Population Phenotypes of Neuropsychiatric Copy Number Variants

Anne Skelton

UMC Utrecht, Netherlands

EuroScholars

Fall 2014

In collaboration with: Dr. Marco Boks, M.D., Ph.D.

Brain Center Rudolf Magnus

Introduction

- Copy Number Variants (CNVs): DNA segments (1 kilobase or larger) present at variable copy numbers as a result of duplication, deletion, or complex rearrangment.
- Genomic Rearrangement results in changes in gene expression[1,2,3,4]:
  - Gene Dosage Effects: Copy numbers influencing expression of recessive/dominant alleles
  - Gene Fusion: Splicing of genes together as a result of intronic deletions
  - Excision or Duplication of Functional Domains
  - Disruption of Coding Sequences
  - Position Effects: Position and coding of a gene has changed
  - Perturbation of Long-Range Gene Regulation
- CNVs can alter human physiological function and produce phenotypic variations leading to disease associations and susceptibility[1,4,5,6]:
  - Duplication/Deletion Disorders
  - Unmasking of Recessive Alleles
  - Functional Polymorphisms
- CNVs have been found to be linked to several "complex" diseases which may be the result of multiple genetic and environmental factors[7].
  - Including Psoriasis, Crohn's Disease, and glomerulonephpritis
- Psychiatric Disorders, classified as a complex disease, are being researched in association to CNVs
  - CNVs may not be fully penetrant in the genome and only contribute a small responsibility individually but in concert with other CNVs, genetic factors, and environmental factors, they may provide part of the etiology for these diseases.[7]

Introduction [cont.]

- Genome Wide Association Studies (GWAS): Studies that look at entire genomes of thousands of participants in hopes of finding genetic etiologies of diseases
  - Common Disease-Common Variant Hypothesis: Common complex traits and diseases are largely caused by common variants of the genome with each having a small to moderate effect size[8,9]
  - Common Disease-Rare Variant Hypothesis: Common complex traits and diseases are largely caused by rare variants of the genome[9]
    - Large number of CNV-trait associations have shown to be more consistent with this hypothesis but this cannot explain phenotypic variability of diseases[10]
  - Results
    - Schizophrenia[11], Autism[12], Bipolar Disorder[13], Major Depressive Disorder[14], Personality Traits[15] have all been found to be associated with various genetic loci
    - Similar CNVs and affected loci across discorders suggests a common underlying mechanism or similar mutations can increase risk for a range of psychopathology[5]
- Phenotypic Variability in Psychiatric Disorders
  - Genomic Causation as a result of GWAS
    - Multiple causative alleles combine and supplement one another and interact with other genomic or environmental factors[16]
    - Altering of Specific Genes/Genomic Segements: changing multiple genes at one locus or a whole genomic segment (as with large variants) may cause overlapping expression or variable penetrance of expression[17]
    - Larger CNVs (more than 100 kb) or multiple CNVs affect larger portions of the genome[6]
      - Large CNVs are most commonly studied in association with complex neurological and psychiatric diseases

Current Study

- Previous GWAS have followed a "Case-Control"design
  - One group of clinically diagnosed patients compared to a group unaffected by psychiatric illness
    - The control group is theoretically unaffected by psychiatric illness; they or a family member have not been diagnosed with an illness or fit the clinical desciptions and cutoffs
  - Does not take into account varying degree of psychiatric symptoms
    - Some degrees of psychiatric symptoms classify as clinical disorders and some occur periodically, to a lesser degree, or are preemptive of a disorder in the future
- General population sample of Dutch participants to undergo genotype and phenotype testing
  - Some participants may have been diagnosed with psychiatric disorders and some of which may have not, but that information is unreported
  - Can compare CNV genotyping data to psychiatric traits and phenotype data determined by various self-report questionnaires rather than classifications of disorders
    - Phenotype data allows a gradient of symptoms to be analyzed
- Hypothesis: exhibiting particular CNVs or an altered frequency of CNVs will predispose participants to display certain psychiatric phenotypes

Participants

- Recruited Via the CannabisQuest Study[18] and GROUP Study[19]
  - CannabisQuest Study: began in 2006 and assessed Dutch-speaking adolescents and young adults online regarding cannabis use and frequency of psychotic experiences
    - 918 of 21,838 individuals where recruited for further genetic assessment
      - Chosen to be representative of four criteria: high cannabis use/high psychotic experience, high use/low experience, low use/high experience, low use/low experience
  - Joined by 339 healthy control individuals from GROUP (Genetic Risk and Outcome of Psychosis) Study
- 1259 Total Individuals
  - Found to have an absence of a history of psychotic disorder or first/second-degree family member with psychotic disorder via the Family Interview for Genetic Studies
  - Underwent comprehensive psychological assessment and blood sampling for genotype testing

