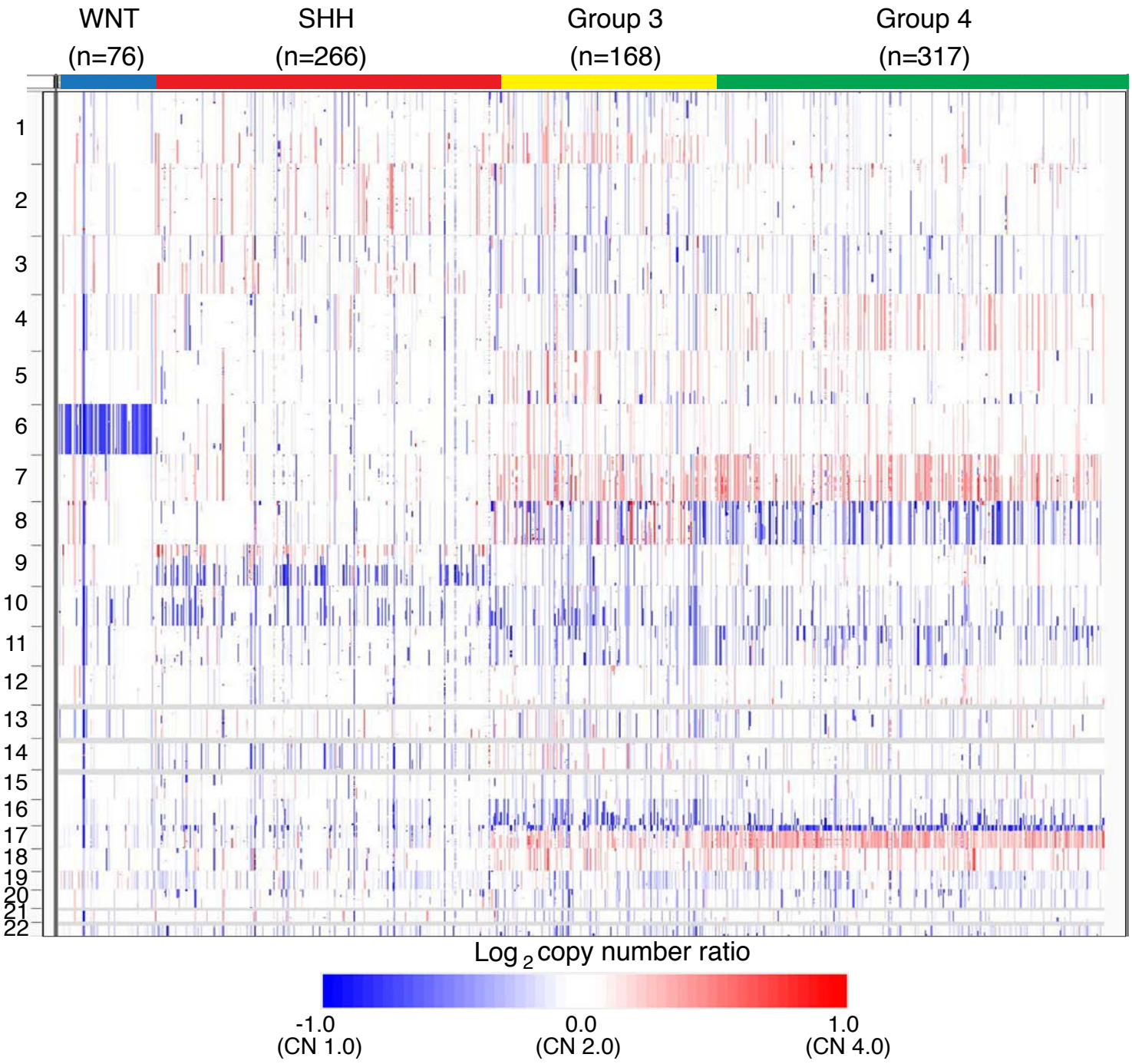
Supplementary Information

Subgroup specific structural variation across 1,000 medulloblastoma genomes

The Medulloblastoma Advanced Genomics International Consortium (MAGIC)

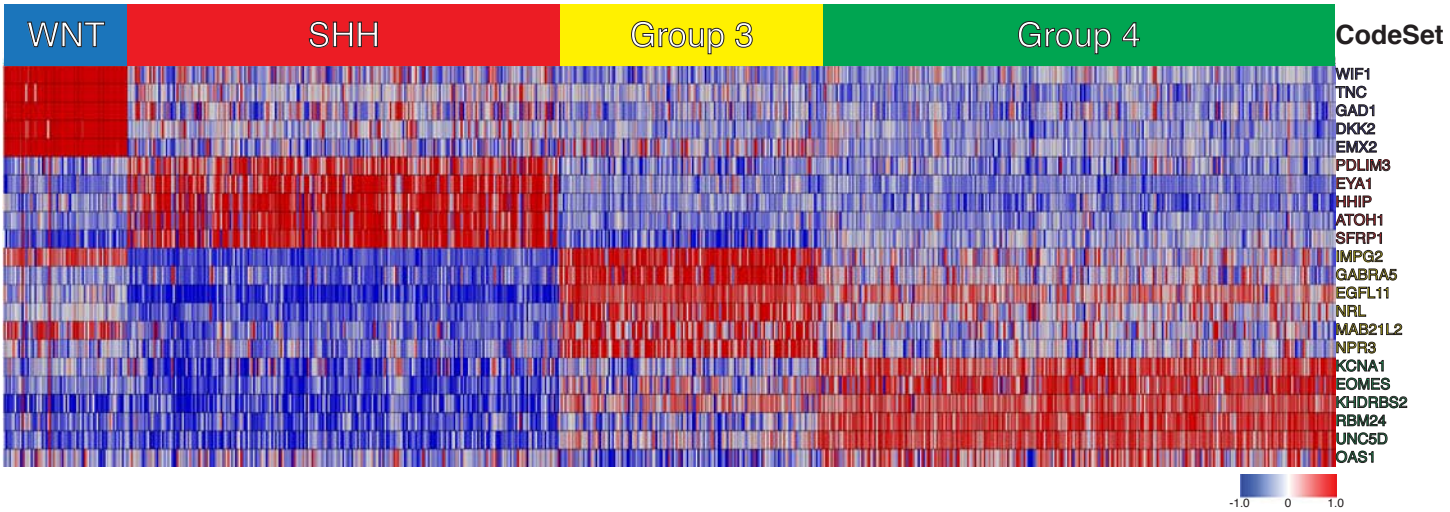
Table of Contents

Supplementary Figure 1, Northcott, Shih et al



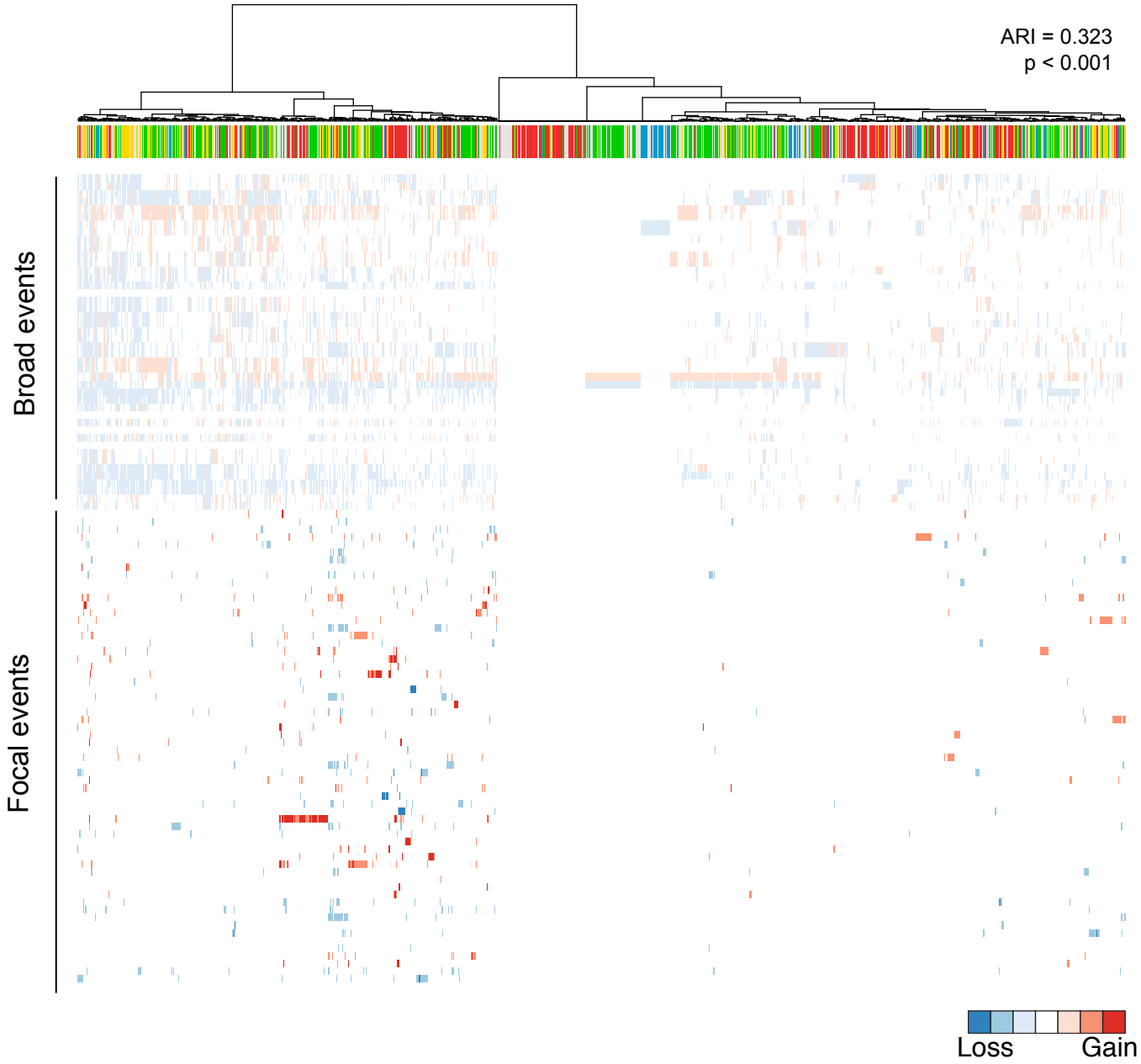
Supplementary Figure 1. Genome-wide copy number profile of medulloblastoma subgroups. SNP6 profiling was performed on 1087 non-overlapping primary medulloblastomas. Shown is a copy number heatmap for 827 cases classified according to medulloblastoma subgroup based on matched gene expression data. Amplifications are shown in red and deletions in blue.

