Integrated small copy number variations and epigenome maps of disorders of sex development

Ina E Amarillo1,2, Isabelle Nievera2, Andrew Hagan3, Vishwa Huchthagowder1, Jennifer Heeley2,4, Abby Hollander2,4, Joel Koenig2,5, Paul Austin2,5 and Ting Wang6

Small copy number variations (CNVs) have typically not been analyzed or reported in clinical settings and hence have remained underrepresented in databases and the literature. Here, we focused our investigations on these small CNVs using chromosome microarray analysis (CMA) data previously obtained from patients with atypical characteristics or disorders of sex development (DSD). Using our customized CMA track targeting 334 genes involved in the development of urogenital and reproductive structures and a less stringent analysis filter, we uncovered small genes with recurrent and overlapping CNVs as small as 1 kb, and small regions of homozygosity (ROHs), imprinting and position effects. Detailed analysis of these high-resolution data revealed CNVs and ROHs involving structural and functional domains, repeat elements, active transcription sites and regulatory regions. Integration of these genomic data with DNA methylation, histone modification and predicted RNA expression profiles in normal testes and ovaries suggested spatiotemporal and tissue-specific gene regulation. This study emphasized a DSD-specific and gene-targeted CMA approach that uncovered previously unanalyzed or unreported small genes and CNVs, contributing to the growing resources on small CNVs and facilitating the narrowing of the genomic gap for identifying candidate genes or regions. This high-resolution analysis tool could improve the diagnostic utility of CMA, not only in patients with DSD but also in other clinical populations. These integrated data provided a better genomic-epigenomic landscape of DSD and greater opportunities for downstream research.

Human Genome Variation (2016) 3, 16012; doi:10.1038/hgv.2016.12; published online 9 June 2016

INTRODUCTION

Over the past decade, the utilization of chromosome microarray analysis (CMA) as a clinical diagnostic tool has accumulated a massive amount of copy number variation (CNV) data, most of which are large or with sizes greater than the arbitrary set reporting limits, e.g., several hundreds of thousands of base pairs (kilobase/kb) to millions of base pairs (megabase/Mb). Although some of these CNVs have already been implicated in neurodevelopmental disorders and contiguous gene syndromes,1–5 many have remained as variants of uncertain clinical significance, with critical regions or causative genes yet to be identified. In clinical CMA, CNVs smaller than the analysis cutoff (e.g., ≤ 25–50 kb, ≤ 25 markers; referred to as ‘small CNVs’ hereafter) have either not been analyzed or not been reported. Thus far, only a few CMA studies have described the detection of such CNVs in human genetic disorders, e.g., 2.2–49 kb in idiopathic autism and/or intellectual disability,6 psoriasis,7 Crohn’s disease,8 autism spectrum disorder9 and congenital heart disease.10 Genome sequencing data have also been utilized in finding CNVs as small as 100 base pairs (bp) to 2.8 kb in patients with autism spectrum disorder and congenital heart disease.11–13 Despite such reports, small CNVs maps for clinical populations remain underrepresented in the existing databases and literature. In contrast, publicly available high resolution CNV maps (with CNVs as small as 50 bp) have been constructed and recently updated,13–15 providing comparable resources for the appropriate annotation and interpretation of small CNVs detected in clinical settings. With the availability of these genomic data, we are now poised to build small CNV maps that could aid in narrowing the genomic gap in human genetic disorders, including disorders of sex development (DSDs). DSDs constitute a group of rare and heterogeneous conditions characterized by atypical manifestations of chromosomal, gonadal and phenotypic sex, resulting in differences in the development of the urogenital and reproductive structures. DSDs are classified into three major groups; XX, DSD; XY, DSD; and sex chromosome DSD.14,15

In a significant portion of DSD cases, the genetic etiology remains elusive. Previous CMA studies in DSD have described a few small CNVs (0.2–50 kb),16–20 most of which have remained uncharacterized. Here, we focused our retrospective CMA on previously unanalyzed small CNVs and other genomic regions, including small regions of homozygosity (ROHs), imprinting and position effects. We used these high resolution data to investigate salient genomic features, e.g., CpG islands (CGIs), structural regions, functional domains and repeat elements, as well as the DNA methylation, histone modification and RNA expression profiles of these predicted regulatory regions in the normal testis and ovary.

MATERIALS AND METHODS

Ethics statement

This retrospective study was approved by the Institutional Review Board of the Washington University in St Louis School of Medicine.

1Cytogenomics and Molecular Pathology Laboratory, Division of Laboratory and Genomic Medicine, Department of Pathology and Immunology, Washington University in St Louis School of Medicine, St Louis, MO, USA; 2Washington University in St Louis School of Medicine DSD Team, St Louis, MO, USA; 3Division of Biology and Biomedical Sciences, Washington University in St Louis, St Louis, MO, USA; 4Department of Pediatrics, Washington University in St Louis School of Medicine, St Louis, MO, USA; 5Division of Biologic and Biomedical Sciences, Washington University in St Louis School of Medicine, St Louis, MO, USA and 6Department of Genetics, Washington University in St Louis School of Medicine, St Louis, MO, USA.

Correspondence: IE Amarillo (iamarillo@path.wustl.edu)

Received 15 January 2016; revised 24 March 2016; accepted 26 March 2016
Patient population
Our retrospective study was performed using CMA data obtained from 52 patients referred to our center from 2008 to 2015 for mild to severe DSDs, atypical secondary sexual development, Turner syndrome or stigmata and sex chromosome anomalies. XY and XX sex chromosome complements accounted for 63% and 21% of the patients, respectively (Table 1). Karyotypes with sex chromosome anomalies were also observed: 6% with loss of Y; 4% with XXY; and 6% with mosaicism, X/X(2) (n = 2) and X/XY (n = 1).

Clinical CMA platform and analysis software
Clinical CMA was performed on two Affymetrix (Affymetrix, Inc., Santa Clara, CA, USA) platform types (Table 1): Whole Genome-Human SNP Array 6.0 (40% of patients; 2008–2012; patients D1-D20, D53) and CytoScan HD (60% of patients; 2012–2015; D21-D52). The latter has a higher resolution (1 probe/880 bp; 2.6 million probes−1.9 million CN and non-polymorphic, 750,000 single nucleotide polymorphism (SNP) probes) than the former (approximately 1.8 million CN and SNP probes), and it encompasses almost all of known OMIM and RefSeq genes. CMA data were compared from control individuals (270 HapMap and 100 in-house) and from 380 phenotypically normal cohorts (284 HapMap and 96 blood samples from 186 female and 194 male patients). Copy number analysis was performed using the Affymetrix Genotyping Console and Chromosome Analysis Suite software programs. All of the genomic linear positions were based on human genome reference version UCSC Genome Assembly GRCh37/hg19.

