Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development

Elena Perenthaler[†], Soheil Yousefi[†], Eva Niggl[†] and Tahsin Stefan Barakat*

Department of Clinical Genetics, Erasmus MC – University Medical Center, Rotterdam, Netherlands

The development of the human cerebral cortex is a complex and dynamic process, in which neural stem cell proliferation, neuronal migration, and post-migratory neuronal organization need to occur in a well-organized fashion. Alterations at any of these crucial stages can result in malformations of cortical development (MCDs), a group of genetically heterogeneous neurodevelopmental disorders that present with developmental delay, intellectual disability and epilepsy. Recent progress in genetic technologies, such as next generation sequencing, most often focusing on all protein-coding exons (e.g., whole exome sequencing), allowed the discovery of more than a 100 genes associated with various types of MCDs. Although this has considerably increased the diagnostic yield, most MCD cases remain unexplained. As Whole Exome Sequencing investigates only a minor part of the human genome (1–2%), it is likely that patients, in which no disease-causing mutation has been identified, could harbor mutations in genomic regions beyond the exome. Even though functional annotation of non-coding regions is still lagging behind that of protein-coding genes, tremendous progress has been made in the field of gene regulation. One group of non-coding regulatory regions are enhancers, which can be distantly located upstream or downstream of genes and which can mediate temporal and tissue-specific transcriptional control via long-distance interactions with promoter regions. Although some examples exist in literature that link alterations of enhancers to genetic disorders, a widespread appreciation of the putative roles of these sequences in MCDs is still lacking. Here, we summarize the current state of knowledge on *cis*-regulatory regions and discuss novel technologies such as massively-parallel reporter assay systems, CRISPR-Cas9-based screens and computational approaches that help to further elucidate the emerging role of the non-coding genome in disease. Moreover, we discuss existing literature on mutations or copy number alterations of regulatory regions involved in brain development. We foresee that the future implementation of the knowledge obtained through ongoing gene regulation studies will benefit patients and will provide an explanation to part of the missing heritability of MCDs and other genetic disorders.

Keywords: malformations of cortical development, gene regulation, *cis*-regulatory elements, enhancers, epigenome, functional genomics, massively parallel reporter assays, bioinformatics

INTRODUCTION

The brain lies at the foundation of what makes us human, as it not only regulates most of our body functions, but it is also central to our cognition and thoughts, defining our personalities, behavior and social interactions. How such a complex organ is formed during development has fascinated biologists for centuries, and we are now living in a technology driven era where knowledge gained through various disciplines such as medicine, biotechnology, computational biology, and neuroscience enables us for the first time to get a glimpse on how these intricate processes are genetically regulated. Understanding the developmental biology of the human brain is not only of utmost importance for satisfying our own natural curiosity of what makes us human, but also promises improvements and therapeutic options for various disorders that affect the human brain, including neurodevelopmental and neurodegenerative diseases, which, together, are a burden on society and have a negative effect on the quality of life of individuals (Genereaux et al., 2015; Jonsson et al., 2017). One of the traditionally best studied parts of the human brain is the cerebral cortex, which is responsible for cognition and sensorimotor activity. The development of the cerebral cortex is a complex and dynamic process organized in three major steps: (I) neural stem cell proliferation, (II) neuronal migration toward the cortical plate, and (III) post-migratory organization (for further review see, Agirman et al., 2017; Van Essen et al., 2018; Seto and Eiraku, 2019). Alterations in any of these steps can be responsible for the development of a group of disorders overall known as malformations of cortical development (MCD), a common cause of neurodevelopmental delay, intellectual disability and epilepsy (Guerrini and Dobyns, 2014).

Malformations of cortical development are a heterogeneous group of disorders originally classified into three main groups, reflecting the stage of cortical development at which the defect arises (Barkovich et al., 1996). To date, more than a 100 genes associated with MCDs have been described (Guerrini and Dobyns, 2014), the majority of which encodes for proteins that function in general neurodevelopmental processes such as cell cycle regulation, neuronal migration, and polarization. These include classically studied MCD genes such as *LIS1*, that, when mutated, causes lissencephaly (Reiner et al., 1993), but also a flood of recently identified genes such as *TLE1* (Cavallin et al., 2018), *GRIN1* (Fry et al., 2018), *CRADD* (Di Donato et al., 2016), *DIAPH1* (Ercan-Sencicek et al., 2015), *WDR62* (Bilguvar et al., 2010), *RTTN* (Kheradmand Kia et al., 2012), *ZIC1* (Vandervore et al., 2018), *MACF1* (Dobyns et al., 2018), and many more (Lu et al., 2018) that could only be implicated in MCD due to the advent of next-generation DNA sequencing technologies, which enable large-scale genomic investigations in an unbiased, hypothesis-free manner.

Potential disease-causing genetic mutations can be detected by whole exome sequencing (WES), a method of sequencing all protein-coding exons in a genome. Although the implementation of WES in the diagnostic process improved the diagnostic yield of Mendelian disorders to ~25–30% (Yang et al., 2013), still many cases of MCDs remain unexplained (de Wit et al., 2008).

This holds true even for cases where multiple affected individuals are found in the same family, or other environmental causes have been excluded, strongly hinting at a genetic cause. Even though the diagnostic yield for some MCD groups displaying very defined features on brain imaging, such as lissencephaly (characterized by a smooth brain surface with absent gyri), can reach up to 80% (Di Donato et al., 2018), for the vast majority of cases of the remaining MCD spectrum the yield is much lower (Wiszniewski et al., 2018). Cases displaying this "missing heritability" are often reasoned to be caused by somatic mutations (Jamuar et al., 2014; Gonzalez-Moron et al., 2017), mosaicism (McMahon et al., 2015; Mirzaa et al., 2016; Zillhardt et al., 2016) or by non-genetic causes, such as viral infections (Bosnjak et al., 2011; Moore et al., 2017). However, as WES only interrogates the 1–2% of the human genome that encodes for proteins (Consortium, 2012), it is tempting to speculate that at least some of this missing heritability of MCD might be caused or influenced by genetic variation in the non-coding genome. This hypothesis is supported by several arguments. First, genome-wide association (GWAS) studies on multiple diseases have shown that more than 90% of disease-associated single nucleotide polymorphisms (SNPs) are located outside of coding genes (Maurano et al., 2012), therefore potentially in regions involved in transcriptional regulation. Second, the last decade has witnessed an enormous progress in our understanding of mechanisms involved in gene regulation that find their origin in the non-coding genome, and it has become clear that aberrant gene regulation can cause a variety of genetic disorders (Zeng et al., 2015; Smith and Flodman, 2018; Spielmann et al., 2018). Key elements in the non-coding genome such as promoters, insulators and enhancers, the latter also referred to as non-coding regulatory elements (NCREs), ensure that genes are turned on or off at the right moment in time. When this tight spatio-temporal regulation is disturbed, gene expression can be affected which could result in a genetic disorder. Although only very few large-scale genetic studies have investigated the role of the non-coding genome in genetic disorders (Doan et al., 2016; Devanna et al., 2018; Short et al., 2018) it is clear from the number of excellent studies that have recently been published (Lettice et al., 2002, 2003; Benko et al., 2009; Smemo et al., 2012; Weedon et al., 2014; Ngcungcu et al., 2017; Protas et al., 2017; Bouman et al., 2018; Gloss and Dinger, 2018; Mehrjouy et al., 2018; Potuijt et al., 2018), that it will only be a matter of time until the community fully realizes the importance of the non-coding genome in health and disease. Finally, one and the same mutation can show different degrees of severity in different patients, and it is likely that this phenotypic variability could be influenced by genetic variations outside of coding genes influencing gene expression (Castel et al., 2018; Niemi et al., 2018).

In this *Review*, we will first provide a concise overview of the current state-of-the-art of the field of gene regulation and the role herein of the non-coding genome, with a particular focus on enhancers. We will review recently developed technology and computational approaches that will facilitate future investigations on non-coding causes of MCDs and will discuss examples of non-coding alterations causing genetic diseases. We conclude this *Review* by providing an example of a strategy to identify regulatory alterations in patients with MCDs. Together, this

will provide a perspective on how "missing heritability" will be identified in the near future and how to move beyond the borders of the exome.

THE NON-CODING GENOME AND ITS ROLE IN GENE REGULATION

According to the central dogma of molecular biology, there are three main processes taking place in the cell: replication of the genetic information, transcription of DNA into RNA, and translation of the RNA molecule into the final functional product, the protein (Crick, 1958). As one of the main surprises from the Human Genome Project, it is now well-established that more than 98% of the human genome does not encode proteins (Consortium, 2012). These non-protein-coding regions were initially considered as junk DNA, which was assumed to be redundant and under no selective pressure, thus allowing for the accumulation of mutations without any harm to the organism (Ohno, 1972; Kuska, 1998). However, several structural elements of non-coding DNA have now been described that regulate gene expression, by determining the 3D genomic organization critical for correct gene regulation. Regulation of gene transcription is particularly crucial during embryonic development, when a single cell needs to differentiate into distinct cell types and to establish diverse gene expression programs in order to acquire a broad range of phenotypes, while maintaining the same genotype. This is achieved by a tight spatio-temporal regulation of gene expression, that allows the transcription of the right gene, at the right level, in the right cell type, and is executed by the interplay between enhancers and gene promoters confined to the "playfield" established by the 3D organization of the genome. It is important to keep in mind in the following paragraphs, that gene regulation, unlike coding DNA, needs to be seen from a non-linear, 3D perspective where regulatory elements need to interact with target genes on long distances.

Genomic organization comprises efficient DNA packaging in the limited space of the nucleus while allowing for DNA replication and gene expression. First, nucleosomes are formed, in which 147 bp of DNA are wrapped around eight histone proteins linked to each other by DNA stretches of various lengths. This beads-on-a-string organization forms the basis of a 10-nm chromatin fiber that is typical of open chromatin, also known as euchromatin. This differs from tightly packaged heterochromatin, where multiple histones wrap into a 30-nm fiber consisting of nucleosome arrays in their most compact form. As a result of this, chromatin is organized into active and inactive compartments that are either open or condensed and which vary in size between 1 to 10 megabases (Mb). Inactive compartments are often found in association with the nuclear lamina, whereas active compartments are more likely to be found in other regions of the nucleus (van Steensel and Belmont, 2017). Regulatory elements such as enhancers and promoters and actively transcribed genes are located in open-chromatin regions, so that they are accessible for the transcriptional machinery. Various post-translational epigenetic modifications of histones put in place by chromatin modifying enzymes can

alter the accessibility of chromatin and can thereby influence how chromatin is packaged and whether it is more or less likely to be active. For example, histone acetylation results in increased chromatin accessibility and makes chromatin more available for the binding of regulatory proteins, such as transcription factors (TFs). Many studies focused on a wide variety of histone modifications (see Bannister and Kouzarides, 2011; Hyun et al., 2017, for *Review*), and have led to a draft of a histone code, where various histone modifications are indicative of the functional role that the chromatin has at those places that are modified. For example, putative enhancers are enriched in chromatin regions surrounded by histone 3 lysine 4 monomethylation (H3K4me1) and lysine 27 acetylation (H3K27ac), while promoters are marked by histone 3 lysine 4 trimethylation (H3K4me3). Insulators are responsible for organizing chromatin at a sub-compartment level. They are often bound by the TF CTCF (also known as 11-zinc finger protein or CCCTC-binding factor) (Bell et al., 1999) and establish the boundaries of so-called topologically associating domains (TADs). TADs are usually <1 Mb in size and delineate those regions of our chromosomes in which sequences interact preferentially with each-other rather than with elements in other regions of the genome. The prevailing model is that these TADs are formed by the dimerization of two CTCF molecules binding the boundaries of a TAD and are stabilized by the interaction with the ring-shaped cohesin complex through a process called loop extrusion (Dowen et al., 2014; Ong and Corces, 2014; Rowley and Corces, 2018). Inside TADs, smaller DNA loops are formed to allow enhancer–promoter interactions and thus regulation of transcription (Dowen et al., 2014; Hnisz et al., 2016). These enhancer-promoter loops, similarly to the CTCF-mediated loops, are thought to be established by the binding and dimerization of the TF YY1 and its interaction with the cohesin complex (**Figure 1**) (Beagan et al., 2017; Weintraub et al., 2017).