Supplementary Figure 2, Northcott, Shih et al



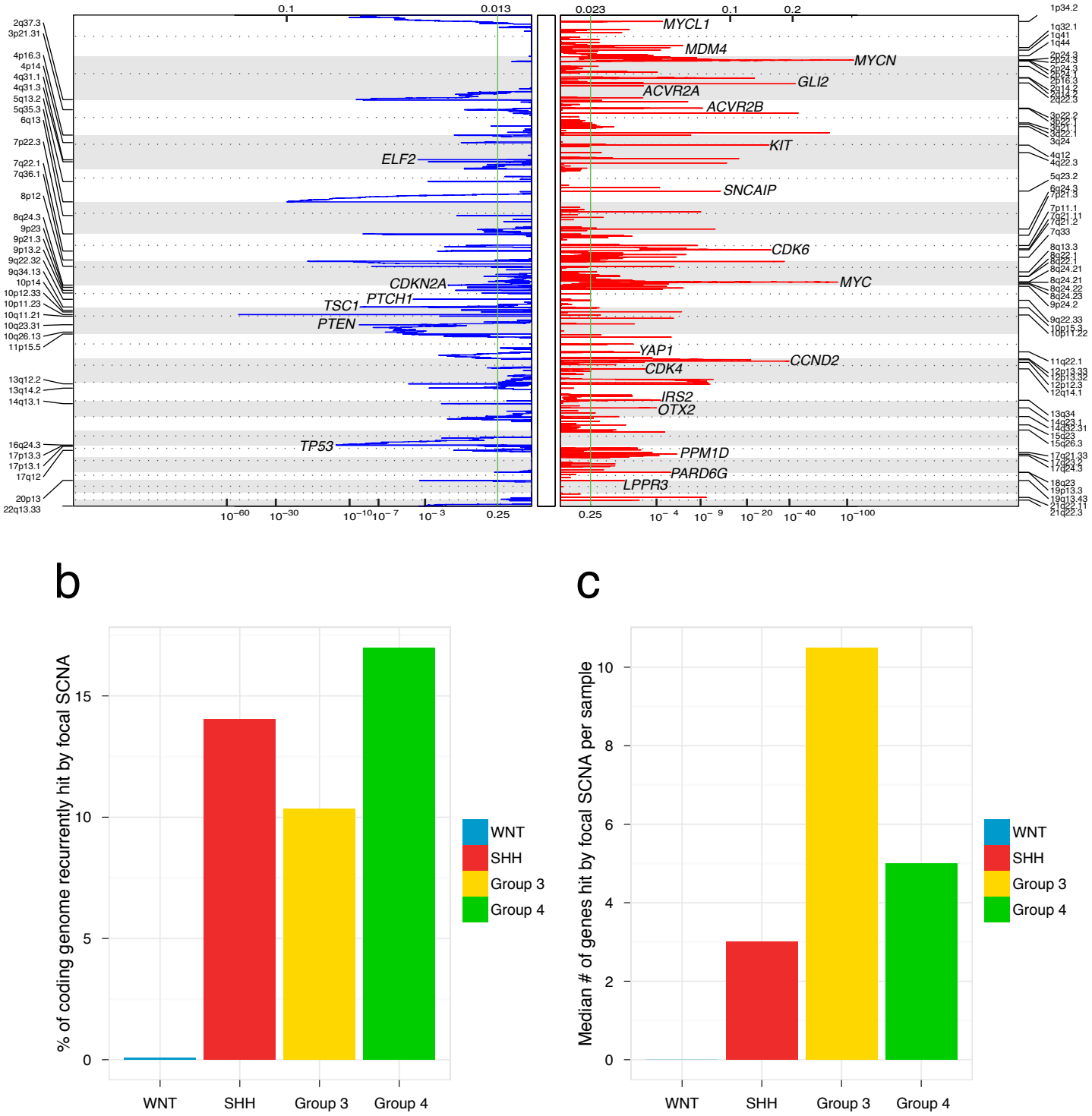
Supplementary Figure 2. nanoString classification of medulloblastoma. Heatmap showing the expression level of 22 medulloblastoma signature genes (CodeSet) across 827 cases in the MAGIC cohort as determined by nanoString. Molecular subgroup affiliation was inferred by class prediction and samples have been sorted according to their assigned class.

Supplementary Figure 3, Northcott, Shih et al



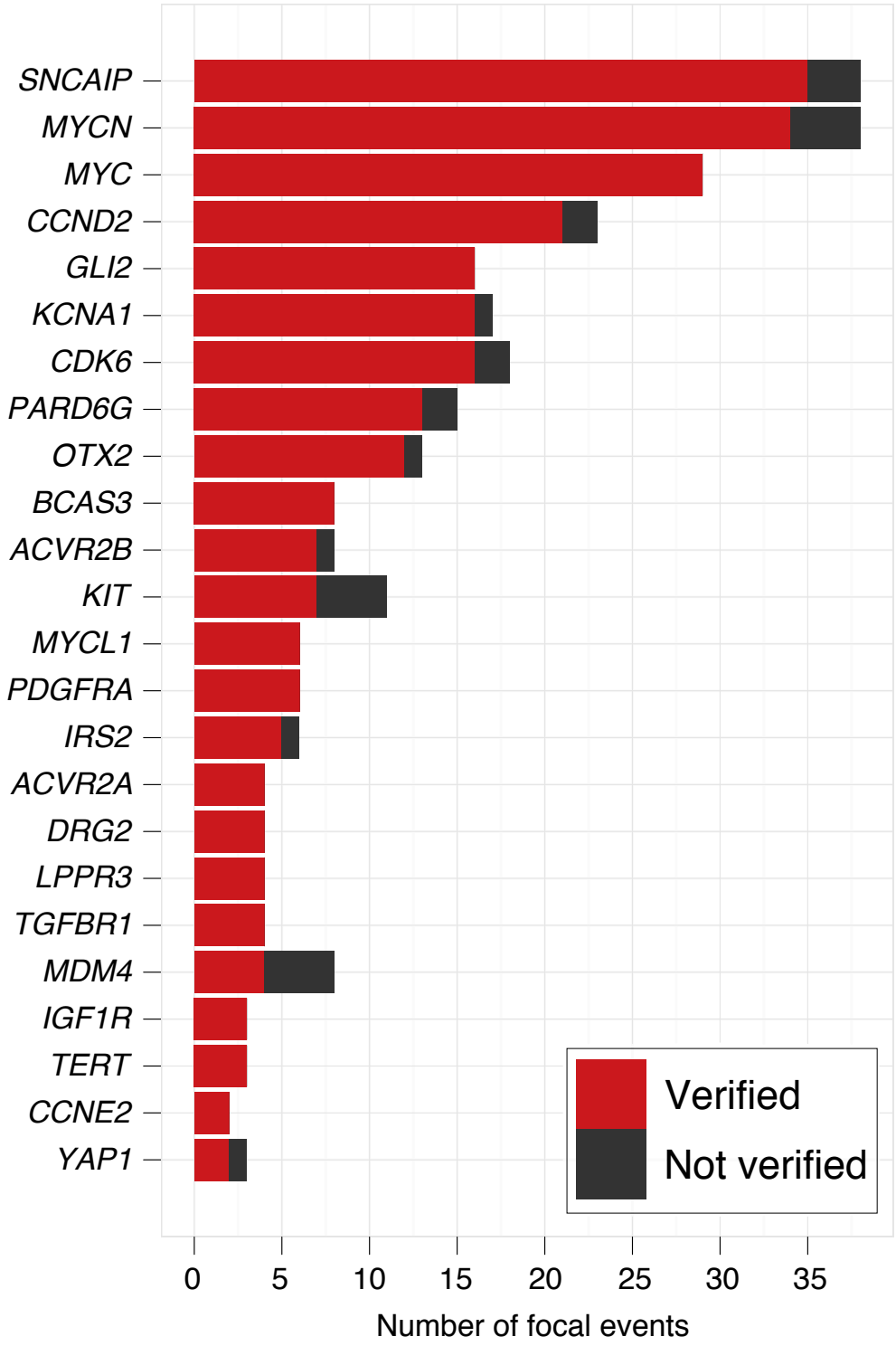
Supplementary Figure 3. Unsupervised hierarchical clustering of broad and focal SCNAs in medulloblastoma. Coloured side bar indicates medulloblastoma subgroups (WNT: blue, SHH: red, Group 3: yellow, Group 4: green). Only focal events identified in the pan-cohort GISTIC2 analysis were included in the clustering. Resulting clusters show significant agreement with known medulloblastoma expression subgroups ($p < 0.001$, Chi-squared test). ARI, Adjusted Rand Index.

Supplementary Figure 4, Northcott, Shih et al



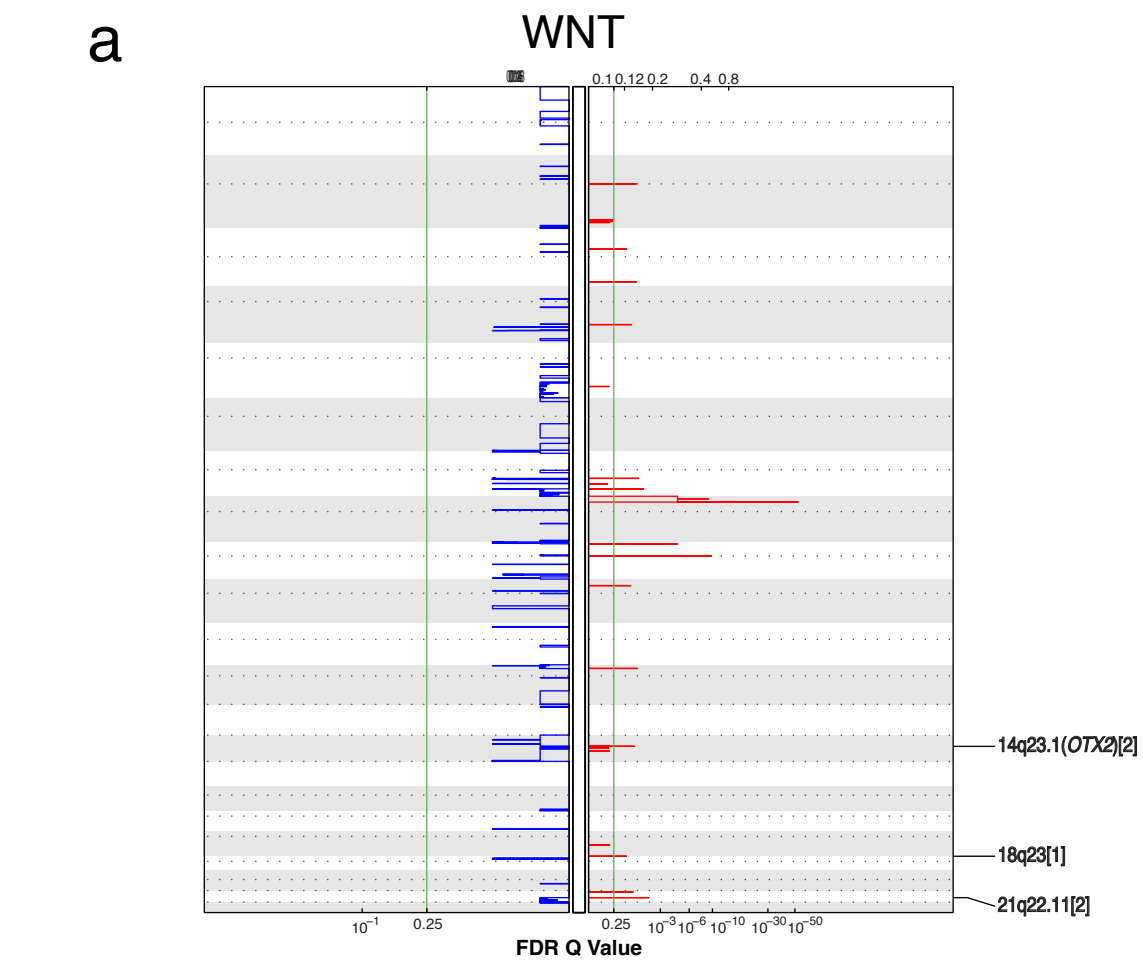
Supplementary Figure 4. Somatic copy number aberrations in medulloblastoma. a, Segmented copy number profiles for 1087 primary medulloblastomas were analyzed using GISTIC2 to identify significant regions of amplification (depicted in red, right plot) and deletion (depicted in blue, left plot). The most significant peaks of interest have been labeled with the candidate driver gene. **b-c,** Genomic coverage and subgroup distribution of recurrent focal somatic copy number aberrations (SCNAs) in medulloblastoma. Both the percentage of the coding genome (**b**) and the median number of genes (per sample) (**c**) affected by SCNAs in medulloblastoma subgroups are summarized.