Retrospective CMA approach
A total of 334 genes obtained from published clinical reports, reviews, CMA studies and the OMIM (Online Mendelian Inheritance in Man) database (http://www.omim.org/) were included in constructing a customized CMA ‘DSD Track’ (Supplementary Table S1; listed by size). These genes have been associated with the development of urogenital and other reproductive structures, as well as with sexual dimorphism. Their sizes and chromosome locations (in bp coordinates) were also noted. CMA data from 52 patients were acquired using a 1 kb filter and 1 probe (versus clinical limits of 25–50 kb, 25 probes). Gene-desert regions known to cis-regulate genes were used to build a ‘position effect track.’ Known imprinting genes (www.geneimprint.com) were also collated and used to build the ‘imprinting track.’ To filter the CNVs and ROHs overlapping with genes (DDGs, imprinted and position effects), the ‘overlap map’ function of the Chromosome Analysis Suite software (Affymetrix) was selected, which facilitated the filtration of CNVs and ROHs of interest from the remaining intervals not overlapping with DSD genes.

Probe coverage
The majority of the genes considered in this study with or without CNVs and/or ROHs showed good or proportional probe coverage (Supplementary Tables S3–S6 and S23). Although no SNPs or CN coverage was seen in approximately 12% (n = 24), multiple probes flanked these genes. Probe coverage was also verified from the Affymetrix website (https://www.affymetrix.com/analysis/netaffx/showresults.affx).

In silico investigations
All of the small CNVs and ROHs that overlapped with DSD genes were further analyzed for detailed genomic and epigenomic features using publicly available databases. We aligned the CNVs detected from our study with the following tracks from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway): RefSeq genes; common CNV data (inclusive and stringent data) by Zarrei et al.,13 obtained from DGV (database of genomic variants; http://dgv.tcag.ca/gb2/gbrowse/dgv2/hg19/); the ClinGen database, which curates benign, uncertain clinical significance and likely benign (UCS LB), UCS likely pathogenic (LP), UCS curated pathogenic genes: DECIPHER22 (https://decipher.sanger.ac.uk/); Segmental Duplications; CGIs, UniProt/SwissProt Protein and Secondary Annotations for regions and domains; and Transcription Factor ChiP for predicted transcription factor binding sites. DECIPHER was also investigated for CNVs overlapping with our data, including their sizes, sex chromosome complements, types of CNVs and proportions of small CNVs with DSD phenotypes. We also aligned our small CNV data set with the WashU Epigenome Browser23 (http://epigenomewidget.wustl.edu/browser/) to investigate the following features: RefSeq genes; CGIs; repeat elements; chromHMM of the ovary for predicted regulatory regions; H3K27Ac and H3K4me1 of the ovary (histone regions predicted to be regulatory); methylIC-seq of the fetal testis and ovary for DNA methylation (Cpg or regulatory); RNA-seq of fetal and adult ovaries for predicted RNA expression; and cis (local) and trans (interchromosomal) interactions (human foreskin fibroblast, high-content genome conformation capture; density map or circle plot representation).

RESULTS
Small CNV discoveries
We designed a customized CMA track targeting 334 genes associated with the development of urogenital and other reproductive structures (referred to as ‘DSD track’ and ‘DSD genes’ hereafter) (Supplementary Table S1). The majority of these genes are mapped on chromosomes X, 9 and 1, and approximately 60% of them are ≤ 50 kb (Supplementary Figures S1A and S2B). Among 14, using 1 kb as the lowest limit of detection, our retrospective CMA of 52 unrelated patients revealed a total of 12,576 CNVs: 6,475 losses (51%) and 6,141 gains (49%) (Table 1). We used the ‘overlap map’ function of the analysis software to expedite the filtration of 301 CNVs overlapping with 68 DSD genes from all of the others, reducing the number of variants for analysis by approximately a hundredfold (Tables 1 and 2, Supplementary Table S2). CN losses accounted for 53% of the CNVs (n = 162; 75% CN = 1, 25% CN = 0), while 47% were CN gains (n = 141; 42% CN = 3, 42% CN = 4, 16% CN = 2). Using the haploinsufficiency index scoring from DECIPHER,24 74% (n = 121) of the total CN losses involving 36 genes were predicted to be haploinsufficient, in the range of 0.06% (ESR1) to 21.2% (CAMK1D) (Supplementary Table S2). From the recently generated small CNV map obtained from normal populations,13 we compared the 21 homozygous losses (CN = 0) involving NOTCH1, SERT1, WWOX and DMD detected in our study with the null CNVs consisting of insignificant, enriched, paralogous CN variable regions, and none were found to be overlapping.

Our CMA approach uncovered 284 (94%) small CNVs involving 63 DSD genes (approximately 93%), and 16 (approximately 24%) of these genes were found to be ≤ 50 kb (Supplementary Table S2). The highest frequencies of CNVs were observed in DMD, WWOX and DHRSX (Table 1) and on chromosomes X, 16 and 9, while none were found on 11, 18, 21 and Y (Supplementary Figure S1A). Eight genes exhibited 14 CNVs with breakpoints disrupting exons (1.8–29.1 kb; 8–37 markers) that involve critical structural and functional regions or domains and biological processes related to sex development (Supplementary Table S3). Thirteen CNVs revealed breakpoints outside of an intact gene, 10 of which included 8 small genes, e.g., a 5.6 kb gain (NRPB1) (Figure 1a) to a 64 kb gain (GPRRI1) (Supplementary Table S2). Identical small CNVs involving 25 genes detected in two or more patients were observed (Table 1): four different recurrent CNVs in DMD (1.4–6.3 kb; 10–40 markers), three in WWOX (7.2–8.8 kb; 12–24 markers), two each in ESR2 (1.8–4.8 kb; 8–26 markers), NOTCH3 (2.8–19.8 kb; 9–46 markers) and TP63 (4.6–8.3 kb; 12–32 markers), and one in 19 genes (1.2–37 kb; 2–66 markers). CNVs with variable sizes (1.1–190 kb) but involved the same intron or exon were observed in 32 genes, while non-recurrent and non-overlapping CNVs were detected in 25 genes.

Genomic and epigenomic landscape
Of the 301 CNVs overlapping with DSD genes, 36% (n = 107) and approximately 60% (n = 184) accounted for CNVs with exonic and intrinsic breakpoints, respectively. Thirteen exonic CNVs involved exon 1 of 9 genes (2.5–95.7 kb; 5–47 markers) (Supplementary Figure S1A).
Table 1. Summary of CNVs and ROH overlapping with DSD genes in our patient cohort

