

Rausch *et al.*, Supplemental Information

Contents

Contents	1
Supplemental Figure Legends	2
Supplemental Dataset Legends.....	2
Supplemental Tables	3
Supplemental Experimental Procedures	4
1. DNA sequencing and single nucleotide variant analysis.....	4
1.1 DNA sequencing, and mapping DNA reads onto the reference genome	4
1.2 Discovery of single nucleotide variants (SNVs) and short InDels	4
1.3 Estimation of <i>TP53</i> mutant allele frequencies.....	5
2. Large-scale somatic structural rearrangement analysis	5
2.1 Mapping genomic structural rearrangements by paired-end mapping.....	6
2.2 Genomic rearrangement fine-mapping by split-read analysis.....	7
2.3 Genomic rearrangement identification by read-depth analysis.....	7
2.4 Simulation of step-wise genomic rearrangements	7
2.5 Mouse DNA copy-number profile data analysis	8
2.6 Analysis of chromothripsis in several LFS-associated malignancies	9
2.7 Analysis of rearrangements in tumors with somatic <i>TP53</i> mutations	9
2.8 Analysis of chromothripsis in AML.....	10
2.8 Cancer gene enrichment analysis.....	11
3. Gene expression analysis	11
3.1 AMLs displaying chromothripsis do not show SHH pathway activation....	11
Supplemental References	13

Supplemental Figure Legends

Figure S1 – Rearrangements resulting from chromothripsis on LFS-MB1 chromosome 3, inferred to form a double minute chromosome

[part of external Supplemental Figure file, linked to Figure 1]

Figure S2 – PCR validations of fourteen inter-amplicon connections in LFS-MB2

[part of external Supplemental Figure file, linked to Figure 3]

Supplemental Dataset Legends

Data S1 – DNA copy-number profiling indicates link between *TP53* mutations and chromothripsis. **(A-B)** SNP-array based copy-number profiles of *TP53* wild-type SHH-MBs. **(C-F)** SNP-array based copy-number profiles of *TP53* mutant SHH-MBs. **(G)** Inference of chromothripsis in several LFS-associated malignancies.

[part of external Supplemental Figure file, linked to Figure 2]

Data S2 – Array-based DNA copy-number profiling in SHH-MB (mouse), in lung cancer (human) and in AML (human), analysis of clinical data in AML, and gene expression profile analysis in AML. **(A-D)** Chromosome rearrangements consistent with chromothripsis are seen in a mouse model of SHH-MB lacking the p53 mouse ortholog (Trp53^{-/-}), but not in a mouse model with wild-type Trp53. **(E)** Example of rearrangements in squamous cell lung cancer case, which were inferred to comprise classical (progressive) chromosome rearrangements. **(F)** Copy-number profile analysis revealed chromothripsis in squamous cell lung cancer case. **(G)** Examples of AMLs with inferred chromothripsis in SNP array data (cohort from Parkin and coworkers). **(H)** Additional examples of AMLs with inferred chromothripsis (second patient cohort). **(I)** In a univariate analysis on 113 AML samples with medical follow-up data chromothripsis was a significant prognostic factor for overall survival ($P < 0.0001$; Wald's Test; see left panel). Furthermore, multivariate analyses revealed chromothripsis ($P = 0.025$), but not "complex" karyotype grouping (Haferlach et al., 2008), as a significant prognostic marker in AML. Similar results were obtained from a second AML cohort with medical follow-up data on 163 individuals (right panel). We further found that the effect of chromothripsis on prognosis is independent of age ($P < 0.0001$). **(J)** Comparative gene expression analysis of canonical SHH pathway genes indicates that chromothripsis in AML occurs in the absence of SHH pathway activation.

[part of external Supplemental Figure file, linked to Figure 5]

Supplemental Tables

Table S1 – DNA sequence variants identified in LFS-MB1

[external file, linked to Figure 1]

Table S2 – Tumor samples analyzed by SNP array-based DNA copy-number profiling

[external file, linked to Figure 2]

Table S3 – DNA sequence variants identified in LFS-MB2 and LFS-MB3

[external file, linked to Figure 3]

Table S4 – DNA sequence variants identified in LFS-MB4

[external file, linked to Figure 4]

Table S5 – Supplemental table related to Figure 5, with copy-number profile and DNA breakpoint junction analyses

[external file, linked to Figure 5]

Supplemental Experimental Procedures

1. DNA sequencing and single nucleotide variant analysis

1.1 DNA sequencing, and mapping DNA reads onto the reference genome

Matched tumor and germline sample pairs were sequenced on Illumina HiSeq 2000 and Genome Analyzer II instruments using paired-end libraries. The raw length of the reads displayed a median of 101bp, and the median insert size of the sequenced libraries was 285-325bp for short insert sizes ('Illumina paired-end [PE] protocol'), and 3900-4545bp for long-range insert sizes ('Illumina mate-pair [MP] protocol'). Reads were aligned to the hg19 assembly of the human reference genome using the Illumina-provided alignment software (ELAND, version 2) and also with BWA (Li and Durbin, 2009). Post-processing of the aligned reads included merging of lane-level data and removal of duplicate read pairs using Picard tools (<http://picard.sourceforge.net>). Only uniquely aligned reads were considered for downstream mutation analysis, and aligned reads were converted to the SAM/BAM format using SAMtools (Li et al., 2009). Coverage calculations, following duplicate removal, considered all informative bases of the reference genome (excluding Ns).

1.2 Discovery of single nucleotide variants (SNVs) and short InDels

We used consensus calls from two different single nucleotide polymorphism analysis pipelines to call somatic SNVs. First, we applied the Genome Analysis Toolkit (GATK) (Depristo et al., 2011) on reads aligned to the reference genome with ELAND. We therefore used the quality recalibration and the local realignment features of GATK, and called SNVs with the *UnifiedGenotyper*. Second, we used SAMtools (Li et al., 2009) mpileup, version 0.1.14, to call SNVs from the BAM files produced by BWA. Candidate calls were then categorized as germline or somatic according to whether there was evidence for the same event at the same locus in the BAM file of the matched control sample. Candidate mutations were subject to the following empirical filters designed to eliminate false positive calls. We excluded SNV calls without sequence coverage in the normal BAM file, as well as calls overlapping simple repeat or microsatellite regions (regions commonly observed to yield false positive SNV and InDel calls). Local sequence context (for example, the GGnnGG motif) can lead to incorrect base calls, but typically involves reads sequenced from a single strand only. Thus, if the variant call was supported by reads from only one strand, we required the opposite strand to be present in at least 30% of all reads mapping to the locus, otherwise the variant call was excluded due to strand bias. Finally, filtered calls

were annotated with RefSeq gene annotations and dbSNP v132, followed by identification of non-synonymous variants using Annovar (Wang et al., 2010) and prediction of the possible functional impact using Polyphen (Adzhubei et al., 2010) version 2.0.22.

We finally compared both call sets (ELAND/GATK and BWA/mpileup) to establish a consensus SNV set. Namely, SNVs that were called by both pipelines were included and regarded as high confidence calls. Furthermore, SNVs affecting protein-coding regions, which were called by only one pipeline, were subject to manual review in both ELAND and BWA alignments, and included if manual review supported a valid alignment. Small (<50bp) insertions and deletions (InDels) were called using the Dindel tool (Albers et al., 2011) with default parameters, separately on the ELAND and BWA alignments. All consensus InDels (inferred from both ELAND and BWA alignments) predicted to affect protein-coding exons were subject to manual inspection to identify possible mis-alignments (*i.e.*, false positive InDels).

1.3 Estimation of *TP53* mutant allele frequencies

We used deep sequencing data to estimate the tumor cell population *TP53* mutant allele frequency of several medulloblastoma samples. Whole-genome sequencing data were used to evaluate mutant allele frequencies of the LFS samples LFS-MB1 – LFS-MB4. In addition, we obtained deep sequencing reads for one additional SHH-MB (MB2034) and two WNT tumors (MB518 and MB496) from the International Cancer Genome Consortium (ICGC) Pediatric Brain Tumor Research Project (www.pedbraintumor.org), which granted early access to Illumina paired-end DNA reads aligned onto the *TP53* gene locus. All reads that were unambiguously mapped onto the genomic base position in question were accounted for. Results are summarized in Table S2.

2. Large-scale somatic structural rearrangement analysis

The detection and analysis of genomic rearrangements ≥ 50 bp in size achieves optimal sensitivity and accuracy when complementary discovery approaches are used (Mills et al., 2011). We thus evaluated the read alignments to detect genomic rearrangements based upon three types of *informative* alignment characteristics: read-depth analysis, paired-end mapping, and split-read analysis (Mills et al., 2011). Whereas read-depth analysis enables assessing the copy-number status of genomic segments of interest (*i.e.*, of such undergoing unbalanced rearrangements), paired-end mapping allows fine-mapping the boundaries of both unbalanced and balanced rearrangements (*e.g.*, including inversions and translocations), and split-read analysis enables inference of rearrangement boundaries at nucleotide resolution. Rearrangement calls without a corresponding variant in the matched normal sample were inferred to

be somatic (*i.e.*, tumor-specific) when identified as unique based on an 80% reciprocal overlap criterion.

2.1 Mapping genomic structural rearrangements by paired-end mapping

We inferred structural rearrangements using an extended version of our earlier approach for paired-end mapping analysis implemented in the PEMer tool (Korbel et al., 2009). Generally speaking, paired-end (or “read pair”) methods rely on comparing mapped paired-ends with an expected mapping distribution, based on expected insert size lengths. The span coverage (also called physical coverage) delineates the sensitivity of read pair methods, and large insert libraries (such as ~4.5kb mate-pair libraries) can thus increase the power of paired-end mapping.

We analyzed paired-end data in the following way: For each input BAM format (Li et al., 2009) file, based on the ELAND version 2 mappings, we first computed the paired-end insert size distribution and determined the median and standard deviation of the insert size distribution. We then identified all discordantly mapped read-pairs that either displayed an abnormal orientation or an insert size greater than the expected range. As a default, we used an insert-size cutoff three standard deviations away from the median insert size to identify discordant paired-ends. Our paired-end mapping analysis module then clustered outliers (discordant paired-ends) into groups corresponding to candidate structural rearrangements (Korbel et al., 2009). We achieved this by computing *connected components* from the discordant paired-ends and identifying a *maximal clique* (candidate rearrangement) for each component. Specifically, we identified instances where two or more discordant read-pairs supported the same structural rearrangement, and hence formed a *clique* – this required both pairs to have the same relative orientation and it required that the absolute distance between both discordantly mapped read pairs was within the expected insert size range. All cliques were required to contain a minimum of two paired-ends to be further considered.

For consistency, non-chromothripsis associated rearrangements were discovered using short-insert size (PE) sequencing data only. Calls were further filtered as follows: For deletions, we required either at least four supporting pairs, or two (or more) supporting pairs and a supporting breakpoint-spanning split-read. For tandem duplications, we required a minimum of two supporting pairs together with a breakpoint-spanning split-read. For rearrangements larger than 100 kb, we required additionally a 50% reciprocal overlap with a read depth-based *gain* or *loss* call (see below) for tandem duplications and deletions, respectively. For inversions and translocations, we required at least two supporting pairs spanning both breakpoints of the event together with a supporting split read that confirmed at least one of the inferred breakpoint junctions.

To analyze chromothripsis-related structural rearrangements we included short (PE) as well as large (MP) insert-size sequencing data. To account for differences in physical coverage we required at least five supporting pairs for LFS-MB1 and LFS-MB2, at least nine supporting pairs for LFS-MB3, and at least thirteen supporting pairs for LFS-MB4.