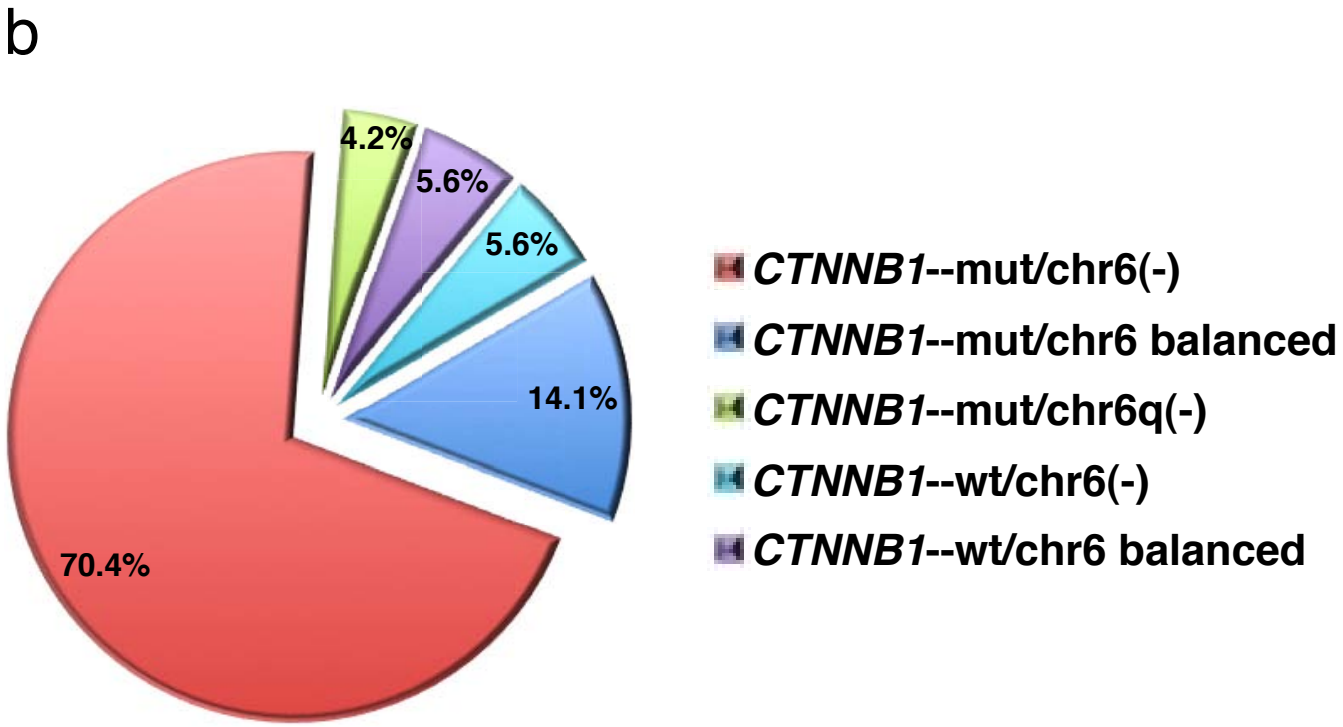
Supplementary Figure 5, Northcott, Shih et al



Supplementary Figure 5. Verification of focal SCNAs by nanoString. Twenty-four genes inferred to be focally gained/amplified by SNP6 were interrogated using a custom nanoString CodeSet across 192 medulloblastomas selected from our cohort. Bar graph shows the number of samples with a verified SCNA in red versus non-verified in black.

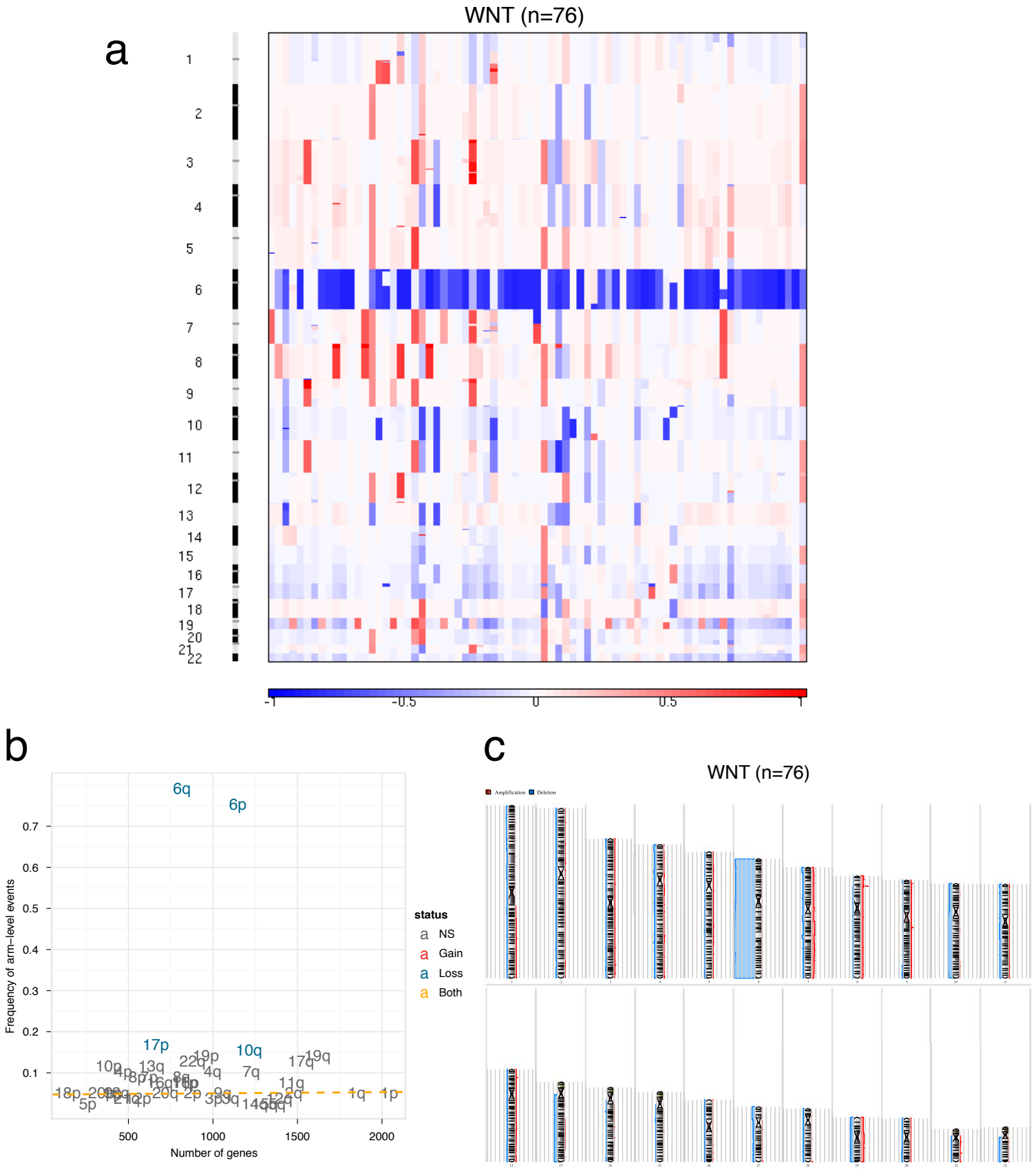
Supplementary Figure 6, Northcott, Shih et al





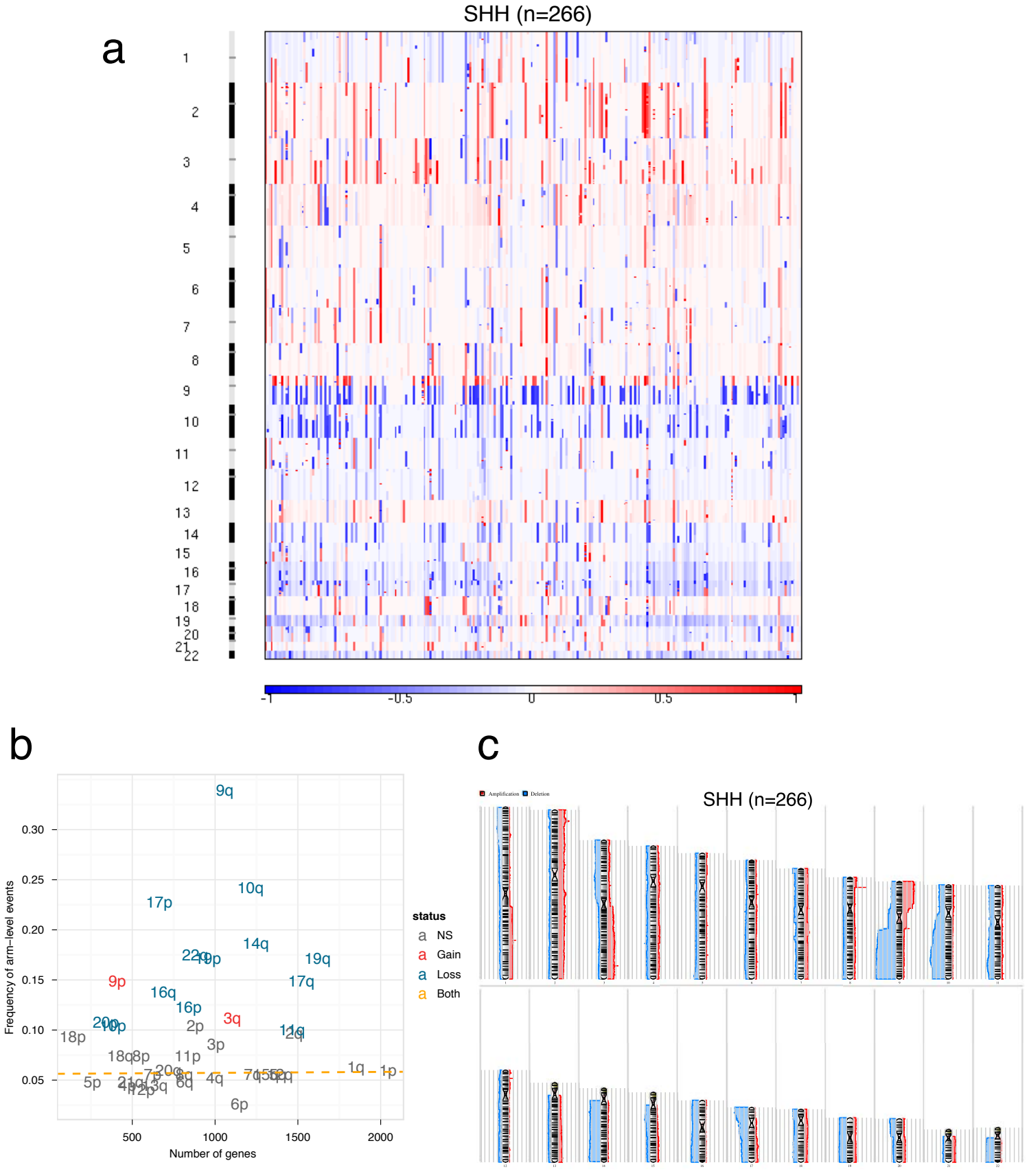
Supplementary Figure 6. WNT medulloblastoma. a, GISTIC2 analysis of WNT medulloblastoma. Significant regions of amplification (right panel) are labeled. No significant regions of deletion were identified (left panel). **b,** Pie chart showing the distribution of *CTNNB1* mutations and chromosome 6 copy number status among WNT cases.

