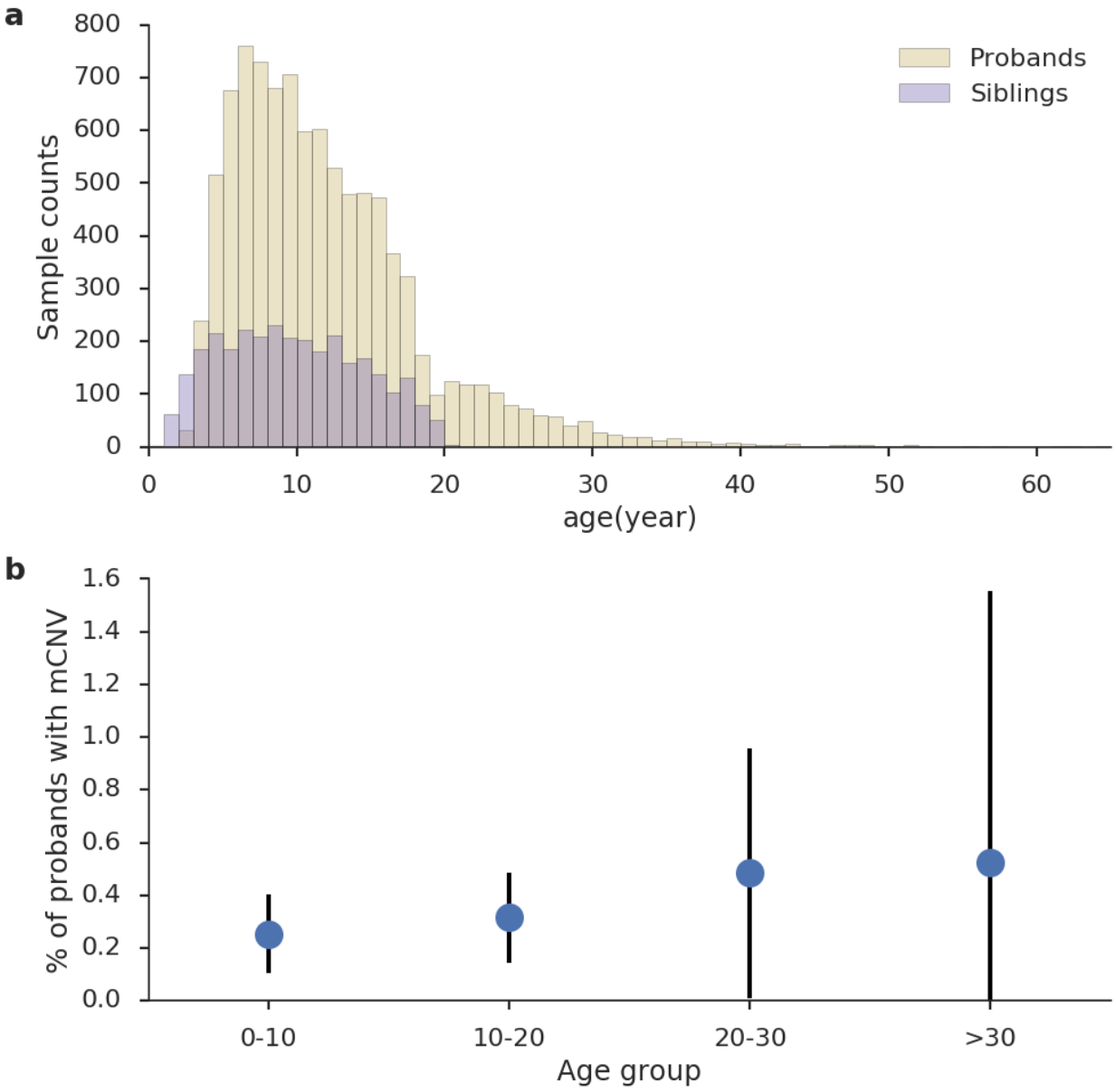
Supplementary Note:

Large mosaic copy number variations confer autism risk

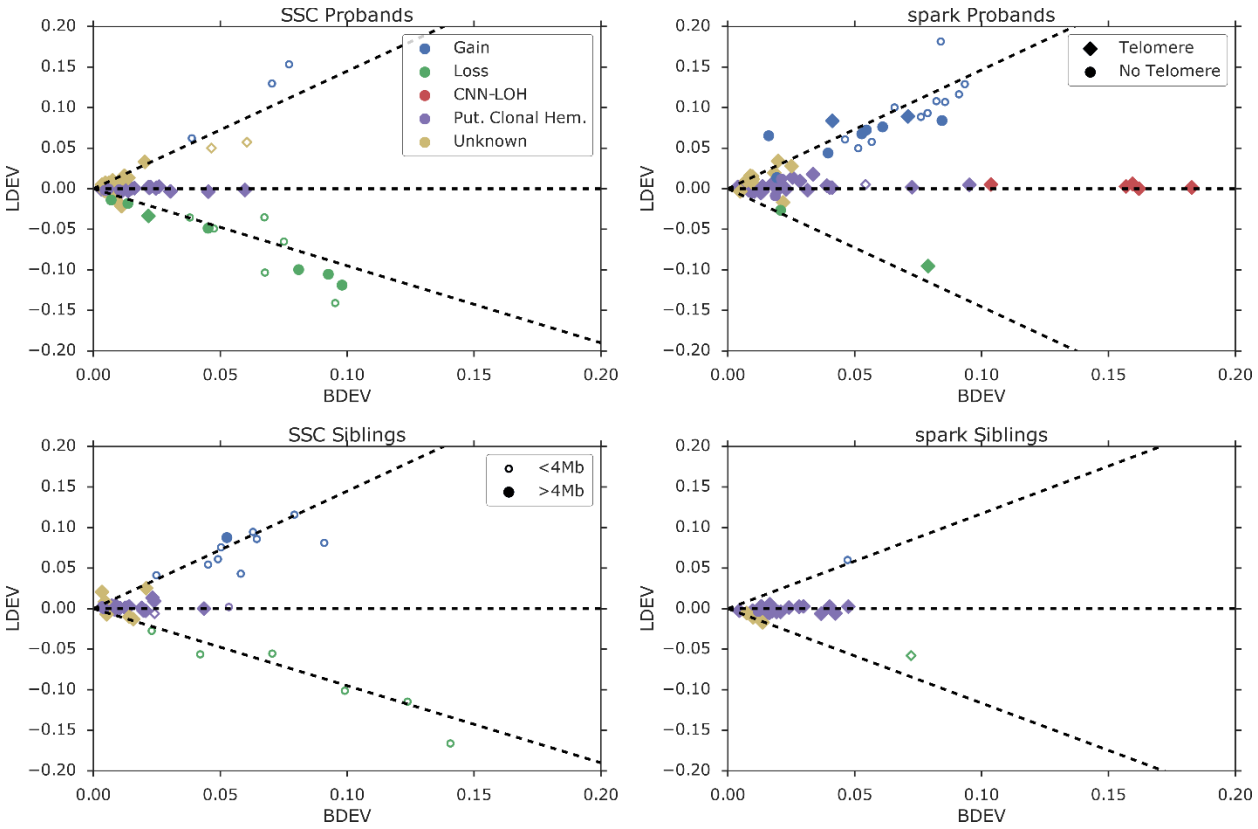
Maxwell A. Sherman[1,2,3,*], Rachel E. Rodin[3,4], Giulio Genovese[3,5,6], Caroline Dias[4,7], Alison R. Barton[2,3], Ronen E. Mukamel[2,3], Bonnie Berger[1,8], Peter J. Park[9,*,§], Christopher A. Walsh[3,4,*,§], Po-Ru Loh[2,3,*,§]

