Supplementary appendix

Presence of pathogenic Copy number variants is correlated with socioeconomic status

George J Burghel (1,2), Unzela Khan (1), Wei-Yu Lin (3) William Whittaker (4), Siddharth Banka (1,5)

Burghel GJ, *et al. J Med Genet* 2019; 57:70–72. doi: 10.1136/jmedgenet-2019-106292

Supplementary Methods

Database: Manchester Centre for Genomic Medicine (MCGM) is a regional genomic diagnostic laboratory and receives referrals for genomic testing from the North-West (NW) of England (http://www.mangen.co.uk, 2019). MCGM offers array-comparative genomic hybridisation (aCGH) as first line of investigation to individuals with developmental disorders and congenital malformations to identify disease causing copy number variants (CNVs).

CNVs are the most common type of structural variation in the human genome and involve more base pairs than any other type of genetic variation [1]. CNVs can be classed as losses (deletions) or gains (e.g. duplications or triplications). Depending on their phenotypic effect, CNVs can be classed as benign (class 1), likely benign (class 2), variant of uncertain significance (class 3), likely pathogenic (class 4) and pathogenic (class 5)[2].

Likely pathogenic (class 4) and pathogenic (class 5) CNVs have been implicated in a wide range of human disorders including intellectual disabilities and neurodevelopmental delays[3]. These disease-causing class 4 and 5 CNVs, like any other CNVs, can be inherited from a parent or they may arise *de novo.* Inherited CNVs generally show reduced penetrance, meaning that the individual carrying the CNVs may not be clinically classed as affected. *De novo* pathogenic CNVs are generally, associated with more severe phenotype when compared with the phenotypes of inherited pathogenic CNVs[4].

The current aCGH platform used at MCGM is the OGT v3 8x60K array platform with a backbone resolution of 180kb and is validated prior to diagnostic use. Analysis is performed using the the CytoSure™ Interpret Software (v4.9). This tool contains tracks linking information from the Database of Genomic Variants, DECIPHER, Online Mendelian Inheritance of Man and the local MCGM patient database (https://www.ogt.com, 2019). Interpretation and classification of CNVs are conducted using the laboratory standard operating procedures and published guidelines[2].

We curated an anonymised departmental database of results from over 17,000 postnatal clinical array-CGH testing performed at MCGM between 2010 and 2017. Our database included information on each identified CNV, its clinical classification, size, loss or gain status, the inheritance status (*de novo* or inherited from a parent) where available and postcode. Details of CNVs are in Table S1.

Data cleaning and filtering: From the database, we selected all pathogenic (Class 5) or likely pathogenic (Class 4) CNVs. From these pathogenic/likely-pathogenic CNVs, we identified CNVs in which complete inheritance status was available (*De novo* vs inherited). From these CNVs we included only one proband from each family (to avoid sample bias by double counting). We also excluded derivative CNVs resulting from parental balanced translocation (as these CNVs cannot be classed as *de novo* or inherited in the traditional sense) and sex chromosomes CNVs (because their phenotype effect is sex dependant). Cases belonging to postcodes for which IMDR data was unavailable were excluded.

Determination of SES: A number of measures and methodologies to study SES in health-related contexts exist[5]. English Indices of Deprivation are one such widely used SES measure in health research. They combine seven domains of deprivation for small geographical areas in England (referred to as Lower-Layer Super Output Areas or LSOAs) into

Burghel GJ, *et al. J Med Genet* 2019; 57:70–72. doi: 10.1136/jmedgenet-2019-106292

an overall weighted aggregation index of multiple deprivation or IMD. These seven domains and their associated weightings are: income deprivation (22.5%), employment deprivation (22.5%), health deprivation and disability (13.5%), education, skills and training deprivation (13.5%), barriers to housing and services (9.3%), crime (9.3%) and living environment deprivation (9.3%) (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015). Each of the LSOAs is scored and ranked from the most deprived (rank 1) to the least deprived (rank 32,844). LSOAs are categorised into 10 equal groups whereby the first and the tenth deciles include the most and the least deprived 10% LSOAs respectively. Each postcode in England falls within an LSOA (http://imd-by-postcode.opendatacommunities.org/). From the postcodes associated of the CNVs we retrieved the index of multiple deprivation rank (IMDR), and its seven constituent domains, using the English indices of deprivation 2015 (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015).