Supplementary Figure 7, Northcott, Shih et al



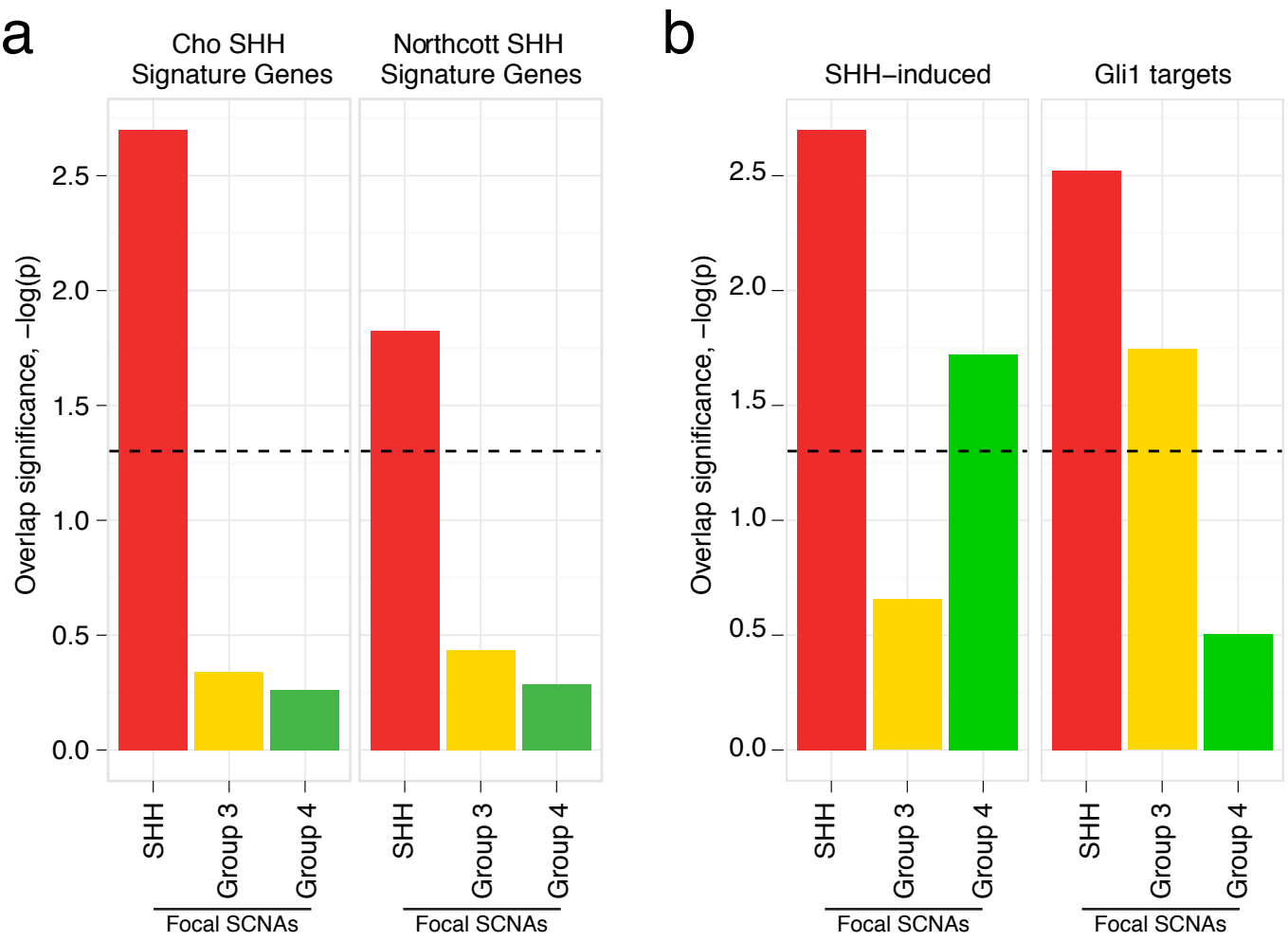
Supplementary Figure 7. SNP6 copy number profile of WNT medulloblastoma. **a,** Copy number heatmap for 76 WNT medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in WNT tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in WNT medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 8, Northcott, Shih et al



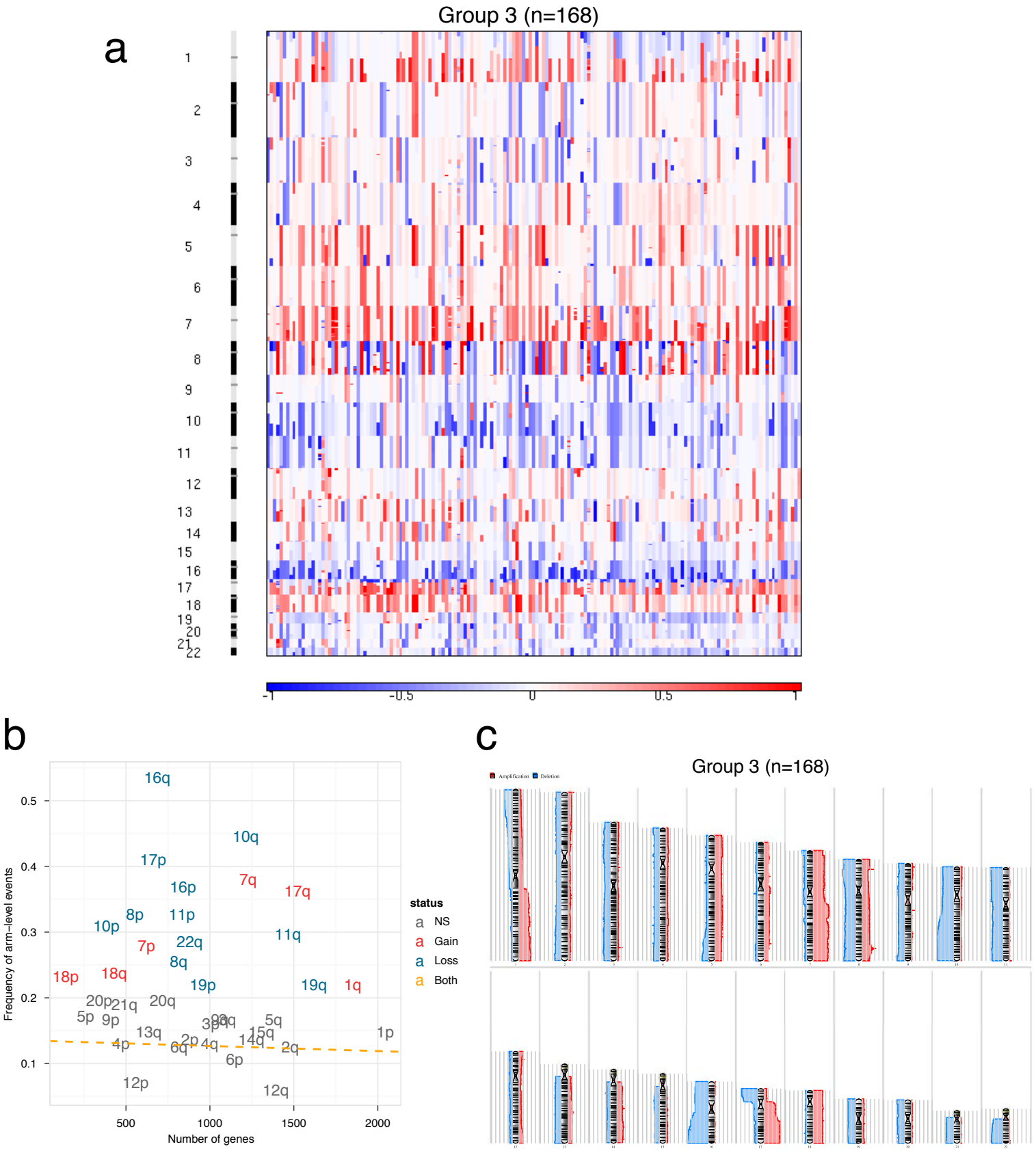
Supplementary Figure 8. SNP6 copy number profile of SHH medulloblastoma. **a,** Copy number heatmap for 266 SHH medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in SHH tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in SHH medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 9, Northcott, Shih et al



