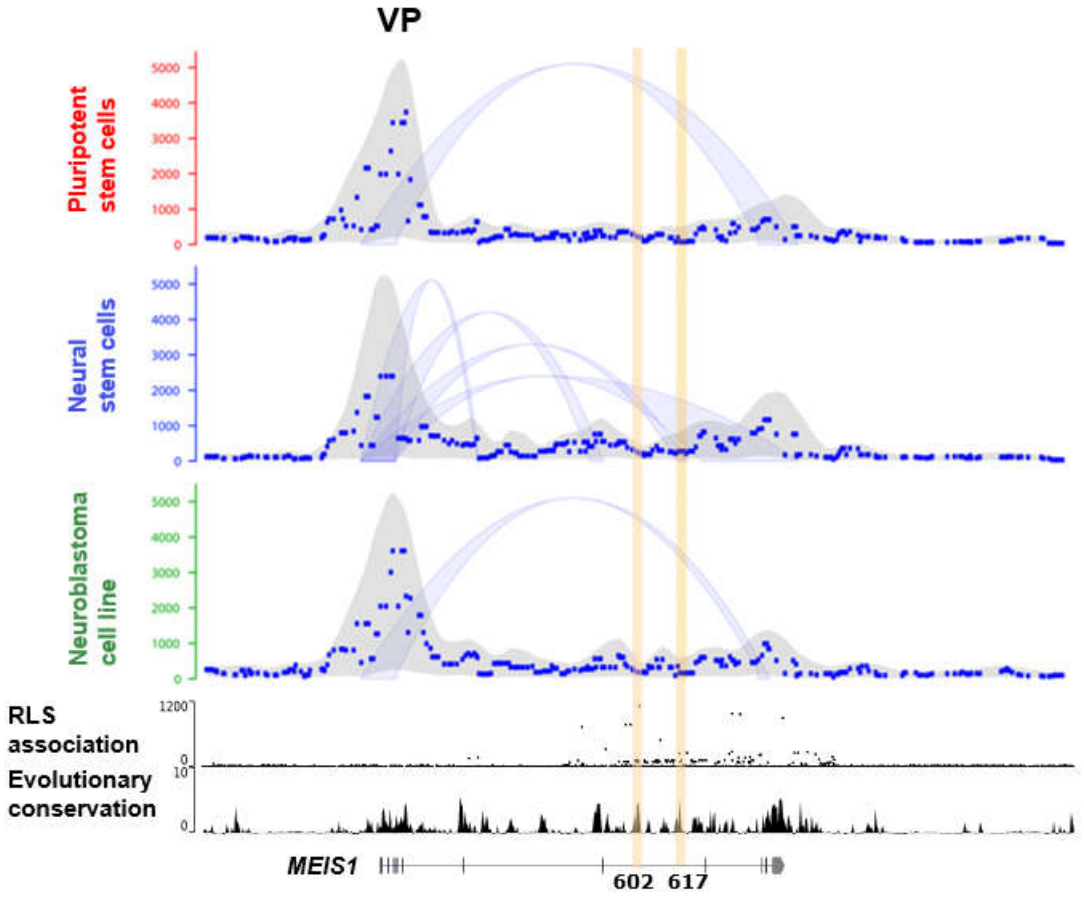


Figure 3.12: 4C-seq at *MEIS1* promoter as a viewpoint (VP). Fragment read count (blue dots) and significant interactions (blue arcs) in pluripotent stem cells - HMGU1 induced pluripotent stem cell line, neural stem cells- H9-derived neural stem cells, and neuroblastoma cell line- SH-SY5Y cell line. RLS association is from unpublished metaGWAS, the scale is -log10P. Evolutionary conservation is phyloP 100-way (Pollard et al. 2010). The positions of HCNR 602 and 617 are highlighted in yellow.

As mentioned, 4C-seq results showed an interaction between *MEIS1* promoter and 3´end in the neuroblastoma cell line. Also, there were two more distant interacting sites detected, one in the proximity of *ETAA1* gene and the other one located within the 120 LD block driving the RLS association signal located 1.3Mb downstream of *MEIS1* (Figure 3.13). The second long-range interacting site is located in the vicinity of tag SNP rs1820987. This finding is the first evidence supporting the hypothesis that this RLS-associated element could regulate *MEIS1*. Moreover, the interaction is specific for neuroblastoma cell type, and it is not detected in the neural stem cell line nor pluripotent stem cell line. The phenotype of the neuroblastoma cell line resembles immature neurons (Krishna et al. 2014) with the potential to differentiate to mature neurons upon treatment with retinoic acid (Påhlman et al. 1984). The expression of *MEIS1* is high in the SH-SY5Y neuroblastoma cell line (https://www.proteinatlas.org/ENSG00000143995-MEIS1/cell) (Uhlén et al. 2015).

However, it should be kept in mind that MEIS1 is implicated in neuroblastoma tumors (Spieker et al. 2001) so this type of cellular model of RLS should be carefully interpreted.

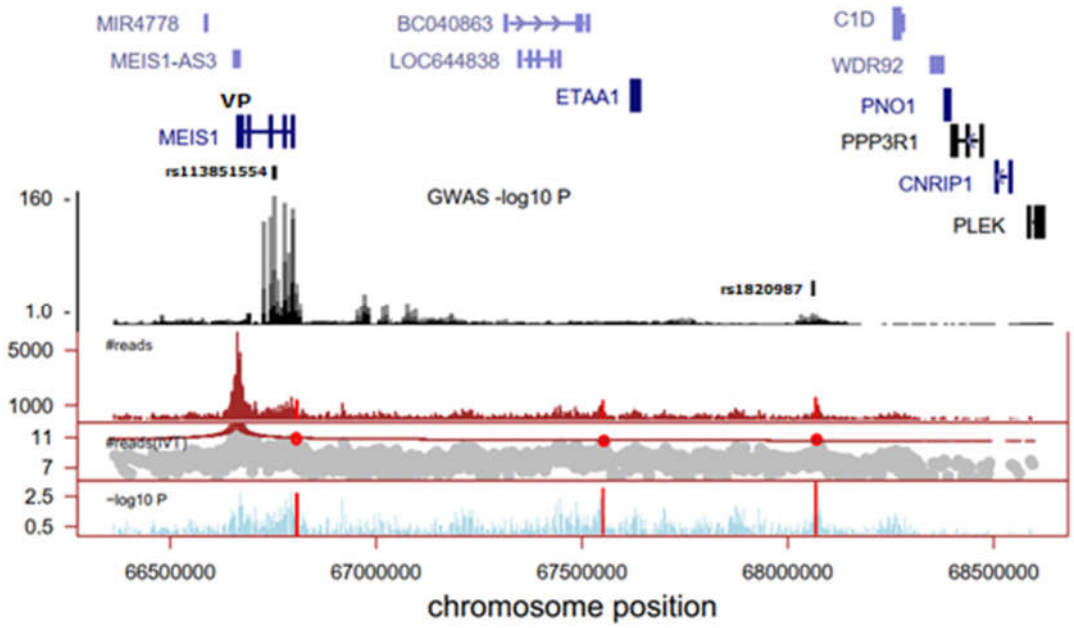


Figure 3.13: 4C-seq at MEIS1 promoter as a viewpoint (VP). 2MB area from the viewpoint is showed in respect to RLS GWAS signals (-log10 P value) from Schormair et al. 2017, marked with black vertical bars. The positions of the tag SNPs rs1138651554 and r1280987 are marked. The fragment counts are shown in dark red and significant interactions are presented as red vertical lines corresponding to the red dots in neuroblastoma cells (SH-SY5Y cell line) Two replicates are merged.

3.4 Targeted deletion of HCNR 617 reveals regulatory effect on *MEIS1* expression

To gain more insight into the regulatory function of HCNR 617 harboring RLS-associated rs12469063 on *MEIS1* expression, I deleted 922 bp fragment in intron 8 of *MEIS1*, including the above-mentioned variant as depicted in Figure 3.14.

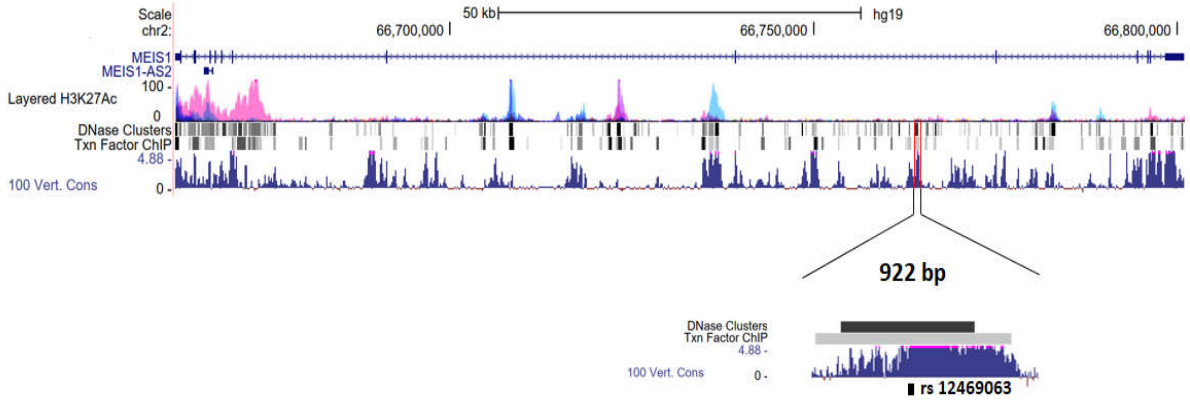


Figure 3.14: *MEIS1* gene. Position and size (922bp) of deleted fragment (HCNR 617) in intron 8 of *MEIS1*. Layered H3K27Ac from track from ENCODE (Dunham et al. 2012; Davis et al. 2018; Navarro Gonzalez et al. 2021) presents levels of enrichment of the H3K27Ac histone mark across the genome. DNase Clusters track from ENCODE (Dunham et al. 2012; Davis et al. 2018; Navarro Gonzalez et al. 2021) presents Dnase I Hypersensitivity Clusters in 125 cell types. Tnx Factor ChIP track from Encode presents Transcription Factor ChIP-seq Clusters (161 factors) (Dunham et al. 2012; Davis et al. 2018; Navarro Gonzalez et al. 2021). 100 Vert. Cons track corresponds to base wise conservation across 100 vertebrates (Pollard et al. 2010). Generated using UCSC Genome Browser, hg19 http://genome.ucsc.edu (Kent et al. 2002). The session URL: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2%3A66628913%2D66819331&hgsid=1135399453_l0PZM2YAWLSecgeJT5j1dhHZRsG1

I employed the CRISPR/Cas9 editing system for targeted deletion with two RNA guides flanking 922 bp long fragment in the human induced pluripotent stem cell lines. As described, two guides were co-nucleofected in iPSC cell line (HMGU1) and after clonal isolation and propagation, clonal populations were genotyped to confirm the deletion. To distinguish heterozygous and homozygous deletion of HCNR 617, two pairs of primers were designed. The first primer pair is positioned outside of deletion, designed to amplify 1499 bp DNA fragments in unedited clonal populations and 577 bp product if the editing was successful (Figure 3.15A). Another primer pair was positioned within the deletion fragment

to confirm the homozygous deletion (Figure 3.15B). For this experiment, I used five clonal cell lines in total, with correctly confirmed genotype. Two clonal cell lines had no deletion band, and three clonal cell lines had a deletion band. One of these three showed heterozygous deletion with wild type and deletion band present as shown Figure 3.15A. Two clonal cell lines had only a deletion band present. In the subsequent PCR, the homozygous deletion was confirmed using a second pair of primers, located within the deletion DNA fragment, as no amplification product was visualized on the gel, as seen in Figure 3.15B. We confirmed the correct deletion size by Sanger sequencing and the results are shown in Figure 3.15C.

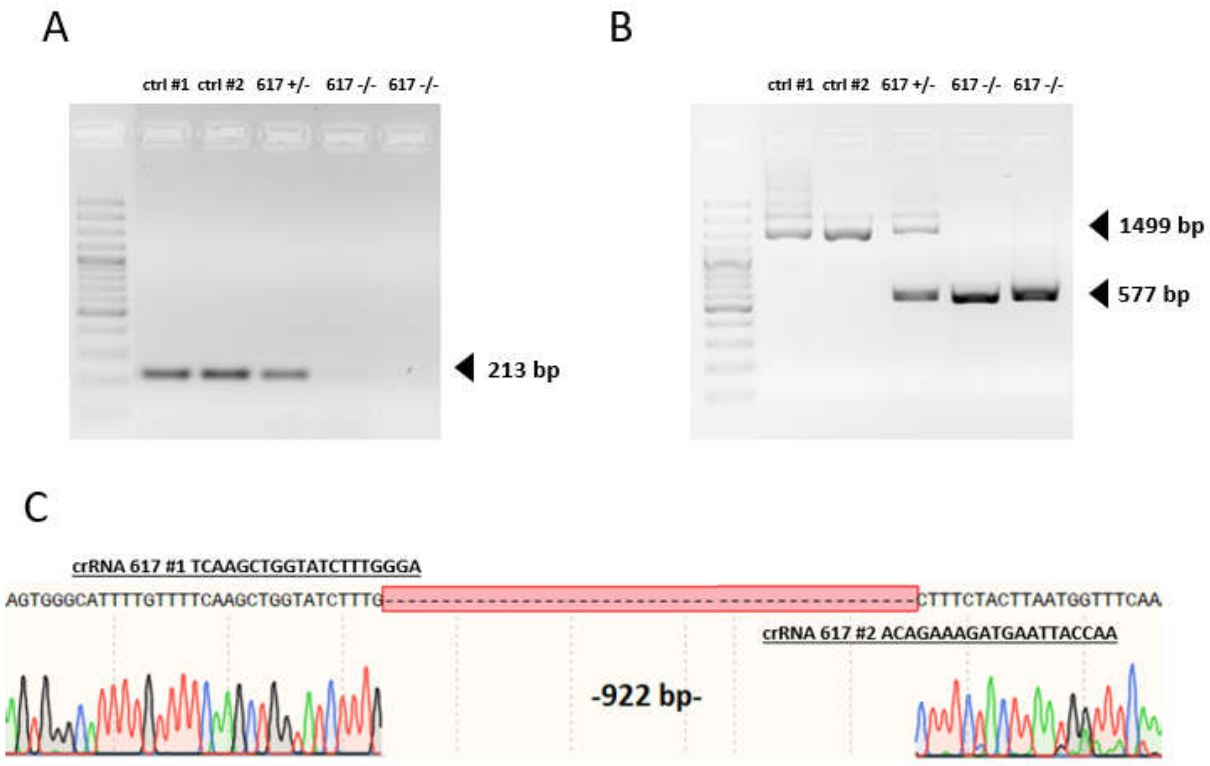


Figure 3.15: Confirmation of correct deletion of HCNR 617. PCR genotyping (A and B) shows two clones with no deletion (ctrl#1 and ctrl#2) and three clones with deletion. Representative Sanger sequencing results in deletion clone with an underlined sequence of crRNAs (C).