2.2 Genomic rearrangement fine-mapping by split-read analysis

The split-read analysis module of our structural rearrangement analysis pipeline uses cliques identified in the paired-end mapping analysis step as an input. For each structural rearrangement interval we screened the putative breakpoint coordinates for single-anchored reads, since the non-mapped reads of such pairs represent suitable candidates for split-read alignments (Ye et al., 2009). We aligned all these previously unmapped reads to the candidate intervals using a fast implementation of the Gotoh-algorithm (Rausch et al., 2008). Using split-read alignments as an input we computed the consensus sequence and aligned the internal consensus sequence to the inferred rearrangement event. Our approach then performed an internal check to assess whether the genomic rearrangement predicted by breakpoint junctions inferred with split-reads agreed with the event inferred by paired-end mapping analysis, by checking whether the event falls into the 99.7% confidence intervals of the respective clique (computed from the insert size distribution). Split-read analysis was used both for fine-mapping rearrangements to nucleotide resolution, as well as to obtain additional support for an inferred event.

2.3 Genomic rearrangement identification by read-depth analysis

We used read-depth analysis to obtain further evidence for unbalanced rearrangements (*i.e.*, tumor-specific deletions or duplications) by applying the CNVnator algorithm (Abyzov et al., 2011) on the ELAND (version 2) mapped tumor and control tissue DNA reads. We used a genomic window size of 1000bp for detecting copy-number alterations with CNVnator. We used CopySeq (Waszak et al., 2010) to obtain copy-number state estimates of genomic segments by read depth analysis.

2.4 Simulation of step-wise genomic rearrangements

Progressively occurring genomic rearrangements were simulated using a strategy in the spirit of a recently developed Monte Carlo simulation approach (Stephens et al., 2011). In brief, starting with homologous pairs of wild-type chromosomes, rearrangements were randomly selected without replacement from the total set of events (*i.e.*, from the set of rearrangements identified based

on paired-end mapping). In each step, the picked rearrangement was applied to the current chromosomal configuration. In cases where a picked rearrangement was impossible (for example, a duplication occurred in a segment previously lost), it was discarded and another selected. If more than one copy of the rearrangement location existed in the current configuration (*e.g.*, the segment had undergone duplication previously), we randomly picked from all possible locations. We monitored both the copy-number of the most amplified segment, as well as the mean copy-number of the segments undergoing amplification. In none, or very few, out of 1000 simulations we performed in each sample did the amplicon segments achieve a similarly high excess in segmental copy-number relative to normal (disomic) regions as we observed in our samples (Figure 5B).

2.5 Mouse DNA copy-number profile data analysis

We analyzed two sets of microarray data from mouse models of SHH-MB to evaluate links between p53 and chromothripsis. We thereby focused our analysis on mice that did not receive irradiation for tumor growth induction, since irradiation can artificially induce chromosome breaks (Ishida et al., 2010). Furthermore, we used the same criteria for inferring chromothripsis as for the SNP microarray-based analysis (≥ 10 changes in copy-number on an individual chromosome, involving three or fewer copy-number states; see Supplemental Experimental Procedures). The first aCGH set we analyzed was from a *Ptch*(+/-);*Trp53*(-/-) background, *i.e.*, generated from an allograft-transplanted mouse medulloblastoma tumor into nude mice (Buonamici et al., 2010). Tumors in recipient mice were treated with a Smo-antagonist (drug) in some cases, which inhibits SHH signaling and stimulates tumor regression (Buonamici et al., 2010). These drug-treated tumors gained resistance, in part by high upregulation of SHH genes, and hence developed into aggressive SHH-MBs (Buonamici et al., 2010). DNA from drug-treated, and alternatively mock-treated, allografted tumors was analyzed by aCGH. For data analysis, we downloaded the aCGH data from GEO and imported them into Nexus 6v10 (Biodiscovery). We analyzed aCGH-based DNA copy-number profiles using FASST2 segmentation (Biodiscovery; default parameters). DNA from drug-treated and also from mock-treated *Ptch*(+/-);*Trp53*(-/-) mice displayed extensive chromosomal rearrangements consistent with chromothripsis in each case (Data S2). Chromosomes, with maximum copy-number state changes per individual chromosome in parenthesis, are as follows: three *Ptch*(+/-);*Trp53*(-/-) mice, mock treated: chr16 (30 copy-number state changes); chr16 (38); chr16 (36). Three *Ptch*(+/-);*Trp53*(-/-) mice, drug-treated: chr8 (94); chr12 (72); chr16 (28).

The second data set we analyzed was from *Ptch*(+/-);*Trp53*(+/+) mice (Ishida et al., 2010). DNA from SHH-MBs arising in these mice was analyzed by aCGH-based

DNA copy-number profiling. In line with our findings for human SHH-MB, the *Ptch*(+/-);*Trp53*(+/+) mice showed no gross chromosomal rearrangements, and thus no evidence for chromothripsis (Data S2). We calculated maximum copy-number state changes per chromosome for all samples in the study that did not receive irradiation, as follows: Three *Ptch*+/-; *Trp53*+/+ mice: chr4 (6 copy-number state changes); chr14 (2); chr2 (6).

2.6 Analysis of chromothripsis in several LFS-associated malignancies

We investigated a potential link between *TP53* mutations and chromothripsis in LFS-associated tumors other than medulloblastoma by collecting copy-number profile data for several tumors from LFS patients, and inferring chromothripsis in the copy-number profiles using the criteria described in the Experimental Procedures. Out of eleven tumor samples assessed, the single neuroblastoma sample, the single glioblastoma sample, one of four adrenocortical carcinoma (ACC) samples, and one of two embryonal rhabdomyosarcoma (RMS) samples displayed rearrangements consistent with chromothripsis (Table S2).

2.7 Analysis of rearrangements in tumors with somatic *TP53* mutations

We sought to evaluate the patterns of genomic rearrangements occurring in several tumors that are known to commonly display *TP53* mutations. Thereby we focused on recent studies that measured copy-number alterations with high-resolution microarray-based platforms. Our analysis included three tumor entities – lung cancer, ovarian cancer, and leukemia: *i.e.*, we analyzed squamous cell lung cancer samples (Weiss et al., 2010) with a reported frequency of somatic *TP53* mutations of 28% (Nelson et al., 2005), ovarian cancer samples with a reported frequency of somatic *TP53* mutations of 97% (The Cancer Genome Atlas Research Network, 2011), and AML samples, 15% of which have been reported to harbor somatic *TP53* mutations (Parkin et al., 2010). We used level 3 (415K) ovarian cancer copy-number profile data deposited without access restrictions.

Most of the ovarian cancer and lung cancer genomes appeared highly rearranged, with patterns of genomic rearrangement that were strikingly different from the patterns of rearrangement we observed in SHH-MBs cases with chromothripsis. Rather than displaying regular switches between copy-number states – a hallmark of chromothripsis – the rearrangements we observed in ovarian cancer and lung cancer samples displayed a wide diversity of different copy-number states (up to a dozen on a single chromosome). This pattern indicates that the ovarian cancer and lung cancer samples mostly rearranged in subsequent, progressive steps in evolving tumor subpopulations (Data S2). Nonetheless, ~5% of the lung cancer samples (an example is in Data S2) and a

similar rate of ovarian cancer cases ($\geq 5\%$) showed massive rearrangements that appeared to be not consistent with a classical progressive rearrangement model, and rather were consistent with the occurrence of chromothripsis. The AML samples are discussed in the following section.

2.8 Analysis of chromothripsis in AML

We obtained compelling evidence for an association between somatically acquired *TP53* mutations and chromothripsis in AML, by examining SNP array-based copy-number profiles in a cohort of 114 AML patients (Parkin et al., 2010). We inferred rearrangements resulting from chromothripsis as described in the Experimental Procedures section. One array was excluded due to high noise level (Table S5). We compared cases with wildtype *TP53* (*i.e.*, *TP53*^{+/+} genotype) to cases harboring somatically acquired *TP53* mutations. We identified rearrangements consistent with chromothripsis in 8/17 AML cases harboring somatic *TP53* mutations, whereas only 1/91 wildtype *TP53* cases showed rearrangements classifiable as chromothripsis (a highly significant association with $p=5.7 \times 10^{-7}$; Fisher's exact test). Furthermore, three out of five cases with *TP53*^{+/-} genotype (homozygous *TP53* deletion) showed chromothripsis.

We confirmed the link between *TP53* and chromothripsis in AML also in a second, previously unpublished, cohort of 311 patients. The patients in this cohort were pre-grouped into 56 "complex" karyotype AMLs based on standard cytogenetics (Haferlach et al., 2008), and 255 cases without "complex" karyotype. The *TP53* mutational status was evaluated by capillary sequencing in the "complex" karyotype AMLs, where 26 (46%) cases showed *TP53* mutations, 10 (18%) showed a *TP53*^{+/-} genotype (hemizygous *TP53*-loss), and 20 (36%) displayed a *TP53*^{+/+} genotype (*i.e.*, wildtype). AMLs without "complex" karyotypes, as defined by cytogenetics, were previously reported to have an extremely low incidence of *TP53* mutations (*i.e.*, $\sim 2\%$ (Haferlach et al., 2008)) and were thus not re-evaluated for *TP53* status studying this cohort. We did not identify any case with chromothripsis amongst the "non-complex" karyotype AMLs (0/255), but detected rearrangements consistent with chromothripsis in 22/56 ($\sim 40\%$) "complex" karyotype AMLs. Most of these 22 cases carried aberrations of *TP53*. Even when analyzing the "complex" karyotype group in isolation, the majority of cases with chromothripsis were observed amongst *TP53* mutant cases (13/26) compared to cases with *TP53*^{+/-} genotype (3/10) and cases with wildtype *TP53* (6/20). Furthermore, an additional 12% (3/26) of *TP53* mutated "complex" karyotype AMLs harbored rearrangements reminiscent of chromothripsis that fell just below our conservative scoring threshold (*i.e.*, harboring eight or nine switches between two or three copy-number states on an individual chromosome; Table S5). These findings further substantiate the link between somatically acquired *TP53* mutations and chromothripsis in AML.

The availability of clinical follow-up data enabled us to test for an association of chromothripsis with prognosis (Data S2) and age at diagnosis. We noticed an increased age of diagnosis in *TP53*-mutated AMLs with chromothripsis compared to *TP53*-mutated AMLs not showing chromothripsis in both of the aforementioned cohorts. This trend was significant when combining clinical data from both cohorts, with an increased age of diagnosis in *TP53*-mutated AMLs with chromothripsis (mean=67 yr), compared to *TP53*-mutated AMLs not showing chromothripsis (mean=58 yr; $P=0.04$; KS-test). Information on age of diagnosis is available in Table S5.

2.8 Cancer gene enrichment analysis

Lists of cancer-related genes associated with SHH-MB and AML were compiled as follows (see Table S5 for complete list). The SHH-MB cancer-related gene set (n=792) was constructed by first collecting cancer genes from a comprehensive list curated by The Cancer Genome Atlas (TCGA) research network (The Cancer Genome Atlas Research Network, 2008). We further added genes from a comprehensive exome sequencing survey of medulloblastoma (Parsons et al., 2011). When compiling the cancer-related genes associated with AML (n=1,665) we similarly started with the aforementioned TCGA gene set. Furthermore, since thus far there has been no published DNA sequencing-based study of AMLs at a similar scale as Parsons et al., we added recurrently gained and lost regions described in the largest survey of DNA copy-number alterations in AML (Walter et al., 2009). We included all genes that were affected by a significant number of focal gains or losses, as assessed by the GISTIC algorithm (Beroukhi et al., 2007; Walter et al., 2009).

