Supplementary Information

Material and Method

Next Generation Sequencing

For the three fresh frozen (FF) recurrence samples and the germline blood sample, the standard TruSeq DNA PCR-Free library preparation was used, while the two samples from formalin fixed paraffin embedded (FFPE) blocks followed a TruSeq DNA Nano protocol with modifications optimized for FFPE derived DNA. This resulted in 4 FF libraries with an average of 298 bp insert size and 2 FFPE libraries with fragments of 163 bp. Whole genome sequencing was performed on the Illumina HiSeq 2500 instrument, generating 100 bp pair-end sequence reads using HiSeq SBS V3 clustering and sequencing chemistry in High output and Rapid run modes. Reads were aligned to the reference genome (hg19, GRCh37.1) using bowtie2[1] with the *sensitive-local* flag and variant calling using ISAAC[2] and Strelka[3] was carried out on all sequencing data in order to monitor standard quality control metrics. The quality of DNA extracted from FFPE samples is variable according to a number of process steps prior to sequencing including fixation protocols and extraction methods. If there is sufficient amount of good quality DNA from such extraction it is possible to obtain good quality whole-genome sequence data, such as the two samples analysed here. The FFPE sample achieved sequencing quality metrics that supported genome-wide variant calling. However somatic SNVs in those FFPE samples showed an unusual number of low frequency somatic variants, clustered at particular loci, and these variants were in linkage disequilibrium on a low number of highly mutated reads at these loci. Further investigations showed that the majority of these loci where the regions of the human genome that are highly conserved among placental mammals, these conserved regions encompass approximating 3% of the human genome. A subset of these aberrant reads was assembled into mini-contigs and searched against the NCBI non-redundant DNA database (nr) using blastn. These mini-contigs were found to have 100% sequence identity to the *Mus musculus* reference genome but only around 90% sequence identity to Homo sapiens. As contamination of mouse genomic DNA was suspected within the FFPE DNA libraries all the reads were re-

aligned to a combined GRCh37.1 and MGSCv37 reference genomes and those reads aligned to the mouse portion of the reference genome were discarded in all downstream analysis.

Mutation Calling

Variant calling was performed as described in[4]. In brief, the "somatic" tool from VarScan 2 v2.33[5] was implemented on matched tumour regions and germ-line. Variants were only accepted if present in ≥ 5% of reads in at least one tumor region and present with ≤2 reads in germ-line and ≥ 2 reads in a tumor region.

Structural Variation Analysis

Inter and intra-chromosomal translocations, inversions, and large (>=10kb) insertions/deletions were identified in WGS data using CREST (v1.0.1)[6], after pre-processing BAMs with the GATK IndelRealigner (v2.1-13-g1706365). To reduce the false positive rate, breakpoint junctions of putative structural variants (SVs) were de novo assembled using TIGRA (v0.3.7)[7] in germline and tumour BAMs individually, and aligned to hg19 using BLAT (v35); breakpoints that were re-constituted from tumour BAMs only were considered valid. Furthermore, SVs with breakpoints mapping to low-copy repeats were removed. In order to assign SVs to specific tumour regions, soft-clipped and discordant paired-end reads in the SV's vicinity were collected, and an SV was called present in a region if at least one of these read-level events was consistent with the SV.

Copy number analysis

Integer copy numbers were computed through the Sequenza R package [8] using a binning size of 200 nt on the whole genome sequencing data. FACS was performed on the recurrence cohort to determine the DNA-index of the specimens, which was used as ploidy measure to determine the copy number profile. To compare the haplotypic origin of shared somatic copy number aberrations (SCNA) between tumour regions, allele frequencies (AF) for each heterozygous SNP were estimated[9]. SNP alleles were classed as "major" (AF > 0.5) or "minor" (AF < 0.5) in the grade IV region, assuming that the "major" allele

represents the higher-copy number haplotype where applicable. Patterns of major/minor allele frequency in the recurrence regions were then compared to these reference alleles; a SCNA with an inverted major/minor allele frequency distribution was interpreted as a recurrent (or secondary) SCNA affecting the alternate haplotype.

Double minute haplotype analysis

To compare the haplotypic origin of shared SCNAs between tumor regions, allele frequencies for each heterozygous SNP were estimated[9]. SNP alleles were classed as "major" (AF > 0.5) or "minor" (AF < 0.5) in the G2 region, assuming that the "major" allele represents the higher-copy number haplotype where applicable. Patterns of major/minor allele frequency in the G4 were then compared to these reference regions; a SCNA with an inverted major/minor allele frequency distribution was interpreted as a recurrent (or secondary) SCNA affecting the alternate haplotype.

Genome doubling and mutations timing

Mutations could be timed based on whether they occurred on the trunk or branches of the tumour phylogenetic tree, and in the case of genome doubling whether they occurred before or after the event. In brief, mutations were timed relative to genome doubling and amplification events based on their mutation copy number. We calculated the mutation copy number, n_{mut}, describing the fraction of tumour cells carrying a given mutation multiplied by the number of chromosomal copies at that locus using the following formula:

$$n_{mut} = VAF \frac{1}{p}[pCN_t + CN_n(1-p)]$$

where VAF corresponds to the variant allele frequency at the mutated base, and p, CN_t, CN_n are respectively the tumour purity, the tumour locus specific copy number, and the normal locus specific copy number. Any mutation with a mutation copy number >=2 was estimated to have occurred prior to a doubling or amplification event where as all mutations with a copy number of 1 were classified as occurring after the doubling or amplification event.

Thus, overall we found four types of mutations in the grade II and grade IV tumour regions:

- Truncal mutations: mutations found both grade II and the grade IV regions.
- Branched mutations:
    - G2 mutations – mutations detected only in the grade II regions
    - G4 early mutations – mutations detected only in the grade IV region exhibiting a mutation copy number >= 2
    - G4 late mutations – mutations detected only in the grade IV region exhibiting a mutation copy number = 1.