Cell lines were established and maintained on Geltrex coated plated in mTeSR™1 culture media. In order to check if the deletion of HCNR 617 affects pluripotency state, immunocytochemical staining for the pluripotency markers OCT4 and SOX2 were performed. Cells were seeded in 24-well plates, fixed and stained according to the protocol described in the method section. As seen in Figure 3.16 all cell lines maintained the pluripotent state characterized by positive staining for OCT4 and SOX2.

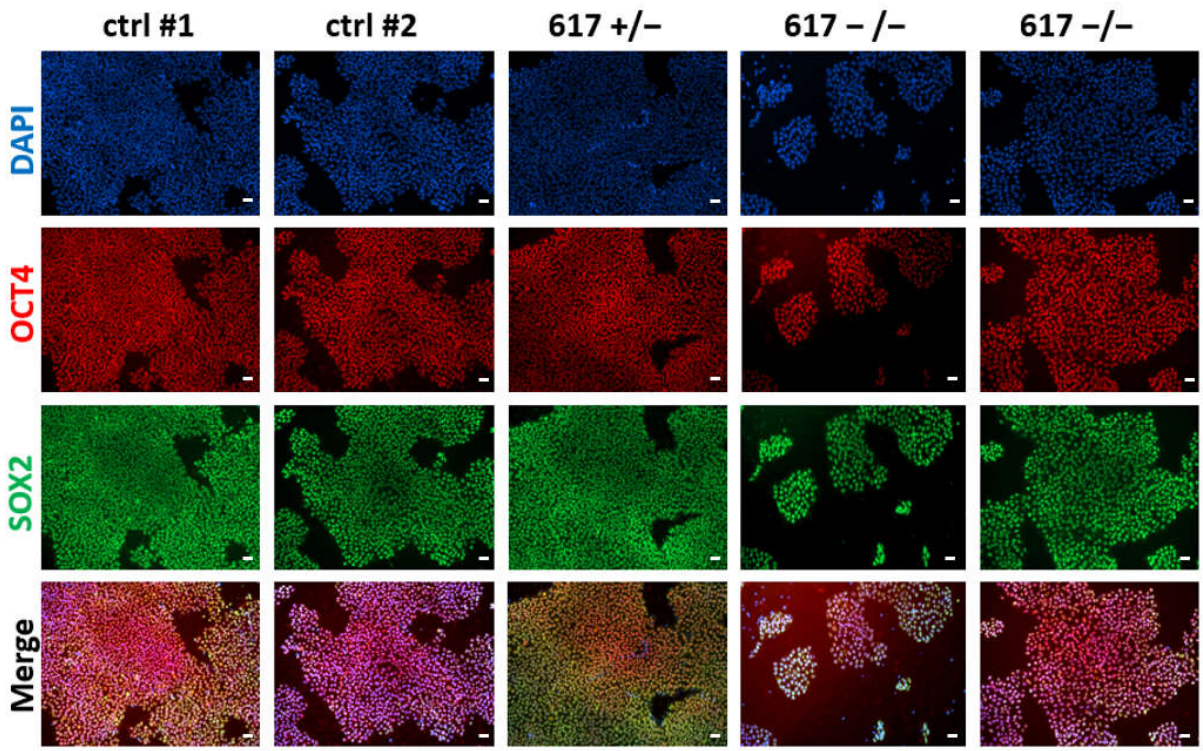


Figure 3.16: OCT4 and SOX2 staining in induced pluripotent stem cell lines. Five clonal lines are presented - ctrl#1 and ctrl#2 control lines with no deletion and three lines with 617 enhancer deletion. Scale bar, 50 µm.

Furthermore, I measured the mRNA expression level of pluripotency marker genes *OCT4* and *NANOG* by qPCR method. OCT4 is a transcription factor, required for the maintenance of pluripotency state (Zeineddine et al. 2014). In combination with other reprogramming factors, OCT4 is used to reprogram differentiated somatic cells into induced pluripotent stem cells (Takahashi et al. 2007). NANOG is another transcription factor relevant for sustaining undifferentiated cell state (W. Zhang et al. 2016). As seen in Figure 3.17, there was no difference in the mRNA expression level of pluripotency genes *OCT4* and *NANOG* between control clones and clones with enhancer deletion.

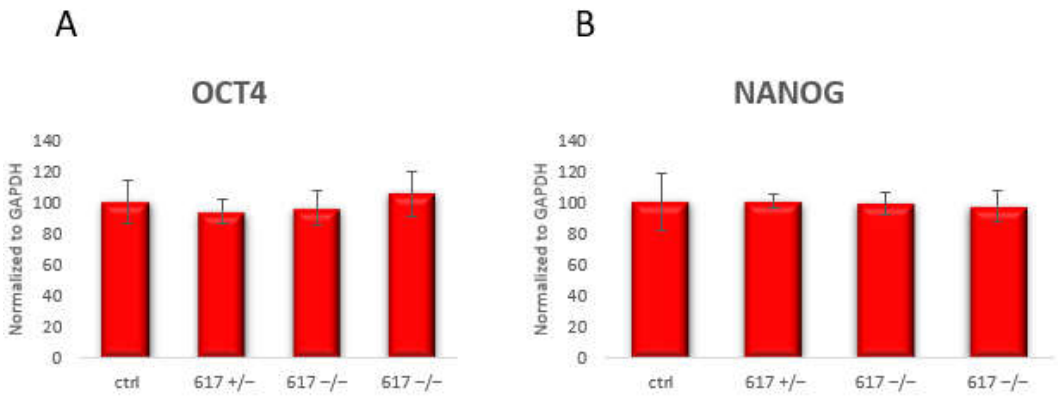


Figure 3.17: qPRC. *OCT4* and *NANOG* in induced pluripotent stem cell lines. Values for two control clones were averaged and compared to the clones with deletion. N=6 biological replicates per cell clone.

To evaluate if the enhancer deletion affected *MEIS1* expression at the level of pluripotent state, we also quantified *MEIS1* mRNA. Finally, we evaluated the *MEIS2* expression level. Interestingly, only one of three clones (617 +/–) with deletion expressed reduced both *MEIS1* and *MEIS2* expression levels whereas, in the other two clones, the levels maintained comparable with control cell lines (Figure 3.18).

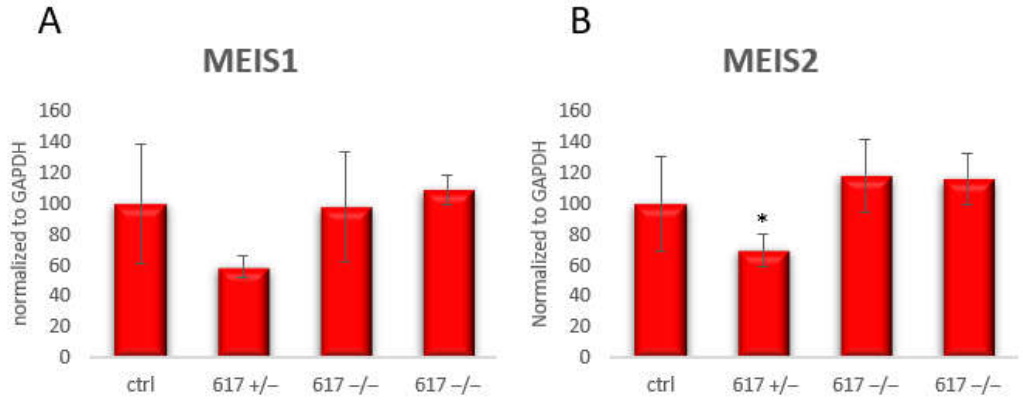


Figure 3.18: qPRC. *MEIS1* and *MEIS2* in induced pluripotent stem cell lines. Values for two unedited clones were averaged and compared to the clones with deletion. N=3-6 biological replicates per cell clone. Asterisks indicate statistically significant interaction *p<0.05 (Student's t-test).

In the next step, we differentiated cells towards GE like progenitors using the previously described protocol (Close et al. 2017). As described, neural induction was initiated using a combination of small molecules SB431542, LDN1933189, and XAV939. On day 10, the cells were collected for RNA extraction and immunocytostaining. As seen in Figure 3.19, neural differentiation was successfully induced, which is indicated by positive immunostaining for PAX6. There was no significant change in immunostaining between

clones which suggests that enhancer deletion had no impact on neural induction and no impact on PAX6 expression.

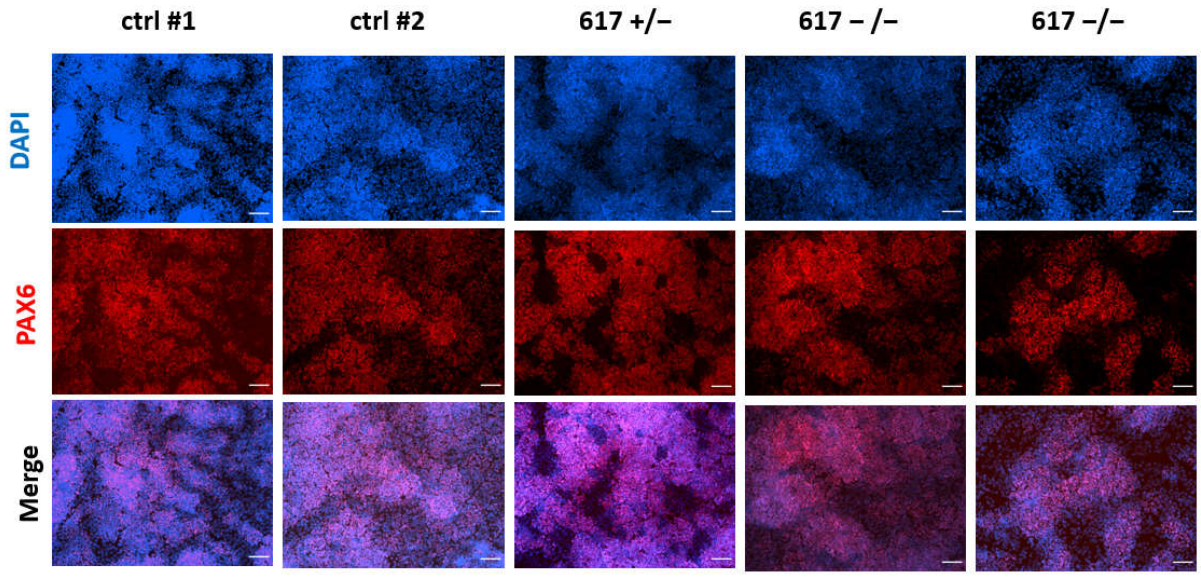


Figure 3.19: Immunofluorescence staining for PAX6 at day 10 of differentiation in neural progenitors. Five clonal lines are presented - ctrl#1 and ctrl#2 control lines with no deletion and three lines with 617 enhancer deletion. Scale bar 100µm.

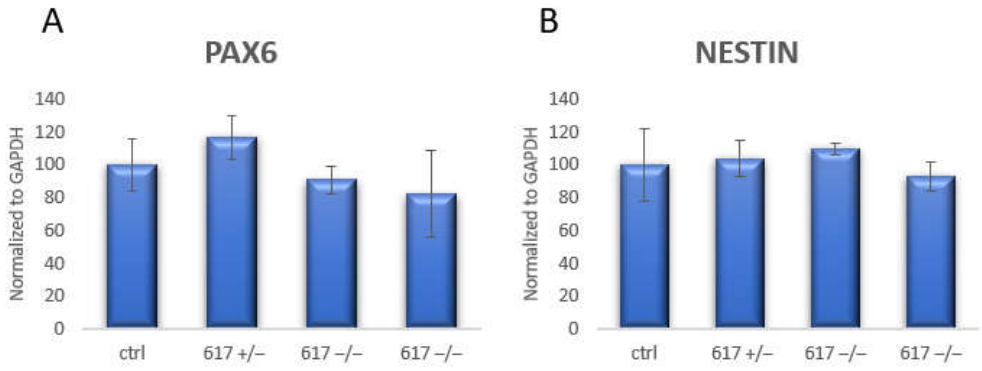


Figure 3.20: qPRC. *PAX6* and *NESTIN* in neural progenitor cell lines. Values for two control clones were averaged and compared to the clones with deletion. N=5 biological replicates per cell clone.

At the same time, I quantified *PAX6* and *NESTIN* by qPRC method. As previously discussed, PAX6 and NESTIN are markers of neural progenitors. Furthermore, it has been reported that Meis1 directly binds to *Pax6* enhancer and upregulates expression during cerebellar development (Owa et al. 2018). However, the results show no significant differences between clones with deletion and control clones (Figure 3.20A). *NESTIN* had a similar expression level in all clonal populations and enhancer deletion had no impact on this gene (Figure 3.20B).

The most important evaluation in this experiment was *MEIS1* expression upon enhancer deletion. As been previously shown, HCNR 617 exhibits feature of activity. ATAC-seq results showed accessibility of this element upon neural induction. This chromatin accessibility progressed and increased and the stage of GE-like progenitors. Furthermore, H9-derived neural stem cell line displayed accessibility at this locus. Finally, in this cell line, physical contact between the *MEIS1* promoter and HCNR 617 was detected. Due to these findings, we reasoned to believe that deletion of this element would affect the expression level of *MEIS1* at the stage of neural progenitors. However, by qPCR, we found that the *MEIS1* expression level had no significant change upon deletion. This might be due to enhancer redundancy who potentially compensates for the loss of enhancer 617. Strikingly, *MEIS2* was significantly reduced in all three clonal populations with enhancer deletion compared to controls (Figure 3.21). These two genes have a high level of sequence conservation and thereby to some extent, functional redundancy (Schulte and Geerts 2019). Compensatory mechanisms between Meis1 and Meis2, which were proposed by some authors, could also play a role in this case (Machon et al. 2015).

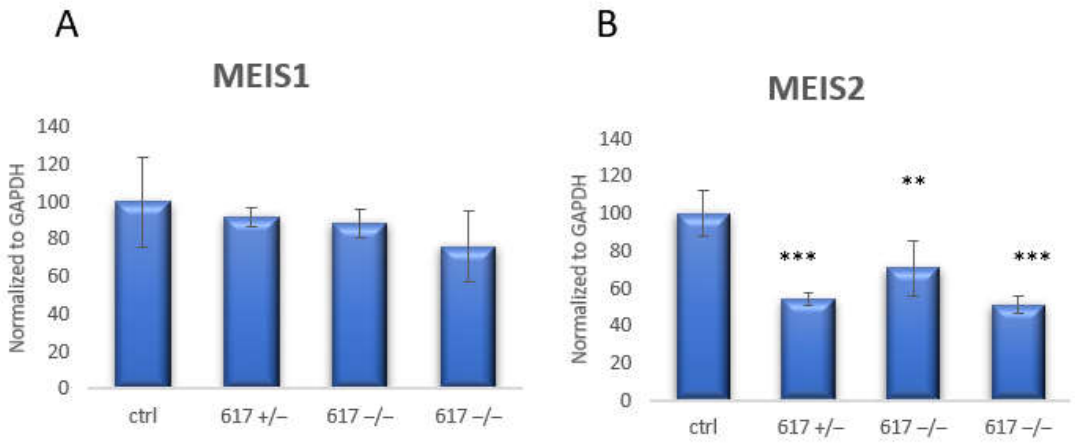


Figure 3.21: qPRC. *MEIS1 and MEIS2* in neural progenitor cell lines. Values for two control clones were averaged and compared to the clones with deletion. N=5 biological replicates per cell clone. Asterisks indicate statistically significant interaction ** $p<0.01$, *** $p< 0.001$ (Student's t-test).

