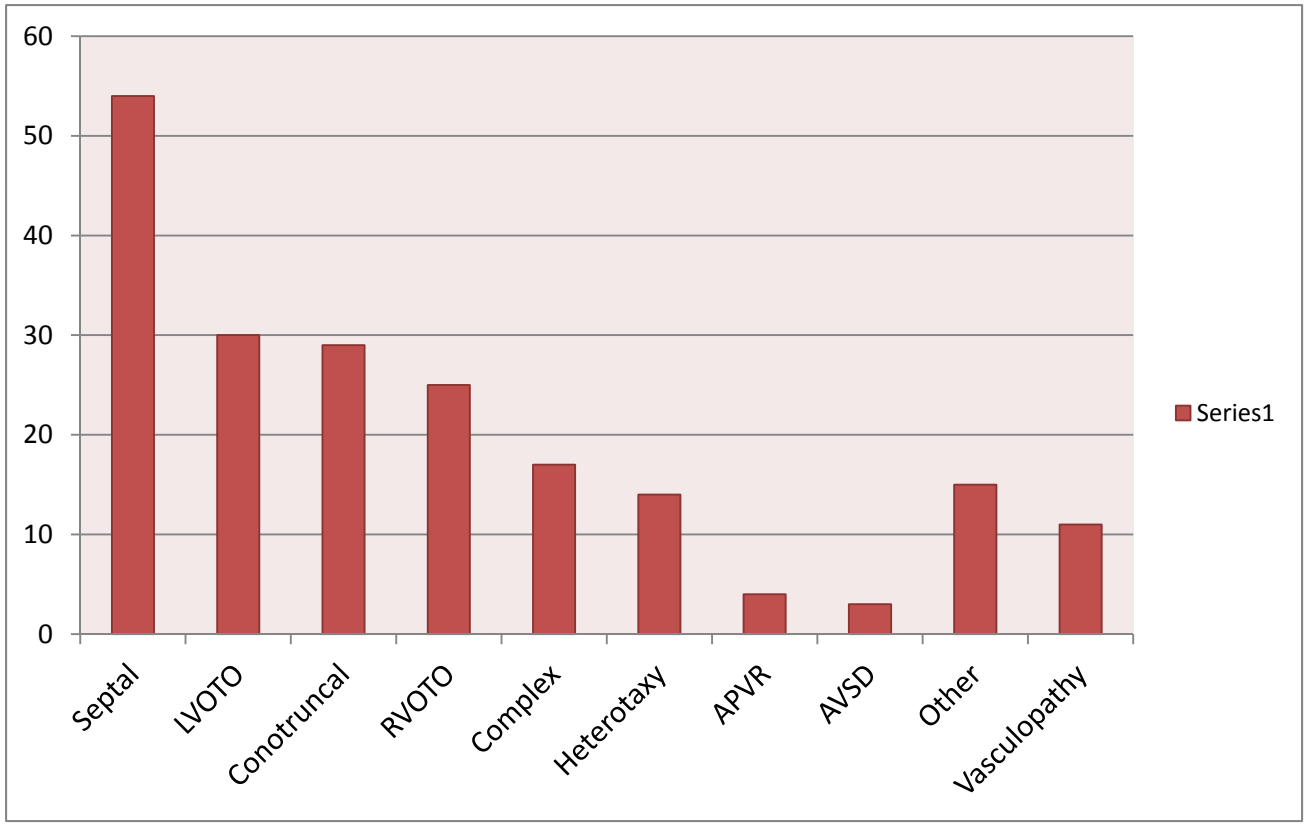
Supplementary figure 1

Summary of CVM classification in the discovery cohort (BCM1)

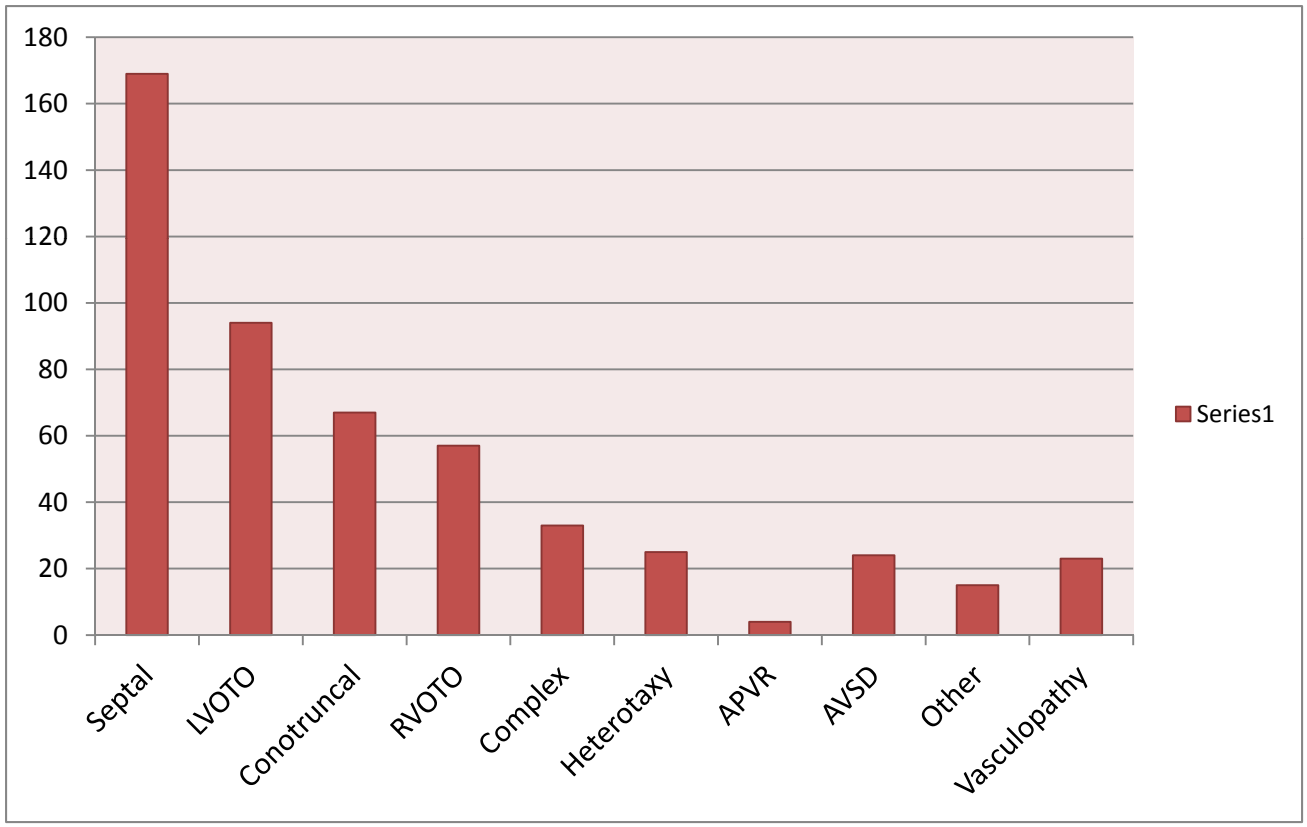
Discovery cohort (n=203)



