Supplemental Index:

Supplemental Methods:

Library construction and exome capture:
All exome samples were prepared by subjecting 2 ug of genomic DNA to a series of shotgun library construction steps, including fragmentation through acoustic sonication (Covaris), end-polishing and A-tailing, ligation of sequencing adaptors, and PCR amplification. Following library construction, 1 μg of shotgun library is hybridized to biotinylated capture probes for 72 hours and then recovered via streptavidin beads. Unbound DNA is washed away, and the captured DNA is PCR amplified for sequencing.

Sequence data processing and alignment:
Raw sequenced reads (from FASTQ files) were first split into 36bp chunks (in order to to avoid interference from indels), and mapped using the mrsFAST (v.2.3.0.2) aligner. Up to two mismatches were allowed per read. To reduce computational overhead, we created a concatenated exome index, consisting of the targeted exons (see below), plus 300bp flanking sequence from the hg19 (NCBI build 37) human reference genome, masked with RepeatMasker and Tandem Repeat Finder. After mapping to this concatenated "exome", we translated mapped coordinates back to to hg19 genome coordinates for further processing.

Exome probe definitions:
For the mrsFAST-based alignments, we developed a probe set (i.e., target regions) by intersecting target definitions of the Roche Nimblegen EZ Exome SeqCap Version 2 (from http://www.nimblegen.com/downloads/annotation/ez_exome_v2/SeqCapEZ_Exome_v2.0_Design_Annotation_files.zip) exome capture kit with RefSeq exons (excluding UTR regions). In addition, we included 4,857 non-exonic targeted regions from the SeqCap Version 2 target definition list. This resulted in 194,080 target probes (available at http://conifer.sourceforge.net)

Initial exon-level normalization:
We calculated RPKM values for the 194,080 target probes individually. The RPKM normalization is given by

$$\text{RPKM} = 10^9 * \text{Read Starts} / \text{Total Mapped Reads} * \text{Target Size (bp)}$$

where the number of Read Starts is defined as the number of reads starting within the target boundaries, and the Total Mapped Reads corresponds to the number of unique reads which had at least one mapping. This initial RPKM normalization step adjusts our read-depth estimates for target (exon) size as well as the overall sequencing coverage in the experiment. To reduce erroneous signal from failed or improperly targeted probes, we excluded 3,964 targets which had a median RPKM ≤ 1 in the 533 ESP samples.

Next, to control for probe-to-probe differences in capture efficiency, we standardized the RPKM values using a z-transformation. The median and standard deviation of each exon were derived from RPKM values of the 533 ESP exomes. The formula for the zRPKM value is:

$zRPKM = (RPKM_{exon,sample} - Median_{exon}) / StdDev_{exon}$

Removing systematic bias between batches:
A previous analysis of exome read-depth values from ~1,700 ESP exomes using principal components analysis (PCA) revealed several strong components, some of which were attributed to "batch" effects (unpublished, Sara Ng and Jay Shendure). We hypothesized that these strong components do not correspond to biological signal, but rather to differences in capture protocol, efficiency and sequencing bias. Using singular value decomposition, a mathematical analog of PCA, we decompose the exon-by-sample (*X*) data matrix into three matrices:

$X = USV^T$

In order to remove the strongest *k* components, we set $S_1...S_k$ to zero to form *S'*, and then recalculate *X* as the dot product of *U*, *S'* and V^T . For computation efficiency, each chromosome is normalized individually across the population. We used an implementation of SVD in the scipy.stats package available for the python programming language.

Discovery of rare CNVs
For discovery of rare CNVs, we removed between 12 and 15 (*k)* singular values, a number which we empirically adjusted based on the inflection point of the "scree plot" (Fig S2), as well as by manual inspection of the final normalized data. To reduce the false positive rate of discovery for rare CNVs, we applied a 15-exon centrally-weighted moving average across exons. We set discovery thresholds at -1.5 or +1.5 for rare deletions and duplications, respectively, and required at least three exome probes to exceed the threshold. To account for the fact that smoothing shrinks the apparent size of discovered events, regions which exceeded this threshold were slightly expanded until the sample's smoothed value crossed within two standard deviations surrounding the population mean of the smoothed values (Fig S1b).

Sample-level quality control:
We excluded ESP exomes from the final background distribution if our algorithm predicted more than 10 calls, as we noted that these samples had a greatly increased total call count (up to 111 calls/sample), and that the calls were largely false positives. This resulted in the exclusion of a total of 80 of 613 initial exomes (87% pass rate) ESP exomes from the background distribution, leaving our final set of 533 exomes. No exomes from the HapMap cohort (range: 1-7 calls per individual) or the autism cohort (range: 0-14 calls per individual) were excluded.

Genotyping CNPs:
For genotyping copy number polymorphic (CNP) regions of the genome, as well as assessing the copy-number of multi-copy genes, we developed a slightly modified approach. Starting from zRPKM values, we again applied the SVD transformation, but opted to remove only five components, in order to prevent the SVD algorithm from

remove *bona fide* signal from the regions of interest. We genotype each individual by determining the average, resulting in the “SVD-ZRPKM value”.

Whole Genome Copy Number Correlations:
To estimate the absolute copy number at CNP loci, read-depth from independent whole-genome sequencing (as previously described in (Sudmant et al., 2010)) was used. Briefly, regions of known copy-number were used to create a copy-number standard curve, and the absolute copy number of tiling 1kb windows across the genome was estimated. For genotyping, the median of the 1kb window estimates was used.

Because we wanted to assess a correlation between exome and whole-genome based methods, we only included loci in the final set if the whole-genome copy number estimate indicated that the locus was polymorphic among the seven HapMap samples tested. We defined a locus to be polymorphic if the absolute range of copy numbers amongst the HapMap samples was greater than 1. Finally, we defined the median copy number of each locus as the median of the absolute copy number estimates among the seven HapMap samples.

Absolute copy number estimation using population frequency information:
To convert relative SVD-ZRPKM values into absolute copy numbers, we used an unsupervised clustering algorithm to cluster SVD-ZRPKM genotype values, and then leveraged genotypes from 43 CNPs in a large set of HapMap samples from (Campbell et al., 2011) to match clusters to absolute copy number.

Unsupervised clustering was done using a mean-shift algorithm implemented in the python package SciKits.learn. The mean-shift algorithm is similar to k-means clustering, but does not require *a priori* information regarding the number of clusters. After clustering, we automatically merged clusters together if their centers were not spaced linearly on the x-axis, as we found that this marginally improved the clustering for some loci. Finally, we fit the most common copy-number state(s) for each locus from (Campbell et al., 2011) to the largest cluster(s) identified by the exome-based SVD-ZRPKM values by maximizing the r^2 value between the two vectors (from each data source) of copy-number states. In other words, we attempted to match the frequencies of each copy number state identified by (Campbell et al., 2011) to consecutive clusters identified by our clustering method. To determine an absolute copy number genotype of a CNP locus for a HapMap sample, we simply determined to which cluster the sample belonged and the matched absolute copy number for that cluster.

Sensitivity call set for HapMap Samples:
To assess sensitivity, we started with CNV calls from the discovery experiment from Conrad and colleagues (Conrad et al., 2010) as a gold standard. This list contained at first 6919 calls for the 5 overlapping hapmap samples in our set. Of these, 486 overlapped at least 3 exome probes (required by our discovery algorithm). Because segmental duplications are prone to array-CGH reference and detection bias, we

removed 416 calls for which 50% of the underlying exome probes were in segmental duplications. Finally, we removed 20 calls found in somatically rearranged regions:

chr2:89156874-89630175	Ig light chain kappa
chr6:32386993-32787910	HLA
chr6:31226231-31328167	HLA
chr14:105994256-107283087	Ig Heavy chain
chr22:22380820-23265082	Ig light chain lambda
chr7:141975722-142519580	T-cell receptor beta subunit

This resulted in 50 calls. For each call, we reviewed several data sources: 1) Illumina i1M or 650Y (for NA15510) SNP array LogR intensities and B-allele frequency, 2) whole genome copy number estimates (from (Sudmant et al., 2010), but not available for NA15510), 3) fosmid-based calls from (Kidd et al., 2008) and 4) SVD-ZRPKM signal across ESP and HapMap samples. We manually curated the 50 calls into four categories: Rare CNVs (5 total), CNPs or CNP-like (42 events), and false positives in the Conrad et al. set (3 calls). False positives had no corroborating evidence in any other data set, and were not counted towards the sensitivity estimates.

Discovery of rare CNVs in ASD trios:

Using the input set of 366 ASD cohort individuals (122 probands) with 366 randomly picked ESP samples, and removing 15 components, our algorithm made a total of 1,043 calls among the 366 individuals in the ASD cohort (with 369 calls in probands), with each sample having between 0 and 14 calls; overall 340 individuals had at least one call. Merging all overlapping calls in the ASD resulted in 282 CNVRs.

As the exome capture reaction targets many genes present in duplicated regions of the genome, and as many exons share homologous sequence, a significant proportion of our calls in probands are due to changes in the copy number of these genes due to independent assortment of parental haplotypes. Starting with the 317 autosomal calls made in the 109 probands for which we also were able to obtain SNP microarray data, we filtered calls to enrich for “rare” CNVs. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of underlying exome probes within the call also in segmental duplications) with segmental duplications were removed (142, or 45%). Next, we calculated the median copy number of calls based on whole-genome read-depth copy-number estimates from ~660 genomes (Sudmant et al., 2010), and additionally filtered 10 calls (3.1%) with more than 3+ copies population-wide (as events stemming from these segmentally-duplicated or higher-copy regions of the genome are likely due to the independent assortment of parental haplotypes, and not “true” rare CNVs). Additionally, we manually curated the calls to remove calls within regions undergoing somatic rearrangement (one call at the *IGH* locus), and merged adjacent or overlapping calls. These steps left 124 calls, and these calls were primarily found in non-duplicated genes and diploid regions of the genome. We categorized each call into one of three bins: de novo, inherited or copy-number polymorphic (Table S3).

Comparison of mrsFAST- and BWA-based read-depth estimation

BWA-based mappings were generated using the default settings for BWA (0.5.6) and post-processed with a pipeline developed specifically for SNP and single nucleotide variant (SNV) discovery. Reads which had more than one high-quality mappings were removed from the alignment and a minimum mapping quality (MAPQ) of 30 was required of all reads. The same method for generating RPKM values from BWA alignments was used as was for mrsFAST-based alignments. We calculated RPKM values for the same 194,080 intervals used elsewhere in this report, and again excluded targets with a median RPKM < 1, a total of 7,117 probes in this experiment.

To make up the sample set for the comparison experiment, we combined 492 ESP samples, for which we had both mrsFAST and BWA-based mapping information, with the 8 HapMap samples. We noticed the the overall variance (as determined by the scree plot) in the BWA-based mapping was lower, and opted to remove only 6 components of variance. For the mrsFAST-based mappings, we removed the usual first 12 components. All other processing steps were done in the same fashion as elsewhere in this paper.

The signal-to-noise ratio for calls was calculated using the formula

$$SNR = |\mu_{call}| / \sigma_{chromosome}$$

where μ_{call} is the mean of the SVD-ZRPKM values for the exons within a call, and $\sigma_{chromosome}$ is the standard deviation of all the SVD-ZRPKM values of the call's chromosome. We calculated the SNR for the seven rare validated calls from table S1 for both mrsFAST-based and BWA-based SVD-ZRPKM values (Table S6). Six of seven rare CNVs showed improved SNR using the mrsFAST-based mappings, with a median improvement of 58% over BWA (mean 38% improvement).

Comparison to ExomeCNV algorithm:

We compared our algorithm to the previously published ExomeCNV (Sathirapongsasuti et al., 2011) in order to better understand the strengths and weaknesses of each. ExomeCNV is designed to detect copy number aberration in the context of cancer, a special case of copy number variation which requires additional parameters to be defined (e.g., the rate of admixture/contamination of tumor and normal), and which must be able to handle samples for which a large fraction of the genome is not diploid. Accordingly, ExomeCNV is designed around a digital comparative hybridization algorithm, which requires that both the test and reference are as closely matched as possible (e.g., tumor-normal pairs of exomes from the same capture and sequence), and includes many features to better characterize cancer exomes. In contrast, ours is designed to discover genic deletions and duplications of exonic regions independently in each sample by first eliminating systematic noise using singular value decomposition.

We compared the ability of both algorithms to detect germline variation in DNA samples extensively analyzed and validated as part of other studies. To assess the sensitivity

and specificity of both algorithms, we used the five HapMap samples for which exome sequence data had been generated and where high-density microarray analyses had been performed previously (Conrad et al., 2010). We set NA19240 as the reference sample, and used ExomeCNV to call CNVs on the remaining four samples (NA12878, NA15510, NA18517, and NA19129). Similar to the authors own use of the NovaAlign alignment package, we used the available BWA alignments for this comparison, and used the same 194,080 probes to generate an interval coverage file using the GATK (version 1.3.8) software package. We left all ExomeCNV parameters at their default values: sensitivity and specificity were set at 0.9999 for exons (maximizing specificity) and 0.99 for calls ("auc" option), and the admixture rate was set at a conservative 0.5 (despite the fact that we did not expect any biological admixture, we found that keeping this setting reduced the number of false positive calls).

Among the four test samples, ExomeCNV predicted 450 CNVs, of which only 63 (14%) overlapped with calls in the Conrad et al. call set by more than 10% reciprocal overlap. In contrast, our algorithm found 24 calls among these four samples, of which 21 (87.5%) overlapped the Conrad et al. set. While both programs were able to find all of the five rare CNVs (Table S3), we note that ExomeCNV predicted 16 CNVs larger than 500kb, which did not have any overlap with the high resolution Conrad et al. set of calls. This low specificity would make it very difficult to find "true positives" in the ExomeCNV output, even when filtering for large CNVs only.

Using exon-level log-ratio output from ExomeCNV, we next compared how sensitive it was to changes in copy-number of duplicated genes. Across the 62 CNP loci genotyped by our algorithm (Table S4), ExomeCNV was able to generate LogR values for 51 loci (82%). Example correlations and a comparison between ExomeCNV and our algorithm are shown for four loci in Figure S8a. Across all loci, when compared to the log-ratio values to the whole-genome estimate for each locus, the median r2 across these loci was 0.57 (c.f. this work's algorithm r2 = 0.92). As with the BWA alignment comparison, the genotyping dynamic range of ExomeCNV was severely limited, and the LogR values from ExomeCNV correlated only poorly with the corresponding whole-genome estimates of absolute copy number for loci with median copy number greater than seven (Figure S8c).

Finally, although the authors of ExomeCNV recognize that their algorithm depends on sample-to-sample consistency, large cohorts of tens to hundreds of exomes cannot be expected to maintain such consistency. Crucially, our algorithm allows for the comparison of samples from different cohorts, and even different iterations of the exome capture reaction itself. To demonstrate this, we examined two ESP samples from two different experimental cohorts (but stemming from the same study, and using the same capture kit version, library preparation steps and sequencing machines). The output from ExomeCNV for chromosome 20 is shown in the top left panel of Figure S7. When we counted the fraction of exome probes which ExomeCNV predicted as copy-number variant, we found that a biologically implausible 96.6% of the exome was detected as changed from diploid copy number (Figure S7, top right panel). In contrast, when we picked an ESP sample from the same experimental batch (and which was closely

matched based on the variance we observed using the SVD decomposition) as the reference, ExomeCNV reported only 0.4% of exome probes as non-diploid (Figure S7, bottom panel). When we applied our algorithm (this work) at a very sensitive setting (± 1 SVD-ZRPKM threshold), we found only that for the same samples, only 0.06% and 0.15% of the exons were altered from diploid. This comparison highlights the strength of singular value decomposition for eliminating batch effects and systematic noise that may arise from exome capture experiments.

Quantitative PCR conditions and primers:

We performed SYBR Green qPCR on 3 loci, using primers listed below. Each reaction was performed in quadruplicate using 10ng of template DNA per reaction. C(t) values averaged for each sample across technical replicates and fold change calculated the $\Delta\Delta C(t)$ method.

Primers:

DOCK6	Fwd	TGCATTTGTTTGATCCGTGT
DOCK6	Rev	TGGGATTTTGTTGGGATGAT
HAVCR1	Fwd	GCAGAAGGGAGACATGAAGC
HAVCR1	Rev	AGACACTGGGAGGGGAAACT
BTNL3/8	Fwd	GTCAGATGGGGTTTTGCTGT
BTNL3/8	Rev	AGGCAAACCGTGAAAACAAC
Albumin Ctl	Fwd	GTGGGCTGTAATCATCGTCT
Albumin Ctl	Rev	TGCTGGTTCTCTTTCACTGAC

References:

Campbell, C. D., Sampas, N., Tsalenko, A., Sudmant, P. H., Kidd, J. M., Malig, M., Vu, T. H., et al. (2011). Population-Genetic Properties of Differentiated Human Copy-Number Polymorphisms. *The American Journal of Human Genetics*, *88*(3), 317–332. doi:10.1016/j.ajhg.2011.02.004

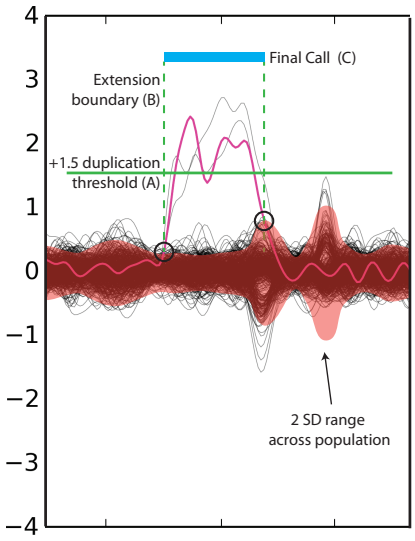
Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., et al. (2010). Origins and functional impact of copy number variation in the human genome *Nature*, *464*(7289), 704–712. doi:10.1038/nature08516

Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., et al. (2008). Mapping and sequencing of structural variation from eight human genomes *Nature*, *453*(7191), 56–64. doi:10.1038/nature06862

Sudmant, P. H., Kitzman, J. O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., et al. (2010). Diversity of human copy number variation and multicopy genes *Science*, *330*(6004), 641–646. doi:10.1126/science.1197005

Sathirapongsasuti, J. F., Lee, H., Horst, B. A. J., Brunner, G., Cochran, A. J., Binder, S., Quackenbush, J., et al. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*, *27*(19), 2648–2654. doi:10.1093/bioinformatics/btr462

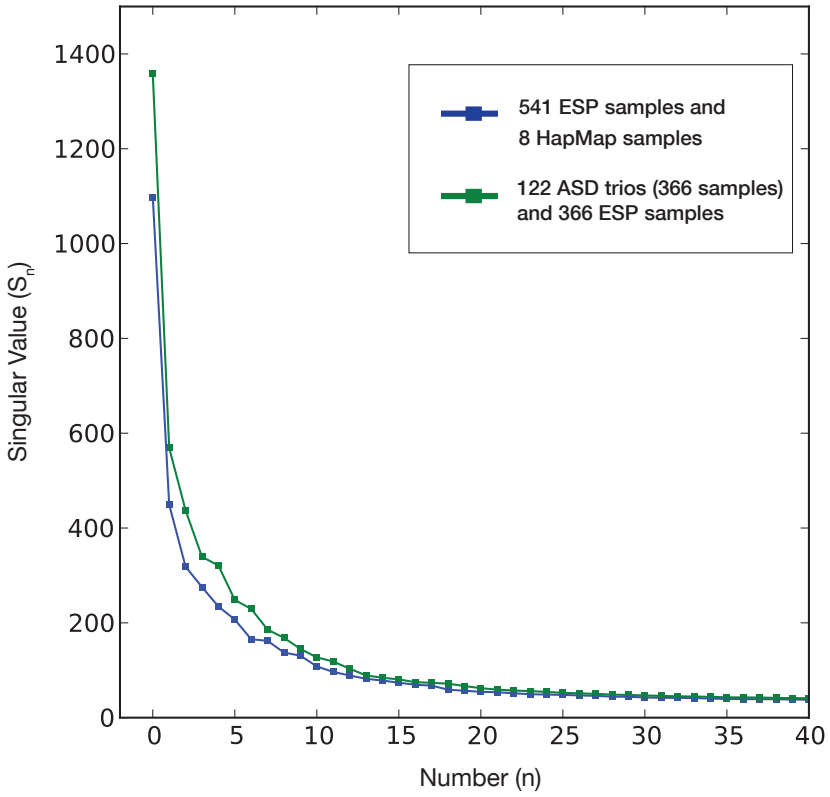
Figure S1: Threshold call overview:



S1: Threshold algorithm

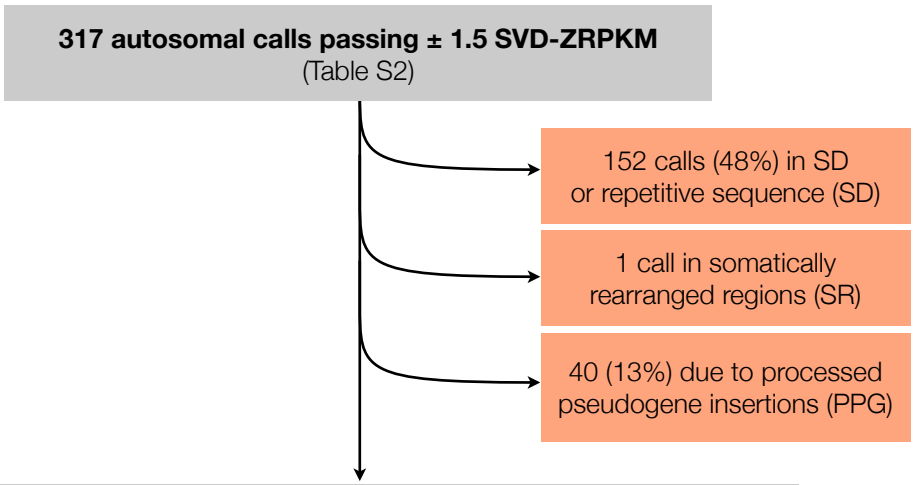
To discover rare CNVs, we found smoothed SVD-ZRPKM values which crossed a threshold (A) of +1.5 or -1.5 for duplications and deletions, respectively. To account for the fact that our smoothed values shrink the apparent size of the call, we extended calls such that the final call (C) better represented the extend of the actual CNV. To do this, we extended calls from the initial supra-threshold event until the smoothed SVD-ZRPKM values dipped below ±2 standard deviations surrounding the population median (red highlight) of the SVD-ZRPKM values (marked in figure by line [B], and by black circles).

Figure S2: Scree Plot



S2: Scree plot
This scree plot shows the first 40 singular values (S_n) from the HapMap (blue) ASD trio (green) samples. The relative contributed variance of each singular value is proportional to its strength indicated on the y-axis.

Figure S3: Calls and validation overview in 122 ASD Probands



124 CNVs to Validate

	SNP Microarray Validation (Sanders et al. 2011)	**Custom array-CGH validation or qPCR** (this work)	**Total Validated**
de novo	6/8 Concordant	0/2 validated	6/8 (75%)
inherited	79/87 Concordant	5/5 validated by CGH 2/2 validated by qPCR	86/87 (99%)
CNP	10/29 Concordant	5/6 validated by CGH 3/3 calls in in *BTNL3/8* locus by qPCR *	25/29 (86%)
All	95/124 (77%)	15/18 (83%)	**117/124 (94%)**

Starting with 317 detected calls in 109 ASD probands, we applied a set of filters to restrict calls to unique/diploid regions of the genome in order to estimate the precision of our method. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of underlying exome probes within the call also in segmental duplications) with segmental duplications or repetitive regions of the genome (152/317, or 48%). One call was located in a somatically rearranged region. Finally, 40 calls were driven exclusively by the insertion of processed pseudogenes elsewhere in the genome (Figure S17, Supplementary note). For the remaining 124 calls, we validated 95 using existing SNP microarray data (Sanders et al. 2011). We next designed a custom array-CGH and qPCR assay to validate additional novel CNVs (Figures S19, S20). Our overall rate for a validated set of CNVs was 117/124.

* For the 10 calls in the *BTNL3/8* CNP, we validated 3 of 3 tested events, and we therefore consider all 10 events at this locus validated.

Figure S4: Filtering of calls from Conrad et al. (2010) array-CGH experiment:

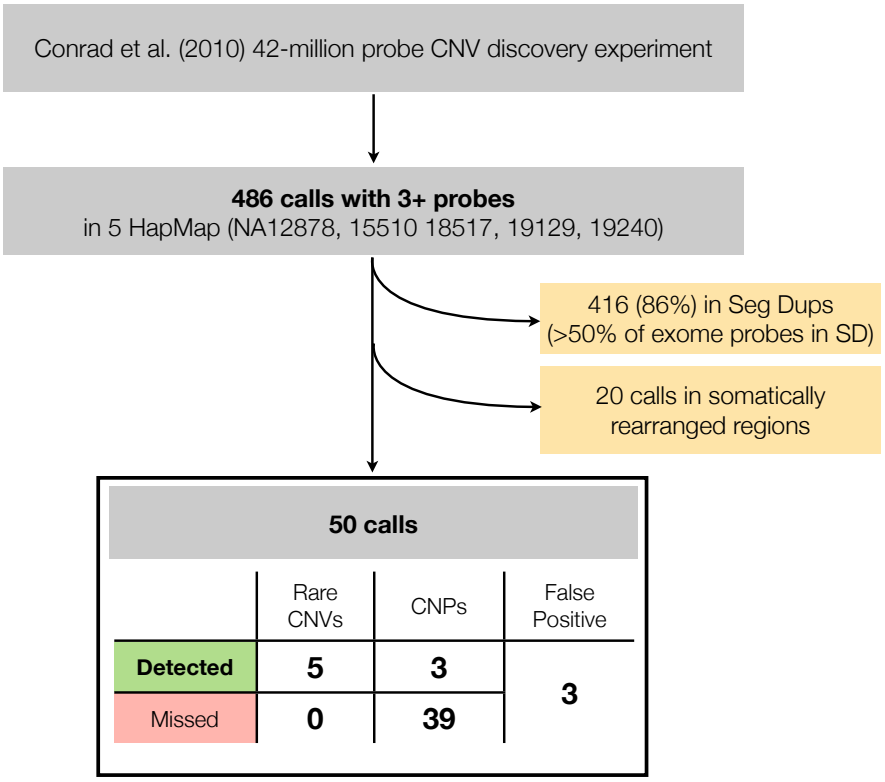


Figure S4: Filtering of calls from Conrad et al. (2010) array-CGH experiment: We estimated the sensitivity of our method using array comparative genomic hybridization calls from Conrad et al. (2010) as a gold standard. Starting with calls from the 42-million probe CNV discovery experiment in (Conrad et al., 2010), there were 486 calls with at least three exome probes in the five HapMap samples for which we had exome sequences. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of exome probes within the call also in segmental duplications) with segmental duplications were removed; additionally, we removed 20 calls in somatically rearranged regions. We manually inspected the remaining 50 calls (Table S2) to assess sensitivity of the method. Five events were rare and all five were detected by the ±1.5 SVD-ZRPKM threshold. There were 36 CNPs, of which only three cross the threshold for rare CNVs. Six of the remaining events were either located in high diversity regions of the genome. Finally, we noted that three of the events were very likely false positive events in the Conrad dataset, as they were not corroborated by Illumina 1M SNP microarray data, nor were they found by a fosmid mapping approach(Kidd et al., 2008).

Figure S5: BWA and mrsFAST comparison – genome view

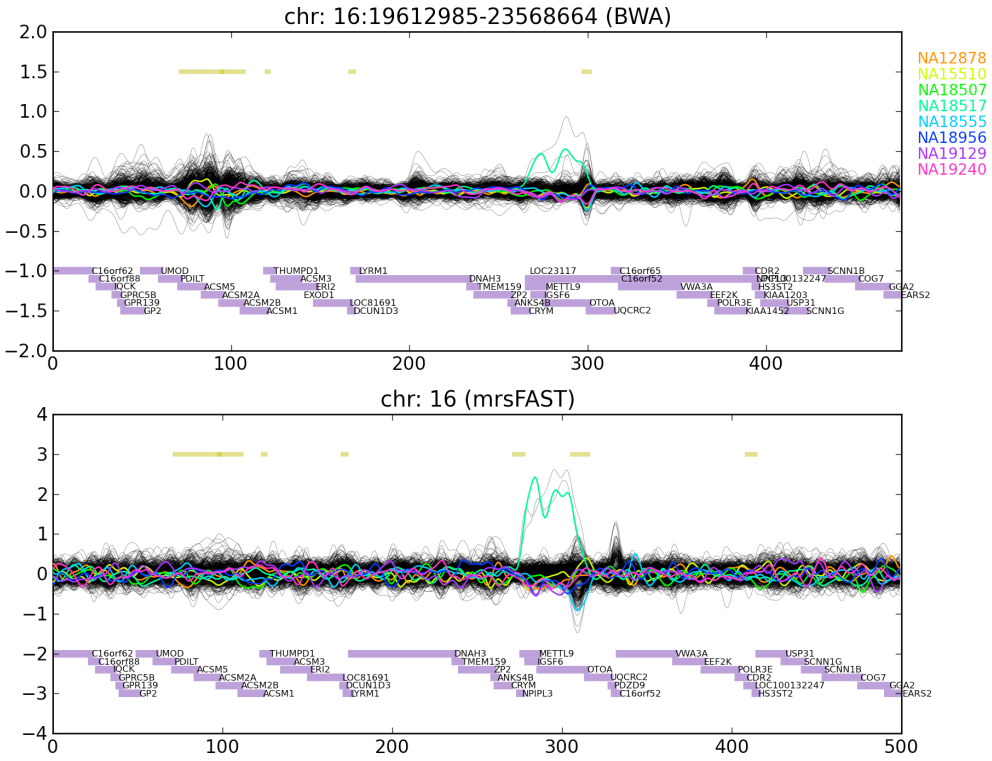


Figure S5:
Visual comparison of BWA and mrsFAST-based mappings on a stretch of chromosome 16. We found that across the seven validated rare CNVs from table S1, the SVD-ZRPKM values derived from BWA mappings had a 57% lower signal-to-noise ratio, as noted by the decreased signal of NA18517 at the *METTL9/OTOA* locus for BWA-based mappings. (Y-axes have different scales to account for the lower standard deviation seen in the BWA-based SVD-ZRPKM values.)

Figure S6a: BWA and mrsFAST comparison – genotyping accuracy

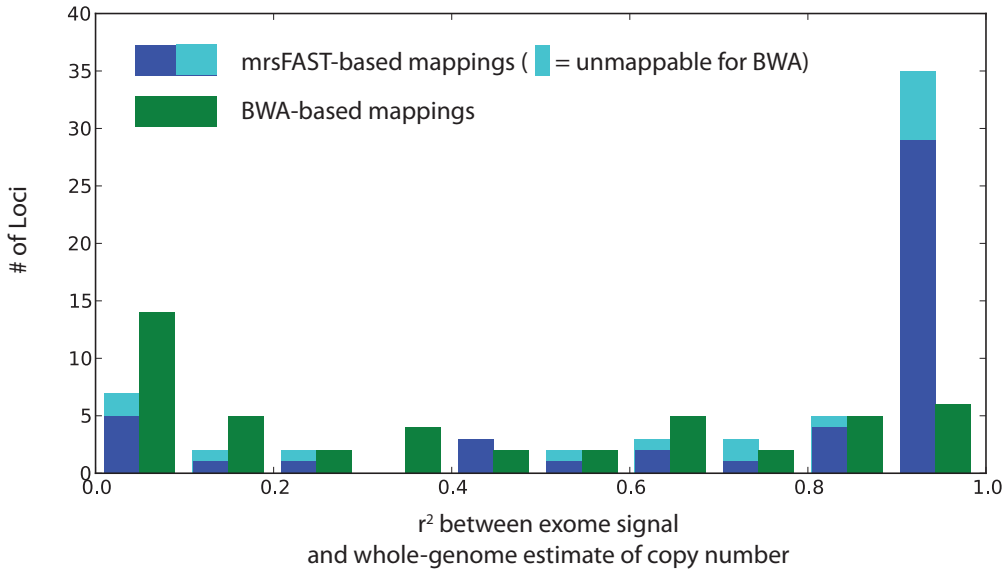


Figure S6a:
Comparison of correlations coefficients of SVD-ZRPKM to whole-genome copy number estimate across 62 CNP loci between BWA- and mrsFAST-based mapping strategies. The median r^2 for the BWA-based experiment is 0.62 (green bars), while for mrsFAST the median r^2 is 0.92 (blue bars). Moreover, for 15 loci, the BWA-based mappings did not have sufficient read-coverage in the loci to be genotyped, making them intractable to BWA-based read-depth genotyping.

Figure S6b: BWA and mrsFAST comparison – by median copy number

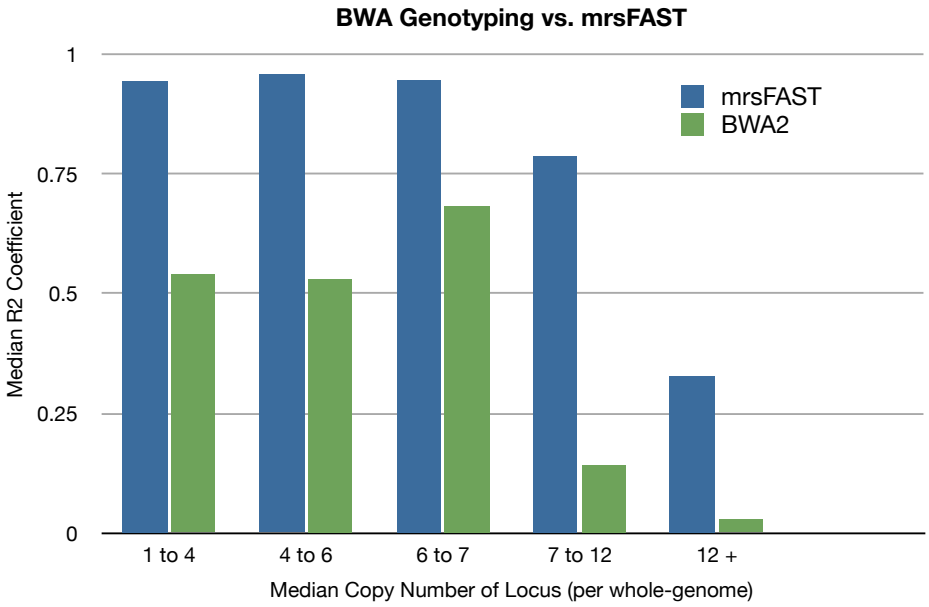


Figure S6b:
Comparison of BWA-based and mrsFAST-based alignments for genotyping of 62 loci, binned by median copy number of each locus. We calculated the median copy number of the 62 loci based on whole-genome read-depth copy-number estimates from ~660 genomes. We note that mrsFAST-based mapping significantly improves the correlation between the SVD-ZRPKM genotyping scores and whole-genome absolute copy number, especially for loci with a median copy number between 7 and 12.