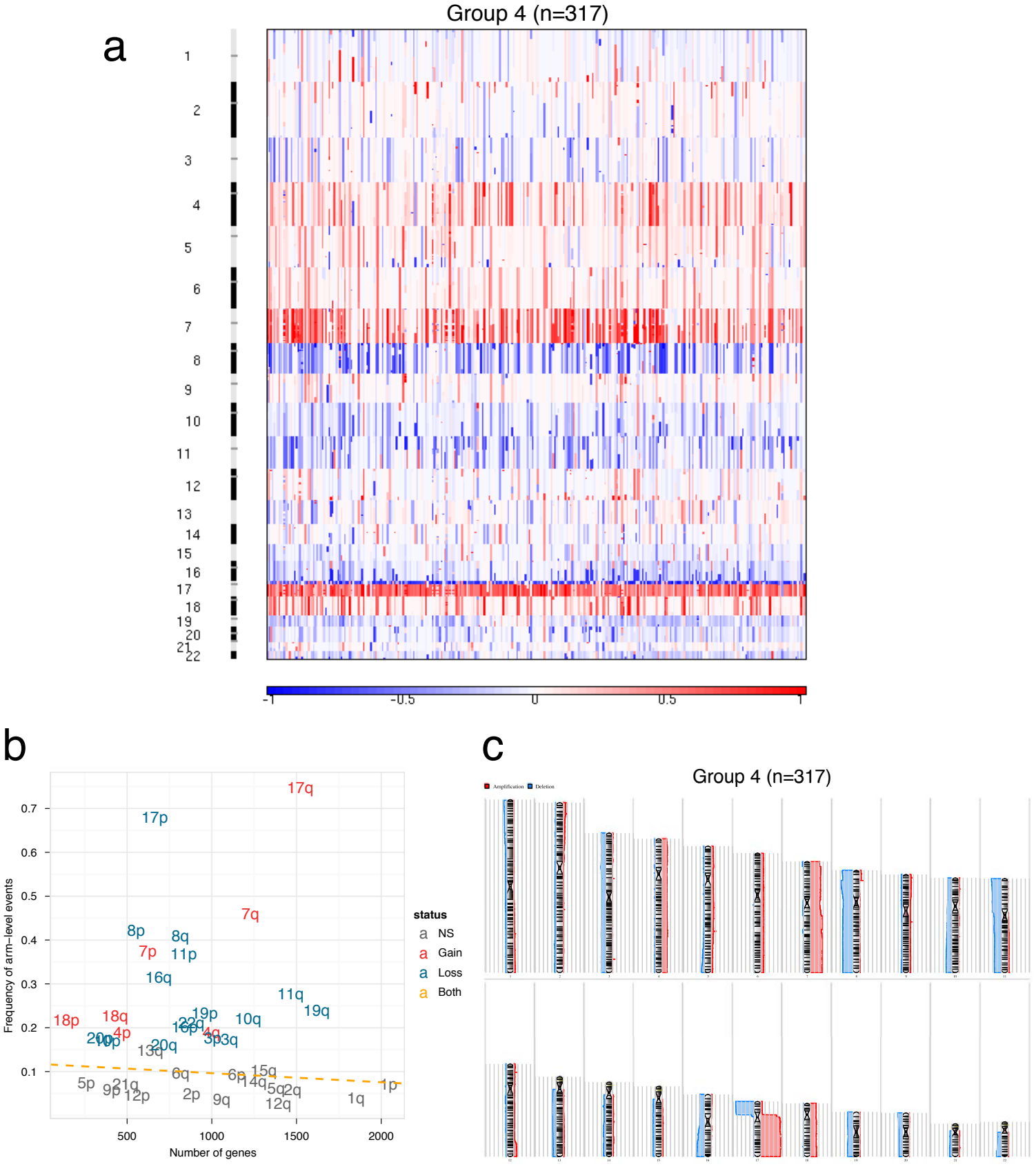
Supplementary Figure 9. Significant overlap between SHH amplification genes and targets of SHH signaling. a, Bar graph shows the overlap significance based on permutation tests between SHH signature genes reported in the Cho (left plot) and Northcott (right plot) studies and genes mapping to focal SCNAs in our dataset. Dashed line indicates significance threshold (P=0.05). **b,** Bar graph of overlap significance between genes focally amplified/gained in medulloblastoma subgroups and genes transcriptionally induced by SHH in cerebellar granule neuron progenitors (CGNPs, left plot) and/or genes targeted by the Gli1 transcription factor (right plot), also in CGNPs.

Supplementary Figure 10, Northcott, Shih et al



Supplementary Figure 10. SNP6 copy number profile of Group 3 medulloblastoma. **a,** Copy number heatmap for 168 Group 3 medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in Group 3 tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in Group 3 medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 11, Northcott, Shih et al

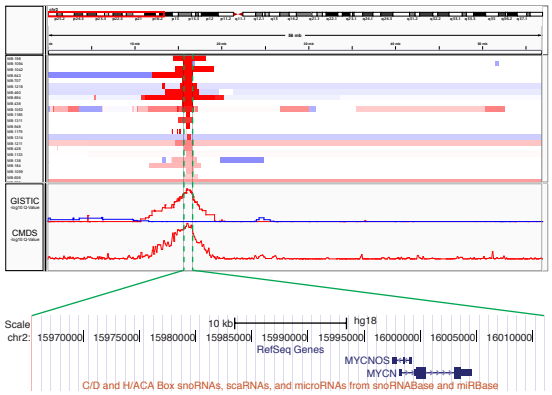


Supplementary Figure 11. SNP6 copy number profile of Group 4 medulloblastoma. **a,** Copy number heatmap for 317 Group 4 medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in Group 4 tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in Group 4 medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 12, Northcott, Shih et al