Table S1: Key sequencing statistics

Sample	Total Reads (Paired)	Median Coverage	Median Insert size	Tumour Purity (%)
G2	1,247,996,776	34	163	44
G4	1,286,353,218	35	163	80
A1	1,238,063,272	31	298	13
A2	1,242,143,531	28	298	10
A3	1,269,214,738	33	298	15

Table S2: Amplifications and deletions summary

Grade II	Grade IV	Recurrence
Focal amplifications		
4q12 (52.7-57.9) 12q14 (58-58.3)	4q12 (52.7-53.0) 4q12 (54.1-55.6) 4q21 (81.1-84.1) 12q14 (58-58.6)	4q12 (52.7-53.0) 4q12 (54.3-55.6) 12q14 (58-58.6)
LOH		
17p	17p	17p
Gains		
4p, 7q, 10q	4p, 7q, 10q, 6p, 19p, 20p	4p, 7q, 10q, 6p, 19p, 20p
Losses		
4q, 5q, 11p	4q, 5q, 11p, 10q, 12q, 13, 16q, 17q	4q, 11p

Bibliography

1. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012; 9(4):357–9.

2. Raczy C, Petrovski R, Saunders CT et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 2013; 29(16):2041–3.

3. Saunders CT, Wong WSW, Swamy S et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012; 28(14):1811–7.

4. De Bruin EC, McGranahan N, Mitter R et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science (80-.). 2014; 346(6206):251–256.

5. Koboldt DC, Zhang Q, Larson DE et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22(3):568–76.

6. Wang J, Mullighan CG, Easton J et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods 2011; 8(8):652–4.

7. Chen K, Chen L, Fan X et al. TIGRA: a targeted iterative graph routing assembler for breakpoint assembly. Genome Res. 2014; 24(2):310–7.

8. Favero F, Joshi T, Marquard AM et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann. Oncol. 2014:mdu479–.

9. Letouzé E, Sow A, Petel F et al. Identity by descent mapping of founder mutations in cancer using high-resolution tumor SNP data. PLoS One 2012; 7(5):e35897.

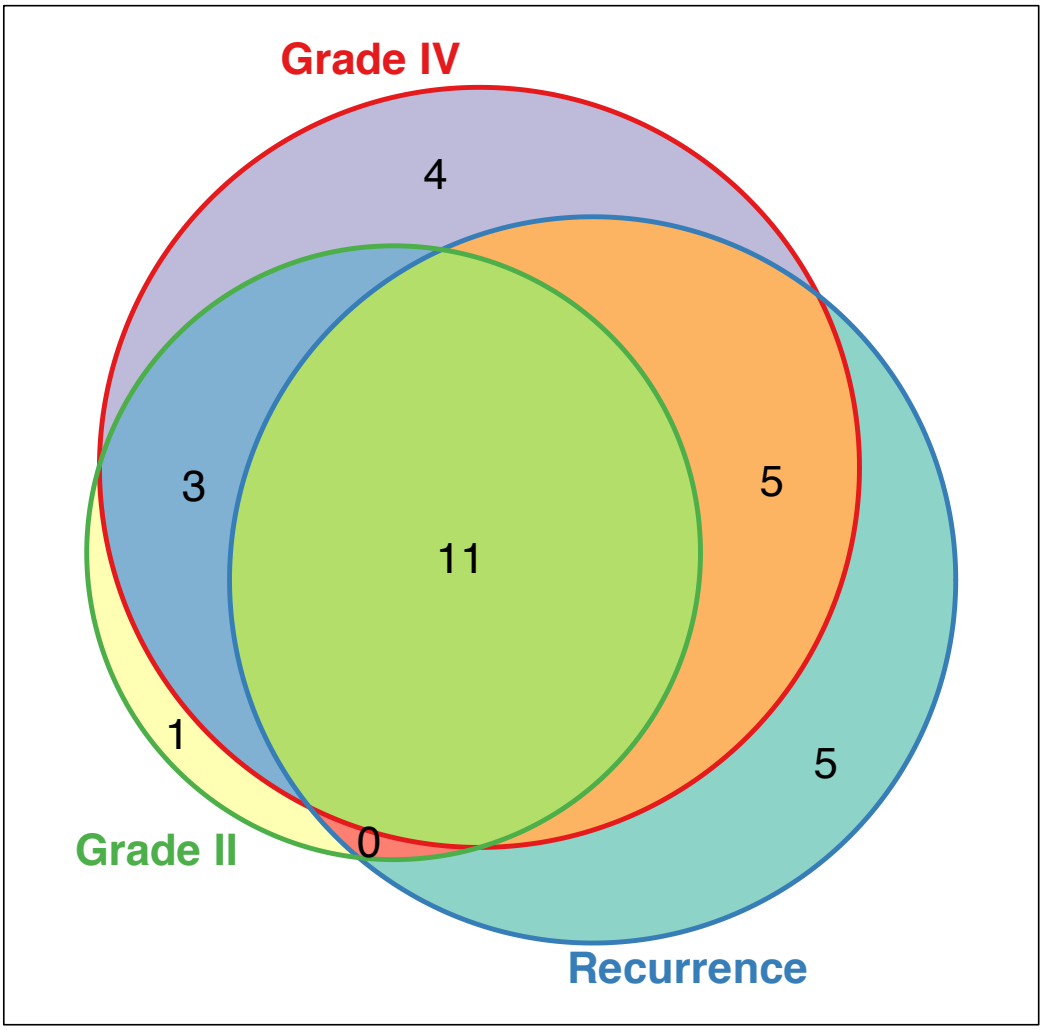


Figure S1: Euler-Venn diagram of coding mutations

A simple Euler-Venn diagram displaying the overlaps of coding mutations in the joint recurrence cohort and the grade II and grade IV samples.