Figure S6c: BWA and mrsFAST comparison – *LRRC37A3* locus

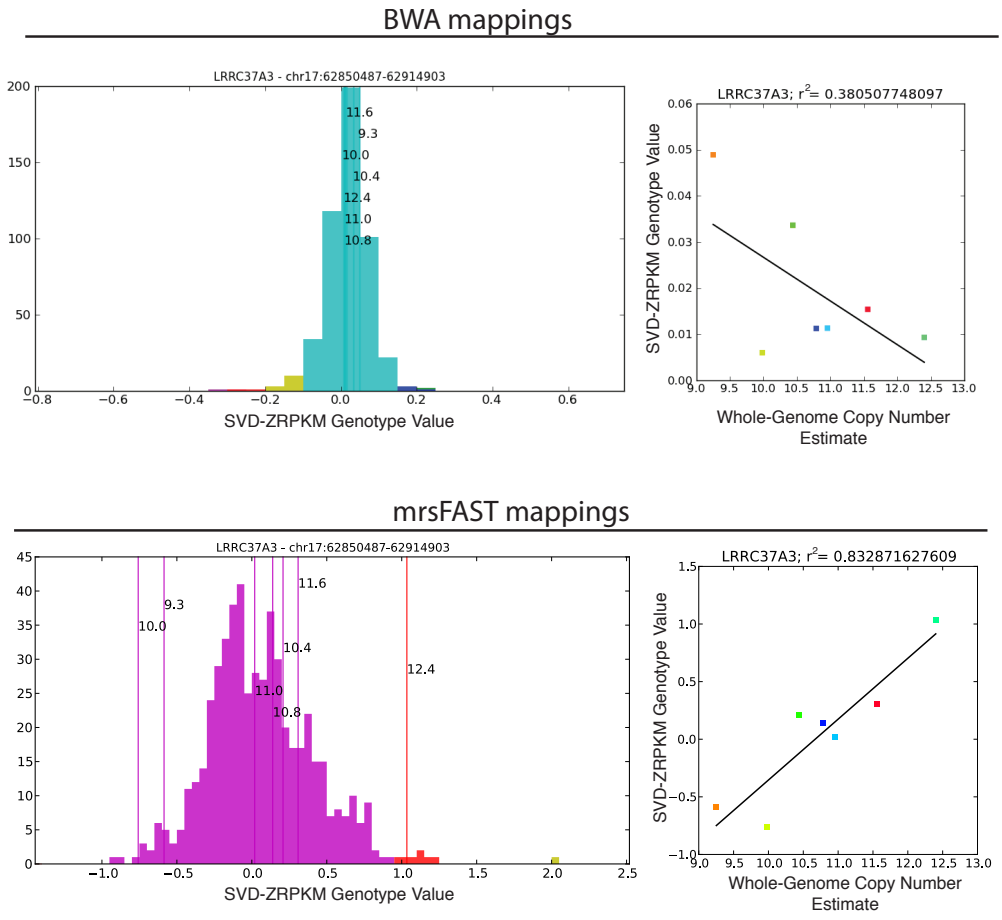
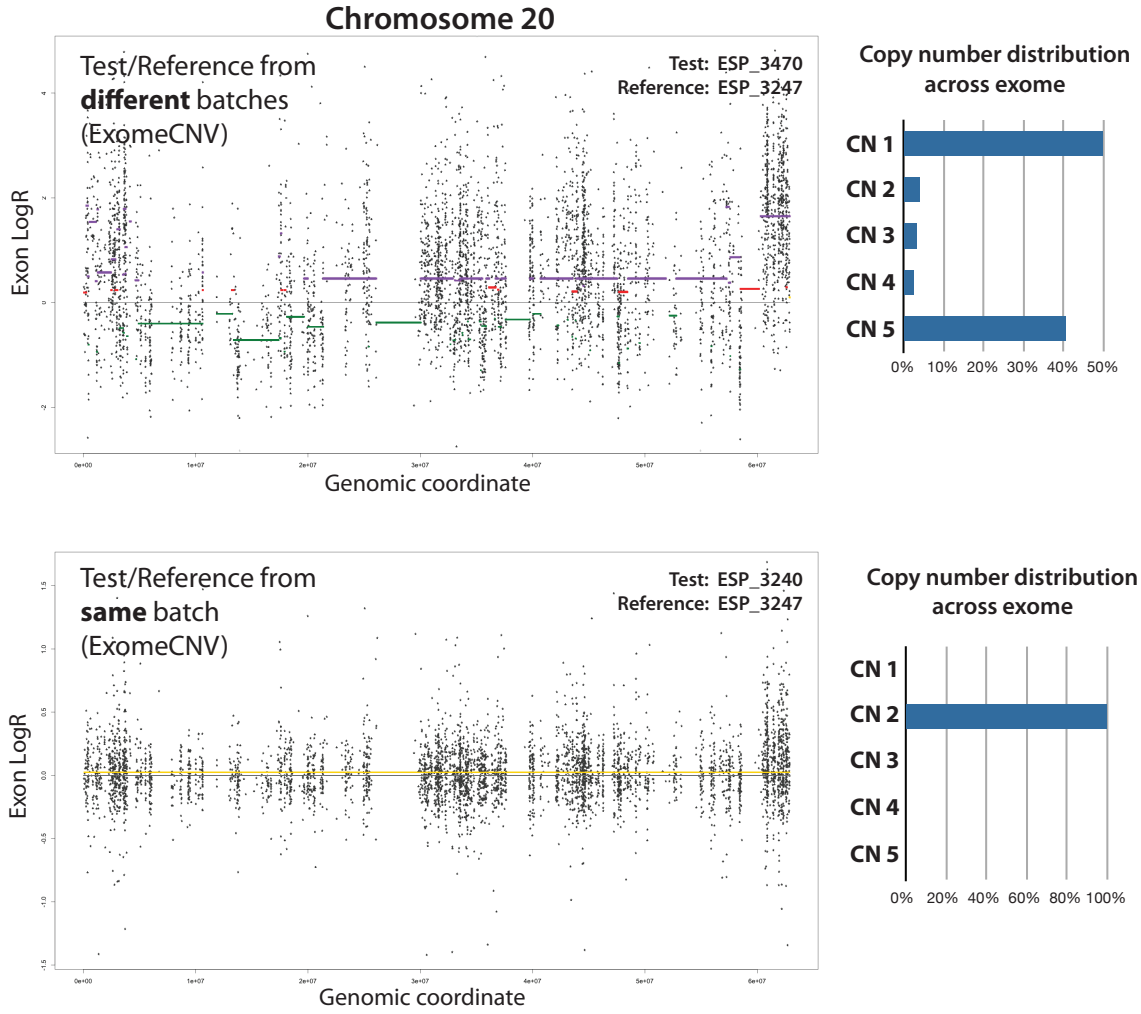


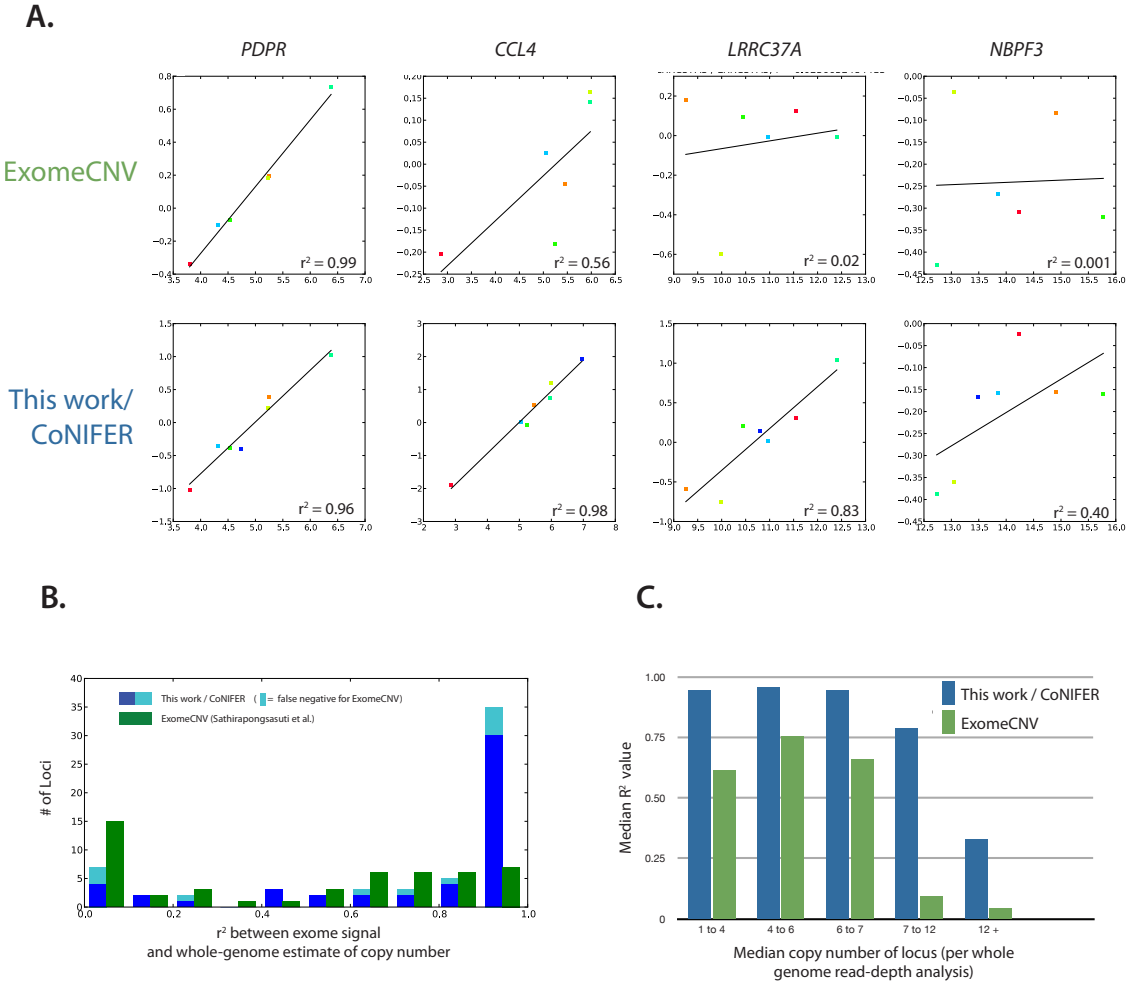
Figure S6c:
Example CNP locus (*LRRC37A3*) representative of difficulty for BWA-based genotyping of loci with median population copy number greater than seven.
(top left): Histogram showing SVD-ZRPKM genotype values of 8 HapMap samples (indicated by horizontal lines) and 492 ESP samples. Annotated numbers on the histogram indicate the absolute copy number, as estimated from whole genome sequencing of HapMap samples.
(top right): Correlation between SVD-ZRPKM values and whole-genome derived absolute copy number for 7 HapMap samples. The poor resolution of BWA-based mappings for this locus contribute to a poor correlation and low accuracy.
(bottom left, right): the same locus for mrsFAST-based mappings. Both the histogram and the scatter plot show markedly increased resolution for distinguishing copy number states and improved SVD-ZRPKM to absolute copy-number correlation.

Figure S7: ExomeCNV results for two references from different cohorts



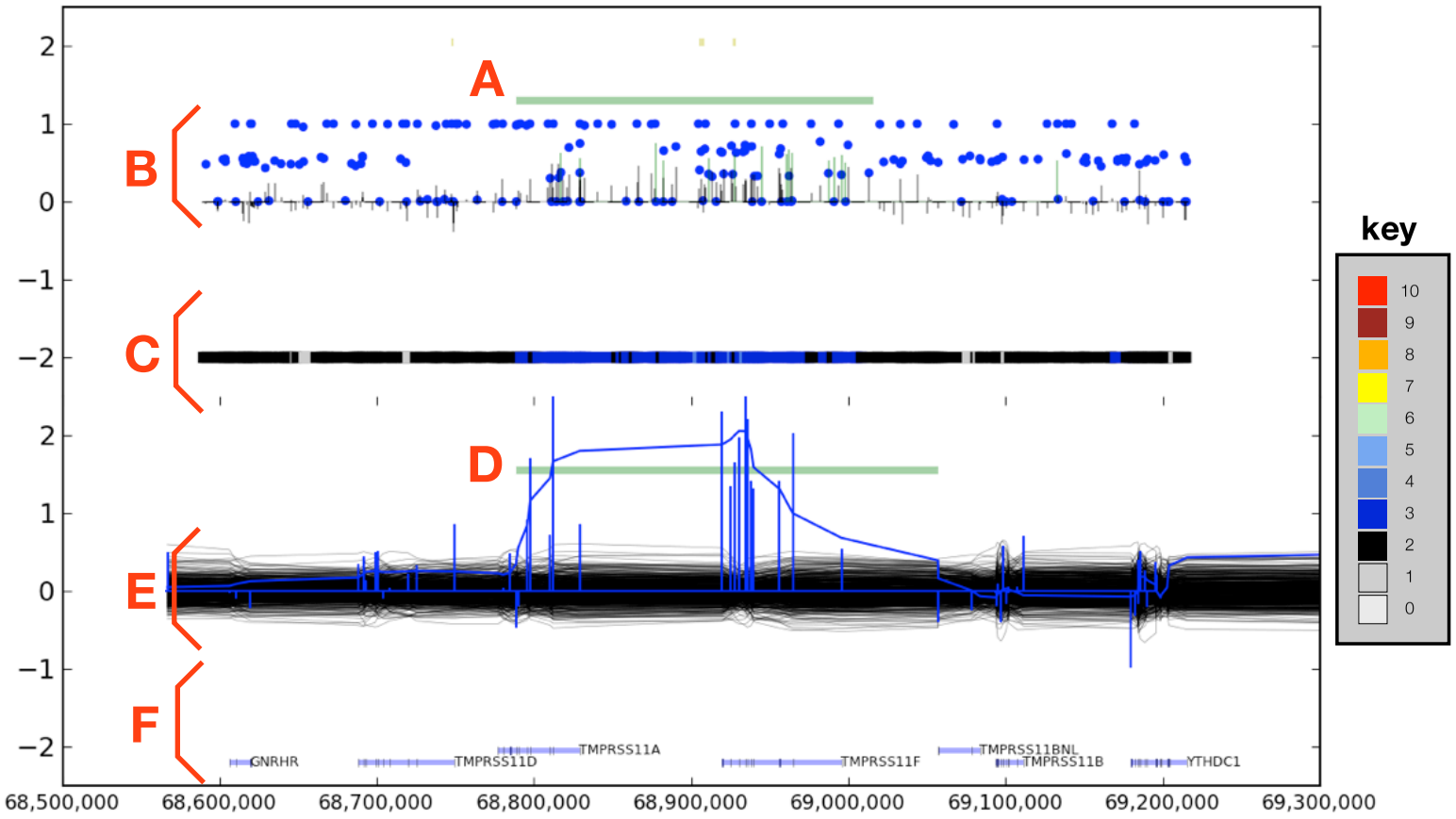
Top: Comparison of two ESP exomes from differing cohorts. Plot shows ExomeCNV LogR output for chromosome 20 and colored bars indicate location of altered copy number. A biologically implausible fraction of the exome (96.6%) is marked as non-diploid (bar chart, top right).
Bottom: Using the SVD algorithm, we matched the same reference (ESP_3247) to a sample from the same cohort/experimental batch. Accordingly, ExomeCNV was less influenced by systematic noise stemming from the exome capture, and marked a much more realistic 99.6% of the exome as diploid.

Figure S8: ExomeCNV and CoNIFER genotyping comparison summary



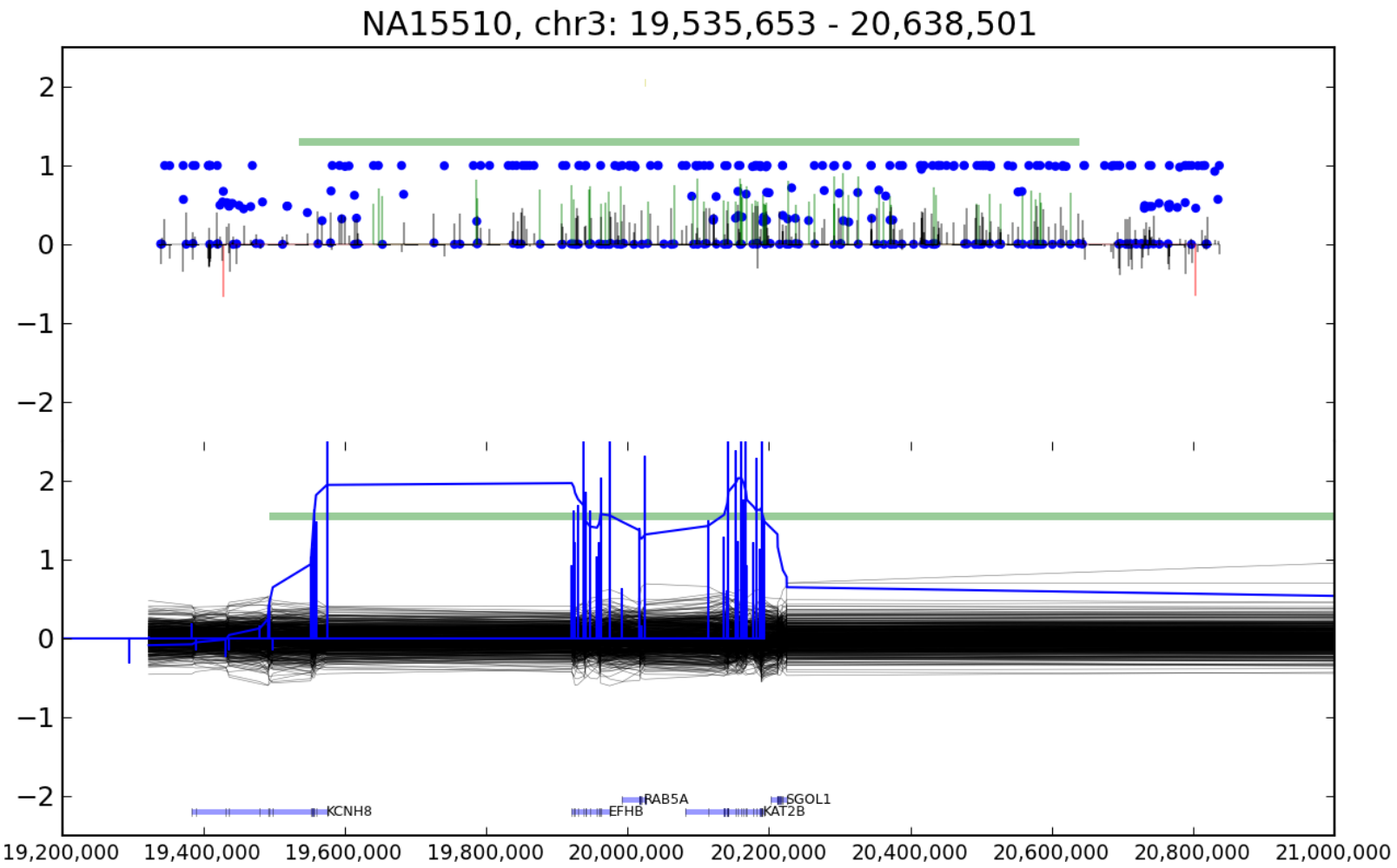
a) Comparison of genotyping correlation between ExomeCNV LogR value (y-axis; top row) and SVD-ZRPKM value (y-axis, bottom row) vs. absolute copy number established by whole-genome read-depth (x-axis, both rows; Sudmant et al., 2010) for four selected loci. b) Distribution of r^2 values across 62 genotyped CNP loci: green bars represent ExomeCNV results (median $r^2 = 0.57$); dark blue bars are the same loci assayed using this work's algorithm (median $r^2 = 0.92$), while light blue bars represent loci which could not be assayed using ExomeCNV (11 loci). c) Median r^2 correlations for ExomeCNV and our algorithm, binned by the median copy number of each CNP locus.

Figure(s) S9

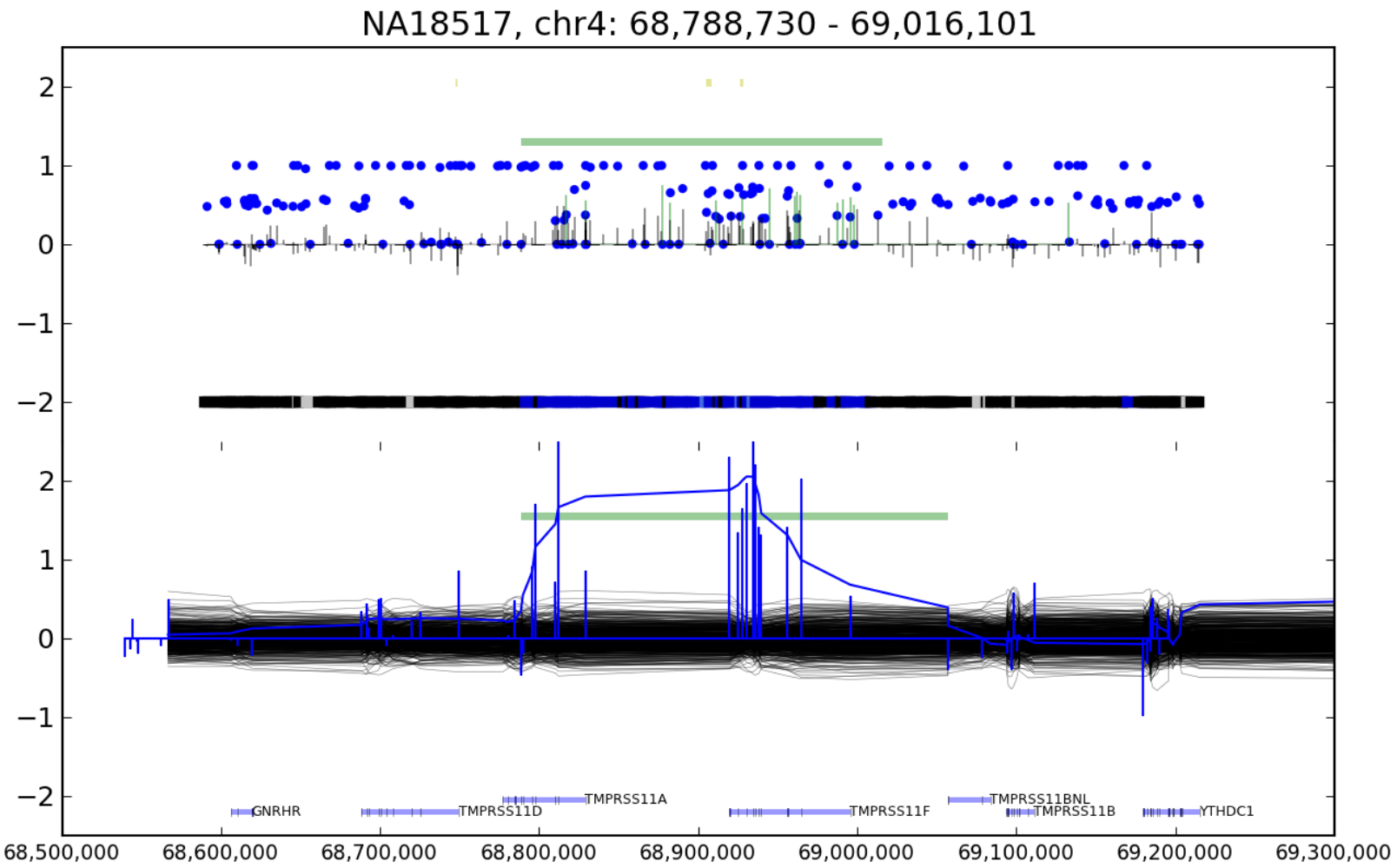


A) CNV call from Conrad et al. (2010)
B) SNP-array data (black lines - LogR; blue dots B-allele frequency)
C) Whole-Genome read depth from Sudmant et al. (2010) – see **key** at right
D) Exome-based CNV call
E) SVD-ZRPKM values (blue line: sample with call; black lines: 533 ESP samples)
F) Refseq Genes

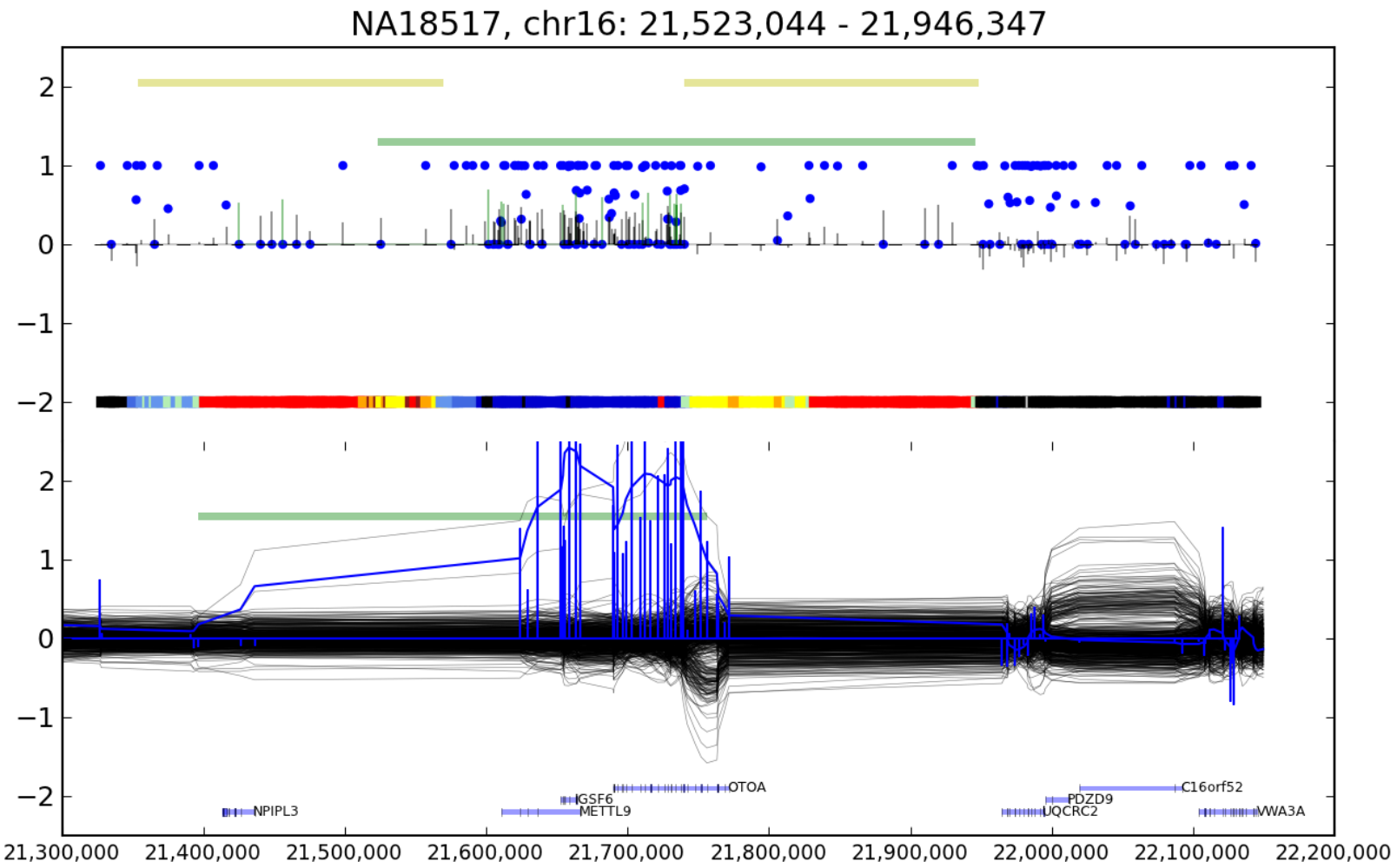
Rare Duplication



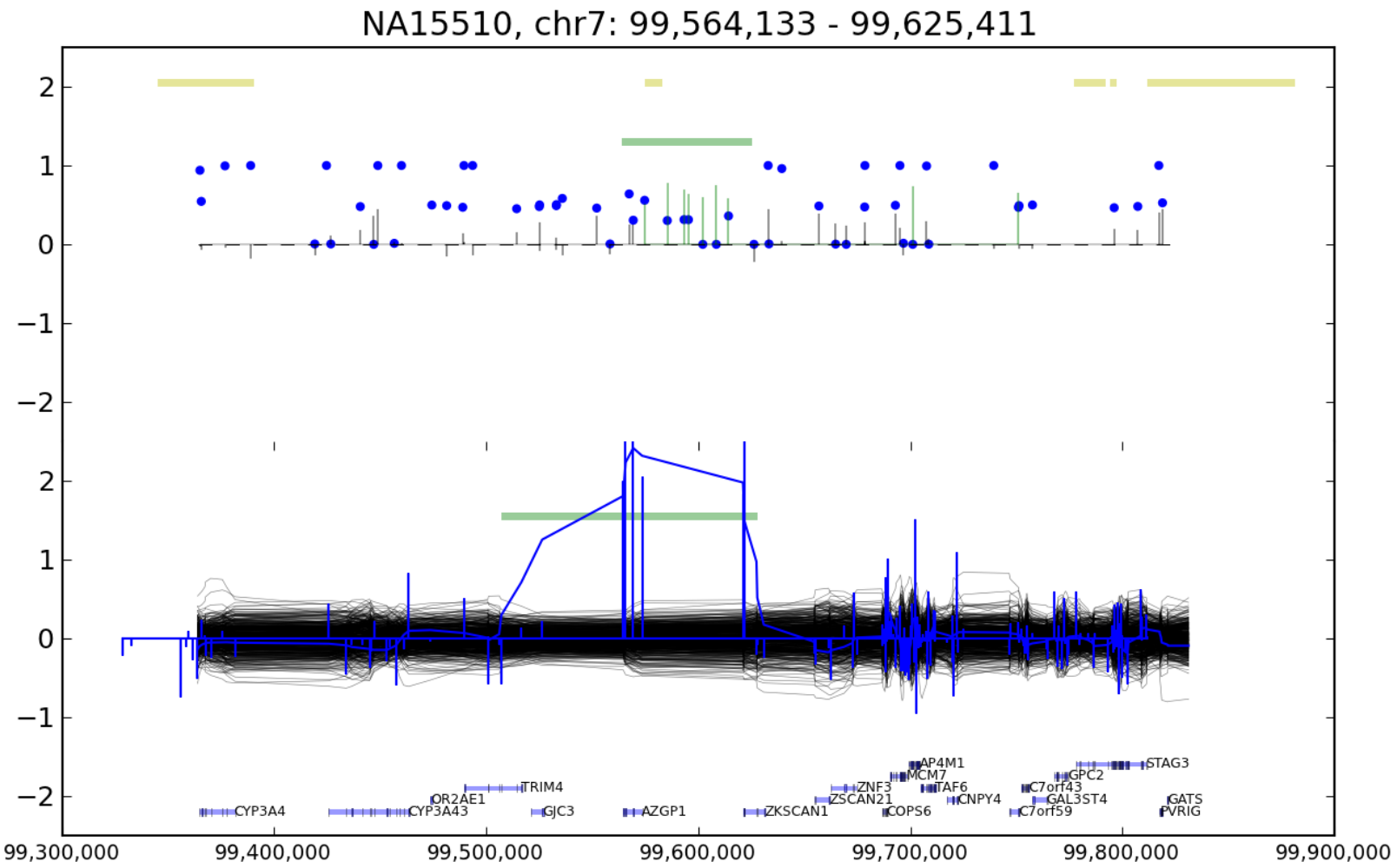
Rare Duplication



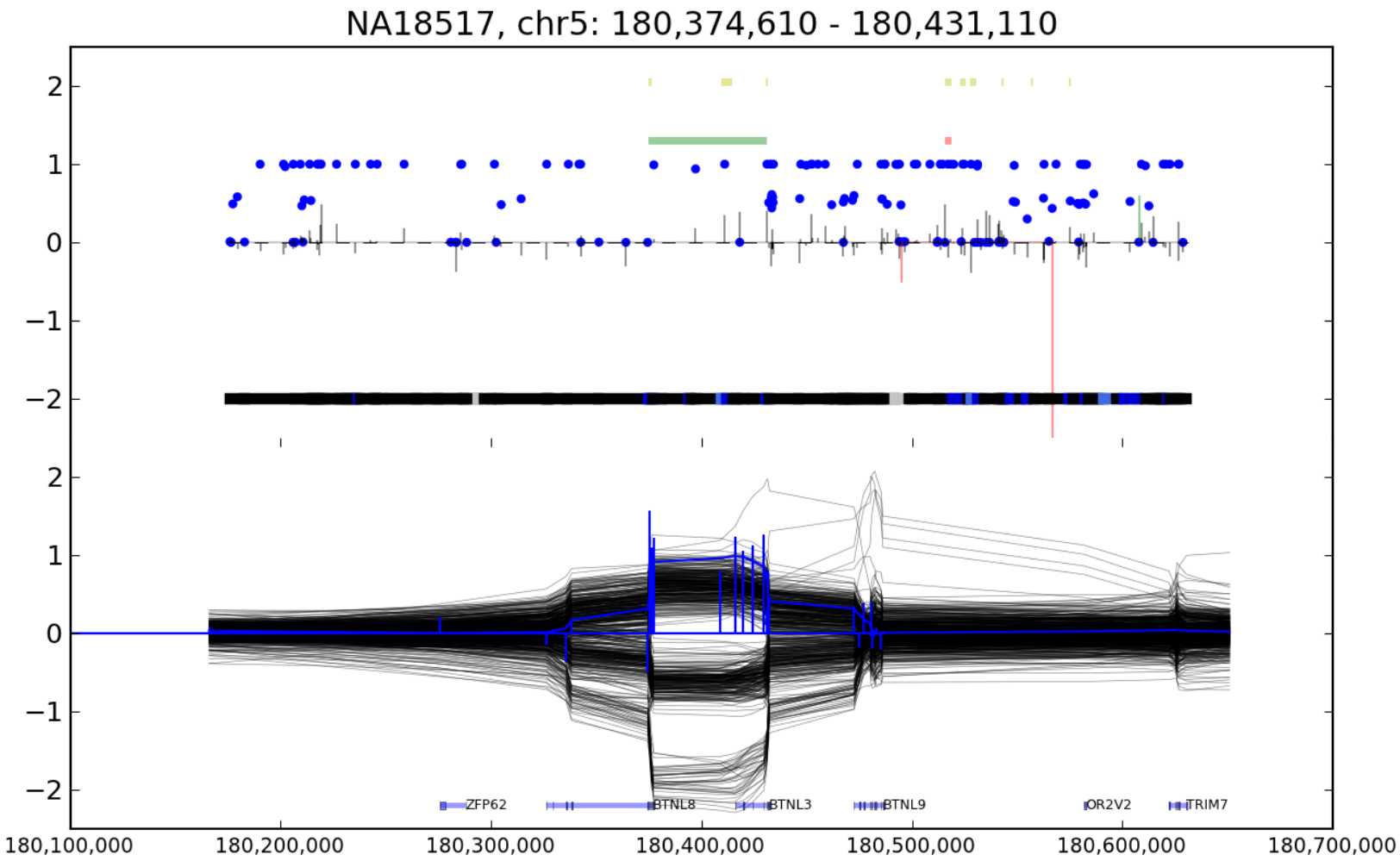
Rare Duplication



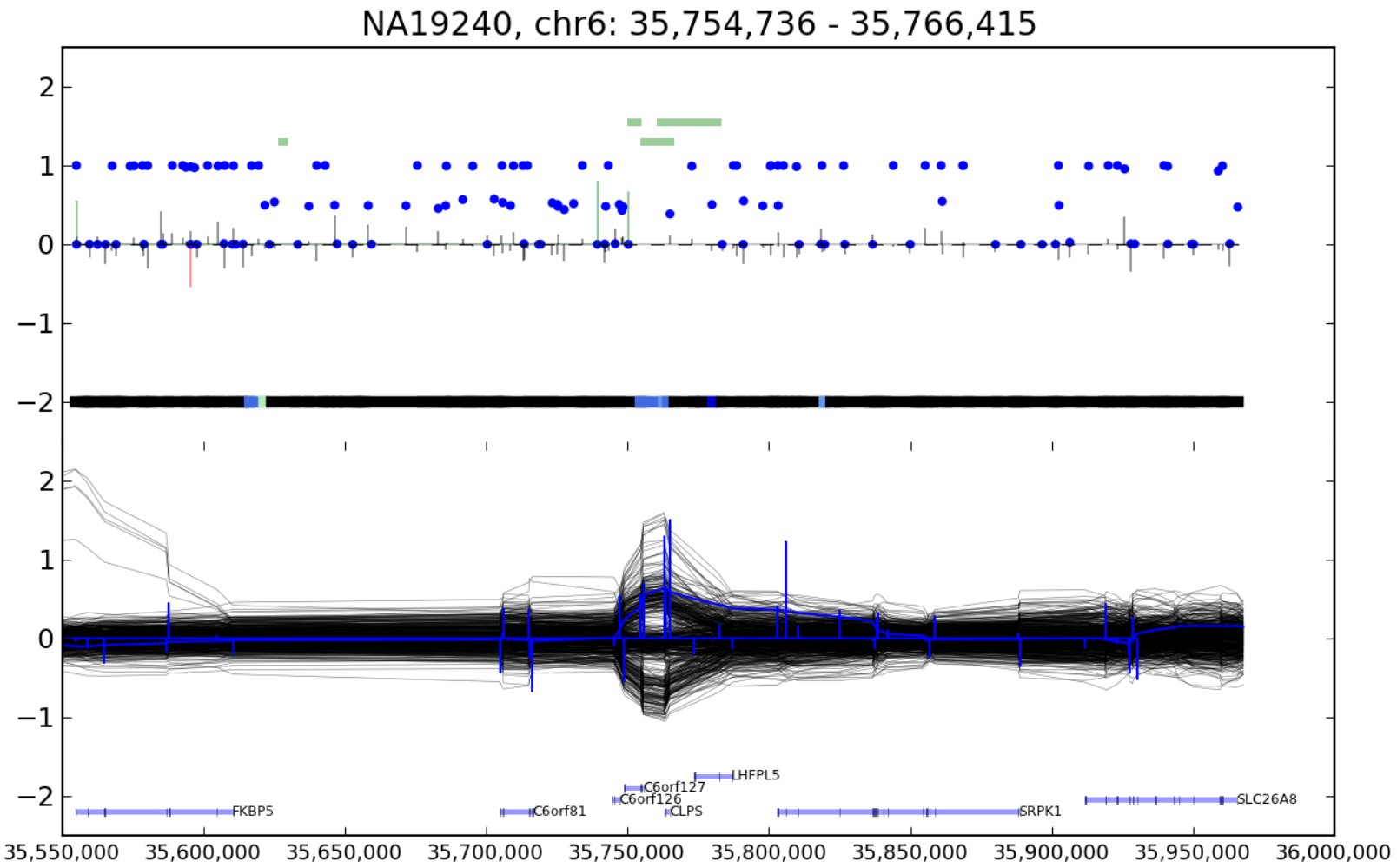
Rare Duplication



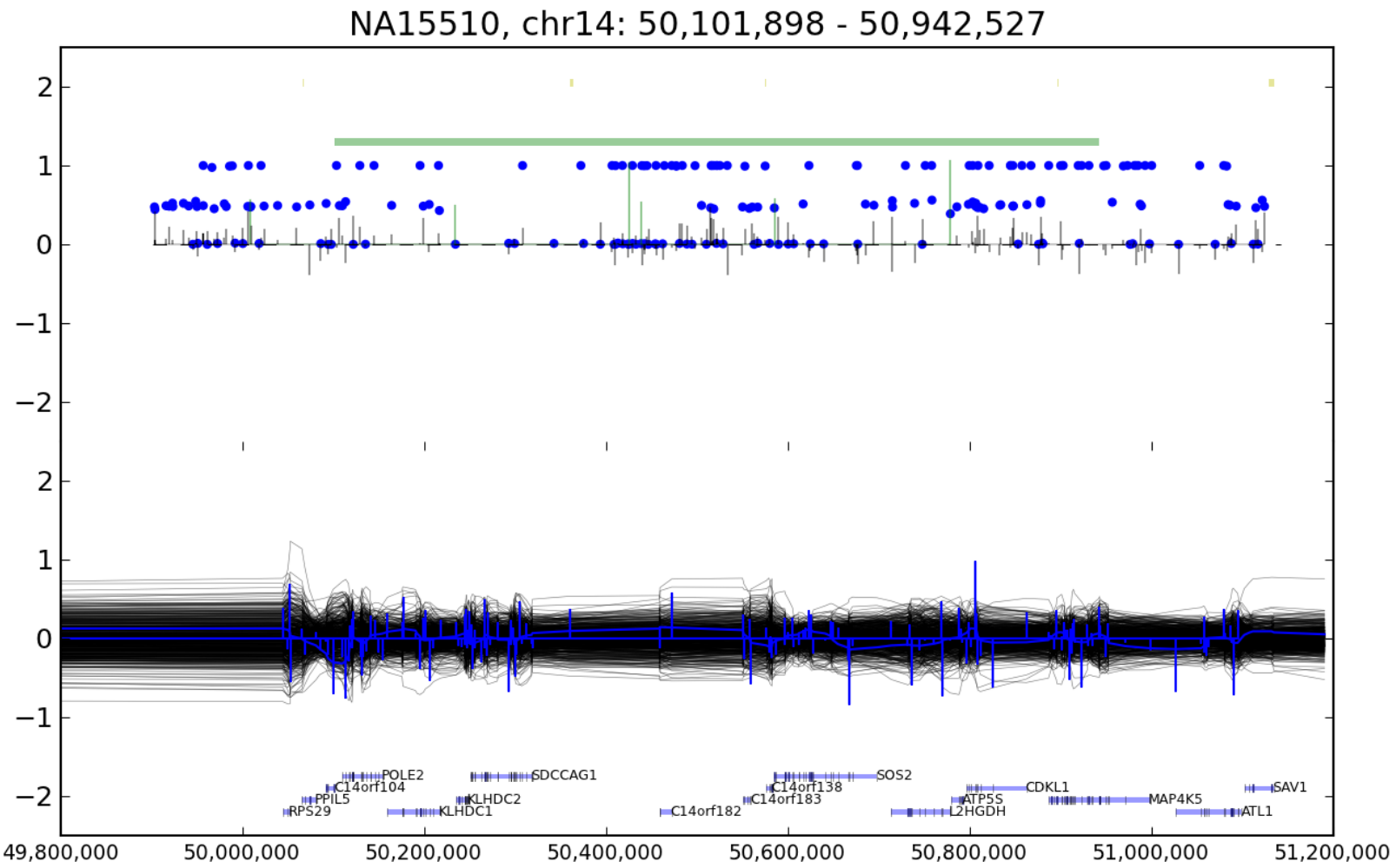
CNP


NA18517, chr5: 180,374,610 - 180,431,110
2
1
0
−1
−2
2
1
0
−1
−2
ZFP62
BTNL8
BTNL3
BTNL9
OR2V2
TRIM7
180,100,000
180,200,000
180,300,000
180,400,000
180,500,000
180,600,000
180,700,000