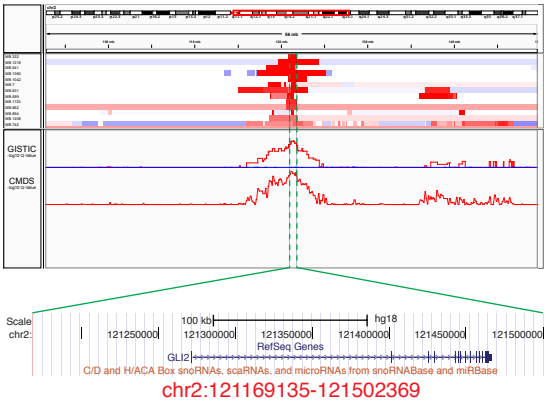
SHH Amplified Regions: Part A

1



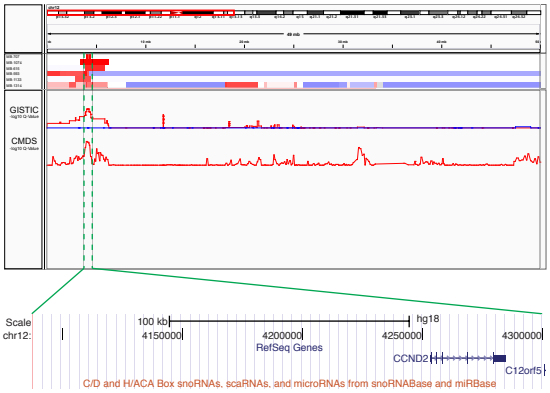
chr2:15965458-16011383

2



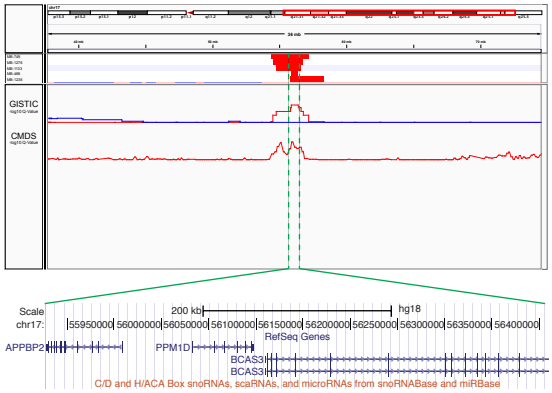
chr2:121169135-121502369

3



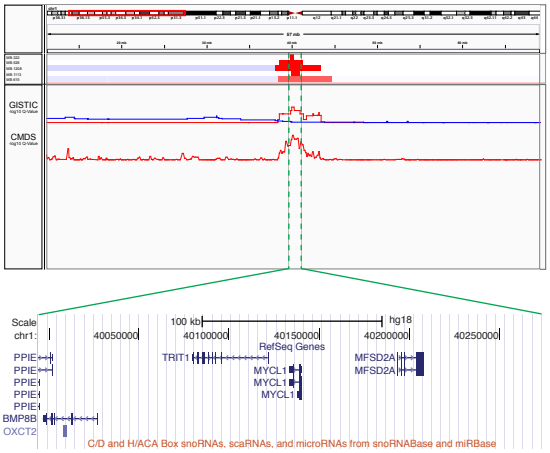
chr12:4087878-4301103

4



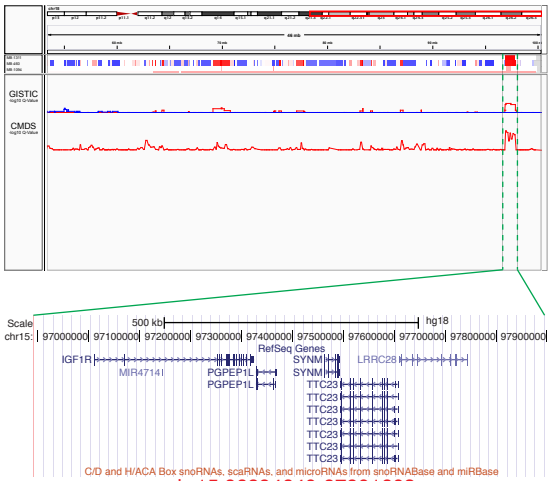
chr17:55878149-56409604

5



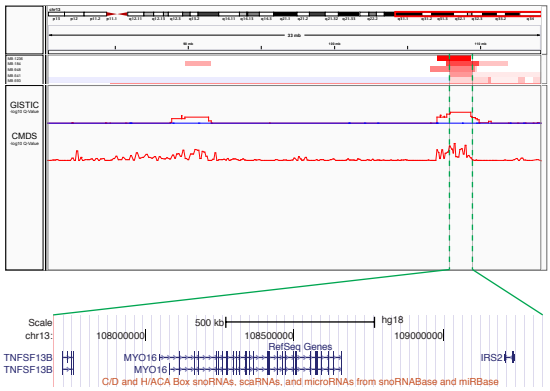
chr1:39994443-40275086

6



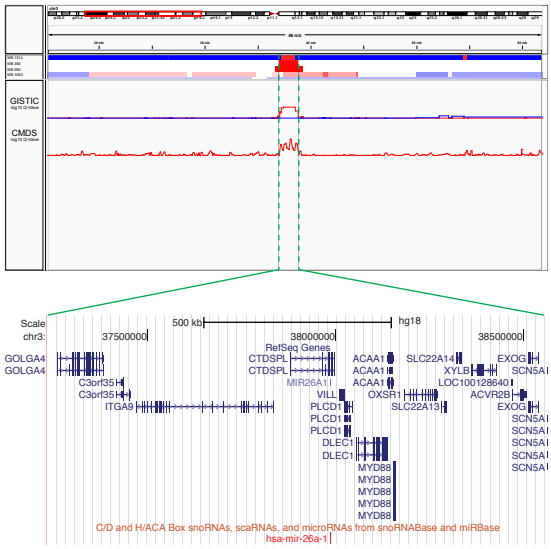
chr15:96894640-97901603

7



chr13:107695688-109344178

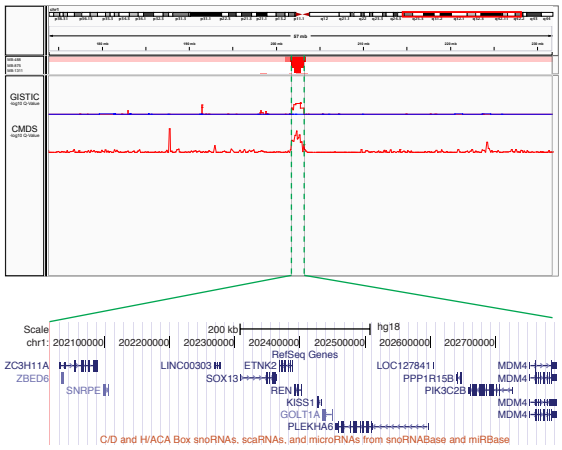
8



chr3:37233861-38565570

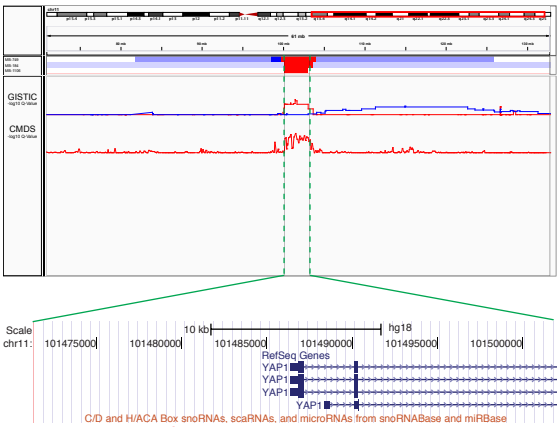
SHH Amplified Regions: Part B

9



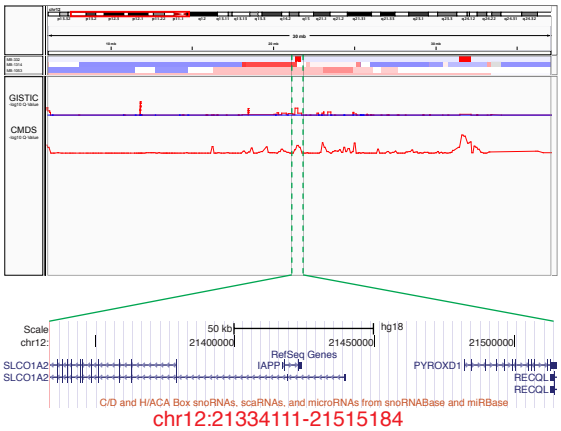
chr1:202017953-202796430

10



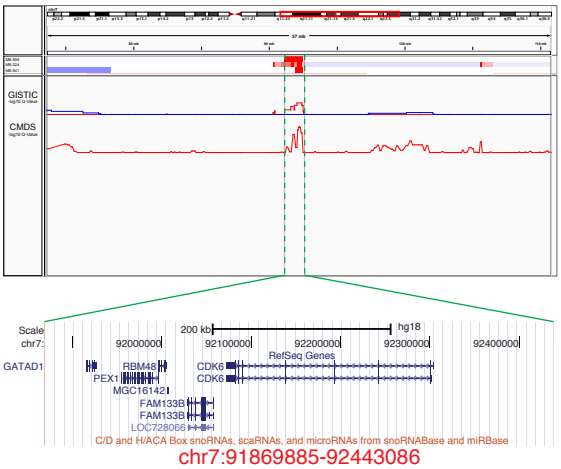
chr11:101471342-101501994

11



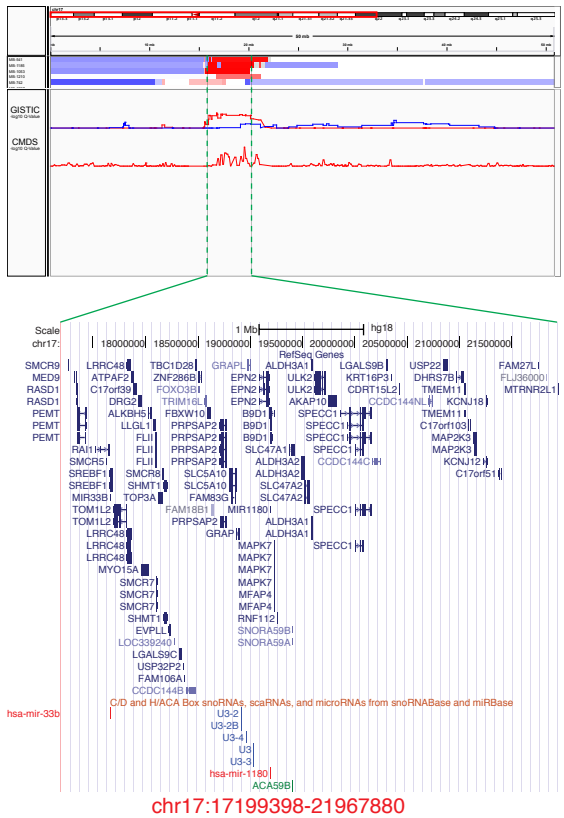
chr12:21334111-21515184

12



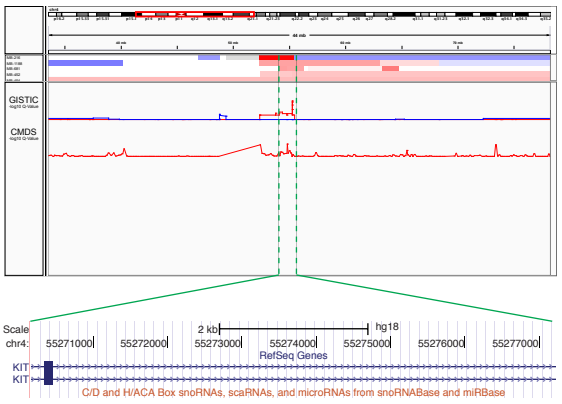
chr7:91869885-92443086

13



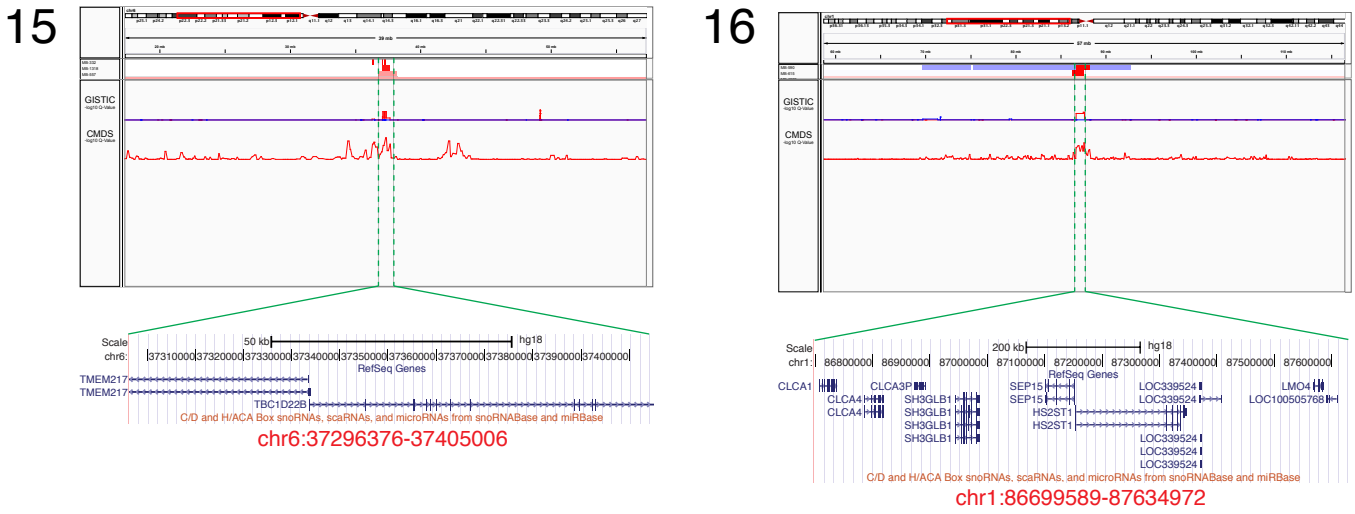
chr17:17199398-21967880

14



chr4:55270164-55277219

SHH Amplified Regions: Part C

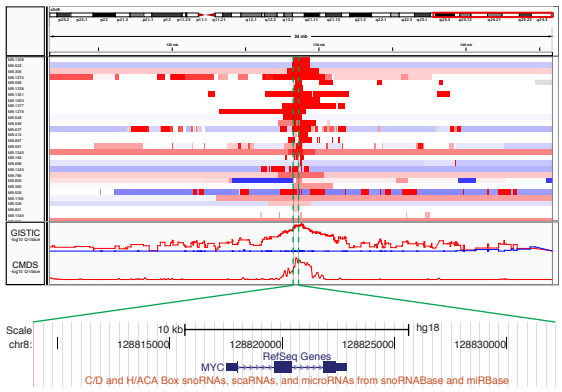


Supplementary Figure 12. Highly significant regions of focal amplification in SHH medulloblastoma. SNP6 copy number output for the top regions of focal amplification identified in SHH as determined by GISTIC2 analysis. Segmented copy number data is shown for SHH samples harboring amplification/gain in the regions of interest. Significance values ($-\log_{10}$ q-value) from GISTIC2 and CMDS analysis are included below the copy number profiles. Output is from the Integrative Genomics Viewer (IGV). UCSC RefSeq genes mapping to each significant region are shown below the IGV output. Regions have been sorted according to significance beginning with the most significant.