Promoters are located around the transcriptional start site (TSS) of genes and are essential to initiate transcription. Enhancers are positive regulators of transcription (Banerji et al., 1981), whose location relative to the TSS of the gene they control varies from adjacent to the promoter, to many kilobases (kb) upstream or downstream (e.g., *in cis*), and can even be located in introns, also of other genes. Moreover, besides acting in a position-independent manner, enhancers can regulate transcription irrespective of their orientation. A classic example of a long-range regulatory element is the limb *SHH* enhancer, which is located ∼1 Mb away from its target gene (Lettice et al., 2003). Finally, one enhancer can regulate several genes, and at the same time each gene can be regulated by multiple enhancers. This creates a redundancy in the system that results in phenotypic robustness, and probably gives advantages during evolution (Osterwalder et al., 2018). Therefore, the positions, identities, and arrangements of *cis*-regulatory sequences ultimately determine the time and place that each gene is transcribed. On a mechanistic level, enhancers directly influence the recruitment of the transcriptional machinery to the TSS of genes (Stadhouders et al., 2012; Coulon et al., 2013). Crucial for this long-range control of gene expression by enhancers is the formation of enhancer-promoter loops which preferentially occur within the neighborhood of a TAD, by DNA bending. The general TFs
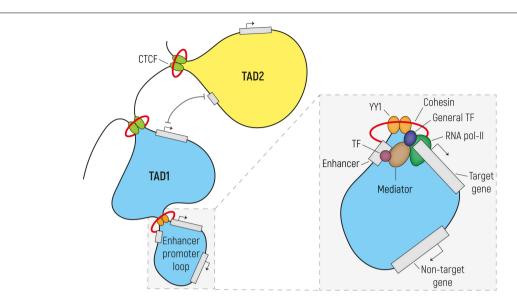
FIGURE 1 | Regulatory enhancer-promoter interactions are restricted within the topological associating domain (TAD) region. The genome (here represented as a black line) is tightly packaged and organized in TADs established by the binding of CTCF to insulator elements, followed by dimerization and interaction with the cohesin complex. In order to establish the enhancer-promoter loops required for transcriptional regulation, enhancers and their target gene should reside in the same TAD. These regulatory loops are formed by the dimerization of YY1 and its interaction with cohesin. In the enlargement is a simplified scheme of transcription initiation (the size does not reflect the actual dimension of each component). Transcription factors (TFs) bind on the enhancer element while the pre-initiation complex formed by the RNA Pol II and the general TFs assembles at the promoter region. Mediator establishes the connection between enhancer and promoter via interactions with TF and components of the transcription preinitiation complex, without binding to DNA. Mediator regulates the phosphorylation of the RNA Pol II in order to release it from the promoter and start transcription.

and the RNA polymerase II bind to the promoter sequence, whereas the distal *cis*-regulatory sequences are bound by TFs, which orchestrate the rate of transcription initiation. Enhancers include TF binding sites (TFBSs) that typically consist of DNA motifs found at multiple sites in the genome, but that are not necessarily all equally likely to be bound by the recognizing TF (Lambert et al., 2018). To provide higher than background activity, homotypic or heterotypic dimerization of transcription regulators increases their DNA binding affinity and specificity (Funnell and Crossley, 2012). TF binding itself can also be influenced by DNA methylation, which is established by DNA methyltransferases (DNMTs) (Yin et al., 2017). Moreover, if TF binding prevents the DNA from rewrapping around the nucleosome, it increases the likelihood that a second transcription regulator binds the DNA, increasing the cooperative effect to the extent of displacing the histone core of the nucleosome (Iwafuchi-Doi et al., 2016; Iwafuchi-Doi, 2019). Multiple TFs have been found to bind in a cooperative manner in TF binding site "hotspots" (Siersbaek et al., 2011), later called stretch enhancers (Parker et al., 2013) or super-enhancers (SEs) (Whyte et al., 2013). The latter are described as long regions with an increased density of enhancer elements characterized by a strong enrichment of H3K27ac, and of TFs and Mediator binding (Hnisz et al., 2013; Whyte et al., 2013). On the one hand, a number of studies suggests that SEs represent a novel class of NCREs that maintain, define, and control mammalian cell identity and whose transcriptional regulatory output is larger than that of the individual enhancer constituents (Young, 2011; Whyte et al., 2012, 2013; Loven et al., 2013). On the other hand,

an increasing number of studies have challenged this view and consider super enhancers as a collection of normal enhancers that together do not have a larger activity than the sum of the individual parts (Hay et al., 2016; Shin et al., 2016). Therefore, the debate on whether SE are a new class of NCREs or whether they simply reflect a clustering of normal NCREs within close proximity remains to be settled.

What is clear from the above, is that our knowledge on complex gene regulatory mechanisms has increased dramatically over the last decade and has provided insights into many sophisticated processes that need to occur correctly for development to proceed normally. Aberrations in many of the steps described above can result in genetic disorders. For example, in recent years a large number of disorders have been described that are caused by mutations in chromatin modifying enzymes or proteins involved in 3D chromatin regulation (Gregor et al., 2013; Ball et al., 2014; Mirabella et al., 2016; Holsten et al., 2018). Given the complexity of gene regulation and the many contributing factors acting at different stages of this process, it seems likely that many more will be discovered in the near future.

GENOME-WIDE IDENTIFICATION OF PUTATIVE ENHANCERS

As introduced in the previous paragraph, transcriptional enhancers were first described as DNA sequences that are able to enhance gene expression on an episomal plasmid
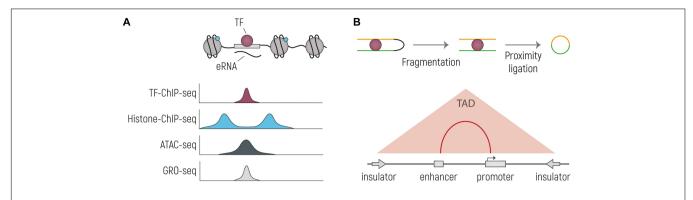
FIGURE 2 | Overview of the main techniques currently used to identify putative enhancer sequences and their interacting genes. **(A)** Schematic drawing of an TF-bound enhancer, located in nucleosome depleted DNA from which eRNA is transcribed. Below are representative genome browser tracks shown, illustrating expected profiles for the same genetic region. Histone-ChIP-seq is illustrative for marks such as H3K27ac and H3K4me1. **(B)** Cartoon representing the main steps of the workflow of Chromosome conformation capture technologies: nuclei are cross-linked, chromatin is then digested and re-ligated by proximity ligation. The two stretches of DNA that are normally located far away from each other (yellow and green), are now ligated together and can be tested by PCR or sequencing. In the bottom part is indicated the output of the experiment, with which TADs and enhancer-promoter interactions can be identified.

(e.g., a non-integrating, extra chromosomal circular DNA), irrespective of their location and orientation relative to the TSS (Banerji et al., 1981; Moreau et al., 1981); thus, enhancer identification was first limited to low-throughput reporter assays, where small fragments of DNA were tested for regulatory activity influencing reporter gene expression. The most widely applied experimental techniques for genome-wide identification of putative enhancers at the endogenous genomic locus today do not rely directly on this functional property, but rather on features that distinguish enhancers from non-regulatory regions at the chromatin level. Indeed, enhancers are bound by TFs and transcription coactivators and are located in open chromatin regions that are depleted from nucleosomes. The surrounding nucleosomes have specific histone tail modifications, such as the previously mentioned H3K4me1 and H3K27ac. Moreover, some enhancers are bi-directionally transcribed in so-called enhancer RNAs (eRNAs). However, even though these features correlate with enhancers, other genomic regions share the same chromatin characteristics, and more functional tests are required to prove that putative enhancers are indeed having a direct functional role in gene regulation (Catarino and Stark, 2018). This led to the development of high-throughput functional screenings, overall known as massively parallel reporter assays (MPRAs) that quantify the enhancer activity of millions of sequences. In the next paragraphs we will discuss the most widely used techniques to identify putative regulatory regions (**Figure 2** and **Table 1**), and in the following section we will focus on high-throughput functional screens.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was first introduced more than 30 years ago to study protein-DNA interactions (Solomon et al., 1988) and it follows three basic steps. First, proteins are covalently cross-linked to their DNA binding-site by treating cells with formaldehyde. Chromatin is then sheared, and protein-DNA complexes are selectively co-immunoprecipitated with an antibody against the protein of interest. Finally, the

cross-linking is reversed, and DNA is isolated and tested to identify the binding sites of the protein of interest. In more recent years, the emergence of next generation sequencing (NGS) technologies, allowed genome-wide mapping of these protein-DNA binding sites (ChIP-seq) (Johnson et al., 2007; Robertson et al., 2007). ChIP-seq is now primarily used to identify putative enhancers across the entire genome by immunoprecipitation of TFs, specific histone-tail post-translational modifications, including H3K4me1 (Heintzman et al., 2007) and H3K27ac (Creyghton et al., 2010), and transcriptional coactivators, such as the histone acetyltransferase p300/CBP (Visel et al., 2009) and Mediator (Whyte et al., 2013). However, neither the binding of a TF nor the presence of histone modifications provide definitive evidence that a sequence acts as a transcriptional enhancer. For example, tissue specific enhancers can have a certain degree of H3K27ac enrichment in tissues where they are not active (Cotney et al., 2012) and not all H3K27ac marked DNA sequences show enhancer activity when functionality tested (Barakat et al., 2018). Several studies have used ChIP-seq for histone modifications to predict enhancers during human brain development (Reilly et al., 2015; Amiri et al., 2018) and in adult brain (Vermunt et al., 2014; Sun et al., 2016; Li et al., 2018; Wang D. et al., 2018), and some have made direct comparisons to brains from other primates, providing important insights in the evolution of humans (Reilly et al., 2015; Vermunt et al., 2016).

Identification of Open Chromatin Regions

As abovementioned, *cis*-regulatory sequences like enhancers are enriched in chromatin regions depleted from nucleosomes (Boyle et al., 2008), as nucleosomes would impede TF binding (Lidor Nili et al., 2010). These accessible DNA regions can be identified in a genome-wide fashion thanks to several techniques such as DNase-seq, FAIRE-seq, and ATAC-seq. DNase-seq takes advantage of the hypersensitivity of open chromatin to nuclease digestion. Briefly, cell nuclei are isolated, and DNA is digested with limiting concentrations of DNase I. Fragments of about

TABLE 1 | Methods for the identification of non-coding regulatory elements (NCRE).

Method	Description	Advantages	Disadvantages
ChIP-seq	Chromatin immunoprecipitation of histone-modifications or TFs coupled with NGS.	Determines genome-wide binding patterns of protein of interest	Not all enhancers are marked by H3K27ac or H3K4me1, or tested TFs. Requires availability of ChIP-grade antibodies. Cannot determine enhancer activity. Cannot identify target gene.
ATAC-seq	Identification of open chromatin regions by the transposon Tn5, that cuts the DNA and inserts sequencing adapters.	Fast. Requires a low number of cells. No need for any *a priori* knowledge.	Other elements are located in open chromatin regions. Cannot determine enhancer activity. Cannot identify target gene.
eRNA detection	Detection of the bidirectionally transcribed eRNA by sequencing the nascent RNA through techniques such as GRO-seq or CAGE.	Identifies enhancer transcription	Not all active enhancers are transcribed.
Chromosome conformation capture	Detection of topological interactions between two loci (3C) or genome wide (4C, 5C, Hi-C).	Identifies enhancer-target gene interactions	Cannot determine enhancer activity.
STARR-seq	Identification of functional enhancers by a massively parallel reporter assay where active enhancers drive their own transcription.	Identifies functional enhancers. Quantitatively measures enhancer activity. High-throughput.	Episomal. Highly complex plasmid libraries requiring substantial number of cells for transfection. Possible false negative results.
CRISPR-Cas9 screenings	Endogenous manipulation of enhancers to force their activation or inactivation.	Identifies functional enhancers. Can be high-throughput. Determines the endogenous effect of enhancer manipulation.	Off-target activity. Possible false negative results.

500 bp are then selected and used for library preparation and sequencing (John et al., 2013). FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) is based on the separation of free and nucleosome-bound DNA. Chromatin is cross-linked with formaldehyde to covalently bind nucleosomes to the DNA, and then sonicated and purified by phenol-chloroform extraction. Nucleosome-bound DNA is sequestered to the interphase, while accessible DNA can be recovered from the aqueous phase and sequenced (Giresi and Lieb, 2009). Finally, the most recently developed method ATAC-seq (Assay for Transposase Accessible Chromatin with high-throughput sequencing) exploits the preference of transposons to land in open chromatin regions. Shortly, the transposon Tn5, loaded with sequencing adapters, is able to simultaneously cut the DNA and insert the adapters in a process known as tagmentation. The open chromatin regions where the transposon preferentially inserts are then amplified with primers binding to the adapters and sequenced. Compared to DNase-seq and FAIRE-seq, ATAC-seq is a simple and fast method that requires less starting material and does not require gel-purification or crosslinking reversal steps and is therefore less prone to loss of material (Buenrostro et al., 2013). However, as mentioned earlier, other regulatory elements such as insulators or promoters are also located in accessible chromatin (Boyle et al., 2008). Therefore, ATAC-seq should be used in combination with other techniques that are more selective for enhancers. Moreover, these methods qualitatively identify putative enhancers and do not allow the quantification of their activity; indeed, also inactive enhancers can be in open-chromatin regions (Arnold et al., 2013; Catarino and Stark, 2018). A major advantage of all techniques assessing chromatin accessibility compared to ChIP-seq is that they screen for putative regulatory regions in an unbiased way, not requiring *a priori* knowledge of enhancer binding factors and not being

restricted to the use of available ChIP-grade antibodies. A recent study has used ATAC-seq and RNA-seq to determine open chromatin regions and gene expression at different gestational weeks, and in different areas of the brain, i.e., the ventricular zone and the neuronal layers, providing a first glimpse on open chromatin dynamics during human fetal brain development (de la Torre-Ubieta et al., 2018).

eRNA

Transcription of enhancer sequences was first reported in the early 90s in the Locus Control Region (LCR) of the β-globin gene cluster (Collis et al., 1990; Tuan et al., 1992; Ashe et al., 1997), where it was found that the expression of the LCR is restricted to the erythroid lineage. Later, transcription of regulatory elements into enhancer RNAs (eRNAs) was validated genome-wide with sequencing, at first, of total neuronal RNA (Kim et al., 2010), followed by sequencing of nascent RNA (GRO-seq, CAGE) in different cell types (Wang et al., 2011; Hah et al., 2013; Kaikkonen et al., 2013; Andersson et al., 2014). Enhancer RNAs are generally bidirectionally transcribed and not polyadenylated (Kim et al., 2010) but reports of unidirectional transcription and polyadenylation of eRNAs exist (Koch et al., 2011). Enhancer transcription was shown to correlate with the presence of other enhancer marks such as histone tail post-translational modifications and p300/CBP and RNApolII binding (Wang et al., 2011; Hah et al., 2013; Kaikkonen et al., 2013), but whether their expression is a cause, or a consequence of gene transcription is still debated (Lam et al., 2014). If eRNA transcription has a direct functional role and is not just noise due to the recruitment of RNApolII, the effect can either be mediated by the transcription process itself or by the transcript produced upon transcription, which might have direct *cis*-regulatory activity similar to other non-coding RNAs such as those involved in X chromosome

inactivation (Barakat and Gribnau, 2010; Natoli and Andrau, 2012). However, even if eRNA presence correlates with enhancer activity at some loci, it seems that it is neither required nor sufficient in all instances (Catarino and Stark, 2018). For example, a recent study assessing eRNAs in brain only found that around 600 intergenic and intronic enhancers are transcribed in eRNAs, and this number even further decreased when considering only those eRNAs replicated in an independent data set or overlapping with enhancer associated histone modifications (Yao et al., 2015). The FANTOM project has found a similar small number of eRNAs in brain, although the majority of those are not overlapping with those from Yao and colleagues (Andersson et al., 2014). The number of predicted brain related enhancers based on other assays by far outnumbers this rather small set of transcribed enhancers, indicating that methods that just take eRNA transcription into account may oversimplify the identification of putative enhancers and may not catch the complete regulatory landscape.