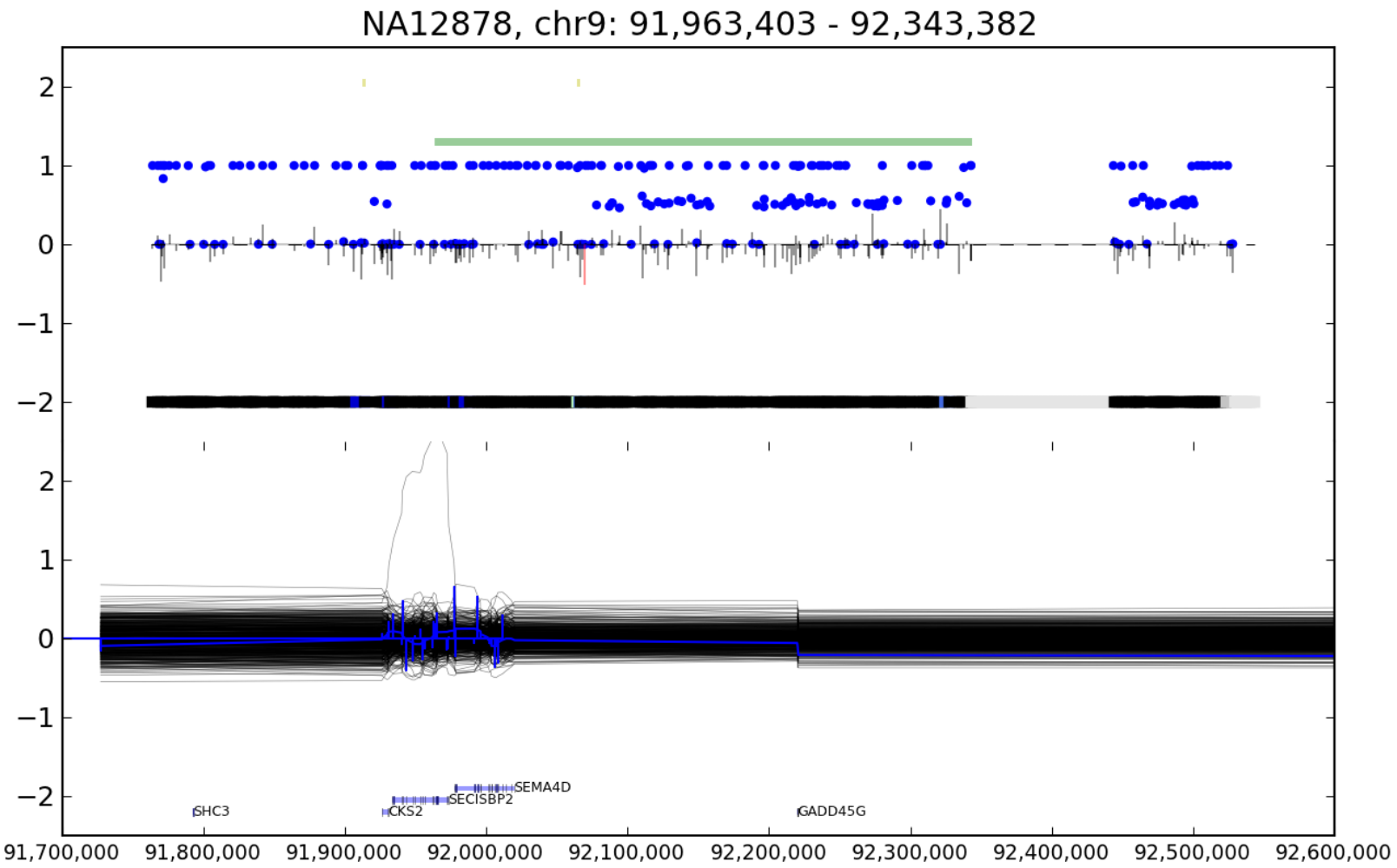
CNP


NA19240, chr6: 35,754,736 - 35,766,415
2
1
0
−1
−2
2
1
0
−1
−2
LHFPL5
C6orf127
C6orf126
CLPS
FKBP5
C6orf81
SRPK1
SLC26A8
35,550,000
35,600,000
35,650,000
35,700,000
35,750,000
35,800,000
35,850,000
35,900,000
35,950,000
36,000,000

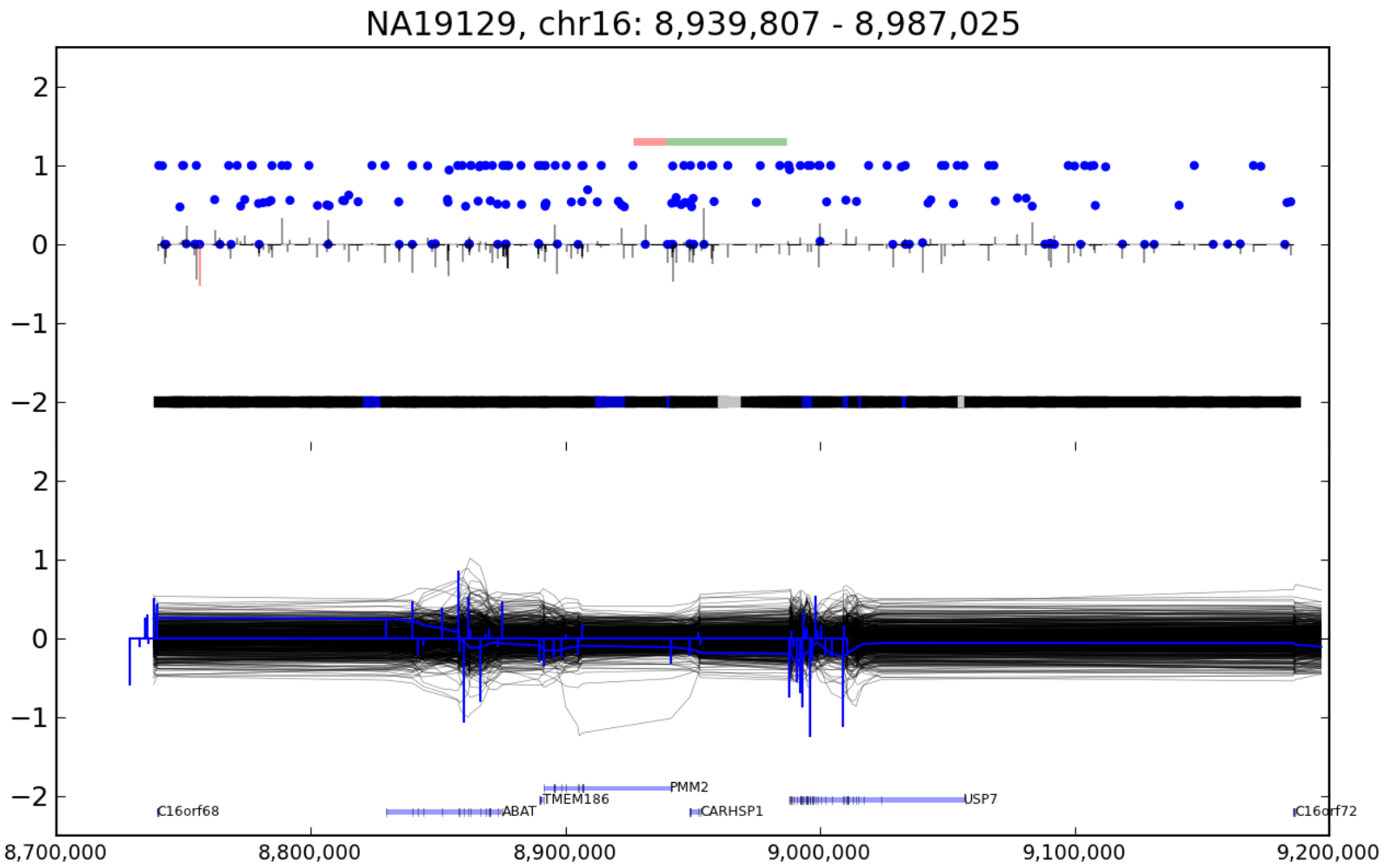
False Positive



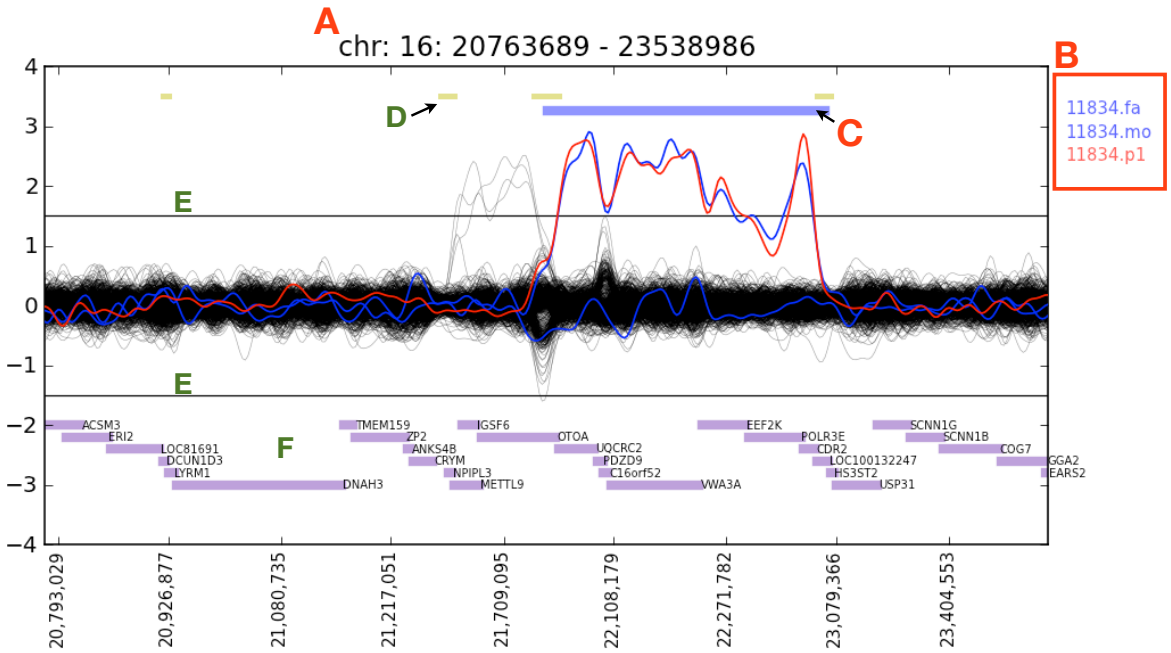
False Positive



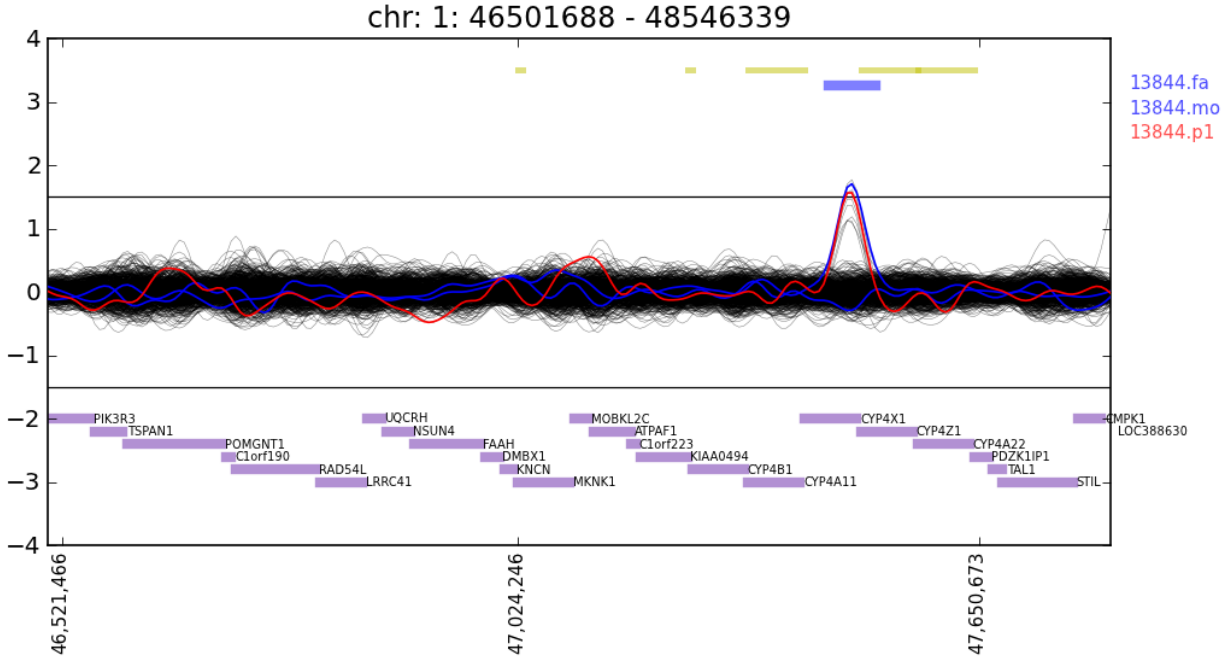
False Positive

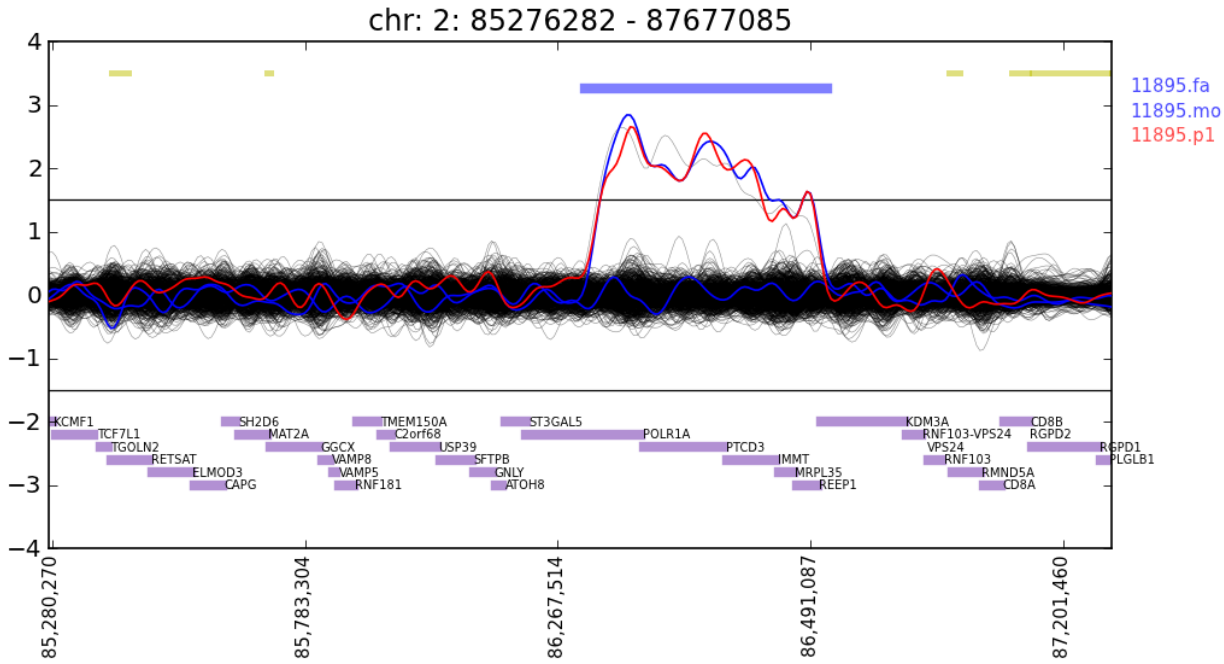


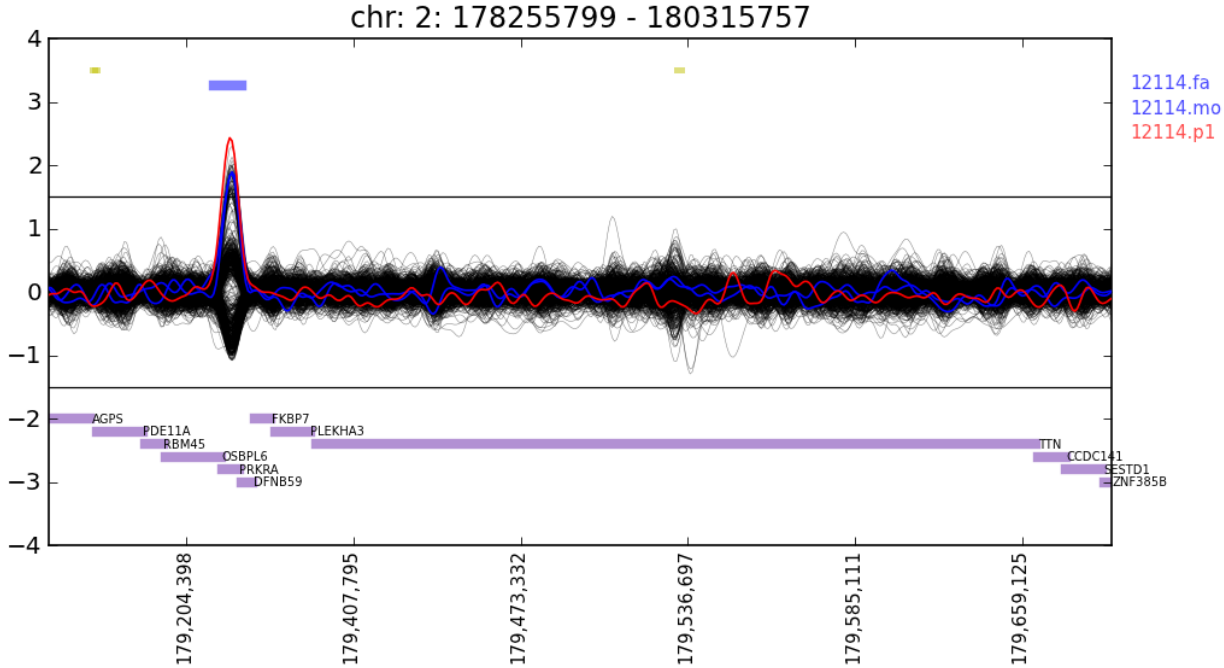
Figure(s) S10

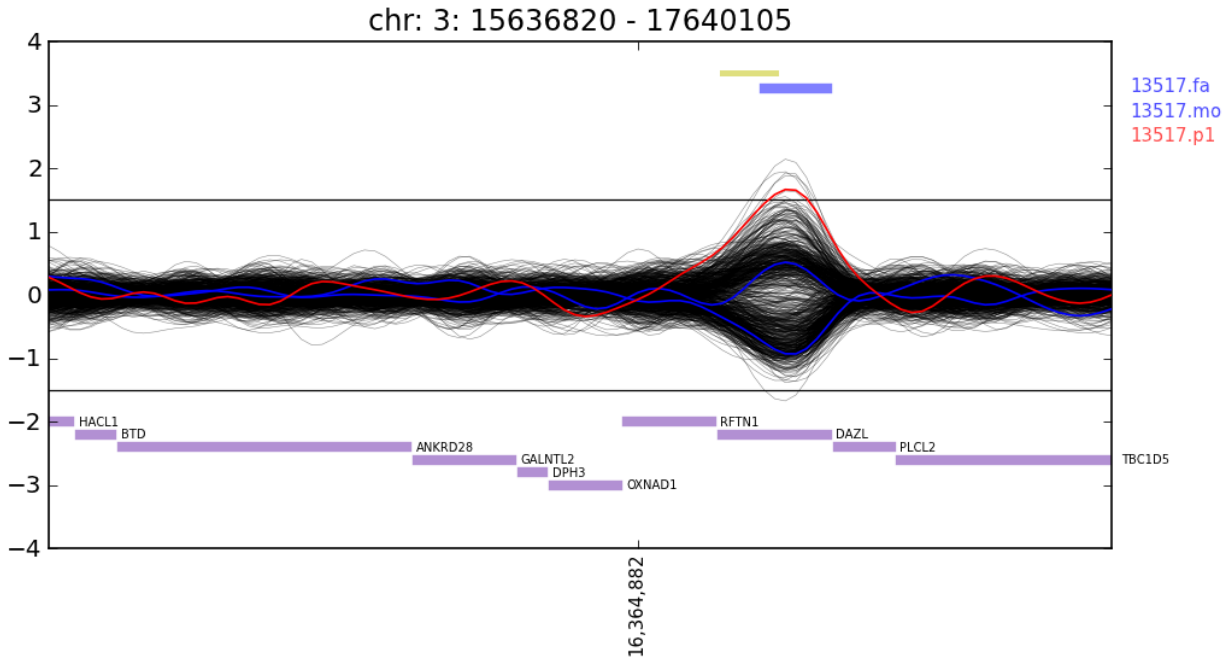


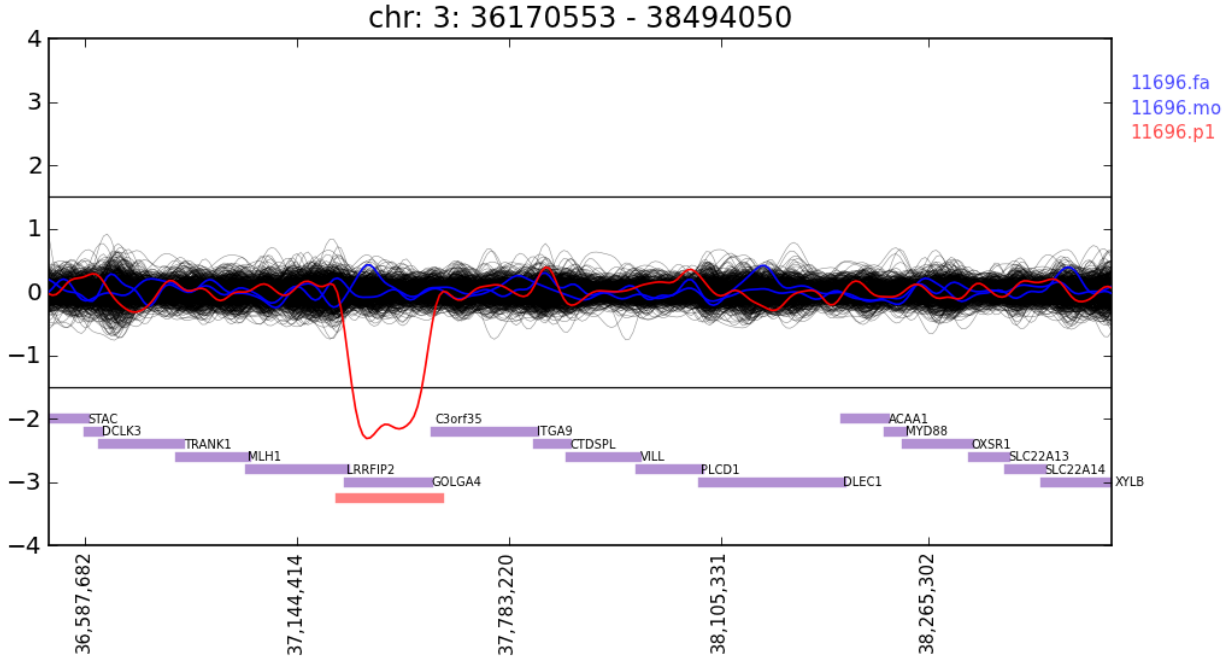
A) Coordinates of genomic interval shown
B) Samples highlighted (blue - parents, red - proband)
C) Call in proband
D) Segmental duplications
E) Threshold used to call deletion/duplications
F) Genes

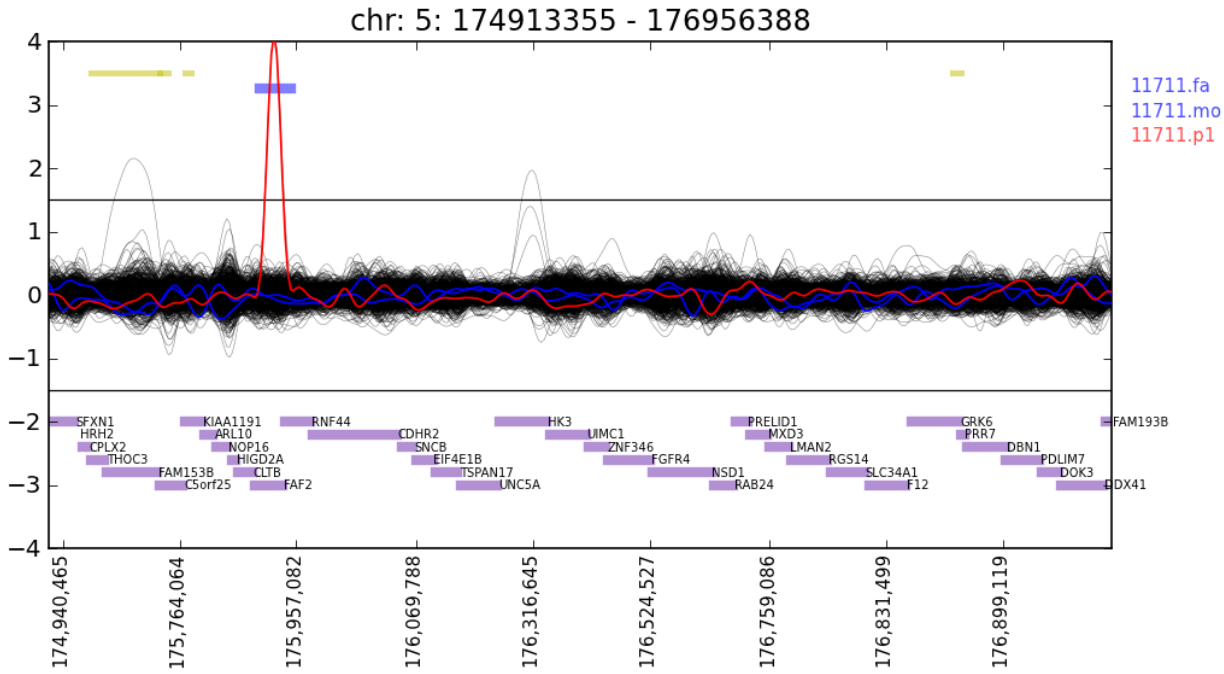


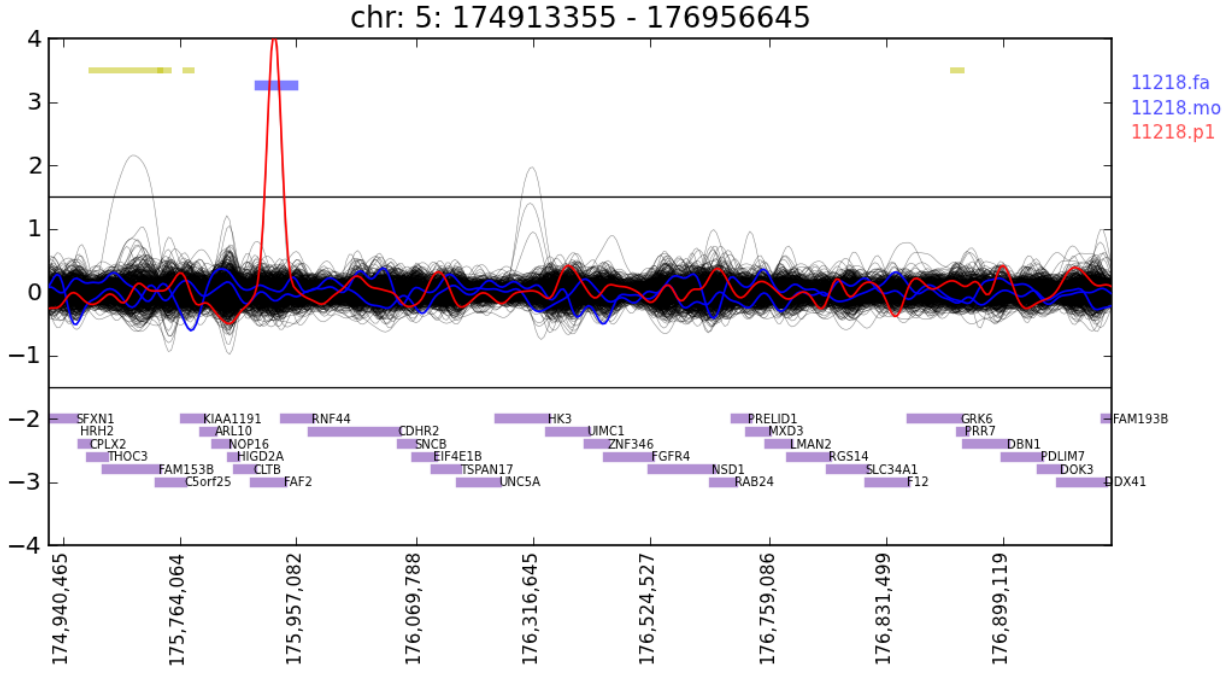

chr: 2: 85276282 - 87677085
11895.fa
11895.mo
11895.p1
KCMF1
TCF7L1
TGOLN2
RETSAT
ELMOD3
CAPG
SH2D6
MAT2A
GGCX
VAMP8
VAMP5
RNF181
TMEM150A
C2orf68
USP39
SFTPB
GNLY
ATOH8
ST3GAL5
POLR1A
PTCD3
IMMT
MRPL35
REEP1
KDM3A
RNF103-VPS24
VPS24
RNF103
RMND5A
CD8A
CD8B
RGPD2
RGPD1
PLGLB1
85,280,270
85,783,304
86,267,514
86,491,087
87,201,460

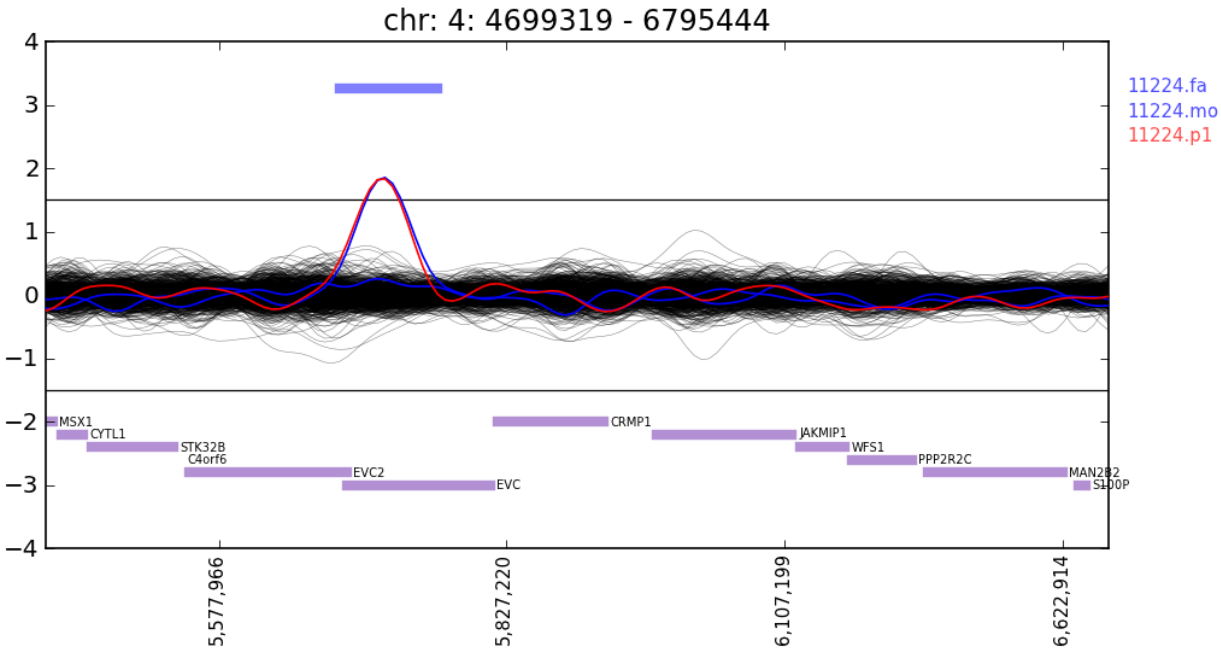

chr: 2: 178255799 - 180315757
12114.fa
12114.mo
12114.p1
AGPS
PDE11A
RBM45
OSBPL6
PRKRA
DFNB59
FKBP7
PLEKHA3
TTN
CCDC141
SESTD1
ZNF385B
179,204,398
179,407,795
179,473,332
179,536,697
179,585,111
179,659,125

