5-2020

Frequency Of Copy Number Variants Involving The Sex Chromosomes In A Clinical Setting

Autumn Vara

FREQUENCY OF COPY NUMBER VARIANTS INVOLVING THE SEX CHROMOSOMES

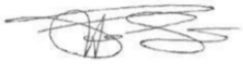
IN A CLINICAL SETTING

By

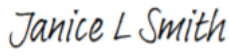
Autumn Elizabeth Vara, B.S.

APPROVED:

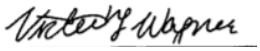
David Rodriguez-Buritica, M.D.
Advisory Professor



Syed S. Hashmi, M.D., MPH, Ph.D



Janice L. Smith, Ph.D



Victoria F. Wagner, M.S.

Kathryn Gunther, M.S.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth
Graduate School of Biomedical Sciences

FREQUENCY OF COPY NUMBER VARIANTS INVOLVING THE SEX CHROMOSOMES IN A CLINICAL SETTING

A

THESIS

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment of

the Requirements for the

Degree of

MASTER OF SCIENCE

by

Autumn Elizabeth Vara, B.S.

Houston, Texas

May 2020

FREQUENCY OF COPY NUMBER VARIANTS INVOLVING THE SEX CHROMOSOMES IN A CLINICAL SETTING

Autumn Elizabeth Vara, B.S.

Advisory Professor: David F. Rodriguez-Buritica, M.D.

Abstract

Copy number variants (CNVs) are a common finding in the clinical setting and contribute to both genetic variation as well as disease. Recently, studies have described the accumulation of multiple CNVs as a disease modifying mechanism. While it has been characterized how additional CNVs may play a role in phenotype, in which ways and to what extent sex chromosomes are involved has not been fully described. We performed a secondary data analysis using the DECIPHER database on 2,273 de-identified individuals with 2 CNVs. CNVs were designated primary and secondary based on our criteria and characteristics of both CNV groups were described. Further analysis was performed identifying differences in CNVs on the sex chromosomes vs autosomes. We found that CNVs on the sex chromosome have a significant difference compared to autosomes when comparing median size (p=0.013), pathogenicity classifications (p<0.001), and variant classification (p=0.001). We identified chromosome combinations for primary and secondary CNVs, and identified the X chromosome was the most common site for a secondary CNV. Additionally, we observed the plurality of secondary CNVs fell in the same chromosome as the primary CNV. From this study, we can conclude that the X chromosome is the most common site for secondary CNVs in a clinical setting. Identification of chromosome combinations for primary and secondary CNVs is essential in explanation of complex phenotypes and highlights areas of importance of the human genome.

Table of Contents

List of illustrations

List of Tables

Introduction

Copy number variants (CNVs) are structural changes of the human genome due to chromosomal anomalies such as deletions, duplications or insertions. (Girirajan, Campbell, and Eichler 2011). CNVs have been well described as contributing to genetic variability as well as disease (Nowakowska 2017). These segments of DNA are found in all individuals and include deleted and duplicated regions ranging from one kilobase pair (kbp) to several megabase pairs (Mbp) (Itsara et al. 2009). CNVs in the human genome often arise in regions containing low copy repeats or segmental duplications, which are blocks of DNA that share >90% sequence identity (Tie-Lin Yang 2018). The most common mechanism from which this occurs is non-allelic homologous recombination (NAHR) and often is responsible for recurrent CNVs in the genome (Hastings et al. 2009). We can predict that with the greater the density of segmental duplications in a chromosome, the more enriched that chromosome may be for CNVs. Another mechanism associated with CNVs, and thought to be responsible for non-recurrent CNVs, is that of non-homologous end joining (Tie-Lin Yang 2018).

It is estimated that on average, a healthy individual harbors approximately 3-7 CNVs with a total average of 540 kbp of copy number variable DNA per person (Itsara et al. 2009). The majority of copy number variants less than ~400 kbp are thought to be benign although exceptions for smaller, pathogenic CNVs exist. (Miller et al. 2010). For instance, 5-10% of healthy individuals have one CNV spanning at least 500 kbp suggesting that size does not always correlate to pathogenicity (Itsara et al. 2009).

Certain rare, pathogenic CNVs have been identified in individuals with a characteristic set of clinical features and contribute to syndromic disorders such as Williams syndrome and 22q11.2 deletion syndrome. More recently, rare, recurrent

CNV's in conjunction with a secondary hit, such as an additional CNV, were found to exacerbate phenotype (Girirajan et al. 2012). A recurrent microdeletion at 16p12.1 is associated with variable expressivity for a spectrum of conditions including neuropsychiatric disorders, autism spectrum disorder, and intellectual disability when in conjunction with other CNVs (Girirajan et al. 2010) To better understand this phenomenon Girirajan et al. conducted a study which looked at 32,587 chromosomal microarrays from children with intellectual disability with or without congenital anomalies and compared that with 8,329 control samples. In this analysis, 10.1% of patients with a rare CNV known to be associated with a neurodevelopmental phenotype or disorder had at least one additional CNV larger than 500 kbp on an autosome, further characterizing the compounding effect that additional CNVs have on disease severity. Other studies also describe a similar observation that a "second" hit, such as a CNV, exacerbates phenotypes (Girirajan et al. 2010, Le Gall et al. 2017). Prakash et al. identified a subset of females with Turner syndrome (TS) with a secondary event, such as a CNV, are significantly more likely to have congenital heart disease than females with TS who did not (Prakash et al. 2016)

While these studies better characterize how additional CNVs may play a role in driving phenotypic expression, how and to what extent sex chromosomes are involved has not been fully described. Sex chromosome aneuploidies are more common than aneuploidies involving the autosomes in, both, neonates and miscarriages, which may support the idea that CNVs involving the sex chromosomes are more common than those on autosomes (Templado, Uroz, and Estop 2013). Based on this information, it is our hypothesis that sex chromosomes will more commonly harbor additional CNVs than autosomes. Although there is no current data supporting this hypothesis, our institution

has identified several individuals found to harbor pathogenic CNVs contributing to the phenotype, in addition to a secondary sex chromosome abnormality. This novel research has the potential to add knowledge in highlighting sex chromosomes in etiology of indications, genomic regions that are of importance, and provide clues for predicting clinical phenotypes.

Methods:

Study sample

This project involves secondary data analysis on a de-identified dataset from DECIPHER (https://decipher.sanger.ac.uk/), a publicly available database in which referring clinicians submit genotypic and phenotypic information on patients who have consented for broad data-sharing. DECIPHER is comprised of data independently submitted by an affiliated institution and at minimum includes chromosome(s) involved, start and end breakpoints for each CNV, and variant class (duplication, deletion, etc.) for each individual. Information such as pathogenicity, inheritance pattern, phenotype, mean ratio, and contribution to phenotype was also available if provided by the referring clinician. Information pertinent to this study include chromosome involved, start/end point, variant class, pathogenicity, inheritance pattern, and phenotype. The dataset received encompasses submissions since the creation of DECIPHER in 2004 until May 2019.

For variants where pathogenicity was provided, DECIPHER categorized into five classes labeled as ‘pathogenic’, ‘likely pathogenic’, ‘variant of uncertain significance (VUS)’, ‘benign’, or ‘likely benign’. We kept these classifications as is with the exception of ‘benign’ and ‘likely benign’ which were combined into one category defined as ‘benign’.

Inheritance patterns for CNVs provided by DECIPHER fell into multiple categories. Variants defined as ‘de novo constitutive’ or ‘de novo mosaic’ in the DECIPHER database were combined and categorized as ‘de novo’ for the purpose of the analysis. Variants in DECIPHER designated ‘biparental’, ‘inherited from normal

parent', 'inherited from a parent with similar phenotype to child', 'inherited from a parent with unknown phenotype', 'maternally inherited, constitutive in mother', 'maternally inherited, mosaic in mother', 'paternally inherited, constitutive in father' or 'paternally inherited, mosaic in father' were further combined into 'inherited'. These two categories were reserved for all studies regarding inheritance.

The size of each CNV was calculated using the provided start and break point for each CNV.

Variants received from DECIPHER were categorized into variant classes labeled 'gain', 'triplication', 'duplication', 'duplication/triplication', 'amplification', or 'deletion'. CNVs labeled as 'duplication/triplication' or 'gain' were excluded due to ambiguity (n=5).

Inclusion/Exclusion criteria

A total of 24,525 unique patients were received from DECIPHER database. For the purpose of this study, patients with strictly one copy number variant were excluded for analysis (n=19,031). Patients were further eliminated if break points were indicative of terminal CNVs that suggested an unbalanced translocation inherited from a balanced carrier parent (n=160). Patient sex was then manually extracted from DECIPHER for this subset of patients using the patient ID provided. Patients with unknown sex were eliminated from the dataset, reaching a final sample of 4,263 unique individuals.

For studies describing the characteristics of secondary CNVs, only individuals with 2 CNVs were analyzed. Therefore, an additional 867 individuals were excluded due to harboring $\geq$ 3 CNVs. For the remaining 3,396 individuals, it was necessitated that one CNV be equal to or larger than 500 kbp. This was designated the primary CNV for all

individuals. The secondary CNV was thus defined as the smaller of the two variants. There were 961 individuals who did not have a CNV $\geq$ 500 kbp. Seven patients were further excluded due to harboring 2 CNVs with the exact size on the same chromosome, therefore we could not determine which would be defined as the primary and secondary CNVs. Additionally, individuals were eliminated if two CNVs were found on the same chromosome less than 300 kbp apart to reduce the possibility of being one large CNV. There is no current literature to suggest how far 2 CNVs on the same chromosome should be to exclude this likelihood, thus 300 kbp was chosen arbitrarily. We also disregarded patients with 2 CNVs that embedded or overlapped one another. This gave a final sample size of 2, 273.

Studies performed on each cohort

For the full cohort of 4,263 individuals, descriptive studies describing number of CNVs (Table 1), sex (Table 2), pathogenicity, inheritance pattern, size of CNV, variant class, and chromosome frequencies were performed. There were 4,912 CNVs with known pathogenicity (Figure 1). Inheritance pattern was known for 3,489 CNVs (Figure 2). Variant classes were compared to pathogenicity classifications for 4,912 CNVs (Figure 3). Median sizes of CNVs were compared to pathogenicity groups. To determine chromosome involvement, we created a graph comparing each chromosome and the frequency of CNVs that were found on the respective chromosome (Figure 4a). We compared frequency of CNVs per chromosome to chromosome length to determine if size was a contributing factor to CNV density (Figure 4b).