Identification of Long-Distance Chromatin Interactions

All methods described until now identify putative enhancers but understanding which genes they regulate remains a challenge. Indeed, despite often regulating nearby genes, enhancers can also be found at long distances from the TSS of their target gene. Moreover, as abovementioned, it is becoming more and more clear that chromatin organization plays an important role in transcription and, as abovementioned, enhancers and promoters need to be brought in close proximity in order for transcription to take place. In the past ~20 years several techniques have been developed to address this question (reviewed in de Wit and de Laat, 2012; Davies et al., 2017). The pioneering method, on which all the later developments are based, is known as chromosome conformation capture (3C) and relies on the formaldehyde cross-linking of chromatin within nuclei, followed by restriction digestion of chromatin and re-ligation by proximity ligation. The obtained fragments represent the junction of two chromatin regions that are normally located far away from each other on the linear genome, but are in close proximity in 3D space, and these junction products can be quantified by PCR (Dekker et al., 2002). 3C was developed to study whether two known regions are interacting with each other and is thus described as a "one vs. one" method (de Wit and de Laat, 2012). Further advances of 3C-based techniques allowed the identification of increasing numbers of contacts; for example, 4C, "one vs. all," allows the identification of all the regions interacting with a specific site of interest (Simonis et al., 2006; Zhao et al., 2006), while 5C, "many vs. many," investigates all contacts that are happening in a specific locus (Dostie et al., 2006). Finally, high-throughput contact identification became possible with Hi-C (Lieberman-Aiden et al., 2009). Hi-C allows the identification of genome-wide interactions thanks to the introduction of biotin-labeled nucleotides at the sites of restriction-digestion. The ends are then ligated, the chromatin is sheared, and the junctions are enriched by streptavidin pull-down and sequenced. By the application of an

algorithm on Hi-C data, TADs can be defined. To investigate all the genome-wide interactions involving a specific protein of interest, HiChIP was developed, by introducing a chromatin immunoprecipitation step (Mumbach et al., 2016). This method has the advantage that it requires less input material and less sequencing reads.

Despite their capacity to identify enhancer-promoter interactions and thereby pieces of chromatin with putative regulatory roles, chromatin conformation techniques have the disadvantage of not directly measuring functional regulatory activity. Moreover, in most cases, interactions are determined on a population level on a high number of cells, which might only provide a snapshot of dynamic regulatory interactions. Finally, the spatial resolution at which interactions can be determined is heavily influenced by the sequencing depth of Hi-C experiments. Hence, there remains a need for more functional tests to validate the regulatory activity of the identified interactions. A recent study has generated Hi-C maps from gestational weeks 15, 16, and 17 of human brain development, a critical time period for cortex development (Won et al., 2016), permitting the large-scale annotation of previously uncharacterized regulatory interactions relevant to the evolution of human cognition and disease. For example, the results of this study have linked several non-coding variants identified in GWAS to genes and pathways involved in schizophrenia, highlighting novel mechanisms underlying neuropsychiatric disorders.

HIGH-THROUGHPUT FUNCTIONAL IDENTIFICATION OF ENHANCERS

As previously highlighted, most of the commonly used techniques to identify regulatory elements are merely predictive, and do not directly measure enhancer activity. Although there is no doubt that techniques such as ChIP-seq, open chromatin mapping and expression analysis have been of tremendous use to globally characterize the gene regulatory landscape of the non-coding genome, it is still clear that there is a need for improved techniques. In many instances, the identified putative enhancer sequences fail to perform as enhancers in functional validation experiments, giving rise to false positive enhancer predictions (Kwasnieski et al., 2014; Halfon, 2019, for an excellent review). Moreover, the resolution of commonly used techniques usually allows the identification of regions in the range of 500–1000 bp as potentially including an enhancer. But this makes it difficult to pinpoint those nucleotides that are of real functional relevance within a given predicted enhancer sequence, and this complicates, for example, the assignment of functional roles of nucleotide variants found in the human population. Finally, many of the currently used techniques take into consideration previously identified knowledge on associations between epigenetic marks and putative enhancers. This potentially excludes other regions of the genome to be functionally assessed as they lack these associations but might nevertheless be functionally relevant (Pradeepa et al., 2016). Direct high-throughput functional tests of enhancer activity, such as massively-parallel reporter assays (MRPAs) and CRISPR-Cas9 based screens have the potential to
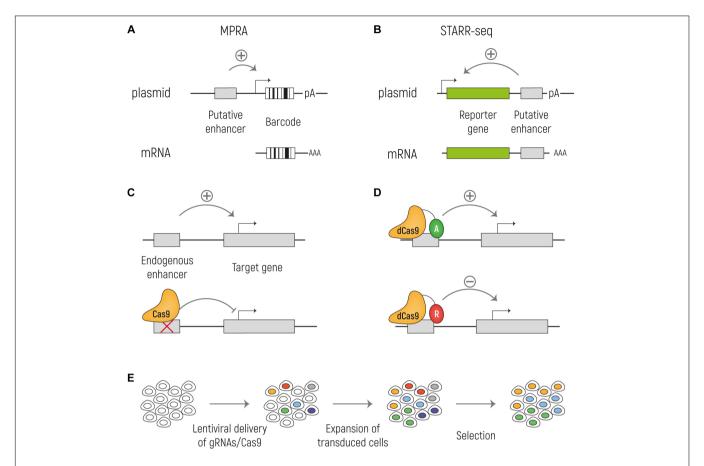
FIGURE 3 | Methods for functional identification of enhancers. **(A)** Massively parallel reporter assays (MPRA) to test enhancer activity in an episomal setup. The putative enhancer sequence is cloned upstream a minimal promoter that drives the expression of a reporter gene and a unique barcode. **(B)** With STARR-seq the putative enhancer sequence is cloned downstream the reporter gene and upstream of the polyA signal. When the enhancer sequence is active, it can drive the expression of the reporter (green) and of itself. In both MPRA and STARR-seq the mRNA is sequenced to identify the active enhancers. **(C)** Cas9 can be used to knock out an enhancer at the endogenous genomic locus to assess its effect on the target gene transcription. **(D)** A catalytically inactive Cas9 (dCas9) can be fused with activators (A: VP64; TET1; p300) or repressors (R: KRAB; SID4X; DNMT3A; KDM1A). **(E)** Cas9 screens can be combined with high-throughput screenings by targeting Cas9 expressing cells with a lentiviral library of gRNA at a low MOI. By doing so, each cell will express a single gRNA and by different selections, such as drug resistance or reporter gene expression, it is possible to investigate the effect of the ablation of a large number of putative enhancers on gene expression in parallel.

address these shortcomings (**Figure 3**), as we will explain in the next section.

Most traditional functional tests for enhancer activity are based on reporter assays, in which a putative enhancer sequence is cloned into a vector with a reporter gene driven by a minimal promoter that alone is not sufficient to induce reporter gene expression. The vectors are then transfected into a cell line or organism of interest, and the reporter gene expression is determined (Banerji et al., 1981). MPRAs are high throughput reporter assays where DNA sequences are inserted before the minimal promoter of a vector with a specific barcode sequence downstream of the open reading frame, which allows the simultaneous assessment of 1000s of sequences for enhancer activity in parallel (Melnikov et al., 2012; Patwardhan et al., 2012; Arnold et al., 2013; Kheradpour et al., 2013; Whyte et al., 2013; Dickel et al., 2014; Murtha et al., 2014; Ernst et al., 2016). After cell transfection, RNA can be purified and sequenced. If the sequence cloned into the vector is a functional enhancer it drives the expression of the corresponding barcode. An adapted approach is Self-Transcribing Active Regulatory Region (STARR) sequencing (Arnold et al., 2013). STARR-seq takes advantage of the fact that enhancers act in a position-independent fashion. Indeed, differently from other MPRAs, STARR-seq does not rely on barcodes, but the candidate sequences are cloned downstream of the TSS and, when active, drive their own transcription. With this assay, millions of sequences can be tested in a single experiment. In both cases, the activity of the enhancer can be measured by the relative abundance of the barcode/sequence transcript from RNA-seq, in comparison to sequencing of the input plasmids. Similar episomal high-throughput approaches have recently been developed to also measure promoter responsiveness to enhancers (Arnold et al., 2017) and autonomous promoter activity (van Arensbergen et al., 2017).

The major advantage of these tests is that they are unbiased, since they are not based on any *a priori* hypothesis about TF binding or histone modifications. Nevertheless, the size of the

human genome requires the construction and transfection of large plasmid libraries, and thus substantial numbers of cells and deeper sequencing and might therefore lead to a lower resolution. To overcome this limitation, it is possible to focus STARR-seq only on putative enhancers, testing only the sequences identified with ChIP (Barakat et al., 2018), ATAC (Wang X. et al., 2018), or other techniques (Vanhille et al., 2015; Shen et al., 2016). In our own application, we have combined ChIP with STARR-seq to generate genome-wide enhancer activity maps in various types of human embryonic stem cells (Barakat et al., 2018).

Despite being incredibly useful to test millions of sequences for enhancer activity in a high-throughput manner, reporter gene assays may have several limitations. First, enhancer activity is tested most often on an episomal background, which might not completely reflect endogenous gene regulation in its native genomic context (Inoue et al., 2017). Interestingly, recent studies suggest that the effects of this might be less strong than initially suggested, as there is a high correlation between episomal enhancer activity and endogenous gene regulation when assessed by CRISPR-based deletions (Barakat et al., 2018), or when a set of enhancers is assessed on both plasmids and integrated at multiple genomic locations (Maricque et al., 2018). Second, MPRAs may potentially give false negative results. Indeed, if a sequence is found inactive in a reporter assay, this does not exclude that it is active as an enhancer in a different cell type, in a different moment in time or has another, but still biologically relevant, role independent on enhancer activity (Shlyueva et al., 2014).

One way to overcome these possible limitations of transgenic reporter assays, is to use the recently developed CRISPR-Cas9 system to manipulate NCREs at the endogenous chromatin context. Cas9 is an RNA-guided DNA endonuclease that is able to induce double-strand breaks that, in the absence of a donor template for homology directed repair, are repaired by the error-prone non-homologous end-joining (NHEJ). The enhancer sequence can thus be deleted, by targeting Cas9 with guide-RNAs (gRNAs) flanking the enhancer sequence or be mutated by the introduction of indels via NHEJ, allowing to test the effect of the enhancer ablation on gene expression in the endogenous chromatin environment. Whereas this approach can be used to study a selected enhancer of interest, as we have done studying enhancers involved in pluripotency of human embryonic stem cells (Barakat et al., 2018), it can also be used in high-throughput screenings with large libraries of gRNAs that are introduced in cells expressing Cas9. Lentiviral transduction of gRNAs at a low multiplicity of infection can result in a single gRNA integration per cell, and in combination with various means of positive or negative selection, such as drug selection or assessment of reporter gene expression, this can be used to investigate in parallel and on a large scale the effect of multiple putative enhancer ablations on gene expression. To this end, large populations of cells are transduced, and the quantitative presence of gRNAs is determined by next generation sequencing of isolated DNA prior and after a selection. If a sequence has an important role in gene regulation, the ablation of that sequence is expected to result in disadvantage for the cells, and therefore gRNAs targeting relevant functional NCREs

will be depleted over time. By comparing sequencing reads after and prior to the selection, it is possible to determine which gRNAs are lost over time, and as the targets of the gRNAs are known, the relevant NCRE can be identified. In one of the first applications, DNA regions around the *TP53* and *ESR1* gene loci were investigated, and it was shown that this approach was feasible to identify functional enhancers and, furthermore, using a dense CRISPR-Cas9 gRNA tilling screen, functional domain within these enhancer sequences were precisely mapped (Korkmaz et al., 2016). Using a similar approach, more than 18.000 gRNAs were used to test around 700 kb of sequence flanking genes involved in BRAF inhibitor resistance in melanoma, finding non-coding regions involved in gene regulation and chemotherapeutic resistance (Sanjana et al., 2016). Other studies investigated putative enhancers involved in oncogene induced senescence (Han et al., 2018), regulation of the *HPRT* gene involved in Lesch–Nyhan syndrome (Gasperini et al., 2017) and regulation of the *POU5F1* gene in embryonic stem cells (Diao et al., 2016, 2017), amongst others (Canver et al., 2015, 2017; Rajagopal et al., 2016; Sen et al., 2016). Besides genome-engineering, CRISPR-Cas9 can also be applied to edit the epigenome, and also this can be coupled to high-throughput screening. Indeed, by fusing a catalytically dead Cas9 (dCas9), that lacks endonuclease activity, to various functional domains it is possible to alter the status of a NCRE forcing its activation or inactivation, referred to as CRISPRa and CRISPRi, respectively. Functional additions to dCas9 leading to NCRE activation include transcription activating domains such as multiple repeats of the herpes simplex VP16 activation domain (VP64) (Maeder et al., 2013; Mali et al., 2013), the nuclear factor-κB (NF-κB) trans-activating subunit activation domain (p65) and human heat-shock factor 1 (HSF1) (Konermann et al., 2015), the 10–11 translocation methylcytosine dioxygenase 1 (TET1) (Liu X. S. et al., 2016), and the p300 acetyltransferase (Hilton et al., 2015). Oppositely, transcription repressive domains that can be used to silence NCREs include Krüppel-associated box (KRAB) domain (Gilbert et al., 2013; Thakore et al., 2015), four concatenated mSin3 domains (SID4X) (Konermann et al., 2013), cytosine-5-methyltransferase 3A (DNMT3A) (Vojta et al., 2016), Histone deacetylase 3 (HDAC3) (Kwon et al., 2017), and the lysine-specific histone demethylase 1A (KDM1A), called dCas9-LSD1 (Kearns et al., 2015). Several of these dCas9 fusion have been used to activate or repress NCREs, and a number of studies have used them in high-throughput screening approaches, most of which focused on NCRE repression (Fulco et al., 2016; Carleton et al., 2017; Xie et al., 2017; Gasperini et al., 2019) but some included also NCRE activation (Klann et al., 2017; Simeonov et al., 2017). It seems only a matter of time till more similar studies editing NCREs in various cell types using the full CRISPR-Cas9 toolbox will be published. Obviously, as all experimental approaches, also CRISPR-Cas9 has its pitfalls and is still far from perfect. For example, reduced on-target activity and off-target effects of gRNAs can introduce experimental noise, and it remains essential that screening results are validated independently. Also, it remains to be seen whether subtle enhancer effects on gene expression, that might still be of biological relevance, can be detected using CRISPR-based screens.