In order to evaluate the effect of enhancer deletion on GE-like neural progenitors, we further differentiated the cells toward ventral fate using the previously described differentiation protocol. Next, we evaluated the expression levels of *MEIS1* and *MEIS2* genes. *MEIS1* was significantly reduced in all clonal populations with deletion compared to non-edited clones. On the other hand, the *MEIS2* expression level remained unchanged (Figure 3.22). As discussed before, ATAC-seq results showed increased open accessibility of 617 regulatory

element at the GE stage. This element showed enhancer activity in ganglionic eminences of developing mice (Spieler et al. 2014). Taken together, this finding demonstrates that RLS associated regulatory element 617 regulates *MEIS1* at the cell-specific stage. What are the consequences of *MEIS1* downregulation and what other genes and pathways are affected remains to be investigated.

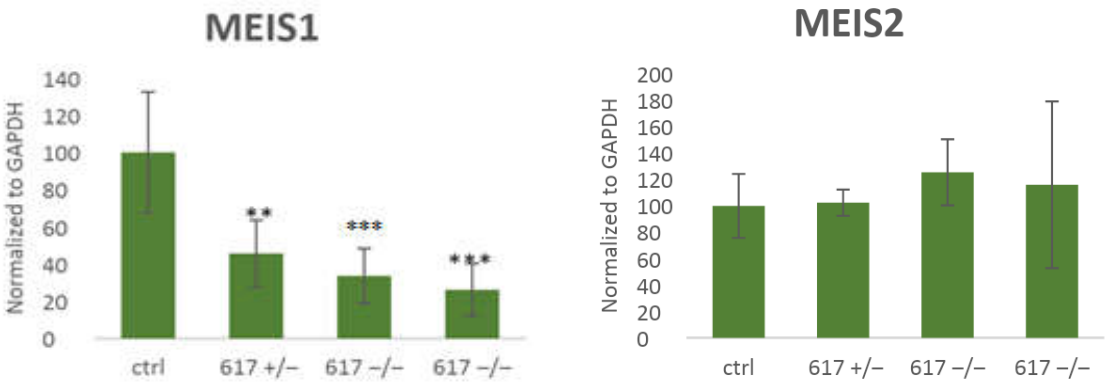


Figure 3.22: qPRC. *MEIS1* and *MEIS2* in induced pluripotent stem cell lines. Values for two unedited clones were averaged and compared to the clones with deletion. N=5 biological replicates per cell clone. Asterisks indicate statistically significant interaction ** $p<0.01$, *** $p< 0.001$ (Student's t-test).

DISCUSSION

4.1 Summary

During my doctoral thesis, I explored the role of RLS-associated risk variants on *MEIS1* regulation at different stages of neural development. I employed complementary methods to investigate regulatory network of *MEIS1* and to elucidate the role of RLS-associated variants. I combined region-specific neural differentiation, an epigenetic assay for chromatin accessibility, three-dimensional chromosomal profiling of the *MEIS1* locus, and CRISPR/Cas9 genome editing to dissect the regulatory landscape of the *MEIS1*. I differentiated an induced pluripotent stem cell line into neural progenitors with properties of ganglionic eminences, a transient structure responsible for the generation of forebrain inhibitory neurons and evaluated *MEIS1* expression dynamics in this system. In search of epigenetic signatures within the *MEIS1* locus, I performed ATAC-seq on *in vitro* differentiated cells as well as commercially available human neural stem cells and excitatory and inhibitory neurons. In-house generated data were combined with published datasets for human and mouse samples to improve the interpretation of cell and tissue-specific regulatory patterns. A highly distinct accessibility pattern involving RLS-associated risk loci at specific cell stages was revieled. HNCR 602 was only accessible in inhibitory neurons and their developmental primordium, the ganglionic eminences. Furthermore, chromatin accessibility within HCNR 602 was detected during mouse developing forebrain. This strongly suggests that HCNR 602 acts as a developmental and adult enhancer with very precise cell type specificity. HCNR 617 exhibited regulatory features in neural cell types including progenitor stages and more mature neural stages. In respect of co-accessibility of enhancers 602 and 617, I identified human lateral ganglionic eminences ganglionic, inhibitory neurons as well as mouse developing forebrain as three sites where these enhancers could regulate *MEIS1* in cooperative fashion.

Furthermore, chromosome conformation capture results showed a direct interaction between the *MEIS1* promoter and HCNR 617 in human neural stem cells. By employing CRISPR/Cas9 genome editing, I demonstrated directly that HCNR 617, carrying the risk variant rs12469063, acts as an enhancer of *MEIS1* expression. This effect was specific to the

stage of ganglionic eminence-like neural progenitors, recapitulating enhancer activity in the developing forebrain.

Finally, chromatin signatures at an RLS-associated region located in an intergenic region on chromosome 2p14, downstream of *MEIS1,* tagged by rs1820897 were explored. There were regions of accessibility with possible functional relevance. In addition, a long-range interaction between the *MEIS1* promoter and the intergenic region was detected, which strongly implicates the intergenic region in the *MEIS1* regulatory network. In summary, this work sheds light on *MEIS1* regulatory elements and cell/tissue-specific patterns of regulation. Moreover, the stage-specific direct effect of an RLS-associated regulatory enhancer on *MEIS1* expression was shown. Table 4.1 summarizes cell type and tissues where investigated RLS-associated elements display features of regulation.

Assays used in this study	***REGULATORY ELEMENTS INVESTIGATED IN THIS STUDY***		
	MEIS1 HNCR 602 tagged by rs11851554	***MEIS1 HCNR 617 tagged by rs12469063***	***2.14 intergenic region tagged by rs12820987***
ATAC-seq DNA accessibility derived from human cell lines	Figure 3.6 • Gabaergic neurons	Figure 3.6 • HMGU12- derived NPC • HMGU12- derived GE • H9-derived neural stem cells • Gabaergic neurons • Glutamatergic neurons	Figure 3.9 Accessibility is detected within the LD block but not within tag SNP in: • H9-derived neural stem cells • Gabaergic neurons
*Public data sets derived from human cells/tissues	Figure 3.8 • Neurons (ChiP-seq 27 Ac) • Neurons (ChiP-seq EP300) • Fetal LGE (ATAC-seq) • Putamen (ATAC-seq)	Figure 3.8 • Neural progenitors (DNAse-seq) • Fetal LGE (ATAC-seq) • Heart (DNAse-seq)	Not studied
*Public data (ENCODE) sets derived from developing mouse	Figure 3.10 • Mouse developing forebrain	Figure 3.10 • Mouse developing forebrain • Mouse developing midbrain (transiently) • Mouse developing hindbrain (transiently) Figure 3.11 • Mouse developing limb (transiently)	Not studied
4C-seq chromosome conformation capture	Figure 3.12 • No physical interaction detected	Figure 3.12 • H9-derived NSC cell line (interaction with MEIS1 promoter)	Figure 3.13 • SHSY-5Y cell line (interaction with MEIS1 promoter)
CRISPR-Cas9 assisted deletion	Not studied	Figure 3.22 • Reduced MEIS1 expression in GE neural progenitors	Not studied

Table 4.1: Summary of cell line and tissues where *MEIS1* associated RLS-associated regulatory elements display features of activity. *Public datasets are listed in Appendix.

4.2 *MEIS1* is the strongest genetic factor for restless legs syndrome

RLS is a sensorimotor neurological disorder with the potential to severely affect the life quality of patients. A considerable effort in genetic studies enabled us to discover the genetic factors associated with the disease. To this date, *MEIS1* is the strongest genetic signal associated with RLS. The lead risk variant rs113851554 located in HCNR 602 of *MEIS1*. The meta-analysis from 2017 showed the remarkable significance of $P = 2x10^{-280}$ and odds ratio 1.92 (Schormair et al. 2017). Another meta-analysis is ongoing and reports a P-value of 10^{-1105} for rs113851554 (unpublished data). Besides, another RLS association signal has been reported in the vicinity of *MEIS1*. An intergenic locus located 1.3 Mb downstream of *MEIS1*, harbors the tag SNP rs1820987 with a significance P-value of $1.39x10^{-58}$ (Schormair et al. 2017). In this doctoral work, this risk locus was found to physically interact with the *MEIS1* promoter in a neuroblastoma cell line, emphasizing the necessity of further experiments that would further clarify the potential regulatory mechanism underlying this risk locus.

Moreover, one recent study discovered an interplay between *MEIS1* and another RLS risk locus encompassing the *SKOR1* gene. They discovered that MEIS1 modulates *SKOR1* expression through direct promoter binding (Catoire et al. 2018). This suggests that MEIS1 possesses a highly complex mode of action in RLS pathology. We did not measure *SKOR1* expression upon *MEIS1* enhancer deletion, but it would be worth evaluating if reduced *MEIS1* expression affects the *SKOR1* in GE-like progenitors. Another interesting candidate to evaluate in deletion experiment is the *MYT1* gene, a newly discovered RLS candidate that was found to be regulated by MEIS1 (Sarayloo et al. 2019).

In addition, GWA studies discovered *MEIS1* association with insomnia symptoms (Lane et al. 2017; Hammerschlag et al. 2017), suggesting a shared genetic basis for RLS and insomnia disorder. However, the pleiotropic effect of *MEIS1* was argued in another study, where they found that *MEIS1* is associated with RLS only and not with insomnia (El Gewely et al. 2018). These opposite findings emphasize the necessity of properly phenotyped cohorts and sufficient sample size (Oexle 2018).

Thanks to GWA studies, substantial progress has been made towards optimal pharmacological treatment tailoring. A recent case report study showed the successful application of thalidomide in a patient with treatment resistant RLS. Thalidomide binds

cereblon and thereby inhibits degradation of RLS-associated candidate MEIS2 which could potentially recompense for MEIS1 dysregulation in RLS (Salminen et al. 2020).

4.3 Implication of *MEIS1* in iron and dopamine pathways

As mentioned before, iron and dopamine abnormalities are two central mechanisms involved in RLS pathology. It remains relatively unknown how exactly genetic factors relate to iron and dopamine. However, there are certain pieces of evidence bringing together *MEIS1* and iron metabolism. Firstly, in the thalamus, ferritin was significantly increased in homozygous risk haplotype carriers those who have reduced expression of *MEIS1,* both on mRNA and protein level. In a Caenorhabditis model to study *MEIS1* ortholog *Unc-62*, the authors found significantly increased ferritin upon *Unc-62* knockdown. In the same study, the authors translated this finding in the human cell model of neuroblastoma and found a reduction of *MEIS1* expression upon iron deficiency (Catoire et al. 2011). Moreover, an increased MEIS1 expression was found in the brain microvasculature isolated from RLS patients compared to controls (Connor et al. 2017). The impaired iron acquisition at blood-brain-barrier (BBB) and altered MEIS1 expression provide new insight into the MEIS1 role in RLS and new possibilities for research of iron interaction with genetic factors contributing to RLS. In this work, I didn't evaluate the relation between iron and *MEIS1* although it would be worth evaluating ferritin levels, as well as genes involved in iron transport and metabolism in GE-like progenitors where deletion of 617 regulatory element reduced *MEIS1* expression. This could further expose how *MEIS1* modulates iron metabolism.

The altered dopaminergic system in RLS is widely studied. *Meis1*-deficient mice had an altered response to the treatment with dopamine agonists (Salminen et al. 2017). Another study where Meis1 heterozygous knockout mice are investigated, found an alteration in dopaminergic system: an increase of tyrosine hydroxylase (TH) in *Meis1* KO mice in the striatum at the mRNA level and decrease of TH at the protein level, which might be a compensatory effect of initial mRNA reduction. In conclusion, the authors suggested impaired dopamine synthesis (Lyu et al. 2020). In this study, we differentiated pluripotent cells to GE-like progenitors, but it would be a promising approach to employ differentiation toward dopaminergic neurons, coupled with overexpression and downregulation of *MEIS1* to examine the connection between MEIS1 and dopamine.

4.4 Prioritizing cell type relevant for RLS pathophysiology

In the post -GWAS era, the great challenge is to prioritize causal variants. Another challenge is to identify tissues and cells where these risk variants act and contribute to the disease. This work showed that *MEIS1* possesses cell and tissue-specific regulatory network, observed from chromatin accessibility data. Even though *MEIS1* is involved in many developmental processes and organogenesis, a very distinct, cell type-specific pattern of regulation is observed across different stages of neural maturity and identity. We employed differentiation protocol to achieve ganglionic eminence-like neural progenitors where Meis1 expression and RLS-associated enhancer activity were detected (Spieler et al. 2014). The protocol allowed us to learn about the stage-specific effect of two regulatory elements, HCNR 602 and HCNR 617, on *MEIS1* regulation. However, the differentiation protocol I applied (Close et al. 2017), provides robust instructions for differentiation to inhibitory neurons. Obtaining these differentiated neurons and assaying them for DNA accessibility would provide us with valuable data generated using *in vitro* model of differentiation. This work did not include the generation of the neurons, however, it would be an important task for the future projects. We profiled commercial cell lines of inhibitory and excitatory neurons which shed a light on *MEIS1* regulation in more mature stages. We identified inhibitory neurons as the only cell type where regulatory element 602 harbouring the lead RLS SNP rs113851554 exhibit accessibility. This provides an evidence for cell prioritization to continue with functional experiments in order to better understand genetics association with RLS.

In addition to in-house generated datasets, we took advantage of published datasets and confirmed selective accessibility of HCNR 602 in a human brain specimen dissected from putamen (Fullard et al. 2018). It is noteworthy to mention that we looked at *MEIS1* accessibility in all brain regions profiled in this study (Fullard et al. 2018) and only in the putamen, the lead RLS SNP was within accessible DNA. This finding further strengthens the hypothesis of gabaergic neurons as a strong candidate cell line where *MEIS1* plays a role in RLS pathology and identifies the striatum as a potential brain region where *MEIS1* might be dysregulated. Moreover, we observed the accessibility of 602 and 617 element in fetal lateral ganglionic eminences (Markenscoff-Papadimitriou et al. 2020).

In addition to conserved elements located within *MEIS1*, I explored chromatin accessibility at an intergenic risk locus 1.3 Mb downstream of *MEIS1*. Within the risk region there are conserved candidate regions, with mapped open chromatin in neural stem cells and neurons,

who could carry the causal risk variant. However, an additional work needs to be done to understand the regulatory impact of this risk locus. Overall, it remains challenging to precisely recapitulate human neurodevelopment using *in vitro* differentiation. However, this work prioritized inhibitory neurons, ganglionic eminences, and putamen as the key regions for potential role of the lead risk variant in RLS pathology.

