Schizophrenia Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Integration of DNA sequence and DNA methylation changes in monozygotic twin pairs discordant for schizophrenia

C.A. Castellani [a,1], M.G. Melka [a,1], J.L. Gui [a], R.L. O'Reilly [b], S.M. Singh [a,b,*]

[a] Department of Biology, The University of Western Ontario, London, Ontario N6A 5B7, Canada
[b] Department of Psychiatry, The University of Western Ontario, London, Ontario N6A 5B7, Canada

ARTICLE INFO

Article history:
Received 29 May 2015
Received in revised form 18 September 2015
Accepted 21 September 2015
Available online xxxx

Keywords:
Integrative genomics
Monozygotic twins
Discordance
Schizophrenia
Threshold model
Psychosis
Genome sequencing
Methylation array

ABSTRACT

Schizophrenia is a complex mental disorder with high heritability (80%), extensive genetic heterogeneity, environmental contributions and only 50% concordance in discordant monozygotic (MZ) twins. Discordant MZ twins provide an exceptional opportunity to assess patient specific genome-wide genetic and epigenetic changes that may account for the disease phenotype. A combined analysis of genetic and epigenetic changes on the same twin pairs is expected to provide a more effective approach for two reasons. First, it is now possible to generate relatively reliable complete genome sequences as well as promoter methylation states on an individual level and second, the unaffected twin that originated from the same zygote provides a near perfect genetic match for contrast and comparison. This report deals with the combined analysis of DNA sequence data and methylation data on two pairs of discordant MZ twins that have been clinically followed for over 20 years. Results on Family 1 show that 58 genes differ in DNA sequence as well as promoter methylation in a schizophrenia-affected twin as compared to her healthy co-twin. The corresponding number for family 2 was 13. The two lists are over represented by neuronal genes and include a number of known schizophrenia candidate genes and drug targets. The results argue that changes in multiple genes via co-localized genetic and epigenetic alteration contribute to a liability threshold that is necessary for development of schizophrenia. This novel hypothesis, although logical, remains to be validated.

1. Introduction

Schizophrenia (OMIM: 181500) is a pervasive and life-altering mental health disorder that presents in early adulthood and affects 1% of the population worldwide (Cardno and Gottesman, 2000). What causes schizophrenia is not known, however genetic and epigenetic determinants must contribute to the development of this disease (Hindorff et al., 2009). Although the molecular results on schizophrenia are accumulating (Dempster et al., 2011; Liao et al., 2015; Wockner et al., 2014), the search for schizophrenia causing genes using traditional approaches has remained challenging. Also, the low reproducibility of results published to date highlight the extensive heterogeneity associated with this complex disease (Ripke et al., 2014). The disappointing success in search of the causations of schizophrenia may be due to the limited scope of current lines of investigation that evaluate genetic, epigenetic or environmental contributors individually. They are not suited to fully unravel the complex interaction between DNA sequence as well as their epigenetic modifications.

Integration across multiple "omic" datasets is now becoming an informative approach for unraveling complex disorders. In particular, evidence is emerging that epigenetic states may serve to directly mediate the relationship between certain genetic polymorphisms and phenotype (McVicker et al., 2013). The proportion of inter-individual variation in methylomes that may be driven by genotype as well as environment is currently not known. What is known is that DNA methylation plays a critical role in the regulation of gene expression (Razin and Kantor, 2005). Also, individual differences in DNA methylation are correlated with DNA sequence polymorphism, labeled as methylation quantitative trait loci (methQTLs). The methQTLs may represent the interaction between genetic variation and environmental influences (Teh et al., 2014). As might be expected, genetic variants at CpG sites are able to alter methylation status at that site (Bell et al., 2011; Gertz et al., 2011; Gibbs et al., 2010; Hellman and Chess, 2010). In addition, genotype at one Single Nucleotide Polymorphism (SNP; a single nucleotide change occurring in ≥1% frequency in the population) may affect methylation status at multiple neighboring CpGs (Eckhardt et al., 2006; Zhi et al., 2013). Accordingly, in humans, inter-individual variation in DNA methylation could be a consequence of nucleotide polymorphism influencing methylation of cytosines, both directly and indirectly (Liu et al., 2013). Here, the genotype alone can explain ~25% of variation in methylation and the best explanation for