To highlight characteristics of secondary CNVs, we described sex, pathogenicity, inheritance pattern, size, variant classification, and chromosome involvement in this

cohort (n=2,273). Pathogenicity was known for 990 primary and1,136 secondary CNVs (Figure 5). Inheritance patterns were known for 1,007 primary CNVs and 851 secondary CNVs. Variant classification was compared between primary and secondary CNVs. Additionally, we compared variant classes and pathogenicity categories for both, primary and secondary CNVs. Chromosome involvement for primary and secondary CNVs are compared Figure 6. A heatmap was then created to identify combinations of chromosome involvement for primary and secondary CNVs (Figure 7a). We then recreated this map to highlight which chromosomes were frequent sites of secondary CNVs (7b). To do this, we calculated the average relative frequency of each secondary CNV. The y-axis was then reorganized from chromosome with the highest relative frequency to the lowest.

To characterize CNVs involving the sex chromosomes, we used the previous sample of 2,273 with 2 CNVs. There were 438 CNVs harbored on the sex chromosomes. Of these CNVs, 205 were classified as primary and 233 were secondary variants. Comparisons were made between location of CNV (sex chromosomes vs autosomes) and size, pathogenicity, and variant classes. Pathogenicity groups are compared between location of CNV (Figure 8). We looked to identify if there were pairs of pathogenicity combinations for primary and secondary CNVs that were recurrent. To perform this, separate investigations were completed based on if the primary CNV fell on an autosome or sex chromosome. For this analysis we focused on individuals with 46,XX karyotype to control for chromosome dosage and size (Figure 9). Variant classes were compared between location of CNV. Further analysis was performed to identify pairs of variant class combinations. Separate investigations were performed based on if the primary CNV fell on an autosomes vs sex chromosome. For this part of the analysis, we

focused again on individuals with 46, XX karyotype to control from chromosome dosage and size. Additionally, we combined variant classes of 'duplication' and 'triplication' into a single category labeled as 'gain' (Figure 10). We further investigated secondary CNV characteristics of individuals who had a pathogenic CNV on the X-chromosome to identify any trends. We looked to identify any recurrent common secondary CNV, variant class, and phenotype. We limited this study again to individuals with 46,XX karyotype (Table 3).

Statistical analysis

All analyses were performed in STATA (v.13.1). P-value <0.05 was set for statistical significance. Continuous variables were not normally distributed and were described using medians and interquartile ranges (IQR). A Kruskal-Wallis test was used to compare continuous variables with categorical. A post-hoc Dunn test was performed to characterize results with statistical significance. Chi-squared tests were used to compare categorical variables.

Results

Full cohort characteristics

There were 4,263 unique individuals from the DECIPHER cohort who had more than one CNV. The number of CNVs per individual ranged anywhere from 2-13 with the majority of individuals (79.7%) having only two (Table 1). Among this sample there were a total of 9,766 CNVs.

Number of CNVs	n	%
2	3,399	79.7
3	624	14.6
4	171	4.0
5	40	0.9
6	12	0.3
7	8	0.2
8	4	0.2
9	2	0.1
10	2	0.1
11	0	0
12	0	0
13	1	0.1

Table 1. Frequency of number of CNVs. This table categorizes percentages based on number of CNVs identified in individuals with 2 or more CNVs.

We did find an enrichment for males to harbor more than one CNV in this cohort. The majority of individuals (58%) had a normal male karyotype of 46,XY and 41% had a normal female karyotype of 46,XX. There were 13 individuals (0.3%) with sex chromosome aneuploidies including 45,X; 47,XXX; 47,XXY; and 47,XYY (Table 2).

Sex	n	%
45,X	2	0.05
47,XXX	1	0.02
47,XXY	4	0.09
47,XYY	6	0.14
46,XX	1,759	41.26
46,XY	2,491	58.43

Table 2. Frequency of sex chromosomal aneuploidies. Represented in this table are the karyotypes observed in individuals with two or more CNVs and the respective frequency.

Due to the nature of the dataset used, half of the CNVs did not have pathogenicity information included as they were not submitted by the referring center and therefore listed as unknown. For 4,912 variants where pathogenicity was known, the largest category was VUS at 44% (Figure 1).

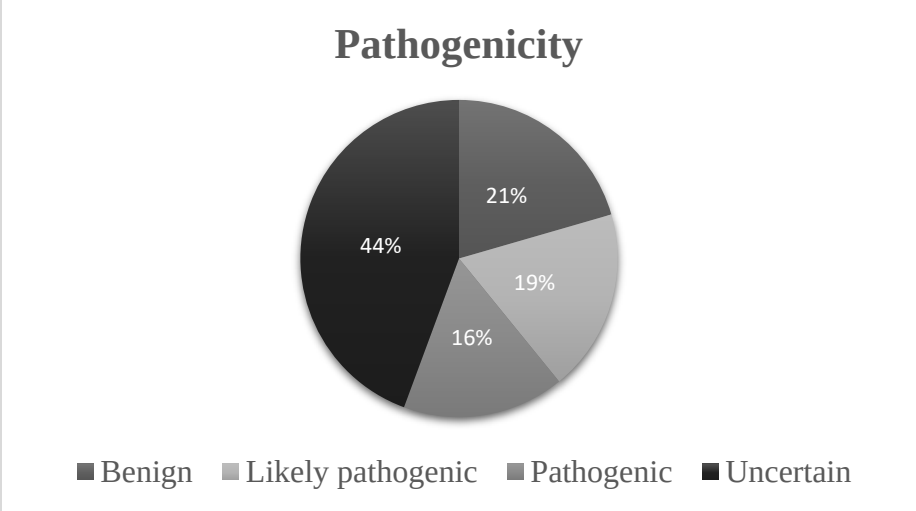


Figure 1. Pathogenicity of CNVs in individuals with ≥2 CNVs. This figure represents the percentage make up of pathogenicity classification for 4,912 CNVs in this cohort.

6,277 variants (67%) did not have inheritance data provided. Of the 3,489 variants with inheritance information, there was a nearly equal distribution with 51% of variants classified as de novo and 49% inherited (Figure 2).

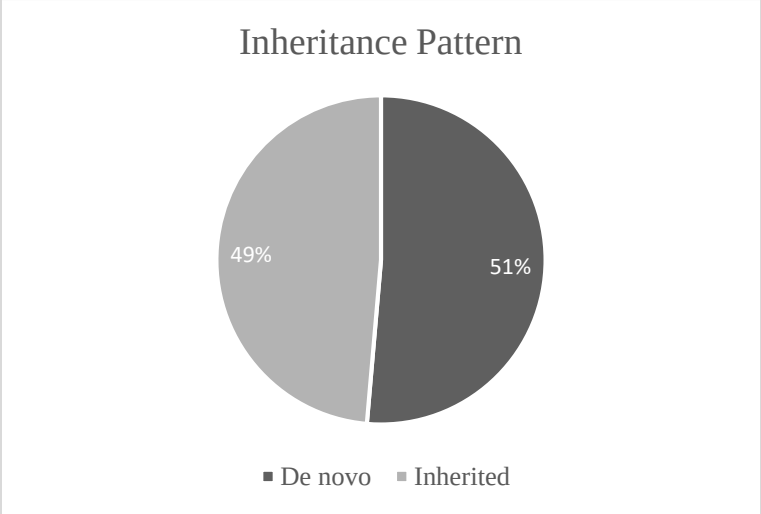


Figure 2. Inheritance patterns of CNVs in individuals with ≥2 CNVs. De novo vs inherited variants in 3,489 CNVs.

There was a statistically significant difference in median sizes amongst pathogenicity groups (p<0.001). CNVs defined as pathogenic had the largest median size of approximately 3.2 Mbp with an interquartile range (IQR) of 7.3 Mbp (8.5 Mbp-1.2 Mbp). The smallest pathogenic variant measured less than 0.001 Mbp (60 base pairs). The largest pathogenic variant was approximately 155 Mbp which is consistent with a sex chromosome aneuploidy of the X chromosome. Variants denoted LP had a median size of approximately 0.6 Mbp and IQR of 1.6 Mbp (1.9 Mbp-0.3 Mbp). The smallest LP CNV was less than 0.001 Mbp (97 base pairs) while the largest was 60 Mbp. Benign variants had the smallest median size of the groups with a median of 0.1 Mbp with an IQR of 0.2 Mbp (0.3 Mbp-0.07 Mbp). The smallest benign variant was less than 0.001 Mbp (36 base pairs) and the largest 100 Mbp. VUS CNVs had a median value which fell between those of LP and benign at approximately 0.4 Mbp and an IQR of 0.5 Mbp (0.7 Mbp-0.2 Mbp). We identified a significantly larger median of pathogenic CNVs compared to all other categories (p<0.001). LP variants had a

significant larger median than VUS and benign (p-<0.001). VUS had a significantly larger median than benign (p<0.001).

Variant classes were compared to pathogenicity classifications and a statistically significant difference was observed (p<0.001). Over 60% of pathogenic variants were found to be deletions. Reciprocally, 55% of benign variants and 61% of VUS were defined as duplications (Figure 3).