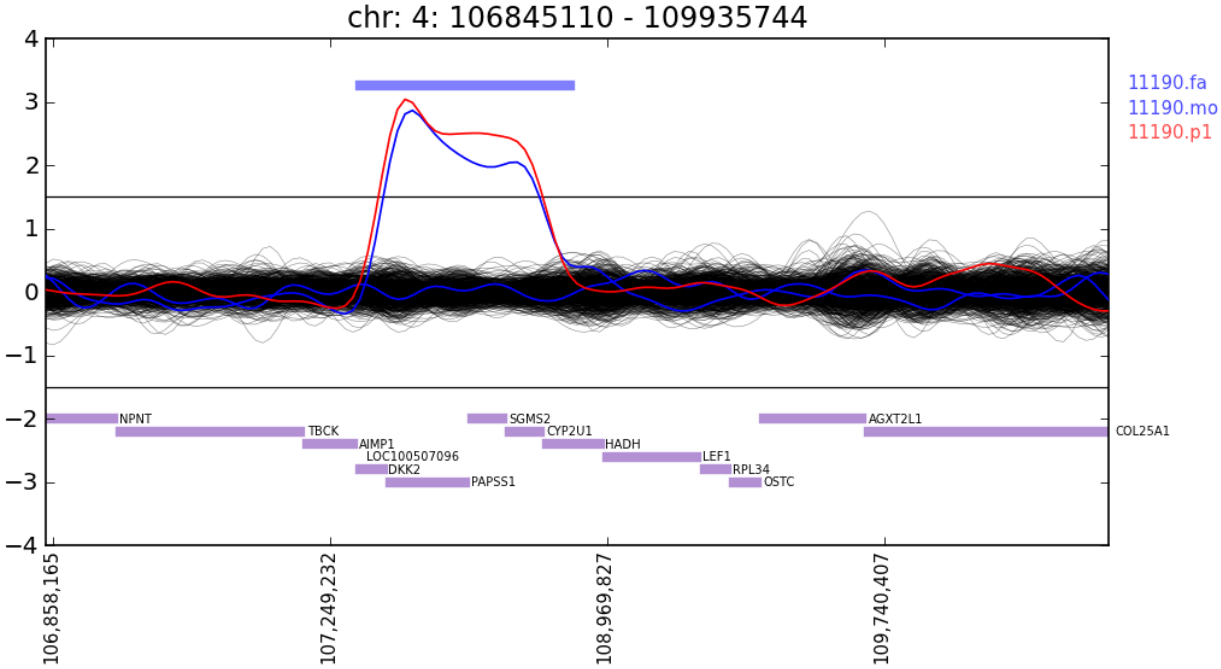

chr: 3: 15636820 - 17640105
13517.fa
13517.mo
13517.p1
HACL1
BTD
ANKRD28
GALNTL2
DPH3
OXNAD1
RFTN1
DAZL
PLCL2
TBC1D5
16,364,882

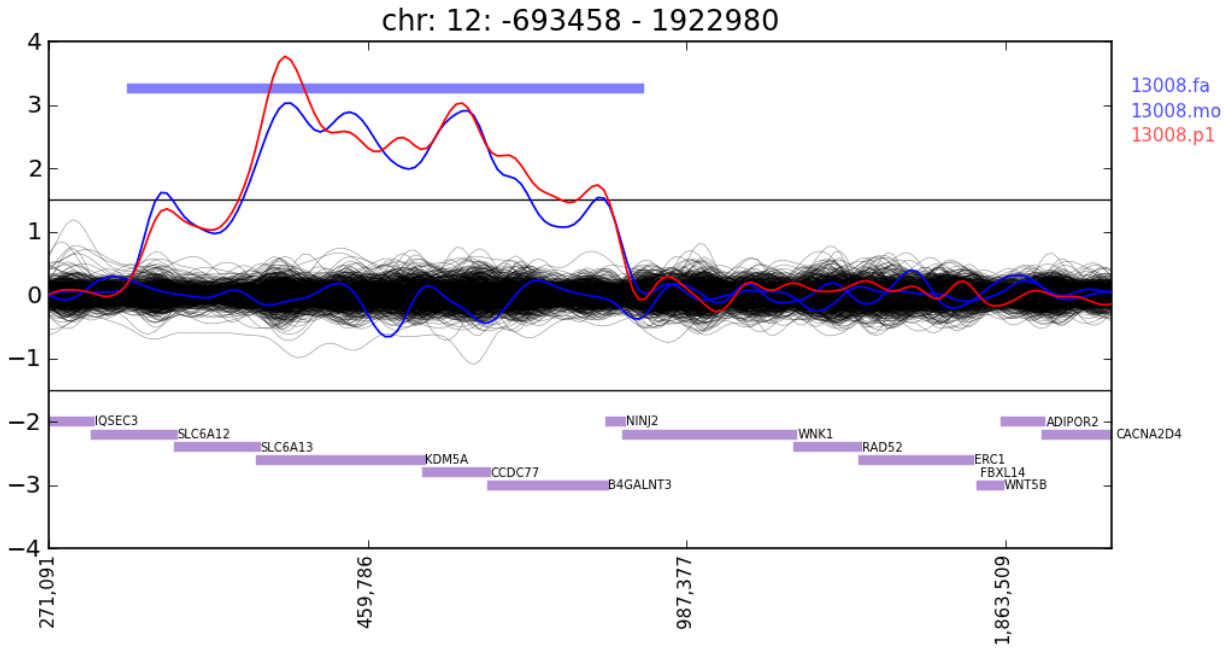

chr: 3: 36170553 - 38494050
11696.fa
11696.mo
11696.p1
STAC
DCLK3
TRANK1
MLH1
LRRFIP2
GOLGA4
C3orf35
ITGA9
CTDSPL
MLL
PLCD1
DLEC1
ACAA1
MYD88
OXSR1
SLC22A13
SLC22A14
XYLB
36,587,682
37,144,414
37,783,220
38,105,331
38,265,302

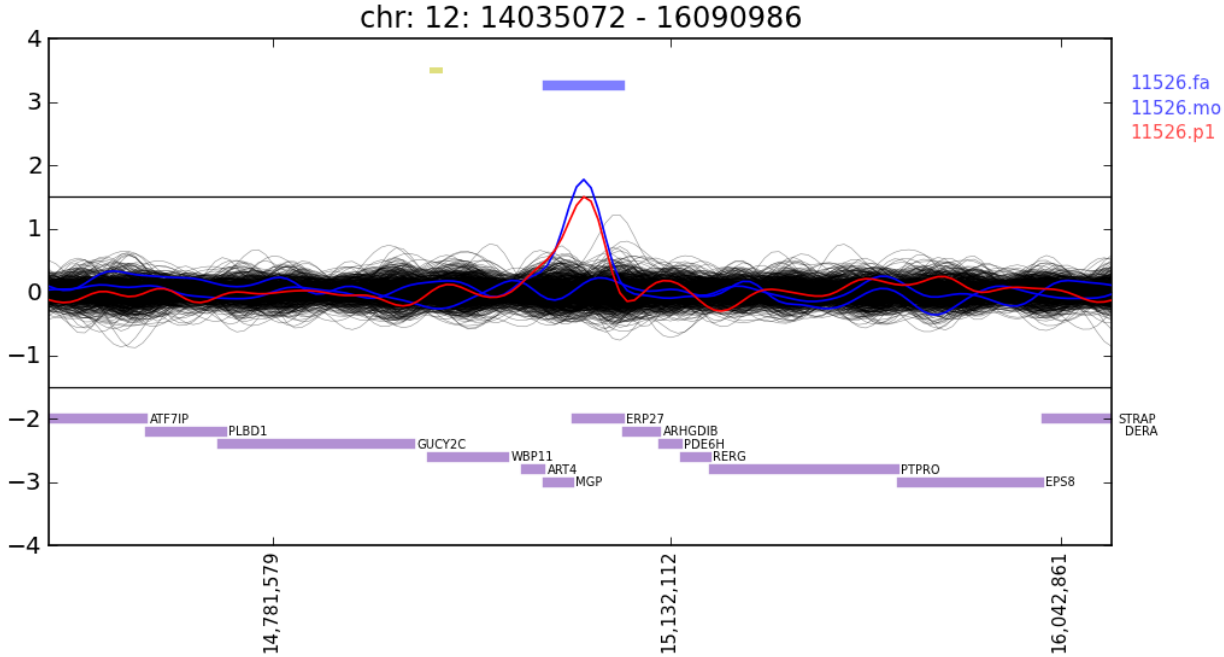

chr: 5: 174913355 - 176956388
11711.fa
11711.mo
11711.p1
4
3
2
1
0
−1
−2
−3
−4
SFXN1
HRH2
CPLX2
THOC3
FAM153B
C5orf25
KIAA1191
ARL10
NOP16
HIGD2A
CLTB
FAF2
RNF44
CDHR2
SNCB
EIF4E1B
TSPAN17
UNC5A
HK3
UIMC1
ZNF346
FGFR4
NSD1
RAB24
PRELID1
MXD3
LMAN2
RGS14
SLC34A1
F12
GRK6
PRR7
DBN1
PDLIM7
DOK3
DDX41
FAM193B
174,940,465
175,764,064
175,957,082
176,069,788
176,316,645
176,524,527
176,759,086
176,831,499
176,899,119

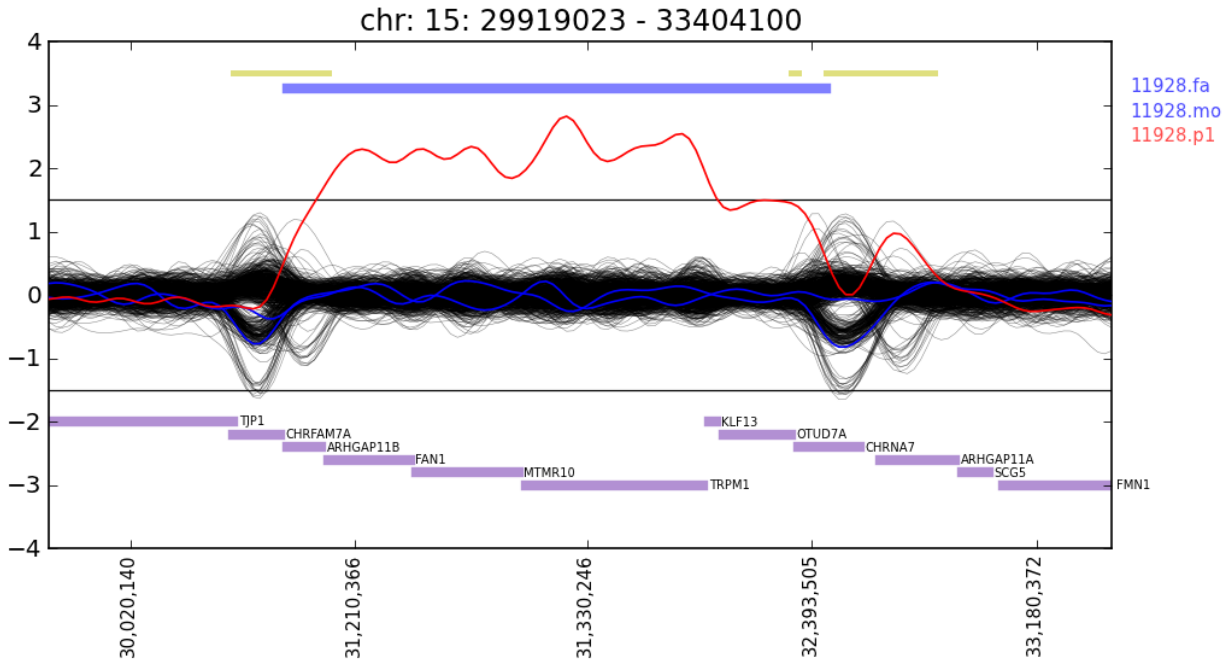

chr: 5: 174913355 - 176956645
11218.fa
11218.mo
11218.p1
4
3
2
1
0
−1
−2
−3
−4
SFXN1
HRH2
CPLX2
THOC3
FAM153B
C5orf25
KIAA1191
ARL10
NOP16
HIGD2A
CLTB
FAF2
RNF44
CDHR2
SNCB
EIF4E1B
TSPAN17
UNC5A
HK3
UIMC1
ZNF346
FGFR4
NSD1
RAB24
PRELID1
MXD3
LMAN2
RGS14
SLC34A1
F12
GRK6
PRR7
DBN1
PDLIM7
DOK3
DDX41
FAM193B
174,940,465
175,764,064
175,957,082
176,069,788
176,316,645
176,524,527
176,759,086
176,831,499
176,899,119

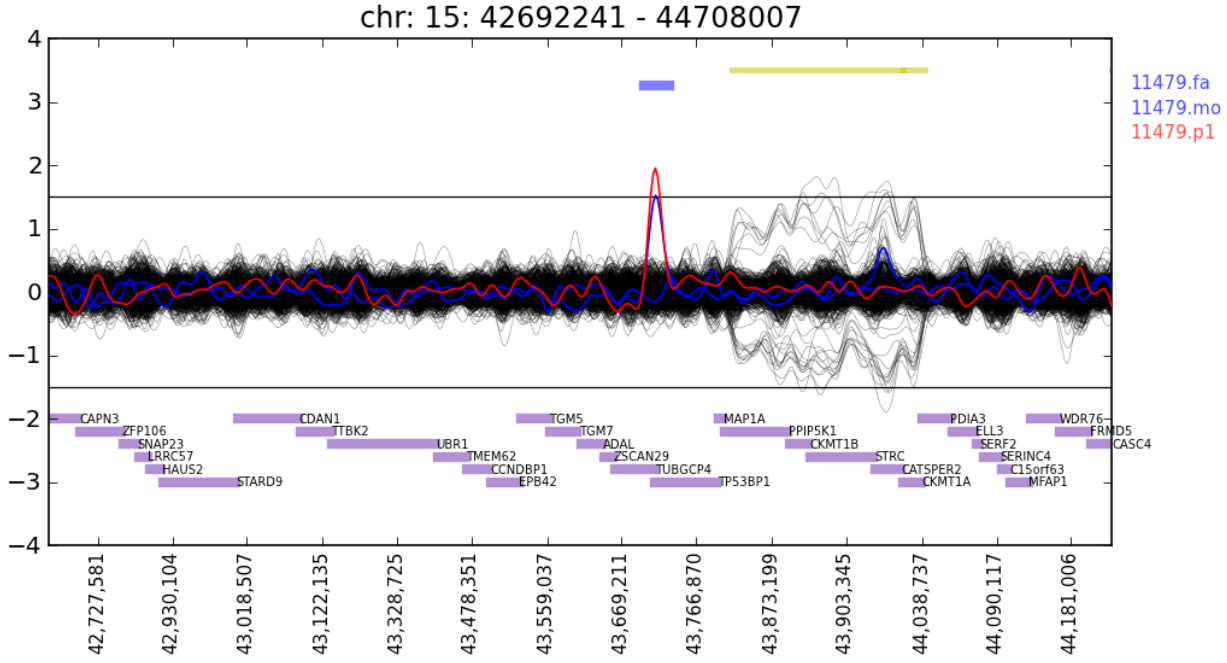

chr: 4: 4699319 - 6795444
11224.fa
11224.mo
11224.p1
MSX1
CYTL1
STK32B
C4orf6
EVC2
EVC
CRMP1
JAKMIP1
WFS1
PPP2R2C
MAN2B2
S100P
5,577,966
5,827,220
6,107,199
6,622,914

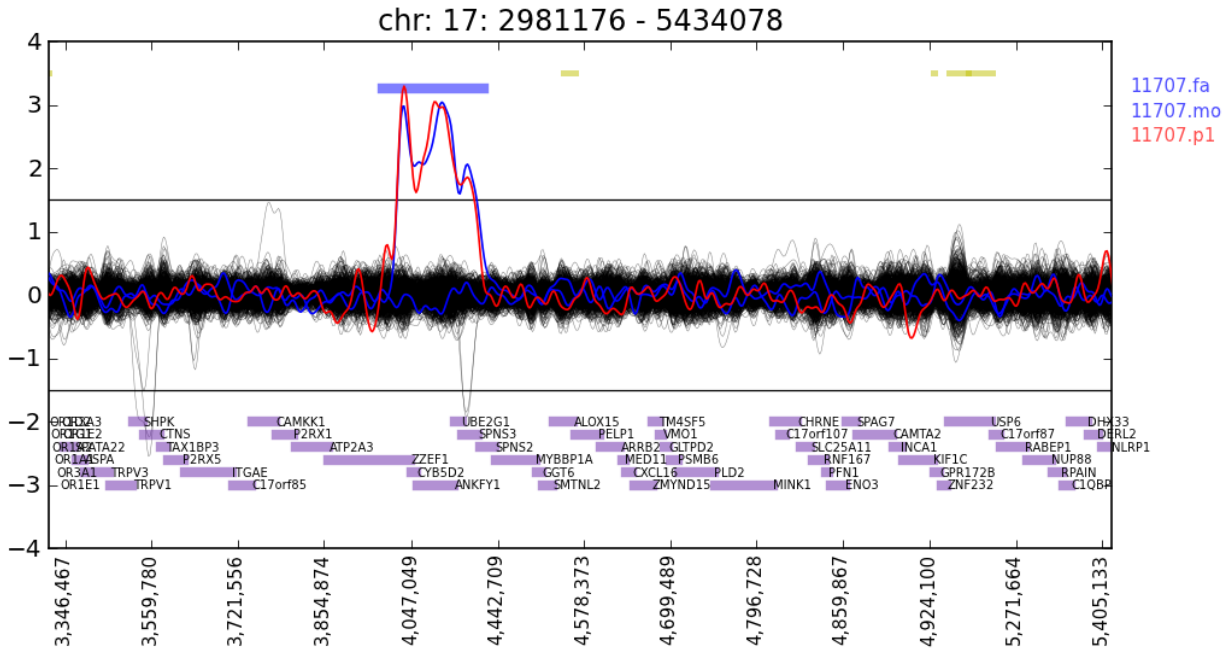

chr: 4: 106845110 - 109935744
11190.fa
11190.mo
11190.p1
NPNT
TBCK
AIMP1
LOC100507096
DKK2
PAPSS1
SGMS2
CYP2U1
HADH
LEF1
RPL34
OSTC
AGXT2L1
COL25A1
106,858,165
107,249,232
108,969,827
109,740,407

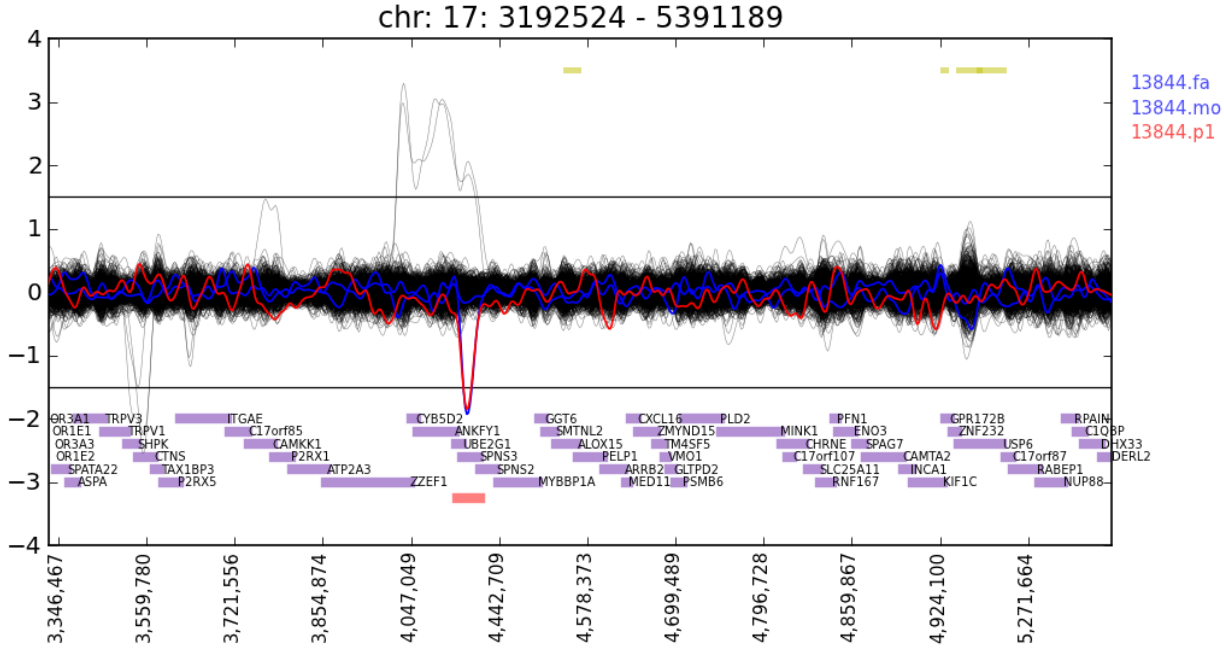

chr: 12: -693458 - 1922980
13008.fa
13008.mo
13008.p1
IQSEC3
SLC6A12
SLC6A13
KDM5A
CCDC77
B4GALNT3
NINJ2
WNK1
RAD52
ERC1
FBXL14
WNT5B
ADIPOR2
CACNA2D4
271,091
459,786
987,377
1,863,509

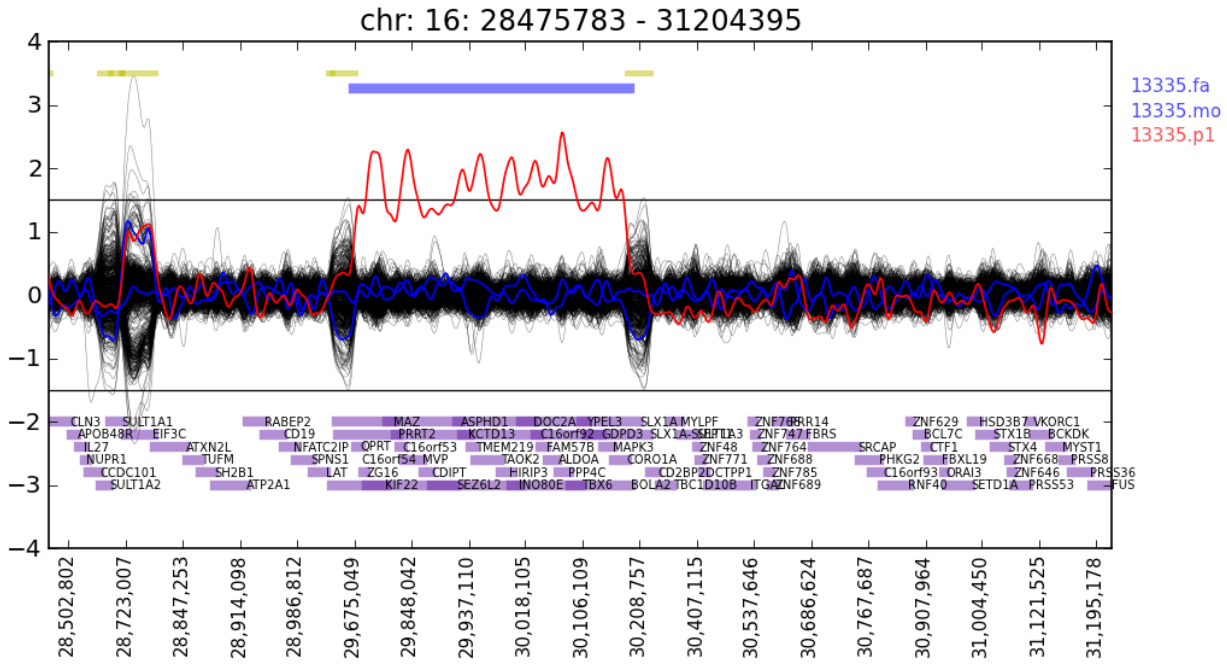

chr: 12: 14035072 - 16090986
11526.fa
11526.mo
11526.p1
ATF7IP
PLBD1
GUCY2C
WBP11
ART4
MGP
ERP27
ARHGDIB
PDE6H
RERG
PTPRO
EPS8
STRAP
DERA
14,781,579
15,132,112
16,042,861

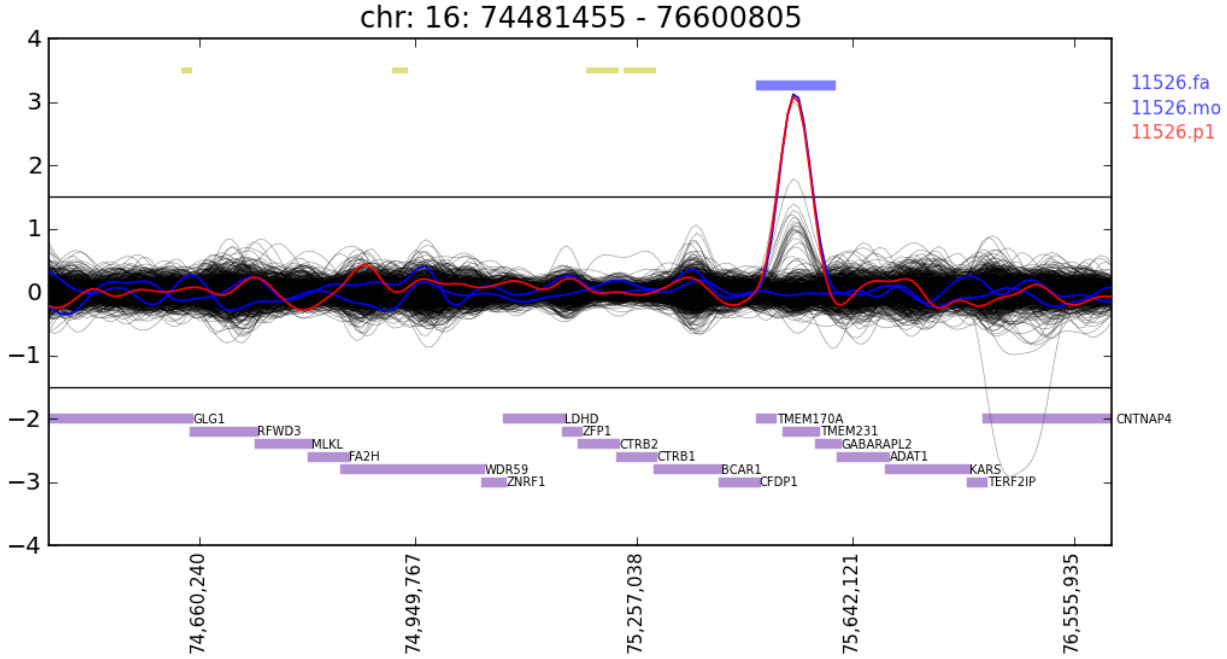

chr: 15: 29919023 - 33404100
11928.fa
11928.mo
11928.p1
TJP1
CHRFAM7A
ARHGAP11B
FAN1
MTMR10
TRPM1
KLF13
OTUD7A
CHRNA7
ARHGAP11A
SCG5
FMN1
30,020,140
31,210,366
31,330,246
32,393,505
33,180,372

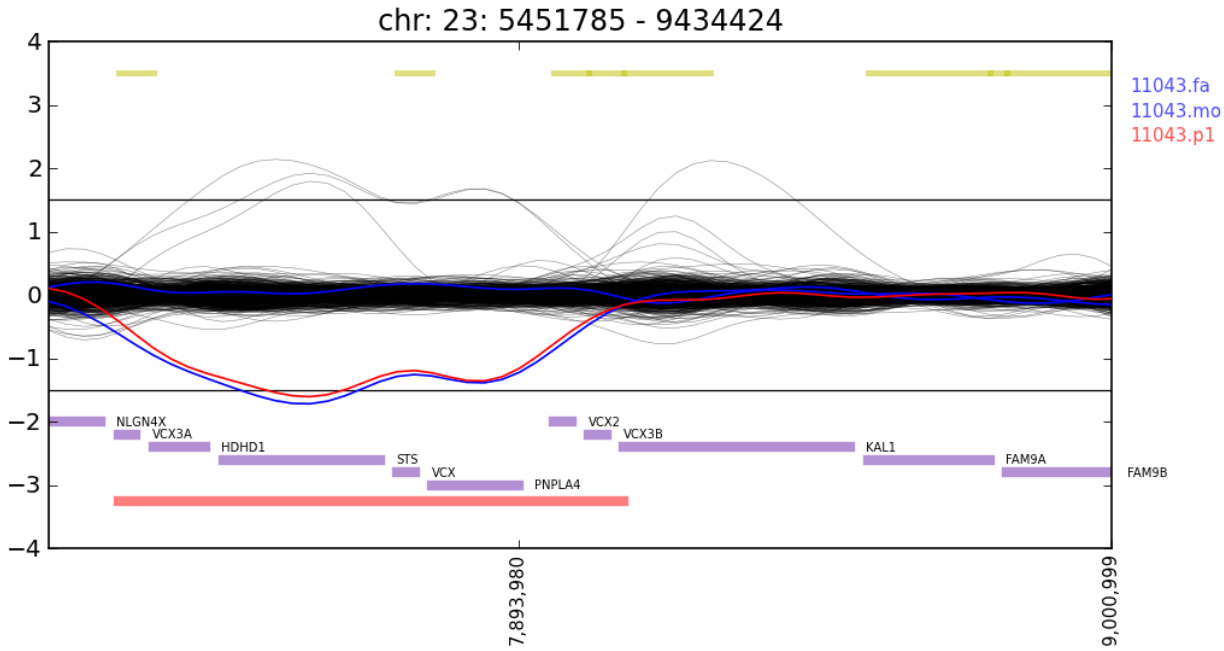

chr: 15: 42692241 - 44708007
11479.fa
11479.mo
11479.p1
CAPN3
ZFP106
SNAP23
LRRC57
HAUS2
STARD9
CDAN1
TTBK2
UBR1
TMEM62
CCNDBP1
EPB42
TGM5
TGM7
ADAL
ZSCAN29
TUBGCP4
TP53BP1
MAP1A
PPIP5K1
CKMT1B
STRC
CATSPER2
CKMT1A
PDIA3
ELL3
SERF2
SERINC4
C15orf63
MFAP1
WDR76
FRMD5
CASC4
42,727,581
42,930,104
43,018,507
43,122,135
43,328,725
43,478,351
43,559,037
43,669,211
43,766,870
43,873,199
43,903,345
44,038,737
44,090,117
44,181,006


chr: 17: 2981176 - 5434078
11707.fa
11707.mo
11707.p1
4
3
2
1
0
−1
−2
−3
−4
3,346,467
3,559,780
3,721,556
3,854,874
4,047,049
4,442,709
4,578,373
4,699,489
4,796,728
4,859,867
4,924,100
5,271,664
5,405,133


chr: 17: 3192524 - 5391189
13844.fa
13844.mo
13844.p1
4
3
2
1
0
−1
−2
−3
−4
3,346,467
3,559,780
3,721,556
3,854,874
4,047,049
4,442,709
4,578,373
4,699,489
4,796,728
4,859,867
4,924,100
5,271,664


chr: 16: 28475783 - 31204395
13335.fa
13335.mo
13335.p1
4
3
2
1
0
−1
−2
−3
−4
CLN3
SULT1A1
RABEP2
MAZ
ASPHD1
DOC2A
YPEL3
SLX1A
MYLPF
ZNF76
RRP14
ZNF629
HSD3B7
VKORC1
APOB48R
EIF3C
CD19
PRRT2
KCTD13
C16orf92
GDPD3
SLX1A-SULT1A3
ZNF747
FBRS
BCL7C
STX1B
BCKDK
IL27
ATXN2L
NFATC2IP
QPRT
C16orf53
TMEM219
FAM57B
MAPK3
ZNF48
ZNF764
SRCAP
CTF1
STX4
MYST1
NUPR1
TUFM
SPNS1
C16orf54
MVP
TAOK2
ALDOA
CORO1A
ZNF771
ZNF688
PHKG2
FBXL19
ZNF668
PRSS8
CCDC101
SH2B1
LAT
ZG16
CDIPT
HIRIP3
PPP4C
CD2BP2
DCTPP1
ZNF785
C16orf93
ORAI3
ZNF646
PRSS36
SULT1A2
ATP2A1
KIF22
SEZ6L2
INO80E
TBX6
BOLA2
TBC1D10B
ITGA2
ZNF689
RNF40
SETD1A
PRSS53
FUS
28,502,802
28,723,007
28,847,253
28,914,098
28,986,812
29,675,049
29,848,042
29,937,110
30,018,105
30,106,109
30,208,757
30,407,115
30,537,646
30,686,624
30,767,687
30,907,964
31,004,450
31,121,525
31,195,178


chr: 16: 74481455 - 76600805
11526.fa
11526.mo
11526.p1
4
3
2
1
0
−1
−2
−3
−4
GLG1
RFWD3
MLKL
FA2H
WDR59
ZNRF1
LDHD
ZFP1
CTRB2
CTRB1
BCAR1
CFDP1
TMEM170A
TMEM231
GABARAPL2
ADAT1
KARS
TERF2IP
CNTNAP4
74,660,240
74,949,767
75,257,038
75,642,121
76,555,935


chr: 23: 5451785 - 9434424
11043.fa
11043.mo
11043.p1
NLGN4X
VCX3A
HDHD1
STS
VCX
PNPLA4
VCX2
VCX3B
KAL1
FAM9A
FAM9B
7,893,980
9,000,999

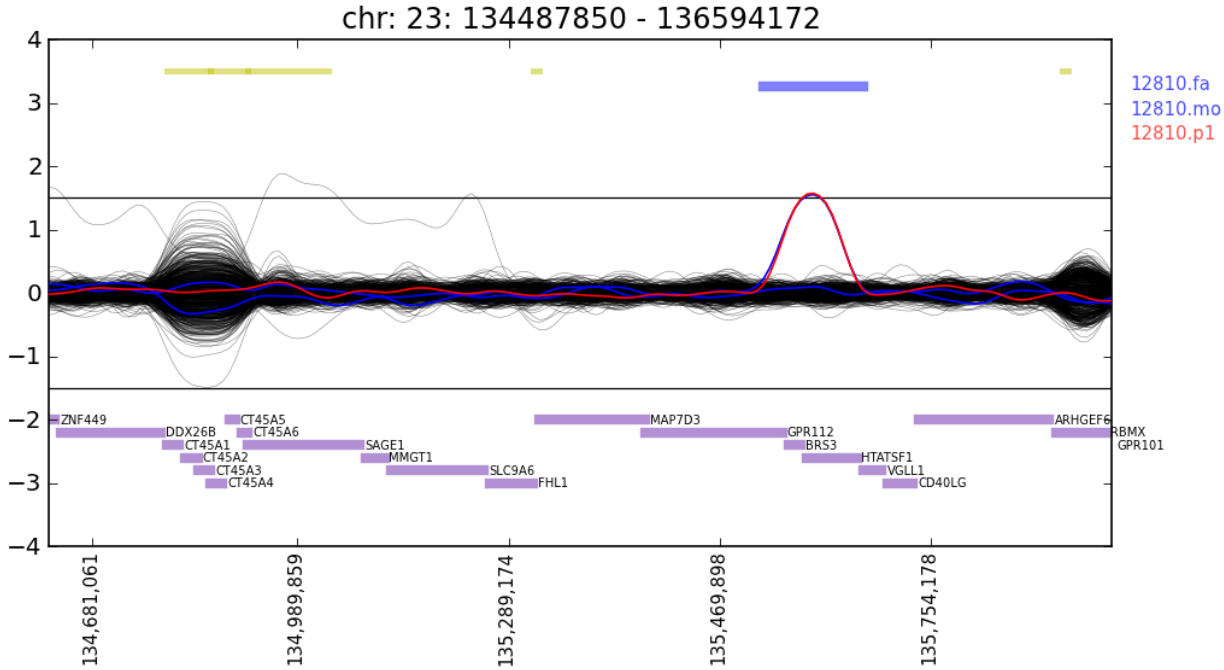
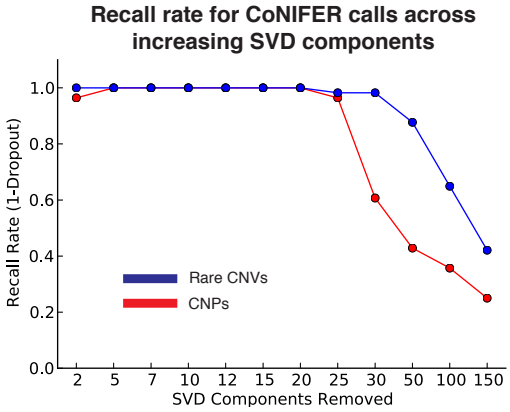
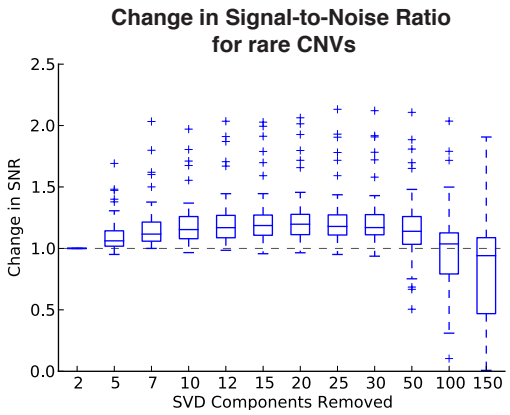

chr: 23: 134487850 - 136594172
12810.fa
12810.mo
12810.p1
ZNF449
DDX26B
CT45A1
CT45A2
CT45A3
CT45A4
CT45A5
CT45A6
SAGE1
MMGT1
SLC9A6
FHL1
MAP7D3
GPR112
BRS3
HTATSF1
VGLL1
CD40LG
ARHGEF6
RBMX
GPR101
134,681,061
134,989,859
135,289,174
135,469,898
135,754,178

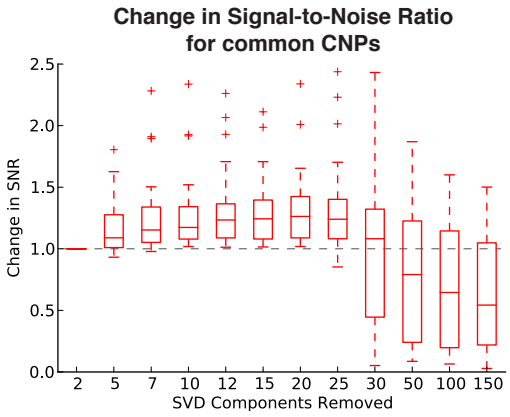
Figure S11: Simulation showing effect of SVD removal on rare and common events



A. Recall rate: We selected 85 SNP-validated CNV calls found within the 122 ASD probands using our algorithm (comprising 57 rare CNVs seen in <1% of cases and 28 CNPs). We iteratively removed SVD components and assessed the proportion of calls that survived the stringent ±1.5 SVD-ZRPKM cutoff. At 30 components removed, over 56 of 57 (98.2%) rare CNVs survive, indicating that biological signal for CNVs in exome read-depth survives the removal widespread systematic noise.