Contents

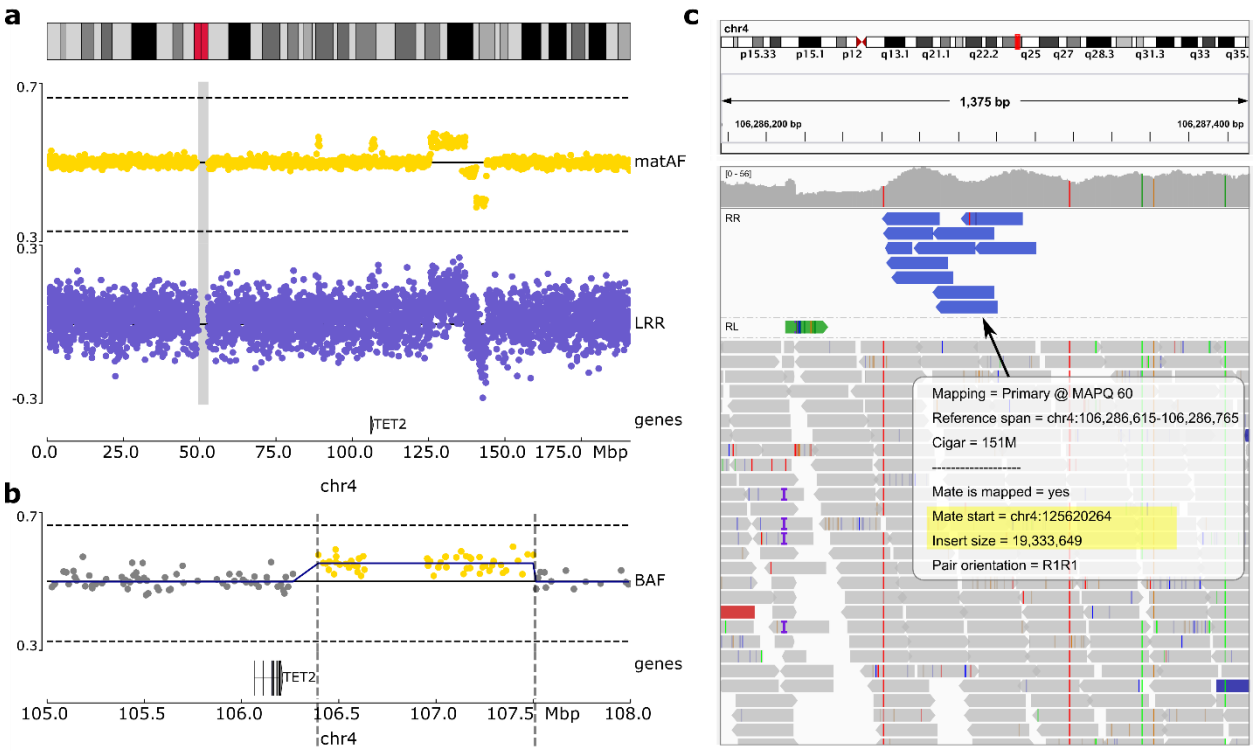
1. Supplementary Figures



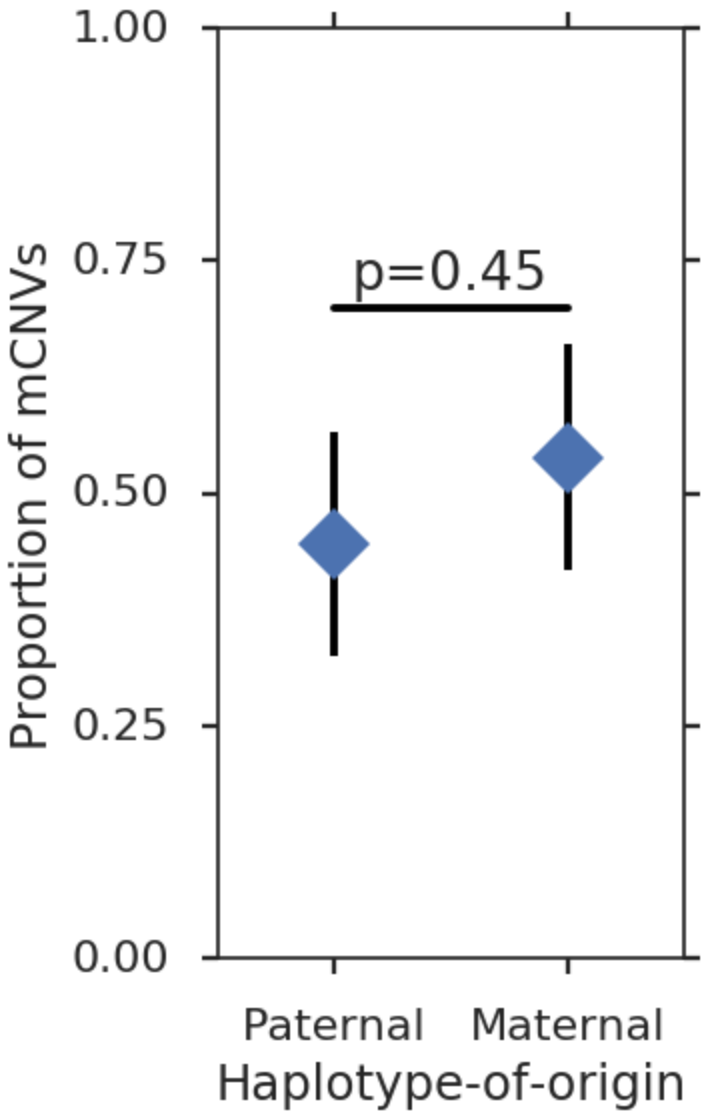
Supplementary Figure 1: **a**, Age distribution of SPARK probands and siblings and **b,** rate of mCNVs in SPARK probands by age. (Probands and siblings in SSC were between age 3-18 at enrollment, and individual-level age information in SSC was not available to us.)



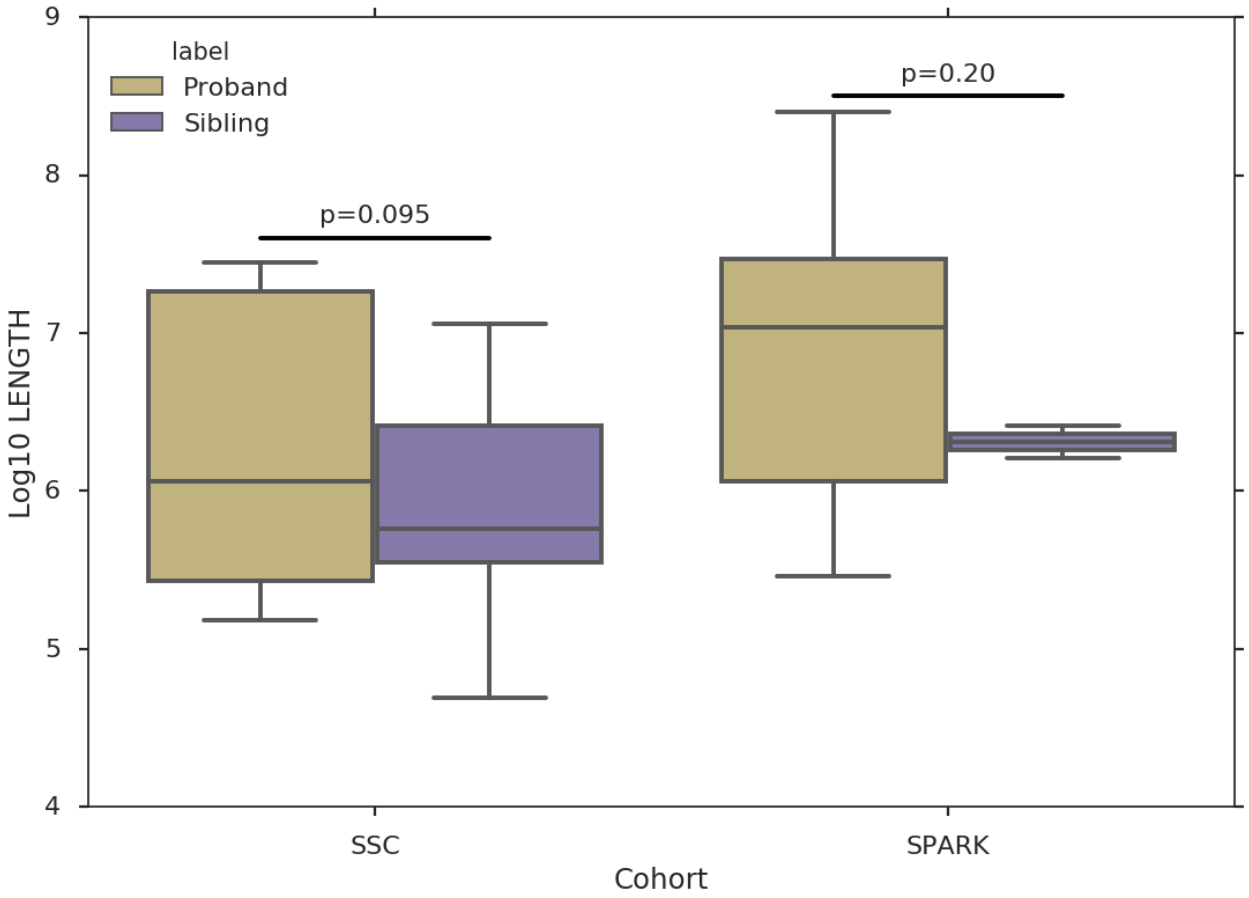
Supplementary Figure 2: Plots of LRR deviation from 0 (LDEV) versus B allele frequency deviation from 0.5 (BDEV) in mosaic events. Gains fall along an upwards diagonal line; losses fall along a downwards diagonal line; and CNN-LOH fall along the horizontal axis. Dashed lines are the expected duplication, deletion, and CNN-LOH trends as inferred by the EM algorithm fit on parental data (Methods). Marker color indicates inferred mosaic copy state. Small, unfilled markers are events <4 Mb and large filled markers are events >4 Mb; circles indicate events which do not extend to telomeres and diamonds indicate those which do extend to telomeres. Event type is indicated by color as listed in the legend of SSC Probands.



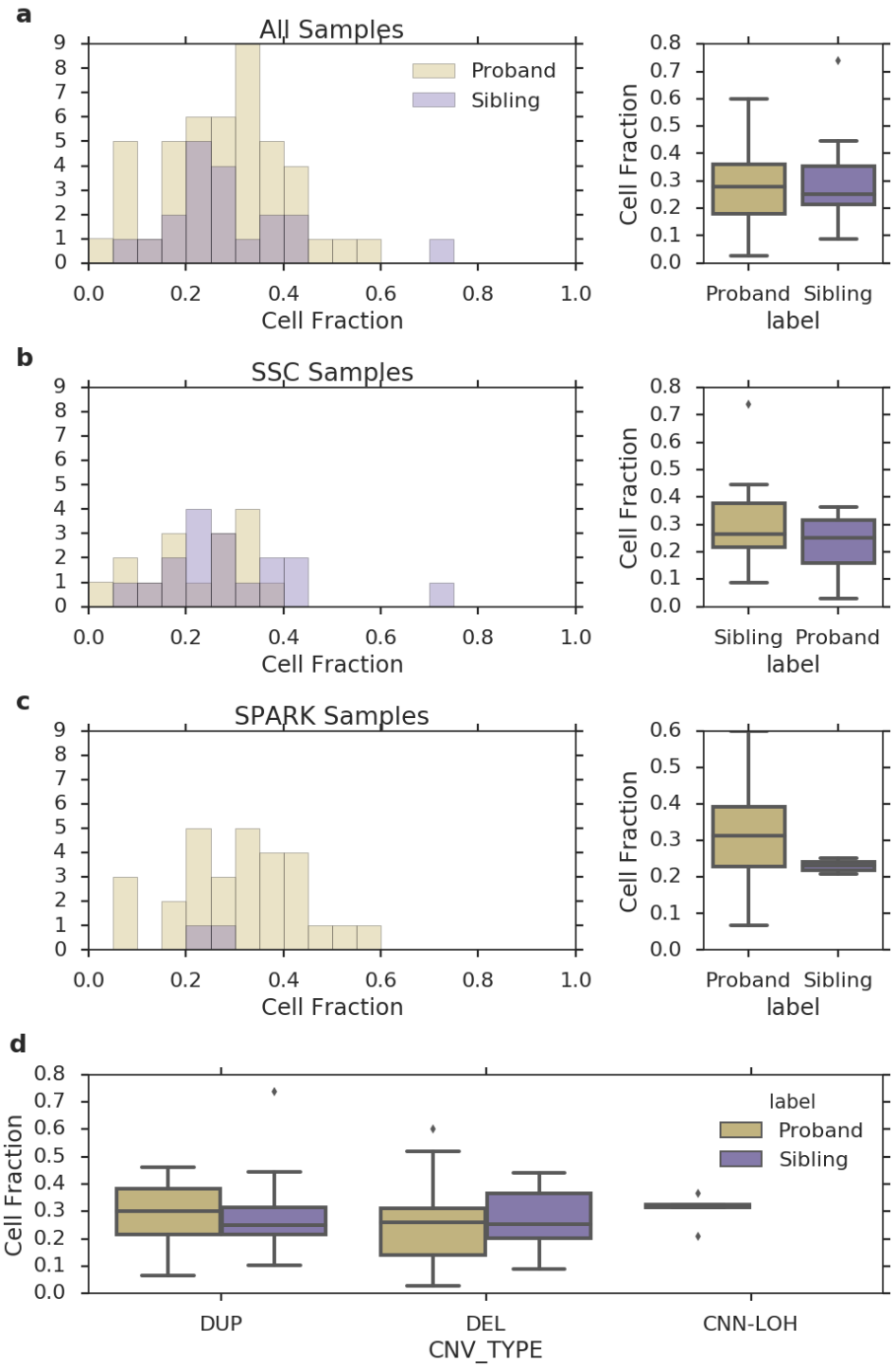
Supplementary Figure 3: SSC sibling 13391.s1 carries a complex rearrangement reminiscent of chromothripsis on chromosome 4. a) Maternal allele frequency (matAF, gold) of heterozygous SNPs and LRR signal (purple) of all array-typed SNPs across chr4 of 13391.s1. Multiple mosaic duplications and deletions are apparent. The location of the gene *TET2*, which has been implicated in myeloproliferative disorders, is as indicated. b) Zoom-in of matAF of heterozygous SNPs in the region around *TET2*. The approximate boundaries of the duplicated segment as determined by MoChA are marked with dashed grey lines. SNPs falling within the duplication are colored gold; SNPs outside the duplication are colored grey. c) IGV plot of left breakpoint of duplicated segment shown in (b). Reads with discordantly mapped mates are colored blue. Mapping information is shown for one representative blue read, demonstrating that it maps to chr4:106,286,615 and its mate maps to the start of the large duplicated segment spanning approximately 4:125610859-137063398 as determined by MoChA. The breakpoint occurs ~85 kb downstream of *TET2*. Reads were aligned to GRCh38 and positions were converted to GRCh37 coordinates via the UCSC liftOver tool.