COMPUTATIONAL APPROACHES

As it has become clear from the discussion above, currently used enhancer prediction and validation techniques heavily depend on computational data analysis, most often involving the analysis of next generation sequencing data. Besides the direct use of computational analysis for biological data processing, more and more efforts are undertaken to use computational power to predict functional NCREs *in silico*. These methods can be broadly summarized in approaches that rely on (1) comparative genomics and evolutionary conservation, (2) clustering of motifs and epigenome features and machine learning approaches, and (3) techniques that deal with the processing of data obtained in functional genomic screens. Here, we mainly focus on the advantages and disadvantages of some of these methods and highlight several resources that can be used to obtain information on genomic enhancer locations. We refer those readers who are interested in a more detailed discussion on the various options for machine learning and other prediction tools to a number of excellent recent reviews (Wang et al., 2013; Suryamohan and Halfon, 2015; Kleftogiannis et al., 2016; Lim et al., 2018).

Comparative Genomics and Evolution in Enhancer Prediction

Functional sequences are expected to be more conserved compared to DNA stretches that are not expected to have any role, as changing of nucleotide composition is expected to alter function. This characteristic is exploited by comparative genomics approaches that aim to identify enhancers by looking at the most conserved sequences across different species. This approach was one of the first computational tools to identify NCREs (Pennacchio et al., 2006; Li et al., 2007). Nevertheless, different studies showed how some NCREs are strongly conserved, while others are rapidly changing also in closely related-species, rendering the solely use of comparative genomics techniques insufficient. For example, Arnold et al. (2014) showed by STARR-seq of different *Drosophila* species how, in the majority of the cases, enhancer function is conserved across species, and the highly conserved enhancers are thought to play an important role during key processes such as embryonic development, and especially in the developing nervous system (Pennacchio et al., 2006). However, several other studies suggest that a portion of enhancers undergo rapid evolution, and that this might be a crucial driver of human evolution (Degner et al., 2012; Shibata et al., 2012; Villar et al., 2015; Glinsky and Barakat, 2019). A subset of active enhancers in human embryonic stem cells is even enriched in human specific transposable elements, and those functional regions would be missed if one were to use only conservation as a key feature for enhancer selection (Barakat et al., 2018). Therefore, although sometimes useful, evolutionary conservation alone for the discovery of NCRE is not recommended as a sole criterion, as it would miss all the newly evolved enhancers. Another extreme example of this are so-called ultraconserved elements, stretches of DNA sequences that are more than 200 bp long and that are 100% identical in multiple species, such as human, rat and mouse

(Bejerano et al., 2004). Whereas some of these sequences were shown to play a role as enhancers (Visel et al., 2008; Dickel et al., 2018), others can be removed from the genome without an obvious phenotype (Ahituv et al., 2007), and it is speculated that some of these sequences might contribute to genome stability (McCole et al., 2018). Enhancers can also be identified by the presence of specific TFBS, as TF binding is a key characteristic of these regulatory sequences. Indeed, combining conservation with TFBS site discovery can further increase the predictive power of comparative genomic approaches. However, even this does not guarantee enhancer identification, as during evolution novel TFBS can appear which execute similar functions as the ones in the ancestry sequence (Ludwig et al., 2000).

Enhancer Prediction Algorithms

Several types of enhancer predicting algorithms have been developed for integrating multiple types of data, such as TF motifs, ChIP-seq, DNase-seq, ATAC-seq, and P300 binding data sets for enhancer prediction by using clustering and machine learning approaches. These algorithms include supervised tools that rely on high-confidence positive and negative training sets (e.g., known- and non-enhancers) to build models that can maximize the differentiation between enhancer and non-enhancer sets, and non-supervised tools that are used to identify hidden and unknown patterns directly from data. Examples of supervised algorithms include CSI-ANN (Firpi et al., 2010), ChromaGenSVM (Fernandez and Miranda-Saavedra, 2012), RFECS (Rajagopal et al., 2013), EnhancerFinder (Erwin et al., 2014), DEEP (Kleftogiannis et al., 2015), DELTA (Lu Y. et al., 2015), PEDLA (Liu F. et al., 2016), REPTILE (He et al., 2017), eHMM (Zehnder et al., 2019), and DBN (Bu et al., 2017). Other unsupervised algorithms such as Segway (Hoffman et al., 2012) and ChromHMM (Ernst and Kellis, 2012) integrate multiple types of epigenome data to define chromatin segmentations that can be used to assign functional roles for various parts of chromatin. One of the main problems of all these prediction programs is that we still lack a detailed understanding of the underlying regulatory code in the non-coding genome. Despite all the advances made over the last decade, we are yet to pinpoint a feature that can identify enhancers (and their activity) in all cell types. As most programs rely on previously generated training sets or on static features such as DNA sequence motifs which by their own do not necessarily predict enhancers in each instance, it is more than logical that despite the large amount of efforts that are undertaken, enhancer prediction programs are far from perfect. For example, although chromatin segmentation is very intuitive and access to these segments can be easily obtained from the UCSC genome browser, it is rather worrying that a recent study testing more than 2000 sequences classified as enhancers using these methods did not detect regulatory activity in 74% of the sequences tested (Kwasnieski et al., 2014). Also the overlap between individual predictions from various programs is rather poor (Kleftogiannis et al., 2016). Quite intuitively, programs that take into account multiple features for enhancer prediction tend to perform better (Erwin et al., 2014; Dogan et al., 2015; He et al., 2017). Therefore, it is tempting to speculate that future large-scale meta-analyses of all currently available enhancer data might

enable the further fine tuning of enhancer prediction tools in the near future. Another area where further progress needs to be achieved is the prediction of enhancer-promoter interactions, that can be used to assign NCREs to their target genes. Several tools for this are currently available, such as ELMER 2 and InTAD. ELMER 2 computes the correlation between the enhancer and target genes by combining both DNA methylation and gene expression data derived from the same data set (Silva et al., 2018), but is limited by the fact that correlations are restricted to the closest neighboring gene, which does not necessarily present the real biological relevant target gene (Kvon et al., 2014). InTAD is a tool to detect genes located upstream and downstream of the enhancer in the same TAD boundary and it can support different types of data as input. The TAD information comes from available Hi-C datasets (Okonechnikov et al., 2019), but the currently available ones have a low resolutions and, up to date, have include a limited number of cell types. Therefore, to fully identify enhancer-gene interactions, it will be important to come up with novel experimental procedures that will enable us to directly test the biological relevance of enhancer-promoter predictions. Finally, a number of programs are developed that aim to predict the functional relevance and possible pathogenicity of variants in NCREs. These include, amongst others, RegulomeDB (Boyle et al., 2012), HaploReg (Ward and Kellis, 2012), CADD (Kircher et al., 2014), GWAVA (Ritchie et al., 2014), GenoCanyon (Lu Q. et al., 2015), Genomiser (Smedley et al., 2016), and INFERNO (Amlie-Wolf et al., 2018). In addition, a number of databases have been generated, including HEDD (Wang Z. et al., 2018), DiseaseEnhancer (Zhang et al., 2018), and EnDisease (Zeng et al., 2019) which have collected NCREs that are related to diseases based on the current literature. It will be crucial to further expand and curate these collections of disease relevant enhancers in the future, and combine them with improved ways of variant interpretation, to fully exploit the relevance of the non-coding genome in disease.

Available Enhancer Databases

As the available information on the non-coding genome is increasing rapidly over the last decade, more and more resources are available online that can help to localize NCREs and interpreting their functional roles. In the next part, we summarize a selection of databases and resources that are currently available (**Table 2**) and can be used to find NCREs of relevance for brain development.

One of the first resources of experimentally tested NCREs was the VISTA enhancer database (Visel et al., 2007). Based on comparative genomics, a large selection of putative NCREs from

TABLE 2 | Enhancer databases.

Database	Source
VISTA	http://genome.lbl.gov/vista/index.shtml
EnhancerAtlas 2.0	http://www.enhanceratlas.org/
FANTOM5	http://slidebase.binf.ku.dk/human_enhancers/presets
PsychENCODE	http://development.psychencode.org/#
dbSUPER	http://asntech.org/dbsuper/

mouse and human was selected and tested in transgenic mouse embryo assay, to determine their *in vivo* enhancer activity, as determined by LacZ expression. The database provides detailed information on the genomic localization of the tested sequences, likely associated genes, and images of transgenic mouse embryos identifying the localization of enhancer driven LacZ expression. Based on the 2/25/2019 update, this database contains 542 tested enhancers active in human forebrain, hindbrain, and midbrain. In addition, the VISTA tool portal can be used as a comparative tool and users can submit their own sequences to conduct comparison against multiple species (Visel et al., 2007), thereby possibly identifying conserved functional NCREs.

EnhancerAtlas 2.0 is a database that has collected putative NCREs based on publicly available data obtained from ChIP-seq for different histone modification, TFs, EP300, and POLII, CAGE and eRNA expression, interaction studies by ChIA-PET (a method that combines 3C with chromatin immunoprecipitation) and chromatin accessibility as determined by FAIRE and DNase-seq. Each putative enhancer is supported by at least three independent high-throughput data sets although the database does not contain any direct functional validations. It contains more than 4,506,217 putative enhancers from 8,573 datasets of 179 human tissue/cells, which through an interactive website can be easily accessed. The 49,925 human fetal brain and 17,103 cerebellum enhancers were predicted using DNa-seq, CAGE, and H3K4me1, and H3K27ac deposition (Gao et al., 2016).

The FANTOM5 database, is the latest version of the FANTOM project, that aims to generate an atlas of mammalian regulatory elements, transcriptomes, and long-noncoding RNAs. NCREs are predicted from sequencing data from cap analysis of gene expression (CAGE) along with RNA-Seq data from multiple tissue and cell types from different developmental time points (Andersson et al., 2014). In total, the database contains more than 43,000 putative enhancers, of which 639 are expressed in brain and 376 were found in neuronal stem cells.

Recently, the PsychENCODE consortium has released data from a large multi-center effort trying to map NCREs during brain development (Amiri et al., 2018; Li et al., 2018; Wang D. et al., 2018). Using analysis of transcriptome, methylation status, histone modifications and even single cell/nucleus-level (transcriptome) genomic data, NCREs were discovered across multiple brain regions over the entire span of human neurodevelopment and from adult brains, and an integrative data analysis was performed. These data, generated from age- and often donor-matched samples, represent the most comprehensive multi-platform functional genomic analysis of the developing human brain performed so far. The analysis resulted in 79,056 enhancers identified from adult brains enriched for H3K27ac and depleted for H3K4me3 (Wang D. et al., 2018). In addition, 96,375 enhancers were shown to interact with protein coding genes during fetal brain development and during *in vitro* differentiation of brain organoids (Amiri et al., 2018). Of the latter, 46,735 enhancers were active only in fetal cortex.

Several super-enhancer databases have been generated such as dbSUPER (Khan and Zhang, 2016), SEA (Wei et al., 2016), and SEdb (Jiang et al., 2019) providing annotation, genomic
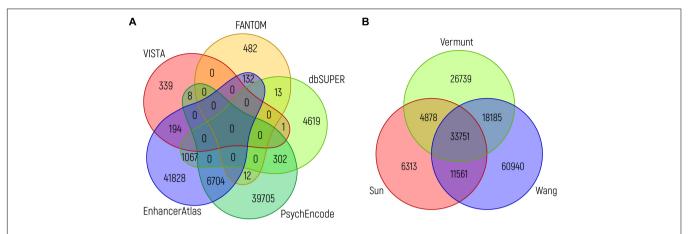
FIGURE 4 | Overlap between brain enhancer databases. **(A)** Venn diagram showing the overlap between brain enhancers (in genome build hg19) from different databases: VISTA (n = 542) (Visel et al., 2007), dbSUPER (n = 6002) (Khan and Zhang, 2016), FANTOM5 (n = 639) (Andersson et al., 2014), PsychENCODE (n = 46731; the 46735 enhancers mentioned in the text are in genome build hg38 and the difference of four loci is due to liftover to hg19) (Amiri et al., 2018) and EnhancerAtlas 2.0 (n = 49925) (Gao et al., 2016). **(B)** Venn diagram showing the overlap between ChIP-seq peaks for H3K27ac from adult brain, as identified in three studies: Sun (n = 56503) (Sun et al., 2016), Vermunt (n = 83553) (Vermunt et al., 2014), and Wang (n = 124437) (Wang D. et al., 2018). The intersection between different sources (in the same order as above) was performed using bedops and bedtools. In both graphs, the minimum overlap of a single nucleotide is required.

coordinates and length of super-enhancers, and their possible associated genes. Among those, dbSUPER is one of the most popular databases with 82,234 super-enhancers from multiple human and mouse cell types. In this database, there are 6,002 and 1,114 super-enhancers detected by H3K27ac enrichment from seven human and three mouse brain tissues and cell types, respectively (Khan and Zhang, 2016).