Self-Report Surveys and Phenotypes

Community Assessment of Psychotic Experiences (CAPE)[20]	Measures positive, negative, and depressive symptoms of psychosis	Increased Frequency of psychotic experienced and distress associated with experiences
Symptom Checklist-90 Revised (SCL-90-R)[21]	Measures nine symptoms of psychopathology and resulting distress: depression, anxiety, somatization, interpersonal sensitivity, hostility, sleep, insufficiency, psychoticism and other remaining symptoms	Broken down by symptom with increased scores meaning increased frequency of symptom
Beck's Depression Inventory (BDI)[22]	Measures both the attitudes and frequency of depressive symptoms	Increased scores result from increased frequency of depressive symptoms
NEO Five-Factor Personality Inventory[23]	The Gold-Standard of personality assessment measures through a five-factor model: neuroticism, extraversion, agreeableness, conscientiousness, and openness to experience	Phenotypes are measured by the intensity of the dichotomy of each personality trait; low scores indicate one end of the dichotomy, high scores the other
State-Trait Anxiety Inventory (STAI)[24]	Measures feelings of anxiety as a state of being and as a personality trait in addition to the frequency of anxious feelings	Increased scores indicate higher frequencies of anxiety and indicate if anxiety is a one-time event versus a continuous state

Table 1. Phenotype Characteristics of Test Sample

Parameter	Combined
Total, N	N=1259
Female Sex, %	52.50%
Age, mean (range), year	20,52 (16-40)
CAPE, mean (range)	67,34 (43-146)
CAPE Positive, mean (range)	27,95 (20-62)
CAPE Negative, mean (range)	25,62 (15-58)
CAPE Depressive, mean (range)	13,77 (8-30)
STAI State, mean (range)	32,01 (20-72)
STAI Trait, mean (range)	34,45 (20-73)
BDI Total, mean (range)	5,34 (0-34)
SCL Anxiety Ever, mean (range)	16,88 (10-50)
SCL Anxiety Weekly, mean (range)	12,76 (10-43)
SCL Depression Ever, mean (range)	29,90 (16-80)
SCL Depression Weekly, mean (range)	22,22 (16-63)
SCL Somatization Ever, mean (range)	21,36 (12-60)
SCL Somatization Weekly, mean (range)	16,09 (12-41)
SCL Sensitivity Ever, mean (range)	31,23 (18-86)
SCL Sensitivity Weekly, mean (range)	24,11 (18-64)
SCL Hostility Ever, mean (range)	9,39 (6-30)
SCL Hostility Weekly, mean (range)	7,31 (6-27)
SCL Sleep Ever, mean (range)	5,64 (3-15)
SCL Sleep Weekly, mean (range)	4,50 (3-15)
SCL Ago Ever, mean (range)	9,41 (7-34)
SCL Ago Weekly, mean (range)	7,77 (7-25)
SCL Insuf Ever, mean (range)	16,77 (9-43)
SCL Insuf Weekly, mean (range)	13,25 (9-37)
SCL Other Ever, mean (range)	13,99 (9-43)
SCL Other Weekly, mean (range)	11,35 (9-36)
NEO Neuroticism, mean (range)	131,01 (68-204)
NEO Extroversion, mean (range)	41,75 (13-58)
NEO Agreeableness, mean (range)	43,71 (22-58)
NEO Conscientiousness, mean (range)	41,46 (23-59)
NEO Openness, mean (range)	39,91 (24-57)

Initial Results showing the population break down of the test sample (age, gender) and the corresponding average of scores on various self-report surveys. Note the breakdown of each self-report survey into its component parts.

Genotype Data

1. Genotypes and nucleotide intensities were generated from blood sample arrays on three different Illumina platforms (Illumina HumanOmniExpress, Illumina Human610-QuadBeadchip, Illumina HumanHap500Array).

2. Each platform was then subjected to separate quality control procedures via PLINKV1.07[25] to exclude genomes . Quality control criteria included missingness, gender errors, and hetero/homozygosity.

3. Linkage disequilibrium pruning was used to select the most informative single nucleotide polymorphisms (SNPs) ($R^2 < 0.2$) as a subsequent quality control step. SNPs were further filtered by genome inclusion (<2% missing) and Hardy Weinberg (inclusion $P > 1e^{-5}$).

4. The three datasets were then separately merged with HapMap Phase 3 individuals to verify ethnicity before finally merging the three datasets together via GenomeStudio.

5. Data was then integrated from each platform into one set using GenomeStudio software. Signal intensities of SNPs were extracted and filtered by minor allele frequency to select the most informative SNPs.

6. Remaining data was run through PennCNV[26] software. PennCNV uses a hidden Markov model which incorporates log R ratio and B allele frequency for each participant to construct a model for the transition between copy number states and determine the location and frequency of copy number variations.

7. The results of the PennCNV software program were then run through quality control to only include CNVs with more than 10 SNPs and with a confidence score to number of SNPs ratio greater than 0.05.