Supplementary Figure 13, Northcott, Shih et al

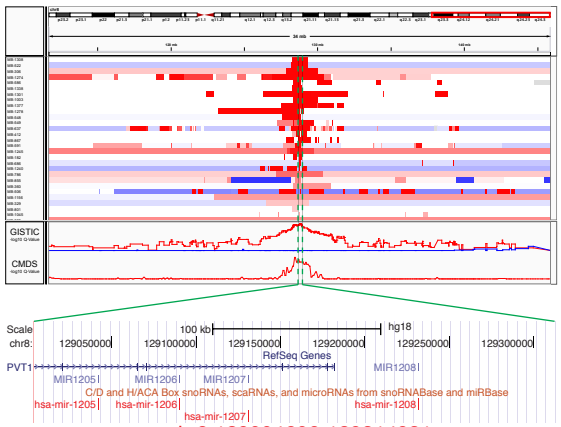
Group 3 Amplified Regions: Part A

1



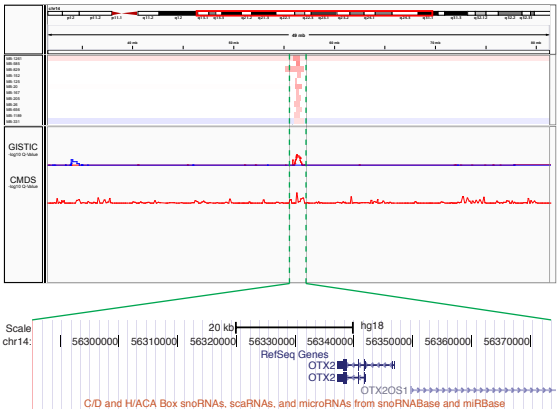
chr8:128808954-128832343

2



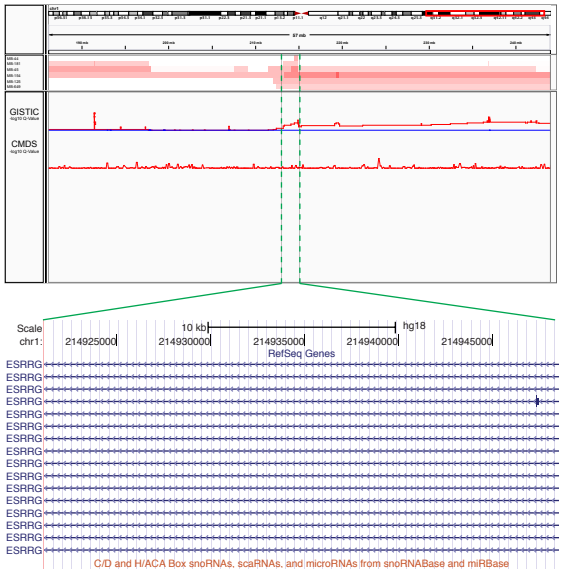
chr8:129004606-129314931

3



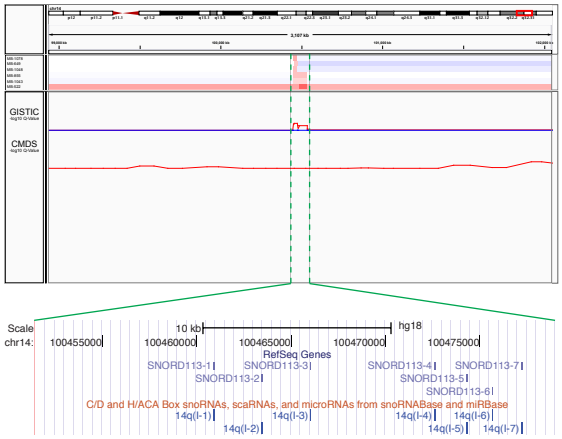
chr14:56285448-56374271

4



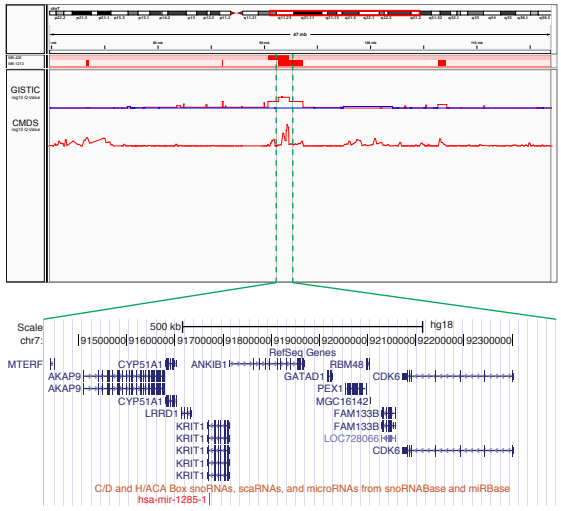
chr1:214921160-214948489

5

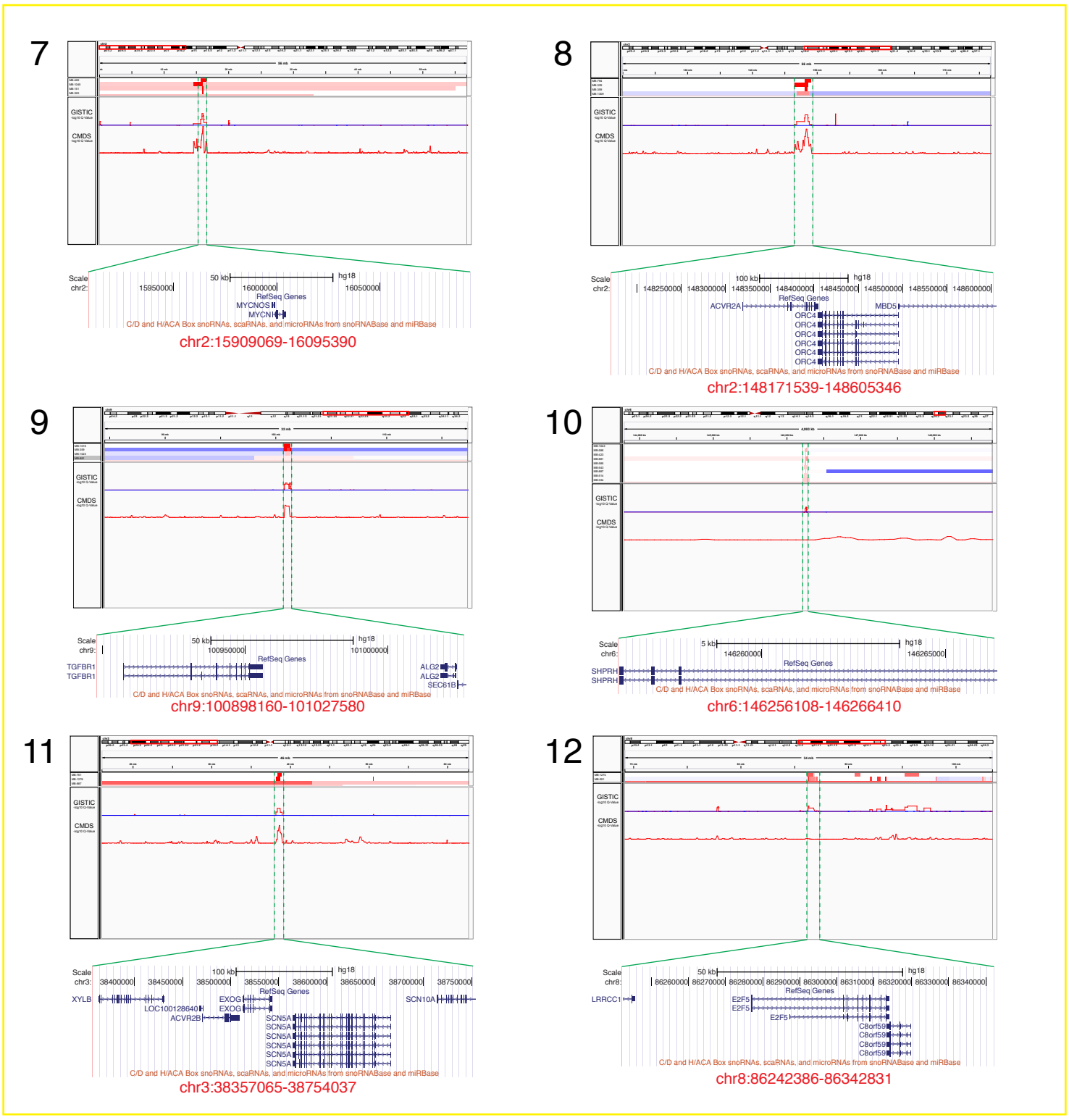


chr14:100451467-100479162

6



chr7:91330040-92397935

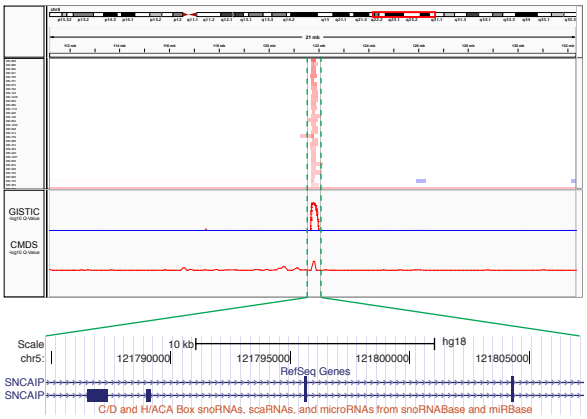


Supplementary Figure 13. Highly significant regions of focal amplification in Group 3 medulloblastoma. SNP6 copy number output for the top regions of focal amplification identified in Group 3 as determined by GISTIC2 analysis. Segmented copy number data is shown for Group 3 samples harboring amplification/gain in the regions of interest. Significance values ($-\log_{10}$ q-value) from GISTIC2 and CMDS analysis are included below the copy number profiles. Output is from the Integrative Genomics Viewer (IGV). UCSC RefSeq genes mapping to each significant region are shown below the IGV output. Regions have been sorted according to significance beginning with the most significant.