| D# | DSD | Sex Chrom | Clinical | CMA Finding | Total CNV | Total CNVs Per Px | Total Ov CNVs Per Px | Total Loss Per Px | Total Ov Loss Per Px | Total Gain Per Px | Total Ov Gain Per Px | Total ROH Per Px | Total ROH Ov Genes Per Px | ROH |
|----|-----|-----------|----------|-------------|-----------|--------------------|----------------------|------------------|---------------------|-------------------|-------------------|----------------|--------------------------|------|
| D1 | SSC | XX N      |          |             | 12,576    | 67–817            | 373                  | 0–18             | 5                   | 160               | 4                 | 6475           | 44–295                    | 162  |
|    |     |           |          |             | 6,141     | 15–623            | 213                  | 0–16             | 1                   | 141               | 1                 | 3653           | 15–817                    | 141  |
| D2 | AG, E, UT, M, SA, KA | X X     |          |             | 376       | 4                  | 244                  | 0–12             | 2                   | 8                 | 1                 | 162           | 44–295                    | 162  |
| D3 | SSC | XX       | P        |             | 355       | 3                  | 119                  | 0–12             | 3                   | 1                 | 236              | 0              | none                      | 236  |
| D4 | H   | H Y V    | V        |             | 200       | 5                  | 127                  | 0–12             | 2                   | 73                | 2                 | 81            | 0–125                     | 81   |
| D5 | H   | H Y V    | V        |             | 158       | 2                  | 73                   | 0–12             | 1                   | 85                | 2                 | none                      | 85   |
| D6 | H   | H Y N    | N        |             | 256       | 3                  | 83                   | 0–12             | 3                   | 173              | 3                 | 142           | 0–115                     | 142  |
| D7 | AG, H, AD | SS X |          |             | 274       | 4                  | 144                  | 0–12             | 2                   | 130              | 3                 | none                      | 130  |
| D8 | SSC | SS X     | N        |             | 195       | 1                  | 70                   | 1                | none                | 125              | 0                 | none                      | 125  |
| D9 | H   | H Y V    | V        |             | 405       | 7                  | 215                  | 4                | none                | 190              | 2                 | FLNA, CHD7                | 192  |
| D10| AG, H, END | H XY | P        |             | 281       | 8                  | 198                  | 4                | none                | 83               | 4                 | none                      | 83   |
| D11| AG, H, END | H Y V | P        |             | 214       | 0                  | 88                   | 0                | none                | 126              | 0                 | none                      | 126  |
| D12| AG, H | H Y X    | N        |             | 339       | 9                  | 179                  | 5                | none                | 160              | 4                 | none                      | 160  |
| D13| H, M, UT, CH | H Y N |             |             | 410       | 7                  | 232                  | 5                | none                | 178              | 2                 | none                      | 178  |
| D14| AG, H | H Y X    | P        |             | 586       | 9                  | 221                  | 4                | none                | 365              | 5                 | none                      | 365  |
| D15| AG, H | H Y N    | N        |             | 711       | 10                 | 295                  | 4                | none                | 416              | 6                 | none                      | 416  |
| D16| H, SA | H Y     | N        |             | 353       | 10                 | 121                  | 4                | none                | 232              | 6                 | none                      | 232  |
| D17| AG, SA | H SSC |             |             | 333       | 7                  | 118                  | 1                | none                | 215              | 6                 | none                      | 215  |
| D18| GA, H | H Y X    | N        |             | 817       | 14                 | 194                  | 4                | none                | 623              | 13                | none                      | 623  |
| D19| H, SA | H Y N    | N        |             | 366       | 3                  | 97                   | 0                | none                | 269              | 3                 | none                      | 269  |
| D20| CH, H | H Y V    | V        |             | 279       | 4                  | 125                  | 3                | none                | 154              | 1                 | none                      | 154  |
| D21| H, SA | H Y V    | V        |             | 800       | 18                 | 177                  | 2                | none                | 623              | 2                 | none                      | 623  |
| D22| END | X X P    | P        |             | 134       | 3                  | 100                  | 3                | none                | 34               | 0                 | none                      | 34   |

Integrated small CNVs and epigenome maps of DSD genes

IE Amarillo et al

© 2016 The Japan Society of Human Genetics

Human Genome Variation (2016) 16012

3
| D# | DSD | Sex Chrom | Clinical CMA Finding | Total CNV Per Px | Total Ov CNVs Per Px | Total Loss Per Px | Total Ov Loss Per Px | Loss Total Gain Per Px | Total Ov Gain Per Px | Gain Total ROH Per Px | Total ROH Ov Genes Per Px | ROH |
|----|-----|-----------|----------------------|------------------|---------------------|------------------|---------------------|----------------------|---------------------|----------------------|------------------------|------|
| D23 | AG, H, END H, UT | XY N | 75 | 3 | 59 | 2 | DMD, POR | 16 | 1 | DMD | 79 | 8 | FRAS1, TP63, WWOX |
| D24 | AG, H, END H, UT | XY N | 104 | 3 | 89 | 3 | FRAS1, INSR, PTPN11 | 15 | 0 | none | 83 | 4 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D25 | H, M, SA | XY N | 67 | 7 | 46 | 5 | DHRSX (2), NRSA2 (2), SOX9 | 21 | 2 | DHRSX, NLGN4X | 80 | 16 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D26 | AG, SA, M, H | XY N | 110 | 4 | 84 | 4 | DHRSX (2), FRAS1, WWOX | 26 | 0 | none | 82 | 14 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D27 | SSC | X P | 177 | 4 | 117 | 1 | CTNNA3 | 60 | 3 | PIP5K1B, TP63, ESR1 | 86 | 4 | OCA2, PCDH17, OCA2, CYP26B1, GHR, WNT2, PCDH17, FGR1, CYP19A1 |
| D28 | GA | XY N | 90 | 3 | 69 | 1 | NRS2A | 21 | 2 | DMD, NR0B1 | 96 | 6 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D29 | AG, C | XX V | 161 | 13 | 130 | 9 | DMD (6), NLGN4X, NOTCH1, WWOX, SRDS5A2, ESR1 | 31 | 4 | DMD (2), EP300, NLGN4X | 29 | 3 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D30 | AG, UT | XXV Y | 164 | 9 | 69 | 2 | HHA1 | 95 | 7 | ATRX, DMD (5), NR0B1 | 102 | 3 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D31 | H, KA | XXV Y | 96 | 6 | 62 | 1 | HHA1 | 34 | 5 | none | 50 | 8 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D32 | M | XY V | 74 | 4 | 58 | 4 | DHRSX, DMD, NOTCH1, WWOX | 16 | 0 | none | 19 | 3 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D33 | TS | X P | 133 | 2 | 99 | 1 | AXIN1, WNT2 | 34 | 1 | WWOX | 79 | 6 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D34 | SSC, TS | XX N | 232 | 10 | 138 | 2 | AXIN1, WNT2 | 94 | 8 | DMD (4), ESR1, NLGN4X, PBX1, TP63 | 111 | 15 | PTPN11, CYPR681, RHOA, WNT2, PCDH17, FGR1, CYP19A1 |
| D35 | UT, END | XY N | 126 | 1 | 94 | 1 | MOVI101 | 32 | 0 | none | 67 | 11 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D36 | UT, KA | XY N | 128 | 4 | 84 | 1 | VAMP7 | 44 | 3 | BNC2, DHRSX, DMD | 78 | 6 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D37 | AG, H, SA | END XY N | 124 | 4 | 94 | 1 | MOVI101 | 30 | 3 | ATRX, CTNNA3, DMD | 75 | 6 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D38 | TS | X P | 177 | 2 | 124 | 2 | SRDS5A2, WWOX | 41 | 1 | PBX1 | 14 | 0 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D39 | TS | X P | 177 | 2 | 124 | 2 | SRDS5A2, WWOX | 53 | 0 | none | 80 | 10 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |
| D40 | AG, H, AD, UT, M, SA | XY N | 245 | 5 | 188 | 4 | DHRSX, DMD (2), NLGN4X | 57 | 1 | KAL1 | 37 | 16 | PTPN11, CYPR681, RHOA, WNT2, STAR, FGR1, CYP19A1 |