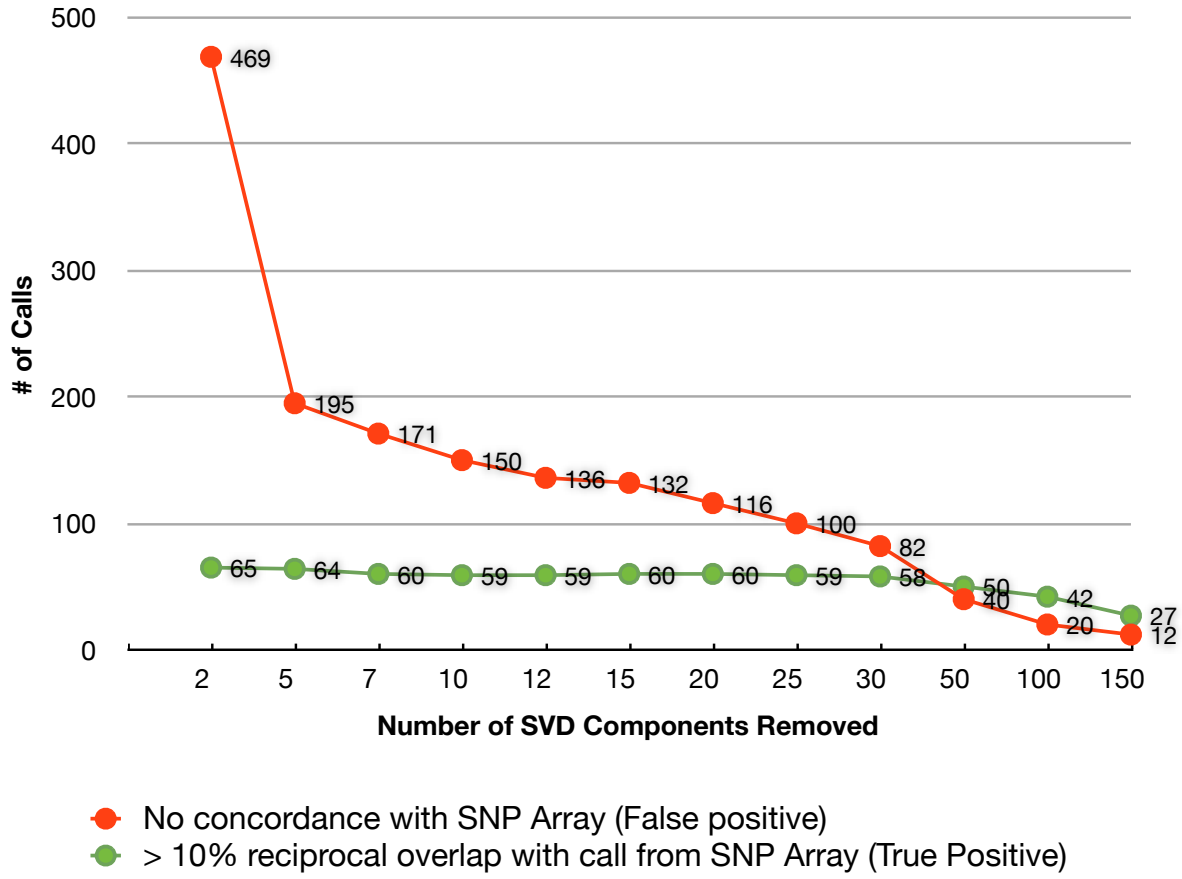


B. Signal to Noise Ratio (SNR) for Rare CNVS: Using the same simulation as in (A) above, we assessed the SNR of each call (defined as the mean of the SVD-ZRPKM values within the call boundaries divided by the standard deviation of the values for call's chromosome). The percent change versus the SNR at 2 SVD components is shown across removal of SVD components. We note that SVD increases the SNR for nearly all of the tested rare CNVs, and the SNR remains robust even when removing significantly more SVD components than necessary.



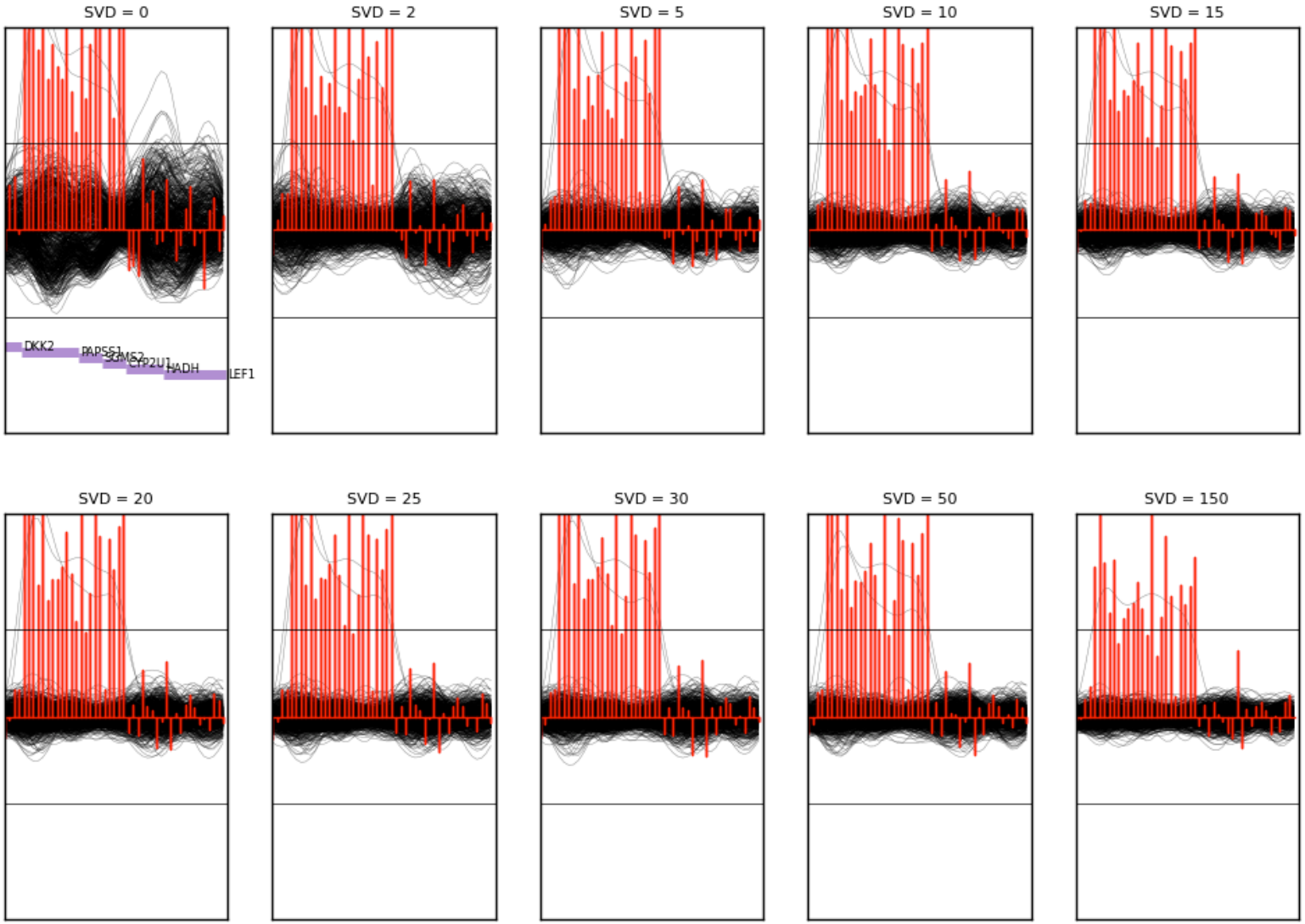
C.SNR for common CNPs: Similar to (B) above, we assessed the SNR for common CNPs. For these events, the increased population variance makes them more susceptible to removal via the SVD algorithm. We suggest genotyping these known CNPs by removing fewer SVD components in order to preserve signal.

Figure S12: Removing SVD components does not impact discovery of rare CNVs



We iteratively removed SVD components from the ASD data set and generated calls at each level. We intersected the resulting calls at each level (using a ±1.5 SVD-ZRPKM threshold) with SNP calls from Sanders *et al*. (2011). A greatly increased rate of false positive calls is seen when fewer SVD components are removed, reflecting prevalent systematic noise found within exome datasets. Removing additional components greatly decreases the number of false positive calls. In contrast, the number of concordant calls remains stable, even when removing SVD components much higher than recommended by the inflection point of the scree plot. We note that the “false positives” seen at 15 components removed are not true false positive, see the text and additional analysis for details. Taken together, these results indicate that the CoNIFER algorithm eliminates systematic noise but preserves a majority of the biological signal.

Figures S13a: Example rare CNV across increasing removal of SVD components



Figures S13b: Example CNP across increasing removal of SVD components

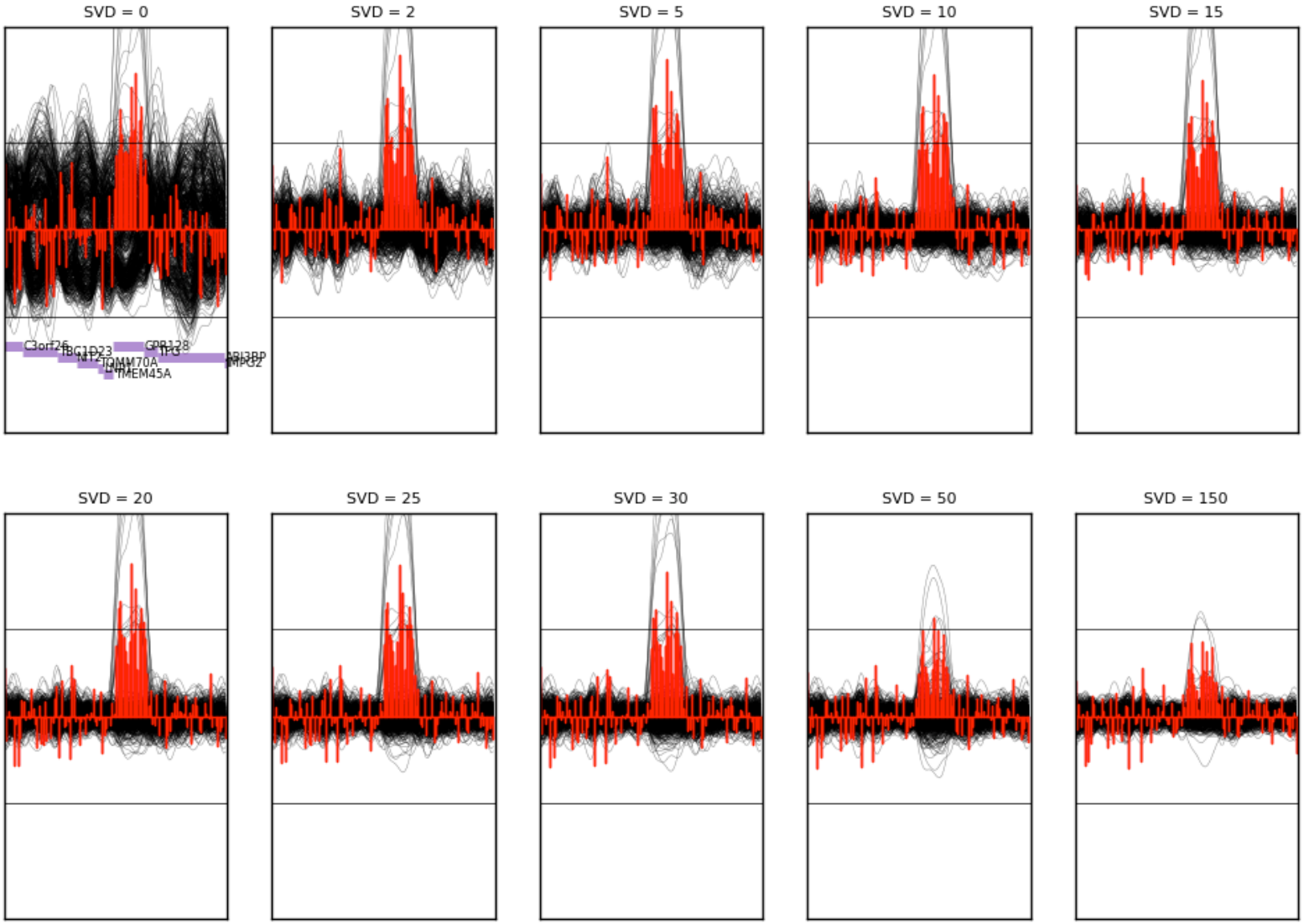
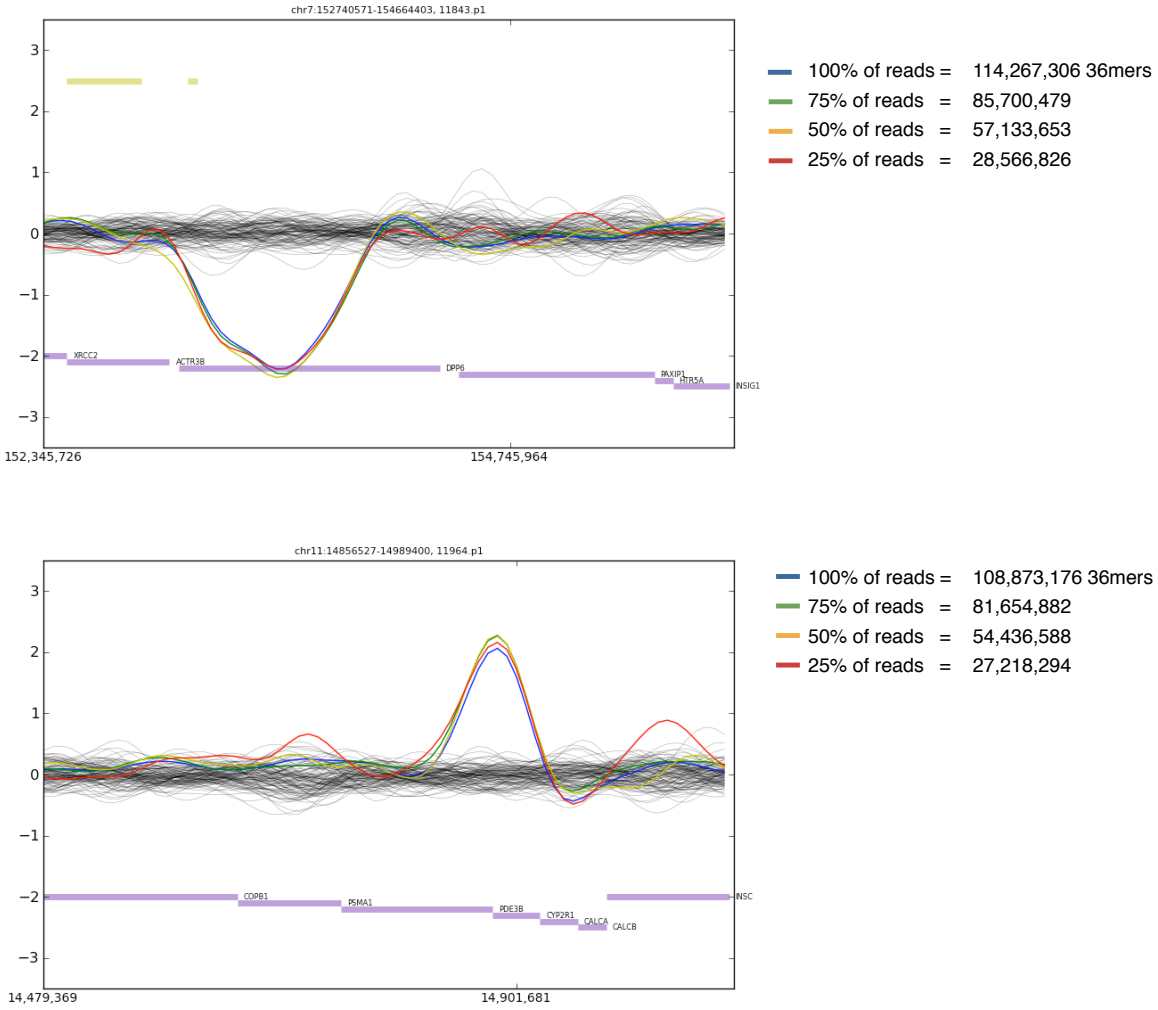
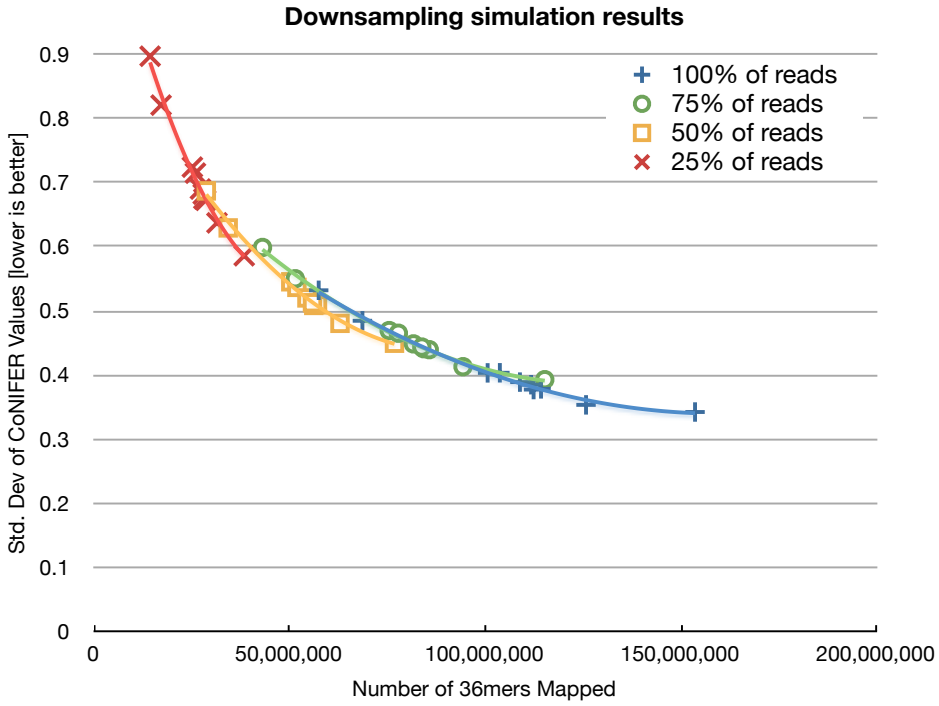


Figure S14: Reduced exome coverage does not attenuate signal for rare CNVs



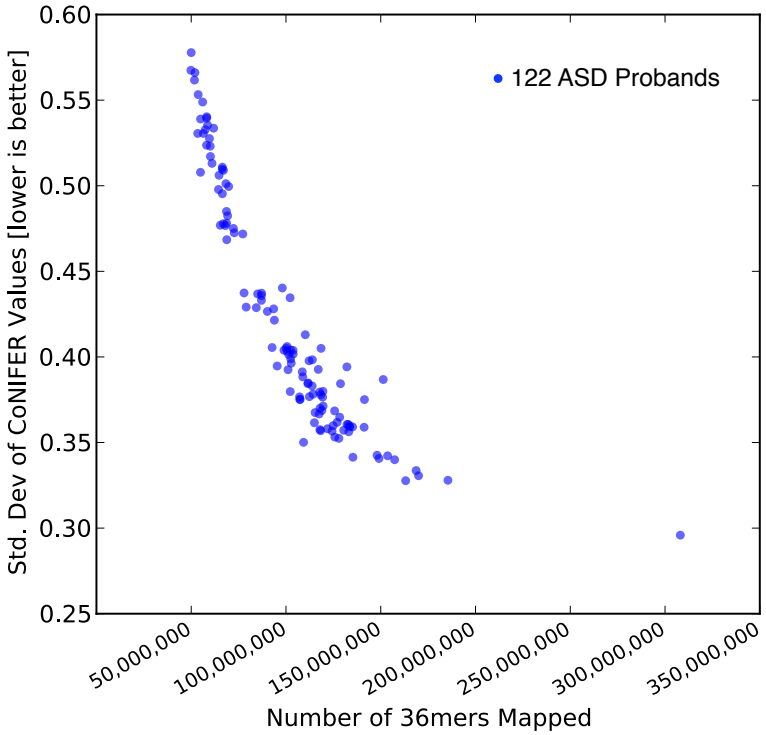
To analyze how lower coverage exomes may affect the CoNIFER algorithm, we randomly down-sampled 10 randomly selected probands from the 122 autism exomes at 75%, 50% and 25% of their original reads. We included each of these exomes in a separate CoNIFER analysis and assessed if there was signal loss for known CNVs. There was virtually no loss of signal across the nine CNVs found in these samples. Two example CNVs from two samples in the simulation are shown above at 100%, 75%, 50% and 25% down-sampling.

Figure S15: Lower exome coverage results in increased random noise



We calculated the noise found within each of the 10 exomes and their down-sampled counterparts by finding the standard deviation across all exons per sample. When mapping fewer than 50 million mapped reads per exome, the noise increases sharply, increasing the false positive rate and decreasing sensitivity to small events.

Figure S16: Random noise in SVD-ZRPKM values increases with reduced coverage



This figure shows the inversely correlated relationship between the total number of 36mers mapped and the standard deviation of the SVD-ZRPKM values for each exome. CoNIFER processing was possible on all displayed exomes, though lower standard deviation for the SVD-ZRPKM values indicate less random noise within the exome which can lead to improved sensitivity and specificity. We suggest a minimum of 50 million mapped reads for optimal performance.

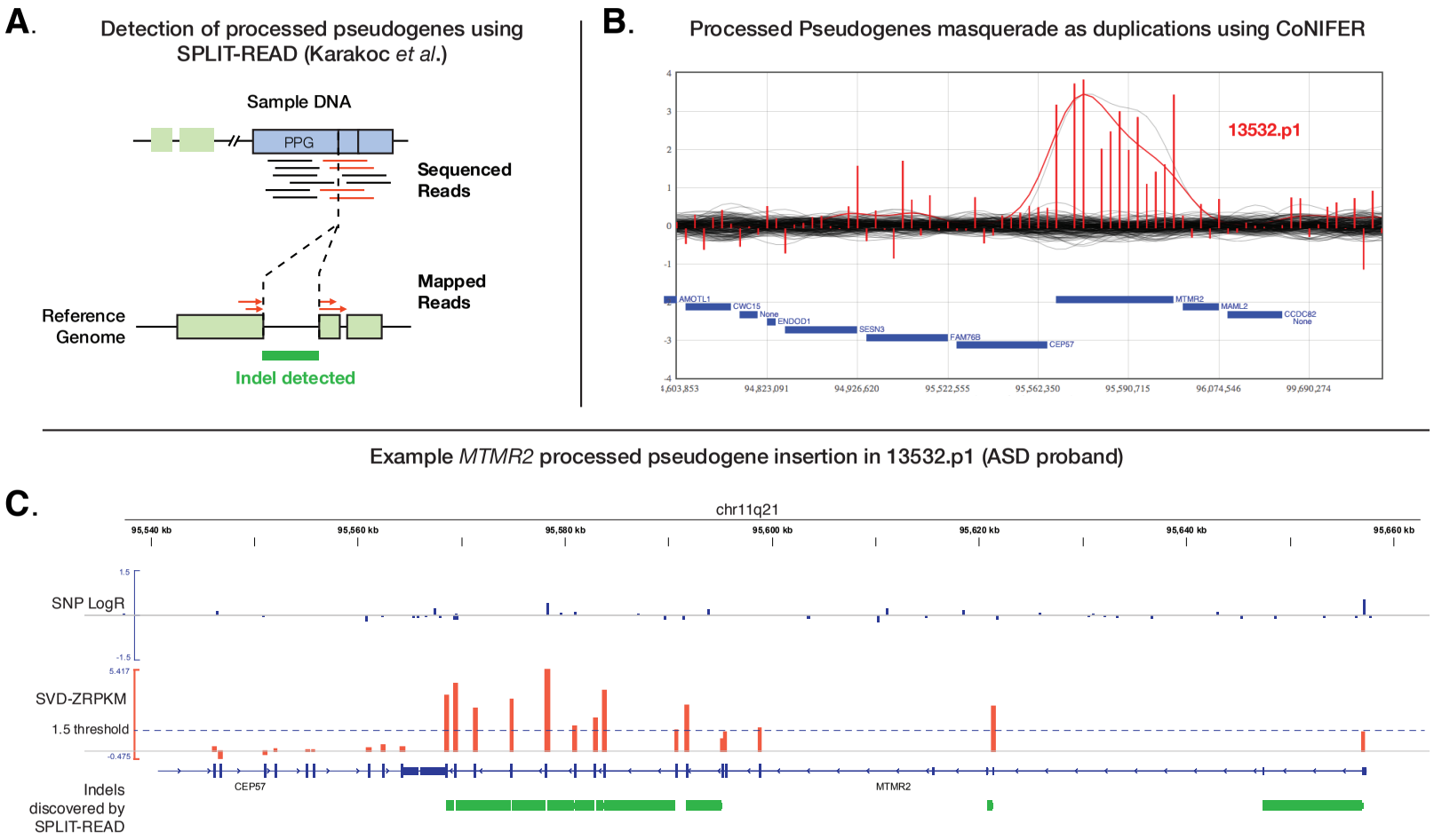
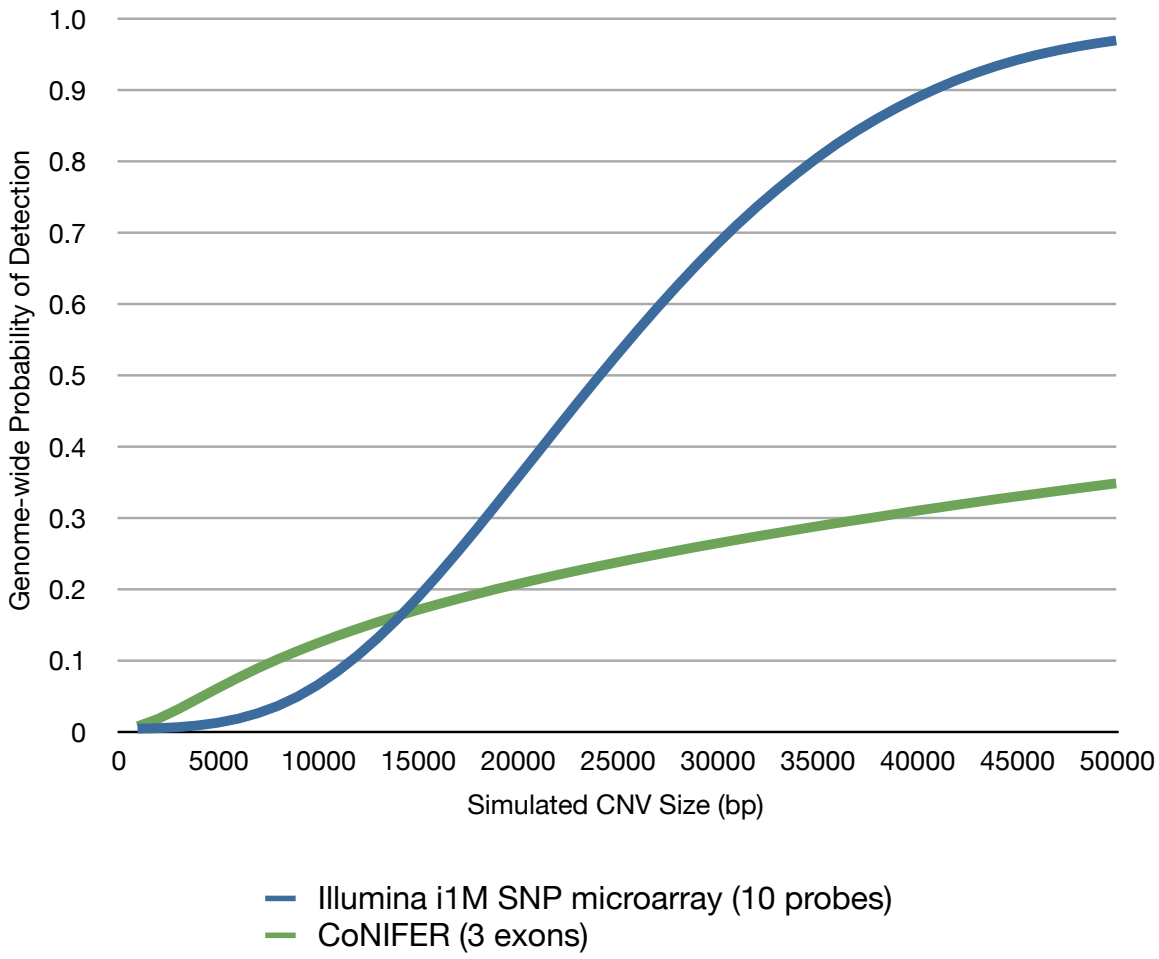


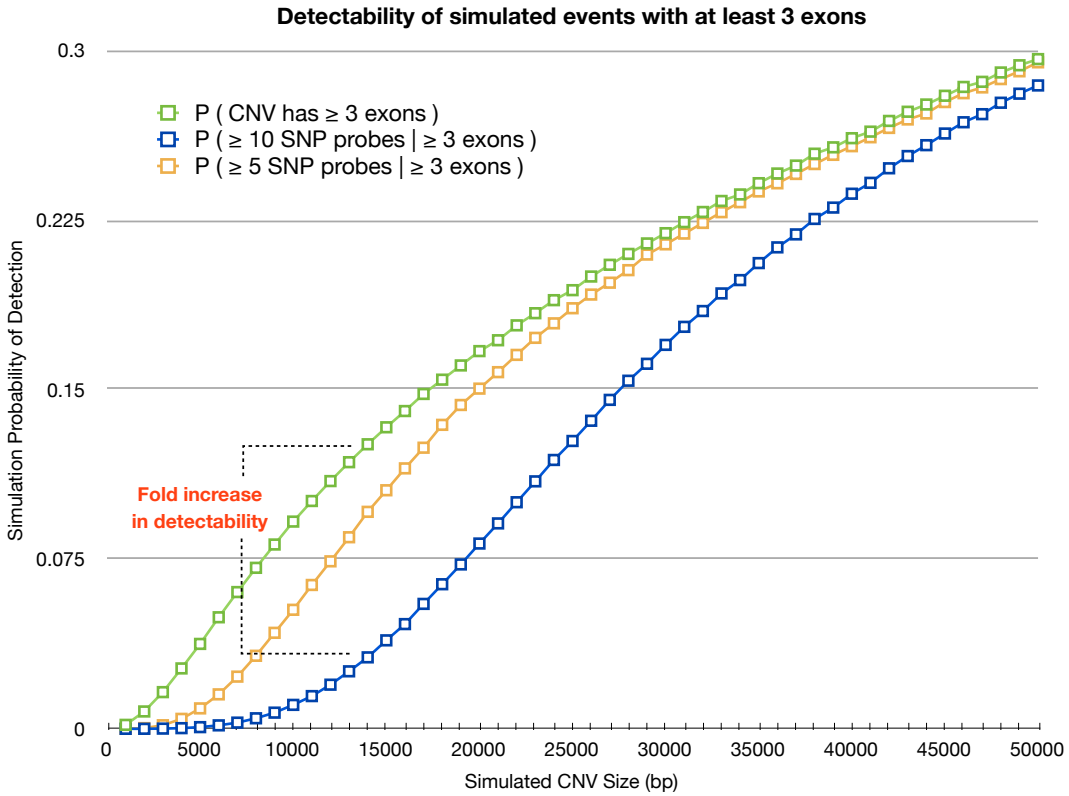
Figure S17: Schematic and example of the *MTMR2* processed pseudogene discovered from exome data:
A. We used SPLIT-READ (Karakoc *et al.*, 2011) to find signatures of processed pseudogenes by looking for deletions that exactly span introns of genes. B. In CoNIFER, processed pseudogenes contribute additional cryptic copies of exons which are not seen in the reference, and masquerade as single-gene duplications. C. Processed pseudogenes are not visible using SNP (top track) or arrayCGH data (not shown), as such assays contain primarily intronic probes, for which the underlying sequences is not duplicated. Middle track (red): SVD-ZRPKM values of 13532.p1 from CoNIFER analysis. Bottom Track (green): deletions found using SPLIT-READ on same sample.

Figures S18a: Genome-wide detectability of CoNIFER and i1M Duo SNP array



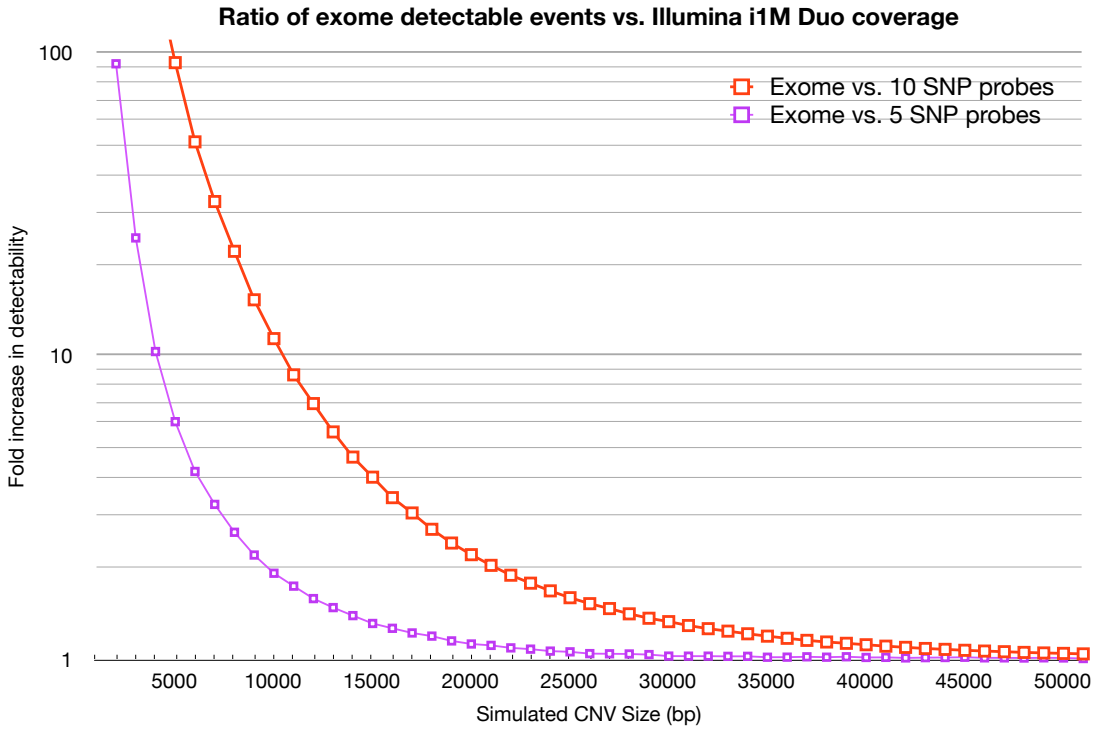
e calculated the theoretical fraction of sites within the genome detectable by either our algorithm (with a minimum of three probes) or a standard Illumina i1M Duo SNP microarray (with a minimum of 10 probes required for detection). Detectability was calculated in binned size ranges of 1kbp and we excluded genomic gaps, centromeres and telomeres from the analysis. The resulting fractions show that the Illumina i1M Duo SNP microarray can detect a large fraction of genomic events larger than 25kbp, as expected for a high-density SNP microarray targeting the entire genome. In contrast, our algorithm has a lower *de facto* fraction of detectable events genome-wide, due the lower overall probe density and targeted nature of the probes. However, for CNVs smaller than ~14kbp, our algorithm has a significant theoretical detection advantage over the Illumina SNP platform.