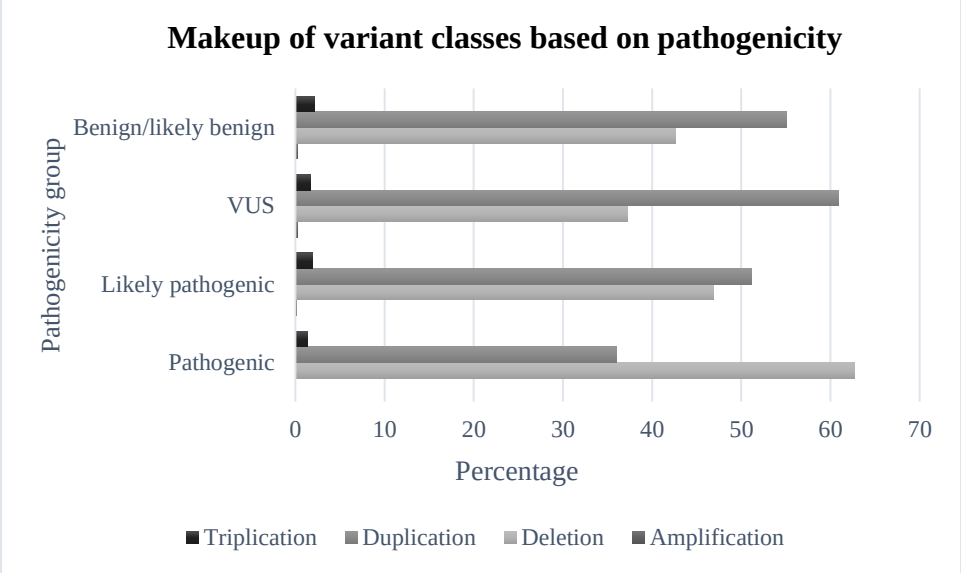


Figure 3. Frequency of CNVs per variant class based on pathogenicity. Percentages that made up each variant class for each pathogenicity group in individuals with two or more CNVs.

Of the 9,766 CNVs, nearly 11% were found between the two sex chromosomes, with 9.4% located on the X chromosome and 1.5% on the Y chromosome. Other chromosomes enriched for CNVs include 1, 16, 15, and 2, each approximately representing 7% of CNVs, respectively (Figure 4a). Chromosome 21 had the smallest amount of CNVs of all the chromosomes. We also identify that chromosomes 22, 16, and 15 have an enriched amount of CNVs per Mbp, whereas chromosome 5 has the least (Figure 4b).

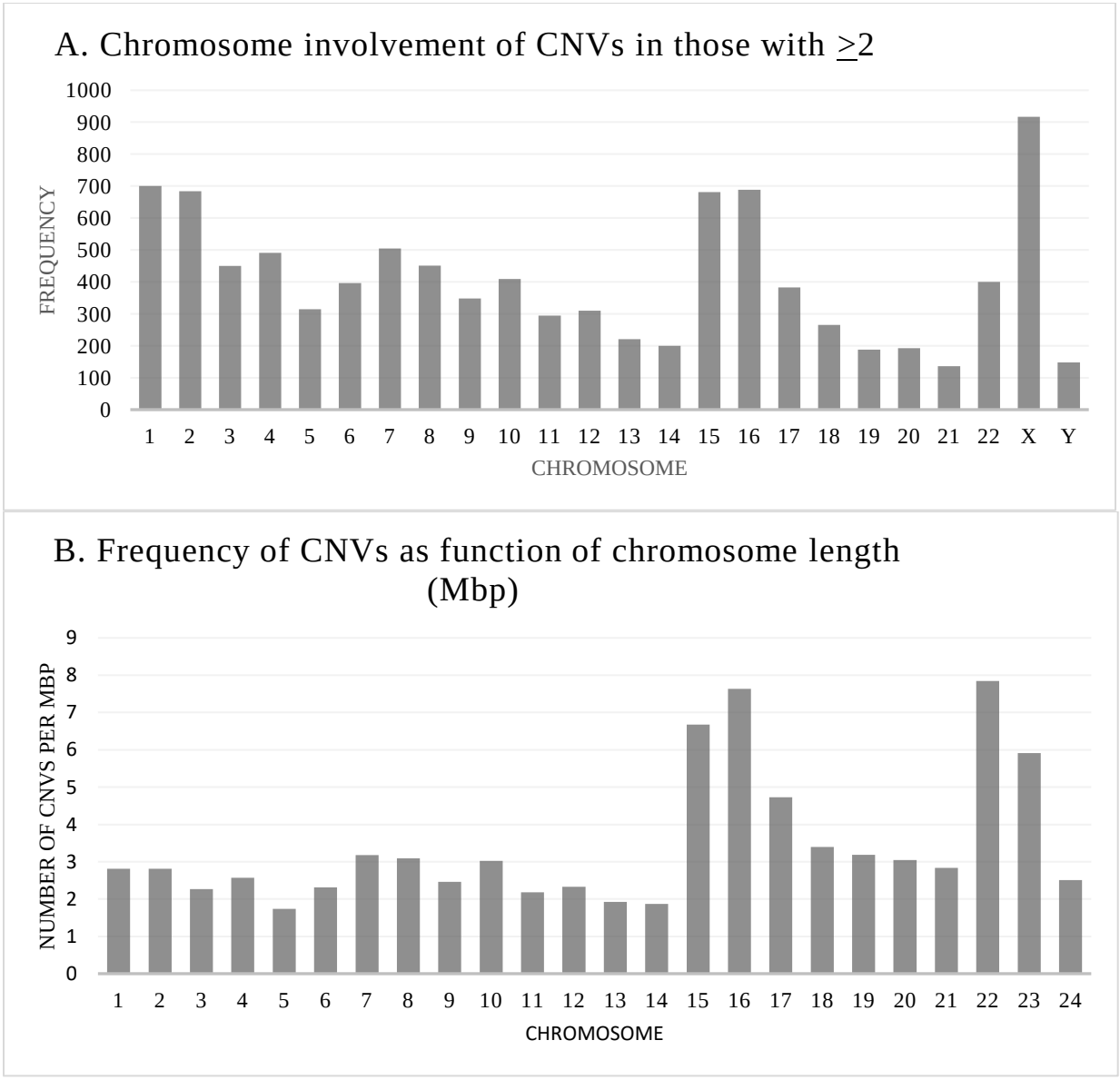


Figure 4. Chromosome involvement of CNVs in those with ≥2 CNVs. Panel A characterizes the frequency of CNVs in individual chromosomes for individuals who have two or more copy number variants. The x-axis lists each chromosome and the y-axis shows the number of CNVs. Panel B illustrates the number of CNVs per Mbp for each chromosome.

Characteristics of secondary CNVs

When further characterizing secondary CNVs, there were 2,273 unique individuals who exclusively had two CNVs on different chromosomes or on same chromosomes with more than 300 kbp apart, with one of the CNVs' size equal to or

large than 500 kbp. Of the 2,273 individuals, 1,269 (56%) had a 46,XY karyotype, and 996 (44%) had a 46,XX karyotype. There were 8 additional patients (less than 1%) with the same sex chromosome aneuploidies as described previously.

We observed a statistically significant difference in pathogenicity classification between the primary and secondary CNVs ($p<0.001$). 1,283 individuals (53%) did not have pathogenicity information for the primary CNV. Of the primary CNVs for which pathogenicity classification was known, 33% were labeled as pathogenic, 27% LP, 4% benign, and 36% were VUS. The same analysis was performed on secondary CNVs. 1,136 individuals (51%) had known data regarding the pathogenicity of the secondary CNV. Similarly to primary CNVs, the greatest proportion of variants was classified as VUS (46%). Pathogenic and LP variants were found in a smaller proportion at 15% and 22% respectively. Benign variants made up 16% of secondary CNVs. Comparisons of primary and secondary CNV pathogenicity can be seen in Figure 5.

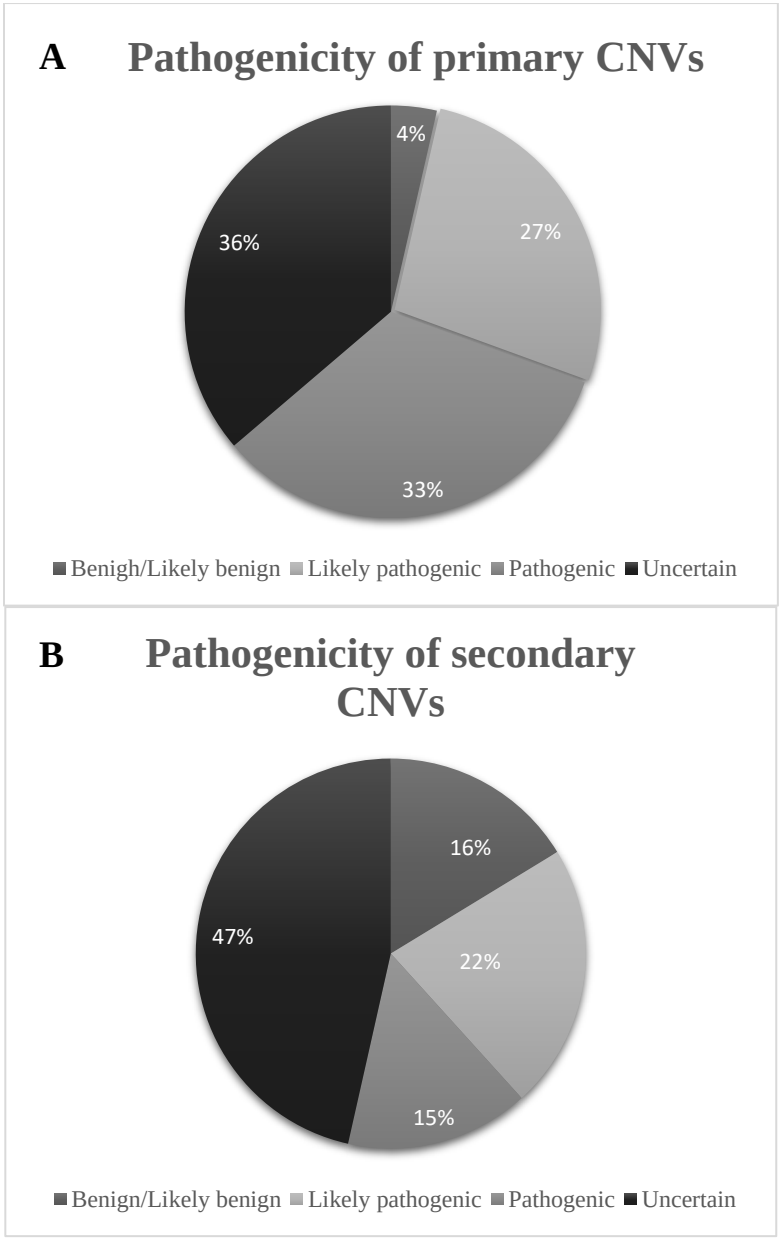


Figure 5. Pathogenicity of primary and secondary CNVs. Panel A illustrates the distribution of pathogenicity classes in primary CNVs. Panel B highlights the pathogenicity classification of secondary CNVs.

Inheritance patterns were compared between primary and secondary CNVs and a statistically significant difference was observed (p<0.001). For the primary CNVs, 1,266 individuals (56%) had unknown or missing inheritance data. Of those CNVs where inheritance pattern was provided, 68% of variants were de novo whereas 32% were inherited. This demonstrates that primary CNVs were largely de novo. Then shifting to

secondary CNVs, there were 851 individuals (37%) with known inheritance in this group. We observed 52% of variants classified as de novo and 48% inherited.

The median sizes for each CNV were further analyzed and a statistically significant difference was observed between primary and secondary CNVs ($p<0.001$). Focusing on primary CNVs, the median size was approximately 2.0 Mbp. The smallest variant in this category was just over 0.5 Mbp and the largest was 155 Mbp. Again, the largest CNV consistent with a sex chromosome aneuploidy of the X chromosome. There was an IQR of approximately 5 Mbp (6.3 Mbp-0.99 Mbp). Median sizes were investigated for each pathogenicity group. Pathogenic primary CNVs had a median size of 4.7 Mbp, LP variants were approximately 1.6 Mbp, VUS had a median size of 1.1 Mbp, and benign CNVs were 0.8 Mbp. The median size for secondary CNVs was smaller than primary CNVs by approximately 1.7 Mbp, with a median size 0.4 Mbp. The smallest variant was just 11 base pairs while the largest was 89 Mbp. There was an IQR of approximately 1 Mbp (1.3 Mbp-0.2 Mbp). Pathogenic variants had a median size of 1.7 Mbp, LP variants were approximately 0.5 Mbp, VUS had a median of 0.3 Mbp, and 0.16 Mbp for benign CNVs.

We compared variant classes between primary and secondary CNVs and did not observe a statistically significant difference in classification of variants between the groups ($p=0.093$). Primary CNVs were comprised of 45% deletions and 52% duplications. Secondary CNVs were made up of 48% deletions and 51% duplications. 3% and 1% of primary CNVs and secondary CNVs were composed of triplications, respectively. We then compared variant classes and pathogenicity categories for both, primary and secondary CNVs. 1,145 primary CNVs were analyzed and a statistically significant difference ($p<0.001$) was observed between variant class and pathogenicity.

85% of benign variants and roughly 60% of VUS's were classified as duplications while 62% of pathogenic variants were classified as deletions. A similar analysis was performed for 1,136 secondary CNVs and similarly a statistically significant difference was observed between variant classes and pathogenicity categories (p=0.003). Parallel to primary CNVs, 60% of pathogenic variants were deletions. Interestingly, 53% of likely pathogenic variants were classified as duplications while 46% were deletions. Benign variants made up 53% of duplications and 44% of deletions.

No significant difference was identified between the chromosome involvement of primary and secondary CNVs (p=0.070). For both primary and secondary CNVs, the most common chromosome to harbor a CNV was the X chromosome making up 7.6% and 8.9% of CNVs, respectively (Figure 6).

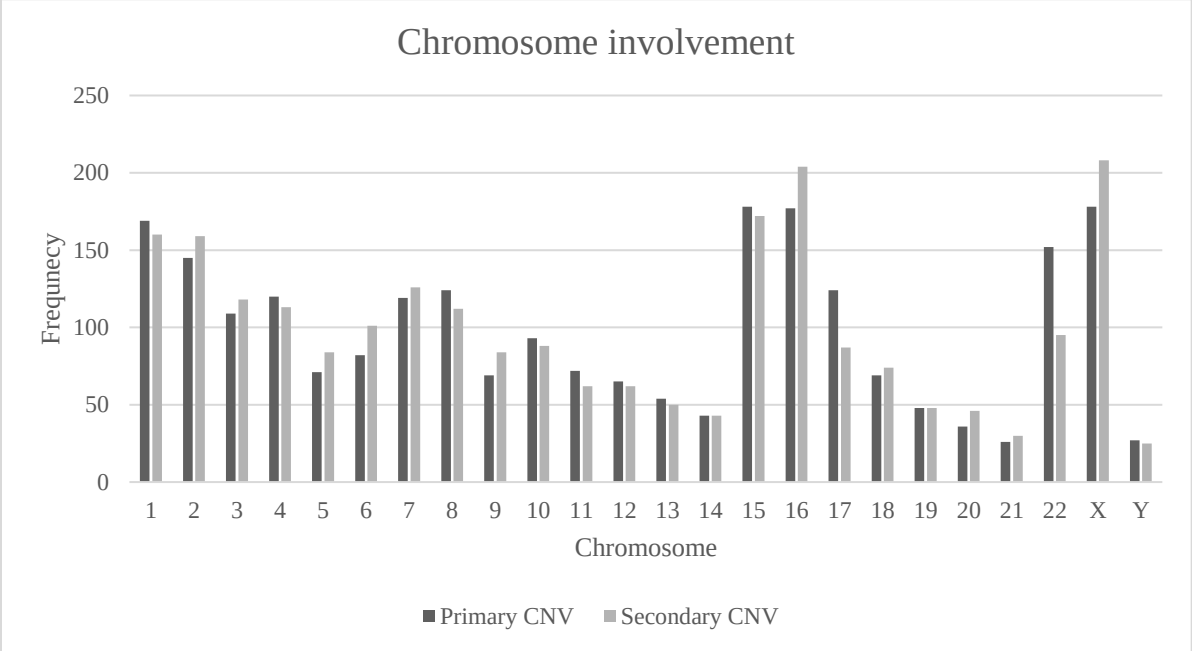


Figure 6. Chromosome involvement in primary and secondary CNVs. This figure compares the frequency of chromosome involvement for primary and secondary CNVs. The chromosome is listed on the x-axis, and the number of CNVs seen is labeled on the y-axis.

We created a map to identify recurrent combinations of chromosome involvement for primary and secondary CNVs (Figure 7a). Diagonally, we are observing a trend that secondary CNVs commonly fall in the same chromosome as the primary CNV. There is evident enrichment of secondary CNVs in chromosomes 15, 16, and the X chromosome. Marked recurrent chromosome combinations involving primary CNVs on an autosome include chromosomes 4/8, 12/2, and 20/4. The X chromosome has a noticeable pattern as common site for secondary CNV when the primary CNV is on chromosomes 3, 12, 16, and Y. When this map was recreated with the y-axis in descending order of relative frequency, we highlight the most common chromosomes for a secondary CNV to occur (7b). Although the Y chromosome is not highly involved in CNVs, we are observing a high combination of Y chromosome being involved in a secondary CNV, when the primary is also on the Y chromosome.

A

Secondary CNV chromosome \ Primary CNV chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
1	16.8	3.6	7.3	5.0	1.5	7.8	4.3	5.9	3.2	6.7	4.4	4.6	10.0	2.5	11.6	5.6	3.2	4.5	7.0	3.1	4.4	7.4	8.7	4.0
2	3.0	15.7	9.2	2.5	5.8	3.9	6.8	5.9	4.8	10.0	1.5	12.3	0.0	5.0	10.4	8.9	8.1	3.0	4.7	9.4	8.7	5.4	7.5	4.0
3	1.8	5.0	7.3	5.9	8.7	7.8	3.4	5.9	9.5	3.3	7.4	4.6	4.0	7.5	4.9	5.0	8.1	0.0	9.3	3.1	4.4	4.0	3.5	0.0
4	4.2	7.9	2.8	10.1	0.0	1.3	9.4	8.4	3.2	5.6	7.4	6.2	6.0	7.5	4.3	1.7	2.4	7.5	7.0	12.5	0.0	5.4	3.5	0.0
5	4.2	2.9	3.7	4.2	13.0	2.6	3.4	0.8	1.6	3.3	0.0	3.1	6.0	0.0	3.7	3.9	1.6	3.0	2.3	6.3	8.7	2.0	6.4	0.0
6	2.4	5.7	1.8	5.0	2.9	1.3	5.1	1.7	7.9	3.3	1.5	1.5	8.0	10.0	4.3	2.8	4.8	4.5	7.0	6.3	4.4	4.7	5.2	8.0
7	6.0	0.7	8.3	1.7	8.7	9.1	13.7	3.4	11.1	10.0	1.5	9.2	6.0	2.5	4.9	3.9	4.8	3.0	7.0	3.1	8.7	3.4	5.2	0.0
8	2.4	2.9	3.7	10.9	4.4	2.6	6.8	21.9	1.6	4.4	5.9	7.7	2.0	2.5	1.2	5.6	2.4	1.5	0.0	0.0	4.4	3.4	2.9	0.0
9	4.2	6.4	2.8	2.5	5.8	3.9	3.4	4.2	20.6	1.1	0.0	0.0	4.0	0.0	4.9	2.8	4.0	3.0	0.0	3.1	0.0	2.0	2.3	4.0
10	2.4	3.6	1.8	5.0	5.8	2.6	4.3	0.8	3.2	12.2	8.8	3.1	4.0	5.0	1.8	6.1	1.6	6.0	0.0	0.0	0.0	5.4	1.2	0.0
11	0.6	0.0	2.8	2.5	4.4	2.6	2.6	3.4	0.0	2.2	11.8	1.5	2.0	0.0	3.7	2.2	2.4	3.0	0.0	6.3	0.0	2.7	4.1	0.0
12	4.2	1.4	1.8	0.0	7.3	1.3	3.4	0.8	0.0	4.4	0.0	9.2	2.0	0.0	3.7	2.8	4.0	0.0	2.3	3.1	4.4	0.7	6.4	0.0
13	3.0	1.4	3.7	0.0	1.5	2.6	2.6	3.4	1.6	0.0	2.9	1.5	8.0	0.0	1.8	0.6	2.4	1.5	0.0	6.3	4.4	2.7	1.7	0.0
14	2.4	0.7	3.7	1.7	1.5	0.0	0.9	1.7	1.6	0.0	5.9	0.0	0.0	12.5	1.2	2.2	2.4	3.0	0.0	0.0	4.4	2.7	0.6	0.0
15	8.4	10.7	6.4	4.2	5.8	10.4	5.1	8.4	6.4	8.9	1.5	9.2	4.0	7.5	8.5	5.6	4.8	9.0	2.3	3.1	13.0	8.1	4.6	12.0
16	10.8	11.4	7.3	13.5	2.9	11.7	2.6	5.9	4.8	6.7	8.8	4.6	18.0	5.0	12.8	12.2	12.1	3.0	7.0	9.4	4.4	12.1	4.1	8.0
17	3.6	4.3	1.8	3.4	4.4	3.9	4.3	4.2	1.6	3.3	1.5	3.1	4.0	7.5	1.8	3.3	12.1	4.5	7.0	0.0	8.7	4.0	3.5	4.0
18	4.8	2.9	5.5	4.2	2.9	2.6	3.4	1.7	4.8	1.1	4.4	1.5	0.0	2.5	1.2	1.7	0.8	22.4	2.3	3.1	0.0	0.0	1.2	4.0
19	0.6	0.7	2.8	3.4	0.0	2.6	1.7	0.8	0.0	2.2	0.0	3.1	0.0	5.0	0.6	3.3	1.6	1.5	16.3	0.0	4.4	1.3	1.7	0.0
20	0.6	2.1	0.9	2.5	1.5	5.2	0.9	0.0	1.6	0.0	4.4	3.1	0.0	0.0	0.0	2.2	2.4	3.0	9.3	12.5	0.0	1.3	1.7	0.0
21	1.8	0.0	0.9	2.5	0.0	1.3	0.9	0.8	0.0	3.3	2.9	0.0	4.0	0.0	0.6	1.7	0.0	1.5	2.3	0.0	8.7	1.3	1.2	0.0
22	6.6	2.9	1.8	0.8	0.0	2.6	3.4	2.5	3.2	1.1	7.4	0.0	4.0	10.0	3.7	5.0	4.0	3.0	4.7	3.1	0.0	11.4	2.9	4.0
X	5.4	7.1	11.0	7.6	8.7	9.1	6.8	5.9	6.4	6.7	10.3	10.8	2.0	7.5	7.9	10.6	8.1	9.0	2.3	6.3	4.4	8.7	17.3	32.0
Y	0.0	0.0	0.9	0.8	2.9	1.3	0.9	1.7	1.6	0.0	0.0	0.0	2.0	0.0	0.6	0.6	1.6	0.0	0.0	0.0	0.0	0.0	2.9	16.0

B

Secondary CNV chromosome \ Primary CNV chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	Average %
X	5.4	7.1	11.0	7.6	8.7	9.1	6.8	5.9	6.4	6.7	10.3	10.8	2.0	7.5	7.9	10.6	8.1	9.0	2.3	6.3	4.4	8.7	17.3	32.0	**8.8**
16	10.8	11.4	7.3	13.5	2.9	11.7	2.6	5.9	4.8	6.7	8.8	4.6	18.0	5.0	12.8	12.2	12.1	3.0	7.0	9.4	4.4	12.1	4.1	8.0	**8.3**
15	8.4	10.7	6.4	4.2	5.8	10.4	5.1	8.4	6.4	8.9	1.5	9.2	4.0	7.5	8.5	5.6	4.8	9.0	2.3	3.1	13.0	8.1	4.6	12.0	**7.0**
2	3.0	15.7	9.2	2.5	5.8	3.9	6.8	5.9	4.8	10.0	1.5	12.3	0.0	5.0	10.4	8.9	8.1	3.0	4.7	9.4	8.7	5.4	7.5	4.0	**6.5**
1	16.8	3.6	7.3	5.0	1.5	7.8	4.3	5.9	3.2	6.7	4.4	4.6	10.0	2.5	11.6	5.6	3.2	4.5	7.0	3.1	4.4	7.4	8.7	4.0	**6.0**
7	6.0	0.7	8.3	1.7	8.7	9.1	13.7	3.4	11.1	10.0	1.5	9.2	6.0	2.5	4.9	3.9	4.8	3.0	7.0	3.1	8.7	3.4	5.2	0.0	**5.7**
3	1.8	5.0	7.3	5.9	8.7	7.8	3.4	5.9	9.5	3.3	7.4	4.6	4.0	7.5	4.9	5.0	8.1	0.0	9.3	3.1	4.4	4.0	3.5	0.0	**5.2**
4	4.2	7.9	2.8	10.1	0.0	1.3	9.4	8.4	3.2	5.6	7.4	6.2	6.0	7.5	4.3	1.7	2.4	7.5	7.0	12.5	0.0	5.4	3.5	0.0	**5.2**
6	2.4	5.7	1.8	5.0	2.9	1.3	5.1	1.7	7.9	3.3	1.5	1.5	8.0	10.0	4.3	2.8	4.8	4.5	7.0	6.3	4.4	4.7	5.2	8.0	**4.6**
8	2.4	2.9	3.7	10.9	4.4	2.6	6.8	21.9	1.6	4.4	5.9	7.7	2.0	2.5	1.2	5.6	2.4	1.5	0.0	0.0	4.4	3.4	2.9	0.0	**4.2**
17	3.6	4.3	1.8	3.4	4.4	3.9	4.3	4.2	1.6	3.3	1.5	3.1	4.0	7.5	1.8	3.3	12.1	4.5	7.0	0.0	8.7	4.0	3.5	4.0	**4.2**
22	6.6	2.9	1.8	0.8	0.0	2.6	3.4	2.5	3.2	1.1	7.4	0.0	4.0	10.0	3.7	5.0	4.0	3.0	4.7	3.1	0.0	11.4	2.9	4.0	**3.7**
5	4.2	2.9	3.7	4.2	13.0	2.6	3.4	0.8	1.6	3.3	0.0	3.1	6.0	0.0	3.7	3.9	1.6	3.0	2.3	6.3	8.7	2.0	6.4	0.0	**3.6**
9	4.2	6.4	2.8	2.5	5.8	3.9	3.4	4.2	20.6	1.1	0.0	0.0	4.0	0.0	4.9	2.8	4.0	3.0	0.0	3.1	0.0	2.0	2.3	4.0	**3.5**
10	2.4	3.6	1.8	5.0	5.8	2.6	4.3	0.8	3.2	12.2	8.8	3.1	4.0	5.0	1.8	6.1	1.6	6.0	0.0	0.0	0.0	5.4	1.2	0.0	**3.5**
18	4.8	2.9	5.5	4.2	2.9	2.6	3.4	1.7	4.8	1.1	4.4	1.5	0.0	2.5	1.2	1.7	0.8	22.4	2.3	3.1	0.0	0.0	1.2	4.0	**3.3**
12	4.2	1.4	1.8	0.0	7.3	1.3	3.4	0.8	0.0	4.4	0.0	9.2	2.0	0.0	3.7	2.8	4.0	0.0	2.3	3.1	4.4	0.7	6.4	0.0	**2.6**
11	0.6	0.0	2.8	2.5	4.4	2.6	2.6	3.4	0.0	2.2	11.8	1.5	2.0	0.0	3.7	2.2	2.4	3.0	0.0	6.3	0.0	2.7	4.1	0.0	**2.5**
20	0.6	2.1	0.9	2.5	1.5	5.2	0.9	0.0	1.6	0.0	4.4	3.1	0.0	0.0	0.0	2.2	2.4	3.0	9.3	12.5	0.0	1.3	1.7	0.0	**2.3**
13	3.0	1.4	3.7	0.0	1.5	2.6	2.6	3.4	1.6	0.0	2.9	1.5	8.0	0.0	1.8	0.6	2.4	1.5	0.0	6.3	4.4	2.7	1.7	0.0	**2.2**
19	0.6	0.7	2.8	3.4	0.0	2.6	1.7	0.8	0.0	2.2	0.0	3.1	0.0	5.0	0.6	3.3	1.6	1.5	16.3	0.0	4.4	1.3	1.7	0.0	**2.2**
14	2.4	0.7	3.7	1.7	1.5	0.0	0.9	1.7	1.6	0.0	5.9	0.0	0.0	12.5	1.2	2.2	2.4	3.0	0.0	0.0	4.4	2.7	0.6	0.0	**2.0**
21	1.8	0.0	0.9	2.5	0.0	1.3	0.9	0.8	0.0	3.3	2.9	0.0	4.0	0.0	0.6	1.7	0.0	1.5	2.3	0.0	8.7	1.3	1.2	0.0	**1.5**
Y	0.0	0.0	0.9	0.8	2.9	1.3	0.9	1.7	1.6	0.0	0.0	0.0	2.0	0.0	0.6	0.6	1.6	0.0	0.0	0.0	0.0	0.0	2.9	16.0	**1.4**

Figure 7. Chromosome combinations of primary and secondary CNVs for individuals. Chromosomes of the primary CNV are found along the top x-axis while secondary CNV chromosomes run along the y axis. The numbers inside boxes represent percentages of column totals. The color of the boxes intensity as the frequency of a chromosome pair increases. In panel A, the secondary CNV chromosomes are listed in numerical order. In panel B, the y-axis is arranged by average relative frequency of secondary CNVs, highlighting which chromosomes averaged more secondary CNVs.

Characterization of CNVs involving the sex chromosomes

We identified a significant difference between the median sizes of CNVs involving the autosomes and sex chromosomes (p=0.013). CNVs on the autosome had a larger median size than those on the sex chromosomes (p=0.0067). Next, the median sizes for primary CNVs were compared based on location of CNV (sex chromosome or autosome). We found no statistical difference in median sizes between the two (p=0.479). A primary CNV located on a sex chromosome was roughly 1.9 Mbp whereas primary CNV on an autosome was approximately 2.0 Mbp. A similar analysis was performed for secondary CNV and we identified a statistically significant difference in median sizes for secondary CNVs (p=0.029). Secondary CNVs had a larger median if found on an autosome (p=0.0104). The median size was 0.35 Mbp on the sex chromosomes. In contrast, the median sizes of secondary CNVs located on an autosome found the median size to be 0.44 Mbp.

We observed a statistically significant difference when comparing pathogenicity groups and location of CNV (p<0.001). VUS made up the largest percentage of pathogenicity groups for both autosomes and sex chromosomes, however makes up a greater proportion of CNVs on the sex chromosomes (Figure 8). CNVs on autosomes had a greater percentage of pathogenic/LP, while we see a slightly higher number of benign variants in sex chromosomes.

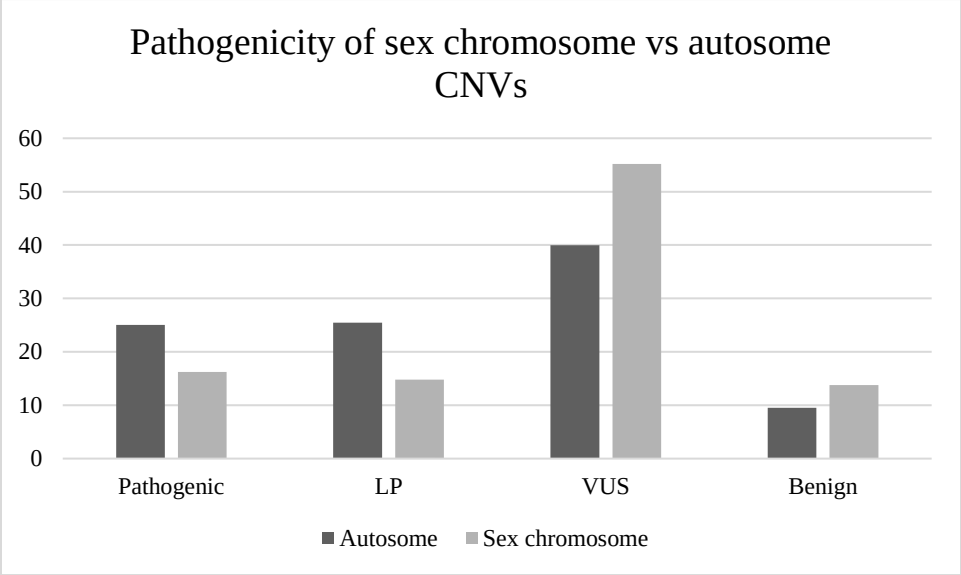


Figure 8. Comparison of pathogenicity groups for sex chromosomes and autosomes. Percentages of CNV pathogenicity classifications for sex chromosomes and autosomes. The x-axis represents the pathogenicity group while the y-axis represents the percentage of CNVs seen.

We also observed a statistical significance between pathogenicity and location of CNV for primary CNVs (p=0.015). There were 90 primary CNVs on the sex chromosomes that had pathogenicity information provided. 49% of those CNVs were classified as VUS. Breakdowns of other pathogenicity classifications include: 23% considered pathogenic, 20% LP, and 8% benign. We then compared this to 990 primary CNVs on the autosomes where 36% were classified as VUS, 34% pathogenic, 27% LP, and 3% benign. A similar analysis was performed for secondary CNVs. A statistical significance was again observed (p=0.003). There were 104 individuals with a secondary CNV with known pathogenicity on the sex chromosomes. Of this cohort, the majority (62%) had CNVs classified as VUS. Pathogenic variants made up 9% of the group, LP 10% and benign CNVs contributed 19%. We then compared this with 969 individuals who had a secondary CNV on an autosome where 46% were classified as

VUS, and 15%, 17%, and 22% made up the pathogenic, benign, and likely pathogenic groups, respectively.

When analyzing pathogenicity combinations, there were 412 females with a primary CNV on an autosome. The most frequent combination seen was a primary CNV classified as a VUS with a secondary CNV of the same classification (Figure 9a). The same analysis was performed for 43 individuals whose primary CNV fell on the X chromosome. Similarly to the previous analysis, the most frequent combination was having 2 CNVs which were both classified as VUS (Figure 9b).

A.

Secondary CNV classification	Primary CNV on autosome pathogenicity classification: Pathogenic	LP	VUS	Benign
Pathogenic	53	3	12	1
LP	28	54	11	5
VUS	45	36	93	1
Benign	32	6	29	3

B.

Secondary CNV classification	Primary CNV on X chromosome pathogenicity classification: Pathogenic	LP	VUS	Benign
Pathogenic	6	2	2	1
LP	0	2	2	0
VUS	3	2	17	0
Benign	1	0	4	1

Figure 9. Frequency of pathogenicity combinations for individuals with 2 CNVs. Panel A represents combinations when the primary CNV falls on an autosome. Panel B represent when the primary CNV falls on the X chromosome. The number of individuals observed are represented in the boxes.

A statistical significance of $p<0.001$ was observed between variant class and location of CNV. We then compared variant classes of primary CNVs to identify if there were differences in class based on location of CNV. A statistically significant difference

was observed (p=0.001). For primary CNVs which fell on a sex chromosome, 65% of CNVs were classified as duplications and 33% were deletions. Comparing this to 2,075 primary CNVs located on the autosomes where a more equal distribution of 51% classified as duplications and 47% deletions was identified. 3% of autosomal primary CNVs were categorized as triplications, and 1% of sex chromosome CNVs. Similarly, a statistical significance was observed (p<0.001) in variant classification of secondary CNVs and location of CNV. The majority (73%) of secondary CNVs on the sex chromosomes were classified as duplications with 25% being labeled as deletions, whereas autosomes harboring a secondary CNV had a more equal distribution of 48% duplications and 50% deletions.

When analyzing variant class combinations, there were 89 females who had a primary CNV on the sex chromosome and 907 individuals with a primary CNV on an autosome, for which variant classifications were known for both CNVs. We observed that regardless of where the primary CNV lies, approximately 1/3 of females in each group harbored a primary CNV classified as gain and secondary CNV classified as a loss (Figure 10).

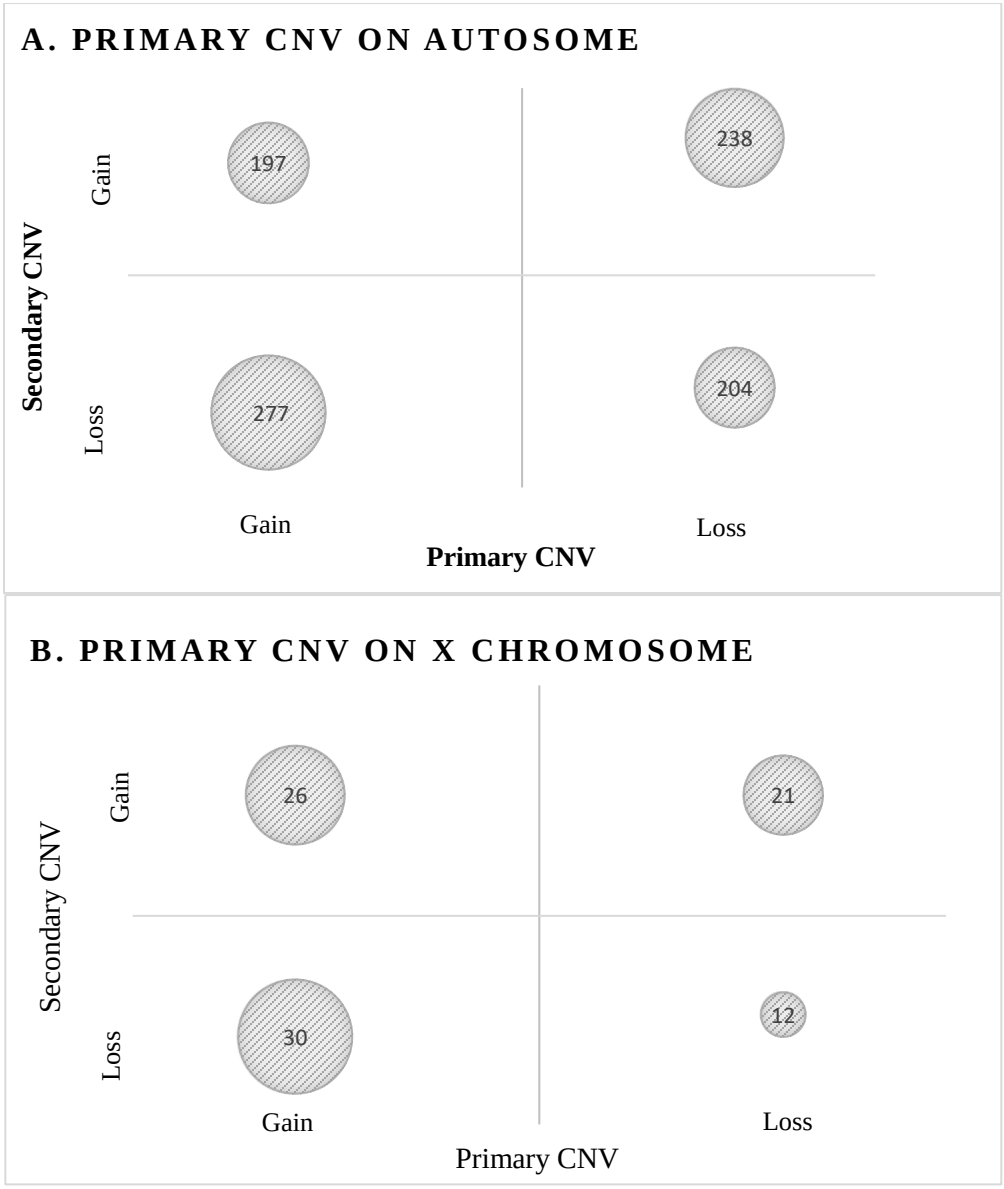


Figure 10. Variant class combinations. The variant class of the primary CNV is located on the x-axis with the secondary CNV on the y-axis. The number of times a specific combination was observed is listed within the bubble. The area of the bubble increases as the observed frequency increases. Panel A represents the combinations of variant class when the primary CNV falls on an autosome. Panel B represents the combinations of variant class when the primary CNV falls on the X chromosome.

Pathogenic X chromosome CNV

There were 12 females who had a pathogenic, primary CNV on the X chromosome. Ten of the 12 individuals had a primary CNV classified as a deletion. For

secondary CNVs, seven individuals had duplications and 5 had deletions. We identified that half of the secondary variants were defined as pathogenic while just 5 CNVs were larger than 0.5 Mbp. Six individuals had a phenotype ranging in severity of intellectual disability (ID).

ID	Size of primary CNV	Primary CNV variant class	Chromosome of secondary CNV	Size of secondary CNV	Variant class of secondary CNV	Pathogenicity of secondary CNV	Phenotype
1	155 Mbp	Deletion	13	18 Mbp	Deletion	Pathogenic	Skeletal abnormalities, short stature, learning disability
2	1.5 Mbp	Deletion	22	1.3 Mbp	Duplication	Pathogenic	Moderate ID
3	7.2 Mbp	Deletion	15	0.3 Mbp	Deletion	Pathogenic	Absent speech, ataxia, severe ID
4	17.8 Mbp	Duplication	3	5.9 Mbp	Deletion	Pathogenic	IUGR, micrognathia
5	15.7 Mbp	Deletion	9	9.9 Mbp	Duplication	Pathogenic	IUGR, micrognathia, cleft palate
6	83 Mbp	Duplication	19	0.023 Mbp	Deletion	Benign	Short stature
7	6.6 Mbp	Deletion	X	0.007 Mbp	Duplication	Benign	Abnormality of the face
8	155 Mbp	Deletion	2	0.07 Mbp	Deletion	VUS	Unknown
9	5.6 Mbp	Deletion	16	4.9 Mbp	Duplication	Pathogenic	Cognitive impairment, global DD
10	2.3 Mbp	Deletion	1	0.47 Mbp	Duplication	VUS	Mild ID, Short stature, skeletal abnormality
11	1.7 Mbp	Deletion	X	0.5 Mbp	Duplication	VUS	ID, obesity, disproportionate short-limb stature
12	10 Mbp	Deletion	1	0.31 Mbp	Duplication	VUS	Unknown

Table 3. Characteristics of secondary CNVs for primary X chromosome CNV. This table characterizes the secondary CNVs seen in 12 females with a pathogenic, primary CNV on the X-chromosome.

Of the 8 individuals with sex chromosome abnormalities (SCA) in our cohort, 5 had phenotypic information provided. Two patients fit the phenotype we would expect with SCA presenting previously described features of developmental delay, delayed speech, and hypotonia (Wattendorf and Muenke 2005, Tartaglia et al. 2010, Bardsley et al. 2013). The phenotype of an individual with 47,XXX (Triple X syndrome) was described as having seizures, severe ID, and tongue thrusting. It has not been previously reported that individuals with Triple X syndrome present with the above phenotype and is likely that the secondary CNV classified as a chromosome 4 duplication is causative of this presentation (Otter, Schrander-Stumpel, and Curfs 2010). Another individual with a SCA of 47,XYY was reported to have features uncommon of this condition including abnormalities of the face and skeletal system, microcephaly, and patent ductus arteriosus (Bardsley et al. 2013). Lastly, a male with 47, XXY was described as having dysmorphic features of low-set ears and clubbing of toes which has not previously been reported as a common characteristic.

Discussion

In an attempt to better characterize secondary CNVs involving the sex chromosomes, we described a large dataset of patients with two or more CNVs. Our analysis greatly focused on those with strictly two CNVs where at least one variant was greater than or equal to 500 kbp. Observation in the clinical setting and frequency of sex chromosome aneuploidies in live births, we hypothesized that secondary CNVs would most frequently involve a sex chromosome. We found that 9.2% of this cohort had a second variant in addition to a primary CNV. Our study demonstrates that the X chromosome is in fact more frequently involved in CNVs than other chromosomes for both primary and secondary CNVs in this cohort. In addition, we also observed that 2 CNVs occur frequently on the same chromosome, and that the increased frequency of Y-Y combination suggest variants are occurring on the same chromosome and not the homolog. Lastly, we identified that CNVs involving the sex chromosomes appear to behave differently than those on the autosome in a variety of categories. The observations of this study are an essential component in better understanding CNV genesis, and can be applied in the future towards proper counseling when additional CNVs are identified.

Along with X chromosome, we observed an enrichment of secondary CNVs in chromosomes 15 and 16. In a study by Girirajan et al. involving a cohort of individuals with developmental delay, there was a similar enrichment for secondary site variants on chromosome 16, although this was specific to known microdeletion/duplications (Girirajan et al. 2012). Here we are working with a larger cohort of affected individuals, without a specific indication and recognize a similar prevalence of chromosome 16 CNVs, overall, indicating this chromosome is frequently involved in CNVs regardless of

indication. This is likely due to the high density (8.2% intrachromosomal, 9.2% interchromosomal) of segmental duplications comprised of this chromosome, making it more susceptible to homologous recombination errors (Zhang et al. 2005). Similarly, chromosome 15 has a higher density of intrachromosomal segmental duplications comparative to others (5.5% intrachromosomal, 7.1% interchromosomal), likely explaining the higher frequency of CNVs seen on this chromosome (Zhang et al. 2005). What remains a question is the reason for increased involvement of the X chromosome in secondary CNVs. In contrast, the X chromosome has a density of approximately 2% intrachromosomal segmental duplications, and 5% interchromosomal which suggests that CNV formation by nonhomologous recombination may be the driving factor of structural change on the X chromosome, compared to autosomes (Zhang et al. 2005).

Of particular interest is the observed relationship between chromosome combinations of primary and secondary CNVs. This observed trend that the plurality of individuals have 2 CNVs on the same chromosome, and high frequency of Y-Y, combination could suggest that both CNVs are occurring on the same chromosome. Based on this, it is a possibility that CNVs larger than 500 kbp have an inefficiency of homologous and/or nonhomologous recombination mechanisms and increases the susceptibility of having an additional CNV on the same chromosome.

CNVs involving the X chromosome have been implicated as a cause of X-linked ID in males of families who had at least two males affected, with asymptomatic heterozygous females. (Whibley et al. 2010). In our study, we identified 12 females with a primary, pathogenic CNV involving the X chromosome, half of whom were described with some form or ID. We did not account for males in this analysis as an attempt to control for chromosome size and dosage. However, our results suggest two hypotheses:

CNVs on the X chromosome can also account for ID in females which may be exacerbated by skewed x-inactivation, or that females with 2 CNVs, one including the X chromosome, are at an increased sensitivity to ID when there is an additional CNV present.

Sex chromosome abnormalities (SCA), with the exception of Turner syndrome, often go undiagnosed related to a mild presentation. (Christian et al. 2000). However, if a SCA is suspected, a karyotype would be the most conventional test to order rather than CMA. In this study we identified 8 individuals with SCA, speculating that a more severe phenotype presented in these individuals. We can conclude that some individuals had a phenotype inconsistent with SCA, which strengthens conclusions drawn from previous research hypothesizing that the addition of CNVs are an independent factor contributing to intellectual disability and more distinct phenotypes in individuals with SCA and neurodevelopmental disorders (Le Gall et al. 2017).

Sex chromosome CNVs do behave differently than CNVs on the autosomes in a few categories. Median sizes of sex chromosome CNVs were smaller than those of autosomes. Specifically, secondary CNVs on the sex chromosomes are smaller than on the autosomes. One explanation may be that, given that the Y chromosome is one of the smallest chromosomes, median sizes are inherently more capable of being larger on autosomes and may explain the size comparison difference. When comparing variant classes between sex chromosome and autosome CNVs, we did observe a statistically significantly difference in variant class for both primary and secondary CNVs ($p<0.001$). For both groups of CNVs, we observed a higher proportion of sex chromosome CNVs to be classified as duplications, while autosomes had a more even distribution of deletions/duplications. A study by Whibley et al. identified that rare

CNVs on the X chromosome showed a slightly increased tendency for deletions, with a greater frequency of deletions found at smaller sizes while duplications dominated larger sized CNVs (Whibley et al. 2010). Duplications on the X chromosome were seen in greater frequency both at larger and smaller variants in our study, making us unable to corroborate this previous study (Whibley et al. 2010). Our study did have a smaller sample size of X chromosome CNVs as well as a broader phenotypic spectrum of individuals which may account for the inconsistencies. However, a similar conclusion as Whitley et al. can be drawn that duplications on the X chromosome may be more tolerable to human compatibility than deletions. Lastly, when analyzing pathogenicity of CNVs on sex chromosomes compared to autosomes, a significant difference for both primary ($p<0.001$) and secondary CNVs ($p<0.001$) was observed. In general, CNVs on the sex chromosomes were less likely to be pathogenic/LP when compared to the autosomes and more likely to be classified as a VUS. One thought is that the enrichment of VUS on the X chromosome can be attributed to x inactivation and therefore the uncertainty of contribution to the phenotype. This observation could also suggest that pathogenic variants on the sex chromosomes are more deleterious to human life compared to the autosomes. This analysis reinforces that there is still much to learn regarding copy variable regions of the sex chromosomes and to what extent they contribute to phenotype. For future directions, it would be essential to take a closer look at those individuals with sex chromosome CNVs and characterize their phenotype, which may lead to identification of copy number variable regions of interest.

We compared combinations of pathogenicity groups and variant classifications depending on whether a primary CNV fell on an autosome or sex chromosome. For pathogenicity combinations, regardless of where the primary CNV fell, most individuals

had two CNVs that were characterized as VUS. This finding may highlight the complexity that goes into classifying copy number variable regions and that more research is needed to elucidate pathogenicity and contribution to phenotypes. Next we looked at combinations of variant classes, again comparing combinations depending on if the primary CNV fell on a sex chromosome or autosome. We observed that regardless of where the primary CNV lie, most individuals had a primary CNV considered a gain, and a secondary CNV that was a loss. This illustrates that the most tolerated combinations of CNVs are those in which a gain is involved. These two analyses suggests that in individuals with two CNVs we can predict similar combination of pathogenicity and variant classes regardless if the primary CNV is on the sex chromosome or autosomes. .While this analysis of sex chromosome CNVs focused on females for simplification, it may have missed the true representation of CNVs on the sex chromosomes, specifically the X chromosome. It is possible that by including males in this analysis we would have identified more primary, pathogenic CNVs on the X chromosome, given that males are more commonly affected than females in X-linked conditions.

We observed a male bias towards multiple CNVs in this study that is worth acknowledging. It is widely accepted that chromosomal microarray analysis is a fist tier testing strategy for individuals with unexplained intellectual disability/developmental delay, autism spectrum disorders (ASD), or multiple congenital anomalies (Miller et al. 2010). Research has shown that there exists a prevalence of males in ASD with a ratio of 4 affected males to 1 affected female (Werling and Geschwind 2013). An explanation for our finding is due to the increased prevalence of ASD in males, requiring more males than females to undergo CMA. If this is the case, then our study is not capturing the true

prevalence of multiple CNVs in females and it cannot be concluded that males are truly more likely to harbor multiple CNVs.