Integrated small CNVs and epigenome maps of DSD
IE Amarillo et al.
Human Genome Variation (2016) 16012 © 2016 The Japan Society of Human Genetics
**Table 1.** (Continued)

| D#  | DSD | Sex Chrom | Clinical CMA Finding | Total CNV Per Px | Total CNV Per Px | Total Loss Per Px | Total Gain Per Px | Loss | Total Gain Per Px | Gain | Total ROH Per Px | Total ROH Ov Genes Per Px | ROH |
|-----|-----|-----------|----------------------|------------------|------------------|------------------|------------------|-----|------------------|-----|------------------|-----------------------------|-----|
| D41 | AG, M, CH, H | XY | N | 169 | 6 | 103 | 5 | DHRSX, DMD (2), WWOX, NOTCH2, AXIN1, DMD (2), ESR2, FRAS1, TP63 | 66 | 1 | TP63 | 87 | 7 | EP300, PCDH17, PTPN11, SRD5A2, BBS4, STAR, FGFR1, PTPN11 |
| D42 | H, UT | XY | V | 267 | 13 | 199 | 7 | DMD, ESR2, OCA2, PBX1, PTPCH1, TP63 | 68 | 6 | none | 16 | 1 |
| D43 | H, UT | XY | N | 92 | 4 | 63 | 3 | CTNNB1, SANKP1, WWOX | 29 | 1 | TP63 | 78 | 10 | AHR, PTPN11, PTC1, STAR, FGFR1, SOX2, SOX5, HSD3B1, HSD3B2, NOTCH2, AR, ATRX, CTNNB1, FN1, FBG1, RHOA, GATA4, PTPN11, E6, MAMLD1 |
| D44 | TS | XX | V | 213 | 10 | 142 | 4 | AXIN1, DMD (2), NLGN4A, TP63, WWOX | 71 | 6 | DMD (3), ESR2, PBX1, ARX | 77 | 6 |
| D45 | GA, KA | XX | V | 126 | 4 | 94 | 2 | EP300, OCA2, SRD5A2, ARX, GATA4, RHOA, ARX, MAMLD1, PTPN11, SRD5A2, FN1A, FN1B, GATA4, PTDG2, MAPK3 |
| D46 | AG, UT, M, CH | X/XY | P | 189 | 8 | 163 | 8 | CTNNA3, DHRSX (2), MOV10L1, PIP5K1B, SHOX, WAP5, SORX9 | 26 | 0 | none | 93 | 8 |
| D47 | H, CH, UT | AG | N | 98 | 0 | 70 | 0 | none | 28 | 0 | none | 75 | 10 | OCA2, PTPN11, CTNN1, SUPT3H, BBS1, BBS4, FMRM2, STAR, FGFR1, ARX |
| D48 | H, CH, UT | AG | N | 192 | 10 | 126 | 7 | DHRX, DMD (4), NLGN4A, SUPT3H | 66 | 3 | EP300, PBX1, TP63 | 23 | 3 | OCA2, PIP5K1B, CTNN1 |
| D49 | H, CH, UT | AG | N | 98 | 0 | 70 | 0 | none | 28 | 0 | none | 75 | 10 | OCA2, PTPN11, CTNN1, SUPT3H, BBS1, BBS4, FMRM2, STAR, FGFR1, ARX |
| D50 | C, UT, END | AD, M, H, UT | XX | N | 125 | 1 | 94 | 1 | DMD | 31 | 0 | none | 71 | 5 |
| D51 | AD, M, H, UT | XX | V | 149 | 4 | 105 | 4 | AXIN1, DHRX, DMD, FGFR | 44 | 0 | none | 108 | 14 | NOTCH1, PTPN11, AR, ATRX, FLNA, FGFR1, PTPCH1, VAMP, ARX, BBS1, CYP2B11, GHR, MAMLD1, PTDG5, NOTCH4, CYP21A2, RHOA |
| D52 | AG, C, UT | XX | N | 217 | 13 | 166 | 12 | AXIN1, BNC2, DMD (7), NOTCH1, TP63, SORX9 | 51 | 1 | DMD | 25 | 10 | PTPN11, PCDH17, OCA2, AR, ATRX, FLNA, VAPM, FGFR1, ARX |
| D53 | M | XX | N | 78 | 0 | 44 | 0 | none | 34 | 0 | none | 85 | 9 | PTPN11, PCDH17, OCA2, AR, ATRX, FLNA, VAPM, FGFR1, ARX |

(i)—number in parenthesis indicates ≥ 1 CNV per gene per patient (loss/loss, gain/gain). *genes with both CNV and ROH. Clinical CMA Finding: N, normal; V, variant of uncertain clinical significance; P, pathogenic. DSD—atypical sex development findings. Abbreviations: AD, adrenal anomalies; AG/GA, ambiguous genitalia or genital anomalies; CH, chordee; CMA, chromosome microarray analysis; CNV, copy number variation; DSD, disorders of sex development; END, endocrine findings; H/E, hypospadias or epispadias; KA, kidney anomalies; M/C, micro penis or clitoromegaly; ROH, region of homozygosity; SA, scrotal anomalies; SSC, secondary sexual characteristics; TS, Turner syndrome or stigmata; UA, urethral anomalies; UT, undescended testis or inguinal gonads. *gene with both CN loss and gain in a patient. *gene with both CNV and ROH in a patient.
### Table 2: Summary of DSD genes with recurrent and overlapping CNVs and ROH