8. A Python script was then used to sort the count the data to determine the number of CNVs, duplications, and deletions for each participant. A separate R script was used to determine the presence of common CNVs with a relation to psychiatric disorders (as found during the literature review) in the participant set.

Projected Statistical Analysis

- Specific CNV Data
  - Independent Samples T-Tests
    - Comparing self-report survey scores between those who have specific CNVs and those who do not
  - Analysis of Covariance (ANCOVA)
    - Comparing self-report survey Scores between those who have specific CNVs and those who do not while controling for age and sex.

- CNV Count Data
  - Correlation
    - Correlations between self-report survey scores and total number of CNVs, total number of deletions, or total number of duplications
  - Analysis of Variance (ANOVA)
    - Comparing quartiles and quintiles of the total number of CNVs, total number of deletions, or total number of duplications and the self-report survey scores
  - Analysis of Covariance (ANCOVA)
    - Comparing quartiles and quintiles of the total number of CNVs, total number of deletions, or total number of duplications and the self-report survey scores while controling for age and sex

Discussion

- Upon completion of statistical analyses, results may accept or reject null hypothesis

- Limitations of project
  - Sample size and composition
  - Power of statistical analyses

- Implications for the genetics of psychopathological disorders, case-control studies, and future research

Acknowledgements

Project Contributors:
- Dr. Marco Boks
- Dr. Kristel Kool-van Eijk
- Annabel Vreeker
- Dr. Jacobine Buizer-Voskamp

Special Thanks:
- Universiteit Medisch Centrum Utrecht
- Ludwig-Maximilians-Universität München
- EuroScholars Program

Works Cited

[1]Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444-454 (2006)

[2]Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848-853 (2007)

[3]Williams, N.M. *et al.* Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplication at 15q13.3. *Am. J. Psychiatry* **169**, 195-204 (2012)

[4]Zhang, F., Gu, W., Hurles, M.E. & Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451-481 (2009)

[5]Henrichsen, C.N., Chaignat, E. & Reymond, A. Copy number variants, diseases, and gene expression. *Hum. Mol. Genet.* **18**, R1-R8 (2009)

[6]Gill, M. Developmental psychopathology: the role of structural variation in the genome. *Dev. Psychopathol.* **24**, 1319-1334 (2012)

[7]Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population based study. *Lancet* **373**, 234-239 (2009)

[8]Frazer, K.A., Murray, S.S., Schork, N.J. & Topol, E.J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10**, 241-251 (2009)

[9]Hsara, A. *et al.* Population analysis of large copy number variants and hotspots of human disease. *Am. J. Hum. Genet.* **84**, 148-161 (2009)

[10]International Schizophrenia Consortium. Common polygenic variation contributes to the risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752 (2009)

[11]Guha, S. *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry* **70**, 253-260 (2013)

[12]Ben-Shacer, S. *et al.* Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J. Med. Genet.* **46**, 382-388 (2009)

[13]Grozeva, D. *et al.* Reduced burden of very large and rare CNVs in bipolar affective disorder. *Bipolar Disord.* **15**, 893-898 (2013)

[14]Degenhardt, F. *et al.* Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **159B**, 263-273 (2012)

[15]Amin, N. *et al.* A genome-wide linkage study of individuals with high scores on NEO personality traits. *Mol. Psychiatry* **17**, 1031-1041 (2012)

[16]Marian, A.J. Molecular genetics studies of complex phenotypes. *Transl. Res.* **159**,64-79. (2012)

[17]Cook, E.H. Jr & Scherer, S.W. Copy number variations associated with neuropsychiatric conditions. *Nature* **455**, 919-923 (2008)

[18]Schubart, C.D. *et al.* Cannabis use at a young age is associated with psychotic experiences. *Psychol. Med.* **41**, 1301-1310 (2010)

[19]Korver, N. *et al.* Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction; objectives, sample characteristics, recruitment, and assessment methods. *Int. J. Methods Psychiatr. Res.* **21**, 205-221 (2012)

Works Cited [cont.]

[20]Konings, M. *et al.* Validity and reliability of the CAPE: a self-report instrument for the measurement of psychotic experiences the general population. *Acta Psychiatr. Scand.* **114**, 55-61 (2006)

[21]Derogatis, L.R. & Unger, R. Symptom Checklist-90-Revised. *Corsini Encyclopedia of Psychology.* (2010)

[22]Beck, A.T. *et al.* An inventory for measuring depression. *Arch. Gen. Psychiatry* **4**, 561-571 (1961)

[23]McCrae, R.R., Costa, P.T. & Martin, T.A. The NEO-PI-3: A more readable revised NEO Personality Inventory. *J. Pers. Assess.* **84**, 261-270 (2010)

[34]Spielberger, C.D. State-Trait Anxiety Inventory. *Corsini Encyclopedia of Psychology* (2010)

[25]Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559-575 (2007)

[26]Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665-1674 (2007)