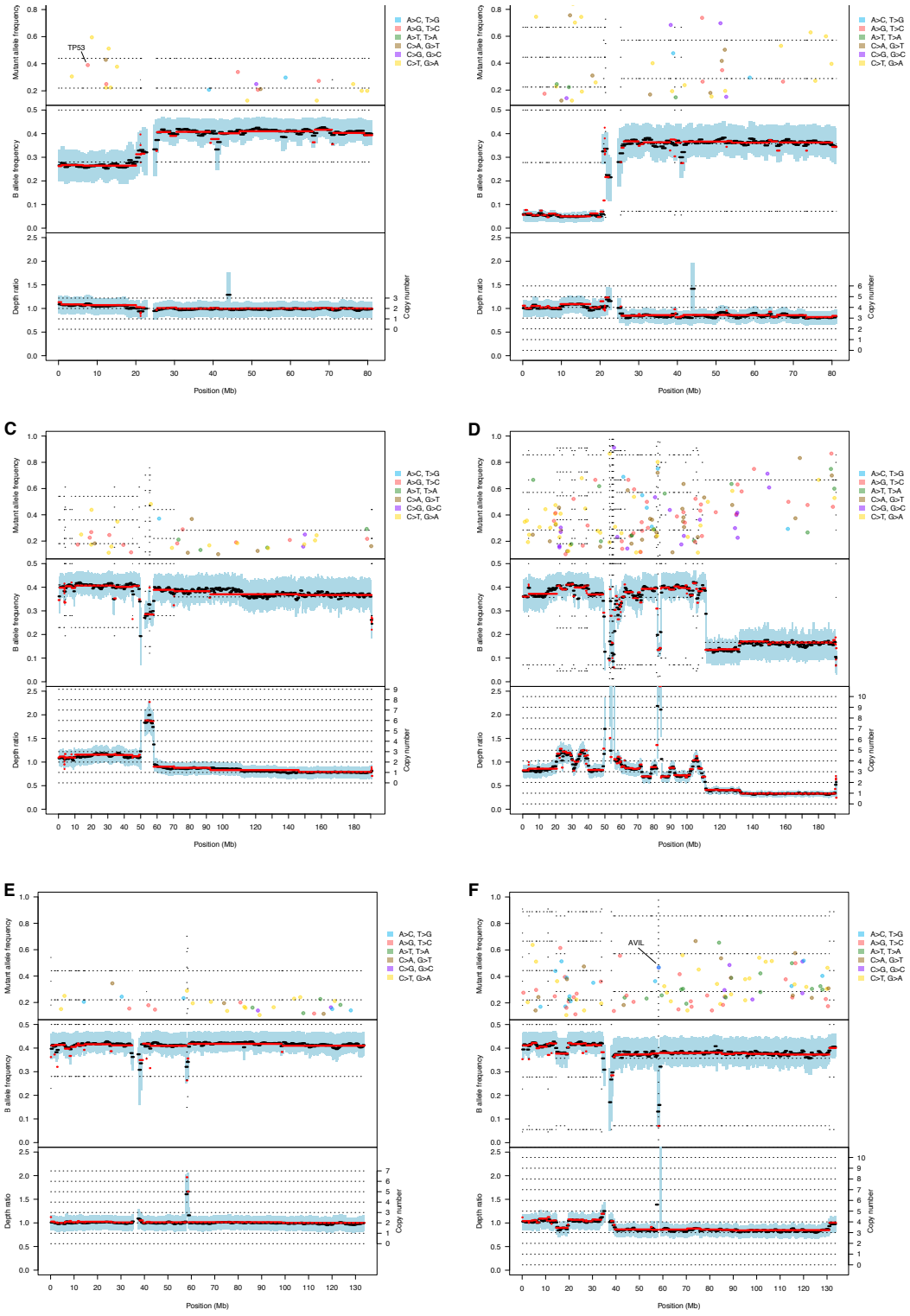


Figure S2: Chromosome view Grade II and Grade IV.
Chromosome view from Sequenza R package for chromosome 17 grade II (**A**) and grade IV (**B**), chromosome 4 grade II (**C**) and grade IV (**D**) and chromosome 12 grade II (**E**) and grade IV (**F**).