| Gene  | Chrom Locus | Gene Size (Kb) | n Px ROH | CNV with ROH | n Px CNV | 0–29 | # Recurrent CNV | CNV Range | Same Exon/Intron CNV (n Px)b | Non-OvC CNV | OvN with Intact Genea |
|-------|-------------|----------------|----------|--------------|----------|------|----------------|-----------|-------------------------------|-------------|------------------------|
| DMD   | Xp21.2      | 2,220.4        | 5–2,220  | 1–32         | 4        | 1    | 1; exons 8–9 (2), 48 (3); 2; exons 1 (4), 2 (8), 7 (4), 9 (4), 44 (4), 55 (4), 59 (6), 60 (6), 62 (2), 74 (2) | none       | 3; exons 18, 40–71, 75–76    | none       | none                   |
| WWOX  | 16q23.1     | 1,113.2        | 22       | 27           | 21       | 3    | 1; inton 5 (7, 2) | none      | 3; inton 5 (20), 8 (3)       | none       | none                   |
| NOTCH3| 19p13.12    | 41.3           | 8        | 8            | 0        | 3    | 1; exons 1 (3), 2–20 (2) | none      | 1; exons 1–33 (6)            | none       | none                   |
| TP53  | 3q28        | 256.9          | 12       | 13           | 7        | 1    | 3; exons 5–7 (4), intron 1 (2) | none      | 3; exons 5, 9 (4), 13 (3)    | none       | none                   |
| ESR2  | 14q23.2     | 111.5          | 10       | 13           | 6        | 3    | 2; exon 9 (2), disrupted exon 13 (2) | none      | 2; exon 9, intron 9         | none       | none                   |
| AXIN1 | 16p13.3     | 65.2           | 5        | 5            | 0        | 1    | 1; exon 5 (3)       | none      | 1; intron 4                  | none       | none                   |
| KRAS  | 12p12.1     | 45.7           | 5        | 4            | 0        | 4    | 1; disrupted exon 5 (2), 4; exon 5 (3) | none      | 1; disrupted exon 4, 5       | none       | none                   |
| TGFb1 | 19q13.2     | 23             | 2        | 2            | 0        | 2    | 1; exon 6–7 (2)   | none      | none                        | none       | none                   |
| EP300 | 22q13.2     | 87.5           | 15       | 3            | 1        | 2    | 1; exon 7–8 (2)   | none      | none                        | none       | none                   |
| CTNNB1| 3p22.1      | 41             | 5        | 5            | 0        | 5    | 1; exon 9–15 (2)  | none      | 1; intron 1                  | none       | none                   |
| PTPTN1| 12q24.13    | 91.2           | 32       | 6            | 1        | 5    | 1; exon 11–16     | none      | 1; intron 1                  | none       | none                   |
| PDGFRA| 4q12        | 69.4           | 4        | 4            | 0        | 4    | 1; exon 12–23 (4), disrupted exon 23 | none      | none                        | none       | none                   |
| MOV101| 22q13.33    | 71.7           | 6        | 6            | 6        | 0    | 1; exons 16–19 (4) | none      | none                        | none       | none                   |
| FRAS1 | 4q21.21     | 486.7          | 5        | 4            | 4        | 0    | 1; exon 30 (4)    | none      | none                        | none       | none                   |
| BNC2  | 9p22.22     | 461.3          | 4        | 4            | 3        | 1    | 1; intron 1 (2)   | none      | 1; intron 1                  | none       | none                   |
| NLGN4K| 22p23.32    | 336.6          | 8        | 9            | 4        | 5    | 2; introns 2 (5), 3 (3) | none      | none                        | none       | none                   |
| NOTCH1| 9q34.3      | 51.3           | 7        | 3            | 3        | 0    | 1; exon 2 (3)     | none      | 1; intron 1                  | none       | none                   |
| PBX1  | 1q23.3      | 292.5          | 5        | 5            | 0        | 5    | 1; exon 2 (4)     | none      | 1; intron 2                  | none       | none                   |
| DHRXS | 22q23.3     | 322.5          | 12       | 15           | 11       | 4    | 1; exon 3 (2); introns 9, 3 (5) | none      | none                        | none       | none                   |
| PCDH17| 13q21.1     | 97.3           | 11       | 2            | 2        | 2    | 1; exon 3 (2)     | none      | 1; intron 3                  | none       | none                   |
| AHRR  | 5p15.33     | 134.1          | 6        | 4            | 4        | 0    | 1; exon 5 (2)     | none      | 1; intron 5                  | none       | none                   |
| PIP5K1| 19q21.11    | 303.8          | 5        | 4            | 1        | 3    | 1; exon 2 (4)     | none      | 2; exon 13 (1), intron 9 (1) | none       | none                   |
| OCA2  | 15q13.1     | 344.4          | 24       | 5            | 0        | 5    | 1; intron 19 (3)  | none      | 2; introns 2, 23             | none       | none                   |
| SOX9  | 17q24.3     | 54             | 4        | 4            | 1        | 0    | 1; 6 Kb (3)       | none      | 1; 1.6 Kb                  | none       | none                   |
| SRD5A2| 2p23.1      | 56.4           | 11       | 2            | 2        | 2    | 1; exon 4-6 (2)   | none      | 6; disrupted exon 24         | none       | none                   |
| DHRXS | 22q23.3     | 281.5          | 12       | 15           | 11       | 4    | 1; 1.4–1.4–1.4; 12–106 | none      | 3; exon 3 (2); introns 1 (9, 3 (5) | none       | none                   |
| CXCL12| 1q11.21     | 14.041         | 3        | 2            | 0        | 2    | 1; exon 4 (2), disrupted | none      | none                        | none       | none                   |
| SHOX  | Xp22.33     | 35.1           | 3        | 3            | 1        | 2    | 1; exon 6 (2)     | none      | 1; exon 6 (2)               | none       | 1; D10                   |
| GFRA1 | 10q25.3     | 216.7          | 5        | 2            | 0        | 2    | 1; exon 5 (2)     | none      | none                        | none       | none                   |
| FLNA  | Xq28        | 26.1           | 8        | 3            | 0        | 3    | 1; 6–48 (3), disrupted exons 22, 48 | none      | none                        | none       | none                   |
| Gene     | Chrom Locus | Gene Size (Kb) | n Px | CNV | ROH | total CNV | Loss | Gain | CN State | CNV Size (Kb) | Probes | # Recurrent CNV (n Px) | ROH | Same Exon/Intron CNV (n Px) | Non-Ov,a CNVs | CNV with Intact Gene |
|----------|-------------|----------------|------|-----|-----|-----------|------|------|----------|----------------|--------|------------------------|-----|-------------------------|----------------|----------------------|
| AR       | Xq12        | 186.6          | 9    | 2   | 1   | 1         | 1, 2 | 1, 2 | 8.4–190.5 | 16–86         | 7      | none                   | 1; intron 1 (2) | none                 | none            |
| FAM167A-AS1 | 8p23.1      | 70.3           | 2    | 2   | 2   | 0         | 1    | 1    | 15.9–19.2| 18–20         | 0      | none                   | 1; intron 1 (2) | none                 | none            |
| RAF1     | 3p25.2      | 80.6           | 2    | 3   | 1   | 2         | 1, 4 | 3.9–13.6| 11–32    | 0             | none              | 1; intron 1 (1) | 1; disrupted exon 22 (ZNF557) | none            |
| INSR     | 19p13.2     | 181.746        | 2    | 2   | 2   | 0         | 1    | 7–29.1| 5–23     | 0             | none              | 1; intron 2 (2) | none                 | none            |
| PTC1     | 9p22.32     | 65.6           | 8    | 6   | 0   | 6         | 4    | 3    | 3.7–16.4| 13–40         | 4      | none                   | 1; intron 2 (3) | 3; exons 1–2, 11–12, 14–15 | none            |
| SNRPN    | 1q11.2      | 154.9          | 2    | 2   | 0   | 2         | 3    | 14.8–36.5| 17–34    | 0             | none              | 1; intron 2 (2) | none                 | none            |
| SPT3H    | 6p21.1      | 551.3          | 7    | 2   | 2   | 0         | 0, 1 | 9.8–10| 8–16     | 5             | none              | 1; intron 2 (2) | none                 | none            |
| RAF1     | 3p25.2      | 80.6           | 2    | 3   | 1   | 2         | 1, 4 | 3.9–13.6| 11–32    | 0             | none              | 1; intron 6 (3) | 2; exon 6, intron 5 | none            |
| CTNNA3   | 10q21.3     | 1776.2         | 14   | 7   | 4   | 3         | 3    | 4    | 6.6–92.2| 5–104         | 7      | none                   | 1; intron 7 (2) | 5; exon 10, introns 5, 9, 12, 13 | none            |
| NR5A2    | 1q32.1      | 149.8          | 4    | 3   | 4   | 3         | 1    | 1.1–4.5| 6–14     | 2             | none              | 1; intron 7 (2) | none                 | none            |
| ESR1     | 6q25.1      | 296            | 3    | 2   | 1   | 1         | 1, 3 | 25–28| 8–28     | 1             | none              | 1; intron 7 (2) | none                 | 1 (D10)         |
| ATRX     | Xq21.2      | 281.4          | 10   | 3   | 0   | 3         | 2, 3 | 2.2–5.2| 10–12    | 7             | none              | 1; intron 23 (2) | none                 | none            |
| GRIP1    | 12q14.3     | 331.7          | 4    | 3   | 2   | 1         | 1, 4 | 8.4–23.9| 11–15    | 1             | none              | none                   | 3; exon 14, introns 1, 5 | none            |
| CNTNAP3  | 9p13.1      | 215.5          | 4    | 3   | 0   | 3         | 3    | 95.7–374.0| 17–66    | 1             | none              | none                   | 2; exon 1–2, 19–24 | 1 (D18)         |
| VAMP7    | Xq28        | 625            | 6    | 2   | 2   | 0         | 1    | 1.5–94| 5–153    | 4             | none              | none                   | 2; exon 6–8 (3), intron 6 | none            |
| FGFR2    | 10q26.13    | 120.1          | 2    | 2   | 1   | 1         | 1, 4 | 4–6   | 6–9      | 0             | none              | none                   | 2; exons 14–15, intron 5 | none            |
| FLT1     | 13q12.2     | 194.8          | 1    | 2   | 2   | 0         | 1    | 2.4–11.4| 5–8      | 0             | none              | none                   | 2; exon 24, intron 3 | none            |
| ROCK2    | 2p25.1      | 163            | 2    | 2   | 2   | 0         | 1    | 28.1–42.2| 15–19    | 0             | none              | none                   | 2; exons 31–33 (PQLC3), intron 1 | none            |
| CHD7     | 8q12.2      | 189.3          | 4    | 2   | 1   | 1         | 1, 4 | 2–11.4| 3–9      | 2             | none              | none                   | 2; exons 35–38, intron 1 | none            |
| NROB1    | Xp21.2      | 80.6           | 2    | 2   | 2   | 0         | 1    | 4.9–5.6| 32–39    | 0             | none              | 1; exon 1–2, disrupted exon 1 | none            |
| PCSK1    | 5q15        | 42.5           | 2    | 2   | 2   | 0         | 1    | 24.8| 18–0     | none           | none              | 1; exon 1–3 | 1 (D9)             | none            |
| ANOS1 (KAL1) | Xp22.31  | 203.3          | 2    | 2   | 1   | 1         | 1    | 2    | 7–18     | 0             | none              | 1; intron 2 | 1 (D3)             | none            |