3. Gene expression analysis

3.1 AMLs displaying chromothripsis do not show SHH pathway activation

To assess whether chromothripsis in AML occurs in the context of SHH pathway activation, we measured the expression of canonical SHH pathway genes in medulloblastoma and AML, using previously published gene expression profiling data for several medulloblastoma samples (Northcott et al., 2011) and for patients from the aforementioned second AML cohort. Gene expression profiling was performed using the Affymetrix Genechip Human Exon 1.0 ST Arrays. Data were normalized (rma-sketch) and analyzed at the gene level (CORE content).

Expression data, shown in boxplots available as Data S2 for several canonical SHH pathway genes, were separately analyzed for the four medulloblastoma subtypes, WNT (n = 8), SHH-MB (n = 33), Group C (n = 27), and Group D (n = 35), and for AML cases showing chromothripsis (n = 12) as well as such showing no chromothripsis (n = 33). The AML cases did not show SHH pathway activation (Data S2), displaying an expression level of these SHH pathway genes that was comparable to the level measured in non-SHH activated medulloblastoma subtypes (similar results were obtained for other SHH pathway genes). Furthermore, there was no difference in SHH pathway gene expression between chromothripsis and non-chromothripsis cases (Data S2).

Supplemental References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249.
- Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., and Durbin, R. (2011). Dindel: accurate indel calls from short-read data. *Genome Res* 21, 961-973.
- Beroukhi, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104, 20007-20012.
- Korbel, J.O., Abyzov, A., Mu, X.J., Carriero, N., Cayting, P., Zhang, Z., Snyder, M., and Gerstein, M.B. (2009). PEm: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol* 10, R23.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Nelson, H.H., Wilkojmen, M., Marsit, C.J., and Kelsey, K.T. (2005). TP53 mutation, allelism and survival in non-small cell lung cancer. *Carcinogenesis* 26, 1770-1773.
- Rausch, T., Emde, A.K., Weese, D., Doring, A., Notredame, C., and Reinert, K. (2008). Segment-based multiple sequence alignment. *Bioinformatics* 24, i187-192.
- The Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609-615.
- Walter, M.J., Payton, J.E., Ries, R.E., Shannon, W.D., Deshmukh, H., Zhao, Y., Baty, J., Heath, S., Westervelt, P., Watson, M.A., et al. (2009). Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc Natl Acad Sci U S A* 106, 12950-12955.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164.
- Weiss, J., Sos, M.L., Seidel, D., Peifer, M., Zander, T., Heuckmann, J.M., Ullrich, R.T., Menon, R., Maier, S., Soltermann, A., et al. (2010). Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med* 2, 62ra93.

Figure S1

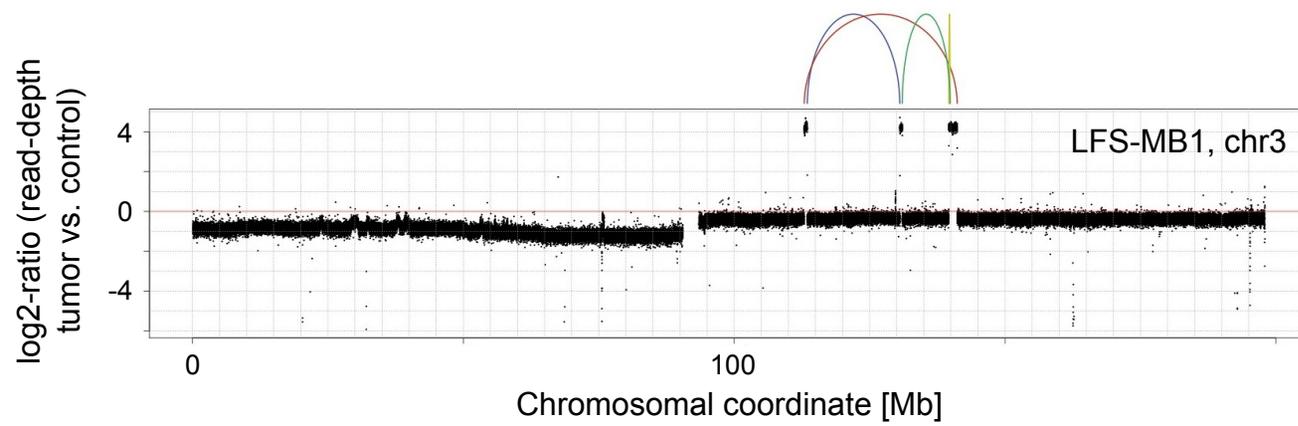
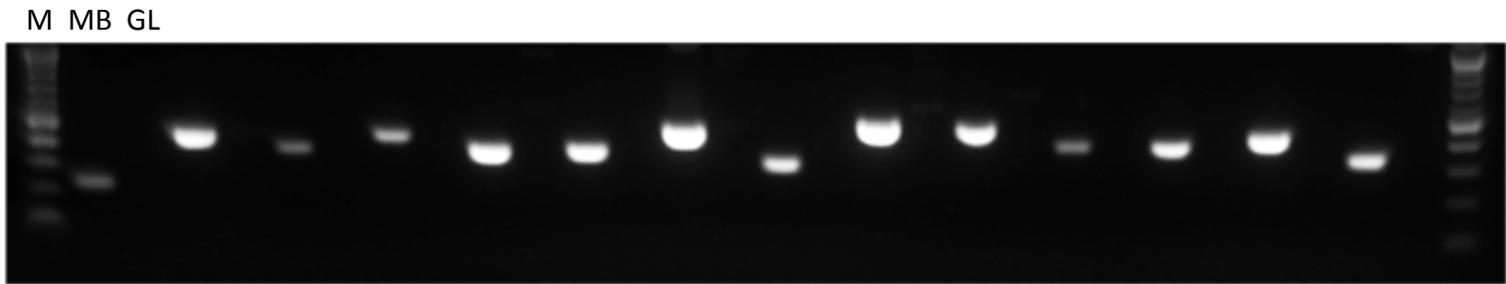
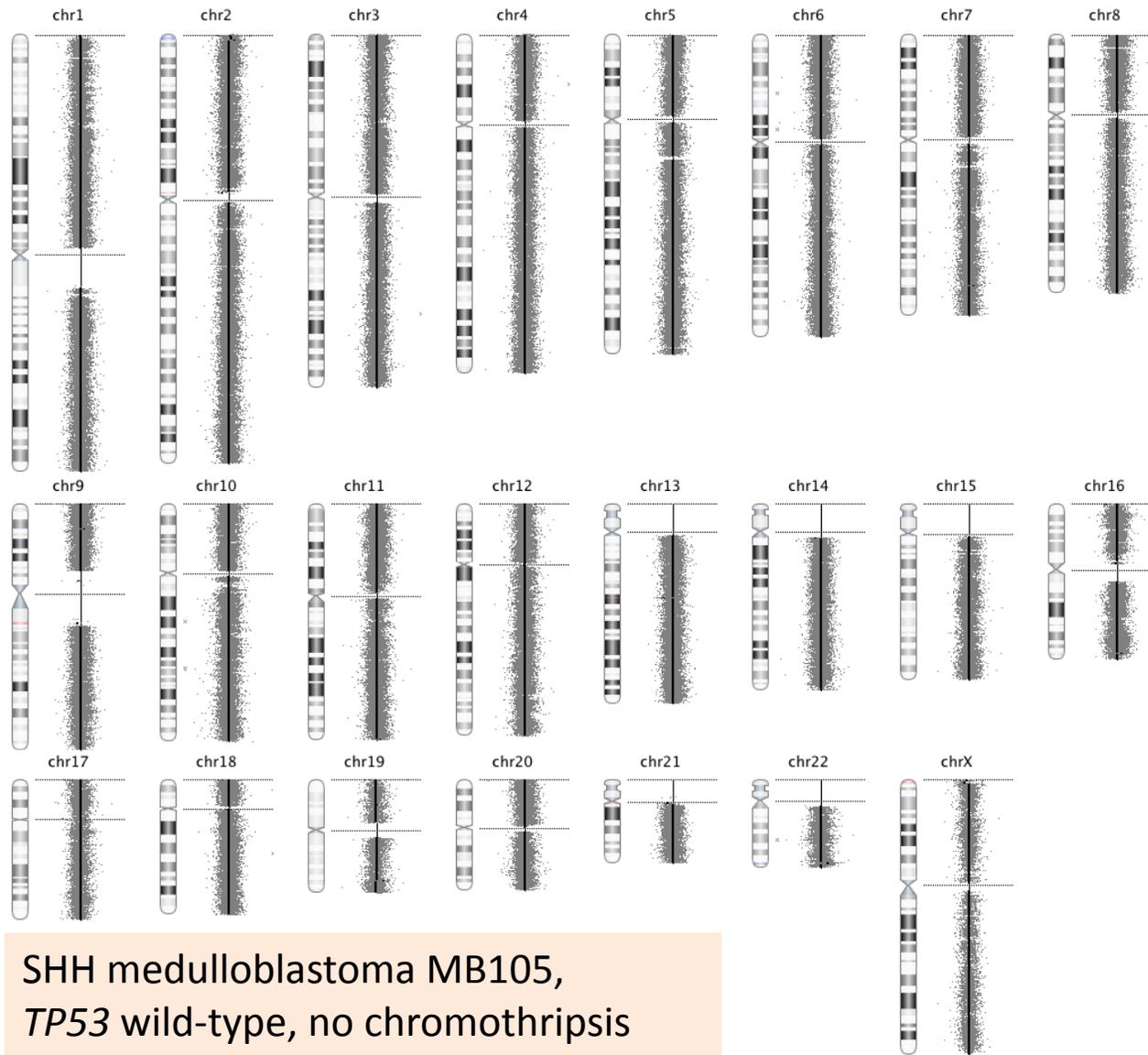