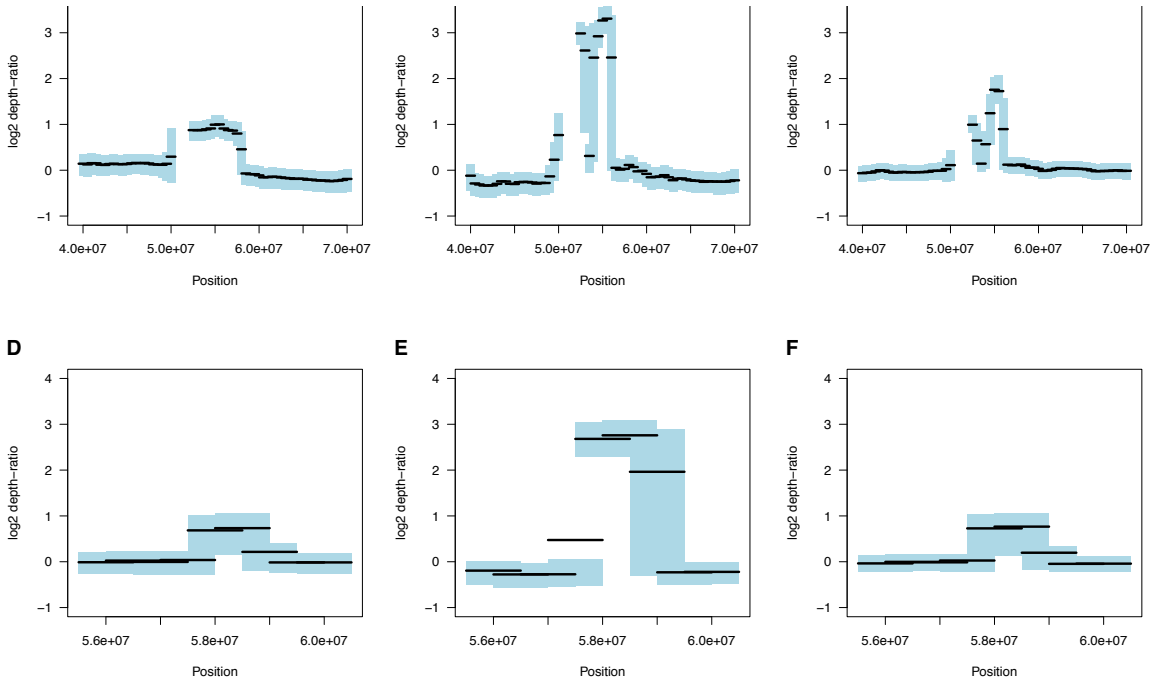


Figure S3: Focal amplification in the primary and the relapse tumour.

Focals amplification in chromosome 4 in grade II (**A**), grade IV (**B**) and the joint relapse cohort (**C**); and in chromosome 12 grade II (**C**), grade IV (**D**) and the joint relapse cohort (**E**)

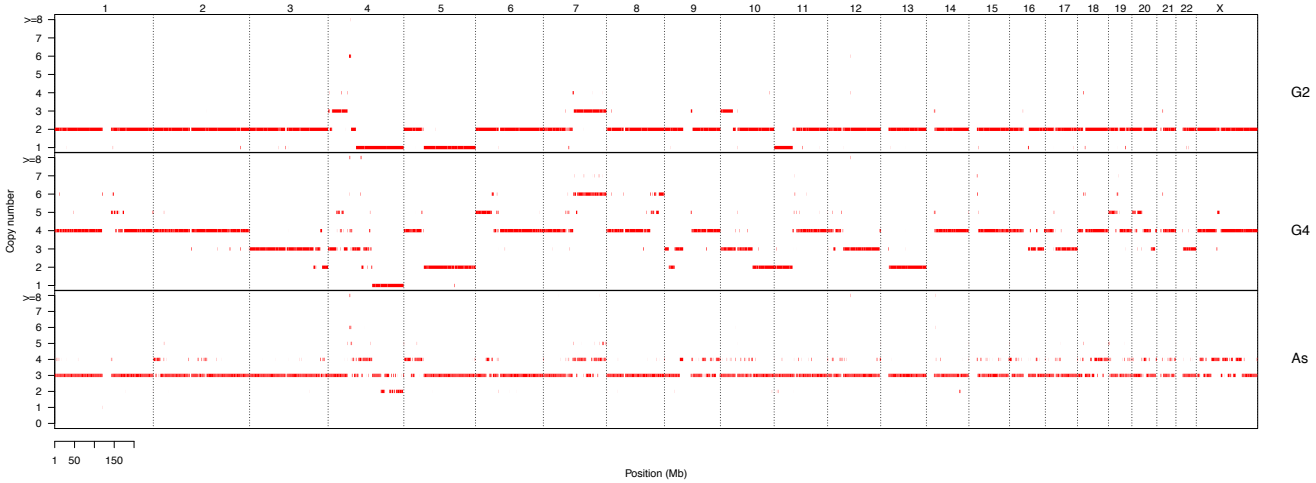


Figure S4: Total copy number segments.
Segments with total copy number for the grade II (upper panel), grade IV (middle panel) and the relapse (bottom panel).