4.5 *MEIS1* regulatory landscape in neural lineages involves RLS risk variants

As known, most of the risk variants are mapped in non-coding regions making it challenging to interpret their contribution to the trait they are associated with. To try to prioritize cell types where *MEIS1*-associated risk variant could express features of enhancers, we employed assay for transposase accessible chromatin coupled with high throughput sequencing (ATAC-seq) to map accessible chromatin and thereby to infer regulatory variants in neural cell types. In this work, I tried to disentangle the regulatory network of *MEIS1* since this gene is the strongest RLS candidate with an odds ratio (OR) = 2.16 95% confidence interval (CI) 2.014–2.49). The results showed that accessibility signature at *MEIS1* locus has distinct features in neural and non-neural lineages, both in human and mouse datasets. Furthermore, we demonstrated progressive accessibility of HCNR 617 in *in vitro* model of differentiation and in commercial lines, neural stem cells, and inhibitory neurons. This underlines the importance of this regulatory element at different stages of neural maturation. Besides, we proved the enhancer activity in GE-like progenitors. Interestingly, this element had no accessibility peak in the neuronal cells of the putamen. This lack of activity could be potentially explained by the fact that the majority of the neurons in the putamen are medium spiny neurons and this element is relevant for less mature stages of neural cell lines. Nevertheless, we did observe open chromatin in inhibitory neurons but these are *in vitro* generated cells, and they cannot mimic faithfully *in vivo* environment. Furthermore, we have observed an open chromatin peak in HCNR 617 in accessibility datasets generated from the human fetal heart and *in vitro* differentiated cardiomyocytes. Finally, mouse datasets showed DNA accessibility in the forebrain and transient accessibility midbrain, hindbrain and limb development. This could potentially speak in favour of the pleiotropic feature of this conserved enhancer, transiently active in certain time points during neurogenesis, cardiogenesis, and limb formation.

On the other side, HCNR 602 displayed very selective accessibility in inhibitory neurons, putamen and fetal lateral ganglionic eminences. Moreover, *in vitro* differentiated neurons were enriched for markers of active enhancers at HCNR 602 according to ChIP seq datasets obtained from ENCODE. Even though we modeled *in vitro* generation of ganglionic eminences, we did not observe the accessibility feature at this element. This could be due to the fact that GE progenitors resemble more medial ganglionic eminences than lateral or simply because *in vitro* differentiation could not model precisely human neurogenesis. This element might be a developmental enhancer, as well as an adult enhancer, strictly controlled spatially and temporally, contributing to RLS which can be considered as a neurodevelopmental disorder with adult-onset. Finally, we searched for an accessibility pattern at the RLS signal in the vicinity of *MEIS1,* tagged by rs1820897. In this dataset, there was no accessibility peak at the tag SNP however, assayed DNA accessibility is not direct evidence of regulatory sequence. Nevertheless, conserved elements surrounding the tag SNP which are part of the LD block driving the RLS signal were enriched for accessible chromatin, which suggest a potential regulatory role.

The advantage of this method is genome-wide accessibility profiling, which means we can use them to explore all candidate loci identified in GWAS and search for the enrichment of risk variants within accessible DNA. This will help for tissue and cell type prioritization for functional follow-up studies.

4.6 Spatial organisation reveals direct contact to RLS risk loci

The spatial organization of the genome is a highly dynamic cellular process and contributes to gene expression. To reveal the regulatory network of the *MEIS1*, we employed circular chromosomal conformation capture followed by high -throughput sequencing 4C-seq. We used three different cell types corresponding to different stages of the neuronal specification. We identified a cell type-specific interaction pattern of *MEIS1* which also involves RLS-associated risk loci. In human neural stem cells, we detected interactions of *MEIS1* promoter with several intronic elements including RLS-associated enhancer 617. This interaction pattern corresponds to the accessibility signature found in the same cell line. This finding strongly supports the role of HCNR 617 in neural development. Furthermore, in the neuroblastoma cell line, we identified interaction with the MEIS1 independent RLS risk locus tagged with rs1820987, located 1.3Mb downstream of *MEIS.* This finding is the very first

evidence that this RLS-associated risk locus engages in long-range interaction with the *MEIS1* and potentially participate in its regulation.

Nevertheless, this work had some limitations. We did not include a spatial profile of iPS-derived neural progenitors, GE-like progenitors, and neurons which were assayed for chromatin accessibility due to insufficient read depth during sequencing, however, this project is still ongoing and interaction profiles of these cells line, especially inhibitory neurons will help to decipher three-dimensional component of *MEIS1* regulation. Another promising approach would be to include patient-derived cells differentiated to neural progenitor and more mature neural stages to see if RLS risk variants disrupt chromatin looping and thereby enhancer-promoter interaction. It would be valuable to evaluate all RLS risk loci simultaneously and search for their interaction patterns as one enhancer element could regulate more genes and could contact other enhancers. This could be achieved by Hi-C method, which interrogates the spatial folding of the entire genome. There are newly developed, adapted methods that require a very low initial input of material and can produce interaction maps at very high resolution (Lu et al. 2020). This could help to prioritize functional variants.

4.7 Targeted deletion of RLS-associated regulatory element proves direct effect on *MEIS1* regulation

Targeted deletion of assumed regulatory elements is a frequently used method to address the question of enhancer functionality. To investigate the enhancer effect on *MEIS1* and the direction of the effect, I deleted HCNR 617 in induced pluripotent stem cells. The expression of *MEIS1* was overall unchanged, and two pluripotency markers, *OCT4* and *NANOG* were stable. Contrary to hypothesized enhancer function in iPSC-derived neural progenitors, we found that HCNR 617 deletion did not affect *MEIS1* expression. The unchanged *MEIS1* expression can be explained by enhancer redundancy. The enhancer redundancy frequently observed in developmental genes and serves to precisely orchestrate complex temporospatial expression. However, the *MEIS2* gene was consistently downregulated in clonal lines with enhancer deletion. This observed phenomenon can be explained by gene redundancy, however, this has to be investigated in more depth. Finally, enhancer deletion affected *MEIS1* expression in GE-like neural progenitor cell line. This speaks in favor of cell type-specific enhancer activity, in agreement with *in vivo* activity observed in a mouse model. Further

study is needed to establish the role of enhancer 617 in neuronal cell types. It is important to mention that clonal cell lines with deletion of regulatory element 602 are obtained. However, upon differentiation to neural progenitors and GE-like progenitors we observed discordant *MEIS1* expression values in two clonal lines with enhancer deletion (data not shown). The future work will be focused on obtaining at least three and more clonal lines with enhancer deletion to be able to produce reliable results. Deletion of both intronic enhancers and distal element 1.3Mb downstream of *MEIS1* identified in the 4C-seq experiment should be evaluated in differentiated neurons, coupled with other functional readouts such as RNA-seq, to disentangle transcriptomic changes and identify downstream targets of MEIS1 who could contribute to the molecular mechanism of RLS.

Generation of mouse lines with enhancer deletion would provide the opportunity to study *in vivo* functionality of risk variants. By employing epigenetic methods coupled with gene expression, different cell types at different time points and overall phenotype can be investigated to determine the effect of enhancers.

4.8 Conclusion

In summary, this work aimed to explore beyond GWAS signal and determine the functionality of RLS-associated risk variants within the *MEIS1*. This gene possesses a very complex regulatory landscape, which is in line with pleiotropic effect in haematopoiesis, neurogenesis, limb, and heart development. This work provided valuable insight into chromatin accessibility at *MEIS1* locus during neural differentiation and pinpointed the cell types and tissues where RLS risk variants exhibit their regulatory features. We revealed physical interactions of *MEIS1* promoter and RLS-associated loci. A targeted deletion of RLS-associated enhancer clearly reduced *MEIS1* mRNA expression level in GE-like progenitors, demonstrating a stage-specific enhancer effect on the *MEIS1*. The results show that a combination of genetic and epigenetic methods can help prioritize functional risk variants. Ultimately, more work will be needed to address the molecular mechanism which causes RLS.

4.9 Future Perspective

Genome-wide association studies rapidly advanced in the past ten years and shed a light on the genetic association of various diseases and traits. However, the complex genetic

architecture of RLS is constantly influenced by environmental factors making it challenging to decipher causal genetic variants. Nonetheless, simultaneously with the advancement of GWAS, genetic and epigenetic technologies coupled with powerful computational approaches emerged, providing a framework for the translation of GWAS findings into functional follow-up studies with the goal to provide better diagnostic and treatment options.

A promising step to further elucidate the role of RLS-associated variants is to adopt the multi-omics approach and simultaneously analyse transcriptomics in combination with systematic epigenetic profiling and spatial genomic organization. Moreover, proteomics and metabolomic profiles should be also integrated into multi-omics framework. Besides, *in vivo,* and *in vitro* models should be studied. Precisely, there are several mouse models developed to investigate RLS syndrome, and they have substantially contributed to understanding the mechanism of underlying Meis1 deficiency. Finally, human derived brain organoids provide great opportunity to establish a platform for human brain development, to study the consequences of RLS variants, and finally to potentially screen for novel druggable targets. This multi-layered approach can serve to study all genetic variants for a given trait and to identify genes and pathways involved in disease mechanism.

Pinpointing the causal variants and mode of action should serve to find optimal therapy, individually tailored for each patient, resulting in better treatment outcomes and life quality.

REFERENCES

Allen, R.P, and C.J. Earley. 2000. 'Defining the Phenotype of the Restless Legs Syndrome (RLS) Using Age-of-Symptom-Onset'. *Sleep Medicine* 1 (1): 11–19. https://doi.org/10.1016/s1389-9457(99)00012-x.

Allen, R. P., P. B. Barker, F. W. Wehrl, F. Wehrl, H. K. Song, and C. J. Earley. 2001. 'MRI Measurement of Brain Iron in Patients with Restless Legs Syndrome'. *Neurology* 56 (2): 263–65. https://doi.org/10.1212/wnl.56.2.263.

Allen, R. P., and C. J. Earley. 1996. 'Augmentation of the Restless Legs Syndrome with Carbidopa/Levodopa'. *Sleep* 19 (3): 205–13. https://doi.org/10.1093/sleep/19.3.205.

Allen, Richard. 2004. 'Dopamine and Iron in the Pathophysiology of Restless Legs Syndrome (RLS)'. *Sleep Medicine* 5 (4): 385–91. https://doi.org/10.1016/j.sleep.2004.01.012.

Allen, Richard P. 2015. 'Restless Leg Syndrome/Willis-Ekbom Disease Pathophysiology'. *Sleep Medicine Clinics* 10 (3): 207–14. https://doi.org/10.1016/j.jsmc.2015.05.022.

Allen, Richard P., Daniel Picchietti, Wayne A. Hening, Claudia Trenkwalder, Arthur S. Walters, Jacques Montplaisi, Restless Legs Syndrome Diagnosis and Epidemiology workshop at the National Institutes of Health, and International Restless Legs Syndrome Study Group. 2003. 'Restless Legs Syndrome: Diagnostic Criteria, Special Considerations, and Epidemiology. A Report from the Restless Legs Syndrome Diagnosis and Epidemiology Workshop at the National Institutes of Health'. *Sleep Medicine* 4 (2): 101–19. https://doi.org/10.1016/s1389-9457(03)00010-8.

Allen, Richard P., Daniel L. Picchietti, Michael Auerbach, Yong Won Cho, James R. Connor, Christopher J. Earley, Diego Garcia-Borreguero, et al. 2018. 'Evidence-Based and Consensus Clinical Practice Guidelines for the Iron Treatment of Restless Legs Syndrome/Willis-Ekbom Disease in Adults and Children: An IRLSSG Task Force Report'. *Sleep Medicine* 41: 27–44. https://doi.org/10.1016/j.sleep.2017.11.1126.

Allen, Richard P., Daniel L. Picchietti, Diego Garcia-Borreguero, William G. Ondo, Arthur S. Walters, John W. Winkelman, Marco Zucconi, et al. 2014. 'Restless Legs Syndrome/Willis-Ekbom Disease Diagnostic Criteria: Updated International Restless Legs Syndrome Study Group (IRLSSG) Consensus Criteria--History, Rationale, Description, and Significance'. *Sleep Medicine* 15 (8): 860–73. https://doi.org/10.1016/j.sleep.2014.03.025.

Amemiya, Haley M., Anshul Kundaje, and Alan P. Boyle. 2019. 'The ENCODE Blacklist: Identification of Problematic Regions of the Genome'. *Scientific Reports* 9 (1): 9354. https://doi.org/10.1038/s41598-019-45839-z.

Ariki, Reina, Satoru Morikawa, Yo Mabuchi, Sadafumi Suzuki, Mayuka Nakatake, Kentaro Yoshioka, Shinya Hidano, et al. 2014. 'Homeodomain Transcription Factor Meis1 Is a Critical Regulator of Adult Bone Marrow Hematopoiesis'. *PloS One* 9 (2): e87646. https://doi.org/10.1371/journal.pone.0087646.

Azcoitia, Valeria, Miguel Aracil, Carlos Martínez-A, and Miguel Torres. 2005. 'The Homeodomain Protein Meis1 Is Essential for Definitive Hematopoiesis and Vascular Patterning in the Mouse Embryo'. *Developmental Biology* 280 (2): 307–20. https://doi.org/10.1016/j.ydbio.2005.01.004.

Barber, Benjamin A., Vichithra R. B. Liyanage, Robby M. Zachariah, Carl O. Olson, Melissa A. G. Bailey, and Mojgan Rastegar. 2013. 'Dynamic Expression of MEIS1 Homeoprotein in E14.5 Forebrain and Differentiated Forebrain-Derived Neural Stem Cells'. *Annals of Anatomy - Anatomischer Anzeiger* 195 (5): 431–40. https://doi.org/10.1016/j.aanat.2013.04.005.

Biedler, J. L., S. Roffler-Tarlov, M. Schachner, and L. S. Freedman. 1978. 'Multiple Neurotransmitter Synthesis by Human Neuroblastoma Cell Lines and Clones'. *Cancer Research* 38 (11 Pt 1): 3751–57.

Bogan, Richard K., and James A. Cheray. 2013. 'Restless Legs Syndrome: A Review of Diagnosis and Management in Primary Care'. *Postgraduate Medicine* 125 (3): 99–111. https://doi.org/10.3810/pgm.2013.05.2636.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bonati, Maria Teresa, Luigi Ferini-Strambi, Paolo Aridon, Alessandro Oldani, Marco Zucconi, and Giorgio Casari. 2003. 'Autosomal Dominant Restless Legs Syndrome Maps on Chromosome 14q'. *Brain: A Journal of Neurology* 126 (Pt 6): 1485–92. https://doi.org/10.1093/brain/awg137.

Bouilloux, Fabrice, Jérôme Thireau, Stéphanie Ventéo, Charlotte Farah, Sarah Karam, Yves Dauvilliers, Jean Valmier, et al. n.d. 'Loss of the Transcription Factor Meis1 Prevents Sympathetic Neurons Target-Field Innervation and Increases Susceptibility to Sudden Cardiac Death'. *ELife* 5. Accessed 4 September 2020. https://doi.org/10.7554/eLife.11627.