Supplementary figure 2

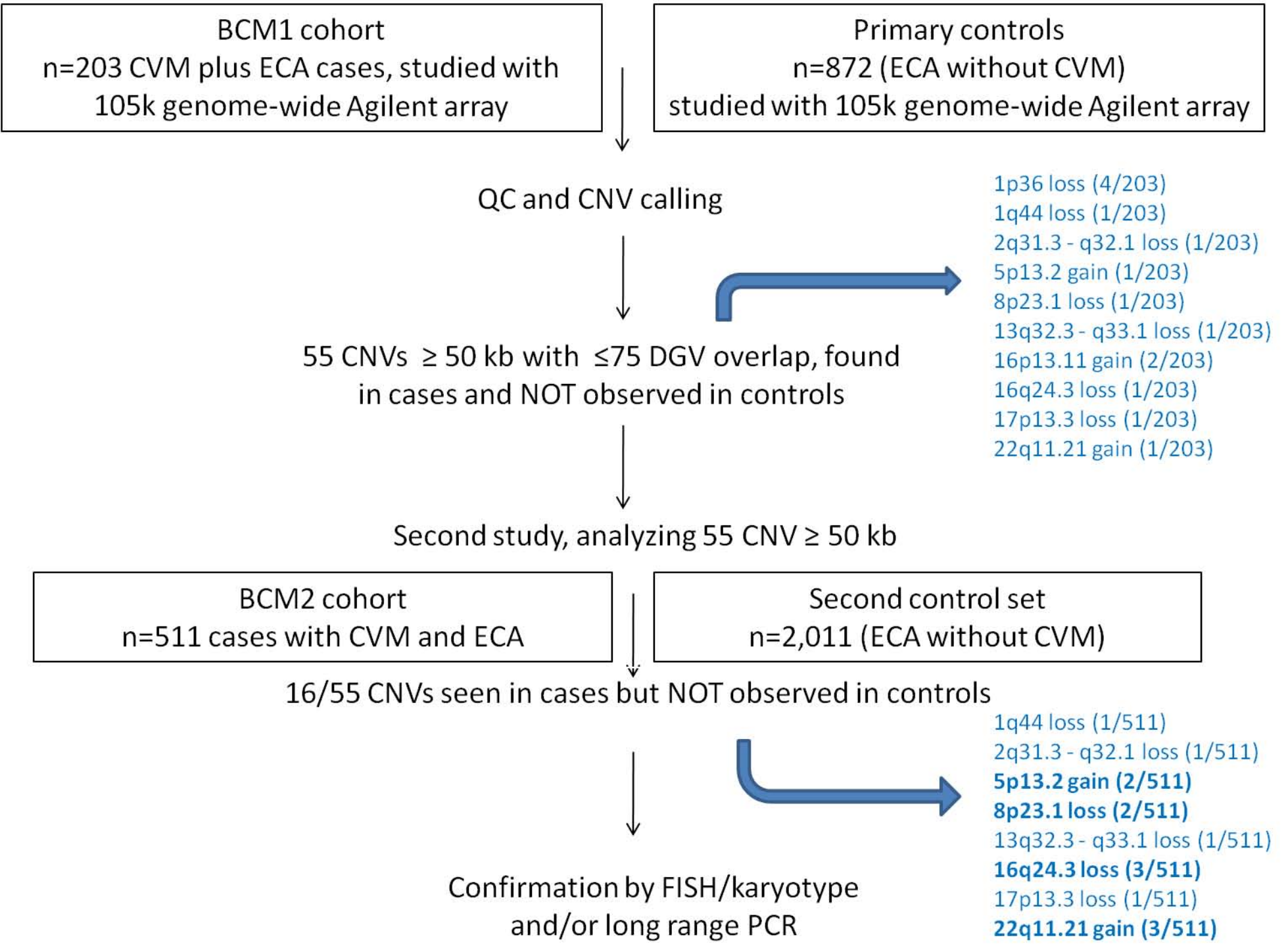
Summary of CVM classification in the second cohort (BCM2)

Replication cohort (n=511)