Figure S2

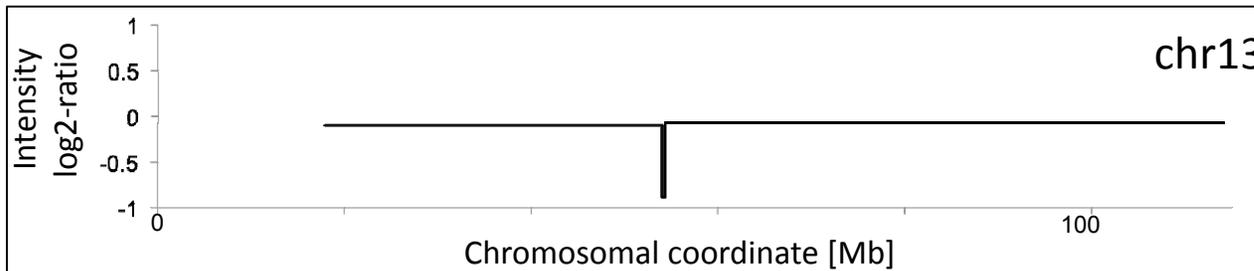


Sample: LFS-MB2

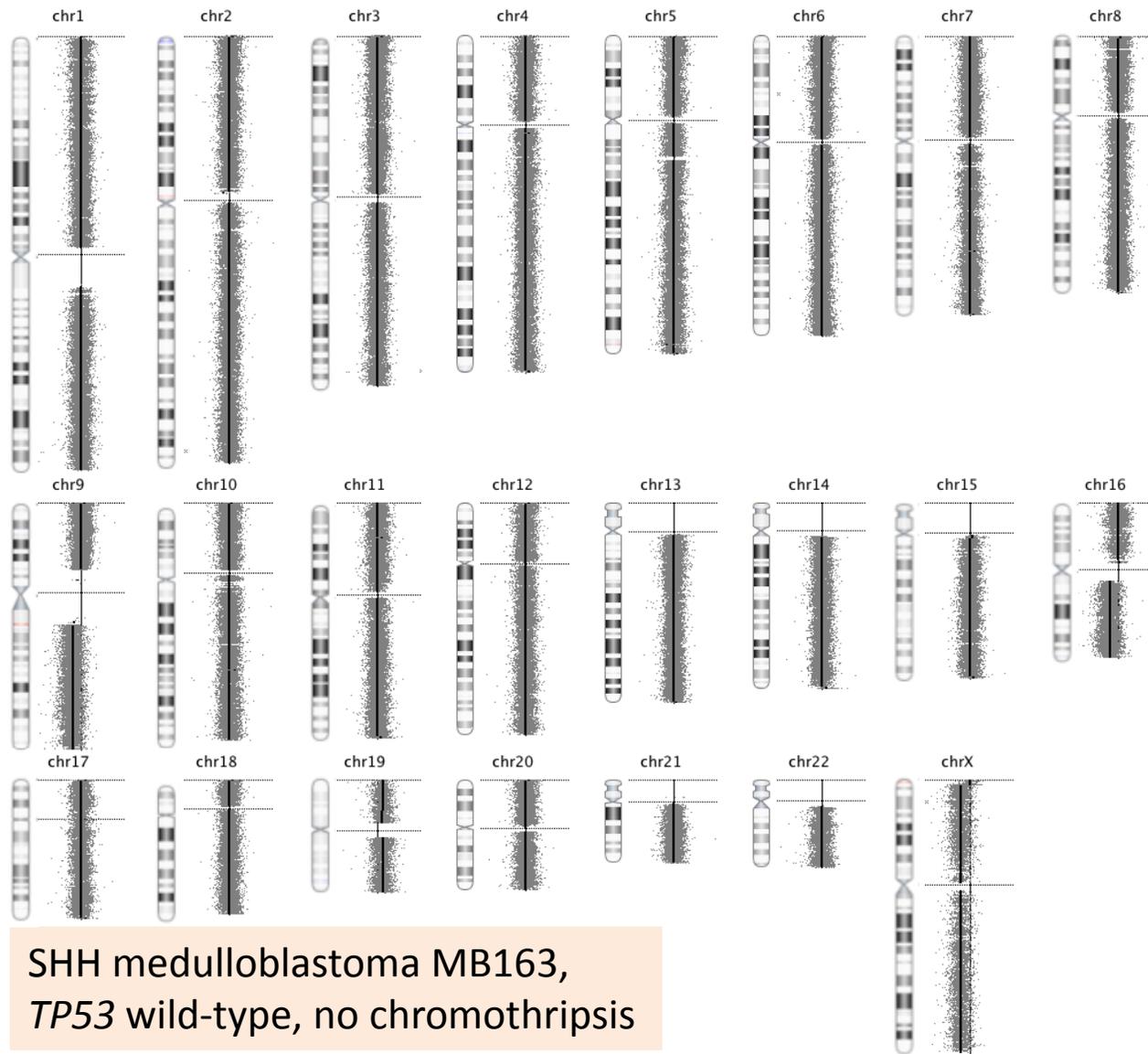
Data S1A



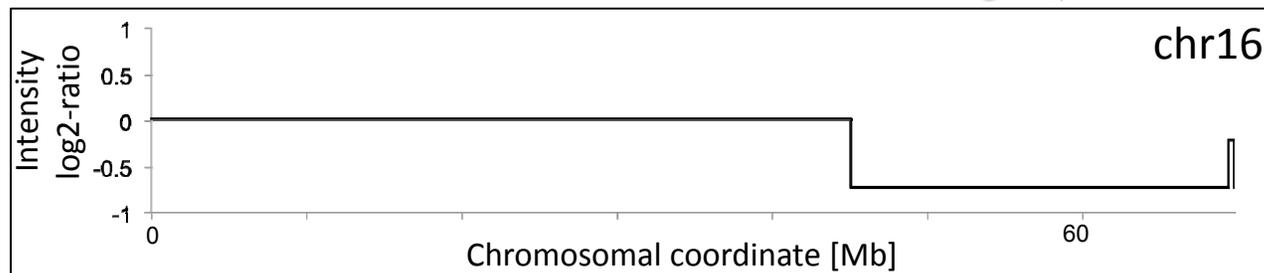
SHH medulloblastoma MB105,
TP53 wild-type, no chromothripsis



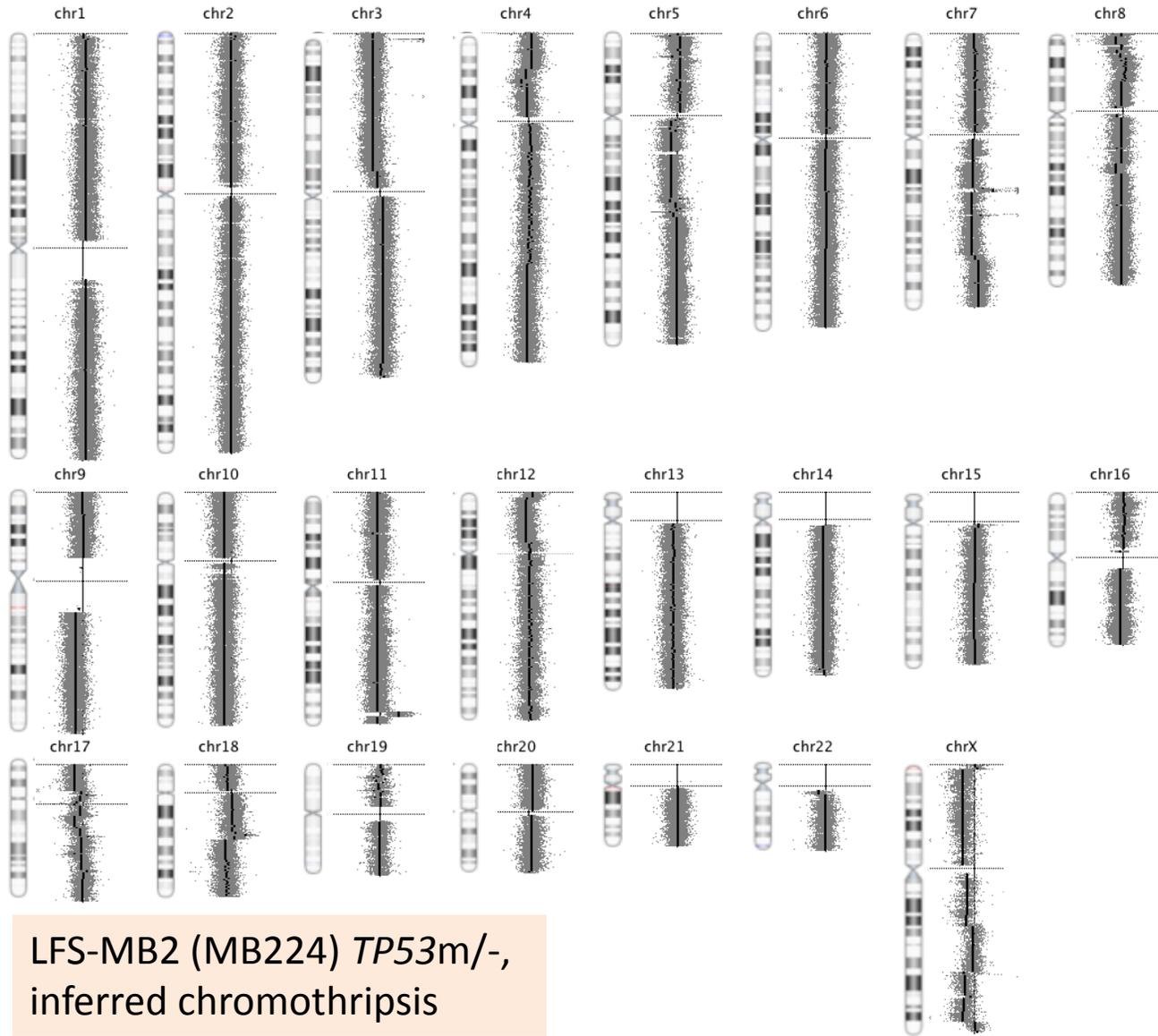
Data S1B



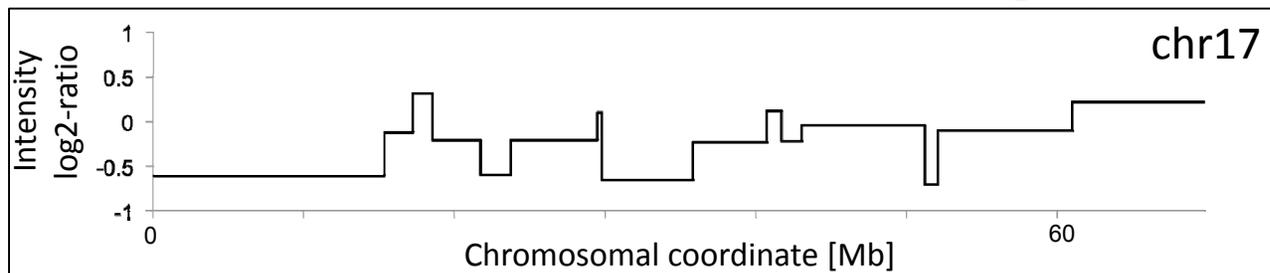
SHH medulloblastoma MB163,
TP53 wild-type, no chromothripsis



Data S1C



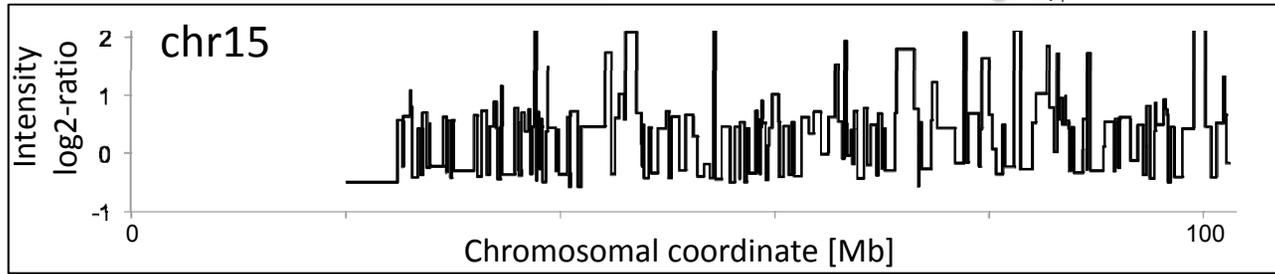
LFS-MB2 (MB224) *TP53m/-*,
inferred chromothripsis