Bradley, Conor A. 2018. 'Mechanistic Insights into a Non-Coding Risk SNP'. *Nature Reviews Urology* 15 (11): 655–655. https://doi.org/10.1038/s41585-018-0079-9.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. 'Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position'. *Nature Methods* 10 (12): 1213–18. https://doi.org/10.1038/nmeth.2688.

Buenrostro, Jason D., Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang, and William J. Greenleaf. 2015. 'Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation'. *Nature* 523 (7561): 486–90. https://doi.org/10.1038/nature14590.

Buniello, Annalisa, Jacqueline A. L. MacArthur, Maria Cerezo, Laura W. Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, et al. 2019. 'The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019'. *Nucleic Acids Research* 47 (D1): D1005–12. https://doi.org/10.1093/nar/gky1120.

Butt, Simon J.B., Vitor H. Sousa, Marc V. Fuccillo, Jens Hjerling-Leffler, Goichi Miyoshi, Shioko Kimura, and Gordon Fishell. 2008. 'The Requirement of Nkx2-1 in the Temporal Specification of Cortical Interneuron Subtypes'. *Neuron* 59 (5): 722–32. https://doi.org/10.1016/j.neuron.2008.07.031.

Callaerts, P., G. Halder, and W. J. Gehring. 1997. 'PAX-6 in Development and Evolution'. *Annual Review of Neuroscience* 20: 483–532. https://doi.org/10.1146/annurev.neuro.20.1.483.

Catoire, Hélène, Patrick A. Dion, Lan Xiong, Mourabit Amari, Rebecca Gaudet, Simon L. Girard, Anne Noreau, et al. 2011. 'Restless Legs Syndrome-Associated MEIS1 Risk Variant Influences Iron Homeostasis'. *Annals of Neurology* 70 (1): 170–75. https://doi.org/10.1002/ana.22435.

Catoire, Helene, Faezeh Sarayloo, Karim Mourabit Amari, Sergio Apuzzo, Alanna Grant, Daniel Rochefort, Lan Xiong, et al. 2018. 'A Direct Interaction between Two Restless Legs Syndrome Predisposing Genes: MEIS1 and SKOR1'. *Scientific Reports* 8 (1): 12173. https://doi.org/10.1038/s41598-018-30665-6.

Chambers, Stuart M., Christopher A. Fasano, Eirini P. Papapetrou, Mark Tomishima, Michel Sadelain, and Lorenz Studer. 2009. 'Highly Efficient Neural Conversion of Human ES and IPS Cells by Dual Inhibition of SMAD Signaling'. *Nature Biotechnology* 27 (3): 275–80. https://doi.org/10.1038/nbt.1529.

Chen, Shenghan, William G. Ondo, Shaoqi Rao, Lin Li, Qiuyun Chen, and Qing Wang. 2004. 'Genomewide Linkage Scan Identifies a Novel Susceptibility Locus for Restless Legs Syndrome on Chromosome 9p'. *American Journal of Human Genetics* 74 (5): 876–85. https://doi.org/10.1086/420772.

Chen, Si-Jing, Le Shi, Yan-Ping Bao, Ye-Kun Sun, Xiao Lin, Jian-Yu Que, Michael V. Vitiello, Yu-Xin Zhou, Yong-Qing Wang, and Lin Lu. 2018. 'Prevalence of Restless Legs Syndrome during Pregnancy: A Systematic Review and Meta-Analysis'. *Sleep Medicine Reviews* 40: 43–54. https://doi.org/10.1016/j.smrv.2017.10.003.

Close, Jennie L., Zizhen Yao, Boaz P. Levi, Jeremy A. Miller, Trygve E. Bakken, Vilas Menon, Jonathan T. Ting, et al. 2017. 'Single-Cell Profiling of an in Vitro Model of Human Interneuron Development Reveals Temporal Dynamics of Cell Type Production and Maturation'. *Neuron* 93 (5): 1035-1048.e5. https://doi.org/10.1016/j.neuron.2017.02.014.

Collins, F. S., L. D. Brooks, and A. Chakravarti. 1998. 'A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation'. *Genome Research* 8 (12): 1229–31. https://doi.org/10.1101/gr.8.12.1229.

Connor, J. R., P. J. Boyer, S. L. Menzies, B. Dellinger, R. P. Allen, W. G. Ondo, and C. J. Earley. 2003. 'Neuropathological Examination Suggests Impaired Brain Iron Acquisition in Restless Legs Syndrome'. *Neurology* 61 (3): 304–9. https://doi.org/10.1212/01.wnl.0000078887.16593.12.

Connor, J. R., X. S. Wang, S. M. Patton, S. L. Menzies, J. C. Troncoso, C. J. Earley, and R. P. Allen. 2004. 'Decreased Transferrin Receptor Expression by Neuromelanin Cells in

Restless Legs Syndrome'. *Neurology* 62 (9): 1563–67. https://doi.org/10.1212/01.wnl.0000123251.60485.ac.

Connor, James R., Xin-Sheng Wang, Richard P. Allen, John L. Beard, Jason A. Wiesinger, Barbara T. Felt, and Christopher J. Earley. 2009. 'Altered Dopaminergic Profile in the Putamen and Substantia Nigra in Restless Leg Syndrome'. *Brain: A Journal of Neurology* 132 (Pt 9): 2403–12. https://doi.org/10.1093/brain/awp125.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. 'Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State'. *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21931–36. https://doi.org/10.1073/pnas.1016071107.

Davis, Carrie A., Benjamin C. Hitz, Cricket A. Sloan, Esther T. Chan, Jean M. Davidson, Idan Gabdank, Jason A. Hilton, et al. 2018. 'The Encyclopedia of DNA Elements (ENCODE): Data Portal Update'. *Nucleic Acids Research* 46 (D1): D794–801. https://doi.org/10.1093/nar/gkx1081.

Delgado, Irene, Alejandra C. López-Delgado, Alberto Roselló-Díez, Giovanna Giovinazzo, Vanessa Cadenas, Laura Fernández-de-Manuel, Fátima Sánchez-Cabo, Matthew J. Anderson, Mark Lewandoski, and Miguel Torres. 2020. 'Proximo-Distal Positional Information Encoded by an Fgf-Regulated Gradient of Homeodomain Transcription Factors in the Vertebrate Limb'. *Science Advances* 6 (23): eaaz0742. https://doi.org/10.1126/sciadv.aaz0742.

Desai, Anup V., Lynn F. Cherkas, Tim D. Spector, and Adrian J. Williams. 2004. 'Genetic Influences in Self-Reported Symptoms of Obstructive Sleep Apnoea and Restless Legs: A Twin Study'. *Twin Research: The Official Journal of the International Society for Twin Studies* 7 (6): 589–95. https://doi.org/10.1375/1369052042663841.

Desautels, A., G. Turecki, J. Montplaisir, A. Sequeira, A. Verner, and G. A. Rouleau. 2001. 'Identification of a Major Susceptibility Locus for Restless Legs Syndrome on Chromosome 12q'. *American Journal of Human Genetics* 69 (6): 1266–70. https://doi.org/10.1086/324649.

Dunham, Ian, Anshul Kundaje, Shelley F. Aldred, Patrick J. Collins, Carrie A. Davis, Francis Doyle, Charles B. Epstein, et al. 2012. 'An Integrated Encyclopedia of DNA Elements in the Human Genome'. *Nature* 489 (7414): 57–74. https://doi.org/10.1038/nature11247.

Dupays, Laurent, Catherine Shang, Robert Wilson, Surendra Kotecha, Sophie Wood, Norma Towers, and Timothy Mohun. 2015. 'Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis'. *Cell Reports* 13 (1): 183–95. https://doi.org/10.1016/j.celrep.2015.08.065.

Earley, C. J., J. R. Connor, J. L. Beard, E. A. Malecki, D. K. Epstein, and R. P. Allen. 2000. 'Abnormalities in CSF Concentrations of Ferritin and Transferrin in Restless Legs Syndrome'. *Neurology* 54 (8): 1698–1700. https://doi.org/10.1212/wnl.54.8.1698.

Edwards, Stacey L., Jonathan Beesley, Juliet D. French, and Alison M. Dunning. 2013. 'Beyond GWASs: Illuminating the Dark Road from Association to Function'. *American Journal of Human Genetics* 93 (5): 779–97. https://doi.org/10.1016/j.ajhg.2013.10.012.

Ekbom, Karl-Axel. 1945. ‘Restless Legs’. *Acta Med. Scand.* 158: 4–122.

El Gewely, Maryam, Mélanie Welman, Lan Xiong, Sophie Yin, Hélène Catoire, Guy Rouleau, Jacques Y. Montplaisir, Alex Desautels, and Simon C. Warby. 2018. ‘Reassessing GWAS Findings for the Shared Genetic Basis of Insomnia and Restless Legs Syndrome’. *Sleep* 41 (11). https://doi.org/10.1093/sleep/zsy164.

Erickson, Timothy, Curtis R. French, and Andrew J. Waskiewicz. 2010. ‘Meis1 Specifies Positional Information in the Retina and Tectum to Organize the Zebrafish Visual System’. *Neural Development* 5 (1): 22. https://doi.org/10.1186/1749-8104-5-22.

Fischer, Eric S., Kerstin Böhm, John R. Lydeard, Haidi Yang, Michael B. Stadler, Simone Cavadini, Jane Nagel, et al. 2014. ‘Structure of the DDB1-CRBN E3 Ubiquitin Ligase in Complex with Thalidomide’. *Nature* 512 (7512): 49–53. https://doi.org/10.1038/nature13527.

Flames, Nuria, Ramón Pla, Diego M. Gelman, John L. R. Rubenstein, Luis Puelles, and Oscar Marín. 2007. ‘Delineation of Multiple Subpallial Progenitor Domains by the Combinatorial Expression of Transcriptional Codes’. *The Journal of Neuroscience* 27 (36): 9682–95. https://doi.org/10.1523/JNEUROSCI.2750-07.2007.

Forrest, Marc P., Hanwen Zhang, Winton Moy, Heather McGowan, Catherine Leites, Leonardo E. Dionisio, Zihui Xu, et al. 2017. ‘Open Chromatin Profiling in HiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci’. *Cell Stem Cell* 21 (3): 305-318.e8. https://doi.org/10.1016/j.stem.2017.07.008.

Freedman, Matthew L, Alvaro N A Monteiro, Simon A Gayther, Gerhard A Coetzee, Angela Risch, Christoph Plass, Graham Casey, et al. 2011. ‘Principles for the Post-GWAS Functional Characterization of Cancer Risk Loci’. *Nature Genetics* 43 (6): 513–18. https://doi.org/10.1038/ng.840.

Fullard, John F., Mads E. Hauberg, Jaroslav Bendl, Gabor Egervari, Maria-Daniela Cirnaru, Sarah M. Reach, Jan Motl, Michelle E. Ehrlich, Yasmin L. Hurd, and Panos Roussos. 2018. ‘An Atlas of Chromatin Accessibility in the Adult Human Brain’. *Genome Research* 28 (8): 1243–52. https://doi.org/10.1101/gr.232488.117.

Gallagher, Michael D., and Alice S. Chen-Plotkin. 2018. ‘The Post-GWAS Era: From Association to Function’. *American Journal of Human Genetics* 102 (5): 717–30. https://doi.org/10.1016/j.ajhg.2018.04.002.

Gao, Peng, Yasin Uzun, Bing He, Sarah E. Salamati, Julie K. M. Coffey, Eva Tsalikian, and Kai Tan. 2019. ‘Risk Variants Disrupting Enhancers of TH1 and TREG Cells in Type 1 Diabetes’. *Proceedings of the National Academy of Sciences* 116 (15): 7581–90. https://doi.org/10.1073/pnas.1815336116.

Garcia-Borreguero, Diego, Ralf Kohnen, Michael H. Silber, John W. Winkelman, Christopher J. Earley, Birgit Högl, Mauro Manconi, Jacques Montplaisir, Yuichi Inoue, and Richard P. Allen. 2013. ‘The Long-Term Treatment of Restless Legs Syndrome/Willis-Ekbom Disease: Evidence-Based Guidelines and Clinical Consensus Best Practice Guidance: A Report from the International Restless Legs Syndrome Study Group’. *Sleep Medicine* 14 (7): 675–84. https://doi.org/10.1016/j.sleep.2013.05.016.

Garcia-Borreguero, Diego, Michael H. Silber, John W. Winkelman, Birgit Högl, Jacquelyn Bainbridge, Mark Buchfuhrer, Georgios Hadjigeorgiou, et al. 2016. 'Guidelines for the First-Line Treatment of Restless Legs Syndrome/Willis–Ekbom Disease, Prevention and Treatment of Dopaminergic Augmentation: A Combined Task Force of the IRLSSG, EURLSSG, and the RLS-Foundation'. *Sleep Medicine* 21 (May): 1–11. https://doi.org/10.1016/j.sleep.2016.01.017.

Giral, Hector, Ulf Landmesser, and Adelheid Kratzer. 2018. 'Into the Wild: GWAS Exploration of Non-Coding RNAs'. *Frontiers in Cardiovascular Medicine* 5. https://doi.org/10.3389/fcvm.2018.00181.

Godau, Jana, Uwe Klose, Adriana Di Santo, Katherine Schweitzer, and Daniela Berg. 2008. 'Multiregional Brain Iron Deficiency in Restless Legs Syndrome'. *Movement Disorders: Official Journal of the Movement Disorder Society* 23 (8): 1184–87. https://doi.org/10.1002/mds.22070.

Gupta, Rajat M., Joseph Hadaya, Aditi Trehan, Seyedeh M. Zekavat, Carolina Roselli, Derek Klarin, Connor A. Emdin, et al. 2017. 'A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression'. *Cell* 170 (3): 522-533.e15. https://doi.org/10.1016/j.cell.2017.06.049.

Hammerschlag, Anke R., Sven Stringer, Christiaan A. de Leeuw, Suzanne Sniekers, Erdogan Taskesen, Kyoko Watanabe, Tessa F. Blanken, et al. 2017. 'Genome-Wide Association Analysis of Insomnia Complaints Identifies Risk Genes and Genetic Overlap with Psychiatric and Metabolic Traits'. *Nature Genetics* 49 (11): 1584–92. https://doi.org/10.1038/ng.3888.

Harismendy, Olivier, Dimple Notani, Xiaoyuan Song, Nazli G. Rahim, Bogdan Tanasa, Nathaniel Heintzman, Bing Ren, et al. 2011. '9p21 DNA Variants Associated with Coronary Artery Disease Impair Interferon-γ Signalling Response'. *Nature* 470 (7333): 264–68. https://doi.org/10.1038/nature09753.

Hartogh, Sabine C. den, Katherine Wolstencroft, Christine L. Mummery, and Robert Passier. 2016. 'A Comprehensive Gene Expression Analysis at Sequential Stages of in Vitro Cardiac Differentiation from Isolated MESP1-Expressing-Mesoderm Progenitors'. *Scientific Reports* 6 (January). https://doi.org/10.1038/srep19386.

Hau, Ann-Christin, Britta Moyo Grebbin, Zsuzsa Agoston, Marie Anders-Maurer, Tamara Müller, Anja Groß, Jasmine Kolb, Julian D. Langer, Claudia Döring, and Dorothea Schulte. 2017. 'MEIS Homeodomain Proteins Facilitate PARP1/ARTD1-Mediated Eviction of Histone H1'. *The Journal of Cell Biology* 216 (9): 2715–29. https://doi.org/10.1083/jcb.201701154.