Finally, GeneHancer is a database of human enhancers and their inferred target genes (Fishilevich et al., 2017). Integrating enhancer predictions from ENCODE, Ensembl, FANTOM and VISTA yielded more than 280,000 candidate regions that where assigned to their target genes based on co-expression correlation, expression of quantitative trait loci and capture Hi-C.

Although all of these databases can easily be accessed and are user friendly, it is important to realize when using them that it is still difficult to judge which of the sources provides the user with those sequences that are indeed most likely to be of functional biological relevance. To illustrate this, it is interesting to compare the overlap between predicted enhancers from the various resources. When we compare putative brain enhancers from VISTA, EnhancerAtlas 2.0, FANTOM5, PsychENCODE and dbSUPER, the overlap between the various enhancer predictions is rather limited, even when considering a single nucleotide as the required overlap (**Figure 4A**). The same holds true when assessing the overlap between ChIP-seq peaks for H3K27ac from key adult brain related data sets (**Figure 4B**). Intuitively, one would expect that those NCREs that are found in multiple data sets are more likely to have a true biological role, although this might be an oversimplification, as the brain is a very heterogeneous organ with many different cell types that might differ in NCRE landscape, and technical limitations might still hinder us from detecting all relevant NCREs in each cell type. Ideally, future studies should generate genome-wide functional activity maps of NCREs for all cell types during brain development. A challenge

that is probably easier to address on paper than in practice in the near future.

ABERRATIONS OF NON-CODING ELEMENTS IN NEURODEVELOPMENTAL DISORDERS AND MALFORMATIONS OF CORTICAL DEVELOPMENT

As previously discussed, an increasing number of studies suggests that a high fraction of causative mutations in neurodevelopmental disorders such as intellectual disability and autism, belong to pathways of transcriptional regulation and chromatin remodeling (Gregor et al., 2013; De Rubeis et al., 2014; Gabriele et al., 2017). Besides mutations in *trans*-acting factors such as TFs or chromatin modifiers, also mutations of NCREs *in cis* have been proven to be causative of disease in an increasing number of cases. A classic example is pre-axial polydactyly caused by alterations of the ZRS, a long-distance enhancer that regulates Sonic hedgehog (*SHH*) expression in the embryonic limb (Lettice et al., 2002; Gurnett et al., 2007). Next to point mutations, also copy number variation (CNVs) such as duplications (Klopocki et al., 2008), and insertions (Laurell et al., 2012) in this region have all been shown to cause polydactyly phenotypes, illustrating the wide range of alterations that can affect enhancer function and thereby result in a phenotype. In this section, we discuss a number of other examples of NCRE alterations, mainly in relation to disorders affecting the brain (**Table 3**).

Holoprosencephaly, a neurodevelopmental disorder characterized by craniofacial malformations, can be caused by coding mutations in the *SHH* gene. However, a point mutation in the *SHH* Brain Enhancer 2 (SBE2) was identified, located 460 kb upstream of the *SHH* gene in a patient with an identical phenotype (Jeong et al., 2008). This mutation was found to

TABLE 3 | Alterations of non-coding regulatory elements in diseases related to the central nervous system.

Disease	Mutation	Affected gene	References
Holoprosencephaly	Point mutation	*SHH*	Jeong et al., 2008
Aniridia	Point mutation	*PAX6*	Bhatia et al., 2013
Polymicrogiria in the Sylvian fissure	Deletion	*GPR56*	Bae et al., 2014
Parkinson's disease	SNP	*SNCA*	Soldner et al., 2016
Schizophrenia	Tandem duplications	*VIPR2*	Vacic et al., 2011
Adult-onset demyelinating leukodystrophy	Deletion of TAD boundary and deletions	*LMNB1*	Giorgio et al., 2015; Nmezi et al., 2019
Intellectual disability	CNV	*ARX*	Ishibashi et al., 2015

be disease-causing, as it disrupts the binding site of the TF SIX3, thereby leading to reduced forebrain SHH expression. In agreement, also mutations in *SIX3* can lead to holoprosencephaly (Wallis et al., 1999). A disease-causing enhancer mutation is also found in the congenital eye malformation aniridia, that is often caused by haploinsufficiency of the TF PAX6, that also plays crucial roles in neural stem cells. A point mutation in the *PAX6* eye-enhancer was found to disrupt PAX6 binding, thereby affecting *PAX6* expression (Bhatia et al., 2013). In another example, a 15-base pair deletion in a regulatory element upstream of an alternative transcript of *GPR56* was found in five individuals from three families (Bae et al., 2014). *GPR56*, when mutant in its coding sequence, leads to widespread cobblestone malformation with cerebellar and white matter abnormalities. In the patients carrying the 15-bp regulatory element deletion, polymicrogyria was bilaterally restricted to the Sylvian fissure, leading to a phenotype of speech delay, intellectual disability and refractory seizures without further motor involvement. The authors could show that the deletion disrupts an RFX binding site, and thereby specifically alters the expression of *GPR56* in the perisylvian and lateral cortex, including the Broca area that is the primary language area.

Besides influencing disorders presenting early in life, also disease emerging later in life, such as neurodegenerative disorders and schizophrenia, are increasingly linked to variants in NCREs. For example, a risk variant in an enhancer regulating α-synuclein expression was recently shown to affect gene expression by altering the binding of the TF EMX2 and NKX6-1 (Soldner et al., 2016). In addition, tandem duplications of the non-coding upstream region of *VIPR2* have been observed in cases of schizophrenia and resulted in upregulated *VIPR2* expression (Vacic et al., 2011). Also, CNVs overlapping with NCREs in other schizophrenia related genes might be implicated in the disease pathogenesis, influencing the disease vulnerability (Piluso et al., 2019).

Multiple CNVs have also been associated with periventricular nodular heterotopia (PNH), a brain malformation in which nodules of neurons are ectopically retained along the lateral ventricles (Cellini et al., 2019). Besides changing gene dosage, CNVs can also change the dosage and position of NCREs, as well as the higher-order chromatin organization of a locus

(Feuk et al., 2006; Spielmann et al., 2018). Similarly, copy-number-neutral structural variants, such as inversions and translocations, can disrupt coding sequences or create fusion transcripts, but these types of variants can also disrupt or create new enhancer landscapes and chromatin domains, resulting in regulatory loss or gain of function. A clinical example of such a structural variant that changes the 3D architecture of the genome is the deletion of a TAD boundary at the *LMNB1* locus, which causes an enhancer to regulate a gene that is normally not regulated by that enhancer (so-called enhancer adoption). In this case, the enhancer adoption leads to adult-onset demyelinating leukodystrophy (ADLD), which is a progressive neurologic disorder affecting the myelination of the central nervous system (Giorgio et al., 2015). More recently, deletions upstream of *LMNB1*, varying in size from 250 to 670 kb, occurring in repetitive elements, have revealed increased LMNB1 expression and an atypical ADLD phenotype (Nmezi et al., 2019). Other rare inherited structural variants in *cis*-regulatory elements might influence the risk for children of developing autism spectrum disorders (ASDs), depending on the parental origin of the structural variant (Brandler et al., 2018). Another study on autism using whole genome sequencing (WGS) on more than 2000 individuals found that probands carry more gene-disruptive CNVs and SNVs resulting in severe missense mutations and mapping to predicted fetal brain promoters and embryonic stem cell enhancers (Turner et al., 2017). In addition, CNVs covering the regulatory elements of the *ARX* gene might cause an intellectual disability phenotype (Ishibashi et al., 2015), and rare non-coding CNVs near previously known epilepsy genes were enriched in a cohort of 198 individuals affected with epilepsy compared to controls (Monlong et al., 2018). Similar findings are reported for multiple system atrophy (Hama et al., 2017) and non-coding variants might influence expression of *GLUT1* causing epilepsy (Liu Y.C. et al., 2016).

Two large-scale analyses focused on NCREs and their role in neurodevelopmental disorders have recently been performed. Using a targeted sequencing approach, Short and colleagues studied *de novo* occurring genomic variants in three classes of putative regulatory elements in 7,930 individuals suffering from developmental disorders from the Deciphering Developmental Disorders (DDDs) study and their parents. The three classes of regulatory elements that they assessed consisted of 4,307 highly evolutionarily conserved non-coding elements (Siepel et al., 2005), 595 experimentally validated enhancers (Visel et al., 2007), and 1,237 putative heart enhancers (May et al., 2011), together covering 4.2 Mb of genomic sequence. In the 6,239 individuals in which exome sequencing did not find a disease cause, they found that conserved non-coding elements were nominally significantly enriched for *de novo* variants (422 observed, 388 expected, $P = 0.04$), whereas in experimentally validated enhancers (153 observed, 156 expected, $P = 0.605$), heart enhancers (86 observed, 86 expected, $P = 0.514$), and intronic controls (901 observed, 919 expected, $P = 0.728$) *de novo* variants were not enriched. When focusing only on conserved non-coding elements that had evidence of activity in brain, they observed an even stronger enrichment. Based on their analysis, the authors estimate that only around 1–3%

of exome-negative individuals will be explained by *de novo* variants in fetal brain-active regulatory elements. However, as in this study only *de novo* variants were assessed, and only a limited set of regulatory elements was used which were already defined in 2010, this is likely an underestimation of the possible impact of the non-coding genome for neurodevelopmental disorders. Doan and colleagues performed a similar targeted sequencing approach assessing so-called human accelerated regions (HARs) (Doan et al., 2016). HARs are conserved regions with elevated divergence in humans and this might reflect potential roles in the evolution of human-specific traits. This study provides evidence that HARs can function as regulatory elements for dosage-sensitive genes expressed in the central nervous system. Using data from a large cohort study investigating 2100 sibling cases of ASD, they found that *de novo* CNV's affecting HARs, or HAR-containing genes, could be implicated in up to 1.9% of ASD cases in simplex families. They then analyzed consanguineous ASD cases using WGS from 30 affected and 5 unaffected individuals and designed a custom capture array to sequence HARs in another 188 affected and 172 unaffected individuals. Individuals with ASD exhibited an excess of rare (AF < 0.5%) bi-allelic HAR alleles (43% excess compared to unaffected, $P = 0.008$), and this enrichment further increased when only taking HARs in to consideration that were likely active as regulatory elements in brain. Using MPRA, 343 bi-allelic HAR variants were functionally tested, and 29% of these were shown to alter the regulatory activity of the reference sequence. Therefore, the enrichment of regulation-altering variants in HARs with predicted activity suggests that many may contribute to the pathogenesis and diversity of ASD. They further functionally validated their findings in three examples of bi-allelic variants in HARs identified in ASD families, regulating the genes *CUX1*, *PTBP2* and *GPC4*, further providing evidence that the investigation of NCREs such as HARs is promising to solve currently genetically unexplained disease cases.

What appears from the examples given above, is that a wide range of NCRE alterations can result in effects on gene transcription, leading to a disease phenotype. This can vary from point mutations affecting the binding of crucial TFs, deletions or duplications of NCRE sequences, shuffling of the genomic location of NCREs affecting their function (e.g., enhancer adoption) or alterations in the global chromatin landscape disrupting borders of TADs, just to mention a few. Given the complexity that can exist in these disease mechanisms, and the current shortcomings in understanding the complete functionality of the non-coding genome, it seems likely that in the next decade many more examples of NCRE alteration in genetic disease will be identified.

THE FUTURE AHEAD

As we have discussed in this *Review*, our understanding of gene regulation has deepened over the last decade. NCREs have been identified as crucial modifiers of gene expression,

and more and more examples of their involvement in human genetic disorders, when mutant, are being reported. In routine clinical practice genetic analysis has mainly focused on the ~2% of the genome that directly encodes for proteins. Most of the non-coding part has been instead neglected, and only recently we could witness a shift of attention toward these sequences. It seems intuitive that, if human evolution resulted in a large and subsequently maintained expansion of the non-coding genome, this should have a functional role, and alterations of these sequences should influence their function and lead to genetic disorders. Given the fact that MCDs are often genetically unexplained despite the routine use of WES, it would be surprising if, in the near future, no genetic alterations of non-coding sequences will be identified in those currently unexplained patients. In order to achieve this, it is crucial to develop novel diagnostic approaches focusing on non-coding regions. Will WGS be useful to find disease causes in those unexplained MCD patients in a clinical setting? Theoretically yes as it will enable the identification of all detectable variants genome-wide, but our current understanding of genomic variation outside exons severely hampers its routine implementation. As a matter of fact, most studies that have used WGS in a clinical setting, have limited their analysis to those nucleotides covering exons, deep intronic variants not covered in WES and copy number and structural variants (Stavropoulos et al., 2016; Lionel et al., 2018; Clark et al., 2019; Scocchia et al., 2019). Therefore, it remains crucial to gain more detailed information on the functional relevance of NCREs and their variants from a basic science point of view. Although the characterization of epigenomic marks such as histone modifications has shown to be useful to identify functional NCREs, it is clear from the discussion above that there are still some pitfalls, as we still lack the perfect mark to identify relevant and active NCREs. One particular concern is that many studies assume that investigating a single histone modification, such as H3K27ac, gives sufficient evidence to call a region a functional NCRE, but this is certainly an oversimplification. As we have argued above, predicted NCREs should remain classified as putative NCREs till they are functionally validated, or at least predicted in multiple studies ideally using different techniques to obtain a higher level of confidence in their function. In current studies, functional validations of putative NCREs is often performed only for a selected number of regions of interest and results of these few validations are extrapolated to the complete data set generated. Even though this is understandable from a pragmatic experimental point of view, it might be one of the reasons for the broad level of variation between predicted enhancers from different sources. Hence, it is crucial to further develop high-throughput approaches for functional validation studies so that more sequences and their variants can be directly functionally tested, leading to a higher confidence in the data resources. The future application of direct functional assays, such as MPRAs and CRISPR-Cas9 based screens, is expected to further add on to our current understanding, even though also these methods are far from perfect yet. Besides these emerging experimental techniques, it is also crucial to

develop novel computational tools that outperform currently available programs for NCRE prediction and disease annotation. Similarly, it is also important to further improve the linking between NCREs and their target genes, going beyond the current resolution of chromatin conformation capture or correlation between activity of putative enhancers and expression of possibly linked genes. Until we will have all these ideal tools widely available, in our opinion the best practice to study the role of the non-coding genome in genetic disorders such as MCDs is to study genetic variation outside exomes in well-defined, exome negative patients and preferably combine this with a direct readout of gene expression in a disease relevant tissue. For the functional annotation of the non-coding variants found in patients, it is essential to use as many sources of information as possible, enabling the highest level of confidence in defining a certain region a regulatory sequence. And last but certainly not least, a detailed clinical phenotyping of patients prior to any genetic investigation remains crucial as it allows the comparison of patients with similar non-coding variants and shared phenotypes. Even in an era where it is cheap to sequence a whole genome, reverse phenotyping of patients remains essential to learn more about the consequences of the genetic variants and to further mature our understanding of the non-coding genome beyond the borders of the exome.