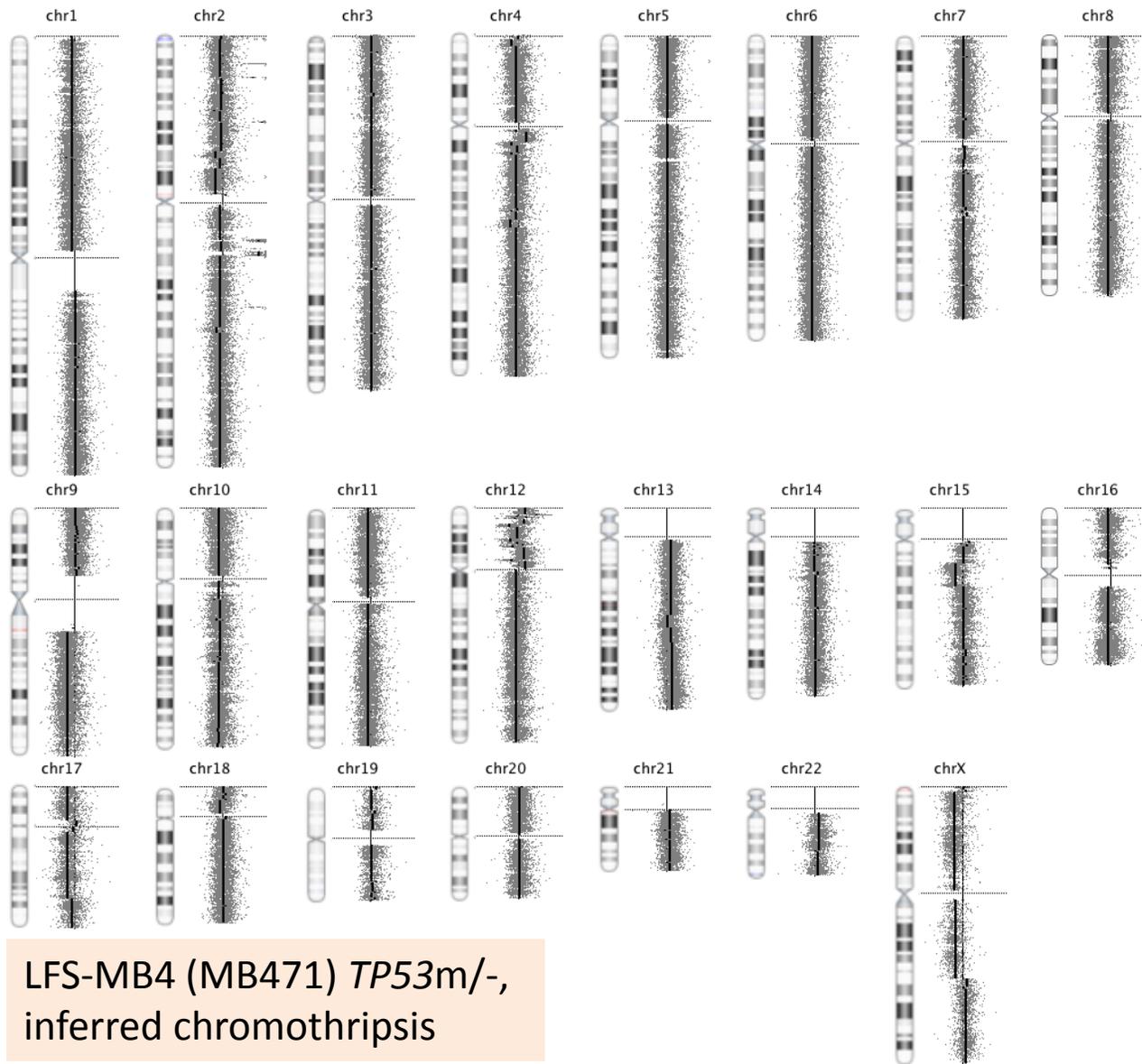
Data S1D



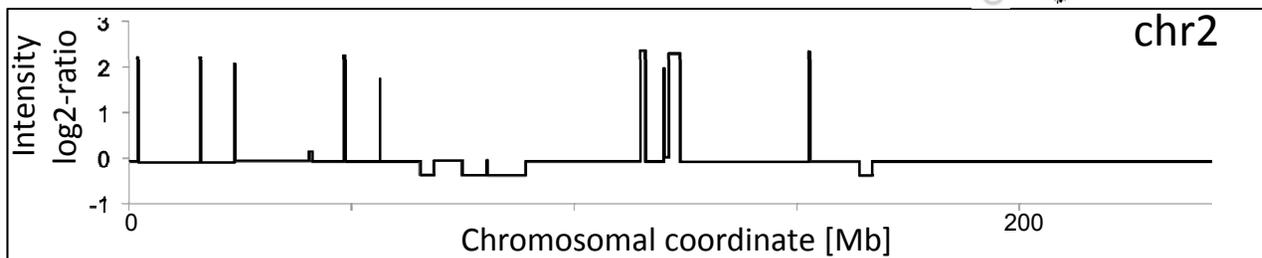
LFS-MB3 (MB460) *TP53*m^{-/-},
inferred chromothripsis



Data S1E

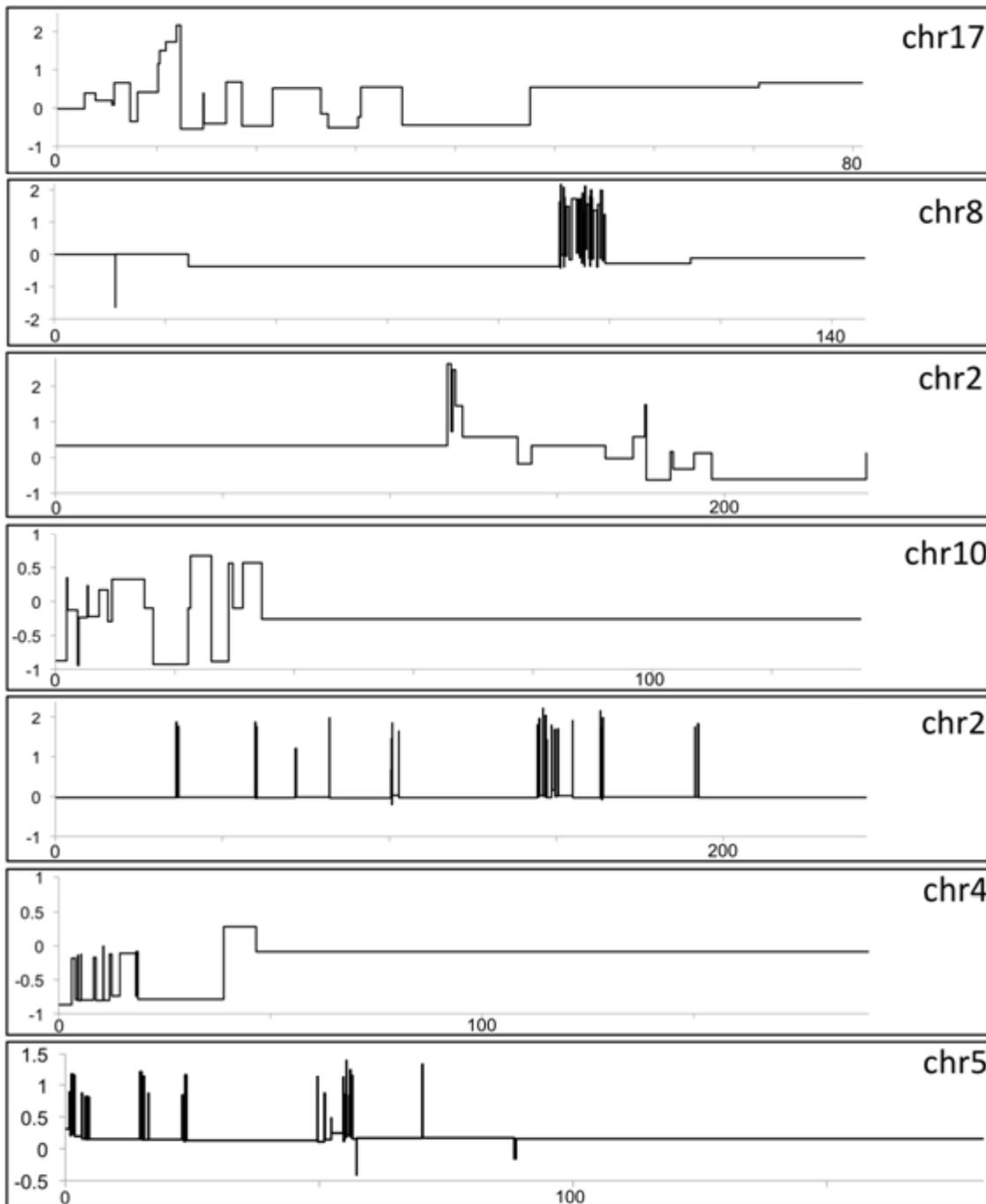


LFS-MB4 (MB471) *TP53*m^{-/-},
inferred chromothripsis



Data S1F

Intensity log₂-ratio



LFS-MB5 (MB391)
TP53+/m, inferred
chromothripsis

MB486
TP53m/-, inferred
chromothripsis

MB7
TP53m/-, inferred
chromothripsis

MB212
TP53m/-, inferred
chromothripsis

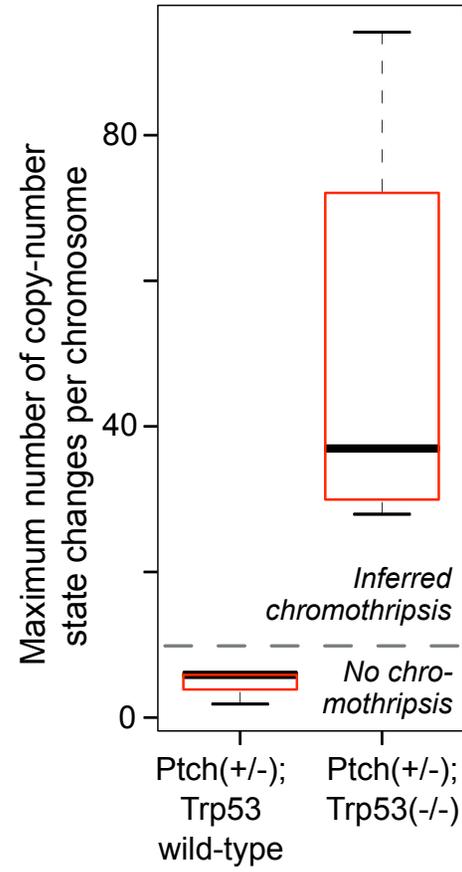
LFS-MB6 (MB387)
TP53+/m, inferred
chromothripsis