Abbreviations: CNV, copy number variation; ROH, region of homozygosity. a n Px—number of patients with both CNV and ROH. b# number of recurrent/overlapping. c number of non-overlapping CNVs; specific exon/intron involved in CNV (number of patients with specific CNV). dCNVs that include an intact smaller gene. e gene names in parenthesis—old name.
Segmental duplications and repeat elements have been exapted as binding sites for transcription factors. Non-coding repeats and regions of low complexity are prone to non-allelic homologous recombinations, often resulting in recurrent, large segmental duplications (segdups). These segdups were found in 16 genes that involved homologous recombinations, often resulting in recurrent, large segmental duplications (segdups). These segdups were found in 16 genes that involved homologous recombinations, often resulting in recurrent, large segmental duplications (segdups). Therefore, these genes may have been selected for functional diversity, potentially providing candidate targets for future studies.
factors, acting as alternative and differentially methylated promoters or enhancers\textsuperscript{31–33} dysregulating the host's gene expression machinery. Non-allelic homologous recombinations involving such repeats have also been described to result in recurrent but smaller deletions/duplications in various human genetic disorders.\textsuperscript{30,34,35} All of the CNVs detected in our study showed variable distribution of these smaller repeat elements: RNA repeats, AT-rich regions, simple repeats or microsatellites, DNA transposons and retroelements, such as long terminal repeats, short interspersed elements, including Alu and mamma- lian wide interspersed repeats, and long interspersed elements (Supplementary Table S7; Figure 1, Supplementary Figure S2). These intervals were also found to harbor CGIs, enhancers and active transcription start sites, which exhibit differential DNA methylation and histone modification, as well as predicting RNA expression in the normal testis or ovary.

Region of homozygosity (ROH). SNP-based CMA could detect genomic copy neutral (CN = 2) stretches of allelic homozygosity. This study detected a total of 3,421 ROHs, and approximately 10% ($n = 333$) overlapped with 92 DSD genes (Table 2; Supplementary Tables S2 and S8). Chromosomes with the most ROHs involving

\textbf{Figure 1.} Continued.
DSD genes were 15, 12 and X (Supplementary Figure S1A). Recurrent ROHs were seen in 58% (n = 53) of these genes, with PTPN11 showing the highest frequency (n = 29 patients) (Table 2, Supplementary Figure S2). Of the ROHs overlapping with DSD genes (~71% were small, 174 showed no corresponding CNVs (exclusively ROH) (n = 174) (Supplementary Table S8). Interestingly, eight X-linked genes seen in nine XX patients with DSD and/or Turner stigmata were found to have copy-neutral ROHs (no CNV) (Supplementary Table S9).

Recently, small ROHs detected by CMA were characterized in 46, XX SRY-negative testicular DSD patients.\(^{13}\) These regions involved 27 genes and exhibited variable and smaller sizes (than the clinical cutoff, i.e., 5–10 Mb) ranging from 200 bp (upstream of ZWINT) to 634 kb (INPP4B, USP38). Comparing these reported regions with our data (Supplementary Table S10), 19 small ROHs (1–8 Mb; 207–2,647 markers) involving 10 genes were also found in 16 of our patients (XX, XY, sex chromosome anomaly). Six of these genes were smaller than 50 kb.

**Position effect.** We customized a “position effect” CMA track to detect gene-desert regions that regulated the expression of a gene at a nearby or distant locus\(^ {36–38} \) e.g., the upstream region of SOX9.\(^ {16,39–41} \) Our study uncovered four recurrent and overlapping small CN losses (1.6–6 kb; 8–17 markers) approximately 1.2 Mb downstream of SOX9 in four patients (Supplementary Table S11). These intervals included DNA transposons (Charlie15A, MER103C), retroelements (Alu and mammalian wide interspersed repeats) and differentially methylated enhancers predicted to interact with other cis/trans regions (Supplementary Figure S2D). In addition to the SOX9 region, 86 CNVs and 44 ROHs involving 15 and 11 regions were also detected, respectively. The regions with the highest frequencies were SHH (n = 11) for CNVs and LCT for ROHs (n = 20). Approximately 52% of the CNVs that involved 15 regions were ≤50 kb (2–81 markers). Although many of the genes regulated by these regions are expressed in adrenal, urogenital and reproductive structures, their functional impact on sex development has yet to be determined.

