Schizophrenia Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Understanding the genetic liability to schizophrenia through the neuroepigenome

John F. Fullard [a], Tobias B. Halene [a,d], Claudia Giambartolomei [a], Vahram Haroutunian [a,c,d], Schahram Akbarian [a,c], Panos Roussos [a,b,d,*]

[a] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[b] Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[c] Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[d] Mental Illness Research, Education, and Clinical Center (VISN 3), James J. Peters VA Medical Center, Bronx, NY, USA

ARTICLE INFO

Article history:
Received 24 November 2015
Received in revised form 14 January 2016
Accepted 18 January 2016
Available online xxxx

Keywords:
Postmortem
Brain
Gene expression
Chromatin
Epigenetics

ABSTRACT

The Psychiatric Genomics Consortium-Schizophrenia Workgroup (PGC-SCZ) recently identified 108 loci associated with increased risk for schizophrenia (SCZ). The vast majority of these variants reside within non-coding sequences of the genome and are predicted to exert their effects by affecting the mechanism of action of *cis* regulatory elements (CREs), such as promoters and enhancers. Although a number of large-scale collaborative efforts (e.g. ENCODE) have achieved a comprehensive mapping of CREs in human cell lines or tissue homogenates, it is becoming increasingly evident that many risk-associated variants are enriched for expression Quantitative Trait Loci (eQTLs) and CREs in specific tissues or cells. As such, data derived from previous research endeavors may not capture fully cell-type and/or region specific changes associated with brain diseases. Coupling recent technological advances in genomics with cell-type specific methodologies, we are presented with an unprecedented opportunity to better understand the genetics of normal brain development and function and, in turn, the molecular basis of neuropsychiatric disorders. In this review, we will outline ongoing efforts towards this goal and will discuss approaches with the potential to shed light on the mechanism(s) of action of cell-type specific *cis* regulatory elements and their putative roles in disease, with particular emphasis on understanding the manner in which the epigenome and CREs influence the etiology of SCZ.

1. Introduction

Recent years have witnessed renewed interest in studying genetic risk for SCZ, largely driven by advances in genomic technologies and a massive increase in sample sizes through the efforts of large consortia. The largest genome-wide association study (GWAS) analysis, conducted by the Psychiatric Genomics Consortium-Schizophrenia Workgroup (PGC-SCZ), comprises a sample set of 36,989 cases and 113,075 controls and identified 108 common variants that show statistical associations with SCZ (PGC-SCZ, 2014). Concurrently, the advent of next generation sequencing technologies has identified rare and de novo mutations conferring a high risk for the disease (Fromer et al., 2014; Purcell et al., 2014). In these exome sequencing studies, rare variants and de novo alleles were spread across a large number of SCZ genes, converging onto common, albeit broad, biological pathways, including genes involved in postsynaptic protein complexes and calcium signaling pathways.

Despite these efforts, a precise variant or target gene for SCZ has not been identified. There are several explanations for this, including unidentified rare variants with high penetrance or somatic mosaicism, and current methodological advances will be able to test these hypotheses in future studies. Here, we focus on findings that emerge from the largest and more recent GWAS in SCZ that set out to identify common risk loci. First, the variants associated with SCZ have small effect sizes that confer moderate risk but that, collectively, contribute to SCZ (i.e. SCZ is a polygenic disease with no single variant accounting for the entire risk). We will, therefore, need to adapt current methods to allow for multiple causal variants and genes to be studied simultaneously. Second, the variants most associated with SCZ often fall within large regions of high linkage disequilibrium (LD) containing multiple variants, any of which may be driving the association. As such, additional information is required to determine which variants are more likely to have functional effects. Third, and, from the perspective of this review, perhaps most importantly, the majority of identified variants are located outside of exons and, as such, do not change the protein coding

* Corresponding author at: Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Department of Genetics and Genomic Science and Institute for Multiscale Biology, One Gustave L. Levy Place, New York, NY 10029, USA.
E-mail address: Panagiotis.roussos@mssm.edu (P. Roussos).

http://dx.doi.org/10.1016/j.schres.2016.01.039
0920-9964/

sequence of genes, suggesting a substantial role for regulatory neuroepigenomic variation in the pathogenesis of SCZ.

In this review, we first describe the neuroepigenome and our current understanding of the ways in which it can be modified. We will then discuss its role in development, how it changes across the lifespan of an individual and its impact on disease. Finally, we provide a perspective for ongoing and future approaches to further our understanding of the neuroepigenome with an emphasis on applications utilizing frozen human postmortem brain tissue. While numerous epigenomic studies have focused on peripheral tissues and animal models, the aim of this review is to discuss studies that pertain to the human brain and, more specifically, to the neuroepigenome.

2. What is the neuroepigenome & why is it important?

Nuclei are between 2 and 10 μm in diameter yet contain approximately 2 m of DNA. In order to fit inside the nucleus, chromosomes are packaged in to a condensed mass consisting of genomic DNA and protein, termed chromatin. Chromatin falls into two broad categories: the more densely packed, transcriptionally repressed, heterochromatin and the less densely packed, transcriptionally active, euchromatin. The basic unit of chromatin is the nucleosome, which is composed of ~147 base pairs of genomic DNA wrapped in sequence around an octamer made up of the core histone proteins, H2A, H2B, H3 and H4. Chromatin consists of arrays of nucleosomes, connected by linker DNA and linker histones, such as histone H1. The combination of histones and DNA constitute the primary building blocks of the epigenome, which comprises a regulatory network that modulates chromatin structure and, ultimately, the accessibility of specific DNA sequences to other factors, such as the molecular machinery involved in transcription. The neuroepigenome refers, specifically, to the epigenetic mechanisms (including those that modify chromatin) that contribute to brain development and function.