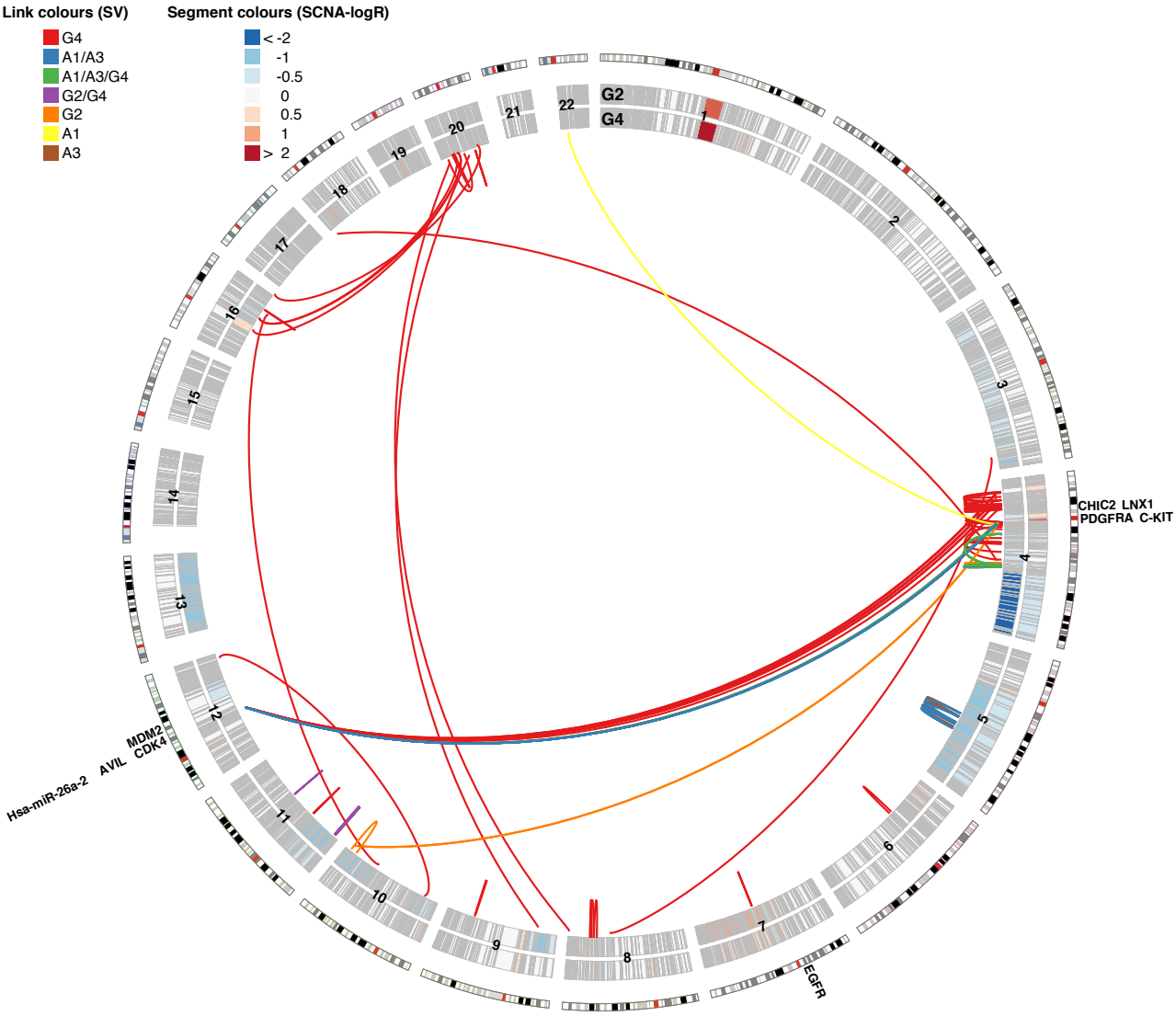


Figure S5: CIRCOS plot of the somatic variations.
Circos plot of the translocations on grade II, grade IV and relapse (coloured lines) and copy number variation for grade II and grade IV (coloured circoular bands)

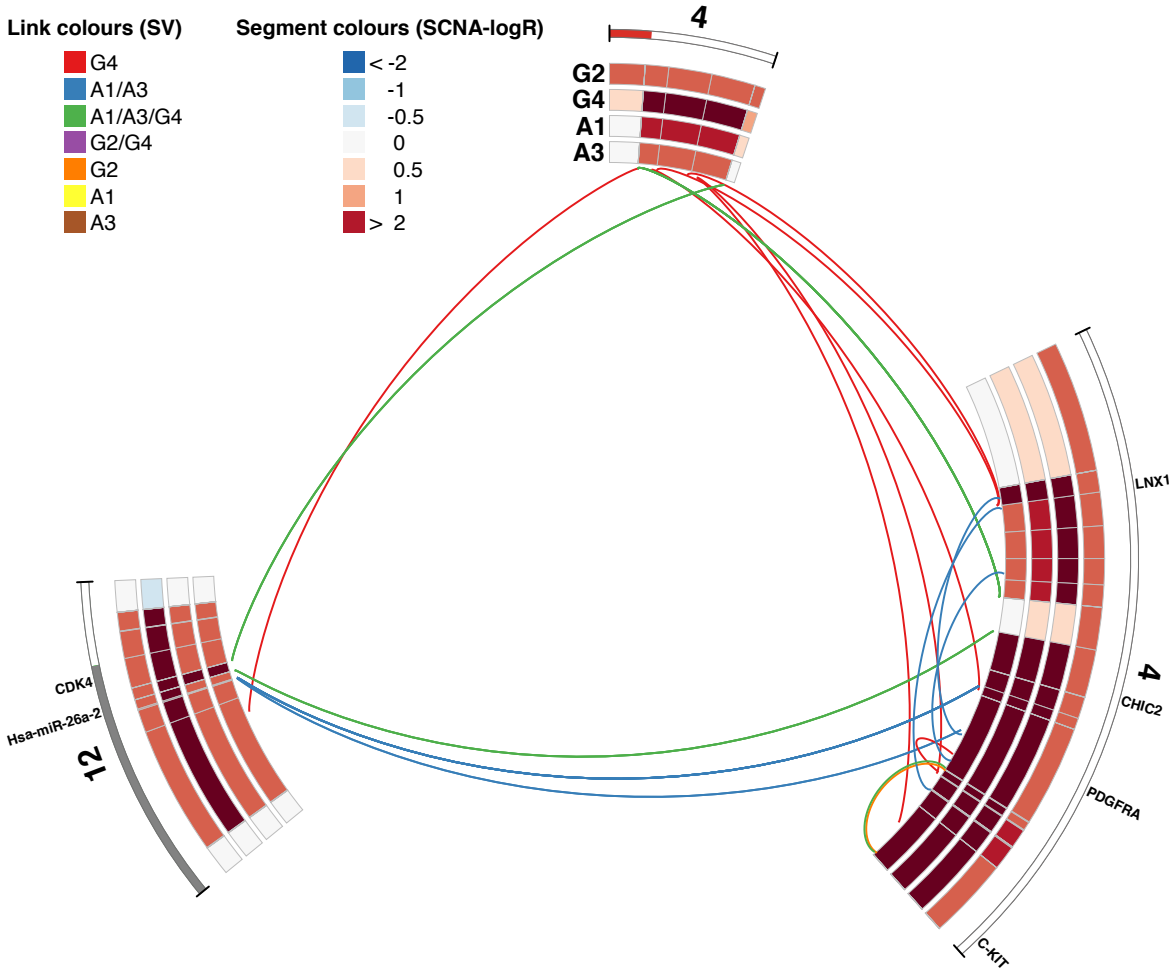


Figure S6: Focus on chromosome 4 and chromosome 12.