Our results suggest that CNVs on the sex chromosomes appear to be a common finding in the clinical setting. It is documented that recurrent de novo and inherited CNVs located on the sex chromosomes have been implicated in individuals with disorders of sexual development, dysmorphic features, and neurologic disorders (Tannour-Louet et al. 2010, Kokalj Vokac et al. 2002, Whibley et al. 2010). However, given the large proportion of sex chromosome CNVs in this cohort, as well as lack of definite variant classification, it is possible that sex chromosomes play a larger role in phenotypic indications than originally thought. It would be important to identify how and to what extent CNVs on the sex chromosomes impact phenotype as this information could be imperative for patient management.

To our knowledge, these results are the first of its kind to describe the involvement of sex chromosomes for multiple CNVs. Strengths of this study include a large sample size allowing for a more accurate generalization. We identified two small cohorts (individuals with SCA plus additional CNV and individuals with pathogenic sex chromosome CNV) of individuals for which phenotypic implications may be drawn and used clinically. Observations of specific chromosome combinations involved in CNVs can be used in future studies to identify phenotypic presentations and applied to clinical settings. We do recognize the limitations in this study as there are no control samples to compare conclusions to, but it would be difficult to find a set on controls with multiple CNVs larger than 0.5 Mbp, as pathogenicity increases with increase CNV size. Limitations secondary to this type of data analysis include no information on the platform of the microarray used, which may be more or less sensitive to current

technology. CNVs falling within the pseudoautosomal region (PAR) are often displayed as a gain or loss on the X chromosome, and FISH technology must be used to determine which sex chromosome is involved. Without knowing whether or not a confirmatory analysis was performed, these results may suggest an enrichment on the X chromosome that is not present. Additionally, we recognized a large number of unknown information due to lack of submission from the referring center to DECIPHER. For sex chromosome CNVs, some studies were completed in only females to control for chromosome dosage and size and the same conclusions may not be applicable to males or all sex chromosome CNVs. It must be noted that rearrangements with an inverted duplication contiguous to a distal deletion exist and were not accounted for in this study. Therefore, enrichment of CNVs on the same chromosome may be present and future studies should consider removal of these variants, such that these CNVs are related to one another and generated by the same mechanism (Zuffardi et al. 2009, Bonaglia et al. 2009).

While this analysis is a step towards better understanding the general characteristics of secondary CNVs and sex chromosomes involvement in CNVs, future studies are indicated. Additional analysis of CNVs as functions of gene and segmental duplication densities may elucidate specific chromosome patterns. The inclusion of a control group for smaller CNVs can further extrapolate if there are certain chromosomes heavily involved in CNVs of unhealthy vs healthy individuals, perhaps highlighting unknown regions of importance in the genome. It would also be interesting to compare similar studies in males vs females investigating if characteristics are similar in both sexes. In general, future studies identifying patterns of CNV associations is imperative to better understanding phenotypes which can be used for more accurate counseling and patient care and to guide variant classification. Multiple studies have identified that the

addition of CNVs is implicated as a novel disease modifying mechanism (Prakash et al. 2016, Girirajan et al. 2012). Continuing to identify CNV combinations and specific patterns will highlight important areas of the genome and be key to understanding complex phenotypes.

Acknowledgements:

This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust. Those who carried out the original analysis and collection of the data bear no responsibility for the further analysis or interpretation of it by the Recipient or its Registered Users.

References:

Bardsley, Martha Zeger, Karen Kowal, Carly Levy, Ania Gosek, Natalie Ayari, Nicole Tartaglia, Najiba Lahlou, Breanna Winder, Shannon Grimes, and Judith L Ross. 2013. "47, XYY syndrome: clinical phenotype and timing of ascertainment." *The Journal of pediatrics* 163 (4):1085-1094.

Bonaglia, Maria Clara, Roberto Giorda, Angelo Massagli, Rita Galluzzi, Roberto Ciccone, and Orsetta Zuffardi. 2009. "A familial inverted duplication/deletion of 2p25. 1–25.3 provides new clues on the genesis of inverted duplications." *European Journal of Human Genetics* 17 (2):179-186.

Christian, SM, D Koehn, R Pillay, A MacDougall, and RD Wilson. 2000. "Parental decisions following prenatal diagnosis of sex chromosome aneuploidy: a trend over time." *Prenatal Diagnosis: Published in Affiliation With the International Society for Prenatal Diagnosis* 20 (1):37-40.

Girirajan, Santhosh, Catarina D Campbell, and Evan E Eichler. 2011. "Human copy number variation and complex genetic disease." *Annual review of genetics* 45:203-226.

Girirajan, Santhosh, Jill A Rosenfeld, Bradley P Coe, Sumit Parikh, Neil Friedman, Amy Goldstein, Robyn A Filipink, Juliann S McConnell, Brad Angle, and Wendy S Meschino. 2012. "Phenotypic heterogeneity of genomic disorders and rare copy-number variants." *New England Journal of Medicine* 367 (14):1321-1331.

Girirajan, Santhosh, Jill A Rosenfeld, Gregory M Cooper, Francesca Antonacci, Priscillia Siswara, Andy Itsara, Laura Vives, Tom Walsh, Shane E McCarthy, and Carl Baker. 2010. "A recurrent 16p12. 1 microdeletion supports a two-hit model for severe developmental delay." *Nature genetics* 42 (3):203.

Hastings, Philip J, James R Lupski, Susan M Rosenberg, and Grzegorz Ira. 2009. "Mechanisms of change in gene copy number." *Nature Reviews Genetics* 10 (8):551-564.

Itsara, Andy, Gregory M Cooper, Carl Baker, Santhosh Girirajan, Jun Li, Devin Absher, Ronald M Krauss, Richard M Myers, Paul M Ridker, and Daniel I Chasman. 2009. "Population analysis of large copy number variants and hotspots of human genetic disease." *The American Journal of Human Genetics* 84 (2):148-161.

Kokalj Vokac, N, P Seme Ciglenecki, A Erjavec, B Zagradisnik, and A Zagorac. 2002. "Partial Xp duplication in a girl with dysmorphic features: The change in replication pattern of late-replicating dupX chromosome." *Clinical genetics* 61 (1):54-61.

Le Gall, Jessica, Mathilde Nizon, Olivier Pichon, Joris Andrieux, Séverine Audebert-Bellanger, Sabine Baron, Claire Beneteau, Frédéric Bilan, Odile Boute, and Tiffany Busa. 2017. "Sex chromosome aneuploidies and copy-number variants: a further explanation for neurodevelopmental prognosis variability?" *European Journal of Human Genetics* 25 (8):930-934.

Miller, David T, Margaret P Adam, Swaroop Aradhya, Leslie G Biesecker, Arthur R Brothman, Nigel P Carter, Deanna M Church, John A Crolla, Evan E Eichler, and Charles J Epstein. 2010. "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies." *The American Journal of Human Genetics* 86 (5):749-764.

Nowakowska, Beata. 2017. "Clinical interpretation of copy number variants in the human genome." *Journal of applied genetics* 58 (4):449-457.

Otter, Maarten, Constance TRM Schrander-Stumpel, and Leopold MG Curfs. 2010. "Triple X syndrome: a review of the literature." *European Journal of Human Genetics* 18 (3):265-271.

Prakash, Siddharth K, Carolyn A Bondy, Cheryl L Maslen, Michael Silberbach, Angela E Lin, Laura Perrone, Giuseppe Limongelli, Hector I Michelena, Eduardo Bossone, and Rodolfo Citro. 2016. "Autosomal and X chromosome structural variants are associated with congenital heart defects in Turner syndrome: The NHLBI GenTAC registry." *American Journal of Medical Genetics Part A* 170 (12):3157-3164.

Tannour-Louet, Mounia, Shuo Han, Sean T Corbett, Jean-Francois Louet, Svetlana Yatsenko, Lindsay Meyers, Chad A Shaw, Sung-Hae L Kang, Sau Wai Cheung, and Dolores J Lamb. 2010. "Identification of de novo copy number variants associated with human disorders of sexual development." *PloS one* 5 (10).

Tartaglia, Nicole R, Susan Howell, Ashley Sutherland, Rebecca Wilson, and Lennie Wilson. 2010. "A review of trisomy X (47, XXX)." *Orphanet journal of rare diseases* 5 (1):8.

Templado, C, L Uroz, and A Estop. 2013. "New insights on the origin and relevance of aneuploidy in human spermatozoa." *Molecular human reproduction* 19 (10):634-643.

Tie-Lin Yang, Ruo-Han Hao, Yan Guo, Christopher J. Papasian, Hong-Wen Deng,. 2018. "Copy Number Variation." In *Genetics of Bone Biology and Skeletal Disease*, edited by Michael P. Whyte Rajesh V. Thakker, John A. Eisman, Takashi Igarashi, 43-54. Academic Press.

Wattendorf, Daniel J, and Maximilian Muenke. 2005. "Klinefelter syndrome." *American family physician* 72 (11):2259-2262.

Werling, Donna M, and Daniel H Geschwind. 2013. "Sex differences in autism spectrum disorders." *Current opinion in neurology* 26 (2):146.

Whibley, Annabel C, Vincent Plagnol, Patrick S Tarpey, Fatima Abidi, Tod Fullston, Maja K Choma, Catherine A Boucher, Lorraine Shepherd, Lionel Willatt, and Georgina Parkin. 2010. "Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability." *The American Journal of Human Genetics* 87 (2):173-188.

Zhang, Liqing, Henry HS Lu, Wen-yu Chung, Jing Yang, and Wen-Hsiung Li. 2005. "Patterns of segmental duplication in the human genome." *Molecular biology and evolution* 22 (1):135-141.

Zuffardi, Orsetta, M Bonaglia, R Ciccone, and R Giorda. 2009. "Inverted duplications deletions: underdiagnosed rearrangements??" *Clinical genetics* 75 (6):505-513.

Vita

Autumn Elizabeth Vara was born in Houston, Texas, the daughter of Felix Rivera Vara and Sharon Ann Hanka. She received her Bachelor of Science with a major in Biology and minor in Health from The University of Houston in 2015. Upon graduation, Autumn spent two years at Sagis Dermatopathology where she worked in the laboratory as a grossing technician. Following her work there, Autumn spent one year at Texas Children's Hospital cultivating her passion for genetic counseling as a genetic counseling assistant in the Cardiovascular Genetics Clinic. In August 2018, she entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences to pursue a Masters Degree in Genetic Counseling.