Hua, Junjie Tony, Musaddeque Ahmed, Haiyang Guo, Yuzhe Zhang, Sujun Chen, Fraser Soares, Jennifer Lu, et al. 2018. 'Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through LncRNA PCAT19'. *Cell* 174 (3): 564-575.e18. https://doi.org/10.1016/j.cell.2018.06.014.

Karroum, Elias Georges, Smaranda Leu-Semenescu, and Isabelle Arnulf. 2012. 'Topography of the Sensations in Primary Restless Legs Syndrome'. *Journal of the Neurological Sciences* 320 (1–2): 26–31. https://doi.org/10.1016/j.jns.2012.05.051.

Kemlink, D., O. Polo, B. Frauscher, V. Gschliesser, B. Högl, W. Poewe, P. Vodicka, et al. 2009. 'Replication of Restless Legs Syndrome Loci in Three European Populations'. *Journal of Medical Genetics* 46 (5): 315–18. https://doi.org/10.1136/jmg.2008.062992.

Kemlink, David, Giuseppe Plazzi, Roberto Vetrugno, Federica Provini, Olli Polo, Karin Stiasny-Kolster, Wolfgang Oertel, et al. 2008. 'Suggestive Evidence for Linkage for Restless Legs Syndrome on Chromosome 19p13'. *Neurogenetics* 9 (2): 75–82. https://doi.org/10.1007/s10048-007-0113-1.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and and David Haussler. 2002. 'The Human Genome Browser at UCSC'. *Genome Research* 12 (6): 996–1006. https://doi.org/10.1101/gr.229102.

Kentepozidou, Elissavet, Sarah J. Aitken, Christine Feig, Klara Stefflova, Ximena Ibarra-Soria, Duncan T. Odom, Maša Roller, and Paul Flicek. 2020. 'Clustered CTCF Binding Is an Evolutionary Mechanism to Maintain Topologically Associating Domains'. *Genome Biology* 21 (1): 5. https://doi.org/10.1186/s13059-019-1894-x.

Khan, Farhan H., Caitlyn D. Ahlberg, Christopher A. Chow, Divya R. Shah, and Brian B. Koo. 2017. 'Iron, Dopamine, Genetics, and Hormones in the Pathophysiology of Restless Legs Syndrome'. *Journal of Neurology* 264 (8): 1634–41. https://doi.org/10.1007/s00415-017-8431-1.

Koo, Brian B. 2015. 'Restless Leg Syndrome Across the Globe: Epidemiology of the Restless Legs Syndrome/Willis-Ekbom Disease'. *Sleep Medicine Clinics* 10 (3): 189–205, xi. https://doi.org/10.1016/j.jsmc.2015.05.004.

Krijger, Peter H. L., Geert Geeven, Valerio Bianchi, Catharina R. E. Hilvering, and Wouter de Laat. 2020. '4C-Seq from Beginning to End: A Detailed Protocol for Sample Preparation and Data Analysis'. *Methods (San Diego, Calif.)* 170 (January): 17–32. https://doi.org/10.1016/j.ymeth.2019.07.014.

Krishna, Abhimanyu, Maria Biryukov, Christophe Trefois, Paul MA Antony, Rene Hussong, Jake Lin, Merja Heinäniemi, et al. 2014. 'Systems Genomics Evaluation of the SH-SY5Y Neuroblastoma Cell Line as a Model for Parkinson's Disease'. *BMC Genomics* 15 (1). https://doi.org/10.1186/1471-2164-15-1154.

Krude, Torsten. 1995. 'Chromatin: Nucleosome Assembly during DNA Replication'. *Current Biology* 5 (11): 1232–34. https://doi.org/10.1016/S0960-9822(95)00245-4.

Kubo, Naoki, Haruhiko Ishii, Xiong Xiong, Simona Bianco, Franz Meitinger, Rong Hu, James D. Hocker, et al. 2021. 'Promoter-Proximal CTCF Binding Promotes Distal Enhancer-Dependent Gene Activation'. *Nature Structural & Molecular Biology* 28 (2): 152–61. https://doi.org/10.1038/s41594-020-00539-5.

Lane, Jacqueline M., Jingjing Liang, Irma Vlasac, Simon G. Anderson, David A. Bechtold, Jack Bowden, Richard Emsley, et al. 2017. 'Genome-Wide Association Analyses of Sleep Disturbance Traits Identify New Loci and Highlight Shared Genetics with Neuropsychiatric and Metabolic Traits'. *Nature Genetics* 49 (2): 274–81. https://doi.org/10.1038/ng.3749.

Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lawrence, H. J., S. Rozenfeld, C. Cruz, K. Matsukuma, A. Kwong, L. Kömüves, A. M. Buchberg, and C. Largman. 1999. 'Frequent Co-Expression of the HOXA9 and MEIS1 Homeobox Genes in Human Myeloid Leukemias'. *Leukemia* 13 (12): 1993–99. https://doi.org/10.1038/sj.leu.2401578.

Lee, Seunggeung, Gonçalo R. Abecasis, Michael Boehnke, and Xihong Lin. 2014. 'Rare-Variant Association Analysis: Study Designs and Statistical Tests'. *American Journal of Human Genetics* 95 (1): 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009.

Levchenko, A., S. Provost, J. Y. Montplaisir, L. Xiong, J. St-Onge, P. Thibodeau, J. B. Rivière, et al. 2006. 'A Novel Autosomal Dominant Restless Legs Syndrome Locus Maps to Chromosome 20p13'. *Neurology* 67 (5): 900–901. https://doi.org/10.1212/01.wnl.0000233991.20410.b6.

Levchenko, Anastasia, Jacques-Yves Montplaisir, Géraldine Asselin, Sylvie Provost, Simon L. Girard, Lan Xiong, Emmanuelle Lemyre, et al. 2009. 'Autosomal-Dominant Locus for Restless Legs Syndrome in French-Canadians on Chromosome 16p12.1'. *Movement Disorders* 24 (1): 40–50. https://doi.org/10.1002/mds.22263.

Lewis, Cathryn M., and Jo Knight. 2012. 'Introduction to Genetic Association Studies'. *Cold Spring Harbor Protocols* 2012 (3): 297–306. https://doi.org/10.1101/pdb.top068163.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Qunhua, James B. Brown, Haiyan Huang, and Peter J. Bickel. 2011. 'Measuring Reproducibility of High-Throughput Experiments'. *Ann. Appl. Stat.* 5 (3): 1752–79. https://doi.org/10.1214/11-AOAS466.

Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. 'Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome'. *Science* 326 (5950): 289–93. https://doi.org/10.1126/science.1181369.

Liebetanz, K. M., J. Winkelmann, C. Trenkwalder, B. Pütz, M. Dichgans, T. Gasser, and B. Müller-Myhsok. 2006. 'RLS3: Fine-Mapping of an Autosomal Dominant Locus in a Family with Intrafamilial Heterogeneity'. *Neurology* 67 (2): 320–21. https://doi.org/10.1212/01.wnl.0000224886.65213.b5.

Lin, Jhih-Rong, Daniel Jaroslawicz, Ying Cai, Quanwei Zhang, Zhen Wang, and Zhengdong D. Zhang. 2018. 'PGA: Post-GWAS Analysis for Disease Gene Identification'. *Bioinformatics* 34 (10): 1786–88. https://doi.org/10.1093/bioinformatics/btx845.

Liu, Jiangying, Ya-Zhen Qin, Shenmiao Yang, Yazhe Wang, Ying-Jun Chang, Ting Zhao, Qian Jiang, and Xiao-Jun Huang. 2017. 'Meis1 Is Critical to the Maintenance of Human Acute Myeloid Leukemia Cells Independent of MLL Rearrangements'. *Annals of Hematology* 96 (4): 567–74. https://doi.org/10.1007/s00277-016-2913-6.

Lu, Leina, Xiaoxiao Liu, Wei-Kai Huang, Paola Giusti-Rodríguez, Jian Cui, Shanshan Zhang, Wanying Xu, et al. 2020. 'Robust Hi-C Maps of Enhancer-Promoter Interactions

Reveal the Function of Non-Coding Genome in Neural Development and Diseases'. *Molecular Cell* 79 (3): 521-534.e15. https://doi.org/10.1016/j.molcel.2020.06.007.

Lyu, Shangru, Hong Xing, Yuning Liu, Pallavi Girdhar, Keer Zhang, Fumiaki Yokoi, Rui Xiao, and Yuqing Li. 2020. 'Deficiency of Meis1, a Transcriptional Regulator, in Mice and Worms: Neurochemical and Behavioral Characterizations with Implications in the Restless Legs Syndrome'. *Journal of Neurochemistry* 155 (5): 522–37. https://doi.org/10.1111/jnc.15177.

Ma, Lixiang, Baoyang Hu, Yan Liu, Scott Christopher Vermilyea, Huisheng Liu, Lu Gao, Yan Sun, Xiaoqing Zhang, and Su-Chun Zhang. 2012. 'Human Embryonic Stem Cell-Derived GABA Neurons Correct Locomotion Deficits in Quinolinic Acid-Lesioned Mice'. *Cell Stem Cell* 10 (4): 455–64. https://doi.org/10.1016/j.stem.2012.01.021.

Machon, Ondrej, Jan Masek, Olga Machonova, Stefan Krauss, and Zbynek Kozmik. 2015. 'Meis2 Is Essential for Cranial and Cardiac Neural Crest Development'. *BMC Developmental Biology* 15 (November). https://doi.org/10.1186/s12861-015-0093-6.

Maeda, Ryu, Kathleen Mood, Teri L. Jones, Jun Aruga, Arthur M. Buchberg, and Ira O. Daar. 2001. 'Xmeis1 , a Protooncogene Involved in Specifying Neural Crest Cell Fate in Xenopus Embryos'. *Oncogene* 20 (11): 1329–42. https://doi.org/10.1038/sj.onc.1204250.

Maher, Brendan. 2008. 'Personal Genomes: The Case of the Missing Heritability'. *Nature* 456 (7218): 18–21. https://doi.org/10.1038/456018a.

Mahmoud, Ahmed I., Fatih Kocabas, Shalini A. Muralidhar, Wataru Kimura, Ahmed S. Koura, Suwannee Thet, Enzo R. Porrello, and Hesham A. Sadek. 2013. 'Meis1 Regulates Postnatal Cardiomyocyte Cell Cycle Arrest'. *Nature* 497 (7448): 249–53. https://doi.org/10.1038/nature12054.

Maisel, Martina, Alexander Herr, Javorina Milosevic, Andreas Hermann, Hans-Jörg Habisch, Sigrid Schwarz, Matthias Kirsch, et al. 2007. 'Transcription Profiling of Adult and Fetal Human Neuroprogenitors Identifies Divergent Paths to Maintain the Neuroprogenitor Cell State'. *Stem Cells (Dayton, Ohio)* 25 (5): 1231–40. https://doi.org/10.1634/stemcells.2006-0617.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. 'Finding the Missing Heritability of Complex Diseases'. *Nature* 461 (7265): 747–53. https://doi.org/10.1038/nature08494.

Manuel, Martine, Ben Martynoga, Tian Yu, John D. West, John O. Mason, and David J. Price. 2010. 'The Transcription Factor Foxg1 Regulates the Competence of Telencephalic Cells to Adopt Subpallial Fates in Mice'. *Development (Cambridge, England)* 137 (3): 487–97. https://doi.org/10.1242/dev.039800.

Marcos, Séverine, Monica González-Lázaro, Leonardo Beccari, Laura Carramolino, Maria Jesus Martin-Bermejo, Oana Amarie, Daniel Mateos-San Martín, et al. 2015. 'Meis1 Coordinates a Network of Genes Implicated in Eye Development and Microphthalmia'. *Development* 142 (17): 3009–20. https://doi.org/10.1242/dev.122176.

Markenscoff-Papadimitriou, Eirene, Sean Whalen, Pawel Przytycki, Reuben Thomas, Fadya Binyameen, Tomasz J. Nowakowski, Arnold R. Kriegstein, et al. 2020. 'A Chromatin

Accessibility Atlas of the Developing Human Telencephalon'. *Cell* 182 (3): 754-769.e18. https://doi.org/10.1016/j.cell.2020.06.002.

Maroof, Asif M., Sotirios Keros, Jennifer A. Tyson, Shui-Wang Ying, Yosif M. Ganat, Florian T. Merkle, Becky Liu, et al. 2013. 'Directed Differentiation and Functional Maturation of Cortical Interneurons from Human Embryonic Stem Cells'. *Cell Stem Cell* 12 (5): 559–72. https://doi.org/10.1016/j.stem.2013.04.008.

Meneely, Samantha, Mai-Lynne Dinkins, Miki Kassai, Shangru Lyu, Yuning Liu, Chien-Te Lin, Kori Brewer, Yuqing Li, and Stefan Clemens. 2018. 'Differential Dopamine D1 and D3 Receptor Modulation and Expression in the Spinal Cord of Two Mouse Models of Restless Legs Syndrome'. *Frontiers in Behavioral Neuroscience* 12 (September). https://doi.org/10.3389/fnbeh.2018.00199.

Mercader, N., E. Leonardo, N. Azpiazu, A. Serrano, G. Morata, C. Martínez, and M. Torres. 1999. 'Conserved Regulation of Proximodistal Limb Axis Development by Meis1/Hth'. *Nature* 402 (6760): 425–29. https://doi.org/10.1038/46580.

Mercader, Nadia, Licia Selleri, Luis Miguel Criado, Pilar Pallares, Carlos Parras, Michael L. Cleary, and Miguel Torres. 2009. 'Ectopic Meis1 Expression in the Mouse Limb Bud Alters P-D Patterning in a Pbx1-Independent Manner'. *The International Journal of Developmental Biology* 53 (8–10): 1483–94. https://doi.org/10.1387/ijdb.072430nm.

Meuleman, Wouter, Alexander Muratov, Eric Rynes, Jessica Halow, Kristen Lee, Daniel Bates, Morgan Diegel, et al. 2020. 'Index and Biological Spectrum of Human DNase I Hypersensitive Sites'. *Nature* 584 (7820): 244–51. https://doi.org/10.1038/s41586-020-2559-3.

Mirza, Aashiq H., Simranjeet Kaur, Caroline A. Brorsson, and Flemming Pociot. 2014. 'Effects of GWAS-Associated Genetic Variants on LncRNAs within IBD and T1D Candidate Loci'. *PLoS ONE* 9 (8). https://doi.org/10.1371/journal.pone.0105723.

Mizuno, Soichi, Takumi Mihara, Tsuyoshi Miyaoka, Takuzi Inagaki, and Jun Horiguchi. 2005. 'CSF Iron, Ferritin and Transferrin Levels in Restless Legs Syndrome'. *Journal of Sleep Research* 14 (1): 43–47. https://doi.org/10.1111/j.1365-2869.2004.00403.x.

Montplaisir, J., S. Boucher, G. Poirier, G. Lavigne, O. Lapierre, and P. Lespérance. 1997. 'Clinical, Polysomnographic, and Genetic Characteristics of Restless Legs Syndrome: A Study of 133 Patients Diagnosed with New Standard Criteria'. *Movement Disorders: Official Journal of the Movement Disorder Society* 12 (1): 61–65. https://doi.org/10.1002/mds.870120111.

Moskow, J. J., F. Bullrich, K. Huebner, I. O. Daar, and A. M. Buchberg. 1995. 'Meis1, a PBX1-Related Homeobox Gene Involved in Myeloid Leukemia in BXH-2 Mice'. *Molecular and Cellular Biology* 15 (10): 5434–43. https://doi.org/10.1128/mcb.15.10.5434.

Navarro Gonzalez, Jairo, Ann S. Zweig, Matthew L. Speir, Daniel Schmelter, Kate R. Rosenbloom, Brian J. Raney, Conner C. Powell, et al. 2021. 'The UCSC Genome Browser Database: 2021 Update'. *Nucleic Acids Research* 49 (D1): D1046–57. https://doi.org/10.1093/nar/gkaa1070.

Nóbrega-Pereira, Sandrina, Nicoletta Kessaris, Tonggong Du, Shioko Kimura, Stewart A. Anderson, and Oscar Marín. 2008. 'Postmitotic Nkx2-1 Controls the Migration of Telencephalic Interneurons by Direct Repression of Guidance Receptors'. *Neuron* 59 (5): 733–45. https://doi.org/10.1016/j.neuron.2008.07.024.

Nott, Alexi, Inge R. Holtman, Nicole G. Coufal, Johannes C. M. Schlachetzki, Miao Yu, Rong Hu, Claudia Z. Han, et al. 2019. 'Brain Cell Type–Specific Enhancer–Promoter Interactome Maps and Disease-Risk Association'. *Science* 366 (6469): 1134–39. https://doi.org/10.1126/science.aay0793.

Oexle, Konrad. 2018. 'Power versus Phenotyping Precision of Genome-Wide Association Studies on Sleep Traits'. *Sleep* 41 (11). https://doi.org/10.1093/sleep/zsy211.

Ohayon, Maurice M., Ruth O'Hara, and Michael V. Vitiello. 2012. 'Epidemiology of Restless Legs Syndrome: A Synthesis of the Literature'. *Sleep Medicine Reviews* 16 (4): 283–95. https://doi.org/10.1016/j.smrv.2011.05.002.

O'Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. 'Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation'. *Nucleic Acids Research* 44 (D1): D733-745. https://doi.org/10.1093/nar/gkv1189.

Ondo, W. G., K. D. Vuong, and Q. Wang. 2000. 'Restless Legs Syndrome in Monozygotic Twins: Clinical Correlates'. *Neurology* 55 (9): 1404–6. https://doi.org/10.1212/wnl.55.9.1404.

Ondo, W., and J. Jankovic. 1996. 'Restless Legs Syndrome: Clinicoetiologic Correlates'. *Neurology* 47 (6): 1435–41. https://doi.org/10.1212/wnl.47.6.1435.

Ou, Minghui, Xia Li, Shibo Zhao, Shichao Cui, and Jie Tu. 2020. 'Long Non-Coding RNA CDKN2B-AS1 Contributes to Atherosclerotic Plaque Formation by Forming RNA-DNA Triplex in the CDKN2B Promoter'. *EBioMedicine* 55 (April). https://doi.org/10.1016/j.ebiom.2020.102694.

Owa, Tomoo, Shinichiro Taya, Satoshi Miyashita, Mariko Yamashita, Toma Adachi, Koyo Yamada, Miwa Yokoyama, et al. 2018. 'Meis1 Coordinates Cerebellar Granule Cell Development by Regulating Pax6 Transcription, BMP Signaling and Atoh1 Degradation'. *Journal of Neuroscience* 38 (5): 1277–94. https://doi.org/10.1523/JNEUROSCI.1545-17.2017.

Påhlman, S., A. I. Ruusala, L. Abrahamsson, M. E. Mattsson, and T. Esscher. 1984. 'Retinoic Acid-Induced Differentiation of Cultured Human Neuroblastoma Cells: A Comparison with Phorbolester-Induced Differentiation'. *Cell Differentiation* 14 (2): 135–44. https://doi.org/10.1016/0045-6039(84)90038-1.

Pei, Zhenglei, Bei Wang, Gang Chen, Motoshi Nagao, Masato Nakafuku, and Kenneth Campbell. 2011. 'Homeobox Genes Gsx1 and Gsx2 Differentially Regulate Telencephalic Progenitor Maturation'. *Proceedings of the National Academy of Sciences* 108 (4): 1675–80. https://doi.org/10.1073/pnas.1008824108.

Pfeufer, Arne, Charlotte van Noord, Kristin D. Marciante, Dan E. Arking, Martin G. Larson, Albert Vernon Smith, Kirill V. Tarasov, et al. 2010. 'Genome-Wide Association Study of PR Interval'. *Nature Genetics* 42 (2): 153–59. https://doi.org/10.1038/ng.517.

Picchietti, Daniel L., Oliviero Bruni, Al de Weerd, Jeffrey S. Durmer, Suresh Kotagal, Judith A. Owens, Narong Simakajornboon, and International Restless Legs Syndrome Study Group (IRLSSG). 2013. 'Pediatric Restless Legs Syndrome Diagnostic Criteria: An Update by the International Restless Legs Syndrome Study Group'. *Sleep Medicine* 14 (12): 1253–59. https://doi.org/10.1016/j.sleep.2013.08.778.

Pollard, Katherine S., Melissa J. Hubisz, Kate R. Rosenbloom, and Adam Siepel. 2010. 'Detection of Nonneutral Substitution Rates on Mammalian Phylogenies'. *Genome Research* 20 (1): 110–21. https://doi.org/10.1101/gr.097857.109.

Pomerantz, Mark M., Nasim Ahmadiyeh, Li Jia, Paula Herman, Michael P. Verzi, Harshavardhan Doddapaneni, Christine A. Beckwith, et al. 2009. 'The 8q24 Cancer Risk Variant Rs6983267 Shows Long-Range Interaction with MYC in Colorectal Cancer'. *Nature Genetics* 41 (8): 882–84. https://doi.org/10.1038/ng.403.

Rataj-Baniowska, Monika, Anna Niewiadomska-Cimicka, Marie Paschaki, Monika Szyszka-Niagolov, Laura Carramolino, Miguel Torres, Pascal Dollé, and Wojciech Krężel. 2015. 'Retinoic Acid Receptor β Controls Development of Striatonigral Projection Neurons through FGF-Dependent and Meis1-Dependent Mechanisms'. *Journal of Neuroscience* 35 (43): 14467–75. https://doi.org/10.1523/JNEUROSCI.1278-15.2015.

Royo, José Luis, José Bessa, Carmen Hidalgo, Ana Fernández-Miñán, Juan J. Tena, Yolanda Roncero, José Luis Gómez-Skarmeta, and Fernando Casares. 2012. 'Identification and Analysis of Conserved Cis-Regulatory Regions of the MEIS1 Gene'. *PLoS ONE* 7 (3). https://doi.org/10.1371/journal.pone.0033617.

Salminen, Aaro V., Lillian Garrett, Barbara Schormair, Jan Rozman, Florian Giesert, Kristina M. Niedermeier, Lore Becker, et al. 2017. 'Meis1: Effects on Motor Phenotypes and the Sensorimotor System in Mice'. *Disease Models & Mechanisms* 10 (8): 981–91. https://doi.org/10.1242/dmm.030080.

Salminen, Aaro V., Nathalie Schandra, Barbara Schormair, Konrad Oexle, and Juliane Winkelmann. 2020. 'Therapeutic Effectiveness of Thalidomide in a Patient with Treatment-Resistant Restless Legs Syndrome'. *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine* 16 (10): 1815–17. https://doi.org/10.5664/jcsm.8696.

Salminen, Aaro V., and Juliane Winkelmann. 2018. 'Restless Legs Syndrome and Other Movement Disorders of Sleep-Treatment Update'. *Current Treatment Options in Neurology* 20 (12): 55. https://doi.org/10.1007/s11940-018-0540-3.

Sarayloo, Faezeh, Alexandre Dionne-Laporte, Helene Catoire, Daniel Rochefort, Gabrielle Houle, Jay P. Ross, Fulya Akçimen, et al. 2019. 'Mineral Absorption Is an Enriched Pathway in a Brain Region of Restless Legs Syndrome Patients with Reduced MEIS1 Expression'. *PLOS ONE* 14 (11): e0225186. https://doi.org/10.1371/journal.pone.0225186.

Schoenfelder, Stefan, and Peter Fraser. 2019. ‘Long-Range Enhancer–Promoter Contacts in Gene Expression Control’. *Nature Reviews Genetics* 20 (8): 437–55. https://doi.org/10.1038/s41576-019-0128-0.

Schormair, Barbara, David Kemlink, Darina Roeske, Gertrud Eckstein, Lan Xiong, Peter Lichtner, Stephan Ripke, et al. 2008. ‘PTPRD (Protein Tyrosine Phosphatase Receptor Type Delta) Is Associated with Restless Legs Syndrome’. *Nature Genetics* 40 (8): 946–48. https://doi.org/10.1038/ng.190.

Schormair, Barbara, Chen Zhao, Steven Bell, Erik Tilch, Aaro V. Salminen, Benno Pütz, Yves Dauvilliers, et al. 2017. ‘Identification of Novel Risk Loci for Restless Legs Syndrome in Genome-Wide Association Studies in Individuals of European Ancestry: A Meta-Analysis’. *The Lancet Neurology* 16 (11): 898–907. https://doi.org/10.1016/S1474-4422(17)30327-7.

Schulte, Dorothea, and Dirk Geerts. 2019. ‘MEIS Transcription Factors in Development and Disease’. *Development (Cambridge, England)* 146 (16). https://doi.org/10.1242/dev.174706.

Schulte, E.C., F. Knauf, D. Kemlink, B. Schormair, P. Lichtner, C. Gieger, T. Meitinger, and J. Winkelmann. 2011. ‘Variant Screening of the Coding Regions of MEIS1</Em> in Patients with Restless Legs Syndrome’. *Neurology* 76 (12): 1106. https://doi.org/10.1212/WNL.0b013e318211c366.

Schulte, Eva C., Maria Kousi, Perciliz L. Tan, Erik Tilch, Franziska Knauf, Peter Lichtner, Claudia Trenkwalder, et al. 2014. ‘Targeted Resequencing and Systematic in Vivo Functional Testing Identifies Rare Variants in MEIS1 as Significant Contributors to Restless Legs Syndrome’. *American Journal of Human Genetics* 95 (1): 85–95. https://doi.org/10.1016/j.ajhg.2014.06.005.

Silber, Michael H., Philip M. Becker, Mark J. Buchfuhrer, Christopher J. Earley, William G. Ondo, Arthur S. Walters, John W. Winkelman, and Scientific and Medical Advisory Board, Restless Legs Syndrome Foundation. 2018. ‘The Appropriate Use of Opioids in the Treatment of Refractory Restless Legs Syndrome’. *Mayo Clinic Proceedings* 93 (1): 59–67. https://doi.org/10.1016/j.mayocp.2017.11.007.

Silber, Michael H., Philip M. Becker, Christopher Earley, Diego Garcia-Borreguero, William G. Ondo, and Medical Advisory Board of the Willis-Ekbom Disease Foundation. 2013. ‘Willis-Ekbom Disease Foundation Revised Consensus Statement on the Management of Restless Legs Syndrome’. *Mayo Clinic Proceedings* 88 (9): 977–86. https://doi.org/10.1016/j.mayocp.2013.06.016.

Sinha, Surajit, and James K. Chen. 2006. ‘Purmorphamine Activates the Hedgehog Pathway by Targeting Smoothened’. *Nature Chemical Biology* 2 (1): 29–30. https://doi.org/10.1038/nchembio753.

Sloan, Cricket A., Esther T. Chan, Jean M. Davidson, Venkat S. Malladi, J. Seth Strattan, Benjamin C. Hitz, Idan Gabdank, et al. 2016. ‘ENCODE Data at the ENCODE Portal’. *Nucleic Acids Research* 44 (D1): D726-732. https://doi.org/10.1093/nar/gkv1160.

Smemo, Scott, Juan J. Tena, Kyoung-Han Kim, Eric R. Gamazon, Noboru J. Sakabe, Carlos Gómez-Marín, Ivy Aneas, et al. 2014. ‘Obesity-Associated Variants within FTO Form Long-

Range Functional Connections with IRX3'. *Nature* 507 (7492): 371–75. https://doi.org/10.1038/nature13138.

Soldner, Frank, Yonatan Stelzer, Chikdu S. Shivalila, Brian J. Abraham, Jeanne C. Latourelle, M. Inmaculada Barrasa, Johanna Goldmann, Richard H. Myers, Richard A. Young, and Rudolf Jaenisch. 2016. 'Parkinson-Associated Risk Variant in Distal Enhancer of α-Synuclein Modulates Target Gene Expression'. *Nature* 533 (7601): 95–99. https://doi.org/10.1038/nature17939.

Song, Lingyun, and Gregory E. Crawford. 2010. 'DNase-Seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells'. *Cold Spring Harbor Protocols* 2010 (2): pdb.prot5384. https://doi.org/10.1101/pdb.prot5384.

Spieker, N., P. van Sluis, M. Beitsma, K. Boon, B. D. van Schaik, A. H. van Kampen, H. Caron, and R. Versteeg. 2001. 'The MEIS1 Oncogene Is Highly Expressed in Neuroblastoma and Amplified in Cell Line IMR32'. *Genomics* 71 (2): 214–21. https://doi.org/10.1006/geno.2000.6408.

Spieler, Derek, Maria Kaffe, Franziska Knauf, José Bessa, Juan J. Tena, Florian Giesert, Barbara Schormair, et al. 2014. 'Restless Legs Syndrome-Associated Intronic Common Variant in Meis1 Alters Enhancer Function in the Developing Telencephalon'. *Genome Research* 24 (4): 592–603. https://doi.org/10.1101/gr.166751.113.

Splinter, Erik, Elzo de Wit, Elphège P. Nora, Petra Klous, Harmen J. G. van de Werken, Yun Zhu, Lucas J. T. Kaaij, et al. 2011. 'The Inactive X Chromosome Adopts a Unique Three-Dimensional Conformation That Is Dependent on Xist RNA'. *Genes & Development* 25 (13): 1371–83. https://doi.org/10.1101/gad.633311.

Stankunas, Kryn, Ching Shang, Karen Y. Twu, Shih-Chu Kao, Nancy A. Jenkins, Neal G. Copeland, Mrinmoy Sanyal, Licia Selleri, Michael L. Cleary, and Ching-Pin Chang. 2008. 'Pbx/Meis Deficiencies Demonstrate Multigenetic Origins of Congenital Heart Disease'. *Circulation Research* 103 (7): 702–9. https://doi.org/10.1161/CIRCRESAHA.108.175489.

