



Whole-genome single nucleotide variant distribution on genomic regions and its relationship to major depression



Chenglong Yu^{a,b,*}, Bernhard T. Baune^c, Julio Licinio^{a,b}, Ma-Li Wong^{a,b}

^a Mind and Brain Theme, South Australian Health and Medical Research Institute, North Terrace, Adelaide, SA 5000, Australia

^b School of Medicine, Flinders University, Bedford Park, SA 5042, Australia

^c Discipline of Psychiatry, School of Medicine, University of Adelaide, Adelaide, SA 5005, Australia

ARTICLE INFO

Keywords:

Major depressive disorder
Whole-genome sequencing
Mexican-American
Kullback-Leibler divergence
Cluster analysis

ABSTRACT

Recent advances in DNA technologies have provided unprecedented opportunities for biological and medical research. In contrast to current popular genotyping platforms which identify specific variations, whole-genome sequencing (WGS) allows for the detection of all private mutations within an individual. Major depressive disorder (MDD) is a chronic condition with enormous medical, social and economic impacts. Genetic analysis, by identifying risk variants and thereby increasing our understanding of how MDD arises, could lead to improved prevention and the development of new and more effective treatments. Here we investigated the distributions of whole-genome single nucleotide variants (SNVs) on 12 different genomic regions for 25 human subjects using the symmetrised Kullback-Leibler divergence to measure the similarity between their SNV distributions. We performed cluster analysis for MDD patients and ethnically matched healthy controls. The results showed that Mexican-American controls grouped closer; in contrast depressed Mexican-American participants grouped away from their ethnically matched controls. This implies that whole-genome SNV distribution on the genomic regions may be related to major depression.

1. Introduction

Recent advances in high-throughput genomic technologies provide unprecedented opportunities for medical research. Patient care is moving towards genotyping/sequencing-based precision medicine. SNP (single nucleotide polymorphism) genotyping is currently the most popular technique for genome-wide association studies (GWAS) which identify significantly associated variations with a trait or disease. However, genome-wide association studies have attracted much criticism, mainly because of their assumption that common variations play a large role in elucidating the heritable variation of common disease (Visscher et al., 2012). The development of new and cheaper whole-genome sequencing (WGS) technology is allowing scientists to search for new channels to provide personalized treatment (Collins and McKusick, 2001; Hamburg and Collins, 2010). WGS can identify single nucleotide variants (SNVs), which are private genetic variants, and determine all the genetic variants within each person. As WGS costs are forecasted to drop further, researchers may have the opportunity to investigate the significance of SNVs, which involve more individual characteristics.

Major depressive disorder (MDD), as a serious chronic condition, is

a main contributor to global disease burden and produces considerable morbidity and mortality (Kessler et al., 1994, 2005; Lopez and Murray, 1998; Wong and Licinio, 2001, 2004). Genetic factors play important roles in the susceptibility to major depression, as indicated by family, twin, and adoption studies (Sullivan et al., 2000; Lohoff, 2010). The heritability of MDD is estimated to range between 40–70% (Lesch, 2004); a meta-analysis of twin studies has estimated a 37% heritability for MDD (Sullivan et al., 2000). Those data support a complex etiology that includes both genetic and environmental factors (Demirkan et al., 2011). Identifying specific genetic factors for the development of MDD remains a challenge. GWAS have identified genetic risk variants for many psychiatric illnesses; however, only recently replicated genome-wide significant loci have been identified for MDD (Amin et al., 2016; CONVERGE Consortium, 2015; Hyde et al., 2016; Peterson et al., 2017). Therefore, the underlying fundamental biology of MDD is largely unknown and much work still needs to be done to fully elucidate the genetic factors that confer susceptibility to this condition (Flint and Kendler, 2014; Sullivan et al., 2012; Yu et al., 2017).

We hypothesized that if two individuals have similar SNV distributions on the genomic regions, then those two individuals would have similar phenotypes, such as traits or diseases. In this study, we

* Corresponding author at: Mind and Brain Theme, South Australian Health and Medical Research Institute, PO Box 11060, Adelaide, SA 5001, Australia.
E-mail address: chenglong.yu@sahmri.com (C. Yu).

investigated the whole-genome SNV distributions on 12 different genomic regions for 25 human subjects including MDD cases and ethnically-matched controls. To accurately compare two SNV distributions, an effective similarity measure may be applied. In our previous work (Yu et al., 2011), we used the Kullback-Leibler divergence (also called relative entropy) to obtain a symmetrised measure and successfully studied the phylogeny of various DNA sequences based on their probability distributions. The Kullback-Leibler divergence, which is a powerful tool to measure the difference between two probability distributions, has been widely used in many fields, such as bioinformatics (Kaitchenko, 2004; Liu et al., 2014; McClendon et al., 2012; Hoang et al., 2015), genomics (Huang et al., 2014; Lin, 2015; Siegel et al., 2015), and machine learning (Lee and Lee, 2006; Rubinstein and Kroese, 2013). By means of this measure, we performed cluster analysis for those subjects including MDD cases and controls. Our results suggest that whole-genome SNV distribution on the genomic regions may have some relationship with MDD.