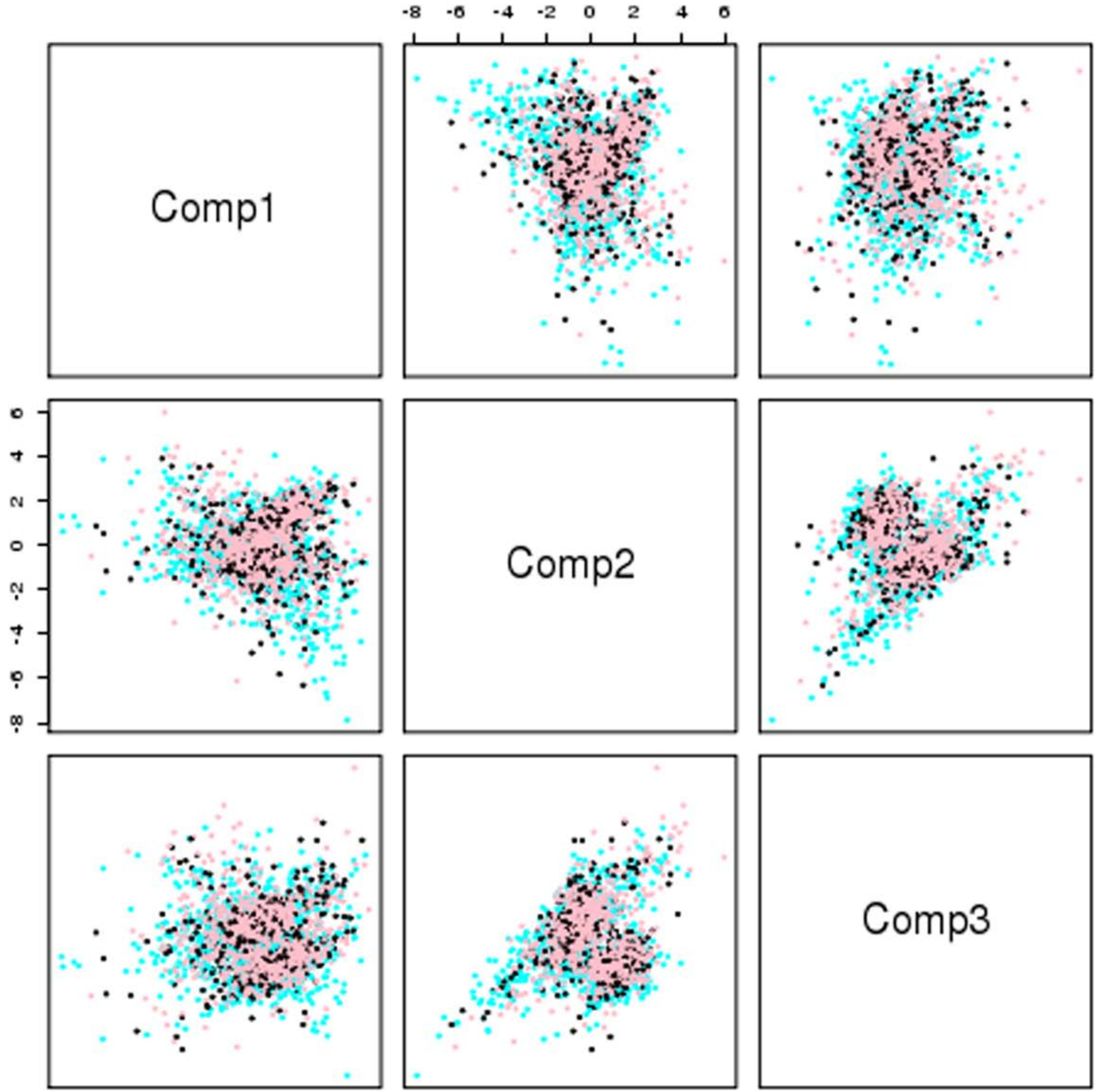
Supplementary figure 3

Flowchart for the study design


BCM1 cohort
n=203 CVM plus ECA cases, studied with 105k genome-wide Agilent array
Primary controls
n=872 (ECA without CVM)
studied with 105k genome-wide Agilent array
QC and CNV calling
1p36 loss (4/203)
1q44 loss (1/203)
2q31.3 - q32.1 loss (1/203)
5p13.2 gain (1/203)
8p23.1 loss (1/203)
13q32.3 - q33.1 loss (1/203)
16p13.11 gain (2/203)
16q24.3 loss (1/203)
17p13.3 loss (1/203)
22q11.21 gain (1/203)
55 CNVs ≥ 50 kb with ≤75 DGV overlap, found in cases and NOT observed in controls
Second study, analyzing 55 CNV ≥ 50 kb
BCM2 cohort
n=511 cases with CVM and ECA
Second control set
n=2,011 (ECA without CVM)
16/55 CNVs seen in cases but NOT observed in controls
1q44 loss (1/511)
2q31.3 - q32.1 loss (1/511)
5p13.2 gain (2/511)
8p23.1 loss (2/511)
13q32.3 - q33.1 loss (1/511)
16q24.3 loss (3/511)
17p13.3 loss (1/511)
22q11.21 gain (3/511)
Confirmation by FISH/karyotype and/or long range PCR

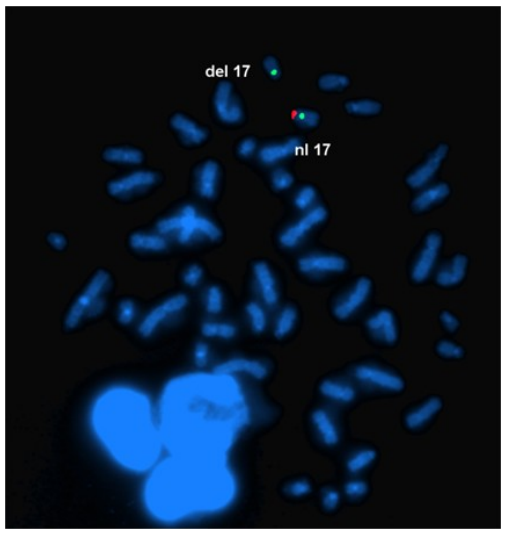
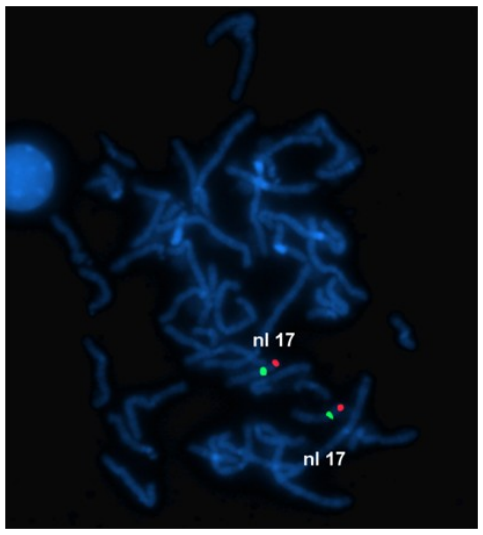
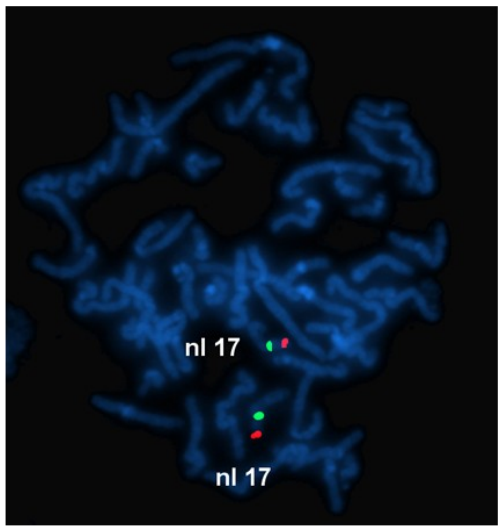
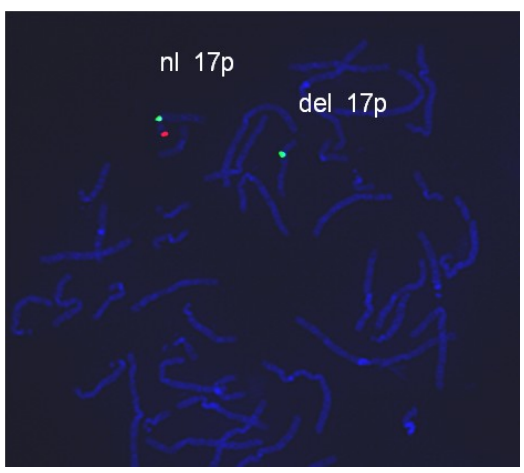
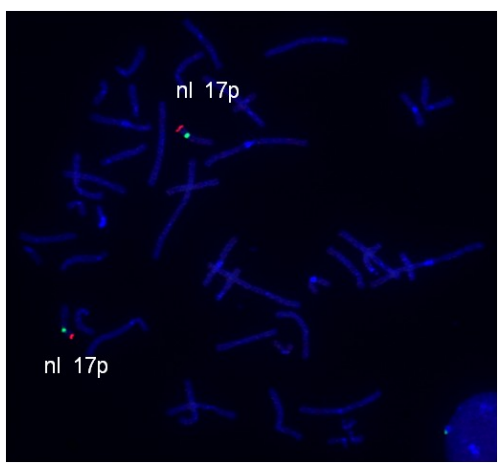
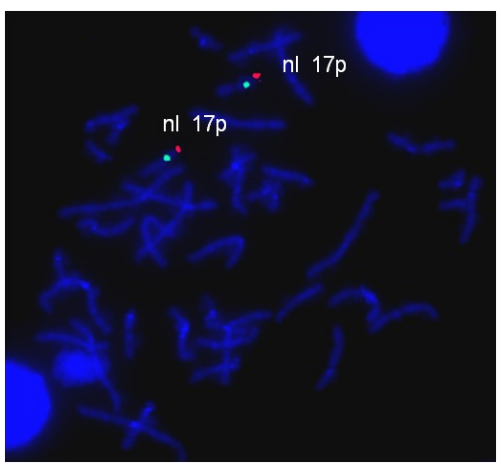
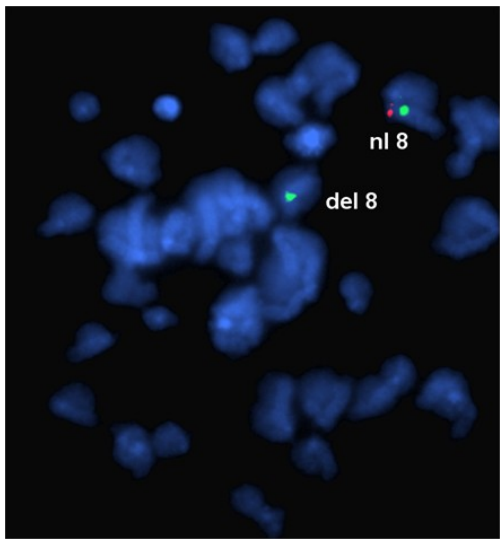
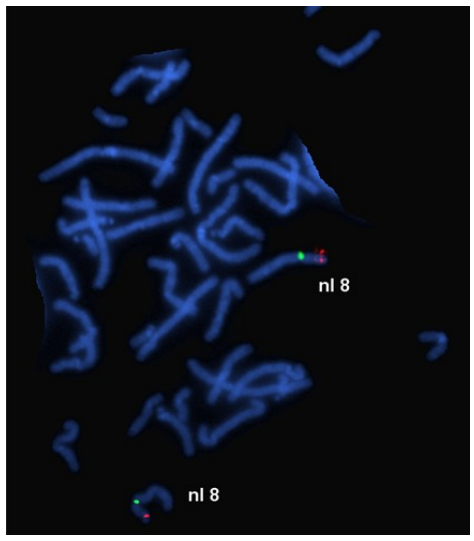
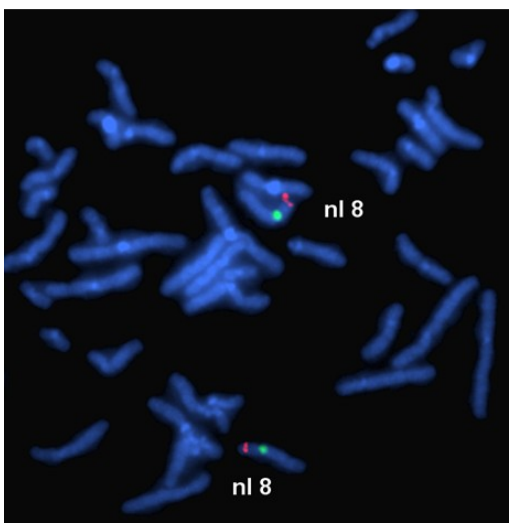
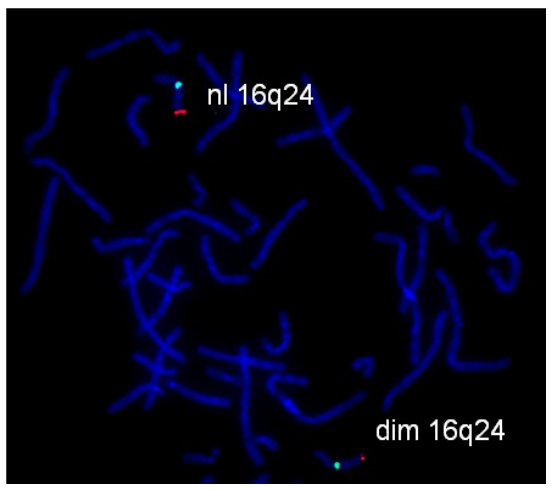
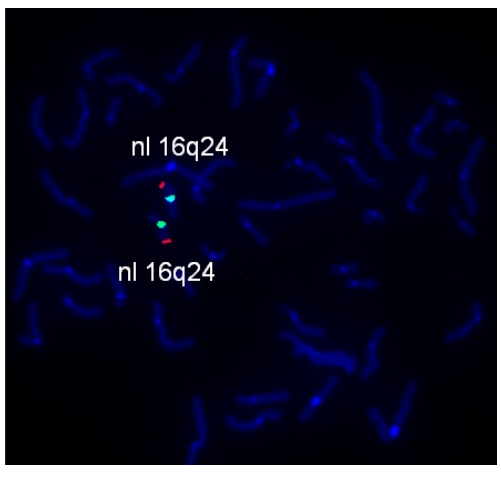
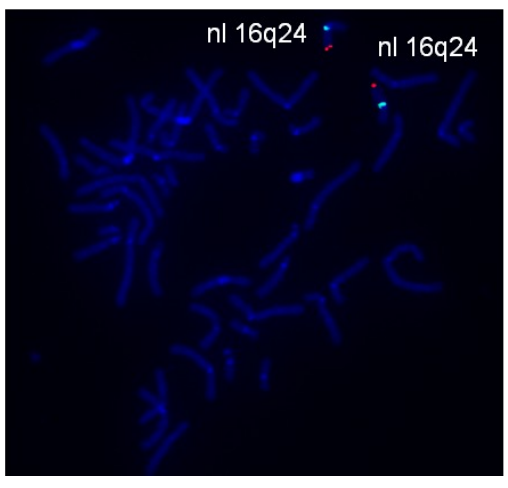
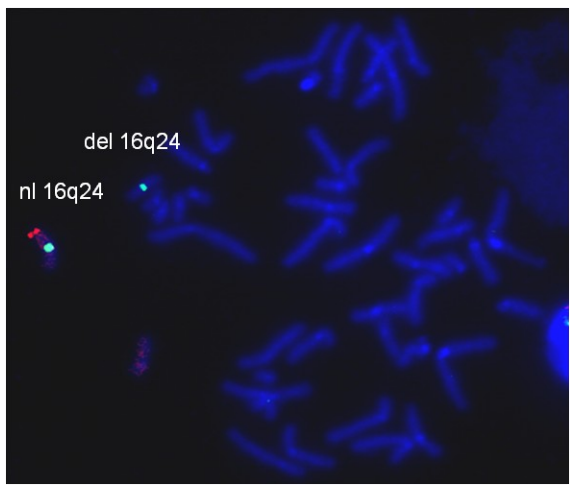
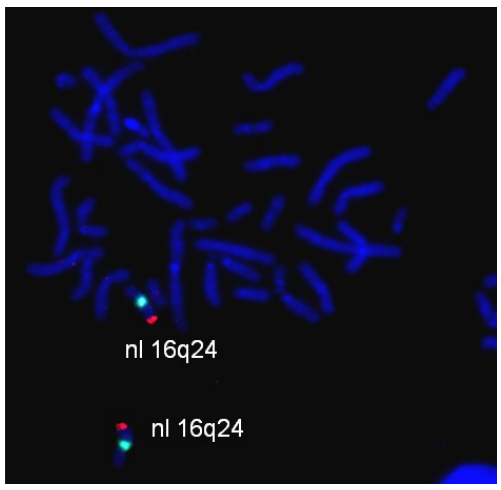
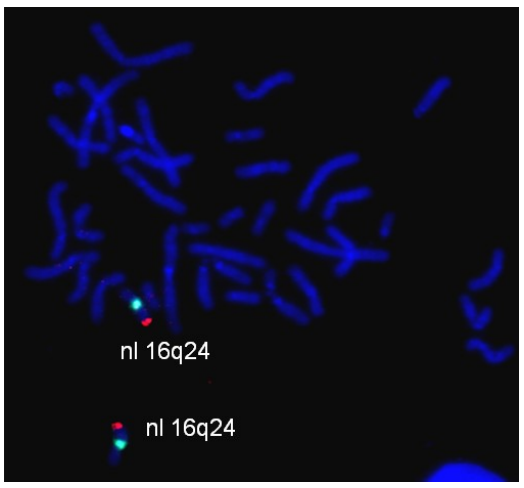
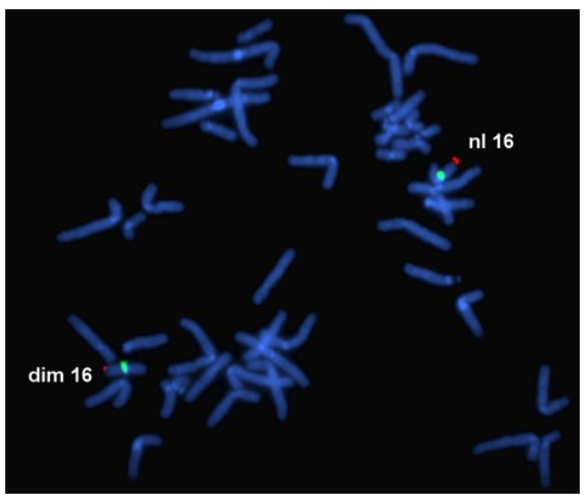
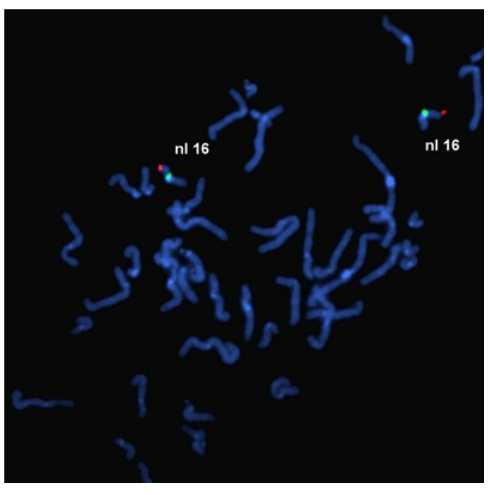
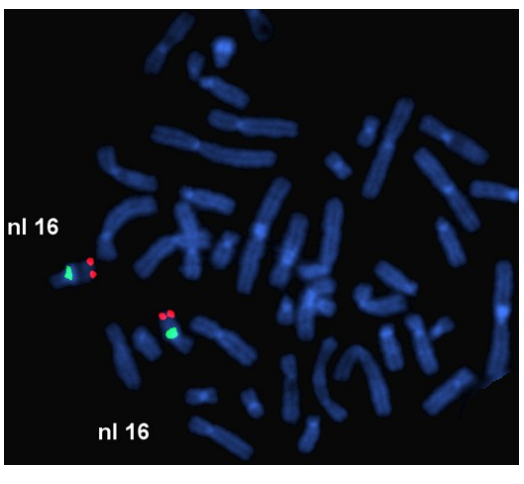
Supplementary figure 4