Stefansson, Hreinn, David B. Rye, Andrew Hicks, Hjorvar Petursson, Andres Ingason, Thorgeir E. Thorgeirsson, Stefan Palsson, et al. 2007. 'A Genetic Risk Factor for Periodic Limb Movements in Sleep'. *The New England Journal of Medicine* 357 (7): 639–47. https://doi.org/10.1056/NEJMoa072743.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. 2007. 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors'. *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Toresson, Håkan, Malin Parmar, and Kenneth Campbell. 2000. 'Expression of Meis and Pbx Genes and Their Protein Products in the Developing Telencephalon: Implications for Regional Differentiation'. *Mechanisms of Development* 94 (1): 183–87. https://doi.org/10.1016/S0925-4773(00)00324-5.

Torre-Ubieta, Luis de la, Jason L. Stein, Hyejung Won, Carli K. Opland, Dan Liang, Daning Lu, and Daniel H. Geschwind. 2018. 'The Dynamic Landscape of Open Chromatin during

Human Cortical Neurogenesis'. *Cell* 172 (1–2): 289-304.e18. https://doi.org/10.1016/j.cell.2017.12.014.

Trenkwalder, Claudia, Richard Allen, Birgit Högl, Walter Paulus, and Juliane Winkelmann. 2016. 'Restless Legs Syndrome Associated with Major Diseases: A Systematic Review and New Concept'. *Neurology* 86 (14): 1336–43. https://doi.org/10.1212/WNL.0000000000002542.

Uhlén, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. 'Tissue-Based Map of the Human Proteome'. *Science* 347 (6220). https://doi.org/10.1126/science.1260419.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Maido Remm, and Steven G. Rozen. 2012. 'Primer3—New Capabilities and Interfaces'. *Nucleic Acids Research* 40 (15): e115. https://doi.org/10.1093/nar/gks596.

Vilariño-Güell, C., H. Chai, B. H. Keeling, J. E. Young, A. Rajput, T. Lynch, J. O. Aasly, et al. 2009. 'MEIS1 p.R272H IN FAMILIAL RESTLESS LEGS SYNDROME'. *Neurology* 73 (3): 243. https://doi.org/10.1212/WNL.0b013e3181ae7c79.

Visel, Axel, Matthew J. Blow, Zirong Li, Tao Zhang, Jennifer A. Akiyama, Amy Holt, Ingrid Plajzer-Frick, et al. 2009. 'ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers'. *Nature* 457 (7231): 854–58. https://doi.org/10.1038/nature07730.

Walter, Carolin, Daniel Schuetzmann, Frank Rosenbauer, and Martin Dugas. 2014. 'Basic4Cseq: An R/Bioconductor Package for Analyzing 4C-Seq Data'. *Bioinformatics* 30 (22): 3268–69. https://doi.org/10.1093/bioinformatics/btu497.

Walters, A. S. 1995. 'Toward a Better Definition of the Restless Legs Syndrome. The International Restless Legs Syndrome Study Group'. *Movement Disorders: Official Journal of the Movement Disorder Society* 10 (5): 634–42. https://doi.org/10.1002/mds.870100517.

Waskiewicz, Andrew Jan, Holly A. Rikhof, Rafael E. Hernandez, and Cecilia B. Moens. 2001. 'Zebrafish Meis Functions to Stabilize Pbx Proteins and Regulate Hindbrain Patterning'. *Development* 128 (21): 4139–51.

Werken, Harmen J. G. van de, Paula J. P. de Vree, Erik Splinter, Sjoerd J. B. Holwerda, Petra Klous, Elzo de Wit, and Wouter de Laat. 2012. '4C Technology: Protocols and Data Analysis'. *Methods in Enzymology* 513: 89–112. https://doi.org/10.1016/B978-0-12-391938-0.00004-5.

Wijemanne, Subhashie, and William Ondo. 2017. 'Restless Legs Syndrome: Clinical Features, Diagnosis and a Practical Approach to Management'. *Practical Neurology* 17 (6): 444–52. https://doi.org/10.1136/practneurol-2017-001762.

Willis, Thomas, and Eugenius Philiatros. 1977. *The London Practice of Physick*. 680, [10] p. Boston: Longwood Press. //catalog.hathitrust.org/Record/007989700.

Winkelmann, J., T. C. Wetter, V. Collado-Seidel, T. Gasser, M. Dichgans, A. Yassouridis, and C. Trenkwalder. 2000. 'Clinical Characteristics and Frequency of the Hereditary Restless Legs Syndrome in a Population of 300 Patients'. *Sleep* 23 (5): 597–602.

Winkelmann, Juliane, Richard P. Allen, Birgit Högl, Yuichi Inoue, Wolfgang Oertel, Aaro V. Salminen, John W. Winkelman, Claudia Trenkwalder, and Cristina Sampaio. 2018. 'Treatment of Restless Legs Syndrome: Evidence-Based Review and Implications for Clinical Practice (Revised 2017)§'. *Movement Disorders: Official Journal of the Movement Disorder Society* 33 (7): 1077–91. https://doi.org/10.1002/mds.27260.

Winkelmann, Juliane, Darina Czamara, Barbara Schormair, Franziska Knauf, Eva C. Schulte, Claudia Trenkwalder, Yves Dauvilliers, et al. 2011. 'Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1'. *PLoS Genetics* 7 (7). https://doi.org/10.1371/journal.pgen.1002171.

Winkelmann, Juliane, Barbara Schormair, Peter Lichtner, Stephan Ripke, Lan Xiong, Shapour Jalilzadeh, Stephany Fulda, et al. 2007. 'Genome-Wide Association Study of Restless Legs Syndrome Identifies Common Variants in Three Genomic Regions'. *Nature Genetics* 39 (8): 1000–1006. https://doi.org/10.1038/ng2099.

Won, Hyejung, Luis de la Torre-Ubieta, Jason L. Stein, Neelroop N. Parikshak, Jerry Huang, Carli K. Opland, Michael Gandal, et al. 2016. 'Chromosome Conformation Elucidates Regulatory Relationships in Developing Human Brain'. *Nature* 538 (7626): 523–27. https://doi.org/10.1038/nature19847.

Xiang, P., W. Wei, C. Lo, P. Rosten, J. Hou, P. A. Hoodless, M. Bilenky, et al. 2014. 'Delineating MEIS1 Cis- Regulatory Elements Active in Hematopoietic Cells'. *Leukemia* 28 (2): 433–36. https://doi.org/10.1038/leu.2013.287.

Xiong, Lan, Hélène Catoire, Patrick Dion, Claudia Gaspar, Ronald G. Lafrenière, Simon L. Girard, Anastasia Levchenko, et al. 2009. 'MEIS1 Intronic Risk Haplotype Associated with Restless Legs Syndrome Affects Its MRNA and Protein Expression Levels'. *Human Molecular Genetics* 18 (6): 1065–74. https://doi.org/10.1093/hmg/ddn443.

Yager, L. M., A. F. Garcia, A. M. Wunsch, and S. M. Ferguson. 2015. 'The Ins and Outs of the Striatum: Role in Drug Addiction'. *Neuroscience* 301 (August): 529–41. https://doi.org/10.1016/j.neuroscience.2015.06.033.

Yamada, Takeyuki, Yumiko Urano-Tashiro, Saori Tanaka, Hirotada Akiyama, and Fumio Tashiro. 2013. 'Involvement of Crosstalk between Oct4 and Meis1a in Neural Cell Fate Decision'. *PLoS ONE* 8 (2). https://doi.org/10.1371/journal.pone.0056997.

Yuan, Fang, Kai-Heng Fang, Shi-Ying Cao, Zhuang-Yin Qu, Qi Li, Robert Krencik, Min Xu, et al. 2015. 'Efficient Generation of Region-Specific Forebrain Neurons from Human Pluripotent Stem Cells under Highly Defined Condition'. *Scientific Reports* 5 (December). https://doi.org/10.1038/srep18550.

Zeineddine, Dana, Aya Abou Hammoud, Mohamad Mortada, and Hélène Boeuf. 2014. 'The Oct4 Protein: More than a Magic Stemness Marker'. *American Journal of Stem Cells* 3 (2): 74–82.

Zhang, Wei, Yi Sui, Jun Ni, and Tao Yang. 2016. 'Insights into the Nanog Gene: A Propeller for Stemness in Primitive Stem Cells'. *International Journal of Biological Sciences* 12 (11): 1372–81. https://doi.org/10.7150/ijbs.16349.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. 'Model-Based Analysis of ChIP-Seq (MACS)'. *Genome Biology* 9 (9): R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhu, Wenliang, Boya Zhang, Mengqi Li, Fan Mo, Tingwei Mi, Yihui Wu, Zhaoqian Teng, Qi Zhou, Wei Li, and Baoyang Hu. 2019. 'Precisely Controlling Endogenous Protein Dosage in HPSCs and Derivatives to Model FOXG1 Syndrome'. *Nature Communications* 10 (1): 928. https://doi.org/10.1038/s41467-019-08841-7.

APPENDIX

Publicly available datasets used for Figures 3.8, 3.11 and 3.12 (Sloan et al. 2016; Dunham et al. 2012; Davis et al. 2018; Fullard et al. 2018; Markenscoff-Papadimitriou et al. 2020).

We thank the ENCODE consortium for genomic data, particularly the Bing Ren laboratory, Cherry laboratory, the John Stamatoyannopoulos laboratory and ENCODE processing pipeline lab for human DNase-seq data, ChIP-seq and mouse ATAC-seq data.

We downloaded the call sets from the ENCODE portal (Sloan et al. 2016) (https://www.encodeproject.org/) with the link and identifiers listed in the table below:

Sample type	Source	Link
Homo sapiens neural progenitor originated from H9 (Neural progenitors DNse seq, Figure 3.8)	ENCODE	https://www.encodeproject.org/files/ENCFF525GHV/
Homo sapiens neural cell originated from H1 (Neurons 27Ac ChiP seq, Figure 3.8)	ENCODE	https://www.encodeproject.org/files/ENCFF574HAY/
Homo sapiens neural cell originated from H1 (Neurons EP300 ChiP seq,Figure 3.8	ENCODE	https://www.encodeproject.org/files/ENCFF644YCV/
Homo sapiens heart tissue embryo (96 days) (Heart DNAse-seq, Figure 3.8)	ENCODE	https://www.encodeproject.org/files/ENCFF164XPV/
Homo sapiens cardiac muscle cell originated from RUES2 (Cardiomyocytes DNAse-seq, Figure 3.8)	ENCODE	https://www.encodeproject.org/files/ENCFF446JBL/
Homo sapiens hematopoietic multipotent progenitor cell (Hematopoetic progenitors DNAse-seq, Figure 3.8)	ENCODE	https://www.encodeproject.org/files/ENCFF157SRM/

Homo sapiens LGE gw19 (Fetal LGE ATAC-seq, Figure 3.8)	Markenscoff-Papadimitriou et al. 2020	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4495220
Homo sapiens putamen (neuronal) Putamen ATAC-seq, Figure 3.8)	Fullard et al. 2018	https://bendlj01.u.hpc.mssm.edu/multireg/ Putamen (PUT) Bigwig Hg19, neuronal

Sample type	**Source**	**Link**
Mus musculus forebrain tissue embryo (11.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF326ULQ/
Mus musculus forebrain tissue embryo (12.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF541FKK/
Mus musculus forebrain tissue embryo (13.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF633ETU/
Mus musculus forebrain tissue embryo (14.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF610RKB/
Mus musculus forebrain tissue embryo (15.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF830GBM/
Mus musculus forebrain tissue embryo (16.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF878NFB/
Mus musculus forebrain postnatal (0 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF561KNB/
Mus musculus midbrain tissue embryo (11.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF434ZPH/
Mus musculus midbrain tissue embryo (12.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF183IJY/

Mus musculus midbrain tissue embryo (13.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF127TSE/
Mus musculus midbrain tissue embryo (14.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF320DPN/
Mus musculus midbrain tissue embryo (15.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF206RZM/
Mus musculus midbrain tissue embryo (16.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF304QSF/
Mus musculus midbrain postnatal (0 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF330GOY/

Sample type	**Source**	**Link**
Mus musculus hindbrain tissue embryo (11.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF893VVC/
Mus musculus hindbrain tissue embryo (12.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF989MKH/
Mus musculus hindbrain tissue embryo (13.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF785ICE/
Mus musculus hindbrain tissue embryo (14.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF675IHZ/
Mus musculus hindbrain tissue embryo (15.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF150ZKN/
Mus musculus hindbrain tissue embryo (16.5 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF246SGJ/
Mus musculus hindbrain postnatal (0 days) Figure 3.10	ENCODE	https://www.encodeproject.org/files/ENCFF950WUT/

Mus musculus heart tissue embryo (11.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF795NRM/
Mus musculus heart tissue embryo (12.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF579BIF/
Mus musculus heart tissue embryo (13.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF299CYY/
Mus musculus heart tissue embryo (14.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF132DQX/
Mus musculus heart tissue embryo (15.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF297GJB/
Mus musculus heart tissue embryo (16.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF448TQN/
Mus musculus heart postnatal (0 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF384BWM/

Sample type	**Source**	**Link**
Mus musculus limb tissue embryo (11.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF569URC/
Mus musculus limb tissue embryo (12.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF441VZK/
Mus musculus limb tissue embryo (13.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF591DUQ/
Mus musculus limb tissue embryo (14.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF527PDO/
Mus musculus limb tissue embryo (15.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF976BNK/

Mus musculus liver tissue embryo (11.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF362ALT/
Mus musculus liver tissue embryo (12.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF906NMD/
Mus musculus liver tissue embryo (13.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF548KDI/
Mus musculus liver tissue embryo (14.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF081HFB/
Mus musculus liver tissue embryo (15.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/experiments/ENCSR465PYP/
Mus musculus liver tissue embryo (16.5 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF539XOU/
Mus musculus liver postnatal (0 days) Figure 3.11	ENCODE	https://www.encodeproject.org/files/ENCFF389JBE/

ACKNOWLEDGMENTS

I would like to thank to Prof. Dr. Juliane Winkelmann for giving me the opportunity to pursue my doctoral thesis at the Institute of Neurogenomics. I am grateful for constant support and motivation to bring my work to the final steps. It was very exciting to enter the world of research!

A very special thank you is for my mentor and supervisor Dr. Daniel Lam who guided me through all stages of my work. I am very grateful for your tremendous support and help. I have learned a lot from you, and it was a great pleasure to work with you!

Furthermore, I would like to thank to Dr. Barbara Schormair for her kind help upon my arrival at the Institute and for having an answer to all my questions and doubts. I would like to express a deep gratitude to Dr. Aaro Salminen for helpful discussions and his help for preparing my talks and to Dr. Erik Tilch for his support in learning basic laboratory techniques at the beginning of my PhD. I am very grateful to Irmgard Zaus, Monika Zimmerman and Julia Vandrey for their assistance every time I needed, especially during my pregnancy. You are great colleagues!

Finally, I would like to thank my husband Aleksandar who motivated me all this time and my daughter Anja for cheering me up. I owe a special thank you to my parents Svetlana and Slavoljub for constant love and support. This work is dedicated to them.

Mama, Slavo hvala vam na svemu!