* Corresponding author at: Molecular Genetics Unit, Western Science Center, Department of Biology, The University of Western Ontario, London, Ontario N6A 5B7, Canada.
E-mail address: ssingh@uwo.ca (S.M. Singh).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.schres.2015.09.021
0920-9964/

the remaining 75% is the interaction of genotype with different environments (Teh et al., 2014).

DNA variations leading to both methylation and expression variation at the same locus appear relatively rare but are reproducible findings (Wagner et al., 2014). For example, Patel et al. (2013) show that candidate SNP and CpG loci with marginal associations in GWAS and EWAS are correlated with type 2 diabetes (Patel et al., 2013). Also, an SNP within the IL4R gene combined with methylation at a CpG site within the same gene is predictive of childhood asthma risk (Soto-Ramírez et al., 2013). Further, widespread relationships exist between DNA methylation and gene expression especially in developmentally significant genes, including HOX clusters (Wagner et al., 2014). Additionally, recent studies describe evidence for gene by environment interactions on DNA methylation (Yousefi et al., 2013). This evidence is consistent with the emerging view that genotype can determine the degree of environmentally induced phenotypic plasticity via allele specific variation.

In humans, where manipulative experiments are out of the question, a novel approach is the use of monozygotic twins. Results on twins have shown that both genetic and environmental factors may affect twin differences in DNA methylation (Castellani et al., 2015; Kaminsky et al., 2009; Ollikainen et al., 2010). Differences in methylation are apparent early in newborn twins (Ollikainen et al., 2010), increase with age (Fraga et al., 2005) and show greater discordance in dizygotic than monozygotic twins (Ollikainen et al., 2010). The assessment of high-resolution genome sequences together with genome-wide methylation in monozygotic twins may offer a better understanding of the genetic and epigenetic contributors to schizophrenia. Here, we report on the joint analysis of sequence variations including structural changes (Complete genome sequences), and genome-wide DNA methylation (NimbleGen Human DNA Methylation) in two families with monozygotic twins discordant for schizophrenia. Monozygotic twins showed differences in DNA sequence as well as DNA methylation and some of these differences were co-localized. The results allow consideration of the interaction of DNA methylation and sequence changes in schizophrenia.

2. Methods

The experiments performed received approval by the University of Western Ontario's Committee on research involving human subjects. The families in the study provided written informed consent for participation. Capacity for consent was ensured using three measures 1) Schizophrenic patients gave consent only during a “normal” phase (no psychosis present), 2) Both twins of the twin pair were present and gave consent at the same time (the normal twin and their Schizophrenic sibling), and, 3) If Dr. O'Reilly felt that the capacity to consent was compromised, the patients were not included in our study. They were interviewed and clinically assessed by a single senior Psychiatrist (R. O'Reilly) using the SCID-I and SCID-II. A second senior psychiatrist independently reviewed videotapes of the structured interviews of the twins and confirmed the diagnoses.

The two discordant twin pairs in this study were female monozygotic twins. The twins from Family 1 were Afro-American females aged 53 and the twins from Family 2 were Caucasian females aged 43. The affected member of twin pair 1 was diagnosed with schizophrenia at age 22 and the affected member of twin pair 2 was diagnosed with schizoaffective disorder at age 27. The twins and their parents contributed whole blood samples for DNA isolation. DNA was extracted from whole blood using the 5 Prime Perfect Pure DNA Blood Kit (Gaithersburg, MD), following the manufacturer's protocol. Zygosity was confirmed by Affymetrix 6.0 microarray and specifically using the Affymetrix Genotyping Console 4.0 concordance feature.