Figure S18b: Simulated detection of small genic CNVs for exome-based and Illumina 1M Duo SNP microarray



We sought to estimate the ability of our algorithm to find small *exonic* (or genic) CNVs in comparison to the detection power of the Illumina i1M SNP Duo (1.1 million probes) microarray for similar events. We randomly simulated the placement of CNVs of a given size within the genome and compared how many of these simulated CNVs intersected at least three exome probes, P(≥ 3 exons), and given this, how many also intersected at least 10 SNP microarray probes, P(≥ 10 SNP probes | ≥ 3 exons). Even when Illumina events with only 5 probes are considered (yellow line, P(≥ 10 SNP probes | ≥ 3 exons)), the targeted nature of the exome probes provides additional power in detecting disruptive genic events.

Figure S18c: Power of exome-based vs Illumina 1M Duo SNP microarray for small genic CNVs

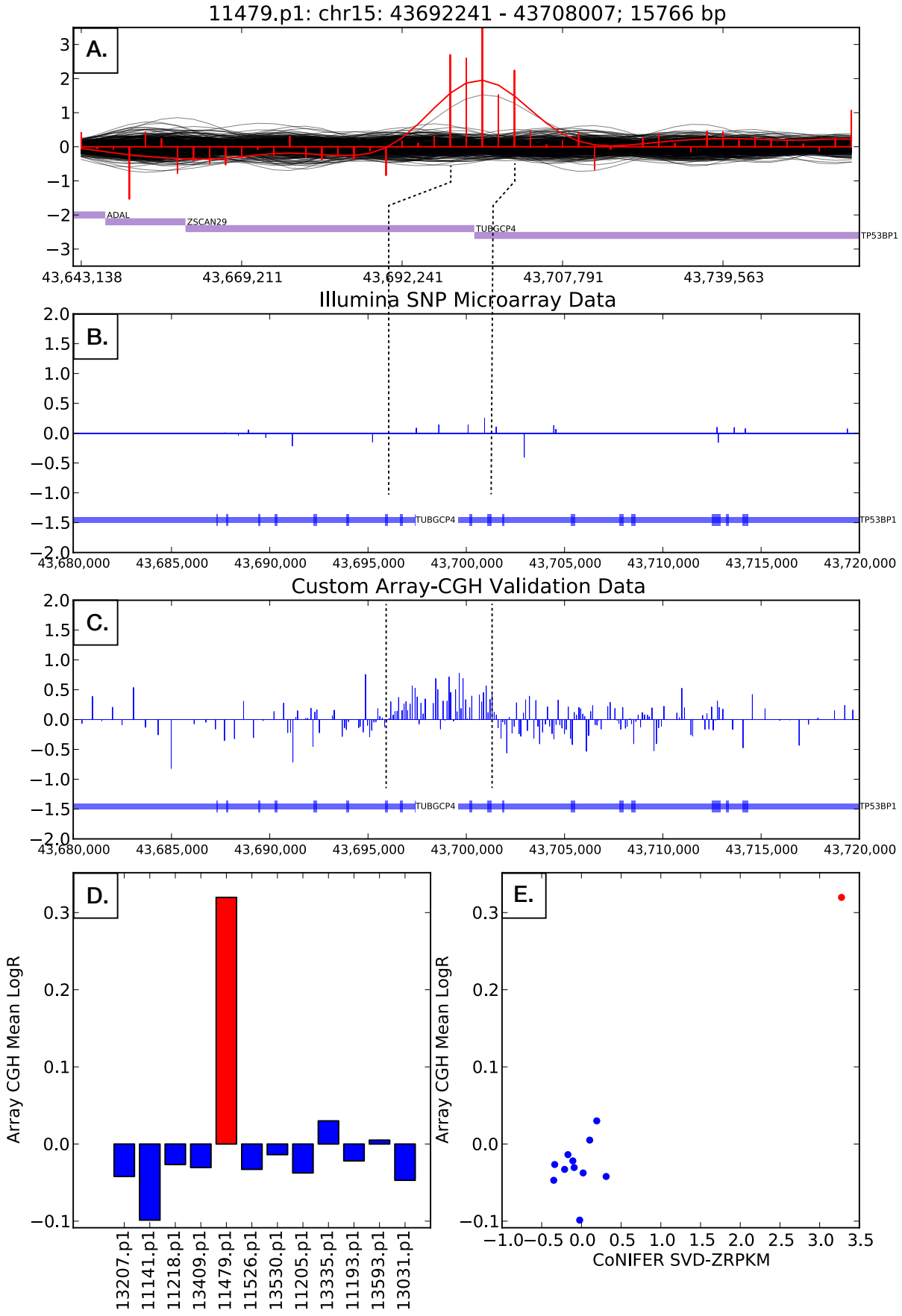


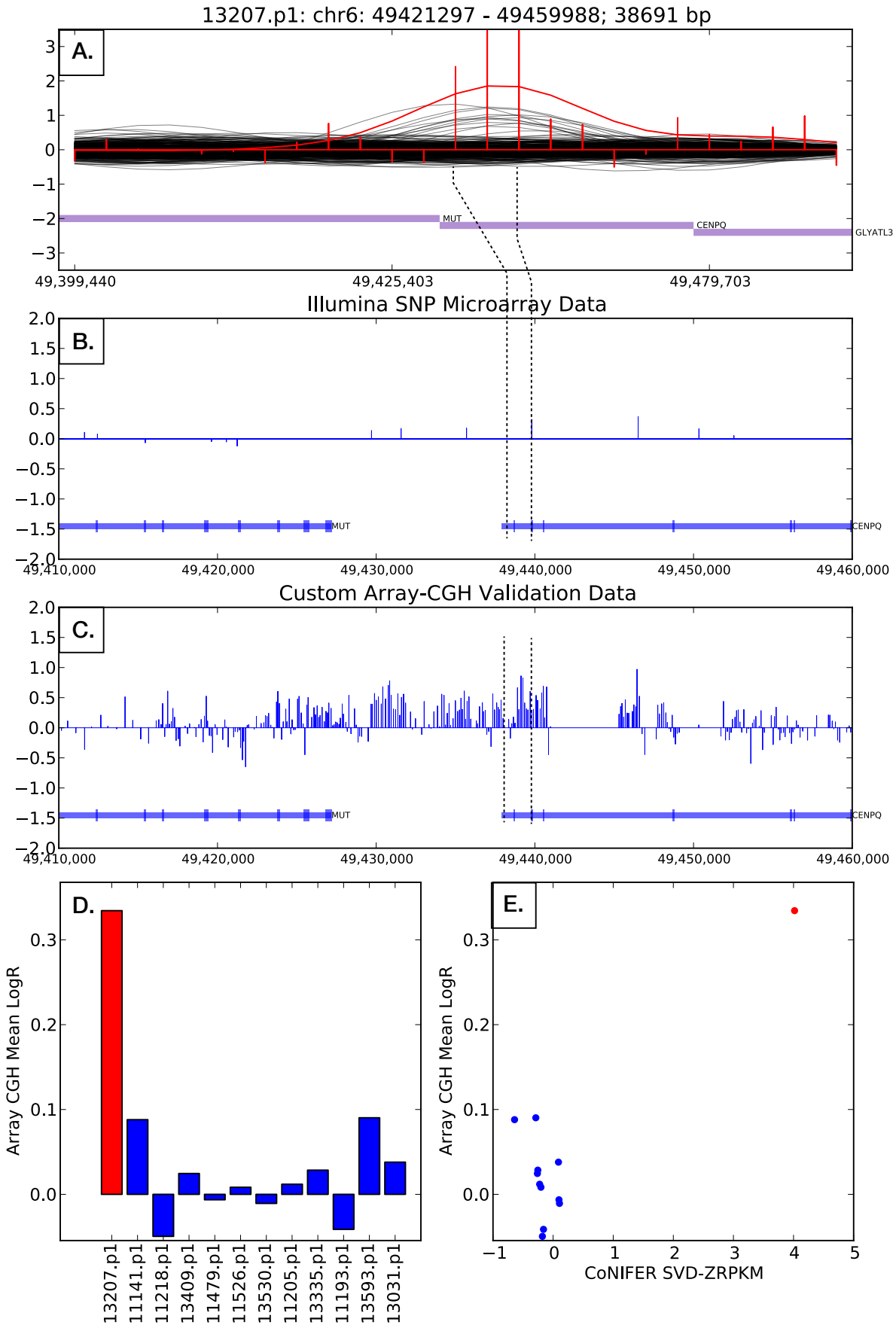
The ratio of fractions in Figure S18b represents the fraction of CNVs within the genome that would only be detected using our algorithm. For example, for 10kb genic CNVs, our algorithm can theoretically detect approximately 8.7-fold more events than the SNP microarray. Owing to the fact that the targeted exome is by default most sensitive to exons, our algorithm still has a significant detection advantage for small genic events of 5kb or less, even if only 3 SNP microarray probes are required.

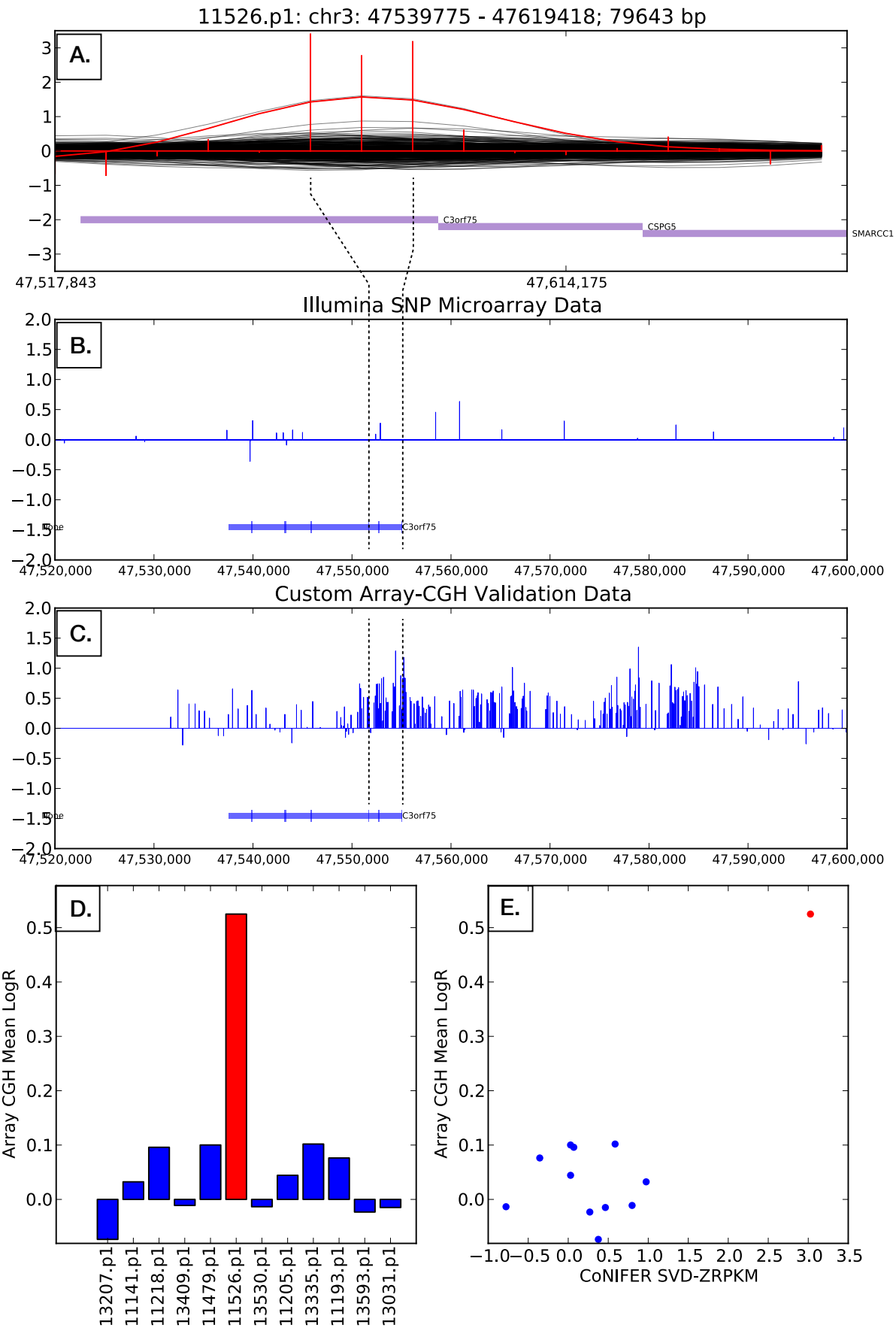
Figure S19(a-j): Custom array-CGH validation of novel CNVs and CNPs

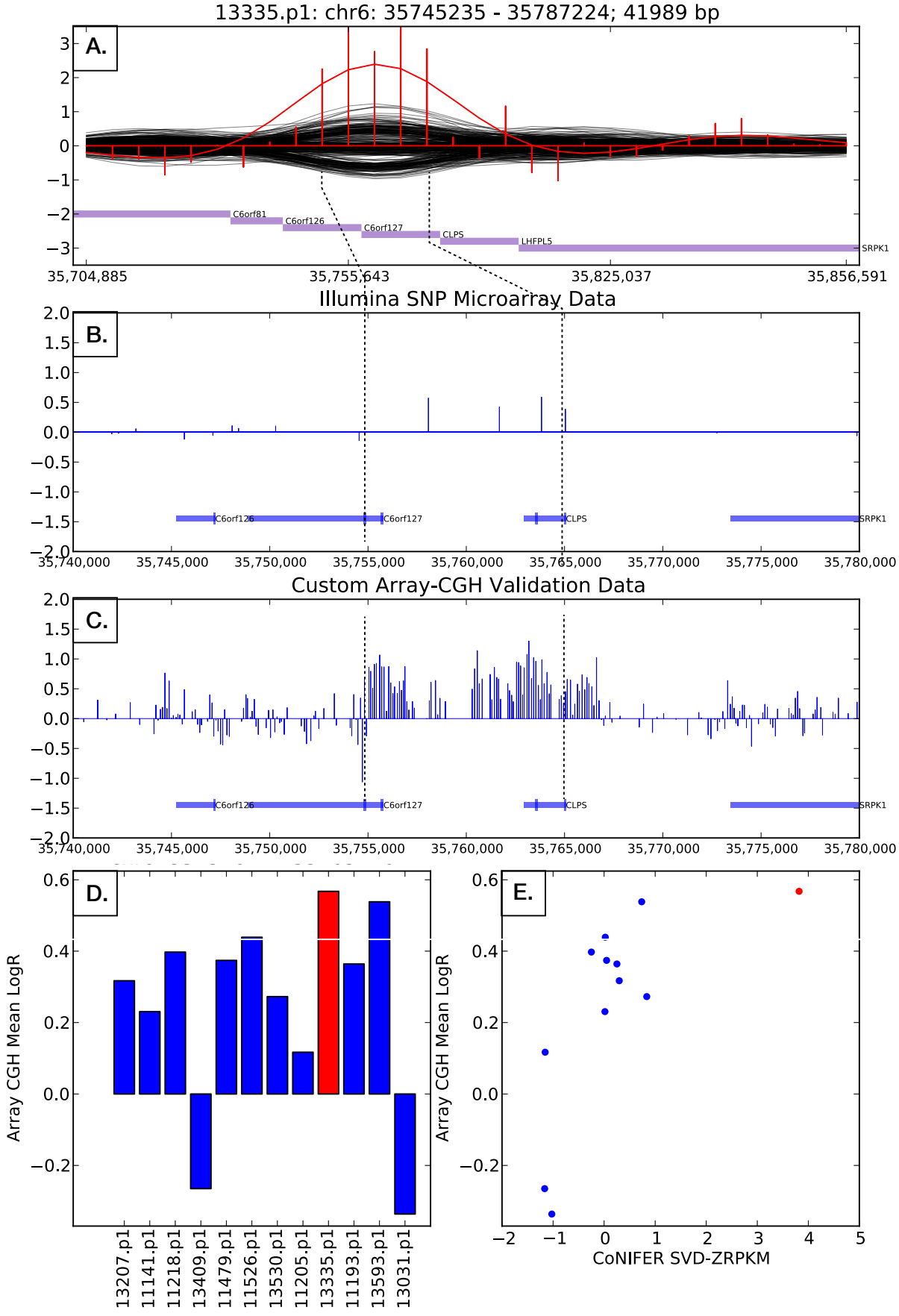
We designed a Nimblegen custom 12x135K CGH array with ~100bp probe spacing near the targeted regions at between 500bp and 30kb probe spacing elsewhere. We used NA18507 (Male) or NA12878 (Female) as reference samples in all experiments, depending on the sex of the test sample. Subplots are as follows: A) SVD-ZRPKM values (red) for the test sample, with all other ASD probands in black. Note that we shows exons in exon-space, not genomic spacing. The approximate location and size of the call is given; however, these boundaries are approximate due to the low resolution of the exome probes. B) Previously generated Illumina 1M or 1M Duo SNP microarray data (Sanders et al, 2011) for the region. The dotted lines indicate the location of the minimal common set of duplicated or deleted exons to their hg19 coordinates. C) Our custom high-density array-CGH data. D) Mean LogR of array-CGH data from (C) as compared to 10 other samples run on the same array. The test sample is highlighted in red. E) Correlation between SVD-ZRPKM values and mean LogR ratio for all 10 samples. Test sample is highlighted in red.

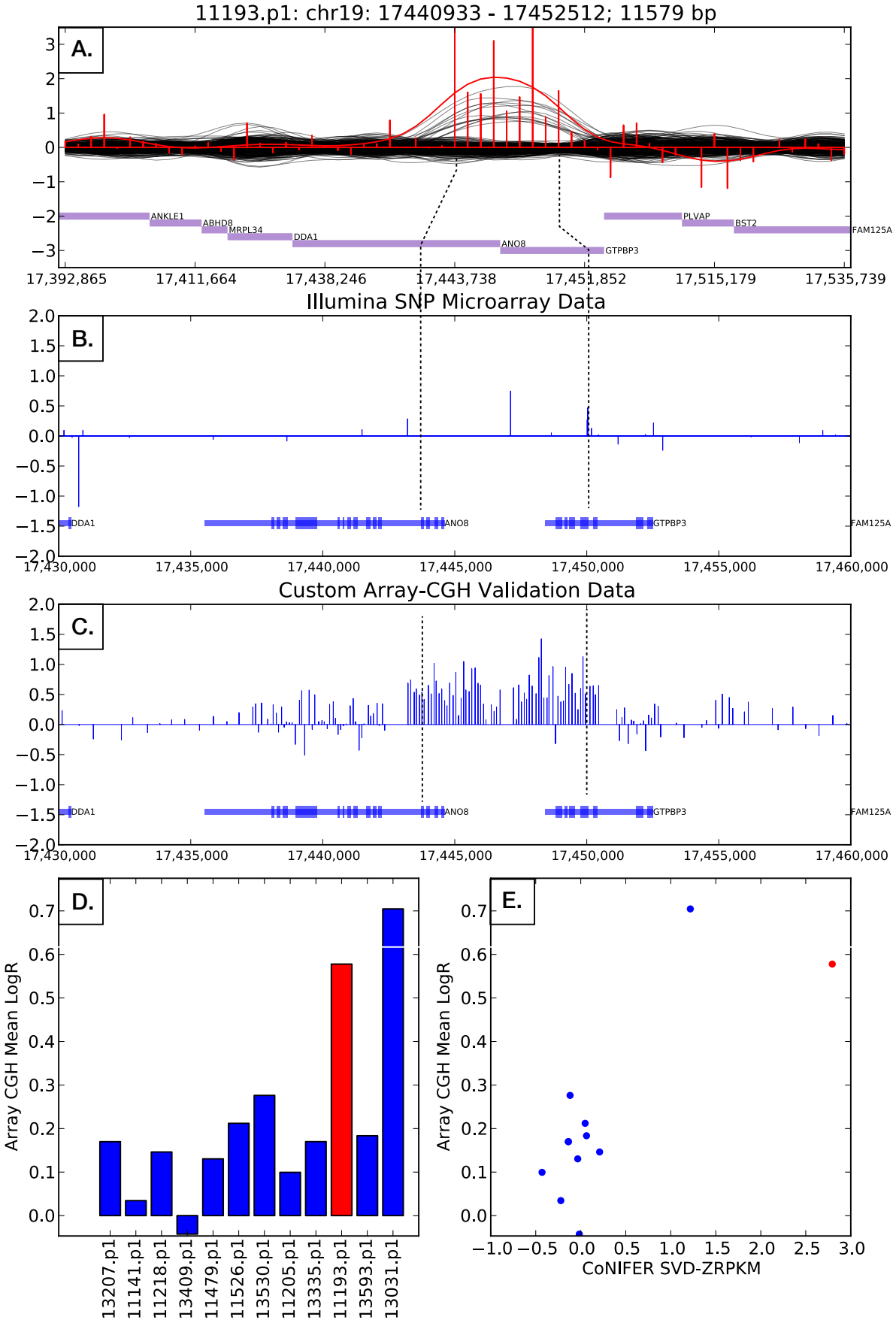
Figure S19a-j

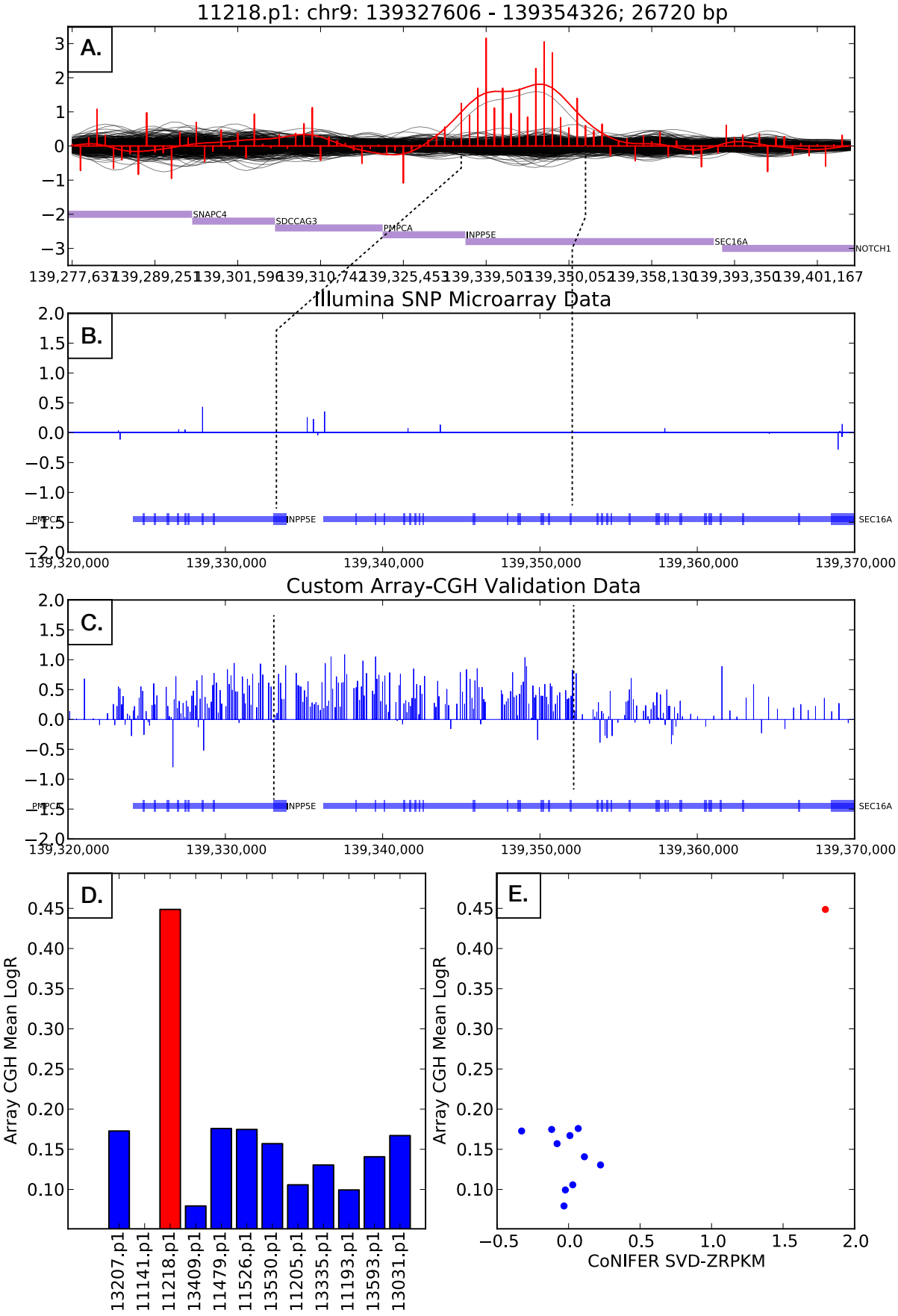


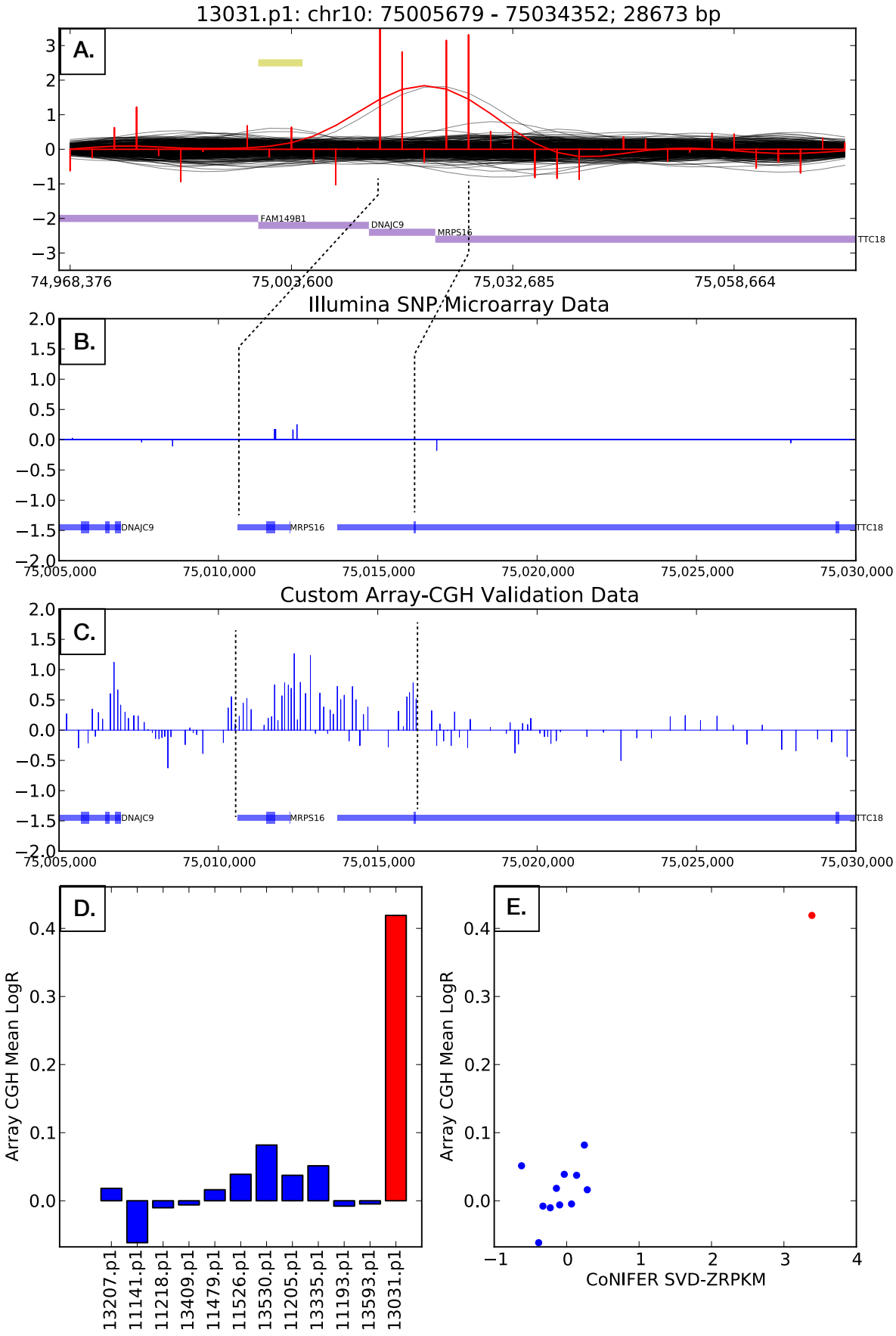

13207.p1: chr6: 49421297 - 49459988; 38691 bp
A.
MUT
CENPQ
GLYATL3
49,399,440
49,425,403
49,479,703
Illumina SNP Microarray Data
B.
MUT
CENPQ
49,410,000
49,420,000
49,430,000
49,440,000
49,450,000
49,460,000
Custom Array-CGH Validation Data
C.
MUT
CENPQ
49,410,000
49,420,000
49,430,000
49,440,000
49,450,000
49,460,000
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
Array CGH Mean LogR
CoNIFER SVD-ZRPKM

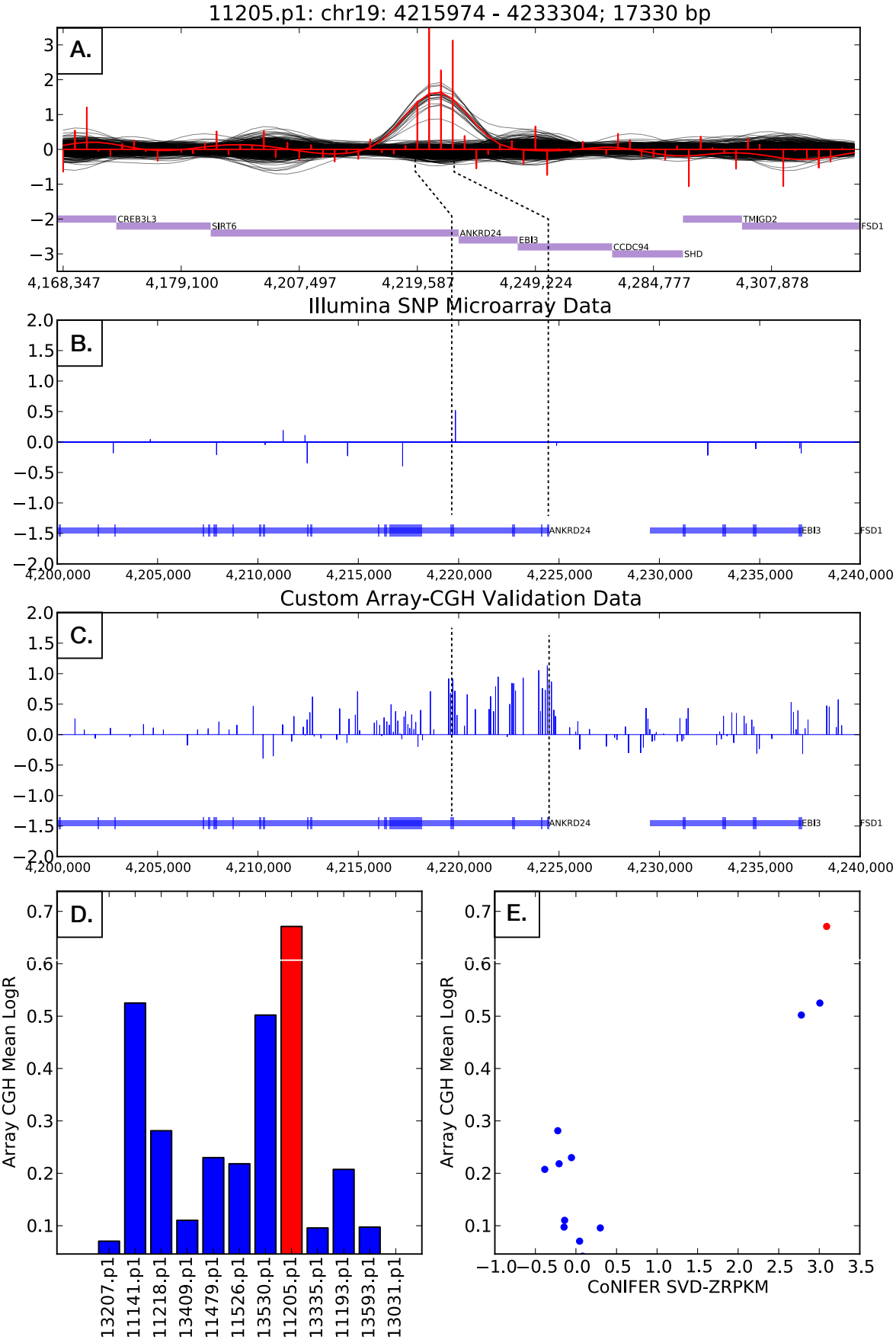

11526.p1: chr3: 47539775 - 47619418; 79643 bp
A.
C3orf75
CSPG5
SMARCC1
47,517,843
47,614,175
Illumina SNP Microarray Data
B.
Gene
C3orf75
47,520,000
47,530,000
47,540,000
47,550,000
47,560,000
47,570,000
47,580,000
47,590,000
47,600,000
Custom Array-CGH Validation Data
C.
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
CoNIFER SVD-ZRPKM