REFERENCES

Agirman, G., Broix, L., and Nguyen, L. (2017). Cerebral cortex development: an outside-in perspective. *FEBS Lett.* 591, 3978–3992. doi: 10.1002/1873-3468. 12924

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A., et al. (2007). Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 5:e234. doi: 10.1371/journal.pbio.0050234

Amiri, A., Coppola, G., Scuderi, S., Wu, F., Roychowdhury, T., Liu, F., et al. (2018). Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* 362:eaat6720. doi: 10.1126/science.aat6720

Amlie-Wolf, A., Tang, M., Mlynarski, E. E., Kuksa, P. P., Valladares, O., Katanic, Z., et al. (2018). INFERNO: inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Res.* 46, 8740–8753. doi: 10.1093/nar/gky686

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461. doi: 10.1038/nature12787

Arnold, C. D., Gerlach, D., Spies, D., Matts, J. A., Sytnikova, Y. A., Pagani, M., et al. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* 46, 685–692. doi: 10.1038/ng. 3009

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–1077. doi: 10.1126/science.1232542

Arnold, C. D., Zabidi, M. A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., et al. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol.* 35, 136–144. doi: 10.1038/nbt.3739

Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N. J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* 11, 2494–2509. doi: 10.1101/gad.11.19.2494

Bae, B. I., Tietjen, I., Atabay, K. D., Evrony, G. D., Johnson, M. B., Asare, E., et al. (2014). Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* 343, 764–768. doi: 10.1126/science. 1244392

Ball, A. R. Jr., Chen, Y. Y., and Yokomori, K. (2014). Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. *Biochim. Biophys. Acta* 1839, 191–202. doi: 10.1016/j.bbagrm.2013.11.002

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27(2 Pt 1), 299–308. doi: 10.1016/0092-8674(81)90413-x

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi: 10.1038/cr.2011.22

Barakat, T. S., and Gribnau, J. (2010). X chromosome inactivation and embryonic stem cells. *Adv. Exp. Med. Biol.* 695, 132–154. doi: 10.1007/978-1-4419-7037-4_10

Barakat, T. S., Halbritter, F., Zhang, M., Rendeiro, A. F., Perenthaler, E., Bock, C., et al. (2018). Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell* 23, 276.e8–288.e8. doi: 10.1016/j.stem. 2018.06.014

Barkovich, A. J., Kuzniecky, R. I., Dobyns, W. B., Jackson, G. D., Becker, L. E., and Evrard, P. (1996). A classification scheme for malformations of cortical development. *Neuropediatrics* 27, 59–63. doi: 10.1055/s-2007-973750

Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., et al. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* 27, 1139–1152. doi: 10.1101/gr. 215160.116

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., et al. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321–1325. doi: 10.1126/science.1098119

Bell, A. C., West, A. G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396. doi: 10.1016/s0092-8674(00)81967-4

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* 41, 359–364. doi: 10.1038/ng.329

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., de Marco, R., et al. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* 93, 1126–1134. doi: 10.1016/j.ajhg.2013.10.028

AUTHOR CONTRIBUTIONS

EP, SY, and EN performed the literature research and wrote the sections of the manuscript. TB conceived the manuscript, wrote the sections, supervised the work, and obtained funding. All authors approved the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

Bilguvar, K., Ozturk, A. K., Louvi, A., Kwan, K. Y., Choi, M., Tatli, B., et al. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 467, 207–210. doi: 10.1038/nature09327

Bosnjak, V. M., Dakovic, I., Duranovic, V., Lujic, L., Krakar, G., and Marn, B. (2011). Malformations of cortical development in children with congenital cytomegalovirus infection - A study of nine children with proven congenital cytomegalovirus infection. *Coll. Antropol.* 35(Suppl. 1), 229–234.

Bouman, A., van Haelst, M., and van Spaendonk, R. (2018). Blepharophimosis-ptosis-epicanthus inversus syndrome caused by a 54-kb microdeletion in a FOXL2 cis-regulatory element. *Clin. Dysmorphol.* 27, 58–62.

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., et al. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322. doi: 10.1016/j.cell.2007.12.014

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112

Brandler, W. M., Antaki, D., Gujral, M., Kleiber, M. L., Whitney, J., Maile, M. S., et al. (2018). Paternally inherited cis-regulatory structural variants are associated with autism. *Science* 360, 327–331. doi: 10.1126/science.aan2261

Bu, H., Gan, Y., Wang, Y., Zhou, S., and Guan, J. (2017). A new method for enhancer prediction based on deep belief network. *BMC Bioinformatics* 18(Suppl. 12):418. doi: 10.1186/s12859-017-1828-0

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218. doi: 10.1038/nmeth.2688

Canver, M. C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E. N., et al. (2017). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* 49, 625–634. doi: 10.1038/ng.3793

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192–197. doi: 10.1038/nature15521

Carleton, J. B., Berrett, K. C., and Gertz, J. (2017). Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor alpha-bound enhancers. *Cell Syst.* 5, 333.e5–344.e5. doi: 10.1016/j.cels.2017.08.011

Castel, S. E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., et al. (2018). Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet* 50, 1327–1334. doi: 10.1038/s41588-018-0192-y

Catarino, R. R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* 32, 202–223. doi: 10.1101/gad.310367.117

Cavallin, M., Maillard, C., Hully, M., Philbert, M., Boddaert, N., Reilly, M. L., et al. (2018). TLE1, a key player in neurogenesis, a new candidate gene for autosomal recessive postnatal microcephaly. *Eur. J. Med. Genet.* 61, 729–732. doi: 10.1016/j.ejmg.2018.05.002

Cellini, E., Vetro, A., Conti, V., Marini, C., Doccini, V., Clementella, C., et al. (2019). Multiple genomic copy number variants associated with periventricular nodular heterotopia indicate extreme genetic heterogeneity. *Eur. J. Hum. Genet.* 27, 909–918. doi: 10.1038/s41431-019-0335-3

Clark, M. M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., et al. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* 11:eaat6177. doi: 10.1126/scitranslmed.aat6177

Collis, P., Antoniou, M., and Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *Embo. J.* 9, 233–240. doi: 10.1002/j.1460-2075.1990.tb08100.x

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Cotney, J., Leng, J., Oh, S., DeMare, L. E., Reilly, S. K., Gerstein, M. B., et al. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* 22, 1069–1080. doi: 10.1101/gr.129817.111

Coulon, A., Chow, C. C., Singer, R. H., and Larson, D. R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* 14, 572–584. doi: 10.1038/nrg3484

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21931–21936. doi: 10.1073/pnas.1016071107

Crick, F. H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.

Davies, J. O., Oudelaar, A. M., Higgs, D. R., and Hughes, J. R. (2017). How best to identify chromosomal interactions: a comparison of approaches. *Nat. Methods* 14, 125–134. doi: 10.1038/nmeth.4146

de la Torre-Ubieta, L., Stein, J. L., Won, H., Opland, C. K., Liang, D., Lu, D., et al. (2018). The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell* 172, 289.e18–304.e18. doi: 10.1016/j.cell.2017.12.014

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215. doi: 10.1038/nature13772

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24. doi: 10.1101/gad.179804.111

de Wit, M. C., Lequin, M. H., de Coo, I. F., Brusse, E., Halley, D. J., van der Graaf, R., et al. (2008). Cortical brain malformations: effect of clinical, neuroradiological, and modern genetic classification. *Arch. Neurol.* 65, 358–366. doi: 10.1001/archneur.65.3.358

Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J. B., Gaffney, D. J., Pickrell, J. K., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394. doi: 10.1038/nature10808

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

Devanna, P., van de Vorst, M., Pfundt, R., Gilissen, C., and Vernes, S. C. (2018). Genome-wide investigation of an ID cohort reveals de novo 3'UTR variants affecting gene expression. *Hum. Genet.* 137, 717–721. doi: 10.1007/s00439-018-1925-9

Di Donato, N., Jean, Y. Y., Maga, A. M., Krewson, B. D., Shupp, A. B., Avrutsky, M. I., et al. (2016). Mutations in CRADD result in reduced caspase-2-mediated neuronal apoptosis and cause megalencephaly with a rare lissencephaly variant. *Am. J. Hum. Genet.* 99, 1117–1129. doi: 10.1016/j.ajhg.2016.09.010

Di Donato, N., Timms, A. E., Aldinger, K. A., Mirzaa, G. M., Bennett, J. T., Collins, S., et al. (2018). Analysis of 17 genes detects mutations in 81% of 811 patients with lissencephaly. *Genet. Med.* 20, 1354–1364. doi: 10.1038/gim.2018.8

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* 14, 629–635. doi: 10.1038/nmeth.4264

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A. Y., Dixon, J., et al. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* 26, 397–405. doi: 10.1101/gr.197152.115

Dickel, D. E., Ypsilanti, A. R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B. J., et al. (2018). Ultraconserved enhancers are required for normal development. *Cell* 172, 491.e15–499.e15. doi: 10.1016/j.cell.2017.12.017

Dickel, D. E., Zhu, Y., Nord, A. S., Wylie, J. N., Akiyama, J. A., Afzal, V., et al. (2014). Function-based identification of mammalian enhancers using site-specific integration. *Nat. Methods* 11, 566–571. doi: 10.1038/nmeth.2886

Doan, R. N., Bae, B. I., Cubelos, B., Chang, C., Hossain, A. A., Al-Saad, S., et al. (2016). Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 167, 341.e12–354.e12. doi: 10.1016/j.cell.2016.08.071

Dobyns, W. B., Aldinger, K. A., Ishak, G. E., Mirzaa, G. M., Timms, A. E., Grout, M. E., et al. (2018). MACF1 mutations encoding highly conserved zinc-binding residues of the GAR domain cause defects in neuronal migration and axon guidance. *Am. J. Hum. Genet.* 103, 1009–1021. doi: 10.1016/j.ajhg.2018.10.019

Dogan, N., Wu, W., Morrissey, C. S., Chen, K. B., Stonestrom, A., Long, M., et al. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* 8:16. doi: 10.1186/s13072-015-0009-5

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309. doi: 10.1101/gr.5571506

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi: 10.1016/j.cell.2014.09.030

Ercan-Sencicek, A. G., Jambi, S., Franjic, D., Nishimura, S., Li, M., El-Fishawy, P., et al. (2015). Homozygous loss of DIAPH1 is a novel cause of microcephaly in humans. *Eur. J. Hum. Genet.* 23, 165–172. doi: 10.1038/ejhg.2014.82

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216. doi: 10.1038/nmeth.1906

Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T. S., et al. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat. Biotechnol.* 34, 1180–1190. doi: 10.1038/nbt.3678

Erwin, G. D., Oksenberg, N., Truty, R. M., Kostka, D., Murphy, K. K., Ahituv, N., et al. (2014). Integrating diverse datasets improves developmental enhancer prediction. *PLoS Comput. Biol.* 10:e1003677. doi: 10.1371/journal.pcbi.1003677

Fernandez, M., and Miranda-Saavedra, D. (2012). Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. *Nucleic Acids Res.* 40:e77. doi: 10.1093/nar/gks149

Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97.

Firpi, H. A., Ucar, D., and Tan, K. (2010). Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinformatics* 26, 1579–1586. doi: 10.1093/bioinformatics/btq248

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017:bax028. doi: 10.1093/database/bax028

Fry, A. E., Fawcett, K. A., Zelnik, N., Yuan, H., Thompson, B. A. N., Shemer-Meiri, L., et al. (2018). De novo mutations in GRIN1 cause extensive bilateral polymicrogyria. *Brain* 141, 698–712. doi: 10.1093/brain/awx358

Fulco, C. P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S. R., Perez, E. M., et al. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769–773. doi: 10.1126/science.aag2445

Funnell, A. P., and Crossley, M. (2012). Homo- and heterodimerization in transcriptional regulation. *Adv. Exp. Med. Biol.* 747, 105–121. doi: 10.1007/978-1-4614-3229-6_7

Gabriele, M., Vulto-van Silfhout, A. T., Germain, P. L., Vitriolo, A., Kumar, R., Douglas, E., et al. (2017). YY1 haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am. J. Hum. Genet.* 100, 907–925. doi: 10.1016/j.ajhg.2017.05.006

Gao, T., He, B., Liu, S., Zhu, H., Tan, K., and Qian, J. (2016). EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* 32, 3543–3551.