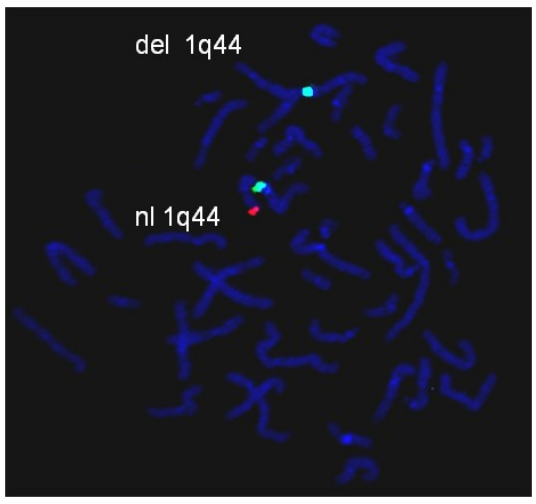
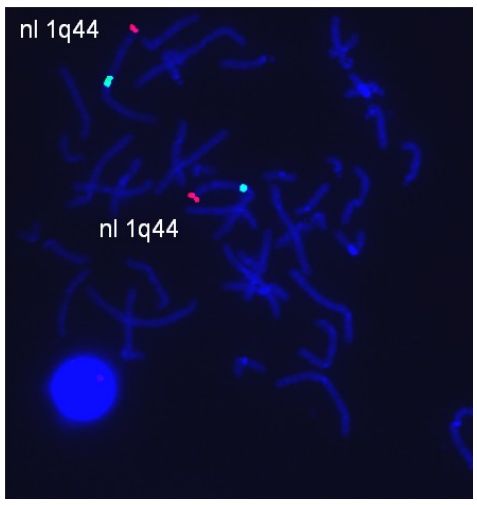
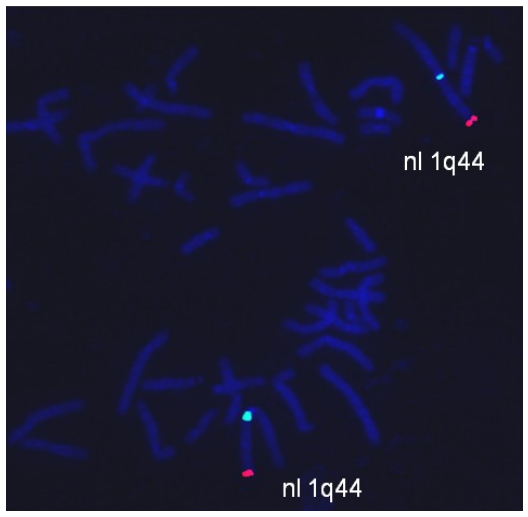
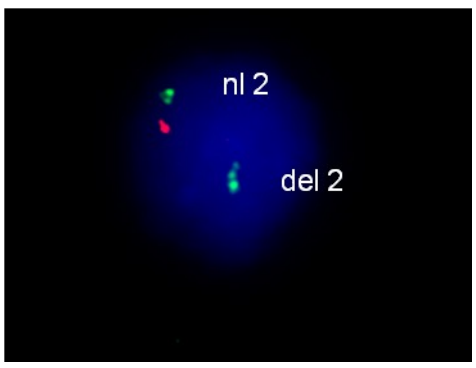
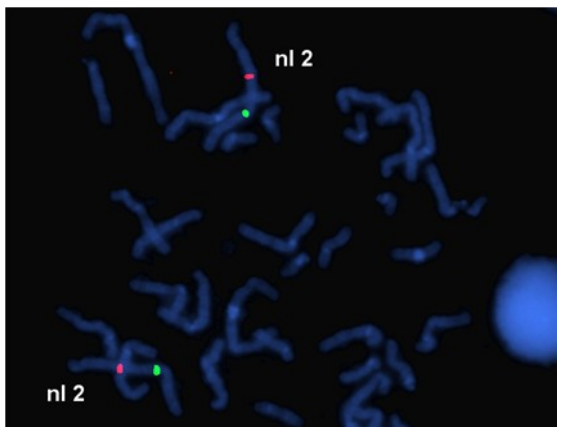
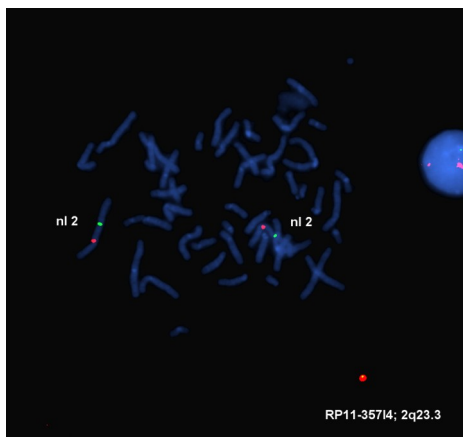
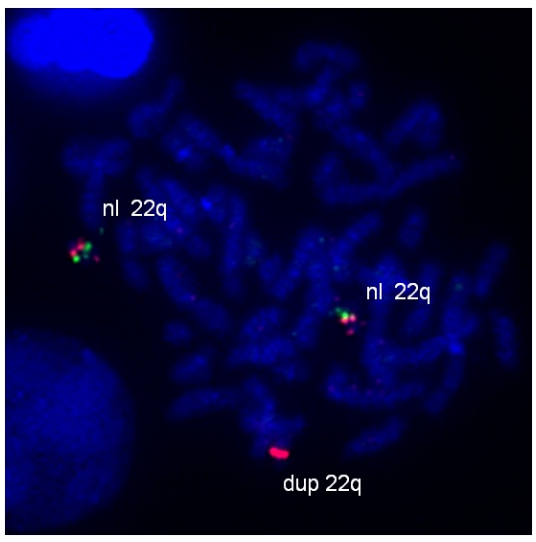
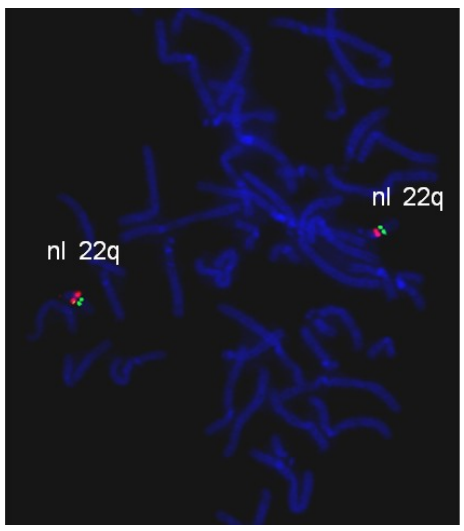
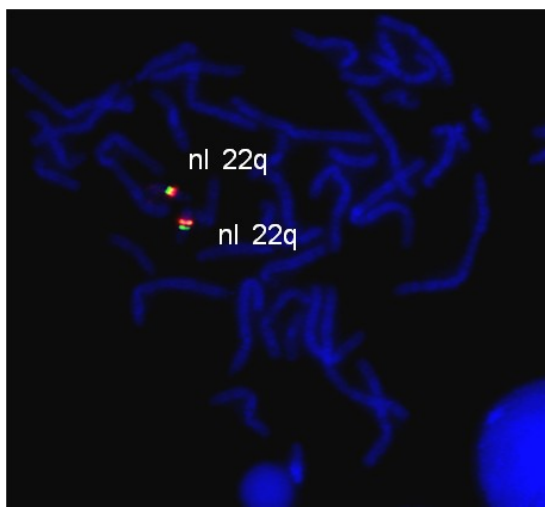
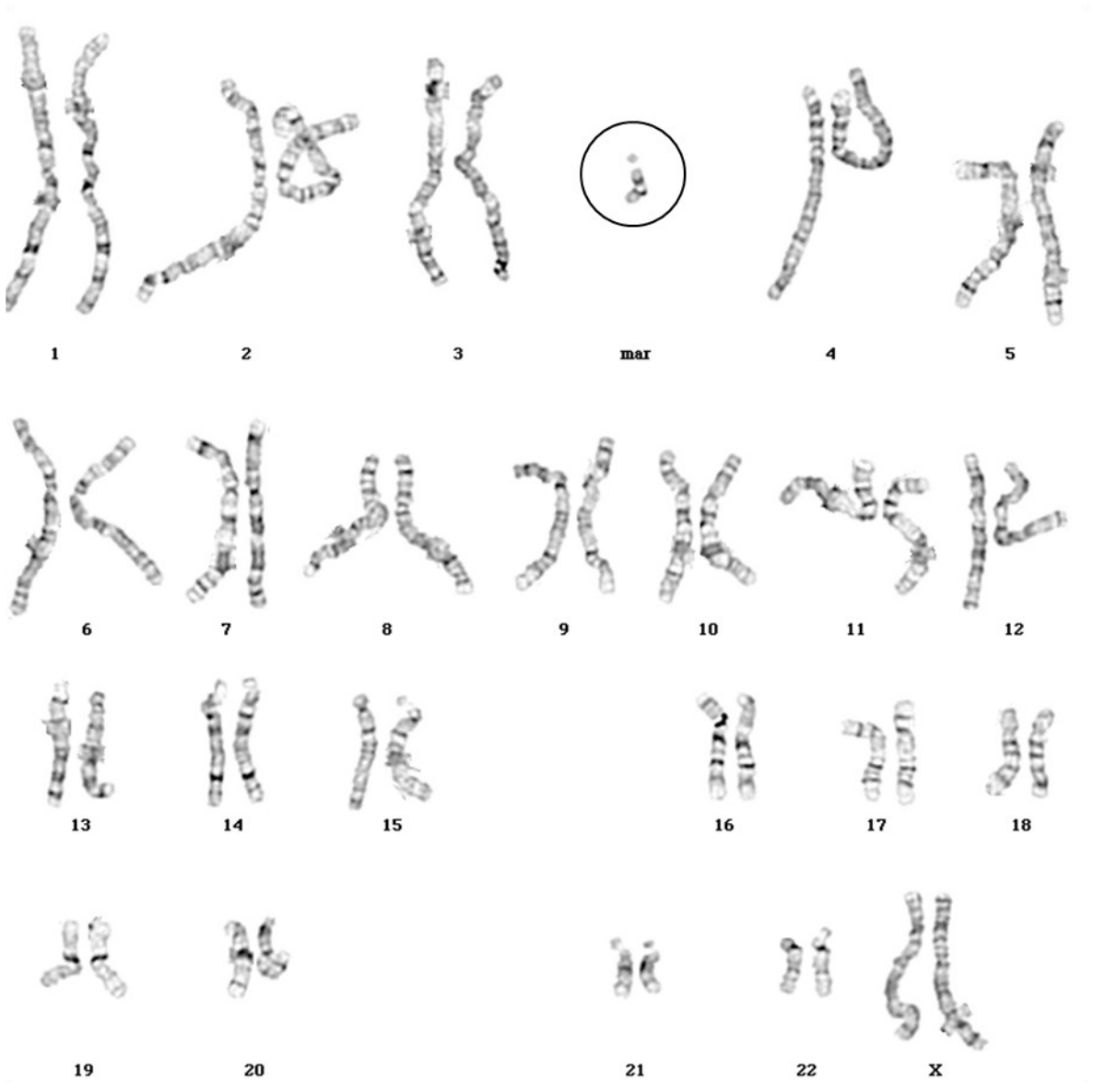
Principal component analysis across all cases and controls using over 10,000 oligonucleotides, illustrating that the cases and controls are matched for gender and ethnicity. Each oligonucleotide and each individual was brought to mean 0 before the decomposition. R package Corpcor was used to compute the singular value decomposition of the centered matrix. Only the first 3 projections are shown. Male controls are shown in blue and female controls are shown in pink. Cases are represented in black. Components 2 and 3 separate the ethnic groups corresponding to Caucasian non-Hispanics and White-Hispanics into approximately equal sized clusters, consistent with the known ethnicity composition of our patient population. The data show that that cases and controls are overlapping, indicating that both genders and ethnicities are well matched between these two cohorts. All cases and controls are pediatric subjects, between age 0 and 18 years.