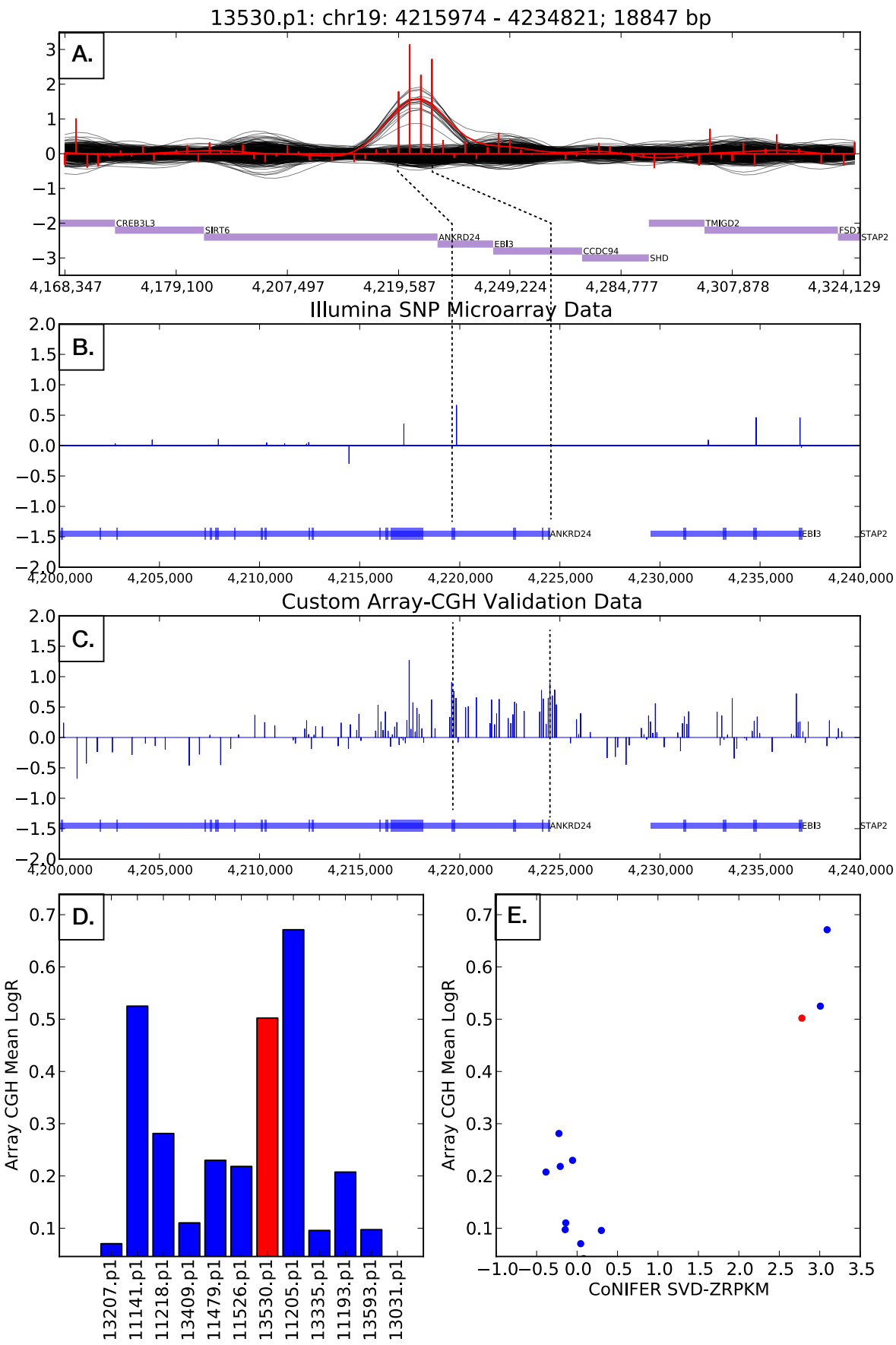

13335.p1: chr6: 35745235 - 35787224; 41989 bp
A.
C6orf81
C6orf126
C6orf127
CLPS
LHFPL5
SRPK1
35,704,885
35,755,643
35,825,037
35,856,591
Illumina SNP Microarray Data
B.
C6orf126
C6orf127
CLPS
SRPK1
35,740,000
35,745,000
35,750,000
35,755,000
35,760,000
35,765,000
35,770,000
35,775,000
35,780,000
Custom Array-CGH Validation Data
C.
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
Array CGH Mean LogR
CoNIFER SVD-ZRPKM


11193.p1: chr19: 17440933 - 17452512; 11579 bp
A.
ANKLE1
ABHD8
MRPL34
DDA1
ANO8
GTPBP3
PLVAP
BST2
FAM125A
17,392,865
17,411,664
17,438,246
17,443,738
17,451,852
17,515,179
17,535,739
Illumina SNP Microarray Data
B.
17,430,000
17,435,000
17,440,000
17,445,000
17,450,000
17,455,000
17,460,000
Custom Array-CGH Validation Data
C.
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
CoNIFER SVD-ZRPKM


11218.p1: chr9: 139327606 - 139354326; 26720 bp
A.
SNAPC4
SDCCAG3
PMPCA
INPP5E
SEC16A
NOTCH1
139,277,637
139,289,251
139,301,596
139,310,741
139,325,451
139,339,503
139,350,052
139,358,130
139,393,350
139,401,167
Illumina SNP Microarray Data
B.
PMPCA
INPP5E
SEC16A
139,320,000
139,330,000
139,340,000
139,350,000
139,360,000
139,370,000
Custom Array-CGH Validation Data
C.
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
CoNIFER SVD-ZRPKM


13031.p1: chr10: 75005679 - 75034352; 28673 bp
A.
FAM149B1
DNAJC9
MRPS16
TTC18
74,968,376
75,003,600
75,032,685
75,058,664
Illumina SNP Microarray Data
B.
DNAJC9
MRPS16
TTC18
75,005,000
75,010,000
75,015,000
75,020,000
75,025,000
75,030,000
Custom Array-CGH Validation Data
C.
DNAJC9
MRPS16
TTC18
75,005,000
75,010,000
75,015,000
75,020,000
75,025,000
75,030,000
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
Array CGH Mean LogR
CoNIFER SVD-ZRPKM


11205.p1: chr19: 4215974 - 4233304; 17330 bp
A.
CREB3L3
SIRT6
ANKRD24
EBI3
CCDC94
SHD
TMIGD2
FSD1
4,168,347
4,179,100
4,207,497
4,219,587
4,249,224
4,284,777
4,307,878
Illumina SNP Microarray Data
B.
ANKRD24
EBI3
FSD1
4,200,000
4,205,000
4,210,000
4,215,000
4,220,000
4,225,000
4,230,000
4,235,000
4,240,000
Custom Array-CGH Validation Data
C.
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
Array CGH Mean LogR
CoNIFER SVD-ZRPKM


13530.p1: chr19: 4215974 - 4234821; 18847 bp
A.
CREB3L3
SIRT6
ANKRD24
EBI3
CCDC94
SHD
TMIGD2
FSD1
STAP2
4,168,347
4,179,100
4,207,497
4,219,587
4,249,224
4,284,777
4,307,878
4,324,129
Illumina SNP Microarray Data
B.
Custom Array-CGH Validation Data
C.
4,200,000
4,205,000
4,210,000
4,215,000
4,220,000
4,225,000
4,230,000
4,235,000
4,240,000
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
CoNIFER SVD-ZRPKM

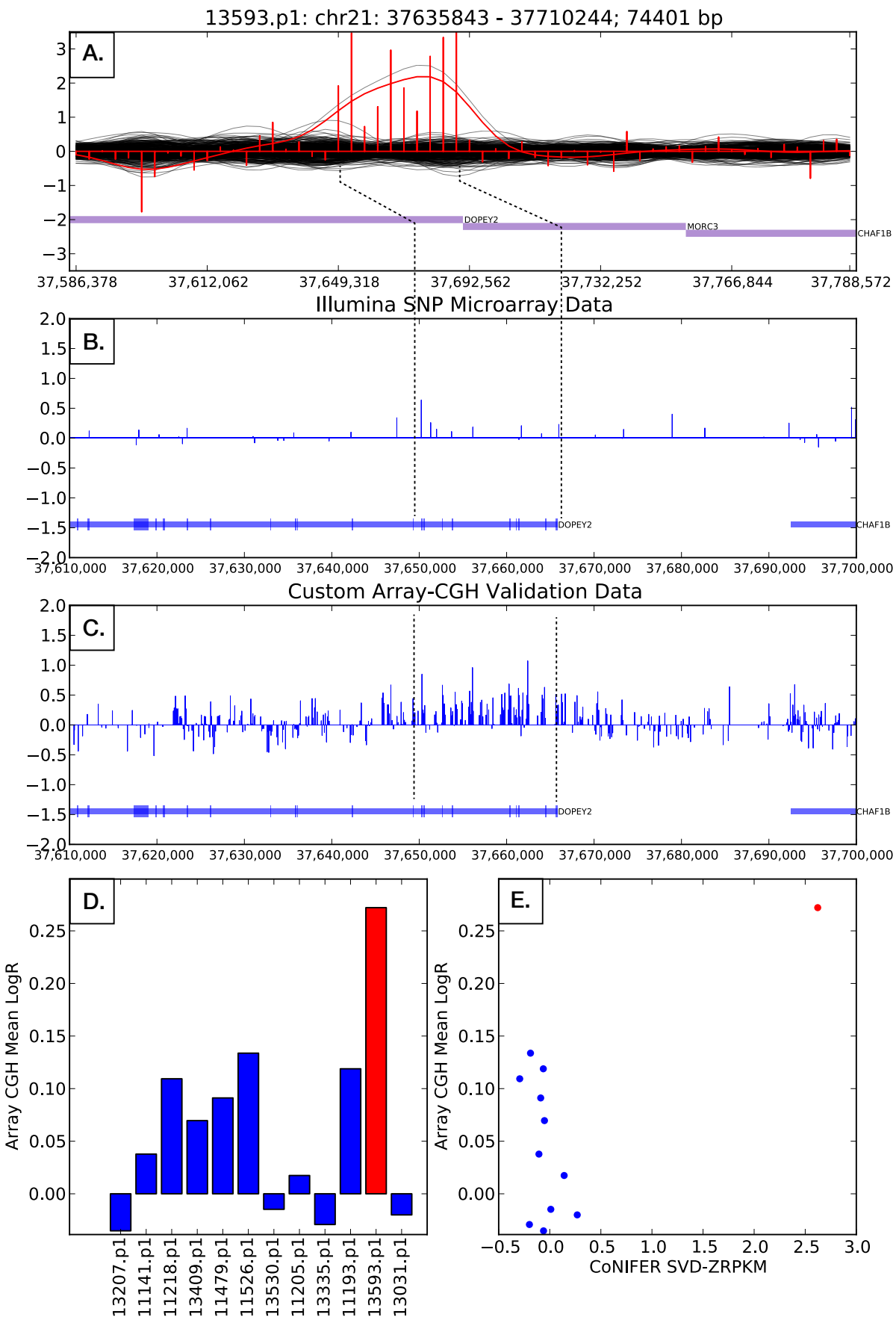
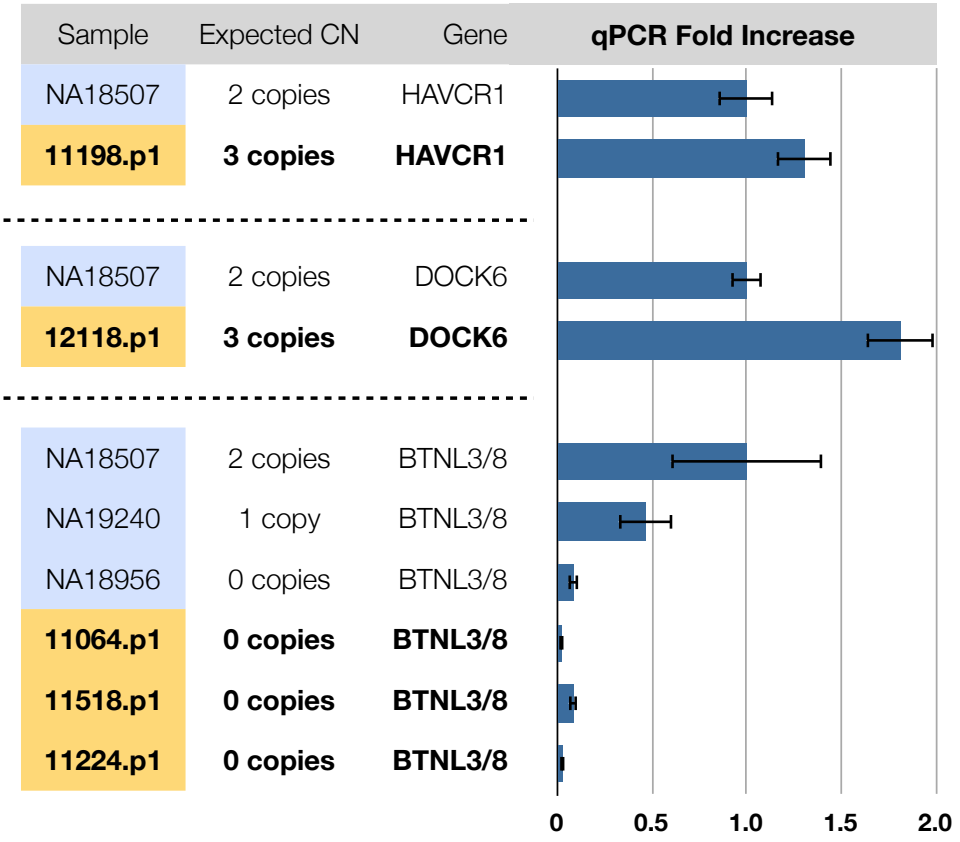
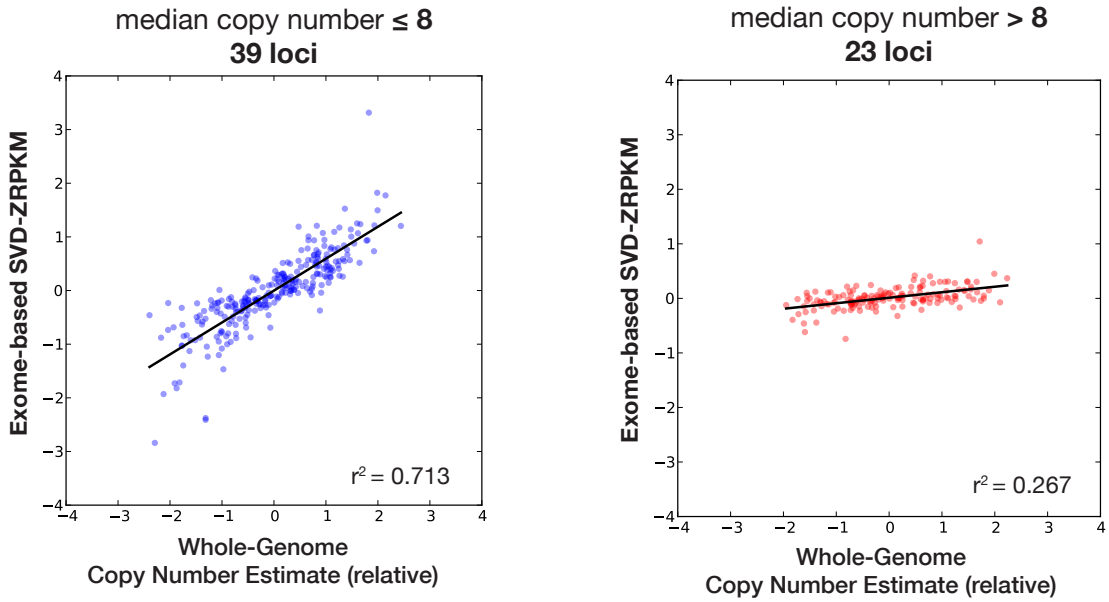

13593.p1: chr21: 37635843 - 37710244; 74401 bp
A.
DOPEY2
MORC3
CHAF1B
37,586,378
37,612,062
37,649,318
37,692,562
37,732,252
37,766,844
37,788,572
Illumina SNP Microarray Data
B.
DOPEY2
CHAF1B
37,610,000
37,620,000
37,630,000
37,640,000
37,650,000
37,660,000
37,670,000
37,680,000
37,690,000
37,700,000
Custom Array-CGH Validation Data
C.
DOPEY2
CHAF1B
D.
Array CGH Mean LogR
13207.p1
11141.p1
11218.p1
13409.p1
11479.p1
11526.p1
13530.p1
11205.p1
13335.p1
11193.p1
13593.p1
13031.p1
E.
Array CGH Mean LogR
CoNIFER SVD-ZRPKM

Figure S21: qPCR results for 2 novel CNVs and 1 CNP



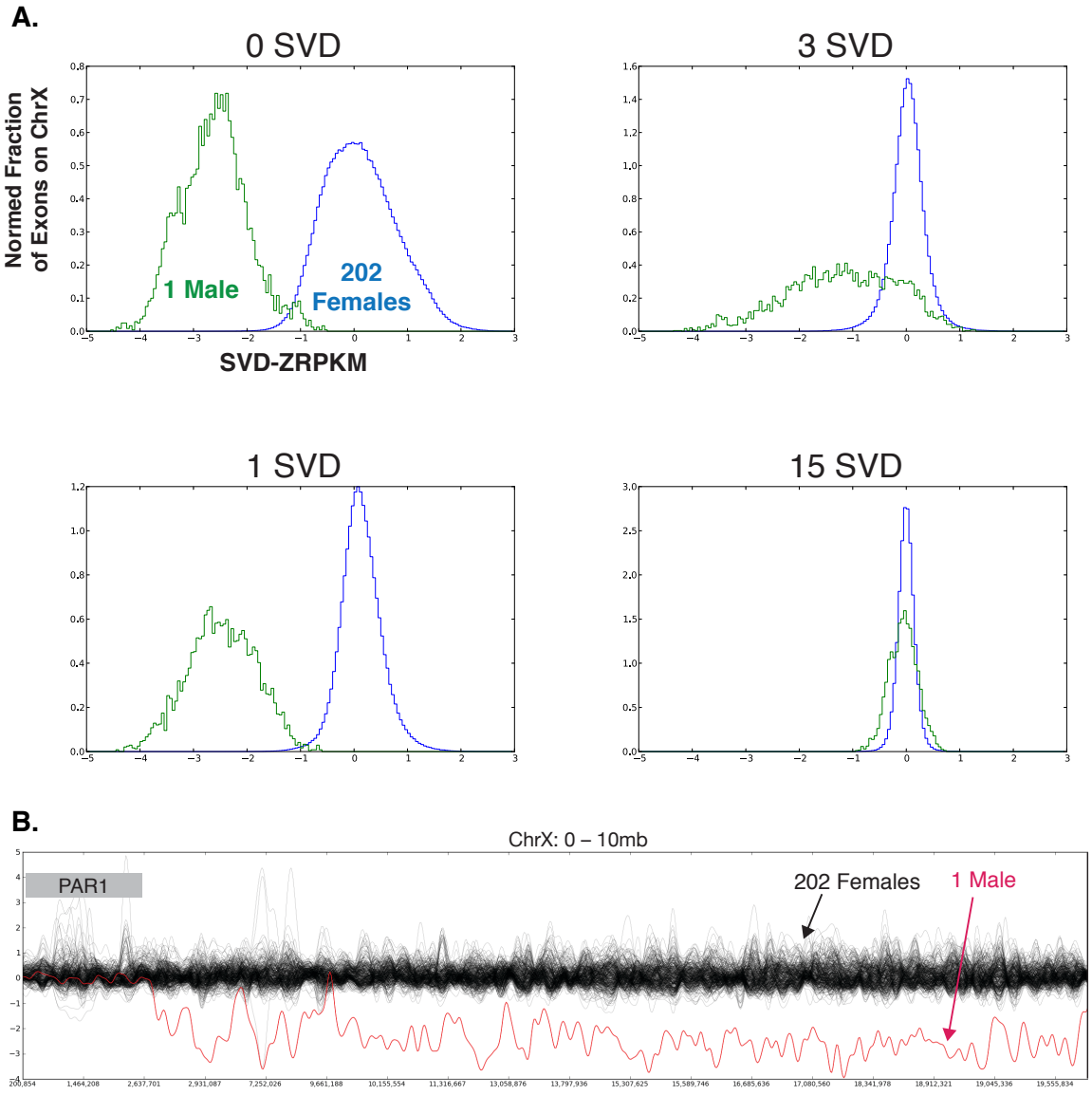
We performed qPCR assays on 3 loci not originally included in our array-CGH experiment (Figure S19). All primers used and conditions are listed in the Supplemental Note. The *HAVCR1* and *DOCK6* events are expected to be rare CNVs and we tested the ASD proband against the NA18507 reference. The BTNL3/8 locus is a polymorphic CNP, and we generated expected copy numbers for the HapMap samples from whole-genome copy number estimates (Sudmant et al. 2010). We validated 3 calls in 3 samples for this locus, and extend this to the remaining 7 calls at this locus.

Figure S21: Correlation between SVD-ZRPKM and copy number



(left) Correlation between SVD-ZRPKM score and relative (by median and standard deviation) whole-genome copy number estimate for 39 loci with ≤ 8 copies; and (right) for 23 loci with > 8 copies. Whole-genome read-depth copy number estimates for these specific sites and genomes were orthogonally validated using single-channel intensity data from previous array-CGH experiments (Sudmant et al. 2010).

Figure S22: Effect of SVD on Chromosome X copy number



We examined how the SVD transformation affected the normalization of a single male individual when compared to 202 female samples. In (A), normalized frequency histograms for each of the non-PAR exons on the X chromosome. The large fraction of variance contributed by the single male forces its normalization by within the first few SVD components. (B) Representative view of the first 20 Mbp of Chr X with one component removed.

Table S1: Precision of HapMap Calls

Sample	hg19 Coordinates (chr - start - stop)			Call Type	Reciprocal Overlap (%)	Annotation
NA15510	1	155,227,075	155,264,176	dup	96%	**Rare**
NA19240	12	133,659,688	133,727,740	dup	15%	**Rare**
NA18517	16	21,426,277	21,756,357	dup	55%	**Rare**
NA15510	3	19,498,259	21,462,939	dup	56%	**Rare**
NA18517	4	68,788,472	69,057,034	dup	85%	**Rare**
NA15510	7	99,507,187	99,627,998	dup	51%	**Rare**
NA15510	9	108,380,239	109,692,040	dup	--	**Rare**
NA19240	1	110,230,495	110,256,383	dup	49%	**CNP**
NA12878	11	60,978,561	60,980,174	del	--	**CNP**
NA12878	11	60,997,390	60,999,003	del	--	**CNP**
NA19240	14	106,405,309	106,758,420	dup	75%	**CNP**
NA19240	15	20,649,180	20,666,895	dup	4%	**CNP**
NA19240	15	22,368,575	22,490,341	dup	40%	**CNP**
NA15510	15	22,368,575	22,738,309	dup	62%	**CNP**
NA12878	17	34,416,528	34,432,035	del	--	**CNP**
NA12878	17	34,523,196	34,539,292	del	46%	**CNP**
NA12878	17	34,624,770	34,640,858	del	33%	**CNP**
NA19240	17	39,535,604	39,551,320	dup	23%	**CNP**
NA18517	19	43,528,842	43,674,290	dup	50%	**CNP**
NA19129	22	21,833,743	21,841,516	dup	8%	**CNP**
NA12878	22	22,328,728	22,989,351	del	15%	**CNP**
NA12878	22	22,989,610	23,249,131	del	70%	**CNP**
NA12878	22	24,373,137	24,384,231	dup	18%	**CNP**
NA18517	22	24,373,609	24,384,231	del	--	**CNP**
NA15510	5	180,376,903	180,430,876	del	96%	**CNP**
NA18517	5	70,297,918	70,337,451	dup	15%	**CNP**
NA19129	6	31,948,780	31,960,321	del	--	**CNP**
NA19240	6	31,983,792	31,992,429	dup	9%	**CNP**
NA19129	6	31,983,792	31,992,729	del	--	**CNP**
NA18517	6	35,762,922	35,787,224	dup	15%	**CNP**
NA19240	7	100,319,584	100,334,703	del	45%	**CNP**
NA15510	7	100,320,286	100,334,703	del	47%	**CNP**

Table S2: Sensitivity vs. Conrad *et al.* aCGH Calls

Conrad CNVR	Sample	hg19 Coordinates (chr - start - stop)			Call Type	# exome probes	Genes	Annotation	Discovered ?	Genotyping possible?
CNVR1313.1	NA15510	3	19,535,653	20,638,501	duplication	47	*KCNH8, EFHB, RAB5A, KAT2B, SGOL1*	**Rare**	Yes	--
CNVR1952.1	NA18517	4	68,788,730	69,016,101	duplication	17	*TMPRSS11A*	**Rare**	Yes	--
CNVR3507.1	NA15510	7	99,564,133	99,625,411	duplication	6	*AZGP1, ZKSCAN1*	**Rare**	Yes	--
CNVR5791.2	NA19240	12	133,717,202	133,779,425	duplication	8	*ZNF140, ZNF10, ZNF268*	**Rare**	Yes	--
CNVR6668.5	NA18517	16	21,523,044	21,946,347	duplication	38	*METTL9, IGSF6, OTOA*	**Rare**	Yes	--
CNVR4393.2	NA12878	9	91,963,403	92,343,382	duplication	27		Conrad False Positive	--	--
CNVR6152.2	NA15510	14	50,101,898	50,942,527	duplication	161		Conrad False Positive	--	--
CNVR6631.1	NA19129	16	8,939,807	8,987,025	duplication	4		Conrad False Positive	--	--
CNVR2861.1	NA18517	6	35,754,791	35,766,680	duplication	4	*CLPS*	**CNP**	Yes	Yes
CNVR3509.1	NA19240	7	100,327,862	100,337,886	deletion	6	*EPO*	**CNP**	Yes	Yes
CNVR8114.1	NA12878	22	24,344,211	24,404,564	duplication	7	*GSTT1*	**CNP**	Yes	Yes
CNVR339.1	NA15510	1	144,950,054	145,080,140	duplication	7	*PDE4DIP*	**CNP**	No	Yes
CNVR339.1	NA12878	1	144,948,283	145,080,140	duplication	7	*PDE4DIP*	**CNP**	No	Yes
CNVR339.1	NA18517	1	144,956,694	145,083,994	duplication	4	*PDE4DIP*	**CNP**	No	Yes
CNVR759.1	NA19129	2	38,956,285	38,972,493	deletion	5	*GALM*	**CNP**	No	Yes
CNVR759.1	NA12878	2	38,955,877	38,972,493	deletion	5	*GALM*	**CNP**	No	Yes
CNVR759.1	NA19240	2	38,955,946	38,972,830	deletion	5	*GALM*	**CNP**	No	Yes
CNVR759.1	NA18517	2	38,956,285	38,972,937	deletion	5	*GALM*	**CNP**	No	Yes
CNVR2719.1	NA18517	5	180,374,610	180,431,110	duplication	11	*BTNL8/3*	**CNP**	No	Yes
CNVR2719.1	NA12878	5	180,376,223	180,430,715	duplication	8	*BTNL8/3*	**CNP**	No	Yes
CNVR2728.1	NA19240	6	257,100	382,983	duplication	8	*DUSP22*	**CNP**	No	Yes
CNVR2728.1	NA12878	6	255,650	382,508	duplication	8	*DUSP22*	**CNP**	No	Yes
CNVR2728.1	NA19129	6	254,458	382,453	duplication	8	*DUSP22*	**CNP**	No	Yes
CNVR2728.1	NA18517	6	257,309	384,408	duplication	8	*DUSP22*	**CNP**	No	Yes
CNVR2861.1	NA19129	6	35,754,996	35,766,570	duplication	4	*CLPS*	**CNP**	No	Yes
CNVR2861.1	NA19240	6	35,754,736	35,766,415	duplication	5	*CLPS*	**CNP**	No	Yes
CNVR2861.1	NA12878	6	35,754,591	35,766,415	duplication	5	*CLPS*	**CNP**	No	Yes
CNVR4912.1	NA18517	10	124,360,402	124,376,587	deletion	8	*DMBT1*	**CNP**	No	Yes
CNVR4912.1	NA12878	10	124,360,512	124,376,427	deletion	7	*DMBT1*	**CNP**	No	Yes
CNVR4912.2	NA18517	10	124,342,529	124,351,752	duplication	6	*DMBT1*	**CNP**	No	Yes

Conrad CNVR	Sample	hg19 Coordinates (chr - start - stop)			Call Type	# exome probes	Genes	Annotation	Discovered ?	Genotyping possible?
CNVR4912.3	NA12878	10	124,342,337	124,360,459	duplication	14	*DMBT1*	**CNP**	No	Yes
CNVR4912.3	NA19240	10	124,342,556	124,360,682	duplication	15	*DMBT1*	**CNP**	No	Yes
CNVR4912.3	NA19129	10	124,341,208	124,358,673	duplication	15	*DMBT1*	**CNP**	No	Yes
CNVR5179.1	NA19129	11	55,366,154	55,452,992	deletion	6	*OR4*	**CNP**	No	Yes
CNVR5179.1	NA12878	11	55,365,742	55,453,061	deletion	6	*OR4*	**CNP**	No	Yes
CNVR6072.4	NA19240	14	20,177,270	20,422,582	duplication	6	*OR4*	**CNP**	No	Yes
CNVR6072.5	NA18517	14	20,289,680	20,424,616	deletion	4	*OR4*	**CNP**	No	Yes
CNVR7095.1	NA18517	17	39,382,871	39,395,430	deletion	3	*KRTAP*	**CNP**	No	No
CNVR7097.1	NA19240	17	39,507,055	39,525,624	deletion	6	*KRT34*	**CNP**	No	Yes
CNVR7098.1	NA19240	17	39,532,301	39,539,205	duplication	7	*KRT34*	**CNP**	No	Yes
CNVR7673.1	NA12878	19	46,622,831	46,628,261	deletion	3	*IGFL3*	**CNP**	No	Yes
CNVR7702.1	NA19129	19	52,131,804	52,148,913	deletion	8	*SIGLEC14*	**CNP**	No	Yes
CNVR7708.1	NA19240	19	53,322,989	53,361,358	duplication	3	*ZNF468*	**CNP**	No	No
CNVR7763.1	NA19129	20	1,552,963	1,595,689	deletion	4	*SIRBP1*	**CNP**	No	Yes
CNVR7763.1	NA18517	20	1,552,963	1,595,689	deletion	4	*SIRBP1*	**CNP**	No	Yes
CNVR7763.1	NA19240	20	1,552,963	1,595,689	deletion	4	*SIRBP1*	**CNP**	No	Yes
CNVR8114.1	NA19129	22	24,344,595	24,404,495	duplication	7	*GSTT1*	**CNP**	No	Yes
CNVR8114.7	NA19240	22	24,364,568	24,404,701	duplication	7	*GSTT1*	**CNP**	No	Yes
CNVR8114.7	NA15510	22	24,371,095	24,404,715	duplication	7	*GSTT1*	**CNP**	No	Yes

Table S3: Calls and validation in ASD probands:

sample	hg19 coordinates (chr, start, stop)			state	type	genes	validation
11205.p1	19	4,215,974	4,233,304	dup	CNP	*ANKRD24*	Yes / Custom aCGH
13530.p1	19	4,215,974	4,234,821	dup	CNP	*ANKRD24*	Yes / Custom aCGH
11141.p1	19	4,212,596	4,249,325	dup	CNP	*ANKRD24*	Yes / Custom aCGH
13409.p1	5	180,218,633	180,430,876	del	CNP	*BTNL8*	Not tested (see note)
11599.p1	5	180,375,919	180,431,443	del	CNP	*BTNL8*	Not tested (see note)
11964.p1	5	180,375,919	180,420,160	del	CNP	*BTNL8*	Not tested (see note)
12249.p1	5	180,375,919	180,431,443	del	CNP	*BTNL8*	Not tested (see note)
11753.p1	5	180,375,919	180,430,876	del	CNP	*BTNL8*	Not tested (see note)
12212.p1	5	180,375,919	180,430,876	del	CNP	*BTNL8*	Not tested (see note)
12667.p1	5	180,376,238	180,430,876	del	CNP	*BTNL8*	Not tested (see note)
11064.p1	5	180,375,919	180,431,443	del	CNP	*BTNL8*	Yes / qPCR
11224.p1	5	180,375,919	180,430,876	del	CNP	*BTNL8*	Yes / qPCR
11518.p1	5	180,375,919	180,431,443	del	CNP	*BTNL8*	Yes / qPCR
13207.p1	6	49,421,297	49,459,988	dup	CNP	*CENPQ*	Yes / Custom aCGH
13335.p1	6	35,745,235	35,787,224	dup	CNP	*CLPS*	No / Custom aCGH
11722.p1	3	16,635,161	16,640,105	dup	CNP	*DAZL*	Not tested
11471.p1	3	16,636,820	16,639,048	dup	CNP	*DAZL*	Not tested
13517.p1	3	16,636,820	16,640,105	dup	CNP	*DAZL*	Not tested
11193.p1	19	17,440,933	17,452,512	dup	CNP	*GTPBP3*	Yes / Custom aCGH
11471.p1	11	55,339,603	55,433,572	del	CNP	*OR4C*	Yes / SNP Microarray
11498.p1	11	55,339,603	55,419,315	del	CNP	*OR4C*	Yes / SNP Microarray
11587.p1	11	55,339,603	55,433,572	del	CNP	*OR4C*	Yes / SNP Microarray
11013.p1	11	55,370,916	55,419,315	del	CNP	*OR4C*	Yes / SNP Microarray
11205.p1	11	55,370,916	55,419,315	del	CNP	*OR4C*	Yes / SNP Microarray
11291.p1	11	55,370,916	55,419,315	del	CNP	*OR4C*	Yes / SNP Microarray
11753.p1	11	55,370,916	55,419,315	del	CNP	*OR4C*	Yes / SNP Microarray
11257.p1	14	20,248,481	20,529,142	dup	CNP	*OR4K*	Yes / SNP Microarray
11653.p1	19	52,132,290	52,149,893	del	CNP	*SIGLEC14*	Yes / SNP Microarray
12641.p1	19	52,133,551	52,149,313	del	CNP	*SIGLEC14*	Yes / SNP Microarray
11711.p1	5	175,913,355	175,956,388	dup	de novo	*FAF2*	No / Custom aCGH
11218.p1	5	175,913,355	175,956,645	dup	de novo	*FAF2*	No / Custom aCGH
13726.p1	11	55,510,303	61,235,941	del	de novo		Yes / Array CGH (O'Roak et al.)
11696.p1	3	37,170,553	37,494,050	del	de novo		Yes / SNP Microarray
12581.p1	9	140,671,069	141,015,333	del	de novo		Yes / SNP Microarray
11928.p1	15	30,919,023	32,404,100	dup	de novo		Yes / SNP Microarray
13335.p1	16	29,475,783	30,204,395	dup	de novo		Yes / SNP Microarray
11526.p1	16	75,481,455	75,600,805	dup	de novo		Yes / SNP Microarray
11526.p1	3	47,539,775	47,619,418	dup	inherited	*C3ORF75/CSPG5*	Yes / Custom aCGH
12118.p1	19	11,319,586	11,363,226	dup	inherited	*DOCK6*	Yes / qPCR
13593.p1	21	37,635,843	37,710,244	dup	inherited	*DOPEY6*	Yes / Custom aCGH
11198.p1	5	156,378,522	156,482,544	dup	inherited	*HAVCR1*	Yes / qPCR
12430.p1	15	100,269,327	100,537,794	dup	inherited	*LYSMD4*	Not tested
13031.p1	10	75,005,679	75,034,352	dup	inherited	*MRSP16*	Yes / Custom aCGH
11218.p1	9	139,327,606	139,354,326	dup	inherited	*SEC16A*	Yes / Custom aCGH
11479.p1	15	43,692,241	43,708,007	dup	inherited	*TP53BP1*	Yes / Custom aCGH
12810.p1	1	86,965,336	87,043,755	del	inherited		Yes / SNP Microarray
11715.p1	1	185,089,514	185,137,530	dup	inherited		Yes / SNP Microarray
12667.p1	1	185,092,988	185,144,245	dup	inherited		Yes / SNP Microarray
11895.p1	1	206,241,532	206,557,431	del	inherited		Yes / SNP Microarray
12130.p1	1	206,241,532	206,557,431	del	inherited		Yes / SNP Microarray

sample	hg19 coordinates (chr, start, stop)			state	type	genes	validation
11707.p1	1	207,307,748	207,640,257	dup	inherited		Yes / SNP Microarray
11064.p1	2	33,622,199	36,691,798	dup	inherited		Yes / SNP Microarray
11472.p1	2	44,508,525	44,549,039	dup	inherited		Yes / SNP Microarray
11895.p1	2	86,276,282	86,677,085	dup	inherited		Yes / SNP Microarray
11023.p1	2	198,285,151	198,593,302	dup	inherited		Yes / SNP Microarray
11023.p1	2	209,027,927	209,104,727	del	inherited		Yes / SNP Microarray
11722.p1	3	100,287,665	100,451,516	dup	inherited		Yes / SNP Microarray
11303.p1	3	100,295,768	100,447,702	dup	inherited		Yes / SNP Microarray
13335.p1	3	141,712,379	142,090,170	dup	inherited		Yes / SNP Microarray
12565.p1	3	151,461,880	152,018,156	del	inherited		Yes / SNP Microarray
11262.p1	3	151,461,880	152,018,156	del	inherited		Yes / SNP Microarray
11224.p1	4	5,699,319	5,795,444	dup	inherited		Yes / SNP Microarray
11190.p1	4	107,845,110	108,935,744	dup	inherited		Yes / SNP Microarray
11788.p1	5	32,093,012	32,235,235	dup	inherited		Yes / SNP Microarray
11056.p1	5	32,093,012	32,235,235	dup	inherited		Yes / SNP Microarray
11480.p1	5	32,097,384	32,242,233	dup	inherited		Yes / SNP Microarray
11469.p1	5	112,899,555	113,740,553	dup	inherited		Yes / SNP Microarray
12130.p1	5	112,902,788	113,740,553	dup	inherited		Yes / SNP Microarray
11193.p1	5	158,523,981	158,634,904	dup	inherited		Yes / SNP Microarray
11459.p1	5	158,600,990	158,697,453	dup	inherited		Yes / SNP Microarray
11480.p1	6	25,923,922	26,368,495	dup	inherited		Yes / SNP Microarray
11425.p1	6	56,882,004	56,993,638	dup	inherited		Yes / SNP Microarray
11459.p1	6	88,311,501	88,374,577	del	inherited		Yes / SNP Microarray
12212.p1	6	107,420,452	107,824,999	del	inherited		Yes / SNP Microarray
11518.p1	6	168,317,768	168,442,831	dup	inherited		Yes / SNP Microarray
12933.p1	6	168,319,414	168,711,126	dup	inherited		Yes / SNP Microarray
11472.p1	6	168,319,414	168,711,964	dup	inherited		Yes / SNP Microarray
12667.p1	6	168,323,535	168,442,831	dup	inherited		Yes / SNP Microarray
11722.p1	6	168,323,535	168,439,409	dup	inherited		Yes / SNP Microarray
11863.p1	6	168,323,535	168,458,019	dup	inherited		Yes / SNP Microarray
13557.p1	6	168,325,684	168,711,126	dup	inherited		Yes / SNP Microarray
11398.p1	7	11,101,590	12,620,846	dup	inherited		Yes / SNP Microarray
11696.p1	7	16,834,559	17,838,777	dup	inherited		Yes / SNP Microarray
12667.p1	7	33,066,428	33,297,022	dup	inherited		Yes / SNP Microarray
11722.p1	7	48,285,108	48,431,736	del	inherited		Yes / SNP Microarray
11526.p1	7	142,659,290	142,961,260	del	inherited		Yes / SNP Microarray
11218.p1	7	142,723,286	142,960,678	del	inherited		Yes / SNP Microarray
11843.p1	7	152,740,571	154,664,403	del	inherited		Yes / SNP Microarray
11141.p1	8	13,071,835	15,480,758	dup	inherited		Yes / SNP Microarray
11556.p1	8	15,601,046	16,035,497	del	inherited		Yes / SNP Microarray
12130.p1	8	15,601,046	16,032,809	del	inherited		Yes / SNP Microarray
12378.p1	9	134,360,072	134,458,089	dup	inherited		Yes / SNP Microarray
12378.p1	10	82,040,435	82,122,829	dup	inherited		Yes / SNP Microarray
12130.p1	10	132,965,059	133,761,295	dup	inherited		Yes / SNP Microarray
12118.p1	10	133,106,473	134,523,960	dup	inherited		Yes / SNP Microarray
11498.p1	10	135,233,529	135,368,588	dup	inherited		Yes / SNP Microarray
11148.p1	10	135,233,529	135,372,455	dup	inherited		Yes / SNP Microarray
11707.p1	10	135,340,899	135,372,455	dup	inherited		Yes / SNP Microarray
11498.p1	10	135,370,262	135,372,455	dup	inherited		Yes / SNP Microarray
11964.p1	11	14,856,527	14,989,400	dup	inherited		Yes / SNP Microarray
12430.p1	11	31,128,044	31,451,948	del	inherited		Yes / SNP Microarray

sample	hg19 coordinates (chr, start, stop)			state	type	genes	validation
13008.p1	12	306,542	922,980	dup	inherited		Yes / SNP Microarray
11526.p1	12	15,035,072	15,090,986	dup	inherited		Yes / SNP Microarray
12581.p1	12	112,167,609	112,323,840	dup	inherited		Yes / SNP Microarray
11083.p1	13	50,118,872	50,237,331	dup	inherited		Yes / SNP Microarray
11257.p1	13	115,004,824	115,048,418	dup	inherited		Yes / SNP Microarray
13530.p1	14	67,940,136	68,276,006	dup	inherited		Yes / SNP Microarray
13533.p1	14	74,512,762	74,551,696	del	inherited		Yes / SNP Microarray
13415.p1	15	57,555,309	57,816,949	dup	inherited		Yes / SNP Microarray
11556.p1	15	89,760,350	89,817,535	del	inherited		Yes / SNP Microarray
11834.p1	16	21,763,689	22,538,986	dup	inherited		Yes / SNP Microarray
11184.p1	16	81,171,041	81,194,510	del	inherited		Yes / SNP Microarray
11964.p1	16	84,402,221	84,474,564	del	inherited		Yes / SNP Microarray
13335.p1	17	644,540	708,487	del	inherited		Yes / SNP Microarray
11707.p1	17	3,981,176	4,434,078	dup	inherited		Yes / SNP Microarray
11947.p1	17	39,502,370	39,553,791	dup	inherited		Yes / SNP Microarray
13409.p1	17	72,322,488	72,733,256	dup	inherited		Yes / SNP Microarray
12667.p1	18	39,613,789	40,503,728	dup	inherited		Yes / SNP Microarray
13494.p1	18	76,873,240	77,132,882	del	inherited		Yes / SNP Microarray
13116.p1	19	45,822,778	45,909,976	dup	inherited		Yes / SNP Microarray
11013.p1	20	6,100,050	8,352,097	dup	inherited		Yes / SNP Microarray
12810.p1	22	32,495,169	32,788,346	dup	inherited		Yes / SNP Microarray
11947.p1	22	40,711,286	41,077,932	dup	inherited		Yes / SNP Microarray
11653.p1	22	41,568,502	41,634,889	dup	inherited		Yes / SNP Microarray

