Investigating the Frequency of Copy Number Variants according to Ancestry in an Unselected Population

Alexys Knighton, COL 2023, Guillaume Huguet, Ph.D., Laura Schultz, Ph.D., David Glahn, Ph.D., Sébastien Jacquemont, M.D., Laura Almasy, Ph.D.
Penn Undergraduate Research Mentorship Program
Department of Biomedical and Health Informatics at the Children's Hospital of Philadelphia, UPenn School of Medicine, University Hospital Center– Saint Justine, Boston Children's Hospital

Introduction

Copy number variations (CNVs) occur when the number of copies of a particular genetic material varies from one segment to another. It has been established that CNVs in a genome may result in a phenotypic expression, if they are large enough (Valsesia 2013). However, the mechanisms and full effect of CNVs are still being explored. Many studies have been conducted in populations that have a disorder. De novo CNVs have been recognized as a significant risk factor for Autism Spectrum Disorder and it has been determined that certain CNVs can contribute to the risk for, or the protection from, schizophrenia (Sanders 2015)(Sebat 2007)(Marshall 2017). The effect of CNVs in unselected populations has yet to be fully explored.

Objective

To look at the burden of CNVs in unselected populations, or those not affected by known disorders, with the intent to observe the unbiased effects of CNVs. Additionally, we stratified the sample set by ancestry to observe the differences of CNV frequencies by heritage.

Methods

Sample: The study population was taken from the UK Biobank.

- 500,000 ages 40-69
- Principal components of ancestry were used to stratify participants into four categories (Figure 1).

CNV calling and CNV burden

- CNVs were called using PennCNV and QuantiSNP
- Overall burden of deleterious CNVs was estimated using pLI and LOEUF scores (Figure 2).
- Eight individual CNVs (Figure 3) had a count of of ≥ 275, meaning that if frequency were even across ancestry groups, we would expect roughly 5 copies in the smallest groups.

Results

Figure 1: UK Biobank Sample Population by Ancestry

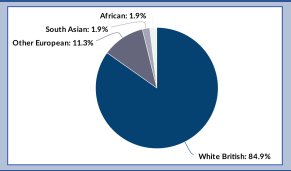
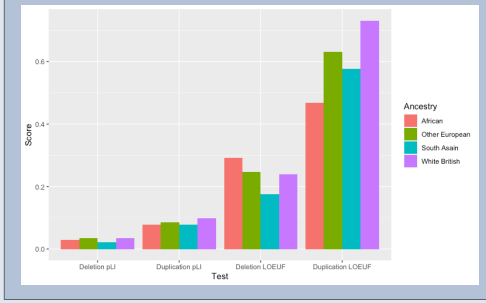
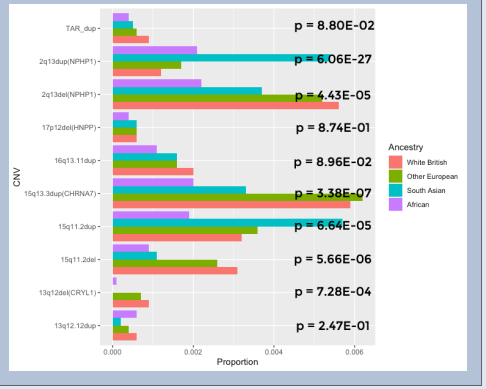


Figure 2: Mean Deletion of pLI and LOEUF by Ancestry



biobank uk
Improving the health of future generations

Figure 3: CNV Frequency by Ancestry Group with Test for Independence



Discussion

- As shown in Figure 2, the mean pLI and LOEUF scores did not differ significantly between the different ancestry groups.
- Four CNVs (13q12del(CRYL1), 15q11.2del, 15q13.3dup(CHRNA7), and 2q13del(NPHP1) were at elevated frequency in the White British population, while being lowered in the other four ancestral groups.
- Two CNVs (15q11.2dup and 2q13dup(NPHP1)) were lower than expected in the White British population, while being elevated in the other four ancestral groups.
- Kendall et. al (2019) investigated cognitive performance and copy number variants in the UK Biobank. Of the six CNVs showing frequency differences by ancestry, four were associated with cognitive performance, including 13q12del(CRYL1), 15q11.2del, 15q11.2dup, and 15q13.3dup(CHRNA7).
- Crawford et al. (2019) looked at the medical consequences of CNVs in adults in the UKBiobank. While none of the six significant CNVs in Table 2 were significantly associated with a medical consequence, some did offer data suggestive of a possible medical outcome. 13q12del(CRYL1) was associated with paralytic symptoms, migraines, ovarian cysts, and arrythmia. 15q11.2del was associated with prostate hyperplasia. 15q11.2dup was associated with neuropathies, gout, atherscleratic vascular disease, high cholestorol, and ischaemic heart disease. 15q13.3dup(CHRNA7) was associated with getting a heart valve replacement, diabetes, ischaemic heart disease, asthma, and inflammatory bowel disease. 2q13del(NPHP1) was associated with varicose veins and osteoporosis.

Methods

Analysis:

- The mean was calculated for pLI scores and LOEUF scores, both for deletions and duplications.
- A chi-squared test for independence was performed to test whether individual CNVs were distributed evenly across the different ancestry groups.
- RStudio software was used for mathematical calculations.

Next Steps

The UK Biobank is mostly a homogeneous sample, so it would be useful to replicate these same analyses in another sample population that is larger and offers more diversity.
Additionally, we want to look at the association of these CNVs with quantitative measures of neuropsychiatric functioning (mood and social function and psychosis).

Acknowledgements

- Supported by NIH U01 MH119690
- A very big thank you to Dr. Laura Almasy for mentoring me this summer, the Velay Women's Science Research Fellowship, and the PURM program for providing the connection between novice undergraduates and researchers.

References

Valsesia A, Macé A, Jacquemont S, Beckmann JS, Kutalik Z. The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. *Front Genet*. 2013;4:92. Published 2013 May 30. doi:10.3389/fgene.2013.00092

Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015;87(6):1215-1233. doi:10.1016/j.neuron.2015.09.016

Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316(5823):445-449. doi:10.1126/science.1138659

Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects [published correction appears in Nat Genet. 2017 Mar 30;49(4):651] [published correction appears in Nat Genet. 2017 Sep 27;49(10):1558]. *Nat Genet*. 2017;49(1):27-35. doi:10.1038/ng.3725

Crawford K, Bracher-Smith M, Owen D, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet*. 2019;56(3):131-138. doi:10.1136/jmedgenet-2018-105477

Kendall KM, Bracher-Smith M, Fitzpatrick H, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *Br J Psychiatry*. 2019;214(5):297-304. doi:10.1192/bjp.2018.301