Statistical analyses: All the statistical tests were performed using R 3.5.2, unless otherwise specified. Statistical significance was set at $p < 0.05$. Chi-squared tests of independence and trend were performed to examine the associations between IMDR deciles (and its seven constituent domains) and CNV inheritance (*de novo*/inherited). To investigate if there was a correlation between age of diagnosis and SES, Jonckheere-Terpstra test was used. To explore the joint effect of the CNV type (losses/gains) and inheritance on IMDR, we further categorised samples into 4 groups based on CNV inheritance and type. Kruskal-Wallis rank sum test was used to test if IMDR are different among 4 groups. Mann Whitney U-tests were used as post-hoc testing for Kruskal-Wallis rank sum test and FDR (false discovery rate) was used to control the inflation of type I error rates. Pairwise fdr-adjusted p values were reported. CNV sizes are defined as the intervals between CNV start and end in the million base scale (Mb). To investigate if CNV sizes differ by IMDR group, Kruskal-Wallis rank sum test was used.

Burghel GJ, *et al. J Med Genet* 2019; 57:70–72. doi: 10.1136/jmedgenet-2019-106292

Supplementary Results

Database: We identified 1,567 pathogenic (Class 5) or likely pathogenic (Class 4) CNVs in our database. Complete inheritance status was available for 614 CNVs (324 *de novo* and 290 inherited). Cleaning and filtering of our data left us a final set of 473 IMDR datasets corresponding to unique individuals with class 4 and 5 autosomal CNVs with full inheritance information. This included 218 inherited (138 losses and 80 gains) ranging in size between 0.002Mb-13.75Mb (median 1.33Mb), 255 *de novo* (193 losses, 62 gains) ranging in size between 0.006Mb-13.93Mb (median 3.45Mb) (Table S1).

IMDR comparisons of pathogenic CNVs against the general population: Chi-squared goodness of fit test showed that the IMDR spread of our patients was significantly different from that of the NW of England population ($p=1.8x10^{-8}$) (Figure1).

IMDR comparisons according to inheritance: The difference above is driven by inherited CNVs rather than *de novo* CNVs; the IMDR of patients with *de novo* CNVs were not too different to that of the NW of England population ($p=5.1x10^{-2}$) while the IMDR of patients with inherited CNVs were significantly different to that in the NW of England ($p=3.1x10^{-10}$) (Figure 1). We also found significant differences in IMDR scores between inherited and *de novo* pathogenic CNVs ($p_{x^2_{df=9}}=3.3x10^{-4}$) (Figure1). Relative to *de novo* CNVs, inherited CNVs are 2.06 times more likely to be living in areas of high deprivation (relative risk ratios = 2.06, $p= 1.1x10^{-5}$) (Figure 1) (Supplementary Figure S1). This difference was significant across following deprivation domains of IMDR – income; employment; health; education, skills and training (Figures S2-S6). There were not significant differences in the following domains - barriers to housing and services; and living environment (Figures S7 and S8).

IMDR comparisons according to age: There was no significant age trend across the different IMDR deciles for *de novo* CNVs (p=0.0615) and for inherited CNVs (p=0.1615) (Figure S9).

IMDR comparisons according to CNV type: IMDR was not significantly different between the two CNV types (Losses and Gains) (p=0.52). Of note, *de novo* CNVs have higher percentages of losses (76%) compared to inherited CNVs (63%) (p=0.005). To check the joint effect of the CNV type and inheritance on IMDR, we further categorised samples based on inheritance (*de novo*/inherited) and type (losses/gains). There are significant differences in IMDR among 4 groups ($p_{\text{Kruskal-Wallis rank sum test}}=2.1x10^{-4}$) (Figure S10). Post-hoc pairwise comparisons showed that both inherited losses and gains were significantly associated with lower IMDR in comparison to *de novo* losses and gains indicating that the effect mainly resulted from mode inheritance rather than CNV type ($p_{\text{false discovery rate}} < 0.05$, Figure S2).

IMDR comparisons according to CNV size: Median CNV sizes ranged from 1.4 to 3.1 Mb across IMDR (Figure S11A). There was no evidence of an effect of CNV size on IMDR ($p_{\text{Kruskal-Wallis rank sum test}}=0.48$) (Figure S11B).

Burghel GJ, *et al. J Med Genet* 2019; 57:70–72. doi: 10.1136/jmedgenet-2019-106292

Supplementary References

1 Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends in Genetics* 2002; **18**: 74–82.

2 Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genetics in Medicine* 2011; **13**: 680–5.

3 Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nature Communications* 2017; **8**: 14366.

4 Veltman JA, Brunner HG. *De novo* mutations in human genetic disease. *Nature Reviews Genetics* 2012; **13**: 565–75.

5 Shavers VL. Measurement of socioeconomic status in health disparities research. *J Natl Med Assoc* 2007; **99**: 1013–23.