Circos plot magnification on chromosome 4 and chromosome 12; displays SV and CNV detected in the grade II grade IV and the relapse samples A1 and A3, each level of the circus represent a sample, starting to the most external level, grade II, grade IV, relapse region A1 and relapse region A3. The colour of the level is proportional to the local copy number. Each segment connecting different genomic position represents a detected SV. The legend indicates the sample where the SV is detected.

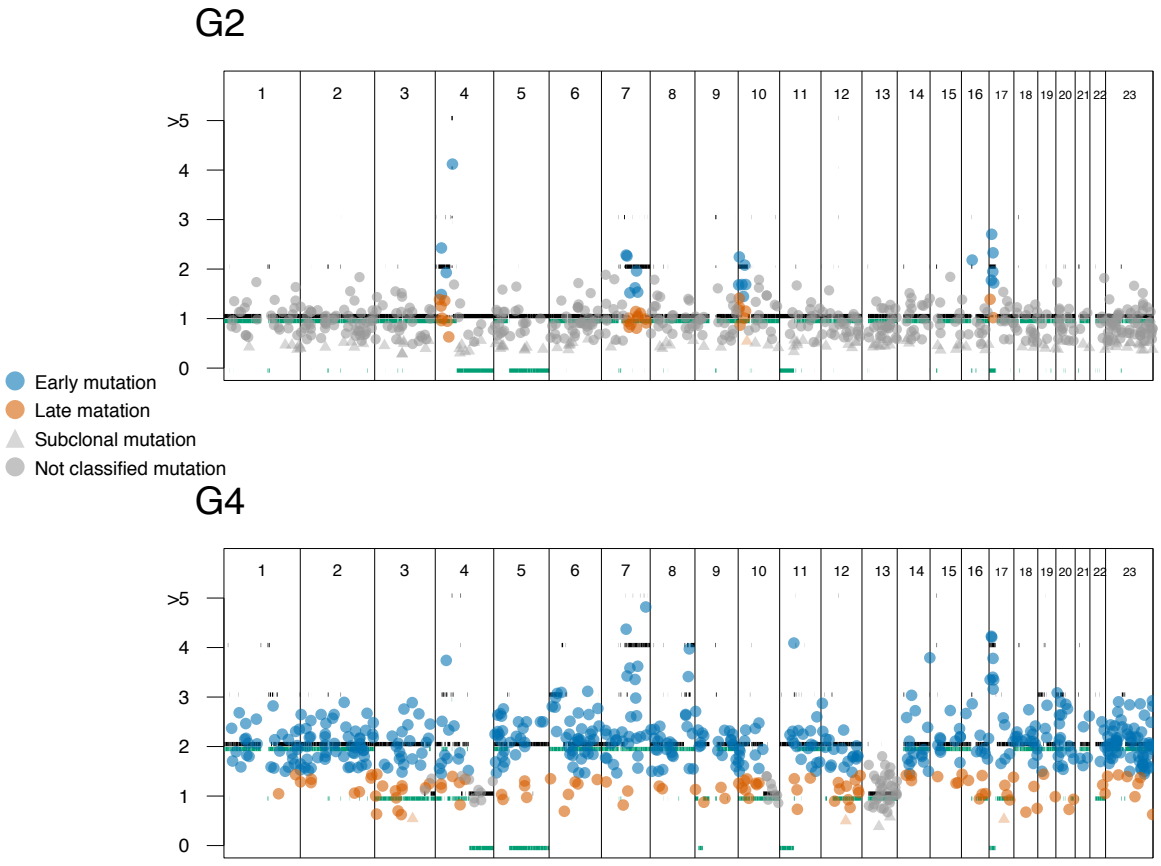


Figure S7: Genome doubling.

Genome doubling analysis was performed using normalised variant allele frequency. Normalisation was carried considering the copy number and the tumor cellularity. Colour of a mutation indicates timing of the given mutation relative to a copy number aberration (local of genome wide aberrations). Early mutation are represented in blue, late mutation are represented in orange, clonal mutation are represented with a grey triangular sign, grey points represent mutation where timing where not possible. Grade II specimen (**G2**) was detected as diploid genome with few copy number variations. Grade IV specimen (**G4**) was detected as tetraploid genome, timing mutation results in a clear pattern of genome doubling, where majority of common mutation between the two specimens are detected as early mutation, and unique mutation of grade II are detected as late mutations, The small number of late mutation in grade IV suggests the close proximity of the grade IV sample to the genome doubling event.