2. Methods

2.1. The Mexican-American sample

In our recent work (Wong et al., 2016) we have investigated whole-exome genotyping of a Los Angeles Mexican-American cohort of 203 MDD patients and 196 healthy controls, aged 19–65 years. Participants in this cohort had three or more grandparents born in Mexico. Participants with MDD were diagnosed using the Structured Clinical Interview for Diagnostic (SCID) and Statistical Manual of mental disorders, and the DSM-IV (Diagnostic and Statistical Manual IV edition) diagnostic criteria for current, unipolar major depressive episode with a HAM-D21 (21-Item Hamilton Depression Rating Scale) score of 18 or greater with item number 1 (depressed mood) rated 2 or greater; they participated in pharmacogenetic studies of antidepressant treatment. MDD was defined as 5 out of 9 criteria in the SCID. A Hamilton score > 18 was used to characterize entry criteria for the pharmacogenetic study but not to diagnose MDD. Raters were experienced bilingual clinical personnel (nurses, social workers and physicians) using Spanish or English versions of questionnaires and rating scales, and diagnosis was confirmed by a research psychiatrist. Controls were in general good health but were not screened for medical or psychiatric illnesses; they were age- and gender-matched Mexican-American individuals recruited from the same community in Los Angeles. Participants provided written informed consent, and detailed demographic, epidemiological, and clinical descriptions were previously described (Dong et al., 2009; Wong et al., 2012, 2014). The study was registered in ClinicalTrials.gov (NCT00265291), and approved by the Institutional Review Boards of the University of California Los Angeles and University of Miami, USA, and by the Human Research Ethics Committees of the Australian National University and Bellbery Ltd, Australia. In this work we obtained complete WGS data for a group of 15 participants selected from the cohort, 10 MDD patients and 5 controls. WGS was performed using Illumina HiSeq 2000 (BGI-Shenzhen, Shenzhen, Guangdong, China). We have confirmed that in the cohort there was no family or population structure among all those individuals and no blood relationship among the 15 selected participants.

2.2. The European-Ancestry sample

We also included WGS data from a group of 10 Australians of European-Ancestry as a comparison group. Those 10 participants, including 5 MDD cases and 5 healthy controls, gave written informed consent and were recruited under the protocol of Characteristics of the Cognitive Function and Mood Study (CoFaM-Study) conducted by the Discipline of Psychiatry, University of Adelaide, South Australia, Australia (Baune and Air, 2016). The MDD diagnoses were made using

Table 1

The descriptive statistics for MDD cases and controls. μ : average value; SD: standard deviation; HAM-D: Hamilton depression rating scale; F: female; M: male.

Group	Status	Number	Sex	Age μ	Age SD	HAM-D μ	HAM-D SD
Mexican-American	MDD	10	10 F	38.8	8.15	23	4.08
	Control	5	5 F	39.6	7.36	N/A	N/A
Australian of European-Ancestry	MDD	5	1 M, 4 F	25	9.82	8.6	3.26
	Control	5	3 M, 2 F	21.2	4.49	2	1.26

Mini International Neuropsychiatric Interview (MINI) for both the cases and the healthy controls. We also assessed HAM-D21 scores for those participants. This study was under approved by Human Research Ethics Committees at the University of Adelaide and Flinders University, South Australia, Australia. WGS was performed using HiSeq X (Garvan Institute, Sydney, New South Wales, Australia).

Table 1 summarizes descriptive statistics of gender, age and HAM-D scores for the 15 Mexican-American subjects and the 10 Australian subjects. The Australians of European-Ancestry sample includes both acutely depressed and remitted depressed patients who have by definition a lower depression severity compared to acutely depressed patients. Thus the HAM-D scores for the Australian sample may be lower on average than the Mexican-American Sample.

2.3. WGS data analysis

We performed the variant calling analysis of high quality paired-end sequencing reads using a previously described pipeline (Wong et al., 2016). Briefly, we aligned the reads of each participant to the human reference genome (hg19, Genome Reference Consortium GRCh37) using Burrows-Wheeler Aligner (Li and Durbin, 2009) to get SAM (sequence alignment/map) format files. SAM files were converted to the BAM (binary version of a SAM file) format files using SAMtools (Li et al., 2009). BAM files were then merged into one BAM file, and the mpileup command in SAMtools was applied to calculate the likelihood of data given each possible genotype, and store the likelihoods in a binary file. The output was supplied to SAMtools/BCFtools (Li, 2011) which created the SNV and INDEL (small insertions and deletions) calling to produce VCF (variant call format) files. Finally, the popular ANNOVAR software (Wang et al., 2010) was applied to annotate SNV information through 12 different genomic regions. WGS data analysis were performed using high-performance computers in eResearch South Australia (www.ersa.edu.au).

2.4. SNV count distribution on genomic regions

After variant annotation, we can categorize the SNV count distribution on 12 genomic regions into: exonic synonyms, exonic nonsynonyms, intronic, intergenic, splicing, UTR3, UTR5, downstream, upstream, ncRNA exonic, ncRNA intronic, and ncRNA splicing. According to ANNOVAR results, all the SNVs are located on those 12 genomic regions. There is no other genomic region to be considered. Thus each subject corresponds to a SNV distribution on the 12 genomic regions. In order to obtain an accurate comparison between two SNV distributions, a proper similarity measure should be carefully employed to test our hypothesis in this study that if two individuals have similar SNV distributions, then those two individuals would have similar phenotypes, such as traits or diseases.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات