**SUPPLEMENTAL MATERIAL**

**Journal name:** Human Genetics

**Title:** Copy-number Variant Analysis of Classic Heterotaxy Highlights the Importance of Body Patterning Pathways

**Authors:** Erin M. Hagen, Robert J Sicko, Denise M. Kay, Shannon L. Rigler, Aggeliki Dimopoulos, Shabbir Ahmad, Margaret H. Doleman, Ruzong Fan, Paul A. Romitti, Marilyn L. Browne, Michele Caggana, Lawrence C. Brody, Gary M. Shaw, Laura L. Jelliffe-Pawlowski, James L. Mills

**Corresponding Author:** Dr. James Mills,

Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, United States jamesmills@nih.gov

# Supplemental Methods

**Genotyping**

Roughly 200-1,200 ng of DNA was extracted from two 3-mm DBS punches for each sample. DNA was extracted at the Wadsworth Center of the New York State Department of Health in Albany, NY. Genotyping was performed at the Biomedical Genomics Center Core Facility at the University of Minnesota using Illumina HumanOmni2.5-8\_v1-1\_B bead arrays and the Infinium LCG assay protocol.

The HapMap sample, a control sample run in duplicate and the trio run in duplicate served as quality control samples. Data were analyzed using Illumina GenomeStudio v2011.1. The genotype no-call threshold was set at <0.15. In the initial stage of analysis, genotypes were called using genotype clusters defined based on the data generated in this project. Genotypes were then manually reviewed, reclustered, edited, and excluded (where appropriate) based on parameters and quality control metrics described in Illumina’s Infinium Genotyping Data Analysis Technical Note (<http://res.illumina.com/documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf>).

# CNV calling and annotation

CNVs were imputed from the SNP genotyping data using Illumina’s cnvPartition algorithm (version 3.2.0) and PennCNV (version 2011/05/03). For both cnvPartition and PennCNV, the data were GC-wave adjusted to reduce the incidence of false positive calls. Additionally, each CNV call required a threshold of three single-nucleotide polymorphism (SNP) probes. Default confidence values were used: 35 for cnvPartition and 10 for PennCNV. CNV call files were compiled and annotated. The percent overlap with each of the following control databases was included: common CNVs in HapMap, Children’s Hospital of Philadelphia (CHOP), and the Database of Genomic Variants (DGV). The following were also noted: the percent agreement between cnvPartition calls and PennCNV calls; the number of cases and controls with the same/overlapping CNVs; and the transcripts and genes encompassed by each CNV. Transcripts and genes were identified using GENCODE Genes track (version 19, December 2013, HAVANA and Ensembl Datasets). CNV calls were reviewed for overlap with: OMIM genes (accessed via the UCSC genome browser, 14 May 2014), pathogenic CNVs defined by the Internal Standards for

Cytogenomic Arrrays database (accessed via the UCSC genome browser, 14 May 2014), CNVs previously reported in heterotaxy cases, genes associated with congenital heart defects, genes involved in embryonic development or signal transduction (defined by reactome version 46, accessed 21 November 2013), and variants in DECIPHER with related phenotypes (search terms included heterotaxy, inversus, situs, or ambiguous (accessed 14 May 2014)).

# CNV Validation

Genomic DNA was extracted from one 3-mm DBS (Saavedra-Matiz et al. 2013) diluted 1:10 in water, and amplified using TaqMan Environmental Master Mix (ABI) in 5µl reaction volumes. A fragment of the RNaseP H1 RNA gene was co-amplified and used as an internal control (TaqMan Copy Number Reference Assay, ABI). Assays were run in quadruplicate on either an ABI 7900HT or an ABI QuantStudio. CopyCaller software v2.0 (ABI) was used to analyze the real-time data using relative quantitation (2-ΔΔCt method). The manual Ct threshold was set to 0.2 with the automatic baseline on.

CopyCaller software parameters were as follows: the median ΔCt for each experiment was used as the calibrator, wells with an RNaseP Ct > 38 were excluded and the zero copy ΔCt threshold was set to six. The average copy number and a software-generated confidence value were calculated for each subject. Samples with confidence values ≥ 0.95 were considered valid; samples with confidence values <0.95 were rerun in quadruplicate.

# Next Generation Sequencing (NGS)

A custom panel targeting 20 genes (Supplemental Table 2) was designed using the Ion AmpliSeq Designer tool V1.2.9 using the 'standard DNA' (225-bp amplicon target sizes) and 'Gene + UTR' options. Two primer pools were used to amplify 760 amplicons, covering 20 target genes, totaling 83.1 kb. DNA was quantified using an RNaseP TaqMan assay on an ABI 7900HT Fast Real-Time PCR System (TaqMan RNaseP Control, Life Technologies). Libraries were constructed using 500 pg DNA, one AmpliSeq primer pool (per reaction) and AmpliSeq library kit 2.0. Amplification was carried out on a GeneAmp PCR System 9700 (Applied Biosystems; ABI; Carlsbad, CA) for 22 cycles. Reaction-specific primers were removed using FuPa reagent. AmpliSeq PCR products from each subject were ligated to P1

adapters and barcodes using IonXpress Barcode Adapter kits. Barcoded libraries were quantified by qPCR using the Ion Library Quantitation Kit. Samples with libraries far below 100pM (14.7 (± 13, (1.7-63)) were re-purified using Gentra Puregene (Qiagen) purification reagents and the manufacturer’s ‘Repurifying DNA Samples’ protocol. The resulting purified DNA was used to successfully prepare libraries above 100pM. Purified libraries were sequenced by the Applied Genomics Technologies Core (AGTC) at the Wadsworth Center, New York State Department of Health. Purified libraries were diluted to 100 pM and pooled. Template preparation was done on the OneTouch system using the Ion OT 200 Template kit v2, DL. Amplified Libraries were sequenced on an Ion Personal Genome Machine (PGM) sequencer (Life Technologies) on Ion 316v2 or 318C chips. Samples were run in three batches. The total aligned output for each of the three runs was 922M bases over 6.8M reads (28 test samples, 318C chip), 826M bases over 5.9M reads (28 test samples, 318C chip), and 529M bases over 3.8M reads (27 test samples, 316v2 chip). Over all runs, coverage uniformity (defined as base coverage at >20% mean coverage) was 85.8% and the proportion of on-target bases (proportion of bases mapping target regions out of total mapped bases per run) was 96.5%, which is consistent with manufacturer specifications. The average number (± standard deviation, range) mapped reads per individual was 197,518 (±78,917, (89,300- 431,869)). Mean read depth was 237X (± 95X, (114X-543X)). 90.8% (± 4.8%, (68.5%-95.7%)) of

bases had ≥20X and 71.1% (± 13.6%, (36.2%-90.2%)) of bases had ≥100X coverage.

Panel information was imported into Torrent Suite (v.4.2) and data were analyzed using Torrent Suite Software. Signal processing and basecalling were carried out using the default basecaller parameters. Sequence data was aligned and mapped to the reference sequence file using the Torrent Mapping Alignment Program (TMAP v.2.18), which is optimized for Ion Torrent data. Variants were called Ion Torrent Variant Caller (TVC 4.2-18) using default parameters for ‘PGM - Germ Line - Low Stringency’, except for the following parameter changes: minimum coverage on either strand = 2 for SNP and INDEL; downsample\_to\_coverage = 400; do\_snp\_realignment = 0; mnp\_min\_cov\_each\_strand = 2; output\_mnv

= 1; allow\_complex = 1.

Following variant calling, individual sample VCFs were merged with the bcftools merge function (bcftools\_mergeVersion=1.2+htslib-1.2.1). Prior to annotating variants with ANNOVAR (Wang et al. 2010) variants were decomposed and left-aligned as recommended by ANNOVAR documentation. Multi-allelic

variants were decomposed using vcflib's vcfbreakmulti function (git cloned on 5/15/2015: https://github.com/ekg/vcflib) and vt’s normalize function was used to left-align variants (version - vt-0.57: https://github.com/atks/vt). Variants were then annotated with ANNOVAR’s table\_annovar function using the following annotation sources: refGene, avsnp142, popfreq\_all\_20150413, clinvar\_20150330 and ljb26\_all (ANNOVAR\_DATE=2015-04-24).

# Potentially Pathogenic Variant Selection

Annotated variants were filtered to select only variants: with a quality value ≥20, with a flow evaluator read depth ≥20, absent from controls run with our samples (N=10), in exonic or splicing regions, allele frequency in any reference population (from ANNOVAR’s popfreq\_all\_20150413) < 0.05 and not labeled non-pathogenic in ClinVar (from ANNOVAR’s clinvar\_2015033). Variants passing the above filters were then manually inspected to remove variants only present in cases not carrying a candidate CNV, variants manually inspected via ClinVar website (date accessed: 7-21-2015) and annotated as non- pathogenic or single heterozygous variants in a primary ciliary dyskinesia gene. The remaining variants (N=25) were considered potentially pathogenic and were Sanger confirmed (N=23, shown in Supplemental Table 3) or ruled out as errors in NGS results (N=2).

# Sequence Variant Validation

Variants selected as potentially pathogenic were validated by Sanger sequencing as previously described with minor modifications (Rigler et al. 2015). Primer sequences were selected using Primer Designer Tool (Thermo Fisher Scientific; Waltham, MA); primer IDs and PCR conditions used are provided in Supplemental Table 4. PCR reactions contained extracted DNA, DNA Master HybProbe master mix (Roche Applied Science; Indianapolis, IN), 1 unit Taq antibody(Clontech; Mountain View, CA),

2.5 mM MgCl2, and 0.2 µM each primer, in a total volume of 25 µl. Standard cycling conditions included an initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at the specified annealing temperature for 30 seconds, elongation at 72°C for 30 seconds, and a final extension at 72°C for 5 minutes. PCR products were cleaned-up using ExoSAP-IT (USB Corporation; Cleveland, Ohio), and sequenced using BigDye Terminator v.3.1 Cycle Sequencing chemistry kits (ABI;

Carlsbad, CA) on an ABI 3730 DNA Analyzer. Sequence chromatograms were analyzed using SeqScape

v.2.1.1 (ABI; Carlsbad, CA), FinchTV v.1.4.0 (Geospiza; Seattle, WA) and Indelligent v.1.2 (Dmitriev and Rakitov 2008).

**Supplemental Table 1.** TaqMan Copy-number Assays Used for CNV Validation

|  |  |  |
| --- | --- | --- |
| **CNV Target****Assay ID Gene****Locus Coordinates** | **# Cases Tested (N=69)** | **# Controls Tested (N=175)** |
|  | Hs05768962\_cn | - | Chr.1:186,262,388 | 69 | 175 |
| 1q31.1 |  |  |  |  |  |
|  | Hs06536115\_cn | *TPR* | Chr.1:186,293,508 | 69 | 10 |
|  | Hs04680339\_cn | *PDCD1* | Chr.2:242,797,714 | 69 | 10 |
| 2q37.3a | Hs00921386\_cn | *CXXC11* | Chr.2:242,815,356 | 69 | 175 |
|  | Hs05835886\_cn | *-* | Chr.2:242,832,875 | 69 | 10 |
| 3p21.31b | Hs04735734\_cn | *IP6K1* | Chr.3:49,795,371 | 69 | 175 |
|  | Hs01269585\_cn | *UBA7* | Chr.3:49,851,182 | 69 | 11 |
|  | Hs05887563\_cn |  | Chr.3:192,327,060 | 69 | 10 |
| 3q29 |  | *FGF12* |  |  |  |
|  | Hs04760355\_cn |  | Chr.3:192,340,045 | 69 | 175 |
| 4p13**c** | Hs05914768\_cn | *-* | Chr.4:42,304,931 | 69 | 10 |
|  | Hs05912981\_cn | *-* | Chr.4:42,323,545 | 69 | 10 |
|  | Hs00422859\_cn | *NIPBL* | Chr.5:36,995,830 | 69 | 175 |
| 5p13.2 | Hs03006002\_cn | *NUP155* | Chr.5:37,351,278 | 69 | 10 |
|  | Hs00470154\_cn | *WDR70* | Chr.5:37,703,208 | 69 | 10 |
|  | Hs02603364\_cn | *KIAA1586* | Chr.6:56,919,811 | 69 | 175 |
| 6p12.1 |  |  |  |  |  |
| 5351-end\_CXAAYT8 *-* Chr.6:56,933,323 Probe Excluded |
|  | Hs05060948\_cn | *ZNF395* | Chr.8:28,239,562 | 69 | 10 |
| 8p21.1 |  |  |  |  |  |
|  | Hs06256337\_cn | *FZD3* | Chr.8:28,391,361 | 69 | 175 |
|  | 5801-sta\_CXPACWW | *-* | Chr.8:5,324,118 | 69 | 175 |
| 8p23.2 |  |  |  |  |  |
|  | 5801-end\_CXFARI4 | *-* | Chr.8:5,341,177 | 69 | 11 |
|  | Hs02750224\_cn | *RUFY2* | Chr.10:70,105,803 | 69 | 10 |
| 10q21.3 | Hs03752906\_cn | *DNA2* | Chr.10:70,193,657 | 69 | 175 |
|  | Hs02732719\_cn | *TET1* | Chr.10:70,450,935 | 69 | 10 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 12p13.33**d** | Hs01439930\_cn | *SLC6A13* | Chr.12:369,208 | 69 | 10 |
|  | Hs00816291\_cn | *KDM5A* | Chr.12:394,718 | 69 | 175 |
|  | Hs03918119\_cn | *SPN* | Chr.16:29,677,039 | 69 | 175 |
| 16p11.2**e** | Hs00856235\_cn | *KCTD13* | Chr.16:29,923,327 | 69 | 11 |
|  | Hs03930440\_cn | *TBX6* | Chr.16:30,098,448 | 69 | 11 |
| 16p13.3**c** | Hs02119086\_cn | *RGS11* | Chr.16:320,518 | 69 | 10 |
|  | Hs00550548\_cn | *AXIN1* | Chr.16:339,565 | 69 | 10 |
|  | Hs03952563\_cn |  | Chr.16:6,757,306 | 69 | 175 |
| 16p13.3 | RBFOX-mi\_CXMSGKF | *RBFOX1* | Chr.16:6,800,447 | 69 | 10 |
|  | Hs03937646\_cn |  | Chr.16:7,065,937 | 69 | 93 |
|  | Hs03927309\_cn |  | Chr.16:83,797,816 | 69 | 175 |
| 16q23.3 |  | *CDH13* |  |  |  |
|  | Hs03952540\_cn |  | Chr.16:83,812,486 | 69 | 11 |
|  | 7829-sta\_CX39QZI | *-* | Chr.17:7,269,257 | 69 | 175 |
| 17p13.1 |  |  |  |  |  |
|  | Hs05480868\_cn | *PLSCR3* | Chr.17:7,294,833 | 69 | 10 |
|  | Hs06504166\_cn | *LDLRAD4* | Chr.18:13,497,645 | 69 | 175 |
| 18p11.21 |  |  |  |  |  |
|  | Hs06453051\_cn | *FAM210A* | Chr.18:13,715,538 | 69 | 10 |
|  | Hs06501555\_cn | *-* | Chr.18:27,880,408 | 69 | 175 |
| 18q12.1 |  |  |  |  |  |
|  | Hs06502278\_cn | *-* | Chr.18:28,053,072 | 69 | 10 |
|  | Hs07220406\_cn |  | Chr.20:52,097,462 | 69 | 175 |
| 20q13.2 |  | *TSHZ2* |  |  |  |
|  | Hs07209539\_cn |  | Chr.20:52,107,864 | 69 | 10 |

aOne case with a duplication detected by microarray was found to be a complex variant. qPCR results were normal (2 copies) at the

location targeted by probe Hs04680339\_cn, 4-5 copies at the location targeted by Hs00921386\_cn and duplicated (3 copies) at Hs05835886\_cn. The region targeted by probe Hs04680339\_cn overlaps with numerous CNV calls in an in-house database and is likely an artifact of the microarray data. bOne control was found to be duplicated in this region (the case identified by microarray also carried a duplication, which was confirmed by qPCR). Minimum predicted overlap between control and case duplication is

91%. cTwo cases with deletion detected by microarray were found to be normal (2 copies) by qPCR at this locus, and are considered false positives. dOne control was found to be duplicated in this region (the case identified by microarray also carried a duplication, which was confirmed by qPCR). Case 11 was found to be normal at Hs01439930\_cn and duplicated at

Hs00816291\_cn; the control identified as duplicated in this region was also normal at Hs01439930\_cn and duplicated at Hs00816291\_cn, suggesting the duplications overlap in the region of *KDM5A*. **e**One control was found to be duplicated (3 copies) in

this region (the case identified by microarray carried a deletion, which was confirmed by qPCR). Minimum predicted overlap between control duplication and case deletion is 68%.

**Supplemental Table 2.** Genes targeted by a custom AmpliSeq Panel

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Chromosome** | **Number of Amplicons** | **Total Bases Targeted** | **Covered Bases** | **Missed Bases** | **Design Coverage (%)** |
| *NKX2-5* | 5 | 15 | 4,428 | 4,428 | 0 | 100.0% |
| *DNAI1* | 9 | 25 | 2,521 | 2,513 | 8 | 99.7% |
| *DNAI2* | 17 | 22 | 2,581 | 2,565 | 16 | 99.4% |
| *ZIC3* | X | 29 | 3,939 | 3,893 | 46 | 98.8% |
| *ACVR2B* | 3 | 83 | 11,373 | 11,172 | 201 | 98.2% |
| *SESN1* | 6 | 35 | 3,631 | 3,548 | 83 | 97.7% |
| *DNAH5* | 5 | 158 | 15,573 | 15,057 | 516 | 96.7% |
| *DNAH11* | 7 | 150 | 14,189 | 13,638 | 551 | 96.1% |
| *TXNDC3* | 7 | 28 | 2,311 | 2,198 | 113 | 95.1% |
| *FOXH1* | 8 | 17 | 2,183 | 2,071 | 112 | 94.9% |
| *GJA1* | 6 | 22 | 3,130 | 2,927 | 203 | 93.5% |
| *NODAL* | 10 | 17 | 2,086 | 1,936 | 150 | 92.8% |
| *LEFTY2* | 1 | 16 | 2,332 | 2,146 | 186 | 92.0% |
| *KTU* | 14 | 21 | 2,963 | 2,700 | 263 | 91.1% |
| *CRELD1* | 3 | 20 | 3,163 | 2,681 | 482 | 84.8% |
| *FOXI2* | 10 | 20 | 3,178 | 2,689 | 489 | 84.6% |
| *GDF1* | 19 | 20 | 2,558 | 1,889 | 669 | 73.8% |
| *RPGR* | X | 50 | 5,869 | 4,183 | 1,686 | 71.3% |
| *NKX6-2* | 10 | 9 | 1,019 | 710 | 309 | 69.7% |
| *CFC1*a | 2 | 3 | 2,018 | 160 | 1,858 | 7.9% |

AmpliSeq 20-gene panel design (i.e., maximum possible experimental coverage).

On average, 90.8% of bases had ≥20X coverage (68.5% min. - 95.7% max.) and 71.1% of bases had ≥100X coverage (36.2% min.

- 90.2% max.)

aExcluded from analysis due to poor sequencing data.

**Supplemental Table 3.** Primer IDs and PCR conditions for Sanger validations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gene** | **Exon** | **Primer ID**a | **Amplicon Size (bp)** | **Annealing Temp (°C)** |
| *NODAL* | 2B | See Rigler et al 2015b | 599 | 55 |
| *CRELD1* | 10 | Hs00363516\_CE | 508 | 60 |
| *DNAH5* | 58 | Hs00257784\_CE | 503 | 60 |
| *DNAH5* | 55 | Hs00257788\_CE | 519 | 60 |
| *DNAH5* | 50 | Hs00257794\_CE | 513 | 60 |
| *DNAH5* | 33 | Hs00333554\_CE | 407 | 60 |
| *DNAH5* | 32 | Hs00588882\_CE | 274 | 60 |
| *SESN1* | 7 | Hs00769360\_CE | 252 | 60 |
| *DNAH11* | 23 | Hs00748997\_CE | 262 | 60 |
| *DNAH11* | 41 | Hs00284750\_CE | 497 | 60 |
| *DNAH11* | 50 | Hs00806318\_CE | 272 | 60 |
| *DNAH11* | 51 | Hs00831757\_CE | 272 | 60 |
| *DNAH11* | 63 | Hs00284773\_CE | 459 | 60 |
| *DNAH11* | 66 | Hs00442604\_CE | 563 | 60 |
| *DNAH11* | 69 | Hs00472077\_CE | 568 | 60 |
| *DNAH11* | 74 | Hs00284785\_CE | 502 | 60 |
| *DNAH5* | 38 | Hs00456566\_CE | 556 | 62 |
| *DNAH11* | 31 | Hs00751629\_CE | 228 | 63 |
| *FOXI2* | 2 | Hs00325293\_CE | 509 | 63 |
| *RPGR* | 15 | Hs00304610\_CE | 539 | TD-PCR |
| *LEFTY2* | 2 | Hs00314112\_CE | 507 | TD-PCR |

TD-PCR = touchdown PCR. Total volume of 25 µl, initial denaturation at 95°C for 5 minutes, 15 cycles of denaturation at 95°C for 30 seconds, annealing at 70°C for 30 seconds (decreasing 1°C each cycle), elongation at 72°C for 30 seconds, 24 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, elongation at 72°C for 30 seconds, final extension at 72°C for 5 minutes.

*NKX6-2* insertion (c.140dupG, p.G47fs in Case 12) could not be attempted because no PCR product could be generated (>80% GC content amplicon); primers (Hs00325582\_CE) were used with multiple PCR conditions/additives attempted.

aM13-tailed predesigned sequencing primers ordered from Thermo Fisher Scientific using the Primer Designer Tool. bPrimer

sequences described in Rigler et al. (2015) were used. Forward: 5'-TTCACTGTCACTTTGTCCCAGGTC-3' Reverse: 5'- TGGAGGTGCTTGAGTAACTGTG-3'.

**Supplemental Table 4.** Non-prioritized CNVs present in individuals with classic heterotaxy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Locus** | **Genomic coordinates**a | **Size (bp)** | **Type** | **Case ID** | **Gene(s)/Transcript(s)** |
| 1p36.11 | 24,495,839–24,520,482 | 24,644 | Dupl | 19 | *IFNLR1* |
| 1p22.2 | 89,401,163–89,441,298 | 40,136 | Dupl | 20 | *CCBL2* |
| 2p13.2 | 72,358,275–72,378,979 | 20,705 | Het Del | 21 | *CYP26B1* |
| 3p25.2 | 13,057,174–13,078,846 | 21,673 | Dupl | 22 | None |
| 3p14.1 | 68,083,964–68,238,985 | 155,022 | Het Del | 23 | *FAM19A1* |
| 3q12.3 | 102,181,131–102,297,540 | 116,410 | Dupl | 6 | *ZPLD1* |
| 3q21.3 | 128,289,541–128,424,456 | 134,916 | Dupl | 24 | *C3orf27; RPN1* |
| 4q13.3 | 71,452,751–71,552,398 | 99,648 | Dupl | 25 | *AMBN; ENAM; IGJ* |
| 5p14.3;p15.1 | 17,692,294–18,515,870b | 823,577 | Dupl | 26 | *RP11-454P21.1; SNORD81* |
| 5q22.1 | 109,960,197–110,429,903 | 469,707 | Dupl | 27 | *CTC-551A13.1; CTC-551A13.2;**SLC25A46; SNORA51; TMEM232; TSLP; WDR36* |
| 5q22.2;q22.3 | 113,067,189–113,126,897 | 59,709 | Dupl | 28 | None |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 5q23.2 | 126,003,817–126,054,668 | 50,852 | Het Del | 29 | None |
| 6p25.1 | 6,148,217–6,181,526c | 33,310 | Het Del | 5 | *F13A1; MIR5683* |
| 6p22.1 | 28,618,341–28,679,945 | 61,605 | Het Del | 20 | None |
| 6q15 | 92,473,270–92,518,321 | 45,052 | Het Del | 4 | None |
| 6q21 | 113,213,997–113,235,177 | 21,181 | Het Del | 30 | None |
| 7q21.11 | 79,614,042–79,644,659 | 30,618 | Dupl | 31 | None |
| 8p12 | 29,537,217–29,570,628 | 33,412 | Het Del | 22 | None |
| 9p24.2 | 3,747,246–4,547,288b | 800,043 | Complex | 32 | *AL162419.1; GLIS3; JAK2;**RNU6-694P; RP11-358M14.2; RP11-70J12.1; SLC1A1* |
| 9p22.2 | 18,182,792–18,219,276 | 36,485 | Het Del | 33 | None |
| 10p14 | 7,037,826–7,118,081 | 80,256 | Het Del | 9 | None |
| 10q25.2 | 113,683,581–113,731,511 | 47,931 | Het Del | 34 | None |
| 11p12 | 37,222,070–37,722,047 | 499,978 | Het Del | 35 | None |
| 11q22.3 | 109,195,249–109,230,743 | 35,495 | Het Del | 26 | *RP11-708B6.2* |
| 12q12 | 44,087,372–44,353,473 | 266,102 | Het Del | 25 | *IRAK4; PUS7L; RP11-**210N13.1; RP11-350F4.2; TMEM117; TWF1* |
| 13q14.13 | 47,156,627–47,196,985 | 40,359 | Het Del | 24 | *LRCH1* |
| 13q31.1 | 80,382,362–80,411,925 | 29,564 | Dupl | 22 | None |
| 13q31.1 | 81,848,881–81,924,660 | 75,780 | Het Del | 15 | *RP11-452B18.2* |
| 13q31.3 | 91,042,605–91,103,197 | 60,593 | Het Del | 36 | None |
| 14q12 | 26,752,004–26,855,298 | 103,295 | Dupl | 13 | None |
| 14q32.12 | 94,560,433–94,585,064 | 24,622 | Dupl | 18 | *IFI27; IFI27L1* |
| 15q14 | 39,518,016–39,587,911 | 69,896 | Het Del | 28 | *C15orf54; RP11-624L4.1* |
| 15q26.2 | 97,352,005–97,380,490c | 28,486 | Het Deld | 23 | None |

aGenomic coordinates were predicted using PennCNV and correspond to DGV build hg19 unless otherwise noted. bCoordinates

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 16q23.1 | 77,169,104–77,191,539 | 22,436 | Het Del | 35 | None |
| 17p12 | 12,043,312–12,131,217c | 87.906 | Dupl | 37 | *MAP2K4; RP11-471L13.2* |
| 17q24.3 | 67,133,367–67,308,358 | 174,992 | Dupl | 38 | *ABCA10; ABCA5; ABCA6* |
| 21q22.12 | 37,016,636–37,041,107 | 24,472 | Het Del | 16 | None |

were predicted using the smallest start coordinate and the greatest stop coordinate of all CNV calls spanning the region and correspond to DGV build hg19. **c**Coordinates were predicted based on cnvPartition calls and correspond to DGV build

hg19. dPennCNV predicted this as two CNVs (a small homozygous deletion followed by a heterozygous deletion), cnvPartition

predicted this as one homozygous deletion; based on the log R ratio and B-allele frequency plots, we believe this CV is most likely a heterozygous deletion (data not shown).

Dupl = duplication; Het Del = heterozygous deletion; Complex = PennCNV called the following variants across the region: duplication (3,752,193-3,845,073), copy-number two (3,845,074-3,945,947), duplication (3,945,948-4,499,945), heterozygous deletion (4,501,352-4,502,848) and duplication (4,504,129-4,547,288) and cnvPartition called the following variants across the region: duplication (3,747,246-3,845,784), copy-number two (3,845,785-3,941,972), duplication (3,941,973-4,297,421), homozygous

deletion (4,297,603-4,298,955), duplication (4,298,955-4,499,945), copy-number two (4,499,946-4,504,128) and duplication

(4,504,129-4,547,288).

**Supplemental Table 5.** Cases with both a sequence variant and candidate CNV

| **Study ID** | **Candidate CNV** | **Sequence Varianta** |
| --- | --- | --- |
| 3 | 3p21.31 Dupl | *FOXI2* - p.G208R |
| 4 | 3q29 Dupl | *LEFTY2* - p.N110N |
| 5 | 5p13.2 Dupl6q15.1 Het Delb | *DNAH5* - p.L1742L |
| 8 | 8p23.2 Het Del | *DNAH11* - p.E3595K*DNAH11* - p.R3756C*DNAH11* - p.A4059T |
| 10 | 10q21.3 Dupl | *DNAH11* - c.11203-1G>C (Homozygous) |
| 11 | 12p13.33 Dupl | *DNAH5* - p.R1761X *DNAH5* - p.R3116X |
| 12 | 16p13.3 Dupl | *NKX6-2* - p.G47fsc |
| 16 | 18p11.21 Dupl21q22.12 Het Delb | *DNAH11* - p.L1787L*DNAH11* - p.L2242L*DNAH11* - p.T3425K*NODAL* - p.G260R |
| 20 | 1p22.2 Duplb6p22.1 Het Delb | *DNAH5* - p.Q3260X*DNAH5* - p.R1761G*DNAH5* - p.L1742L |
| 21 | 2p13.2 Het Delb | *NODAL* - p.G260R |
| 25 | 4q13.3 Duplb12q12 Het Delb | *DNAH5* - p.S2756S*DNAH5* - p.V2128I |
| 28 | 5q22.2-q22.3 Duplb15q14 Het Delb | *RPGR* - p.E1117D (Hemizygous) |
| 33 | 9p22.2 Het Delb | *DNAH11* - p.R1375H*DNAH11* - p.C2756R *DNAH11* - p.H2788D |
| 35 | 11p12 Het Delb16q23.1 Het Delb | *CRELD1* - p.I417V |
| 37 | 17p12 Duplb | *SESN1* - p.E318del |

aHeterozygous unless otherwise noted. bThis candidate CNV was not selected for validation by qPCR. cThis variant could not be validated nor ruled out via Sanger sequencing (the variant was in a region difficult to sequence).

Het Del=heterozygous deletion; Dupl=duplication.

# Supplemental References

Dmitriev DA, Rakitov RA (2008) Decoding of Superimposed Traces Produced by Direct Sequencing of Heterozygous Indels. PLoS Comput Biol 4(7): e1000113. doi:10.1371/journal.pcbi.1000113

Rigler SL, Kay DM, Sicko RJ, Fan R, Liu A, Caggana M, Browne ML, Druschel CM, Romitti PA, Brody LC, Mills JL (2015) Novel copy-number variants in a population-based investigation of classic heterotaxy.

Genet Med 17:348-357. doi:10.1038/gim.2014.112

Saavedra-Matiz CA, Isabelle JT, Biski CK, Duva SJ, Sweeney ML, Parker AL, Young AJ, DiAntonio LL, Krein LM, Nichols MJ, Caggana M (2013) Cost-Effective and Scalable DNA Extraction Method from Dried Blood Spots. Clinical Chemistry 59:1045-1051. doi:10.1373/clinchem.2012.198945

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high- throughput sequencing data. Nucleic Acids Research 38:e164-e164. doi:10.1093/nar/gkq603