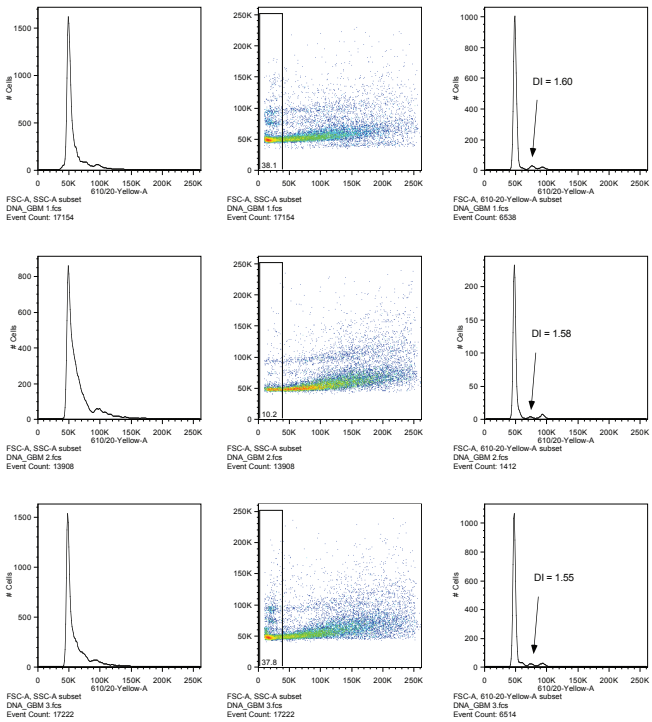
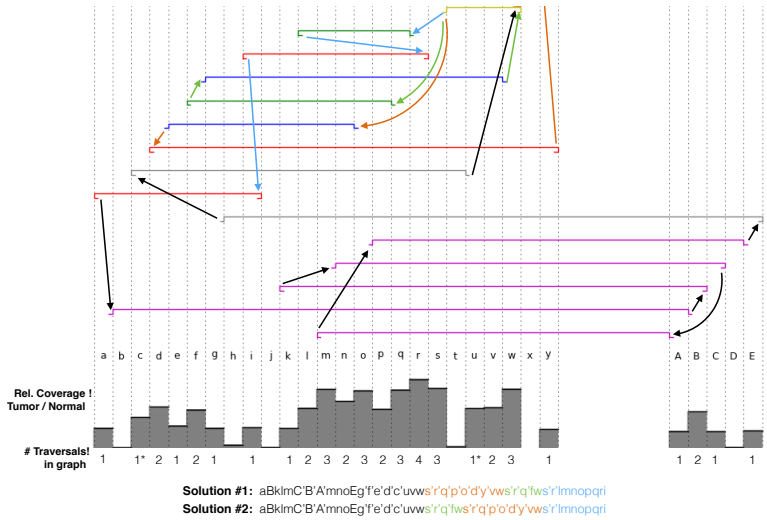


Figure S8: DNA index by Flow cytometry.

Flow cytometry results for the 3 relapse sectors. DNA index results are 1.60 for the sector A1 (upper panel), 1.58 for A2 (middle panel) and 1.55 for A3 (bottom panel). The results confirm the relapse to have a triploid genome.



GBM01-G4

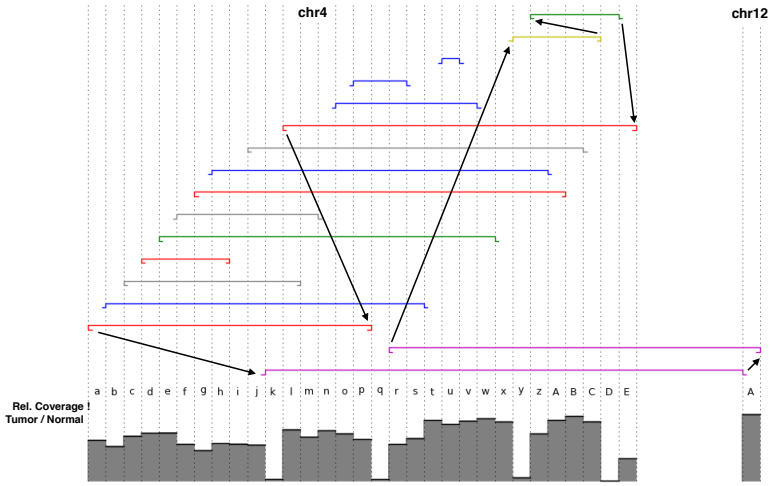


Figure S9: Precise reconstruction of the double minute.
Reconstruction of the double minute in the relapse (upper panel) and in the grade IV (bottom panel) using the method described in Sanborn at 2014.