Calls in segmental duplications and processed pseudogenes:

sample	hg19 coordinates (chr, start, stop)			state	type	genes
12114.p1	2	179,255,799	179,315,757	Dup	PPG	*PRKRA*
11141.p1	2	179,296,823	179,315,757	Dup	PPG	*PRKRA*
11190.p1	2	179,296,823	179,315,757	Dup	PPG	*PRKRA*
12744.p1	2	179,296,823	179,315,757	Dup	PPG	*PRKRA*
11788.p1	2	179,296,823	179,318,347	Dup	PPG	*PRKRA*
12810.p1	2	179,300,871	179,315,757	Dup	PPG	*PRKRA*
11013.p1	2	179,300,871	179,320,878	Dup	PPG	*PRKRA*
11571.p1	2	179,300,871	179,315,757	Dup	PPG	*PRKRA*
11707.p1	2	179,300,871	179,312,313	Dup	PPG	*PRKRA*
11834.p1	2	179,300,871	179,315,757	Dup	PPG	*PRKRA*
11414.p1	2	179,300,871	179,315,170	Dup	PPG	*PRKRA*
11452.p1	2	179,300,871	179,315,170	Dup	PPG	*PRKRA*
11009.p1	2	179,300,871	179,315,757	Dup	PPG	*PRKRA*
11346.p1	2	179,300,871	179,315,170	Dup	PPG	*PRKRA*
11504.p1	2	179,300,871	179,315,757	Dup	PPG	*PRKRA*
11587.p1	2	179,306,336	179,315,170	Dup	PPG	*PRKRA*
11843.p1	2	179,306,336	179,312,313	Dup	PPG	*PRKRA*
11193.p1	3	196,454,793	196,626,933	Dup	PPG	*PAK2*
11303.p1	5	138,643,104	138,700,432	Dup	PPG	*MATR3*
12933.p1	8	29,197,614	29,959,489	Dup	PPG	*TMEM66*
11753.p1	8	29,197,614	29,953,044	Dup	PPG	*TMEM66*
13222.p1	8	29,197,614	29,959,489	Dup	PPG	*TMEM66*
11660.p1	8	29,202,886	29,959,489	Dup	PPG	*TMEM66*
11722.p1	8	29,940,362	30,335,353	Dup	PPG	*TMEM66*
11303.p1	8	98,725,889	98,973,758	Dup	PPG	*LAPTM4B*

sample	hg19 coordinates (chr, start, stop)			state	type	genes
11638.p1	8	98,731,276	98,863,702	Dup	PPG	*LAPTM4B*
11479.p1	8	98,731,276	98,943,750	Dup	PPG	*LAPTM4B*
11414.p1	8	98,735,106	98,954,127	Dup	PPG	*LAPTM4B*
11023.p1	8	98,735,106	98,900,470	Dup	PPG	*LAPTM4B*
11827.p1	8	98,735,106	98,954,127	Dup	PPG	*LAPTM4B*
11141.p1	11	84,822,704	85,366,752	Dup	PPG	*TMEM126B*
13532.p1	11	95,555,662	95,724,887	Dup	PPG	*MTMR2*
11452.p1	11	95,560,949	95,724,887	Dup	PPG	*MTMR2*
13409.p1	12	53,291,211	53,410,394	Dup	PPG	*EIF4B*
13409.p1	12	54,639,898	54,718,965	Dup	PPG	*CBX5*
12249.p1	12	104,376,576	104,387,282	Del	PPG	*TDG*
11193.p1	13	21,720,943	21,950,794	Dup	PPG	*C13ORF3*
13409.p1	13	27,679,867	27,847,631	Dup	PPG	*RPL21*
11638.p1	17	45,201,251	45,297,419	Dup	PPG	*CDC27*
13409.p1	17	45,201,251	45,297,419	Dup	PPG	*CDC27*
12114.p1	10	124,360,506	124,377,856	Dup	PPG + SD	*DMBT1*
13031.p1	1	104,088,869	107,691,461	Dup	SD	
13532.p1	1	104,114,731	107,691,461	Dup	SD	
12114.p1	1	104,115,684	104,120,230	Dup	SD	
12603.p1	1	104,116,329	104,120,467	Dup	SD	
13557.p1	1	104,116,329	104,160,230	Dup	SD	
12114.p1	1	104,160,062	104,166,606	Dup	SD	
13557.p1	1	104,161,532	104,199,120	Dup	SD	
12603.p1	1	104,162,175	104,166,843	Dup	SD	
12114.p1	1	104,198,952	104,236,798	Dup	SD	
13557.p1	1	104,200,415	104,238,261	Dup	SD	
12603.p1	1	104,201,061	104,205,633	Dup	SD	
13415.p1	1	104,235,921	104,295,431	Dup	SD	
12114.p1	1	104,293,091	104,299,535	Dup	SD	
13557.p1	1	104,293,604	104,299,535	Dup	SD	
12603.p1	1	104,295,200	104,297,436	Dup	SD	
12073.p1	1	110,231,294	110,233,186	Dup	SD	
11504.p1	1	110,231,669	110,235,917	Dup	SD	
13222.p1	1	110,231,846	110,233,186	Dup	SD	
12212.p1	1	110,231,846	110,233,186	Dup	SD	
11184.p1	1	120,572,528	144,618,296	Dup	SD	
12703.p1	1	120,611,947	143,912,295	Dup	SD	
11184.p1	1	144,881,429	144,952,689	Dup	SD	
12703.p1	1	148,806,089	149,804,560	Dup	SD	
11184.p1	1	149,281,755	149,812,729	Dup	SD	
12667.p1	1	161,475,257	161,647,155	Dup	SD	
11480.p1	1	161,475,776	161,677,133	Dup	SD	
11660.p1	1	161,475,776	161,677,133	Dup	SD	
11722.p1	1	196,716,240	196,801,129	Del	SD	
11498.p1	1	196,716,240	196,801,129	Del	SD	
11205.p1	1	196,757,345	196,794,801	Del	SD	
11526.p1	1	202,391,747	202,403,896	Dup	SD	
11526.p1	2	89,156,574	95,539,853	Del	SD	
11083.p1	2	89,246,519	95,537,822	Del	SD	
11571.p1	2	89,416,533	90,109,382	Del	SD	
11452.p1	2	89,629,573	90,139,880	Dup	SD	

sample	hg19 coordinates (chr, start, stop)			state	type	genes
11660.p1	2	110,656,305	111,225,831	Del	SD	
11043.p1	2	110,663,456	111,223,204	Del	SD	
11948.p1	2	112,526,867	112,560,083	Dup	SD	
11109.p1	5	666,084	843,839	Dup	SD	
12335.p1	5	68,800,071	68,862,452	Dup	SD	
11184.p1	5	69,328,141	69,372,398	Del	SD	
11523.p1	5	69,361,791	69,372,398	Dup	SD	
11184.p1	5	70,203,560	70,266,295	Del	SD	
11523.p1	5	70,234,665	70,270,120	Dup	SD	
12641.p1	6	32,486,332	32,630,025	Dup	SD	
13409.p1	6	74,201,956	74,310,164	Dup	SD	
13494.p1	7	71,868,235	72,338,613	Dup	SD	
13532.p1	7	143,140,545	143,658,017	Dup	SD	
11291.p1	7	143,555,916	143,559,606	Dup	SD	
13031.p1	7	143,929,003	144,072,768	Dup	SD	
12444.p1	7	143,955,788	144,071,982	Dup	SD	
13532.p1	7	144,015,217	144,071,982	Del	SD	
11788.p1	9	84,302,248	84,610,116	Del	SD	
13031.p1	9	84,543,428	84,605,401	Del	SD	
11262.p1	9	84,545,014	85,597,697	Dup	SD	
11141.p1	9	117,068,799	117,104,405	Dup	SD	
11660.p1	9	117,072,828	117,094,209	Dup	SD	
11257.p1	9	117,085,413	117,093,954	Dup	SD	
12641.p1	9	117,085,413	117,095,414	Dup	SD	
11093.p1	9	117,085,413	117,094,209	Dup	SD	
11083.p1	9	117,085,413	117,094,209	Dup	SD	
11452.p1	9	117,085,413	117,094,209	Dup	SD	
11472.p1	9	117,085,413	117,094,209	Dup	SD	
11425.p1	9	117,085,942	117,093,954	Dup	SD	
12430.p1	9	117,085,942	117,093,954	Dup	SD	
11069.p1	9	117,085,942	117,093,954	Dup	SD	
12118.p1	9	117,086,297	117,093,142	Dup	SD	
11375.p1	9	117,086,297	117,093,142	Dup	SD	
13048.p1	9	117,086,297	117,092,856	Dup	SD	
11653.p1	9	135,932,152	135,974,149	Dup	SD	
13116.p1	9	135,933,200	135,977,149	Del	SD	
13409.p1	9	136,215,098	136,218,997	Dup	SD	
11257.p1	10	46,321,362	46,964,019	Dup	SD	
13517.p1	10	47,161,250	47,164,387	Dup	SD	
11257.p1	10	48,428,640	49,393,688	Dup	SD	
12444.p1	11	3,400,267	3,756,554	Dup	SD	
12603.p1	11	60,971,578	61,016,007	Del	SD	
11262.p1	11	60,971,578	61,032,049	Dup	SD	
11414.p1	11	60,971,578	61,017,486	Dup	SD	
11190.p1	11	60,974,968	60,977,430	Dup	SD	
11190.p1	11	60,989,872	60,997,535	Dup	SD	
11190.p1	11	61,008,698	61,016,007	Dup	SD	
12581.p1	12	8,290,735	8,608,738	Dup	SD	
13532.p1	13	54,885,795	58,299,428	Dup	SD	
13116.p1	13	57,722,008	57,736,839	Dup	SD	
11141.p1	14	19,582,970	20,389,737	Dup	SD	

sample	hg19 coordinates (chr, start, stop)			state	type	genes
11928.p1	14	20,215,586	20,389,737	Dup	SD	
13415.p1	14	20,215,586	20,483,352	Dup	SD	
11291.p1	14	73,985,726	74,059,120	Dup	SD	
11556.p1	14	105,964,953	106,234,580	Dup	SD	
11291.p1	14	105,964,953	106,234,580	Dup	SD	
11523.p1	14	106,329,088	106,376,628	Del	SD + SR	
11510.p1	14	106,231,942	107,114,522	Del	SD	
11141.p1	15	22,073,120	22,490,341	Dup	SD	
11193.p1	15	22,368,575	22,483,611	Dup	SD	
11947.p1	15	30,092,848	30,919,152	Del	SD	
12744.p1	15	30,659,620	30,922,976	Del	SD	
11947.p1	15	32,323,100	32,465,110	Del	SD	
12603.p1	15	43,826,871	44,046,100	Dup	SD	
11109.p1	15	43,904,574	44,053,729	Del	SD	
13557.p1	16	28,330,315	28,489,198	Dup	SD	
11224.p1	16	28,330,315	28,474,490	Del	SD	
11190.p1	16	28,394,438	28,474,490	Del	SD	
11218.p1	16	28,401,840	28,411,990	Dup	SD	
13557.p1	16	28,620,028	28,746,834	Dup	SD	
11224.p1	16	28,711,197	28,746,834	Del	SD	
11190.p1	16	28,723,007	28,743,513	Del	SD	
13177.p1	16	70,162,686	70,182,456	Dup	SD	
11141.p1	17	18,286,584	18,498,828	Dup	SD	
13207.p1	17	18,370,032	18,487,151	Dup	SD	
11141.p1	17	20,200,273	20,799,333	Dup	SD	
13207.p1	17	20,209,335	20,363,756	Dup	SD	
11141.p1	17	25,958,291	26,088,257	Dup	SD	
13207.p1	17	25,965,288	26,091,170	Dup	SD	
11193.p1	17	29,556,852	29,580,018	Dup	SD	
11093.p1	17	36,294,031	36,347,081	Dup	SD	
13557.p1	17	44,632,650	44,788,485	Dup	SD	
11722.p1	17	44,632,896	44,782,220	Dup	SD	
11480.p1	17	44,707,776	44,771,287	Dup	SD	
12073.p1	17	44,707,776	44,770,434	Dup	SD	
11834.p1	17	44,714,741	44,770,434	Dup	SD	
13177.p1	17	44,718,008	44,751,979	Dup	SD	
11518.p1	17	45,517,796	45,681,415	Dup	SD	
11895.p1	17	45,608,666	45,681,415	Dup	SD	
11571.p1	17	61,914,554	62,006,835	Del	SD	
11064.p1	17	61,914,800	62,006,835	Del	SD	
11257.p1	17	61,915,341	62,006,684	Del	SD	
11948.p1	17	61,915,341	62,006,835	Del	SD	
12933.p1	18	44,497,284	44,580,809	Dup	SD	
11184.p1	19	7,032,323	7,056,914	Dup	SD	
11459.p1	19	7,032,323	7,056,914	Dup	SD	
12603.p1	19	7,037,871	7,051,633	Dup	SD	
11141.p1	19	33,464,331	33,503,630	Dup	SD	
11193.p1	19	33,467,336	33,502,709	Dup	SD	
11948.p1	19	43,097,692	43,858,145	Dup	SD	
12565.p1	19	43,098,916	43,857,918	Dup	SD	
11055.p1	19	43,228,133	43,858,145	Del	SD	

sample	hg19 coordinates (chr, start, stop)			state	type	genes
11141.p1	19	43,864,416	43,969,723	Del	SD	
11246.p1	19	55,177,849	55,247,339	Dup	SD	
11459.p1	19	55,240,958	55,294,475	Dup	SD	
11246.p1	19	55,295,088	55,341,435	Dup	SD	
11006.p1	19	55,299,350	55,317,699	Dup	SD	
13530.p1	19	55,299,350	55,336,533	Dup	SD	
11398.p1	19	55,301,177	55,320,338	Del	SD	
13593.p1	19	55,309,063	55,317,699	Del	SD	
11013.p1	19	55,309,063	55,317,699	Del	SD	
11141.p1	19	55,309,063	55,316,532	Del	SD	
12565.p1	19	55,309,063	55,317,699	Del	SD	
13335.p1	19	55,309,063	55,325,197	Del	SD	
12581.p1	19	55,315,345	55,317,699	Dup	SD	
11257.p1	22	16,282,477	17,489,004	Dup	SD	
13533.p1	22	16,287,253	17,444,719	Dup	SD	

Table S4: SNP calls in ASD probands and sensitivity:

sampleID	chr	start (hg19)	stop (hg19)	state	i1M Probe Count	Exome Probe Count	Detected by Exome?
12118.p1	13	39,420,663	39,424,878	del	5	3	No
11599.p1	4	73,508	142,550	dup	20	3	No
11504.p1	7	100,968,363	101,127,455	dup	48	3	No
11498.p1	3	155,481,097	155,509,663	dup	19	3	No
11257.p1	3	155,481,097	155,518,835	dup	21	3	No
11711.p1	1	170,917,459	170,937,400	del	13	4	No
11479.p1	20	47,246,127	47,251,687	del	11	4	No
11414.p1	7	64,621,664	65,081,242	del	140	4	No
11093.p1	4	57,538	127,452	dup	22	4	No
11013.p1	7	124,503,189	124,556,473	del	17	4	No
11948.p1	9	137,292,505	137,315,293	del	16	5	No
11425.p1	1	115,399,741	115,417,093	del	9	5	No
11587.p1	8	17,819,812	17,830,005	del	8	6	No
11526.p1	17	909,998	923,916	dup	6	6	No
11948.p1	2	241,703,960	241,713,646	del	6	7	No
11523.p1	16	87,446,053	87,461,969	del	7	8	No
12212.p1	12	53,573,903	53,586,822	del	14	10	No
11948.p1	19	3,196,667	3,207,646	del	12	10	No
11928.p1	19	1,220,004	1,235,071	del	8	10	No
11069.p1	17	5,418,799	5,462,805	del	26	10	No
13031.p1	1	103,339,272	103,376,862	del	30	13	No
11948.p1	20	42,293,880	42,350,811	del	24	13	No
12130.p1	14	105,173,211	105,180,565	del	5	14	No
11472.p1	7	4,823,971	4,841,349	del	9	14	No
11948.p1	19	1,207,204	1,245,700	del	27	19	No
11472.p1	8	145,654,794	145,675,491	del	24	23	No
12378.p1	10	82,100,428	82,112,873	dup	7	3	Yes
11827.p1	13	21,728,320	21,732,348	dup	9	3	Yes
11711.p1	2	160,540,261	160,604,936	dup	22	3	Yes
11556.p1	15	89,784,681	89,804,111	del	11	3	Yes
11545.p1	10	82,100,428	82,112,488	dup	6	3	Yes
11526.p1	19	43,991,980	44,001,379	del	7	3	Yes
11504.p1	8	146,023,923	146,031,702	del	4	3	Yes
12444.p1	5	32,107,084	32,167,220	dup	38	4	Yes
12114.p1	5	32,107,084	32,159,517	dup	37	4	Yes
11863.p1	2	38,955,977	38,971,095	dup	14	4	Yes
11788.p1	5	32,107,084	32,167,220	dup	38	4	Yes
11722.p1	2	38,955,977	38,971,095	dup	14	4	Yes
11696.p1	10	54,524,658	54,536,551	del	25	4	Yes
11653.p1	2	38,955,977	38,971,095	dup	14	4	Yes
11587.p1	2	38,955,977	38,964,531	dup	8	4	Yes
11526.p1	16	75,539,436	75,577,559	dup	34	4	Yes
11526.p1	12	15,063,995	15,074,313	dup	9	4	Yes
11480.p1	5	32,107,084	32,169,547	dup	39	4	Yes
11472.p1	16	75,539,436	75,579,233	dup	35	4	Yes
11469.p1	7	150,553,475	150,560,322	del	14	4	Yes
11246.p1	2	38,955,977	38,965,076	dup	12	4	Yes
11184.p1	16	81,181,180	81,187,852	del	16	4	Yes

sampleID	chr	start (hg19)	stop (hg19)	state	i1M Probe Count	Exome Probe Count	Detected by Exome?
11109.p1	5	32,107,084	32,159,517	dup	37	4	Yes
11056.p1	5	78,377,334	78,389,912	dup	6	4	Yes
11056.p1	5	32,107,084	32,167,220	dup	38	4	Yes
12810.p1	11	32,699,987	32,815,580	del	28	5	Yes
12565.p1	3	151,511,085	151,561,598	del	28	5	Yes
11834.p1	13	114,513,673	114,530,395	dup	15	5	Yes
11526.p1	7	142,827,954	142,889,936	del	24	5	Yes
11472.p1	2	44,519,142	44,545,576	dup	14	5	Yes
11469.p1	5	112,916,398	112,945,992	dup	20	5	Yes
11375.p1	5	157,073,947	157,118,579	del	17	5	Yes
11262.p1	3	151,512,694	151,554,749	del	26	5	Yes
11218.p1	7	142,827,954	142,889,936	del	24	5	Yes
13031.p1	9	135,942,204	135,957,452	del	10	6	Yes
12810.p1	1	87,028,669	87,038,695	del	9	6	Yes
12667.p1	7	33,131,729	33,187,279	dup	27	6	Yes
12212.p1	6	107,493,418	107,667,248	del	62	6	Yes
12130.p1	8	15,948,235	16,021,468	del	40	6	Yes
12130.p1	5	112,911,165	112,947,050	dup	23	6	Yes
12118.p1	11	4,406,483	4,456,562	dup	45	6	Yes
11346.p1	7	33,127,539	33,187,279	dup	28	6	Yes
11224.p1	4	5,735,303	5,773,055	dup	48	6	Yes
11083.p1	13	50,124,621	50,185,204	dup	34	6	Yes
11056.p1	16	29,879,215	29,885,866	dup	7	6	Yes
11043.p1	7	33,131,729	33,187,279	dup	27	6	Yes
11964.p1	11	14,875,154	14,903,636	dup	14	7	Yes
11023.p1	2	209,034,715	209,054,928	del	10	7	Yes
12430.p1	11	31,177,108	31,428,202	del	80	8	Yes
12130.p1	1	206,317,334	206,329,651	del	18	8	Yes
11895.p1	1	206,317,334	206,329,651	del	18	8	Yes
11556.p1	8	15,948,235	16,029,094	del	44	8	Yes
11141.p1	8	13,357,501	14,660,575	dup	766	8	Yes
12667.p1	1	185,103,113	185,123,630	dup	16	9	Yes
11964.p1	16	84,433,034	84,470,158	del	48	9	Yes
11715.p1	1	185,103,113	185,123,630	dup	16	9	Yes
11707.p1	1	207,403,840	207,533,155	dup	103	9	Yes
11459.p1	6	88,317,583	88,367,635	del	21	9	Yes
11707.p1	17	4,306,099	4,422,090	dup	76	13	Yes
11257.p1	13	115,007,056	115,045,729	dup	18	13	Yes
11571.p1	1	2,524,205	2,539,400	del	17	14	Yes
11364.p1	22	35,711,098	35,748,208	dup	21	14	Yes
11722.p1	3	100,339,588	100,443,732	dup	41	17	Yes
11303.p1	3	100,335,088	100,443,732	dup	42	17	Yes
11653.p1	22	41,577,964	41,627,073	dup	24	18	Yes
11013.p1	20	7,549,585	8,317,018	dup	363	18	Yes
11722.p1	7	48,294,575	48,417,856	del	77	20	Yes
11190.p1	4	108,493,334	108,876,094	dup	133	21	Yes
11696.p1	7	16,839,086	17,746,655	dup	512	22	Yes
11696.p1	3	37,282,070	37,457,208	del	78	22	Yes
11510.p1	14	105,564,734	105,623,612	dup	27	23	Yes
11064.p1	2	33,733,554	34,505,480	dup	384	24	Yes
12810.p1	22	32,530,256	32,703,072	dup	88	26	Yes

sampleID	chr	start (hg19)	stop (hg19)	state	i1M Probe Count	Exome Probe Count	Detected by Exome?
12581.p1	12	112,181,078	112,315,172	dup	74	29	Yes
12444.p1	11	3,624,237	3,750,628	dup	84	32	Yes
11023.p1	2	198,295,171	198,534,514	dup	124	37	Yes
11480.p1	6	25,969,958	26,267,800	dup	237	39	Yes
11947.p1	22	40,720,027	40,893,364	dup	99	46	Yes
12581.p1	9	140,680,073	141,072,194	del	173	60	Yes
11398.p1	7	11,203,796	12,473,521	dup	738	65	Yes
11928.p1	15	30,936,285	32,451,488	dup	551	76	Yes
12118.p1	10	133,729,749	134,343,062	dup	315	82	Yes
11834.p1	16	21,963,364	22,449,883	dup	173	94	Yes

Table S5: Genotyping Correlation with Whole-Genome Absolute Copy Number

Location	Genes	mrsFAST r^2	BWA r^2	Median Copy Number
chr1:104230039-104238912	*AMY1A*	0.98		8.11
chr1:110222301-110242933	*GSTM2,GSTM1*	0.99	0.86	3.12
chr1:144951760-145076079	*PDE4DIP*	0.99	0.61	6.66
chr1:145209110-145285912	*NOTCH2NL*	0.92	0.17	8.73
chr1:145293370-145368682	*NBPF10*	0.03	0.14	258.86
chr1:196788860-196801319	*CFHR1*	0.88	0.64	2.66
chr1:196825137-196896065	*CFHR4*	0.54	0.50	2.53
chr1:202415009-202496465	*PPP1R12B*	0.99	0.05	2.03
chr1:21766630-21811393	*NBPF3*	0.40	0.01	13.85
chr1:25598980-25656936	*RHD*	0.98	0.87	4.01
chr11:55403116-55451172	*OR4P4,OR4S2,OR4C6*	0.94	0.88	1.04
chr11:61008668-61018915	*PGA5*	0.98	0.21	5.98
chr12:11505418-11542473	*PRB1*	0.46	0.83	4.39
chr14:20202606-20420924	*OR4Q3,OR4M1,OR4N2,OR4K2,OR4K5*	0.96	0.93	3.75
chr14:74035771-74042359	*ACOT2*	0.98	0.08	3.00
chr15:22304656-22588026	*OR4N4*	0.97	0.61	4.23
chr15:30605924-30675622	*CHRFAM7A*	0.84	0.07	3.92
chr16:14766404-14788526	*PLA2G10*	0.22	0.11	8.57
chr16:15068832-15131552	*PDXDC1*	0.93	0.36	4.73
chr16:22524883-22547861	*LOC100132247*	0.08		48.81
chr16:32684848-32688053	*TP53TG3B,TP53TG3*	0.89		8.28
chr16:70148739-70196427	*PDPR*	0.96	0.91	4.73
chr17:18362101-18425291	*LGALS9C*	0.96	0.80	6.74
chr17:20353175-20370848	*LGALS9B*	0.93	0.04	6.57
chr17:34431219-34433014	*CCL4*	0.98	0.01	5.45
chr17:34522268-34524156	*CCL3L1*	0.95		6.88
chr17:34746118-34808091	*TBC1D3H,TBC1D3G,TBC1D3C*	0.75		47.70
chr17:36337711-36348666	*TBC1D3*	0.92		49.01
chr17:39506594-39525574	*KRT33A,KRT33B*	0.67	0.61	2.19
chr17:39531902-39536694	*KRT34*	0.60	0.58	2.46
chr17:39738532-39743147	*KRT14*	0.92	0.05	4.35
chr17:44165239-44800231	*KIAA1267,LRRC37A,ARL17A,LRRC37A2,NSF*	0.97	0.35	3.95
chr17:45608443-45700642	*NPEPPS*	0.79	0.07	8.42
chr17:62850487-62914903	*LRRC37A3*	0.83	0.24	10.79
chr19:49535129-49536495	*CGB2*	0.43	0.01	19.19
chr19:54799854-54804238	*LILRA3*	0.85	0.76	7.05
chr2:97779232-97915915	*ANKRD36*	0.06	0.11	18.23
chr22:16256331-16287937	*POTEH*	0.58		18.93
chr22:23043312-23249272	*IGLL5*	0.90	0.10	2.13
chr22:24376138-24384284	*GSTT1*	0.98	0.44	1.09
chr22:25677318-25911586	*LRP5L*	0.99	0.93	2.96
chr3:197879236-197907728	*FAM157A*	0.03	0.36	8.71
chr3:75786028-75834255	*ZNF717*	0.10	0.03	39.36
chr4:69366860-69554789	*UGT2B17,UGT2B15*	0.91	0.00	2.72
chr4:70127619-70235027	*UGT2B28*	0.91	0.94	5.50
chr5:180377202-180416706	*BTNL3*	0.99	0.97	2.36
chr5:68821588-68854548	*OCLN*	0.92		3.19

Location	Genes	mrsFAST r^2	BWA r^2	Median Copy Number
chr5:69316084-69343660	*SERF1A*	0.77		4.01
chr5:69345316-69374572	*SMN1*	0.95		3.84
chr5:795743-825341	*ZDHHC11*	0.92	0.78	3.98
chr6:257332-380527	*DUSP22*	0.99	0.99	4.01
chr6:32455238-32493130	*HLA-DRB5*	0.92	0.09	1.41
chr7:101986192-101996889	*SPDYE6*	0.02		38.21
chr7:102114556-102332921	*POLR2J,SPDYE2,POLR2J3,RASA4,UPK3BL,POLR2J2*	0.95	0.12	10.58
chr7:143223558-143541003	*CTAGE15P,FAM115C*	0.98		3.66
chr7:144052488-144077725	*ARHGEF5*	0.95	0.45	5.29
chr7:43980493-44058748	*UBE2D4,SPDYE1*	0.05	0.06	9.20
chr8:11946846-11973025	*ZNF705D*	0.18		9.32
chr9:141106636-141134172	*FAM157B*	0.00	0.61	9.16
chr9:14510-29739	*WASH1*	0.26		21.12
chr9:33795558-33799229	*PRSS3*	0.96	0.33	5.75
chr9:67926760-67969840	*ANKRD20A1*	0.66		28.15

Table S6: Accuracy of absolute copy number prediction

location	genes	# correct HapMap genotypes (of 7; from Campbell et al 2011)
chr1:25592663-25663607	*RHD*	7
chr1:110222301-110242933	*GSTM1,GSTM2*	7
chr1:144959523-145081011	*PDE4DIP*	5
chr1:196738897-196801697	*CFHR3,CFHR1*	4
chr1:196825137-196896065	*CFHR4*	5
chr1:202389905-202402001	*PPP1R12B*	7
chr1:202415009-202496465	*PPP1R12B*	2
chr2:89160037-89262733	*Ig Light chain locus*	4
chr3:100547342-100670846	*ABI3BP*	6
chr3:151511518-151550270	*AADAC*	7
chr3:189364074-189538586	*TP63*	5
chr4:68793517-68833125	*TMPRSS11A*	6
chr4:69386965-69483317	*UGT2B17,UGT2B15*	7
chr4:70127619-70235027	*UGT2B28*	5
chr4:144921494-145040886	*GYPA,GYPB*	7
chr5:795743-825341	*ZDHHC11*	5
chr5:32107113-32169449	*PDZD2,GOLPH3*	7
chr5:68821588-68854548	*OCLN*	5
chr5:69316084-69343660	*SERF1A*	4
chr5:69345316-69374572	*SMN1*	6
chr5:180377202-180416706	*BTNL3*	6
chr6:257332-380527	*DUSP22*	5
chr6:32455238-32493130	*HLA-DRB5*	6
chr7:143223558-143541003	*FAM115C,CTAGE15P*	4
chr9:115383227-115585827	*KIAA1958,C9orf80,SNX30*	7
chr10:51008386-51114434	*PARG*	7
chr10:135232058-135377386	*MTG1,CYP2E1,SYCE1*	7
chr11:55403116-55451172	*OR4P4,OR4S2,OR4C6*	3
chr12:11505418-11542473	*PRB1*	5
chr14:20202606-20420924	*OR4Q3,OR4M1,OR4N2,OR4K2,OR4K*	6
chr14:88400031-88414591	*GALC*	6
chr15:22304656-22588026	*OR4N4*	4
chr15:30605924-30675622	*CHRFAM7A*	7
chr16:70148739-70196427	*PDPR*	5
chr17:18362101-18425291	*LGALS9C*	5
chr17:34416411-34496071	*CCL3,CCL4,TBC1D3B*	2
chr17:39506594-39525574	*KRT33A,KRT33B*	6
chr17:39531902-39536694	*KRT34*	7
chr19:54724572-54740148	*LILRB3*	4
chr22:22754320-23038160	*PRAME,GGTLC2*	6
chr22:23043312-23249272	*IGLL5*	6
chr22:24347958-24395540	*GSTT1,LOC391322*	5

Table S7: Signal-to-Noise ratios for mrsFAST and BWA calls

Sample	Chrom	Start	Stop	mrsFAST Signal	mrsFAST StdDev	mrsFAST SNR	BWA Signal	BWA StdDev	BWA SNR	mrsFAST SNR improvement
NA18517	16	21396577	21756357	1.927	0.183	10.516	0.373	0.059	6.314	67%
NA19240	12	133659688	133727740	0.881	0.138	6.383	0.212	0.052	4.046	58%
NA15510	7	99507187	99627998	1.670	0.160	10.463	0.284	0.048	5.916	77%
NA19129	6	29910533	30043566	0.894	0.173	5.183	0.377	0.056	6.748	-23%
NA18517	4	68788472	69057034	1.646	0.152	10.841	0.470	0.055	8.520	27%
NA15510	3	19492646	21465556	1.565	0.169	9.240	0.246	0.042	5.833	58%
NA15510	1	155227075	155264543	1.487	0.157	9.496	0.459	0.051	9.025	5%