MB131
TP53m/-, inferred
chromothripsis

MB2034
TP53m/- (somatic),
inferred
chromothripsis

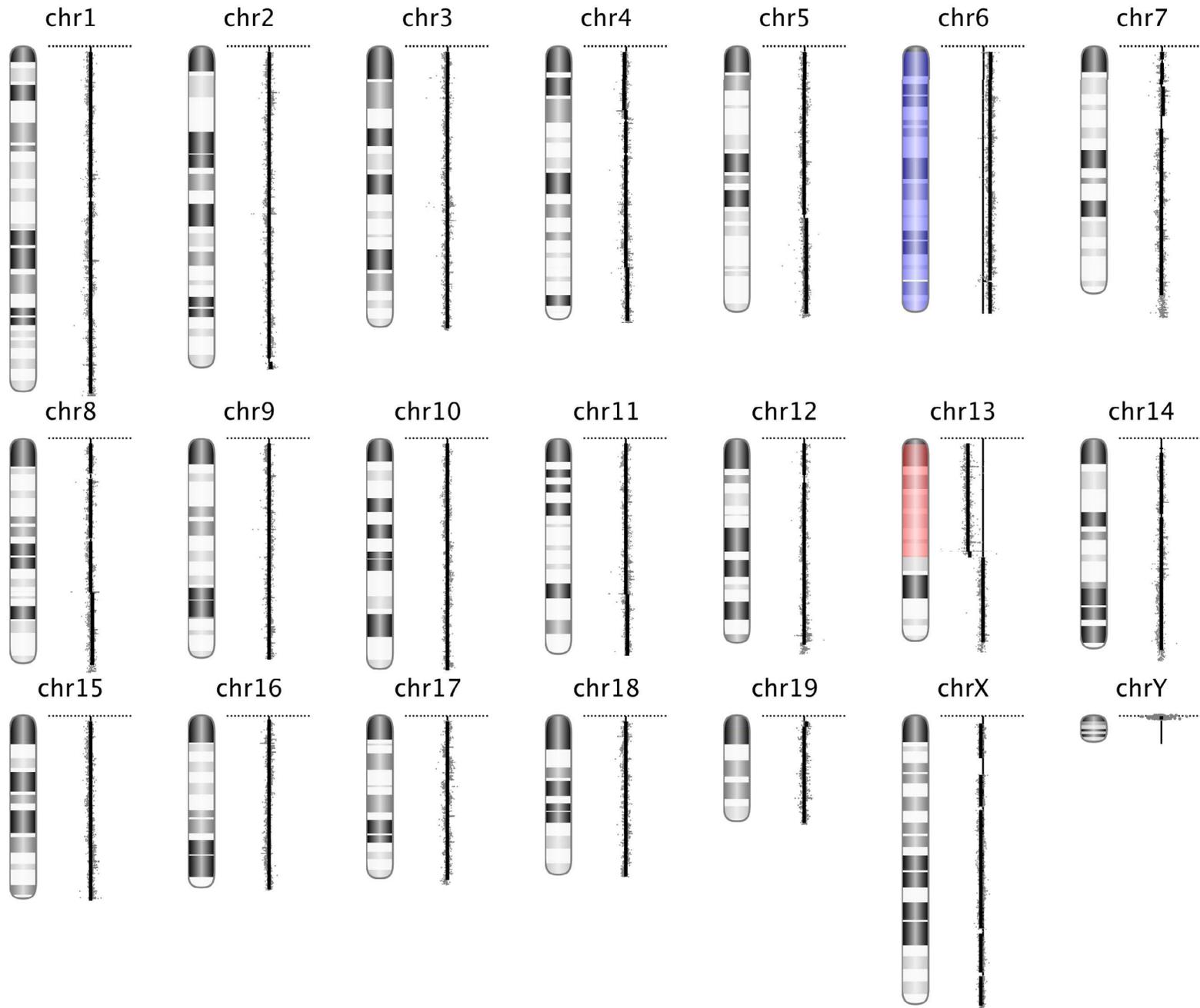
Chromosomal coordinate

Data S2A



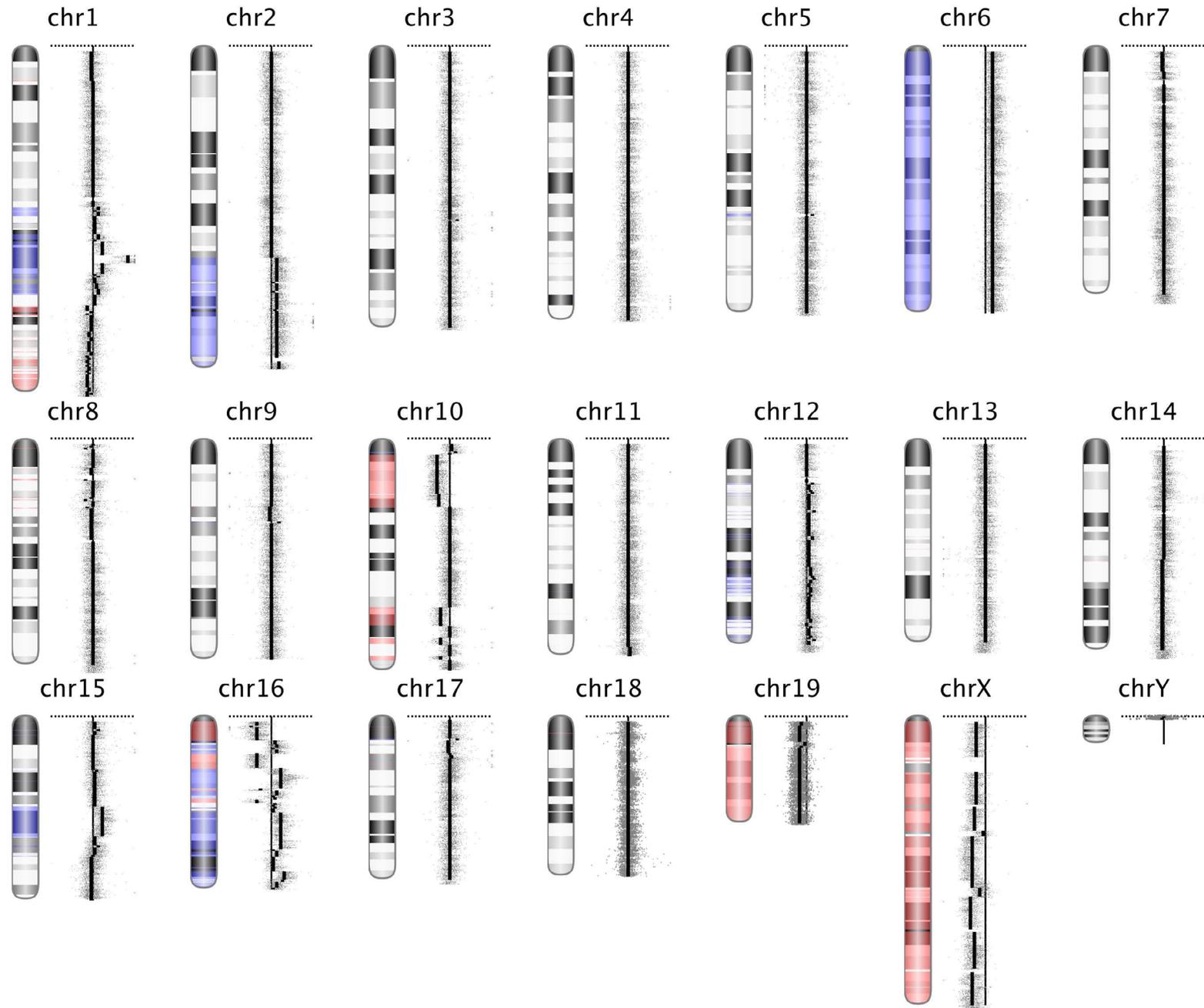
GSM480975, Ptch+/-;Trp53+/, no chromothripsis

Data S2B



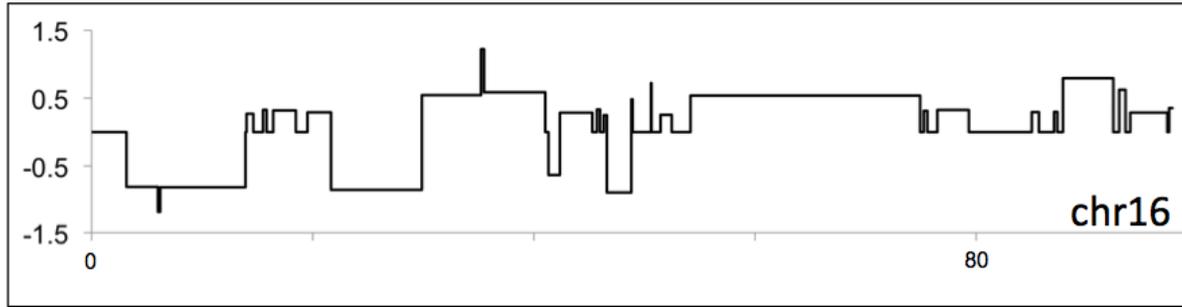
GSM490325, Ptch+/-;Trp53-/-; inferred chromothripsis