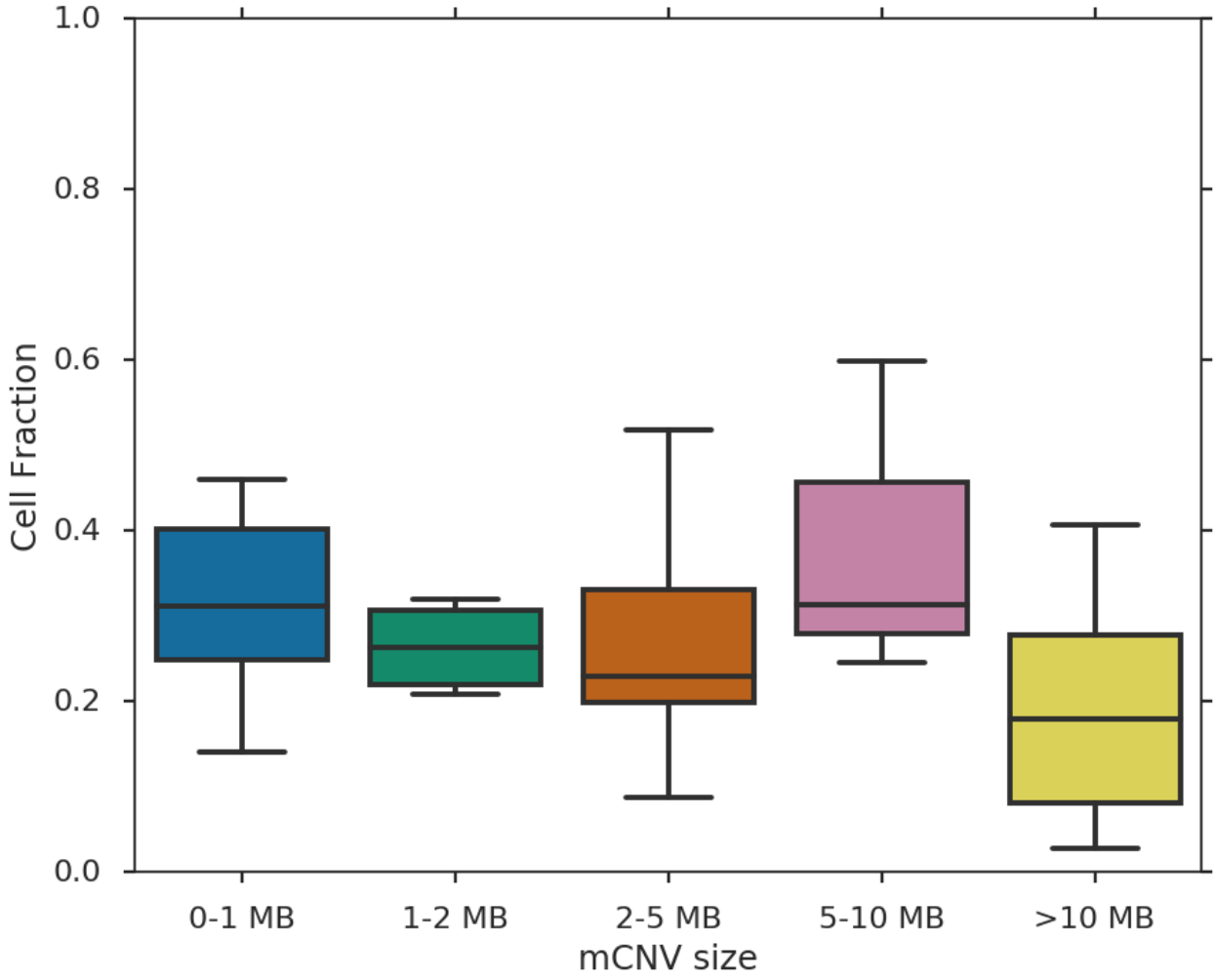
Supplementary Figure 4: Proportion of mCNVs that were located on the paternal haplotype and maternal haplotype, respectively; error bars, 95% CI. For this analysis, CNN-LOH events are considered to be located on the haplotype which is duplicated (i.e. the haplotype that becomes homozygous in the mosaic state). Of the 64 mCNVs detected, 28 were located on the paternal haplotype and 35 on the maternal haplotype. The haplotype of one event could not be established because parental genotypes were unavailable. The p-value is calculated under a binomial model that assumes mCNVs arise with equal probability on either parental haplotype.



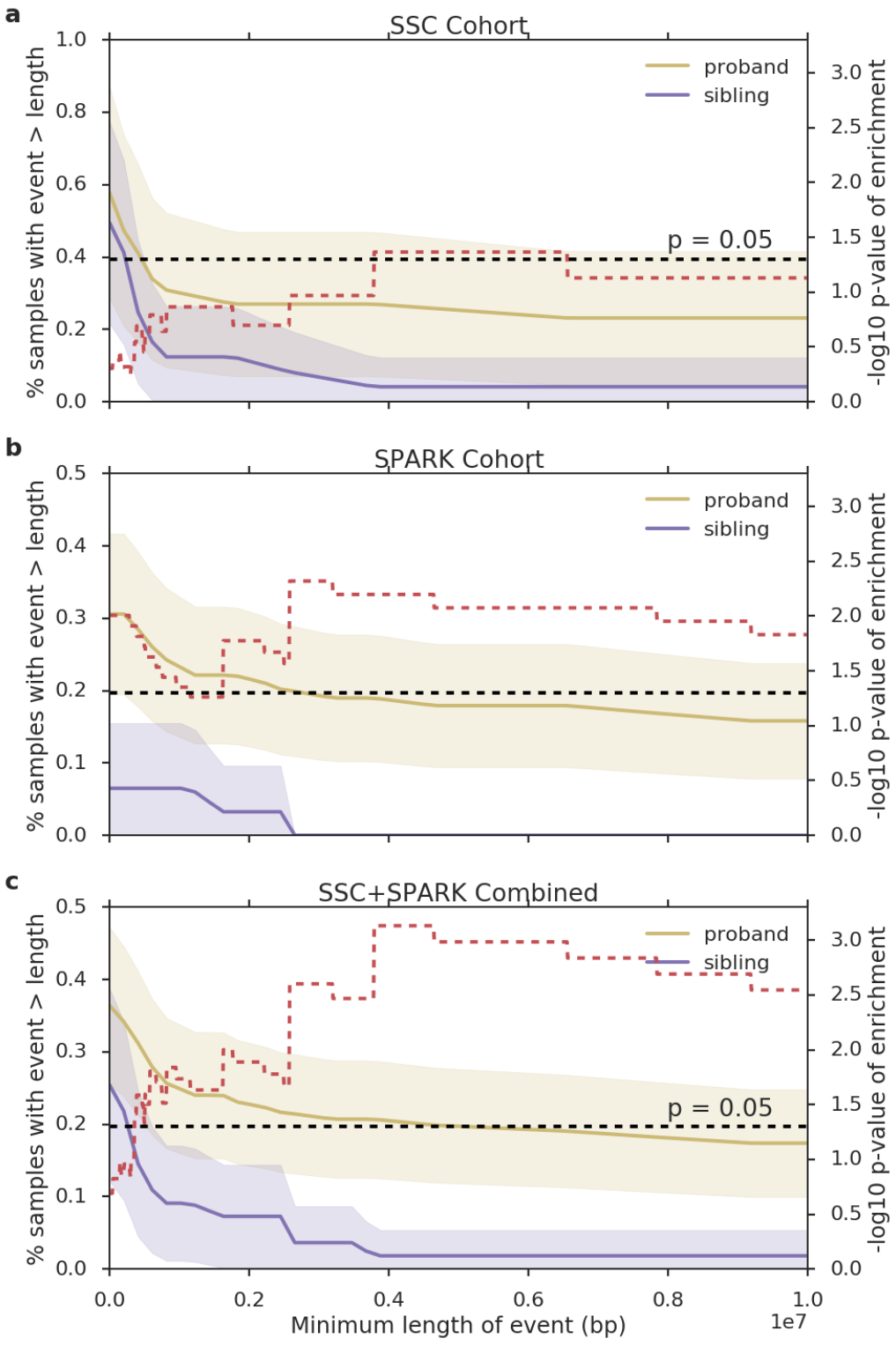
Supplementary Figure 5: Length distribution of mosaic CNVs stratified by cohort.



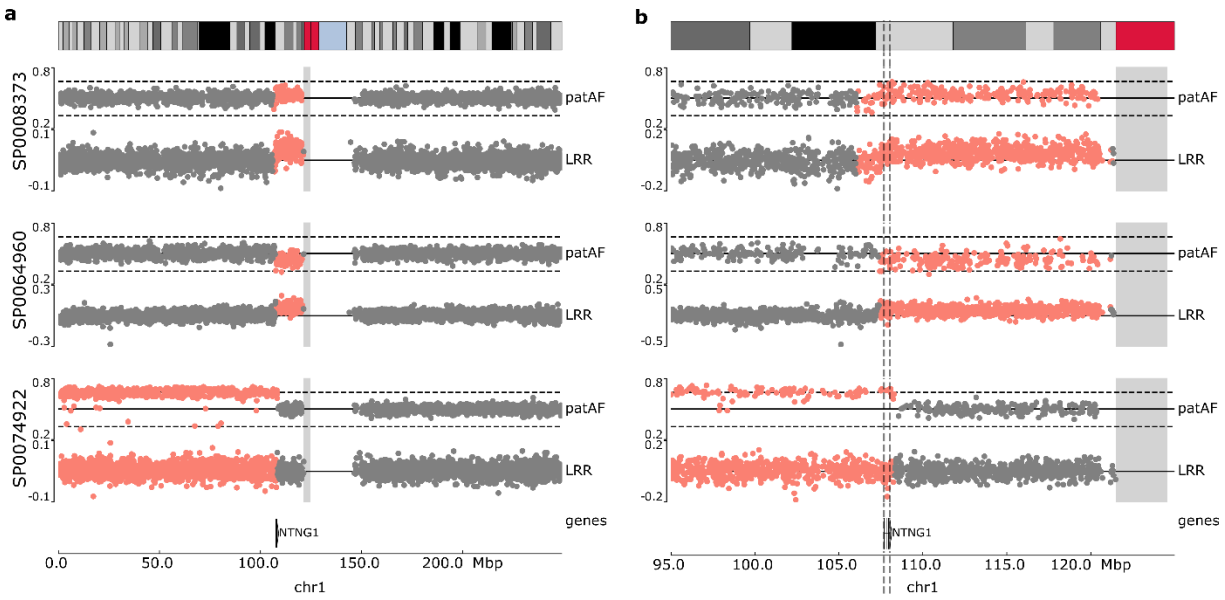
Supplementary Figure 6: **a,** Cell fraction distribution of mCNVs across all events, **b,** events detected in SSC, and **c**, events detected in SPARK. The right column provides a boxplot to provide greater clarity in comparing the distribution across probands and siblings. **d,** Cell fraction distribution stratified by event type. No differences are statistically significant between probands and siblings.



Supplementary Figure 7: Cell fraction distribution of mCNVs stratified by length of event. CNN-LOH events were excluded because we imposed a strict minimum cell fraction of 0.2 for CNN-LOH events in order to filter potential clonal hematopoiesis events



Supplementary Figure 8: **a,** Burden by length in SSC, **b,** SPARK, and **c,** a combined analysis. The dashed red line provides the -log10 p-value (corresponding to the y-axis on the right) of the burden test at a given minimum mCNV length. The black line indicates the P = 0.05 significance level. P-values are not corrected for choice of size threshold.



Supplementary Figure 9: **a**, Paternal allele fraction (patAF) and LRR plots from three individuals with events which either start (SP0008373, SP0064960) or end (SP0074922) immediately adjacent to *NTNG1*. The former two are duplications while the latter is a CNN-LOH. **b,** Zoom in of neighborhood around *NTNG1* showed that all three events encompass *NTNG1* (boundaries marked by dashed grey lines) but do not terminate within *NTNG1*.

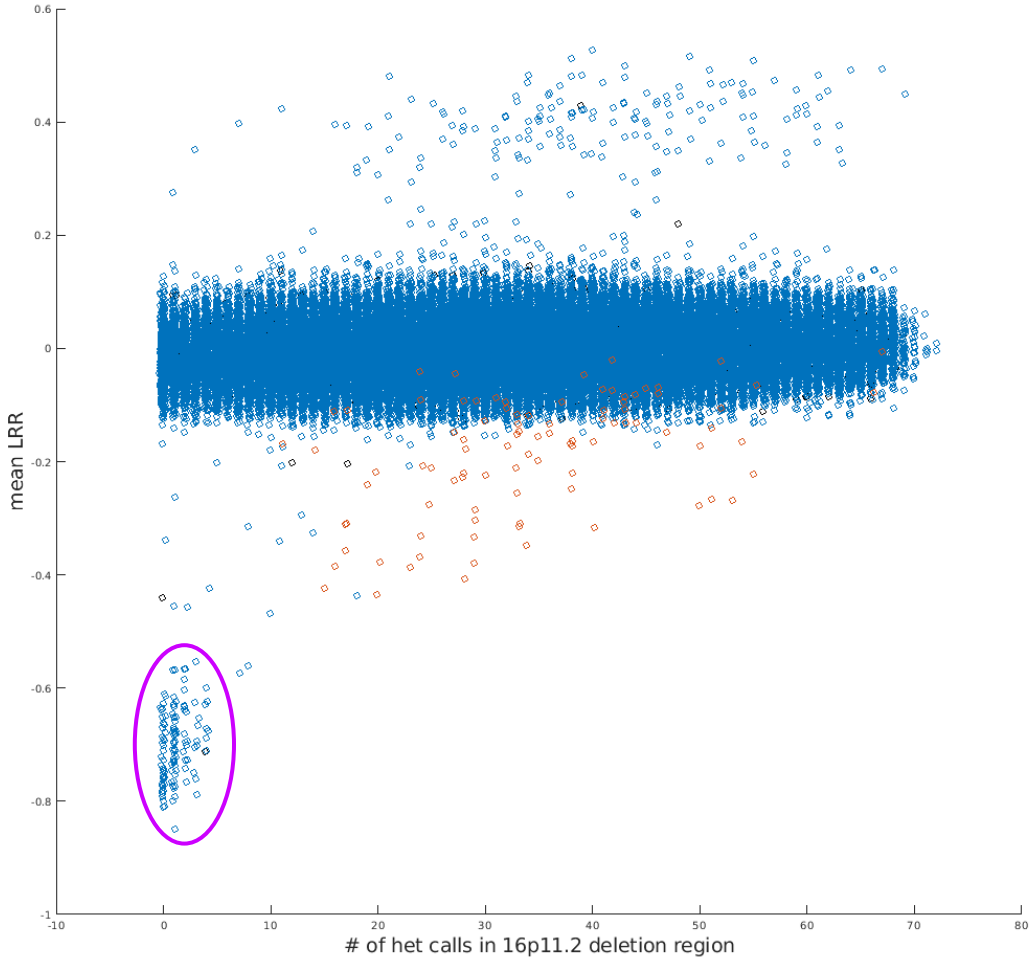
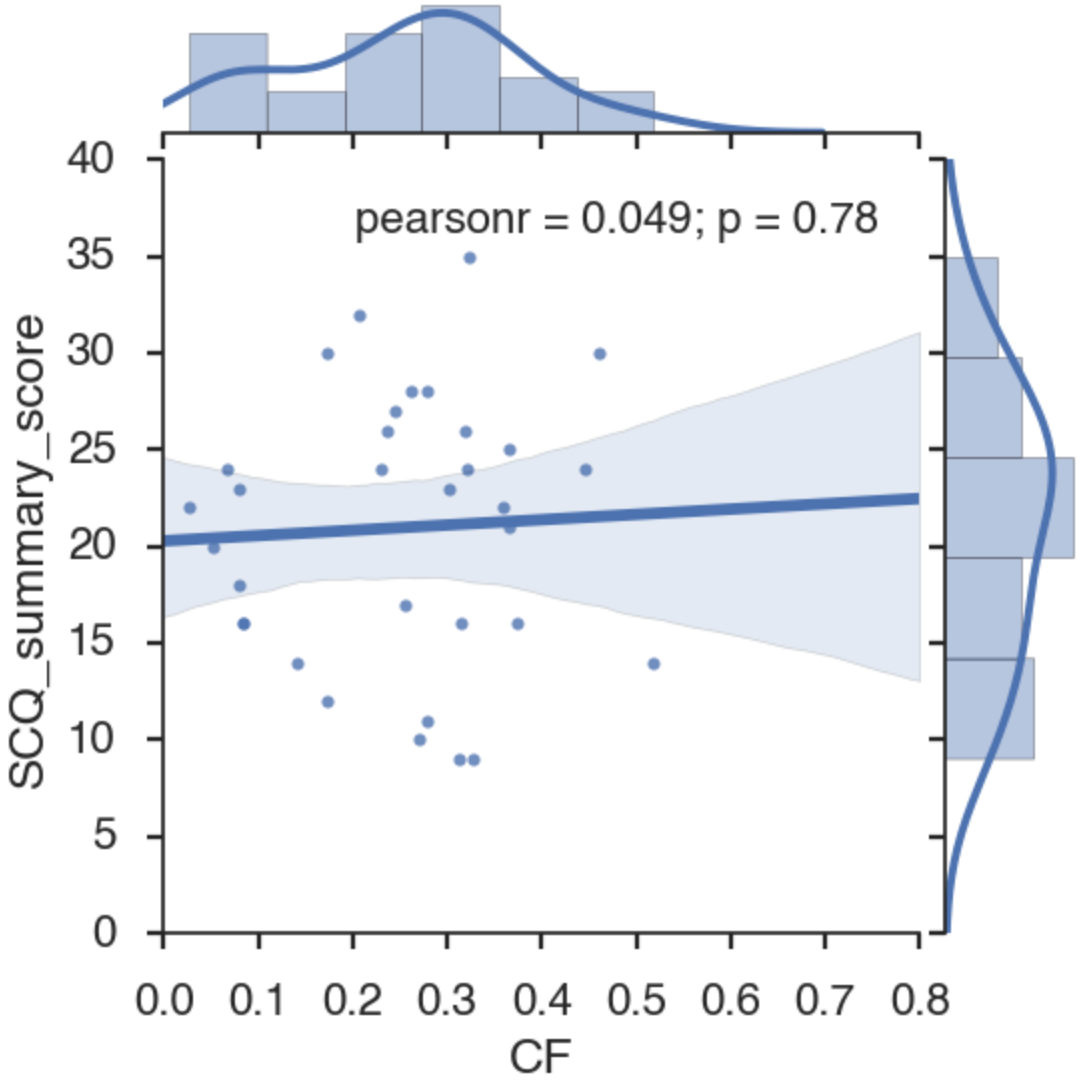
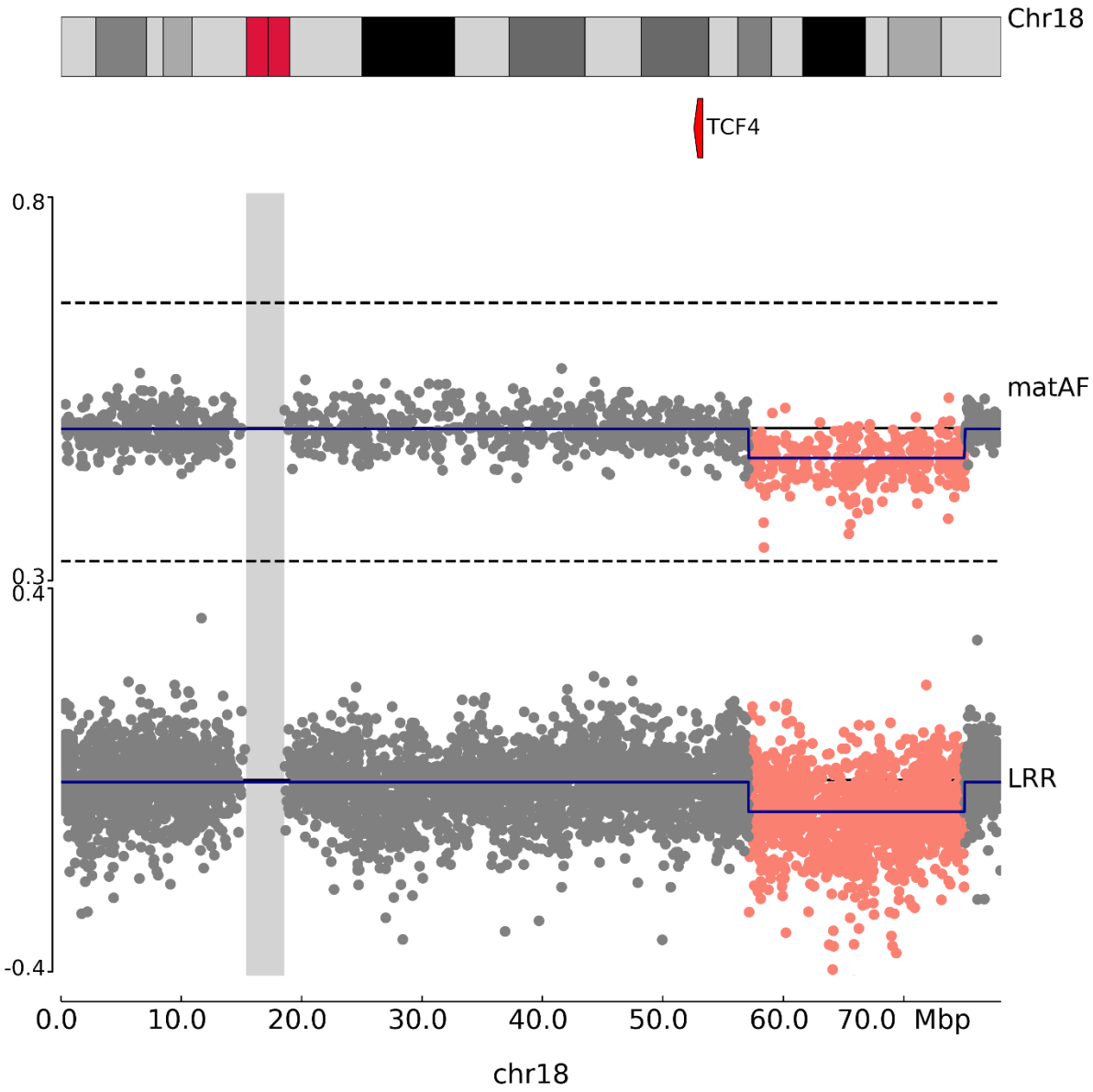


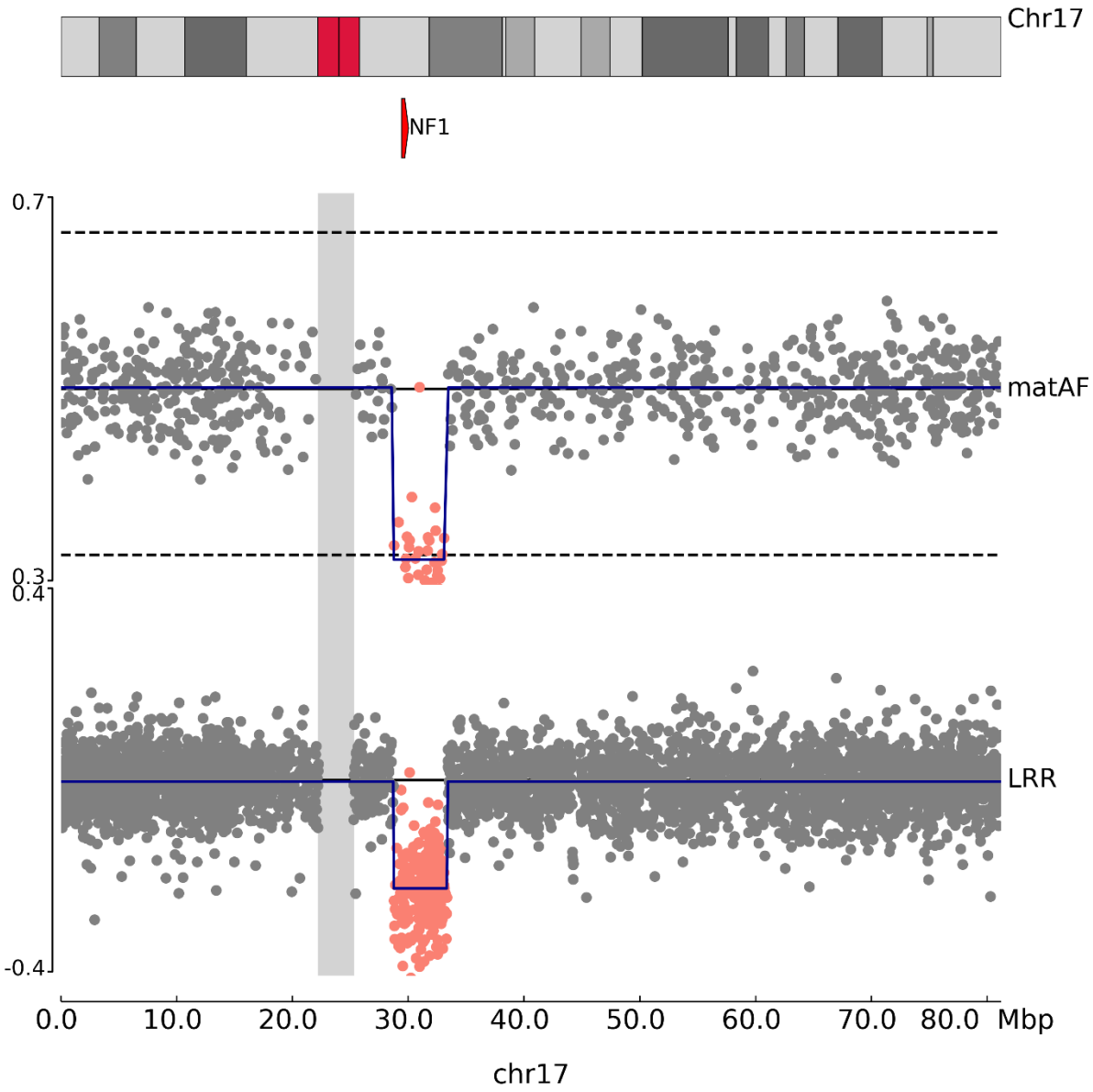
Figure 10: Mean genotyping intensity (LRR) vs. number of heterozygous SNP calls in the 16p11.2 deletion region (chr16:29,655,864-30,195,048, hg19) in UK Biobank samples. Samples with <5 hets and mean LRR <-0.5 were identified as germline 16p11.2 deletion carriers (purple circle). Red markers indicate carriers of mosaic 16p11.2 deletions previously identified[1].



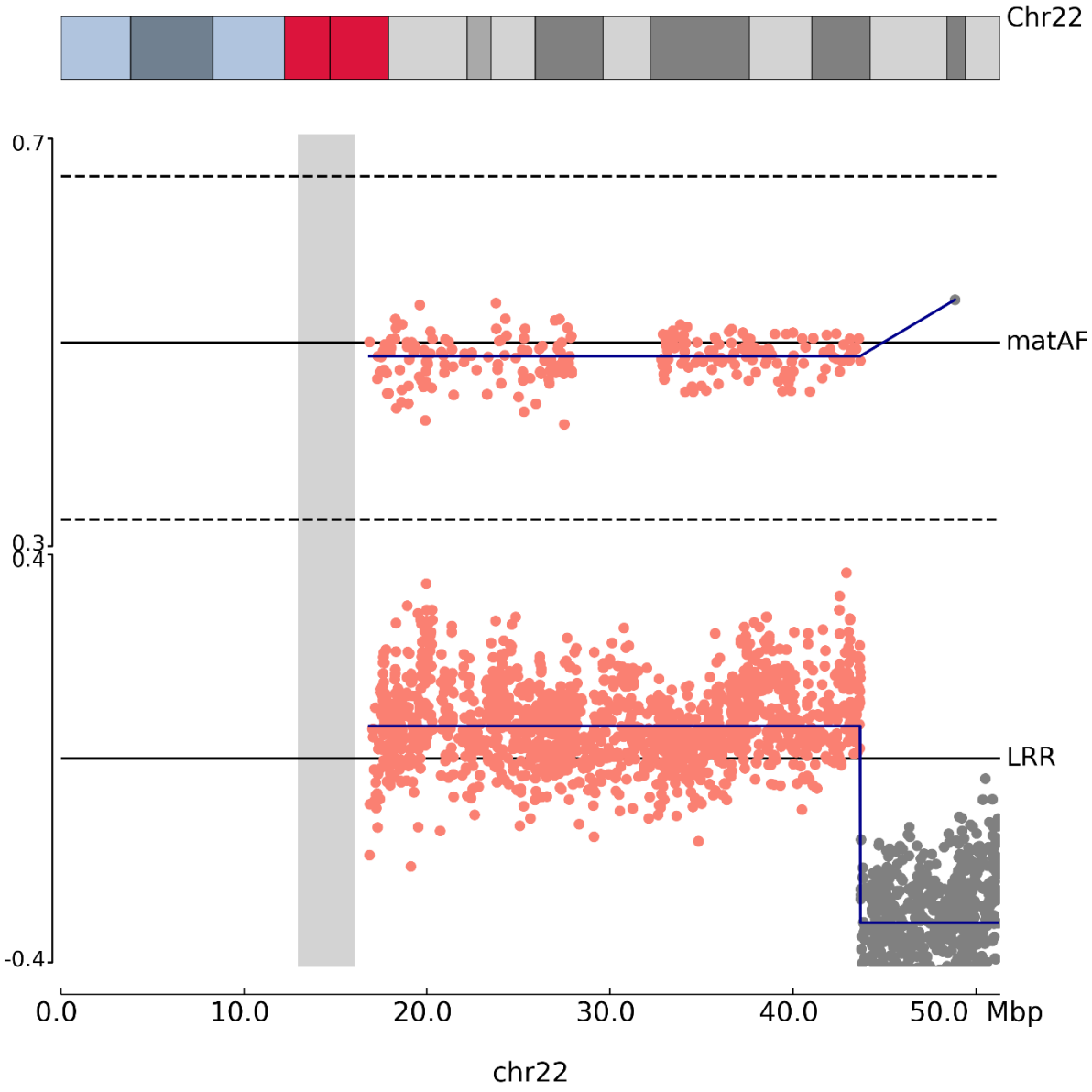
Supplementary Figure 11: Correlation between mosaic event cell fraction in probands and ASD phenotypic severity as quantified by SCQ summary score. Shaded region, 95% CI.



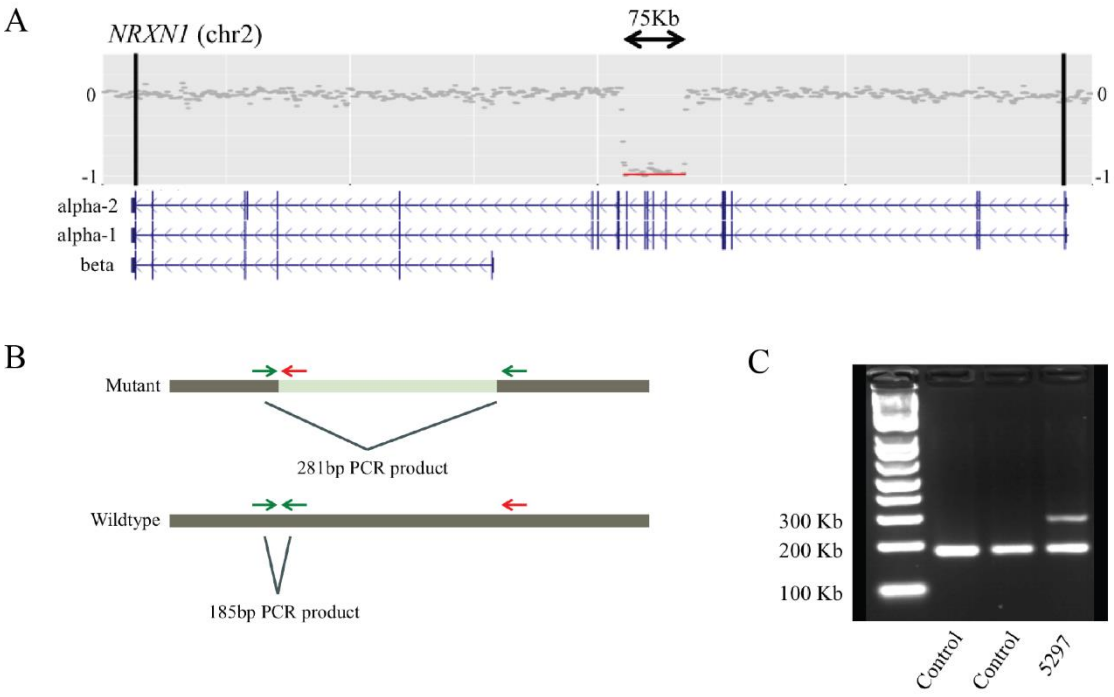
Supplementary Figure 12: Maternal allele fraction (matAF) and LRR plots of the chr18q distal deletion carried by SSC proband 12246.p1. The matAF plot includes only heterozygous SNPs. The boundaries of the event estimated by MoChA are chr18:57,102,326-75,041,151. The location of *TCF4* (chr18:52,942,850-53,068,756) is indicated with a red triangle.



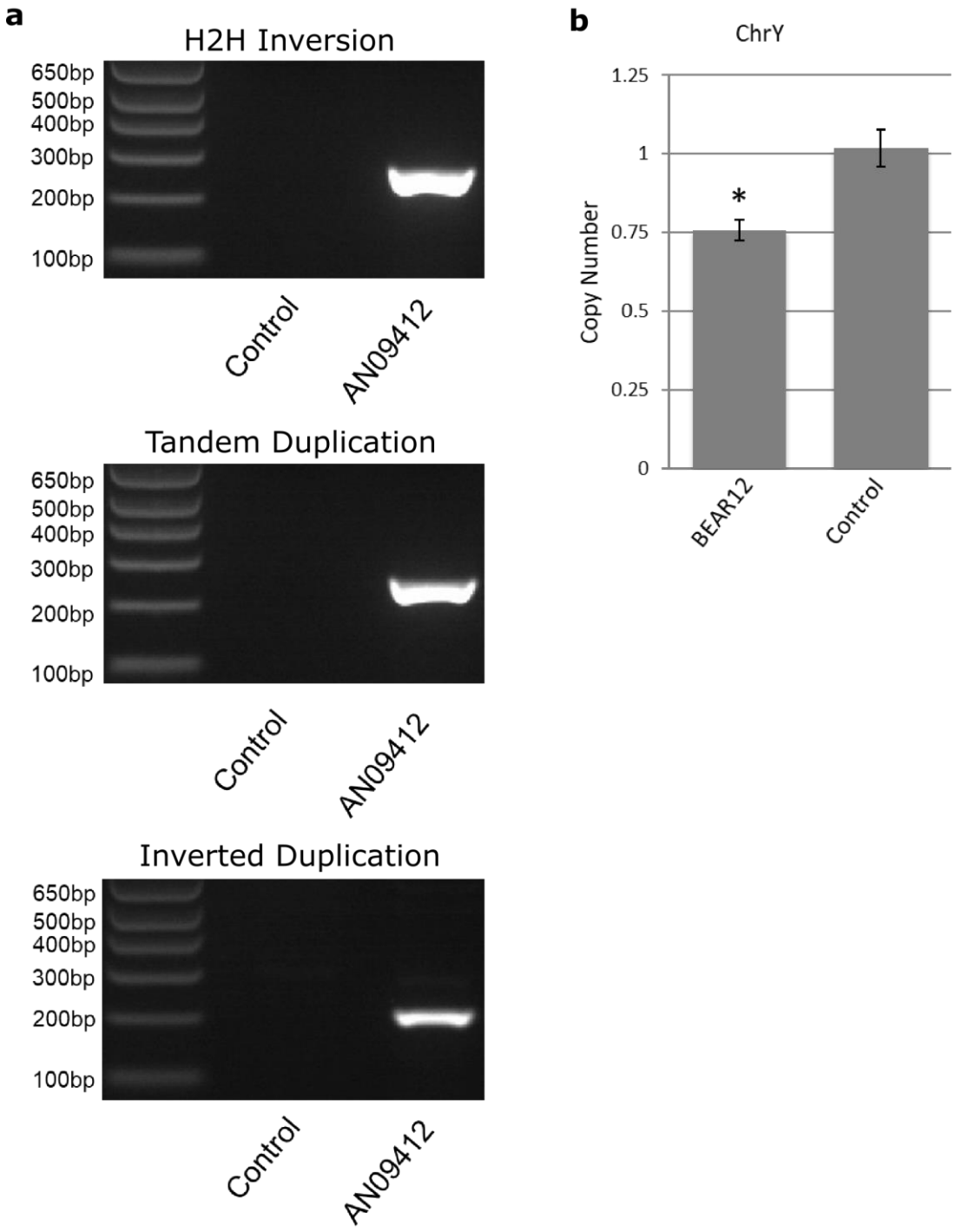
Supplementary Figure 13: Maternal allele fraction (matAF) and LRR plots of mosaic deletion in SPARK proband SP0095456 overlapping *NF1* (indicated by red triangle). The matAF plot contains only heterozygous SNPs.



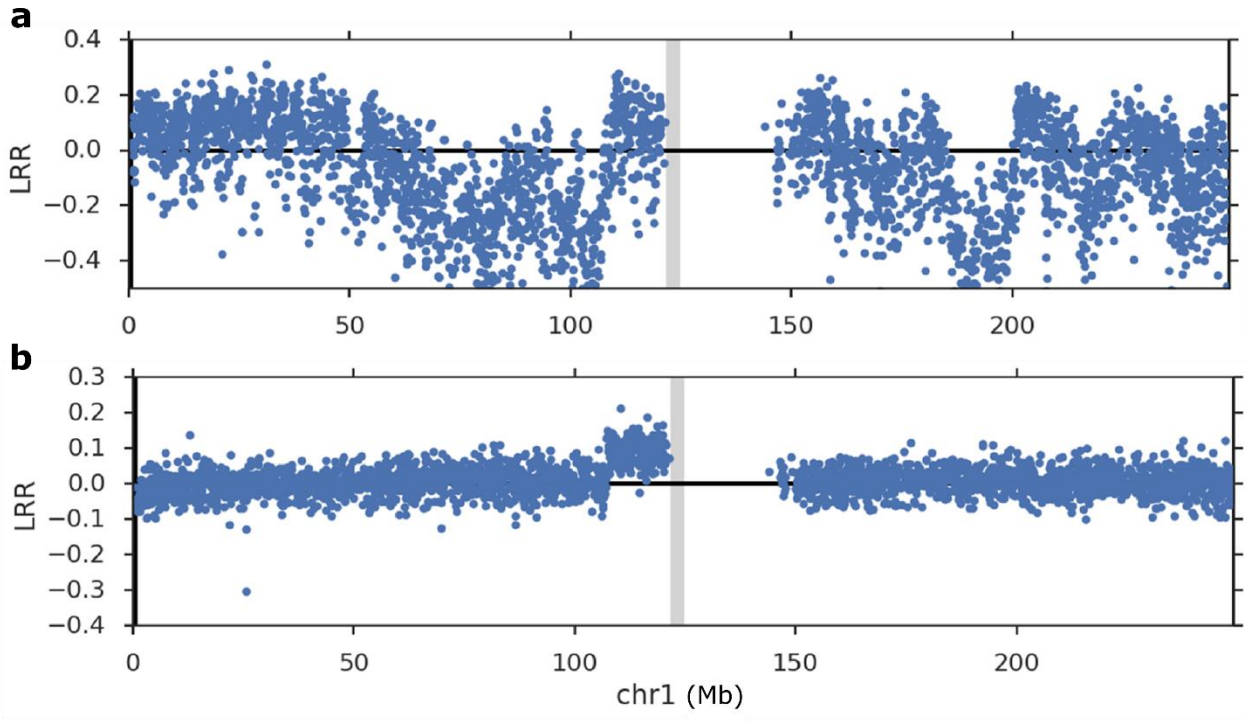
Supplementary Figure 14: Maternal allele fraction (matAF) and LRR plots of chr22 in SPARK proband SP0025588. The proband carries a mosaic duplication of paternal chr22 in 6.7% of cells (indicated by pink dots), and a germline de novo deletion on his maternal 22q haplotype (indicated by grey dots in LRR plot); a germline deletion converts all SNPs to a hemizygous state, and thus no heterozygous SNPs are present within the germline deletion except a few genotyping errors.



Supplementary Figure 15: Validation of a germline *NRXN1* deletion in brain WGS sample UMB5297. A, Difference from diploid copy number as calculated from WGS read depth. The red line indicates the ~75 kb germline deletion (2:50,731,007-50,805,387) within the alpha form of *NRXN1*. B, design of primers for mutant sequence resulting from the germline deletion and wildtype sequence present in the human reference genome. C, Gel electrophoresis of PCR products from mutant and wildtype primers. An extra band is visible in UMB5297, resulting from PCR amplification of the mutant sequence.



Supplementary Figure 16: Additional validation experiments for mCNVs discovered in WGS data. a) Gel electrophoresis of PCR products from primers for the breakpoints corresponding to the T2T inversion (top panel), tandem duplication (middle panel) and inverted duplication (bottom panel) in DNA from AN09412 and DNA from a control brain. White lines are visible when a successful PCR product was formed. b) copy number of chrY in brain tissue from BEAR12 and tissue from a control brain inferred from ddPCR. Error bars, approximate 95% CI.



Supplementary Figure 17: Log-R Ratio (LRR) across chromosome 1 in SPARK proband SP0064960 **a,** before and **b,** after applying principal-components denoising to the signal (Methods). After denoising, a mosaic duplication is evident from 107 Mb to 121 Mb.

2. Supplementary Notes

1. 13391.s1 chr4 event and its relationship to *TET2*

We investigated whether the complex, chromothripsis-like event found in SSC sibling 13391.s1 (Supplementary Table 1, Supplementary Fig. 3a) might be a clonal hematopoietic event. One duplicated segment (chr4:106,390,734-107,509,199) occurred in the neighborhood of *TET2* (chr4:106,067,842-106,200,960), a common target of driver mutations in clonal hematopoiesis[2,3] (Supplementary Fig. 3b); we hypothesized that the event might affect *TET2* and thus have resulted in clonal expansion in blood. We thus obtained whole-genome sequencing for this individual from SFARI Base and manually inspected the region for the exact location of the left-hand breakpoint depicted in Supplementary Fig. 3c. We identified split reads and discordant reads indicating a rearrangement in which chr4:106,286,478 is connected to chr4:125,620,145, consistent with the breakpoints estimated from BAF and LRR genotyping-intensity measures by MoChA (chr4:106,390,734 and chr4:125,610,859, respectively). While *TET2* does not appear to be altered by the event, we cannot rule out that regulation of *TET2* may be disrupted, resulting in aberrant expression.

We also contacted the Simons Foundation Autism Research Initiative to obtain additional information on the individual. They noted the individual was ~6 years old at time of assessment. An assessing clinician noted mild delay in gross motor and personal care per the Vineland Adaptive Behavior Scales but no other clinically significant behavioral findings. The family did not note any significant medical or psychiatric history for the child. The child did not receive an IQ test.

We thus found no conclusive evidence either for the event arising due to clonal hematopoiesis or for it being an early embryonic event, which we suspect would manifest phenotypically. Given the age of the child at assessment, it is possible that a developmental or neuropsychiatric phenotype was not yet identifiable. We were unable to reach the family for further follow-up.

Given that we could not conclusively determine that the event was clonal hematopoietic, we opted to be conservative and include the event in all main figure analyses. Exclusion of the event results in burden p-values that are slightly more significant and does not change any findings reported in the main text.

2. Cell fraction distribution of mCNVs.

The mosaic fraction of mCNVs in probands and siblings had a median of ~0.25 (Supplementary Fig. 6. We suspected this feature may reflect statistical limitations of our methods and not underlying biology because 1) MoChA's sensitivity to low cell-fraction mCNVs increases as mCNV length increases; 2) the copy number state of low-cell fraction mCNVs is not easily inferred (Supplementary Fig. 2); and 3) we had imposed a strict minimum cell fraction of 0.2 on CNN-LOH events taken forward to analysis. We therefore stratified the cell-fraction distribution by mCNV size (Supplementary Fig. 7) after removing CNN-LOH events. Consistent with the hypothesis that the cell fraction distribution is a result of methodological limitations, the median cell fraction decreased with increasing length ($P = 1.3\times10^{-3}$, comparison of cell fractions between mCNVs 0-1 Mb in size vs. mCNVs >10 Mb in size, Mann-Whitney U-test).

3. Mosaic CNV recurrence analysis

While most mCNVs we detected had unique, non-recurrent breakpoints (Supplementary Table 1), we observed two nearly identical mosaic duplications of 1pcen in SP0008373 and SP006490 (Supplementary Fig. 9a, top two panels). The event in SP0008373 had an estimated cell fraction of 17.1% while the event in SP006490 had an estimated cell fraction of 27.8%. We confirmed using `PLINK --genome`[4] that the two samples were unrelated (PI_HAT = 0.0398). Note, to calculate this statistic, we first calculated the minor allele frequency for each SNP across all genotyped individuals in SPARK and supplied these estimates to `PLINK –genome` using the `--read-freq` option. Additionally, we observed a CNN-LOH event in SP0074922 (Supplementary Fig. 9a, bottom panel) which ended in the same vicinity as the two duplications started. Thus, we observed three breakpoints within a 3 Mb region (chr1:106,052,011-108,310,224). While this region includes *NTNG1*, a gene that has been associated with ASD[5], the probability of observing three or more breakpoints within a 3 Mb region is P=0.107 under the assumption that breakpoints are distributed uniformly at random across the genome.

4. Analysis of mosaic CNVs in 16p11.2 in the UK Biobank

16p11.2 *de novo* germline CNVs (both gains and losses) are strongly associated with ASD[6,7], yet we observed no mosaic analogues of such events in either SSC or SPARK probands or siblings.

We looked for mosaic analogues of the germline 16p11.2 CNVs (duplications or deletions) among mCNVs identified in the larger UK Biobank cohort[1]. We observed 73 events contained within the boundaries the extended 16p11.2 CNV locus (chromosome 16: 28,000,000-31,000,000, hg19), of which all were deletions. The carriers of these events were heavily biased to be female (Observed: 56 females, 17 males; Expected: 40 females, 33 males; P = 8.7e-5, Fisher's Exact Test) as has been previously reported[8]. We next checked whether these events could have arisen due to age-related clonal hematopoiesis; we observed a small, non-significant increase in prevalence of 16p11.2 losses with age (mean age = 57.3 (s.e.m. = 0.89) years in carriers; mean age = 56.5 years in the full cohort), in contrast with the strong increase in prevalence of other mosaic events with age (mean age = 59.5 (s.e.m. = 0.1) years in carriers of other events), suggesting an early-developmental origin of these events.

5. Putative damaging variants within mosaic CNVs

Beyond directly disrupting a gene by deletion, bifurcation, or dosage alterations, a mosaic CNV can convert a damaging variant that is heterozygous in the germline state into a hemizygous or homozygous variant in the mosaic state: suppose an individual inherited a damaging variant on paternal 1p; if the individual acquires a mosaic UPD of paternal 1p (CNN-LOH in which the paternal haplotype replaces the maternal haplotype), the damaging variant will be homozygous (i.e. present on both haplotypes) in the mosaic cells. If the individual acquires a mosaic deletion of 1p on the maternal haplotype, the damaging variant will be hemizygous in the mosaic cells. We thus searched for putative damaging variants (stop-gain, start-loss, frameshift, splice-site, or missense variant with CADD Phred score >20, ref. 8) converted to a hemizygous or homozygous state by a mCNV (Methods).

We found several examples of this phenomenon (Supplementary Table 6). SPARK proband SP0069140 carried a *NRXN1* start-loss variant on the maternal haplotype that was converted to a hemizygous state through a mosaic deletion of the paternal *NRXN1*. This example is described in detail in that main text. Additionally, every CNN-LOH event converted at least one putative damaging variant to a homozygous state. However, none of these variants occurred within known ASD genes, and their clinical relevance was of unknown significance.

6. Additional mosaic CNVs with plausible connections to proband phenotype

ASD, related neuropsychiatric disorders, and co-morbid medical conditions are phenotypically diverse, and the limited, standardized phenotypic information provided for each proband in SSC and SPARK is generally not detailed enough to allow clinical diagnosis of particular syndromes. We were therefore unable to confirm with high confidence that four mCNVs which clearly disrupted known ASD genes were likely responsible for the observed phenotype. However, in each case, we found that the observed phenotypes were fairly consistent with the mCNV being causative.

Rare mutations in *FOXP1* are associated with ASD[5] and intellectual disability[9,10]. SSC individual 11270.p1 carries a 19.0 Mb deletion encompassing *FOXP1* in 28% of cells. Consistent with germline disruption of *FOXP1*, the proband has evidence of significant intellectual disability (Non-verbal IQ, NVIQ=49; Vineland Adaptive Summary Score, VSS=63, reported phrased speech delay).

Loss of function mutations and microdeletions affecting *SETD5* have been implicated in ASD[11], ID[12], and developmental delay[13]. SSC proband 13362.p1 carries a 6.6 Mb deletion encompassing *SETD5* in 31% of cells. While the proband has reported speech delay and cognitive function below the population average (FSIQ=82, NVIQ=83, VSS=86), they do not meet the criteria for ID. We hypothesize that this may represent a mosaic phenotype which is milder than an equivalent germline analogue.

De novo and mosaic LOF variants in *BAZ2B* have been associated with ASD[14,15] and moderate intellectual disability[16]. SSC proband 11671.p1 carries an 11.9 Mb deletion encompassing *BAZ2B* in 33% of cells. Consistent with previous reports, the proband has a medical diagnosis of mild ID (DSM IV diagnosis code 317).

SPARK proband SP0095456 carried a 4.6 Mb deletion encompassing *NF1* in 52% of cells (Supplementary Fig. 13). Mosaic *NF1* microdeletions have previously been reported; unlike their germline counterparts, mosaic *NF1* losses are not associated with increased risk of intellectual disability or medical conditions including congenital heart abnormalities or neurofibroma tumors[17]. The mCNV we detected is significantly larger than most mosaic *NF1*-microdeletions (typically 1.2 Mb)[17,18]; nonetheless the proband's phenotype is mild, consistent with those previous reports: a learning disability but no cognitive impairment or medical conditions. While *NF1* deletions are often observed in clonal hematopoiesis, the mCNV we

detected seems unlikely to be present only in leukocytes due to the young age of the proband and the high cell fraction of the event. Indeed, in the UK Biobank, we observed 49 focal *NF1* deletions (occurring between 29.42 and 39.70 Mb on chr17). Of these, only eight (16%) had cell fraction estimated to exceeded that of the event observed in SP0095456 (>52%). Seven of the eight individuals were older than 60 years of age, and all eight were older than 55 years of age. The proband was 7 years of age at time of sample evaluation.

7. Other events with unverified disruption of ASD genes or connection to phenotype

Two probands carry events whose breakpoints appeared to bifurcate ASD genes. In this case, the CNV would act like a loss-of-function variant. However, we were unable to confirm the exact location of the breakpoints because genome sequencing (either whole-exome or whole-genome) was not available for these samples. A third proband carried an event which appeared to be a mosaic rescue of a germline deletion, but the phenotypic data available for the proband was not sufficiently detailed to confirm if the rescue resulted in a milder than expected phenotype.

SPARK proband SP0016887 carried a 667 kb duplication in 26% of cells whose left breakpoint appeared to bifurcate *TRIO*. *TRIO* has been previously reported to be enriched for variants (inherited and *de novo*) likely to affect ASD risk[11]. The proband exhibits speech delay (single words at 31 months). However, neither WES nor WGS sequencing was available to confirm the exact location of the left breakpoint of this event.

SSC proband 13674.p1 carried a 22.9 Mb deletion of 4p in 8.3% of cells. Simultaneously we also detected a 23.7 Mb mosaic event of 12p in 8.5% of cells (Supplementary Table 12). While we were unable to classify the chr12 event due to its low cell-fraction, we suspect that it is a duplication arising from an unbalanced translocation wherein the first 23 Mb of chr12 was duplicated and replaced the first 23 Mb of 4p. The chr12 event contains multiple genes previously implicated in ASD including C12orf57, *CACNA1C*, *GRIN2B*; additionally, the left breakpoint appears to bifurcate the ASD gene *SOX5*. Neither WES nor WGS was available to confirm this hypothesis.

SPARK proband SP0025588 carried a 7.5 Mb *de novo* germline deletion of maternal 22q (22q13.2-q13.33). Such events have been reported in multiple cases of individuals with developmental disorders[19] and typically cause Phelan-McDermid syndrome[20] (OMIM: 606232).

Interestingly, the proband also exhibited a mosaic rescue genotype: 6.7% of cells carried a duplicated paternal chr22 and thus returned 22q to a diploid state (Supplementary Fig. 14). We hypothesize that this mosaic rescue should decrease the severity of the individual's phenotype. While the proband has reported symptoms consistent with Phelan-McDermid Syndrome including being non-verbal at 14 years old, having intellectual disability and motor delay, the limited phenotype data provided no evidence either for or against decreased disease severity.