Gasperini, M., Findlay, G. M., McKenna, A., Milbank, J. H., Lee, C., Zhang, M. D., et al. (2017). CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* 101, 192–205. doi: 10.1016/j.ajhg.2017.06.010

Gasperini, M., Hill, A. J., McFaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., et al. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* 176, 377.e19–390.e19. doi: 10.1016/j.cell.2018.11.029

Genereaux, D., van Karnebeek, C. D., and Birch, P. H. (2015). Costs of caring for children with an intellectual developmental disorder. *Disabil. Health J.* 8, 646–651. doi: 10.1016/j.dhjo.2015.03.011

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044

Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., et al. (2015). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum. Mol. Genet.* 24, 3143–3154. doi: 10.1093/hmg/ddv065

Giresi, P. G., and Lieb, J. D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* 48, 233–239. doi: 10.1016/j.ymeth.2009.03.003

Glinsky, G., and Barakat, T. S. (2019). The evolution of Great Apes has shaped the functional enhancers' landscape in human embryonic stem cells. *Stem Cell Res* 37:101456. doi: 10.1016/j.scr.2019.101456

Gloss, B. S., and Dinger, M. E. (2018). Realizing the significance of noncoding functionality in clinical genomics. *Exp. Mol. Med.* 50:97. doi: 10.1038/s12276-018-0087-0

Gonzalez-Moron, D., Vishnopolska, S., Consalvo, D., Medina, N., Marti, M., Cordoba, M., et al. (2017). Germline and somatic mutations in cortical malformations: molecular defects in argentinean patients with neuronal migration disorders. *PLoS One* 12:e0185103. doi: 10.1371/journal.pone.0185103

Gregor, A., Oti, M., Kouwenhoven, E. N., Hoyer, J., Sticht, H., Ekici, A. B., et al. (2013). De novo mutations in the genome organizer CTCF cause intellectual disability. *Am. J. Hum. Genet.* 93, 124–131. doi: 10.1016/j.ajhg.2013.05.007

Guerrini, R., and Dobyns, W. B. (2014). Malformations of cortical development: clinical features and genetic causes. *Lancet Neurol.* 13, 710–726. doi: 10.1016/S1474-4422(14)70040-7

Gurnett, C. A., Bowcock, A. M., Dietz, F. R., Morcuende, J. A., Murray, J. C., and Dobbs, M. B. (2007). Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am. J. Med. Genet. A* 143A, 27–32. doi: 10.1002/ajmg.a.31563

Hah, N., Murakami, S., Nagari, A., Danko, C. G., and Kraus, W. L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 23, 1210–1223. doi: 10.1101/gr.152306.112

Halfon, M. S. (2019). Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases. *Trends Genet.* 35, 93–103. doi: 10.1016/j.tig.2018.11.004

Hama, Y., Katsu, M., Takigawa, I., Yabe, I., Matsushima, M., Takahashi, I., et al. (2017). Genomic copy number variation analysis in multiple system atrophy. *Mol. Brain* 10:54. doi: 10.1186/s13041-017-0335-6

Han, R., Li, L., Ugalde, A. P., Tal, A., Manber, Z., Barbera, E. P., et al. (2018). Functional CRISPR screen identifies AP1-associated enhancer regulating FOXF1 to modulate oncogene-induced senescence. *Genome Biol.* 19:118. doi: 10.1186/s13059-018-1494-1

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L., et al. (2016). Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat. Genet.* 48, 895–903. doi: 10.1038/ng.3605

He, Y., Gorkin, D. U., Dickel, D. E., Nery, J. R., Castanon, R. G., Lee, A. Y., et al. (2017). Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1633–E1640. doi: 10.1073/pnas.1618353114

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318. doi: 10.1038/ng1966

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., et al. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517. doi: 10.1038/nbt.3199

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., et al. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947. doi: 10.1016/j.cell.2013.09.053

Hnisz, D., Day, D. S., and Young, R. A. (2016). Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 167, 1188–1200. doi: 10.1016/j.cell.2016.10.024

Hoffman, M. M., Buske, O. J., Wang, J., Weng, Z., Bilmes, J. A., and Noble, W. S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* 9, 473–476. doi: 10.1038/nmeth.1937

Holsten, T., Bens, S., Oyen, F., Nemes, K., Hasselblatt, M., Kordes, U., et al. (2018). Germline variants in SMARCB1 and other members of the BAF chromatin-remodeling complex across human disease entities: a meta-analysis. *Eur. J. Hum. Genet.* 26, 1083–1093. doi: 10.1038/s41431-018-0143-1

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* 49:e324. doi: 10.1038/emm.2017.11

Inoue, F., Kircher, M., Martin, B., Cooper, G. M., Witten, D. M., McManus, M. T., et al. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* 27, 38–52. doi: 10.1101/gr.212092.116

Ishibashi, M., Manning, E., Shoubridge, C., Krecsmarik, M., Hawkins, T. A., Giacomotto, J., et al. (2015). Copy number variants in patients with intellectual disability affect the regulation of ARX transcription factor gene. *Hum. Genet.* 134, 1163–1182. doi: 10.1007/s00439-015-1594-x

Iwafuchi-Doi, M. (2019). The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 11:e1427. doi: 10.1002/wsbm.1427

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., et al. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* 62, 79–91. doi: 10.1016/j.molcel.2016.03.001

Jamuar, S. S., Lam, A. T., Kircher, M., D'Gama, A. M., Wang, J., Barry, B. J., et al. (2014). Somatic mutations in cerebral cortical malformations. *N Engl. J. Med.* 371, 733–743. doi: 10.1056/NEJMoa1314432

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* 40, 1348–1353. doi: 10.1038/ng.230

Jiang, Y., Qian, F., Bai, X., Liu, Y., Wang, Q., Ai, B., et al. (2019). SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.* 47, D235–D243. doi: 10.1093/nar/gky1025

John, S., Sabo, P. J., Canfield, T. K., Lee, K., Vong, S., Weaver, M., et al. (2013). Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol.* 130, 21.27.1–21.27.20. doi: 10.1002/0471142727.mb2127s103

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.

Jonsson, U., Alaie, I., Lofgren Wilteus, A., Zander, E., Marschik, P. B., Coghill, D., et al. (2017). Annual research review: quality of life and childhood mental and behavioural disorders - a critical review of the research. *J. Child Psychol. Psychiatry* 58, 439–469. doi: 10.1111/jcpp.12645

Kaikkonen, M. U., Spann, N. J., Heinz, S., Romanoski, C. E., Allison, K. A., Stender, J. D., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51, 310–325. doi: 10.1016/j.molcel.2013.07.010

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M., et al. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* 12, 401–403. doi: 10.1038/nmeth.3325

Khan, A., and Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.* 44, D164–D171. doi: 10.1093/nar/gkv1002

Kheradmand Kia, S., Verbeek, E., Engelen, E., Schot, R., Poot, R. A., de Coo, I. F., et al. (2012). RTTN mutations link primary cilia function to organization of the human cerebral cortex. *Am. J. Hum. Genet.* 91, 533–540. doi: 10.1016/j.ajhg.2012.07.008

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., et al. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* 23, 800–811. doi: 10.1101/gr.144899.112

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187. doi: 10.1038/nature09033

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., Song, L., Hilton, I. B., et al. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* 35, 561–568. doi: 10.1038/nbt.3853

Kleftogiannis, D., Kalnis, P., and Bajic, V. B. (2015). DEEP: a general computational framework for predicting enhancers. *Nucleic Acids Res.* 43:e6. doi: 10.1093/nar/gku1058

Kleftogiannis, D., Kalnis, P., and Bajic, V. B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinformatics* 17, 967–979. doi: 10.1093/bib/bbv101

Klopocki, E., Ott, C. E., Benatar, N., Ullmann, R., Mundlos, S., and Lehmann, K. (2008). A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J. Med. Genet.* 45, 370–375. doi: 10.1136/jmg.2007.055699

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18, 956–963. doi: 10.1038/nsmb.2085

Konermann, S., Brigham, M. D., Trevino, A., Hsu, P. D., Heidenreich, M., Cong, L., et al. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472–476. doi: 10.1038/nature12466

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588. doi: 10.1038/nature14136

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., et al. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* 34, 192–198. doi: 10.1038/nbt.3450

Kuska, B. (1998). Should scientists scrap the notion of junk DNA? *J. Natl. Cancer Inst.* 90, 1032–1033. doi: 10.1093/jnci/90.14.1032

Kvon, E. Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J. O., Pagani, M., Schernhuber, K., et al. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* 512, 91–95. doi: 10.1038/nature13395

Kwasnieski, J. C., Fiore, C., Chaudhari, H. G., and Cohen, B. A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* 24, 1595–1602. doi: 10.1101/gr.173518.114

Kwon, D. Y., Zhao, Y. T., Lamonica, J. M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* 8:15315. doi: 10.1038/ncomms15315

Lam, M. T., Li, W., Rosenfeld, M. G., and Glass, C. K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* 39, 170–182. doi: 10.1016/j.tibs.2014.02.007

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172, 650–665. doi: 10.1016/j.cell.2018.01.029

Laurell, T., Vandermeer, J. E., Wenger, A. M., Grigelioniene, G., Nordenskjold, A., Arner, M., et al. (2012). A novel 13 base pair insertion in the sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Hum. Mutat.* 33, 1063–1066. doi: 10.1002/humu.22097

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735. doi: 10.1093/hmg/ddg180

Lettice, L. A., Horikoshi, T., Heaney, S. J., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7548–7553. doi: 10.1073/pnas.112212199

Li, L., Zhu, Q., He, X., Sinha, S., and Halfon, M. S. (2007). Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol.* 8:R101.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O. V., Gulden, F. O., Pochareddy, S., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362:eaat7615. doi: 10.1126/science.aat7615

Lidor Nili, E., Field, Y., Lubling, Y., Widom, J., Oren, M., and Segal, E. (2010). p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. *Genome Res.* 20, 1361–1368. doi: 10.1101/gr.103945.109

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Lim, L. W. K., Chung, H. H., Chong, Y. L., and Lee, N. K. (2018). A survey of recently emerged genome-wide computational enhancer predictor tools. *Comput. Biol. Chem.* 74, 132–141. doi: 10.1016/j.compbiolchem.2018.03.019

Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., et al. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* 20, 435–443. doi: 10.1038/gim.2017.119

Liu, F., Li, H., Ren, C., Bo, X., and Shu, W. (2016). PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Sci. Rep.* 6:28517. doi: 10.1038/srep28517

Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., et al. (2016). Editing DNA methylation in the mammalian genome. *Cell* 167, 233.e17–247.e17.

Liu, Y. C., Lee, J. W., Bellows, S. T., Damiano, J. A., Mullen, S. A., Berkovic, S. F., et al. (2016). Evaluation of non-coding variation in GLUT1 deficiency. *Dev. Med. Child Neurol.* 58, 1295–1302. doi: 10.1111/dmcn.13163

Loven, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., et al. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334. doi: 10.1016/j.cell.2013.03.036

Lu, I. L., Chen, C., Tung, C. Y., Chen, H. H., Pan, J. P., Chang, C. H., et al. (2018). Identification of genes associated with cortical malformation using a transposon-mediated somatic mutagenesis screen in mice. *Nat. Commun.* 9:2498. doi: 10.1038/s41467-018-04880-8

Lu, Q., Hu, Y., Sun, J., Cheng, Y., Cheung, K. H., and Zhao, H. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Sci. Rep.* 5:10576. doi: 10.1038/srep10576

Lu, Y., Qu, W., Shan, G., and Zhang, C. (2015). DELTA: a distal enhancer locating tool based on AdaBoost algorithm and shape features of chromatin modifications. *PLoS One* 10:e0130622. doi: 10.1371/journal.pone.0130622

Ludwig, M. Z., Bergman, C., Patel, N. H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564–567. doi: 10.1038/35000615

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., and Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* 10, 977–979. doi: 10.1038/nmeth.2598

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838. doi: 10.1038/nbt.2675

Maricque, B. B., Chaudhari, H. G., and Cohen, B. A. (2018). A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nat. Biotechnol.* doi: 10.1038/nbt.4285 [Epub ahead of print].

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195. doi: 10.1126/science.1222794

May, D., Blow, M. J., Kaplan, T., McCulley, D. J., Jensen, B. C., Akiyama, J. A., et al. (2011). Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* 44, 89–93. doi: 10.1038/ng.1006

McCole, R. B., Erceg, J., Saylor, W., and Wu, C. T. (2018). Ultraconserved elements occupy specific arenas of three-dimensional mammalian genome organization. *Cell Rep.* 24, 479–488. doi: 10.1016/j.celrep.2018.06.031

McMahon, K. Q., Papandreou, A., Ma, M., Barry, B. J., Mirzaa, G. M., Dobyns, W. B., et al. (2015). Familial recurrences of FOXG1-related disorder: evidence for mosaicism. *Am. J. Med. Genet. A* 167A, 3096–3102. doi: 10.1002/ajmg.a.37353

Mehrjouy, M. M., Fonseca, A. C. S., Ehmke, N., Paskulin, G., Novelli, A., Benedicenti, F., et al. (2018). Regulatory variants of FOXG1 in the context of its topological domain organisation. *Eur. J. Hum. Genet.* 26, 186–196. doi: 10.1038/s41431-017-0011-4

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30, 271–277. doi: 10.1038/nbt.2137

Mirabella, A. C., Foster, B. M., and Bartke, T. (2016). Chromatin deregulation in disease. *Chromosoma* 125, 75–93. doi: 10.1007/s00412-015-0530-0

Mirzaa, G., Timms, A. E., Conti, V., Boyle, E. A., Girisha, K. M., Martin, B., et al. (2016). PIK3CA-associated developmental disorders exhibit distinct classes of mutations with variable expression and tissue distribution. *JCI Insight* 1:87623. doi: 10.1172/jci.insight.87623

Monlong, J., Girard, S. L., Meloche, C., Cadieux-Dion, M., Andrade, D. M., Lafreniere, R. G., et al. (2018). Global characterization of copy number variants in epilepsy patients from whole genome sequencing. *PLoS Genet.* 14:e1007285. doi: 10.1371/journal.pgen.1007285

Moore, C. A., Staples, J. E., Dobyns, W. B., Pessoa, A., Ventura, C. V., Fonseca, E. B., et al. (2017). Characterizing the pattern of anomalies in congenital zika syndrome for pediatric clinicians. *JAMA Pediatr.* 171, 288–295. doi: 10.1001/jamapediatrics.2016.3982

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* 9, 6047–6068. doi: 10.1093/nar/9.22.6047

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., et al. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922. doi: 10.1038/nmeth.3999

Murtha, M., Tokcaer-Keskin, Z., Tang, Z., Strino, F., Chen, X., Wang, Y., et al. (2014). FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat. Methods* 11, 559–565. doi: 10.1038/nmeth.2885

Natoli, G., and Andrau, J. C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* 46, 1–19. doi: 10.1146/annurev-genet-110711-155459

Ngcungcu, T., Oti, M., Sitek, J. C., Haukanes, B. I., Linghu, B., Bruccoleri, R., et al. (2017). Duplicated enhancer region increases expression of CTSB and segregates with keratolytic winter erythema in south african and norwegian families. *Am. J. Hum. Genet.* 100, 737–750. doi: 10.1016/j.ajhg.2017.03.012

Niemi, M. E. K., Martin, H. C., Rice, D. L., Gallone, G., Gordon, S., Kelemen, M., et al. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* 562, 268–271. doi: 10.1038/s41586-018-0566-4

Nmezi, B., Giorgio, E., Raininko, R., Lehman, A., Spielmann, M., Koenig, M. K., et al. (2019). Genomic deletions upstream of lamin B1 lead to atypical autosomal dominant leukodystrophy. *Neurol. Genet.* 5:e305. doi: 10.1212/NXG.0000000000000305

Ohno, S. (1972). So much "junk" DNA in our genome. *Brookhaven Symp. Biol.* 23, 366–370.