Data S2C



Data S2D

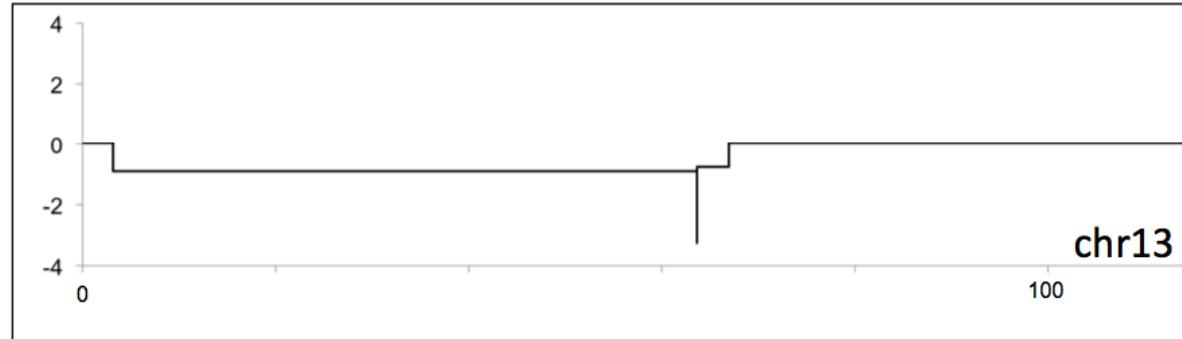
Intensity log2-ratio



GSM490325
Ptch+/-;Trp53-/-,
inferred
chromothripsis

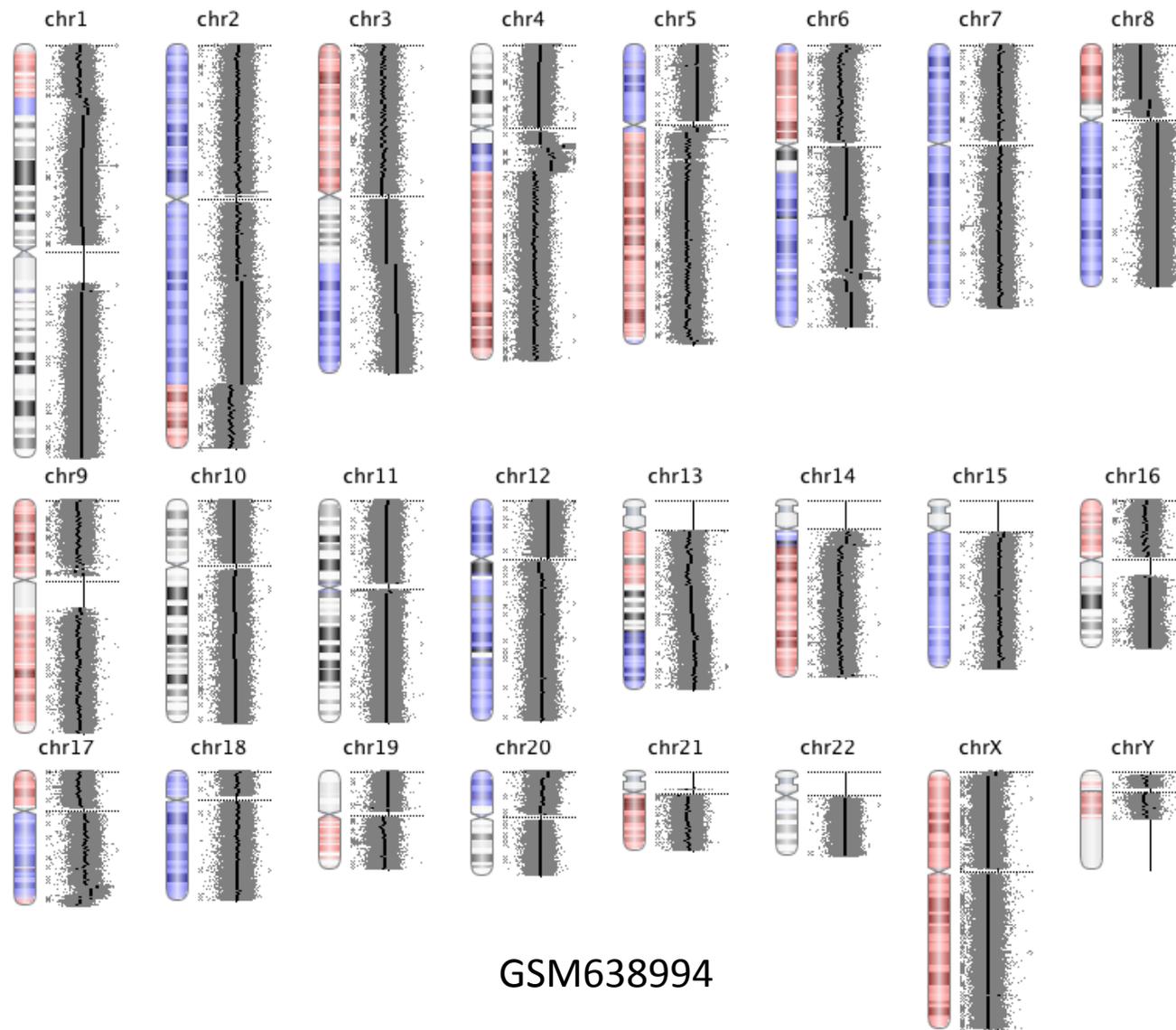


GSM490325
Ptch+/-;Trp53-/-,
inferred
chromothripsis

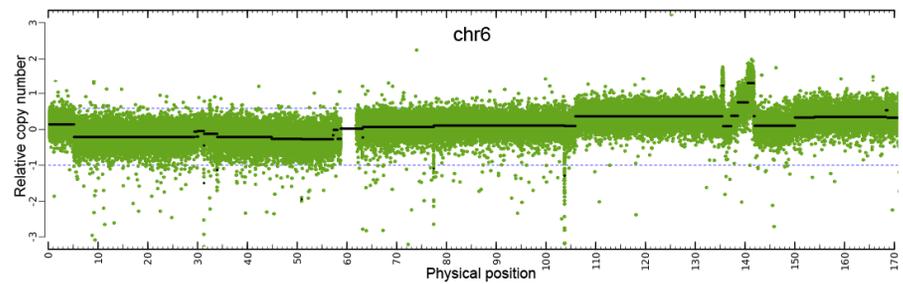
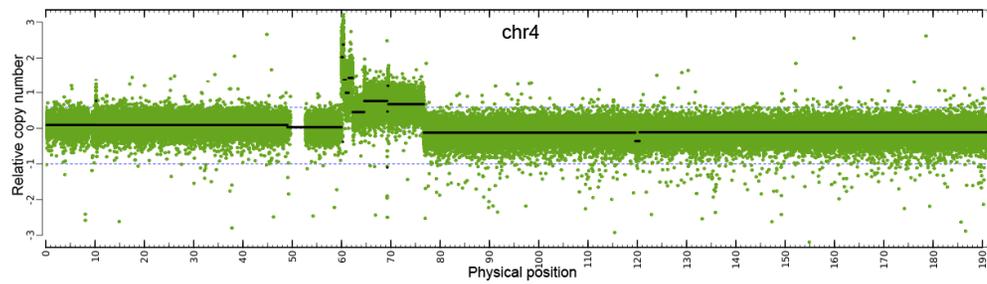


GSM480975
Ptch+/-;Trp53+/-,
no chromothripsis

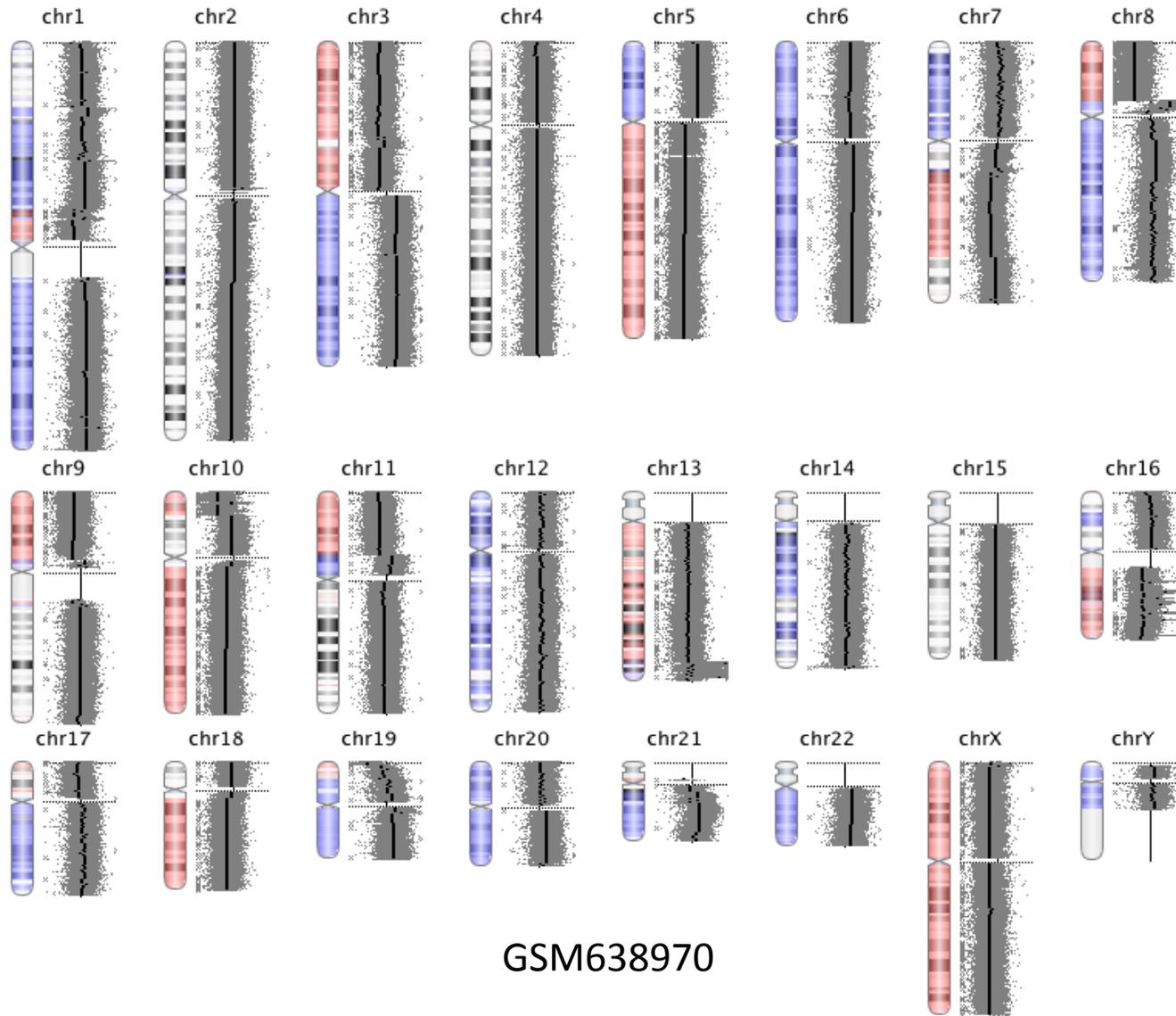
Data S2E



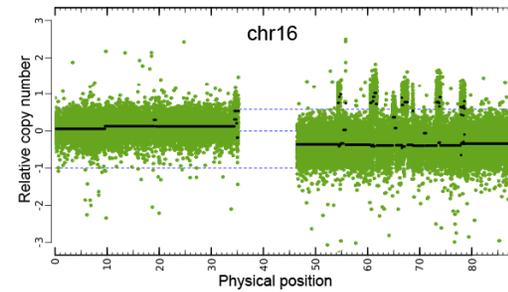
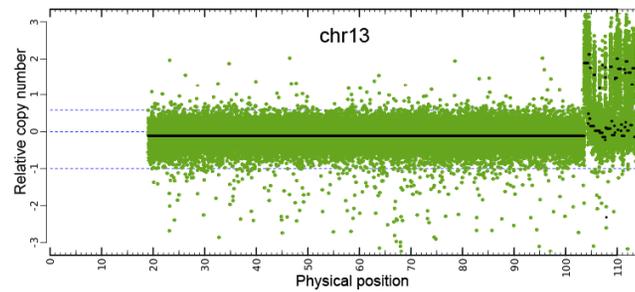
GSM638994