2.1. Complete genomics whole genome sequencing

The genome sequence of the twin pairs was generated at Complete Genomics Inc. (Mountain View, CA) in the form of ~2 billion overlapping 70-base nucleotide sequences. They allowed reconstruction of six individual genomes. The sequences met the criteria of high accuracy (99.999%) and were considered suitable for identification of rare variants including somatic mutations as described by Drmanac et al. (2010). These variants included single nucleotide variants (SNVs; single nucleotide changes that differ from reference), indels and block substitutions as well as larger variants classified as Copy Number Variants (CNVs) and Structural Variants (SVs) that were called in comparison to reference sequence (NCBI Build 37). CNVs were called based on a read depth or depth-of-coverage algorithm provided by Complete Genomics. Sequence coverage was averaged and then GC bias was corrected for over a fixed window (2 kb) and normalized relative to a set of standard (CG 45 genome reference) genomes sequenced by Complete Genomics. A Hidden Markov Model (HMM) was used to determine the integer copy number state (0–10). SVs were detected by identifying discordant mate pair mappings found during the assembly process. Mate pair mappings where each arm maps to the reference genome but with either an unexpected length between them or an anomalous orientation were subjected to local *de novo* assembly to refine junction breakpoints at single base pair resolution. The in depth methods have been further discussed elsewhere (Castellani et al., 2015, unpublished data). Given the potential for false positives in genome sequencing, a stringent read depth of 50 and a call quality of 100 (calculated by complete genomics and based on a phred scale) were chosen as parameters for initial variant filtering. Variants were annotated with overlapping genes, cytoband, gene region (Exonic, Intronic, Promoter, 5′UTR, 3′UTR, Splice Site), translational impact (if applicable), SIFT function prediction (if applicable), SIFT score (if applicable), dbSNP ID (if applicable) and frequency in the 1000 Genomes as well as frequency in the Complete Genomics Public Genome dataset. This dataset allowed for identification of a variety of DNA sequence differences within the twin pairs.

2.2. NimbleGen human DNA methylation promoter plus CpG island 720 k array

We assessed genome-wide methylation in the two twin pairs and one set of parents using the NimbleGen Methylation Promoter Microarray**.** The NimbleGen Human DNA Methylation 3x720k CpG Island Plus RefSeq Promoter Microarray is a multiplex slide with 3 identical arrays per slide. Each Roche Nimblegen Inc. (Madison, MI) array covers 27,728 annotated CpG islands as well as 22,532 promoters of the RefSeq genes derived from the UCSC RefFlat files. Median-centering, quantile normalization, and linear smoothing was performed by Bioconductor packages *Ringo*, *limma*, and *MEDME* at ArrayStar. The detailed methods of this analysis are described elsewhere (Castellani et al., 2015). Briefly, the pair files were analyzed with the tiling workflow in Partek Genomics Suite® version 6.6 (St. Louis, Missouri). Nimblegen scan pair files (635 nm and 532 nm) for each sample were annotated against hg18 and enriched regions were detected using a two-way ANOVA between an affected twin and the respective unaffected co-twin. The enriched regions settings were set at a minimum p-value of 0.001 and the number of probes to call a region was set at a minimum of 4. Genes that overlapped differentially methylated regions (up to 5000 bp upstream or 3000 bp downstream) were annotated. The genome regions that showed differences in methylation between co-twins were selected for further analysis.

2.3. Datasets and analysis

This analysis was based on within-pair differences between two pairs of monozygotic twins discordant for schizophrenia. It included complete genome sequence features as well as genome-wide DNA promoter methylation. The results identified gene specific sequence differences that are associated with corresponding differences in promoter DNA methylation. Differences between the affected and healthy MZ twins in twin pair 1 included sequence variations (1896) and promoter

Please cite this article as: Castellani, C.A., et al., Integration of DNA sequence and DNA methylation changes in monozygotic twin pairs discordant for schizophrenia, Schizophr. Res. (2015), http://dx.doi.org/10.1016/j.schres.2015.09.021