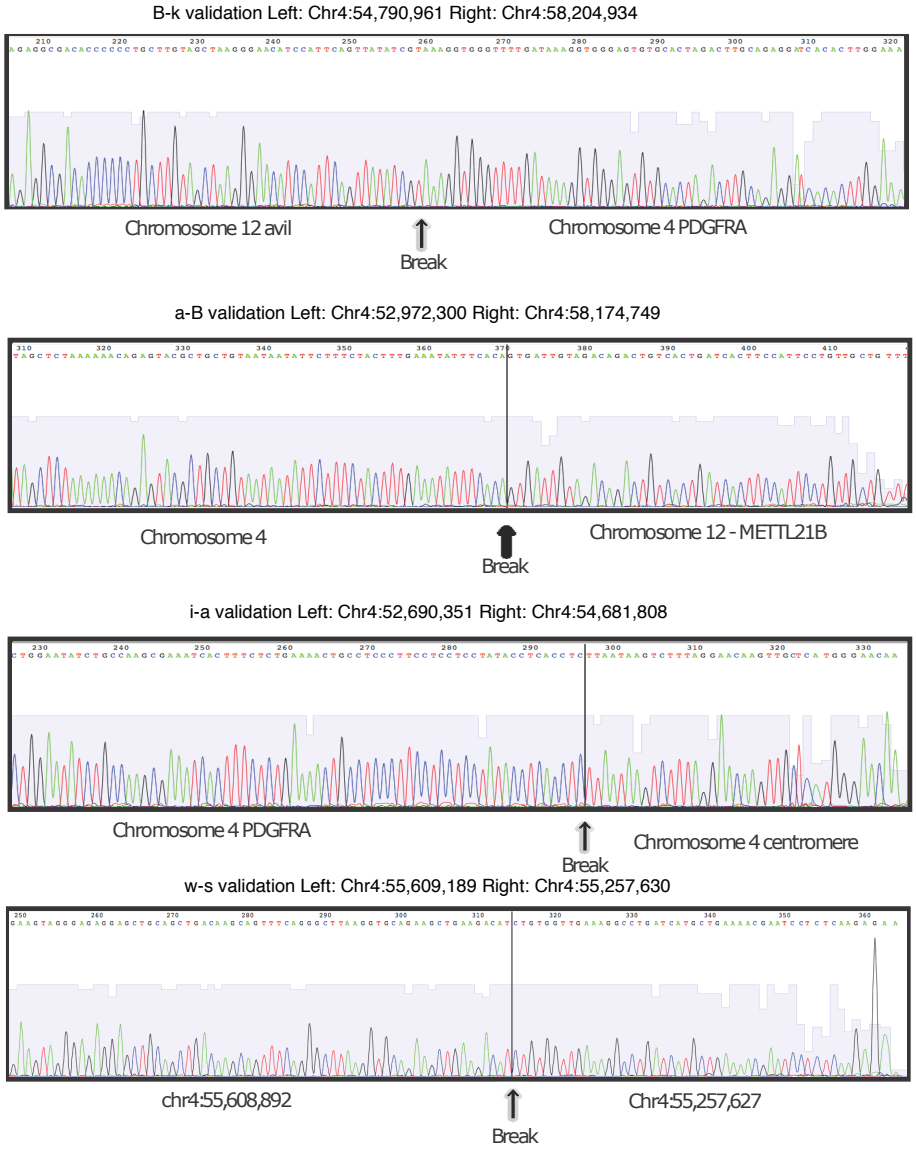


Figure S10: Sanger sequencing validation of the DM breakpoint:
Validation by Sanger sequencing of the break points forming the double minute.

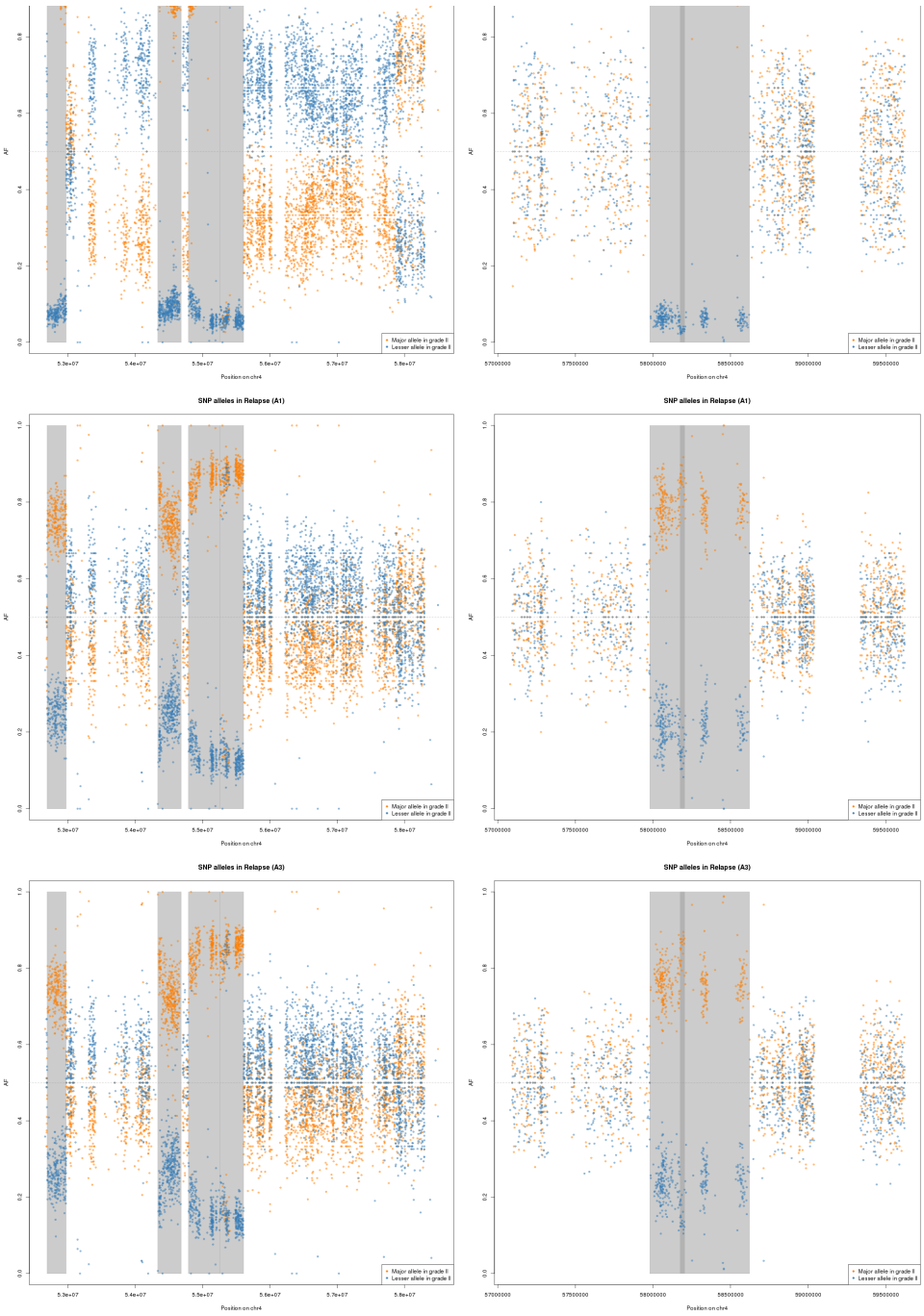


Figure S11: Allele-specific analysis of heterozygous single-nucleotide polymorphisms in the grade IV and relapse samples.

For SNPs mapping to the DM interval, SNP alleles were classed as “major” (AF > 0.5, orange) or “minor” (AF < 0.5, blue) in the grade II sample. Using this alleleic classification, allele frequency (AF) was plotted for the grade IV and relapse samples. Loci forming apart from a ~200Kb interval on 4q, chr4:55,282,288-55,499,750, the DM segments in all higher grade samples appear to share haplotypes, which may reflect a common origin from the amplified haplotypes in the grade II sample. Light grey rectangles represent loci mapping to the double minute.