8. Germline CNVs in brain tissue with plausible connection to ASD

We identified 15 disease-relevant germline SVs in ASD brains (Supplementary Table 10), revealing potential causes of disease in several previously unsolved cases. These include a case of germline 15q-duplication in case AN06365, a well-documented syndromic cause of autism usually occurring in the de novo state[21], as well as a 10Mb germline copy gain in case UMB1638 in chromosome 20q13.2-13.33. Duplications of this region have been previously identified in individuals with ASD[22,23] . A small germline deletion in the ASD risk gene *NRXN1* was found in one ASD case, UMB5297 and validated via PCR Supplementary Fig. 15). Importantly, this 75kb deletion affects five critical exons and is likely causative of disease in this case[24–26].

3. Supplementary Table Descriptions

Supplementary Table 1:

- **Description:** mosaic CNVs in probands and unaffected siblings
- **Sheet name:** ST1_mosaic_CNVs
- **Columns:**
    - **SAMPLE**: SSC or SPARK sample ID
    - **Cohort**: whether sample from SSC or SPARK
    - **SEX:** sex as reported in phenotype data
    - **AGE:** in years (only available for SPARK samples)
    - **CNV_TYPE:** DUP (gain), DEL (loss), or CNN-LOH (copy-number neutral loss of heterozygosity)
    - **CHROM:** chromosome number
    - **BEG (GRCh37):** start position of mosaic CNV (GRCh37 coordinates)
    - **END (GRCh37):** end position of mosaic CNV (GRCh37 coordinates)
    - **LENGTH:** size in base pairs of mosaic CNV
    - **CYTOBAND:** chromosomal cytobands overlapped by mosaic CNV
    - **CF:** mosaic cell fraction
    - **NSITES:** number of array probes within the CNV
    - **HETS:** number of heterozygous SNPs within the CNV
    - **BDEV:** absolute B allele frequency deviation from 0.5
    - **BDEV_SE:** standard error of BDEV
    - **LDEV:** Log R ratio deviation from 0
    - **LDEV_SE:** standard error of LDEV
    - **LOD_LRR_BAF:** Log-odds ratio of MoChA's LRR + BAF model
    - **LOD_BAF_PHASE:** log-odd ratio of MoChA's BAF + Phase model
    - **PATERNAL_BAF:** average allele frequency of heterozygous SNPs inherited from the father
    - **MATERNAL_BAF:** average allele frequency of heterozygous SNPs inherited from the mother
    - **HAPLOTYPE_OF_EVENT:** parental haplotype on which the CNV occurs
    - **N_RefSeq:** number of RefSeq genes within the CNV
    - **ASD_GENES:** curated high confidence ASD genes within the CNV
    - **FSIQ:** Full-scale IQ
    - **NVIQ:** Non-verbal IQ
    - **SCQ_summary_score:** Social communication questionnaire summary score
    - **SCQ_normed:** Z-normalized SCQ summary score

Supplementary Table 2:

- **Description:** *de novo* CNVs or loss-of-function variants in mosaic carriers from SSC
- **Sheet name:** ST2_SSC_dnCNVs
- **Columns:**
    - **SAMPLE:** as in Supplemental Table 1
    - **CHROM:** as in Supplemental Table 1
    - **BEG (GRCh37):** as in Supplemental Table 1
    - **END (GRCh37):** as in Supplemental Table 1

    - **LENGTH:** as in Supplemental Table 1
    - **Type:** dnCNV type (Del or Dup) or LoF SNV type

Supplementary Table 3:

- **Description:** counts of carriers of ASD-associated CNVs across cohorts.
- **Sheet name:** ST3_ASD_CNV_counts
- **Columns:**
    - **LOCATION:** location of ASD-associated CNV
    - **ASD Probands:** counts of carriers in ASD cohorts
        - Germline: SSC & Autism Genome Project from Sanders et al.[7]
        - SSC+SPARK for mosaic (this analysis)
    - **UK Biobank:** counts of carriers in the UK Biobank
        - Germline: From Crawford et al.[27]
        - Mosaic: From Loh et al.[1]
    - **Germline carrier:** germline carrier counts
    - **Mosaic carrier:** mosaic carrier counts

Supplementary Table 4:

- **Description:** Mosaic CNVs and associated mental health phenotypes in the UK Biobank
- **Sheet name:** ST4_ASD_mCNVs_UKB
- **Columns:**
    - **LOCATION:** location of ASD-associated CNV
    - **CHROM:** Chromosome
    - **BEG (GRCh37):** Beginning of region in which mosaic CNVs were identified
    - **END (GRCh37):** End of region in which mosaic CNVs were identified
    - **N:** total individuals analyzed
    - **UKB_LOSS:** # individuals with a mosaic loss
    - **UKB_GAIN:** # individuals with a mosaic gain
    - **N_mentalQ:** # carriers who completed the mental health questionnaire
    - **N_ASD:** # carriers with self-reported ASD diagnosis
    - **N_SCZ:** # carriers with self-reported schizophrenia diagnosis
    - **N_BP:** # carriers with self-reported bipolar diagnosis
    - **N_ADHD:** # carriers with self-reported ADHD diagnosis
    - **N_DEPRESSION:** # carriers with self-reported depression diagnosis
    - **N_ANXIETY:** # carriers with self-reported anxiety diagnosis

Supplementary Table 5:

- **Description:** Phenotype associations of germline and mosaic 16p11.2 deletions
- **Sheet name:** ST5_16p11.2_UKB_assoc
- **Columns:**
    - **Phenotype:** continuous or binary phenotype
    - **Mosaic:** Results for all mosaics
    - **Mosaic (high CF):** Results for high cell fraction (>0.3) mosaics
    - **Germline:** Results for germline deletions
    - **N carrier:** number of carriers
    - **N controls:** number of controls

    - **Coeff:** regression coefficient of carrier status
    - **Coeff_se:** standard error of carrier status regression coefficient
    - **P:** p-value of regression coefficient different than zero

Supplementary Table 6:

- **Description:** Likely gene disrupting variants made homozygous by a mosaic CNV
- **Sheet name:** ST6_LGD_variants_in_mCNVs
- **Columns:**
    - **SAMPLE (event):** Sample ID and mosaic CNV description
    - **CHROM:POS:** Chromosome and start position of LGD variant
    - **TYPE:** variant description
    - **HAPLOTYPE:** parental haplotype of LGD variant
    - **GENE:** gene containing the variant
    - **EFFECT:** predicted effect from Variant Effect Predictor

Supplementary Table 7:

- **Description:** mosaic CNVs in SPARK siblings with known genetic diagnosis
- **Sheet name:** ST7_sibs_w_genetic_diag
- **Columns:** as in Supplementary Table 1

Supplementary Table 8:

- **Description:** list of WGS samples
- **Sheet name:** ST8_WGS_sample_list
- **Columns:**
    - **Case ID:** Sample ID from brain bank
    - **Age (years):** Age in years at time of death
    - **Sex:** biological sex
    - **Brain Region Sequenced:** brain location of sequenced tissue
    - **Cause of Death:** cause of death
    - **PMI (hours):** post-mortem interval
    - **Brain Bank:** brain bank that provided tissue
    - **Sequencing Company:** company that sequenced tissue
    - **Library Preparation Method:** method used to prepare tissue for sequencing
    - **Average Coverage:** average read depth of sequencing

Supplementary Table 9:

- **Description:** mosaic CNVs in WGS samples
- **Sheet name:** ST9_WGS_mCNVs
- **Columns:** as in Supplementary Table 1

Supplementary Table 10:

- **Description:** germline CNVs overlapping ASD genes in WGS samples
- **Sheet name:** ST10_WGS_germline_CNVs
- **Columns:** as in Supplementary Table 1

Supplementary Table 11:

- **Description:** breakpoints in AN09412 WGS data supporting mosaic CNV on chromosome 2
- **Sheet name:** ST11_AN09412_breakpoints
- **Columns:**
    - **CHROM:** chromosome number
    - **START:** left-hand side of fusion breakpoint
    - **END:** right-hand size of fusion breakpoint
    - **# SUPPORTING DISCORDANT READS:** number of discordant reads consistent with the breakpoint
    - **# SUPPORTING SPLIT READS:** number of splits reads consistent with the breakpoint
    - **AVERAGE READ DEPTH:** Average read depth at the breakpoint

Supplementary Table 12:

- **Description:** Unknown mosaic copy state event in SSC proband 13674.p1
- **Sheet name:** ST12_13674.p1_UKN_event
- **Columns:** as in Supplementary Table 1

4. References

1. Loh, P.-R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *bioRxiv* 653691 (2019) doi:10.1101/653691.
2. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* **371**, 2477–2487 (2014).
3. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* **371**, 2488–2498 (2014).
4. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
5. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
6. Sanders, S. J. *et al.* Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* **70**, 863–885 (2011).
7. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
8. Loh, P.-R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).
9. Hamdan, F. F. *et al.* De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671–678 (2010).
10. Horn, D. *et al.* Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Human Mutation* **31**, E1851–E1860 (2010).

11. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
12. Grozeva, D. *et al.* De Novo Loss-of-Function Mutations in SETD5, Encoding a Methyltransferase in a 3p25 Microdeletion Syndrome Critical Region, Cause Intellectual Disability. *The American Journal of Human Genetics* **94**, 618–624 (2014).
13. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
14. Iossifov, I. *et al.* Low load for disruptive mutations in autism genes and their biased transmission. *PNAS* **112**, E5600–E5607 (2015).
15. Krupp, D. R. *et al.* Exonic Mosaic Mutations Contribute Risk for Autism Spectrum Disorder. *The American Journal of Human Genetics* **101**, 369–390 (2017).
16. Bowling, K. M. *et al.* Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Med* **9**, 43 (2017).
17. Kehrer-Sawatzki, H., Mautner, V.-F. & Cooper, D. N. Emerging genotype-phenotype relationships in patients with large NF1 deletions. *Hum. Genet.* **136**, 349–376 (2017).
18. Vogt, J. *et al.* SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biology* **15**, R80 (2014).
19. Luciani, J. J. *et al.* Telomeric 22q13 deletions resulting from rings, simple deletions, and translocations: cytogenetic, molecular, and clinical analyses of 32 new observations. *Journal of Medical Genetics* **40**, 690–696 (2003).
20. Phelan, M. C. *et al.* 22q13 deletion syndrome. *American Journal of Medical Genetics* **101**, 91–99 (2001).

21. Finucane, B. M. *et al.* 15q Duplication Syndrome and Related Disorders. in *GeneReviews*® (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
22. Guo, H. *et al.* Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. *Sci Rep* **7**, 44155 (2017).
23. Maini, I. *et al.* Prematurity, ventricular septal defect and dysmorphisms are independent predictors of pathogenic copy number variants: a retrospective study on array-CGH results and phenotypical features of 293 children with neurodevelopmental disorders and/or multiple congenital anomalies. *Ital J Pediatr* **44**, 34 (2018).
24. Ching, M. S. L. *et al.* Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **153B**, 937–947 (2010).
25. Lowther, C. *et al.* Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. *Genet Med* **19**, 53–61 (2017).
26. Viñas-Jornet, M. *et al.* A common cognitive, psychiatric, and dysmorphic phenotype in carriers of NRXN1 deletion. *Mol Genet Genomic Med* **2**, 512–521 (2014).
27. Crawford, K. *et al.* Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *Journal of Medical Genetics* **56**, 131–138 (2019).