-8 -6 -4 -2 0 2 4 6
Comp1
Comp2
Comp3
-8 -6 -4 -2 0 2 4 6

Supplementary figures 5 and 6

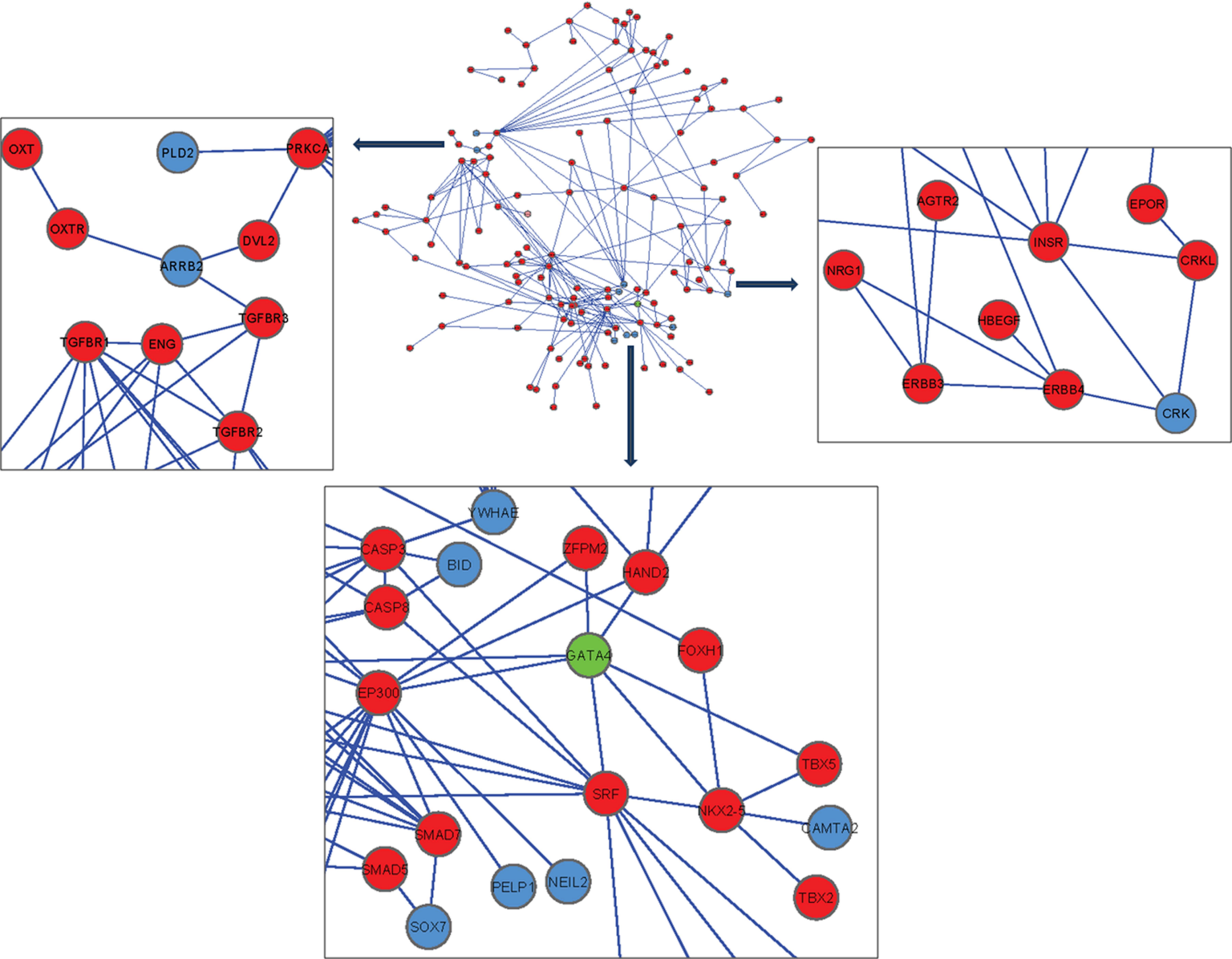
FISH validation and *de novo* ascertainment of the enriched structural variants in cases

	Child	Parent	Parent	
CVM3843				17p13.3 loss DN
CVM9035				17p13.3 loss DN
CVM3839				8p23 loss (*GATA4*) DN
CVM7658				16q24.3 loss DN
CVM0585				16q24.3 loss DN
CVM4535				16q24.3 loss DN

	Child	Parent	Parent	
CVM1260				1q44 loss DN
CVM2630				2q31.3 loss (*SSFA2, PPP1R1C, PDE1A*) DN
CVM1421				22q11 gain (*CECR1*) DN
CVM0498				22q11 gain (marker chromosome) Parental samples N/A