Okonechnikov, K., Erkek, S., Korbel, J. O., Pfister, S. M., and Chavez, L. (2019). InTAD: chromosome conformation guided analysis of enhancer target genes. *BMC Bioinformatics* 20:60. doi: 10.1186/s12859-019-2655-2

Ong, C. T., and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246. doi: 10.1038/nrg3663

Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243. doi: 10.1038/nature25461

Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17921–17926. doi: 10.1073/pnas.1317023110

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* 30, 265–270. doi: 10.1038/nbt.2136

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502. doi: 10.1038/nature05295

Piluso, G., Monteleone, P., Galderisi, S., Giugliano, T., Bertolino, A., Rocca, P., et al. (2019). Assessment of de novo copy-number variations in Italian patients with schizophrenia: detection of putative mutations involving regulatory enhancer elements. *World J. Biol. Psychiatry* 20, 126–136. doi: 10.1080/15622975.2017.1395072

Potuijt, J. W. P., Baas, M., Sukenik-Halevy, R., Douben, H., Nguyen, P., Venter, D. J., et al. (2018). A point mutation in the pre-ZRS disrupts sonic hedgehog expression in the limb bud and results in triphalangeal thumb-polysyndactyly syndrome. *Genet. Med.* 20, 1405–1413. doi: 10.1038/gim.2018.18

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C., Schneider, R., et al. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* 48, 681–686. doi: 10.1038/ng.3550

Protas, M. E., Weh, E., Footz, T., Kasberger, J., Baraban, S. C., Levin, A. V., et al. (2017). Mutations of conserved non-coding elements of PITX2 in patients with ocular dysgenesis and developmental glaucoma. *Hum. Mol. Genet.* 26, 3630–3638. doi: 10.1093/hmg/ddx251

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., et al. (2016). High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* 34, 167–174. doi: 10.1038/nbt.3468

Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoyannopoulos, J., et al. (2013). RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* 9:e1002968. doi: 10.1371/journal.pcbi.1002968

Reilly, S. K., Yin, J., Ayoub, A. E., Emera, D., Leng, J., Cotney, J., et al. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* 347, 1155–1159. doi: 10.1126/science.1260943

Reiner, O., Carrozzo, R., Shen, Y., Wehnert, M., Faustinella, F., Dobyns, W. B., et al. (1993). Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats. *Nature* 364, 717–721. doi: 10.1038/364717a0

Ritchie, G. R., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nat. Methods* 11, 294–296. doi: 10.1038/nmeth.2832

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657. doi: 10.1038/nmeth1068

Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. doi: 10.1038/s41576-018-0060-8

Sanjana, N. E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., et al. (2016). High-resolution interrogation of functional elements in the noncoding genome. *Science* 353, 1545–1549. doi: 10.1126/science.aaf7613

Scocchia, A., Wigby, K. M., Masser-Frye, D., Del Campo, M., Galarreta, C. I., Thorpe, E., et al. (2019). Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. *NPJ Genom. Med.* 4:5. doi: 10.1038/s41525-018-0076-1

Sen, D. R., Kaminski, J., Barnitz, R. A., Kurachi, M., Gerdemann, U., Yates, K. B., et al. (2016). The epigenetic landscape of T cell exhaustion. *Science* 354, 1165–1169.

Seto, Y., and Eiraku, M. (2019). Human brain development and its in vitro recapitulation. *Neurosci. Res.* 138, 33–42. doi: 10.1016/j.neures.2018.09.011

Shen, S. Q., Myers, C. A., Hughes, A. E., Byrne, L. C., Flannery, J. G., and Corbo, J. C. (2016). Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* 26, 238–255. doi: 10.1101/gr.193789.115

Shibata, Y., Sheffield, N. C., Fedrigo, O., Babbitt, C. C., Wortham, M., Tewari, A. K., et al. (2012). Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 8:e1002789. doi: 10.1371/journal.pgen.1002789

Shin, H. Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G., et al. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* 48, 904–911. doi: 10.1038/ng.3606

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286. doi: 10.1038/nrg3682

Short, P. J., McRae, J. F., Gallone, G., Sifrim, A., Won, H., Geschwind, D. H., et al. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* 555, 611–616. doi: 10.1038/nature25983

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050. doi: 10.1101/gr.3715005

Siersbaek, R., Nielsen, R., John, S., Sung, M. H., Baek, S., Loft, A., et al. (2011). Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *Embo. J.* 30, 1459–1472. doi: 10.1038/emboj.2011.65

Silva, T. C., Coetzee, S. G., Gull, N., Yao, L., Hazelett, D. J., Noushmehr, H., et al. (2018). ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* 35, 1974–1977. doi: 10.1093/bioinformatics/bty902

Simeonov, D. R., Gowen, B. G., Boontanrart, M., Roth, T. L., Gagnon, J. D., Mumbach, M. R., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* 549, 111–115. doi: 10.1038/nature23875

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains

uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38, 1348–1354. doi: 10.1038/ng1896

Smedley, D., Schubach, M., Jacobsen, J. O. B., Kohler, S., Zemojtel, T., Spielmann, M., et al. (2016). A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease. *Am. J. Hum. Genet.* 99, 595–606. doi: 10.1016/j.ajhg.2016.07.005

Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C., and Nobrega, M. A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* 21, 3255–3263. doi: 10.1093/hmg/dds165

Smith, M., and Flodman, P. L. (2018). Expanded insights into mechanisms of gene expression and disease related disruptions. *Front. Mol. Biosci.* 5:101. doi: 10.3389/fmolb.2018.00101

Soldner, F., Stelzer, Y., Shivalila, C. S., Abraham, B. J., Latourelle, J. C., Barrasa, M. I., et al. (2016). Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* 533, 95–99. doi: 10.1038/nature17939

Solomon, M. J., Larsen, P. L., and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53, 937–947. doi: 10.1016/s0092-8674(88)90469-2

Spielmann, M., Lupianez, D. G., and Mundlos, S. (2018). Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467. doi: 10.1038/s41576-018-0007-0

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., et al. (2012). Transcription regulation by distal enhancers: who's in the loop? *Transcription* 3, 181–186. doi: 10.4161/trns.20720

Stavropoulos, D. J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., et al. (2016). Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom. Med.* 1:15012. doi: 10.1038/npjgenmed.2015.12

Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., et al. (2016). Histone acetylome-wide association study of autism spectrum disorder. *Cell* 167, 1385.e11–1397.e11. doi: 10.1016/j.cell.2016.10.031

Suryamohan, K., and Halfon, M. S. (2015). Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdiscip. Rev. Dev. Biol.* 4, 59–84. doi: 10.1002/wdev.168

Thakore, P. I., D'Ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., et al. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149. doi: 10.1038/nmeth.3630

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11219–11223. doi: 10.1073/pnas.89.23.11219

Turner, T. N., Coe, B. P., Dickel, D. E., Hoekzema, K., Nelson, B. J., Zody, M. C., et al. (2017). Genomic patterns of de novo mutation in simplex autism. *Cell* 171, 710.e12–722.e112. doi: 10.1016/j.cell.2017.08.047

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., et al. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 471, 499–503. doi: 10.1038/nature09884

van Arensbergen, J., FitzPatrick, V. D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H. J., et al. (2017). Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* 35, 145–153. doi: 10.1038/nbt.3754

Van Essen, D. C., Donahue, C. J., and Glasser, M. F. (2018). Development and evolution of cerebral and cerebellar cortex. *Brain Behav. Evol.* 91, 158–169. doi: 10.1159/000489943

van Steensel, B., and Belmont, A. S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* 169, 780–791. doi: 10.1016/j.cell.2017.04.022

Vandervore, L. V., Schot, R., Hoogeboom, A. J. M., Lincke, C., de Coo, I. F., Lequin, M. H., et al. (2018). Mutated zinc finger protein of the cerebellum 1 leads to microcephaly, cortical malformation, callosal agenesis, cerebellar dysplasia, tethered cord and scoliosis. *Eur. J. Med. Genet.* 61, 783–789. doi: 10.1016/j.ejmg.2018.10.018

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T., Fernandez, N., et al. (2015). High-throughput and quantitative assessment of enhancer

activity in mammals by CapStarr-seq. *Nat. Commun.* 6:6905. doi: 10.1038/ncomms7905

Vermunt, M. W., Reinink, P., Korving, J., de Bruijn, E., Creyghton, P. M., Basak, O., et al. (2014). Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep.* 9, 767–779. doi: 10.1016/j.celrep.2014.09.023

Vermunt, M. W., Tan, S. C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., et al. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* 19, 494–503. doi: 10.1038/nn.4229

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., et al. (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566. doi: 10.1016/j.cell.2015.01.006

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858. doi: 10.1038/nature07730

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L. A. (2007). VISTA enhancer browser–a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–D92.

Visel, A., Prabhakar, S., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., et al. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40, 158–160. doi: 10.1038/ng.2007.55

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., et al. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* 44, 5615–5628. doi: 10.1093/nar/gkw159

Wallis, D. E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., et al. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nat. Genet.* 22, 196–198. doi: 10.1038/9718

Wang, C., Zhang, M. Q., and Zhang, Z. (2013). Computational identification of active enhancers in model organisms. *Genomics Proteomics Bioinformatics* 11, 142–150. doi: 10.1016/j.gpb.2013.04.002

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390–394. doi: 10.1038/nature10006

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362:eaat8464. doi: 10.1126/science.aat8464

Wang, X., He, L., Goggin, S. M., Saadat, A., Wang, L., Sinnott-Armstrong, N., et al. (2018). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.* 9:5380. doi: 10.1038/s41467-018-07746-1

Wang, Z., Zhang, Q., Zhang, W., Lin, J. R., Cai, Y., Mitra, J., et al. (2018). HEDD: human enhancer disease database. *Nucleic Acids Res.* 46, D113–D120. doi: 10.1093/nar/gkx988

Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934. doi: 10.1093/nar/gkr917

Weedon, M. N., Cebola, I., Patch, A. M., Flanagan, S. E., De Franco, E., Caswell, R., et al. (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* 46, 61–64. doi: 10.1038/ng.2826

Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., et al. (2016). SEA: a super-enhancer archive. *Nucleic Acids Res.* 44, D172–D179. doi: 10.1093/nar/gkv1243

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell* 171, 1573.e28–1588.e28. doi: 10.1016/j.cell.2017.11.008

Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., et al. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225. doi: 10.1038/nature10805

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319. doi: 10.1016/j.cell.2013.03.035

Wiszniewski, W., Gawlinski, P., Gambin, T., Bekiesinska-Figatowska, M., Obersztyn, E., Antczak-Marach, D., et al. (2018). Comprehensive genomic analysis of patients with disorders of cerebral cortical development. *Eur. J. Hum. Genet.* 26, 1121–1131. doi: 10.1038/s41431-018-0137-z

Won, H., de la Torre-Ubieta, L., Stein, J. L., Parikshak, N. N., Huang, J., Opland, C. K., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538, 523–527. doi: 10.1038/nature19847

Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G. C. (2017). Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell* 66, 285.e5–299.e5. doi: 10.1016/j.molcel.2017.03.007

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl. J. Med.* 369, 1502–1511. doi: 10.1056/NEJMoa1306555

Yao, P., Lin, P., Gokoolparsadh, A., Assareh, A., Thang, M. W., and Voineagu, I. (2015). Coexpression networks identify brain region-specific enhancer RNAs in the human brain. *Nat. Neurosci.* 18, 1168–1174. doi: 10.1038/nn.4063

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356:eaaj2239. doi: 10.1126/science.aaj2239

Young, R. A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940–954. doi: 10.1016/j.cell.2011.01.032

Zehnder, T., Benner, P., and Vingron, M. (2019). Predicting enhancers in mammalian genomes using supervised hidden Markov models. *BMC Bioinformatics* 20:157. doi: 10.1186/s12859-019-2708-6

Zeng, W., Min, X., and Jiang, R. (2019). EnDisease: a manually curated database for enhancer-disease associations. *Database* 2019:baz020. doi: 10.1093/database/baz020

Zeng, Y., Wang, G., Yang, E., Ji, G., Brinkmeyer-Langford, C. L., and Cai, J. J. (2015). Aberrant gene expression in humans. *PLoS Genet.* 11:e1004942. doi: 10.1371/journal.pgen.1004942

Zhang, G., Shi, J., Zhu, S., Lan, Y., Xu, L., Yuan, H., et al. (2018). DiseaseEnhancer: a resource of human disease-associated enhancer catalog. *Nucleic Acids Res.* 46, D78–D84. doi: 10.1093/nar/gkx920

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347. doi: 10.1038/ng1891

Zillhardt, J. L., Poirier, K., Broix, L., Lebrun, N., Elmorjani, A., Martinovic, J., et al. (2016). Mosaic parental germline mutations causing recurrent forms of malformations of cortical development. *Eur. J. Hum. Genet.* 24, 611–614. doi: 10.1038/ejhg.2015.192