Importantly, the epigenome is not static and can be modified, providing a temporal dimension to gene expression and, ultimately, to cell function. Histones can undergo an array of post-translational modifications, including, but not limited to, mono-, di- and tri-methylation, acetylation and serine phosphorylation and these modifications can have a variety of impacts on genome structure and function. For example: Histone H3 methylation at lysines 4, 9, and 27, are marks associated, respectively, with active transcription, heterochromatin formation, and transcriptional repression (Li and Reinberg, 2011). Histone H3 trimethylation at lysines 27 and 9 are associated with polycomb repression and heterochromatin silencing, respectively, whereas acetylation at either residue is a characteristic of active enhancers and regulatory sequences (Ernst et al., 2011; Pasini et al., 2010; Zhu et al., 2013). Together, these provide illustrative examples of how the same residue can have a diametrically opposed influence on gene expression depending on how it is modified. For review of the different histone modifications and their impacts see (Jakovcevski and Akbarian, 2012) and references therein.

DNA sequence can also be chemically modified, leading to a variety of effects on the activity of a given gene. An example is DNA methylation, which typically results in the suppression of gene expression e.g. methylation of CpG dinucleotide islands, which are usually found in proximity to (or within) promoters. Although some genes become hypermethylated over time, there is a trend towards global loss of DNA methylation (hypomethylation) throughout life (Gonzalo, 2010), a trend that may be a contributory factor in age related neurodegenerative disorders (Akbarian et al., 2013; Johnson et al., 2012). The importance of DNA methylation in the regulation of gene expression is further demonstrated by the fact that hypermethylation and hypomethylation, relative to normal tissue, have been implicated in a variety of human cancers where, typically, there is hypermethylation of tumor suppressor genes and hypomethylation of oncogenes (Gokul and Khosla, 2013).

In addition to methylation, cytosine residues in DNA are also susceptible to modification through hydroxymethylation, in which hydrogen 5 of cytosine is replaced by a hydroxymethyl group (5hmC). Whereas methylation occurs in promoters and is associated with lower gene expression, 5hmC, conversely, affects intragenic regions and, although it's precise role is unknown, is associated with elevated gene expression (Kato and Iwamoto, 2014; Nestor et al., 2012). During early postnatal development the neuronal genome accumulates uniquely high levels of non-CpG methylation and 5hmC (Kinde et al., 2015). Whole genome analysis has revealed that the content of 5hmC is particularly high in the brain, where it constitutes the primary modification of many enhancers and regions actively undergoing transcription (Wen et al., 2014). In addition, 5hmC peaks are found at the 5′ splice sites of exon–intron boundaries where it is thought to influence splicing and gene expression (Khare et al., 2012; Wen et al., 2014). Due to the presence of high levels of 5hmC in the brain, and in neurons, hydroxymethylation has been speculated to play a pivotal role in controlling neuronal differentiation, neural plasticity and brain functions (Wen and Tang, 2014). Genomic DNA from mouse adult brain contains high levels of 5-methylcytosine (5mC) in a non-CG context compared with other tissues (Xie et al., 2012). High levels of 5hmC have also been observed in humans, where the content of 5hmC between normal tissues appears to be highly variable, is associated with the body of transcribed genes, and is directly proportional to levels of transcription of those genes (Nestor et al., 2012).

Epigenetic modification of DNA has also been identified as a key mechanism for environmental regulation of gene expression (Jirtle and Skinner, 2007) and environmental factors can trigger lifelong molecular changes to the epigenome with a profound impact on the health and, perhaps, behavior of the organism later in life (Klengel and Binder, 2015). Although the majority of epigenetic research has focused on modifications of histones and DNA, RNA is also extensively modified (Satterlee et al., 2014). RNA methylation has been observed in both prokaryotic and eukaryotic organisms and in numerous types of RNA molecules, including mRNA, tRNA, and non-coding RNA (Wang and He, 2014). Although the function of RNA methylation remains unclear, it has been proposed to play roles in, among others, post transcriptional regulation of gene expression (Yue et al., 2015) and RNA biogenesis and splicing (Alarcon et al., 2015a, 2015b; Dominissini et al., 2012).

3. The genome in 3-dimensions

Importantly, DNA methylation and its variants (hydroxymethylation, etc.), multiple post-translational histone modifications and other types of epigenetic regulation, fail to fully characterize the epigenome and localized chromatin architecture at any given genomic locus. This is because the chromosomal arrangements in the interphase nucleus are not random and it is now generally accepted that genetic information is not only encoded in nucleotide sequence but also in the dynamic 3-dimensional organization of the genome. For example, loci at sites of active gene expression are more likely to be clustered together and positioned towards a central position within the nucleus, while heterochromatin and silenced loci are located towards the nuclear periphery (Cremer and Cremer, 2001; Duan et al., 2010). Thus, the spatial position of genomic sequences provides a critically important layer of regulation in eukaryotic cells. Furthermore, chromosomal loopings are associated with transcriptional regulation by permitting direct interaction between distal DNA elements, often separated by many kilobases along the linear genome (Gaszner and Felsenfeld, 2006; Sanyal et al., 2011; Wood et al., 2010).

Some interactions influence fundamental biological processes such as imprinting (Zhang et al., 2014) and dysregulated higher order chromatin is also thought to contribute to disease, for example Cornelia de Lange Syndrome (CdLS). With an estimated incidence of 1:10–30,000 live births, CdLS is among the more frequent genetic disorders (source http://ghr.nlm.nih.gov). CdLS is associated with a range of

Please cite this article as: Fullard, J.F., et al., Understanding the genetic liability to schizophrenia through the neuroepigenome, Schizophr. Res. (2016), http://dx.doi.org/10.1016/j.schres.2016.01.039