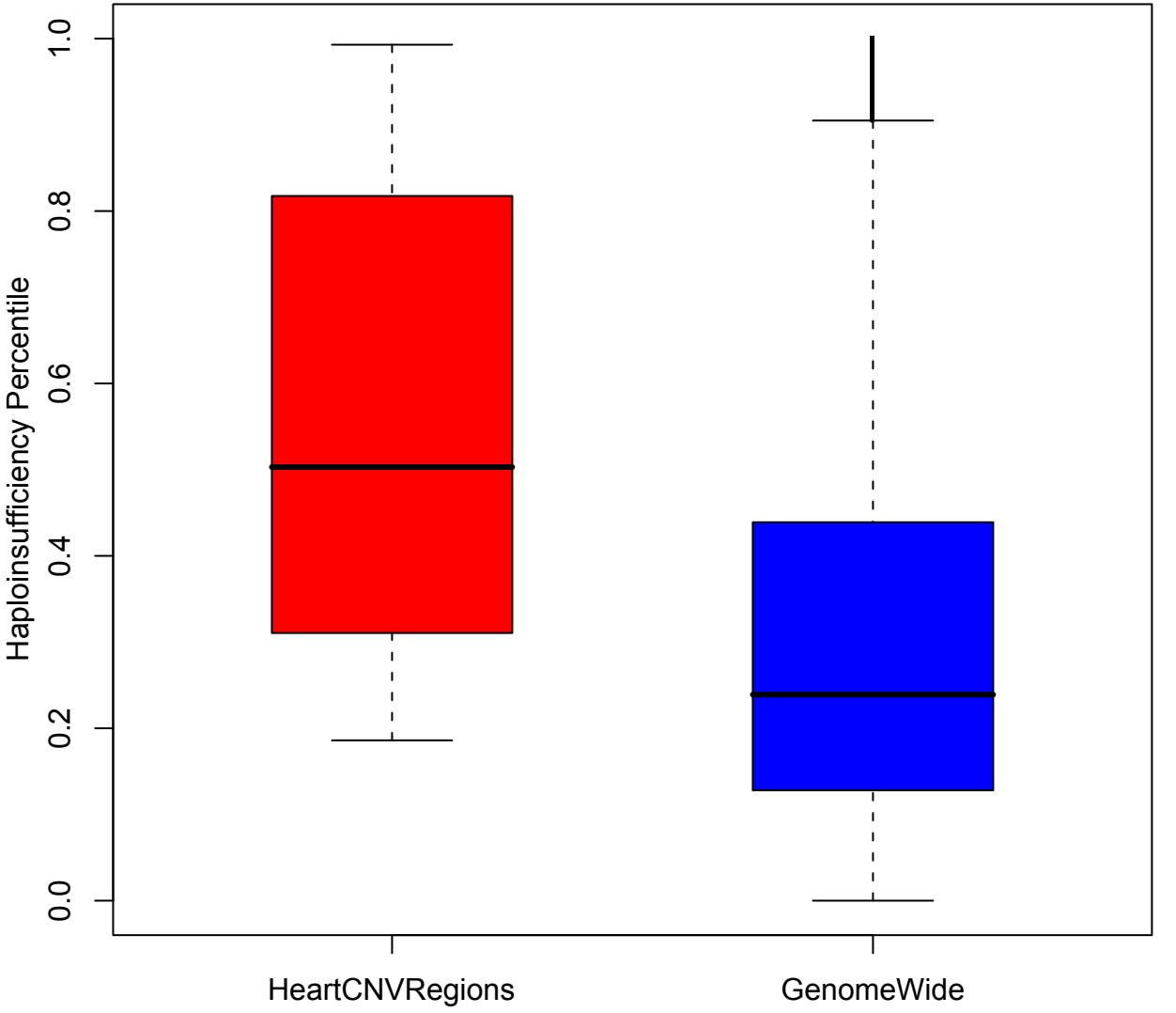
Supplementary figure 7

Protein-protein interactions of the candidate proteins encoded within the enriched CNVs with 276 proteins specific for human cardiac development in the Human Protein Reference Database (HPRD). Subnetwork comprised of genes harbored by the enriched structural variants (blue) and annotated human cardiac genes (red). Connections in the network are determined by annotations documented in the HPRD. Note that the candidate proteins NEIL2 and PELP1, a corepressor of SRF directly interact with EP300, responsible for Rubinstein-Taybi syndrome. SOX7, a transcriptional factor directly interacts with SMAD7 and SMAD5. NKX2-5 and HAND2 regulate the expression of transcription factor GATA4 (represented in green, as in both categories). CAMTA2 acts as a coactivator of Nkx2-5 and the associated 17p13 variant is enriched in the CVM cohort. These results show that the genes within the enriched CNVs in this study encode proteins that directly interact with proteins implicated in human cardiac development.


OXT
PLD2
PRKCA
OXTR
DVL2
ARRB2
TGFBR3
TGFBR1
ENG
TGFBR2
AGTR2
EPOR
INSR
CRKL
NRG1
HBEGF
ERBB3
ERBB4
CRK
YWHAE
CASP3
BID
ZFPM2
HAND2
CASP8
GATA4
FOXH1
EP300
TBX5
SRF
NKX2-5
SMAD7
SMAD5
PELP1
NEIL2
TBX2
SOX7

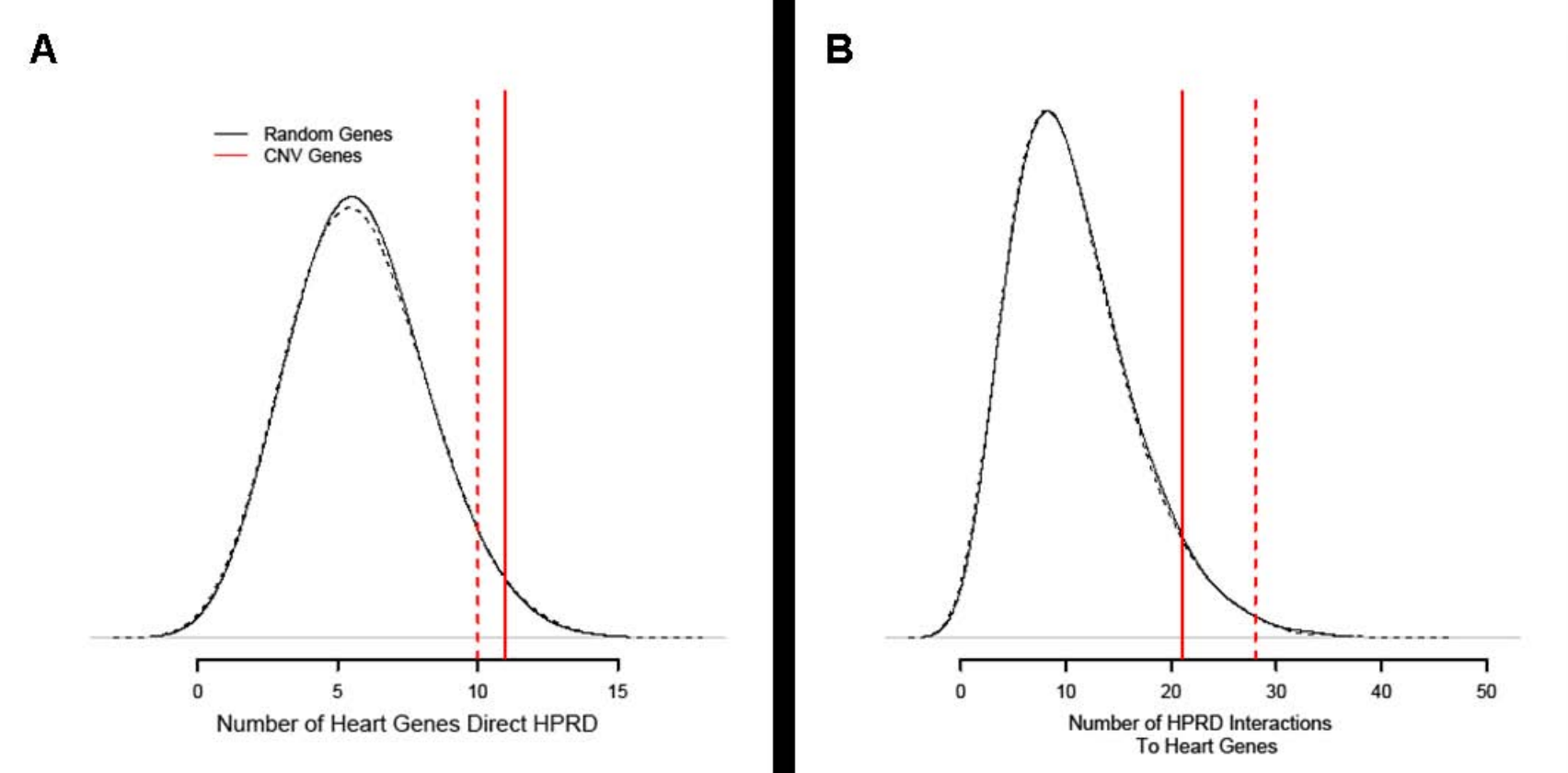
Supplementary figure 8

Haploinsufficiency scores for the genes within the 16 enriched structural variants


Haploinsufficiency Percentile
0.0
0.2
0.4
0.6
0.8
1.0
HeartCNVRegions
GenomeWide

Supplementary figure 9

Monte Carlo analysis to determine the statistical significance of the number of direct interactions and size of the subnetwork of the candidate proteins with 276 human cardiac specific genes, using a random set of proteins with 10,000 simulations. A subset of the candidate genes identified in the study, have direct interactions with the annotated human cardiac specific genes in the HPRD. The number of distinct genes directly connected to annotated human cardiac genes (A) was larger than would be expected by chance, with *P* = .0288. The total number of interactions (B) between the candidates and the annotated cardiac genes was also significant, with *P* = .0112. This protein interaction network analysis indicates that the genes within the enriched CNVs encode proteins that directly interact with proteins implicated in human cardiac development.


A
Random Genes
CNV Genes
0
5
10
15
Number of Heart Genes Direct HPRD
B
0
10
20
30
40
50
Number of HPRD Interactions
To Heart Genes