**Imprinted genes.** We investigated 34 genes known to be imprinted (list obtained from the GeneImprint database: http://www.geneimprint.com/) for ROHs and CNVs (Supplementary Table S12). Approximately 65% of such genes were ≤50 kb. SNRPN is one of these genes, which has been implicated in Prader-Willi syndrome with or without DSD. Using an “imprinting CMA track,” 118 CNVs and 20 ROHs were found to involve 22 and 14 genes (paternal, maternal or isoform independent), respectively. The CNVs ranged from 1.6 kb to 2.4 Mb (2–728 markers), 73% of which were ≤50 kb. The ROHs ranged from 1 to 2.5 Mb (229–969 markers). GRB10 showed the highest CNV frequency (n = 19 patients). Concurrent gain and loss were exhibited by some genes, including GRB10 (13 patients) and SNRPN (7 patients). These genes are expressed in the adrenal gland, as well as the urogenital and reproductive structures, and their contributions to DSDs remain to be explored.

**Linked genes and gene interactions**

We explored the chromosome distribution of DSD genes, and we found several adjacent genes with recurrent ROH or CN gains (Supplementary Table S13), e.g., HSD3B1/HSD3B2/NOTCH2 (1p11.2-p12; ROH) in three patients with hypospadias (H/E), FGF8/CYP17A1 (10q24.32-q24.33; ROH) in three patients also with H/E, and NR0B1/DMO (Xp21.2; gain) in two patients with genital anomalies.

We utilized different public databases to investigate the interactions (e.g., genetic and physical) of the genes included in this study (Supplementary Table S14). Multiple interacting genes (DSD, imprinted, position) with CNVs and ROHs were revealed, with CTNNB1 showing the greatest number of interacting partners (Figure 2). Many other genes within the same pathway or network have also been found to exhibit local (cis) or interchromosomal (trans) gene interactions\(^ {12} \) (Supplementary Table S14).

---

**Figure 2.** Genetic (a) and physical (b) interactions, pathway associations (c) and co-expression (d) of CTNNB1 with other 125 genes (DSD, imprinted, position effect) with CNVs and ROH (GeneMANIA; http://www.genemania.org/). See Supplementary Table 24 for complete list of genes. CNV, copy number variation; DSD, disorders of sex development. Genes with predicted interactions/associations are in black circles, and without interaction/association are in gray circle.
Overlapping cases in DGV, ClinGen and DECIPHER
We compared our data with the most recent small CNV map (≥ 50 bp; 2,057,368 variants) constructed from a normal cohort (n = 2,647 subjects from 23 studies), and approximately 75% (225/301) of our CNVs were not overlapping (Supplementary Table S15; Figure 1a–c; Supplementary Figure S2D). We also searched the ClinGen database for comparable size CNVs overlapping with our data, and approximately 16% were documented as benign or likely benign, 7 CNVs (26–223 kb) involving 5 genes (CREBBP, CHD7, EP300, NRS A1, S UPT3H) were documented as pathogenic, and most (approximately 80%) showed no entry (Supplementary Table S15; Figure 1; Supplementary Figure S2). Approximately 61% of small CNVs that are common in healthy populations (DGV) or categorized as benign (ClinGen) were found to include salient genomic and epigenomic features characteristic of tissue-specific gene regulation and expression (Supplementary Tables S3–S6). The DECIPHER database also revealed CNVs with variable sizes overlapping with our data, and 19 DSD genes were documented to have small CNVs (33–50 kb), including CHD7, ANOS1 and WWOX, in patients with various DSD conditions (Supplementary Table S16). Although small ROHs (≥ 4 kb) (Supplementary Table S17) overlapping with DSD genes and CNVs overlapping with position effect regions (≥ 4.2 kb) (Supplementary Table S18) and imprinted genes (≥ 1.8 kb) (Supplementary Table S19) were also documented in DECIPHER, none were reported in patients with DSD.

Small CNVs and ROHs correlated with DSD phenotype
Pathogenic CMA findings were reported originally in only 13% of patients, and all of the cases involved aneuploidy, polyplody, mosaic sex chromosome complement or large structural anomalies (Table 1). Variants of uncertain clinical significance were reported in 23% of patients, none of which were recurrent. Normal CMA accounted for 62% of patients, approximately 12% of whom revealed ROHs in one or more chromosomes. CNVs and/or ROHs involving DSD genes were detected in all of the patients regardless of sex chromosome complement (XX, XY, aneuploid, mosaic) and original CMA finding (normal, pathogenic, variants of uncertain clinical significance).

Using our high-resolution and gene-targeted CMA approach, this retrospective study revealed individual genes with concurrent CNVs (loss/loss, gain/gain, loss/gain combinations) or multiple genes with concurrent CNVs and/or ROHs (Table 2). Some of these genes were found to be recurrent in two or more patients with similar DSD phenotypic findings (summarized in Supplementary Tables S20 and S21). The 28 patients with ambiguous genitalia with or without hypospadias/epispadias (H/E) revealed CNVs overlapping with a region known to regulate gene expression.25 Importantly, we found that most of the CNVs demonstrated that, regardless of size, the clinical relevance of the detected CNV is determined by gene content and other layers of evidence.49,50 Current clinical CMA practice follows arbitrarily set limits such that CNVs smaller than these cutoffs are typically not analyzed or reported.

In this study, we uncovered small CNVs from unanalyzed CNVs previously obtained from DSD patients. Although relaxing our analysis filter to its 1 kb limit of detection generated an extensive list of variants, our customized DSD gene-targeted CMA track expeditiously filtered CNVs (with proportional probe coverage) overlapping with DSD genes, thereby allowing us to circumvent the daunting task of investigating all of the variants. The majority of the small CNVs detected in our investigations were not documented in comparable resources obtained from normal cohorts. Although the DECIPHER and ClinGen databases have documented overlapping CNVs, most of them were found to be relatively larger in size, and very few small CNVs were documented in patients with DSDs. These findings suggested that these variants are not common in healthy populations and remain underrepresented in clinical populations. Many deletions or duplications of partial regions of DSD genes have been described in various disorders, some of which were reported in DSDs (Supplementary Table S22). All of these findings, however, warrant further studies to determine whether these ‘rare’ CNVs are specific to either syndromic or isolated DSDs or to various clinical populations, including those at risk for gonadoblastoma or germ cell neoplasms and especially those CNVs involving genes implicated in cancer.