Data S2F

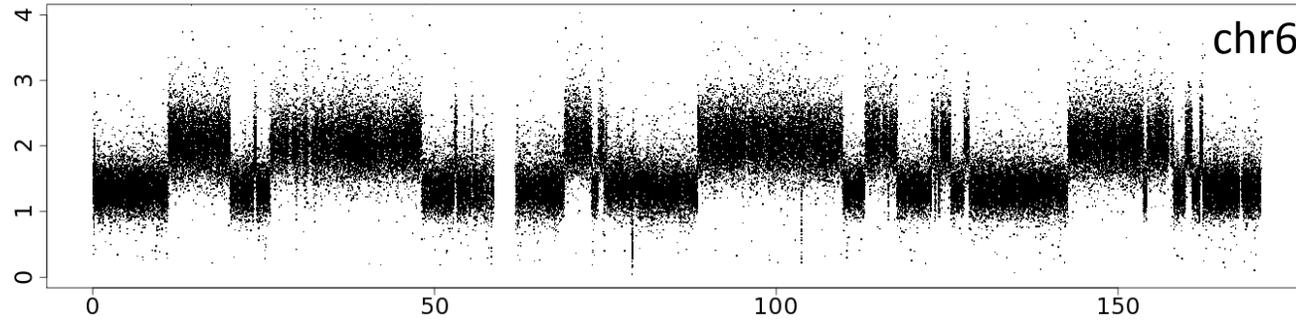


GSM638970

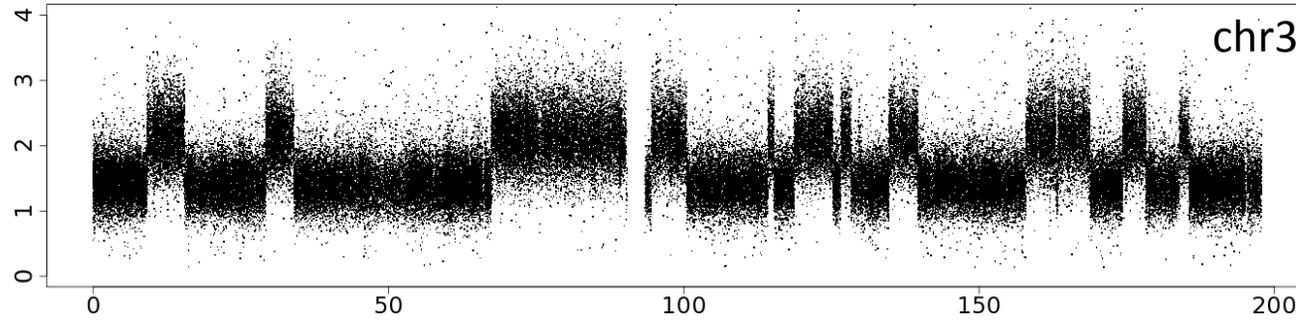


Data S2G

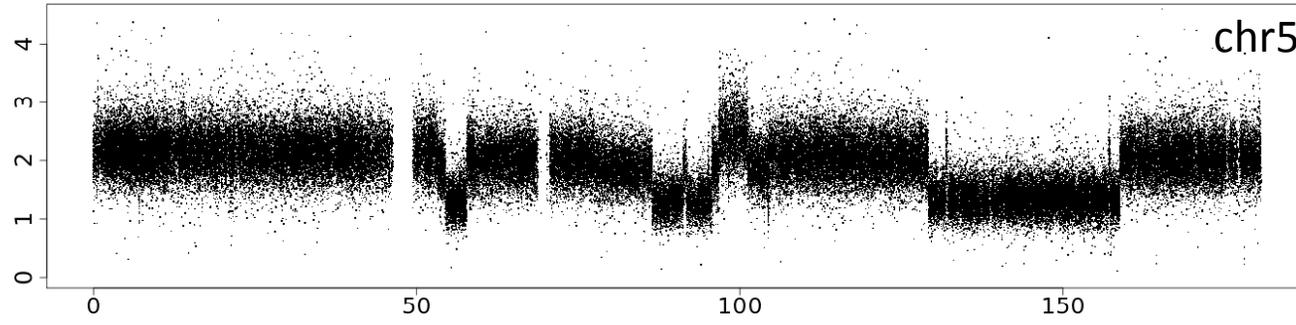
Inferred copy number



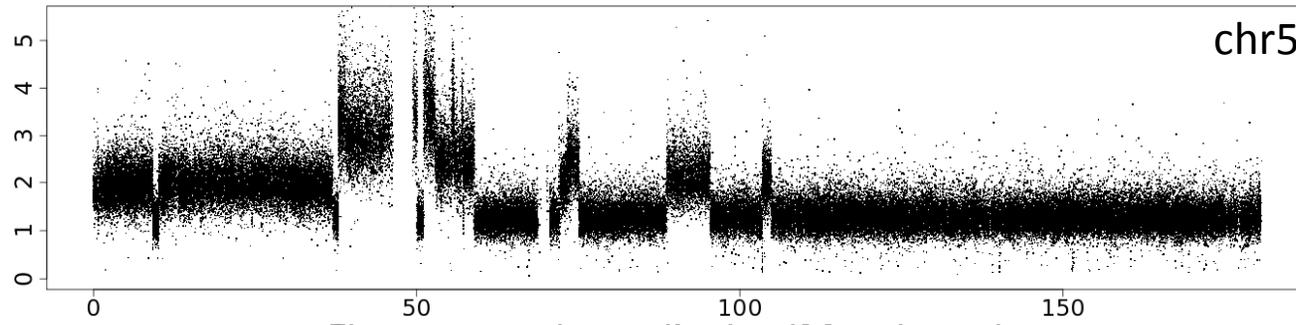
AML (GSM630977)
*TP53*m/-, inferred
chromothripsis



AML (GSM631073)
*TP53*m/-, inferred
chromothripsis



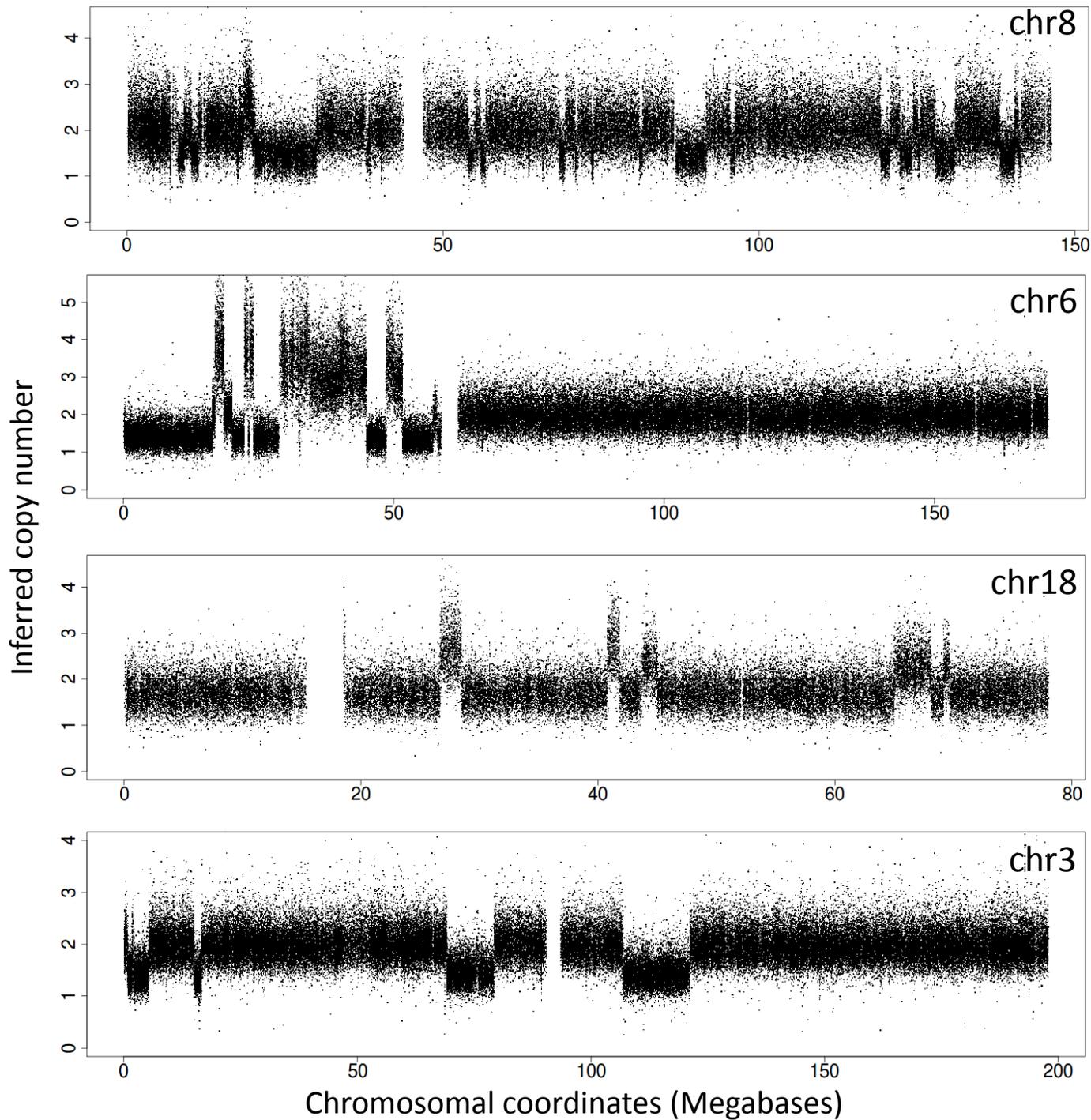
AML (GSM631075)
*TP53*m/-, inferred
chromothripsis



AML (GSM631103)
*TP53*m/-, inferred
chromothripsis

Chromosomal coordinates (Megabases)

Data S2H



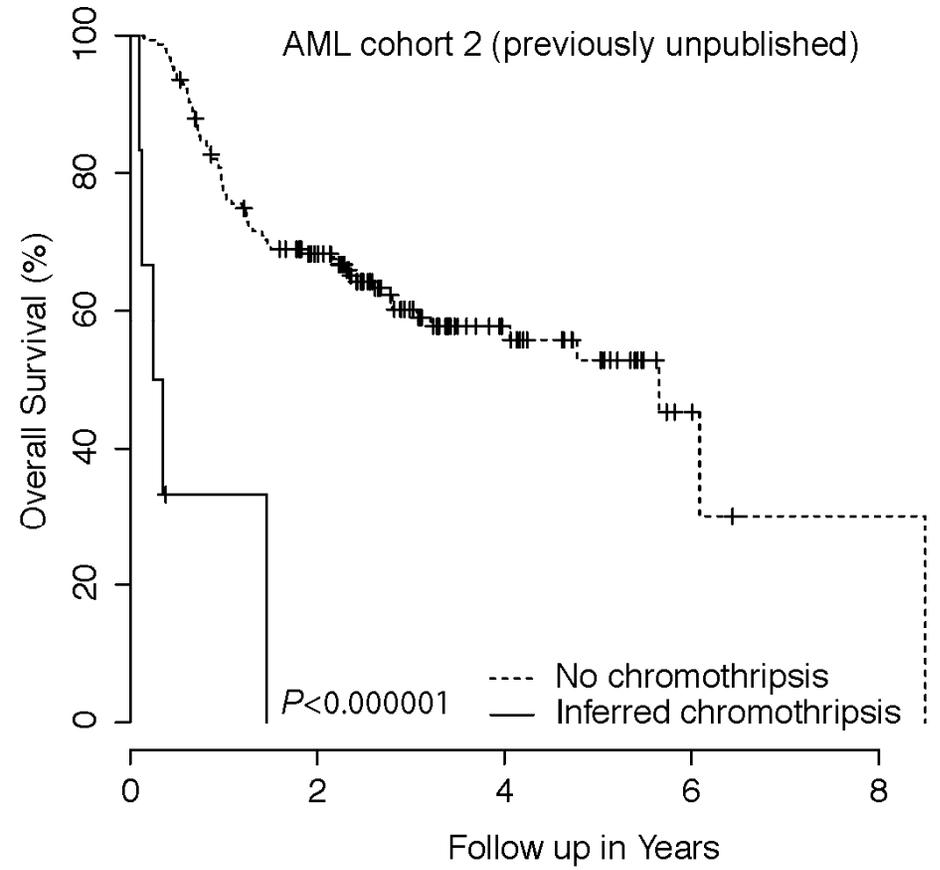
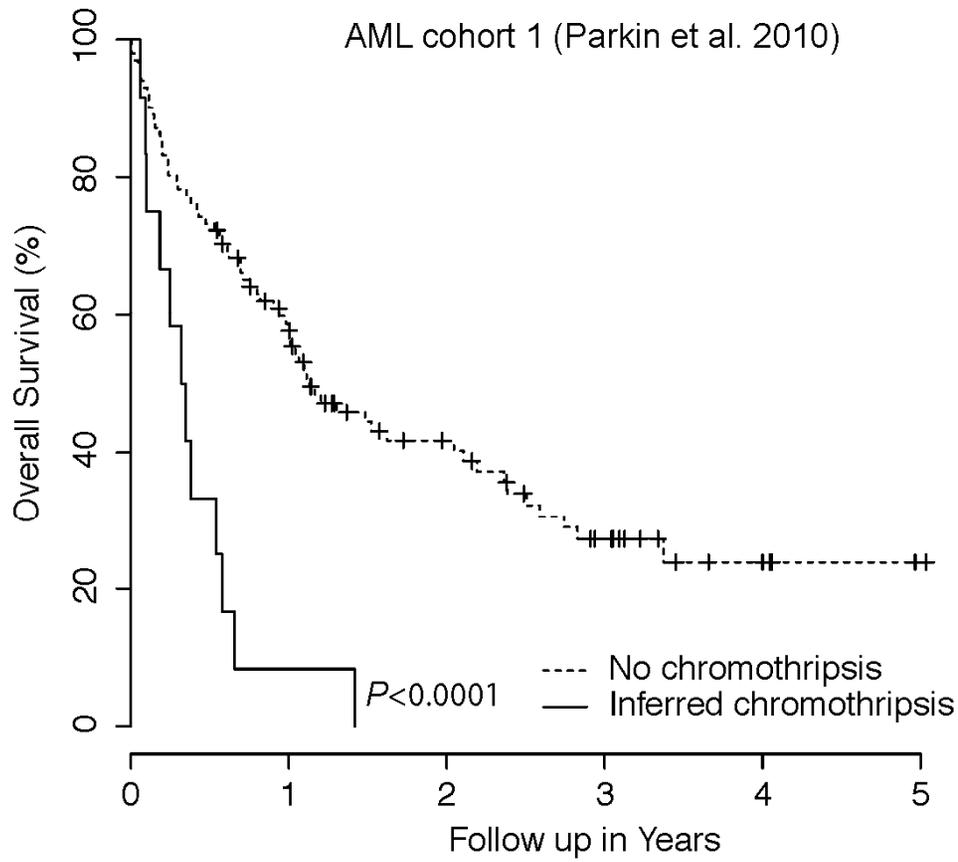
AML (08KM1077)
TP53+/*m*, inferred
chromothripsis

AML (08KM0227)
TP53+/*m*, inferred
chromothripsis

AML (08PB3645)
TP53+/*m*, inferred
chromothripsis

AML (08KM3747)
*TP53**m*/*-*, inferred
chromothripsis

Data S2I



Data S2J