Our CMA approach demonstrated the ability to detect not only CNVs that are ≤ 50 kb but also relevant genes that are smaller than the clinical CMA cutoff size. For example, small CN gains (4.9 and 5.6 kb) involving an intact small NR0B1 gene (approximately 5 kb in size) were detected in two patients with DSDs. These small CNVs and genes would be masked in clinical settings, unless it was one of the contiguous genes included within a reportable large CNV. We also identified not only potentially relevant private or overlapping small CNVs but also several CNVs that were recurrent in two or more patients with similar phenotypic findings. The breakpoints of these CNVs involved different regions or 5′ or 3′ ends, were intragenic or were upstream or downstream of a gene. High-resolution data obtained from this study provided us with a close-up view of the genomic details of deleted or duplicated structural and functional domains or regions. Although there are limited epigenomic data for the normal testis and ovary and none for atypical tissues, we were able to obtain DNA methylation, histone modification and RNA expression profiles in silico. Exon 1 CNVs exhibited the typical features of a 5′ promoter, such as the presence of unmethylated or differentially methylated CGIs, regulatory enhancers, active transcription start sites and regions coding for signal peptides. Intron CNVs were usually not reported in the past; however, their roles in various diseases have been gaining significance in recent years, e.g., disease-associated SNP variations were found in non-coding regions defined by regulatory H3K27Ac marks.52 The differential H3K27Ac marks and methylation and RNA expression profiles observed in some of the detected exon/intron 1 CNVs, as well as in other CNVs with orphan CGIs and enhancers acting as remote or cryptic promoter regions, suggested tight spatiotemporal tests/ovary-specific gene regulation.25 Importantly, we found that most of the CNVs...
involved genes that are highly expressed in adrenal, urogenital and other specific reproductive structures, which are key effector or recipient sites responsible for DSD phenotypes. In addition, these CNVs occur in genes that perform critical functions in sex development, including germ cell migration and development and gonadal and genital development, and in androgen and other signaling pathways. Whether CNVs or ROHs involving these regions and genes in these affected tissues would significantly result in improper peptide transport, aberrant protein structure, dysregulation of expression and ultimately a DSD phenotype remains to be elucidated.

Exaptation of transposable elements into promoter or enhancer regions has been correlated with spatiotemporal or tissue-specific gene regulation of the host genome. Our study focused on some of these elements within these small CNVs. Similar to segdups, these elements have been known to result in recurrent or overlapping small CNVs via non-allelic homologous recombination, and they have been implicated in various human disorders, e.g., the GTF2IRD1 and GTF2I genes distal to the ELN (7q11.23) locus that harbor DNA transposon CHARLIE-like region in Williams syndrome. Short interspersed elements are non-long terminal retroelements that are typically found in GC and gene-rich, early replicating euchromatin, as well as in cyogenetic G-banded light bands. It has been demonstrated that CNVs involving clusters of Alu short interspersed elements dysregulate gene expression within the promoter regions, they alter splicing and they shift reading frames, resulting in aberrant phenotypic findings. Although CNVs with or without differential methylation patterns for these DNA transposons and other repeats have been implicated in several human disorders, the significance of the repeat element-harboring CNVs that also showed regulatory genomic and epigenomic profiles remains to be pursued in DSDs.

An ROH region is typically reported in clinical settings if it is $\geq 5$ Mb and harbors a causative gene. However, regardless of how large the detected ROH interval is, such regions are further investigated for genes that might be inherited in an autosomal recessive pattern or involved in a disorder associated with uniparental disomy or imprinting. If detected, reflex studies, such as methylation or microsatellite analysis, are performed on the gene of interest. Our findings from gene-targeted approaches demonstrated that causative genes within an ROH less than the cutoff size could easily be revealed. Although copy neutral ROHs have been described in many neoplastic conditions, they are yet to be investigated in DSD. In a recent CMA study, in patients with 46,XX testicular DSD, small ROH-harboring candidate DSD genes were uncovered. Similarly, our study revealed small CNVs and ROHs overlapping with some of the genes previously reported, as well as with DSD genes and imprinted genes. Although it has yet to be determined whether the CNVs and ROHs involving these genes would result in DSDs, our data provided several avenues for further investigations in DSD patients, including those at risk for gonadoblastoma or germ cell neoplasms.

Our CMA track specific for position effect regions easily detected small CNVs and/or ROHs in gene-desert regions. SOX9 is one of the genes that is regulated by upstream regulatory elements via position effect, and disruption of this region results in craniofacial, skeletal and sex development anomalies. On the basis of several reported cases, a more refined interval upstream of SOX9 was described as the smallest critical region for XX (68 kb) and XY (32.5 kb) DSDs. The recurrent, small deletions detected downstream of SOX9 in our DSD patients overlapped with the reported 1.3 Mb position effect region described in a patient with campomelic dysplasia (with DSD). Whether the much smaller CNVs overlapping with position effect regions that we revealed in this study would result in DSDs through mechanisms similar to SOX9 remains to be elucidated. It is also possible that the other genes mapped within the interval downstream of SOX9 contribute to the DSD phenotype (or not at all), although this relationship also remains to be investigated.

Mapping the relative distributions of CNVs and ROHs overlapping with DSD genes revealed that CNVs were mostly frequently observed on chromosomes X, 16 and 9, and ROHs were commonly found on chromosomes 15, 12 and X. Neither CNVs nor ROHs were observed on chromosomes 21 or Y. Some of these CNVs involve dose-sensitive haploinsufficient genes that typically exhibit highly conserved coding and promoter regions, and they are tissue-specific and embryonically expressed. These genes are known to interact strongly with other haploinsufficiency genes in the same network. As genomic tools have become more sophisticated, more combinations of variants obtained from CMA and/or sequencing, both common and rare, have been revealed, e.g., compound heterozygote, triploinsufficiency and CNV/SNV combination, and some have been implicated in human diseases. Whether CNVs or ROHs involving DSD genes (including linked, imprinted or position effect genes) that are within the DSD network result in pathway dysregulation and subsequently in isolated or syndromic DSDs remains to be elucidated and mapped. The chromosome distribution of DSD genes and their associated CNVs and ROHs, as well as their local and interchromosomal interactions, could provide a better picture of the genomic landscape and perhaps of some of the tools needed to draw the genomic and epigenomic maps of the spatiotemporal regulation of interactions, pathways and networks in DSDs.

Our retrospective CMA identified recurrent small CNVs and ROHs overlapping with DSD genes in patients with sex chromosome aneuploidy or mosaicism, variants of uncertain clinical significance and clinically normal CMA. These findings emphasized our gene-targeted and disease-specific CMA approach to identifying small CNVs overlapping with genes or small regions of overlap that might be relevant to DSDs, supporting this approach in improving the diagnostic power and utility of CMA. In addition, these findings provided shorter intervals for detailed genomic and epigenomic analyses. The results obtained from this study will contribute significantly to the accumulating resources for small CNVs, as well as in narrowing the genomic gap in both normal and clinical populations. We are encouraged that this study will provide many avenues for downstream genomic and epigenomic research investigations and that our DSD-specific CMA approach will advance our understanding regarding DSDs, as well as provide a model analysis tool for finding small CNVs in other human disorders.

ACKNOWLEDGEMENTS
The authors thank the Washington University in St Louis School of Medicine (WUSTL SOM) DSD Clinic and Research team and Mike Evens and Jennifer Stamm-Birck of Cytogenomics Lab (WUSTL SOM) for assisting in initial patient data gathering. We also thank Amy Dodson and Dr Ma Xenia Ilagan (WUSTL SOM), Dr Julius Militante and Dr Christian Nievera for their helpful comments on the manuscript.

AUTHOR CONTRIBUTIONS
IA designed the study, performed the data analysis and wrote the manuscript. IN, VH and AH curated the data. JH reviewed the patient data. TW reviewed the epigenomic data. IN, VH, AH, JH, TW, AH and PA reviewed the manuscript.

COMPETING INTERESTS
The authors declare no conflict of interest.

REFERENCES
1 Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y et al. Detection of large-scale variation in the human genome. Nat Genet 2004; 36: 949–951.
Supplementary Information for this article can be found on the Human Genome Variation website (http://www.nature.com/hgv).