Supplementary Figure 14, Northcott, Shih et al

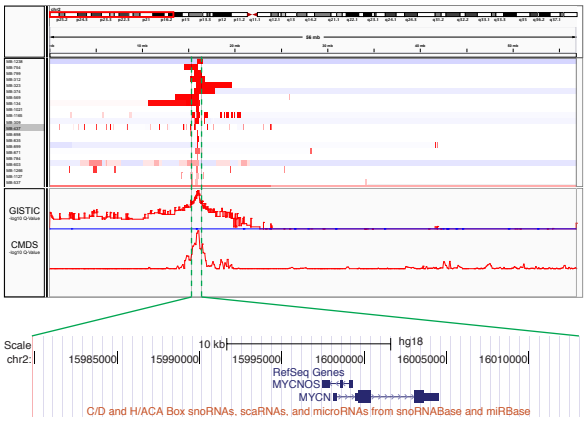
Group 4 Amplified Regions: Part A

1



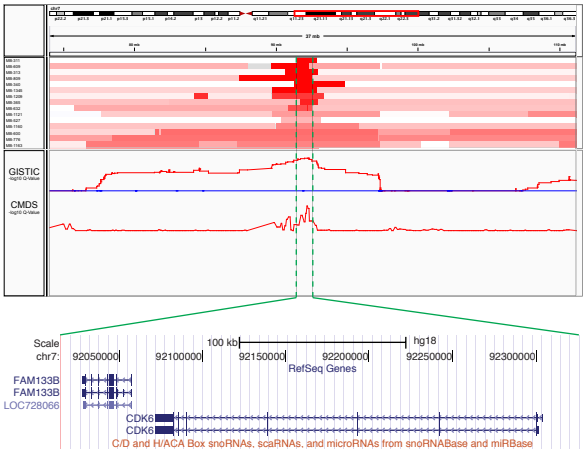
chr5:121784784-121807288

2



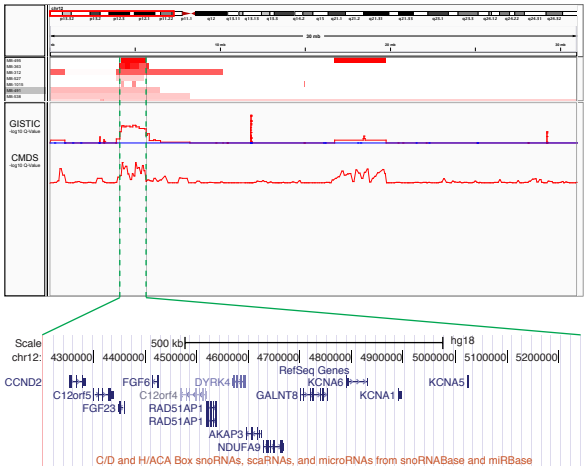
chr2:15979915-16013340

3



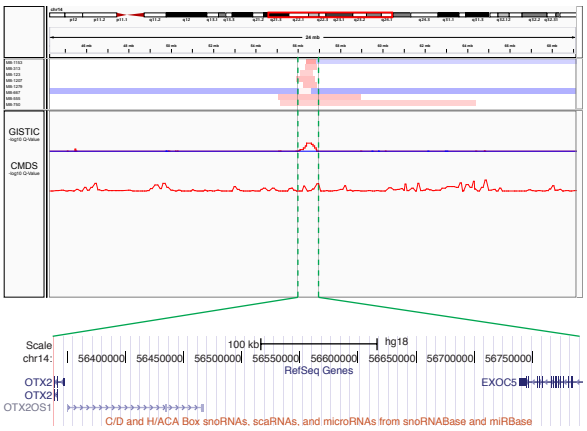
chr7:92015629-92327963

4



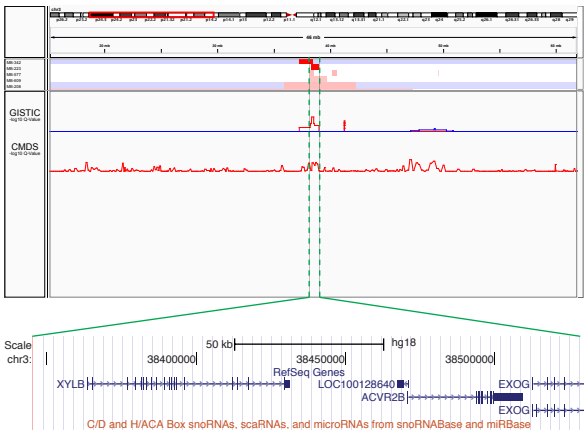
chr12:4204570-5249490

5



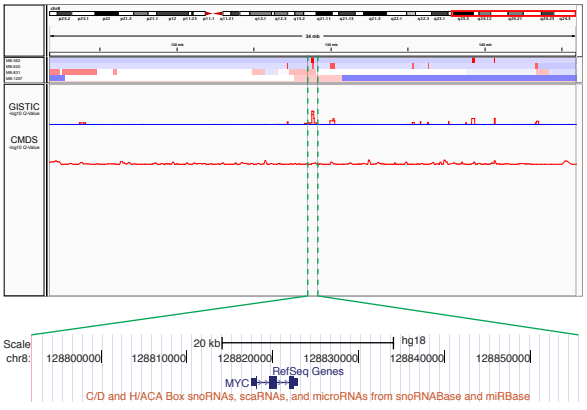
chr14:56338942-56793422

6



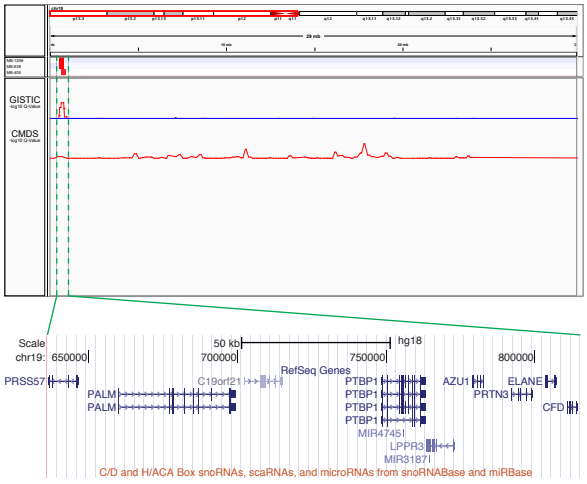
chr3:38345561-38529959

7



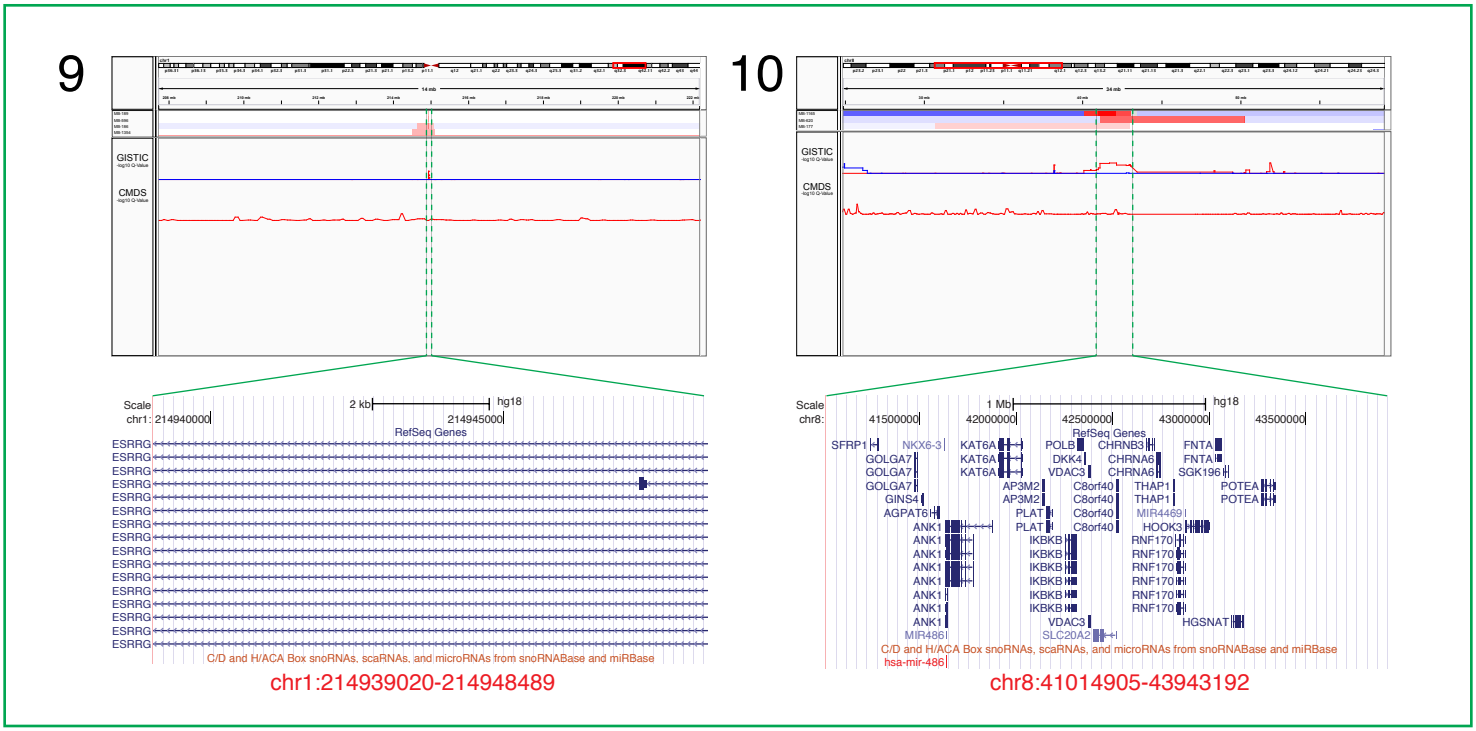
chr8:128792106-128856003

8



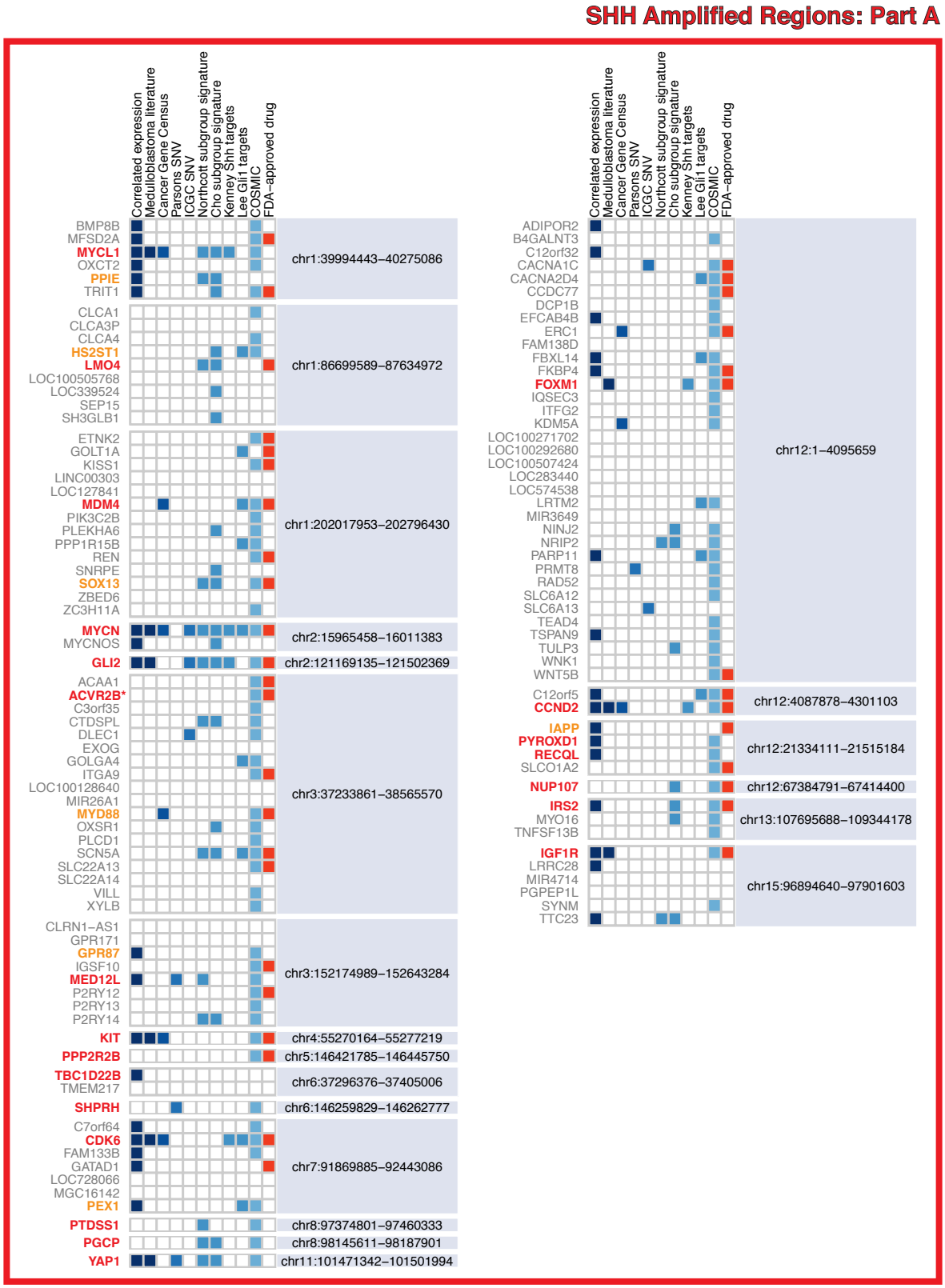
chr19:636205-816405

Group 4 Amplified Regions: Part B

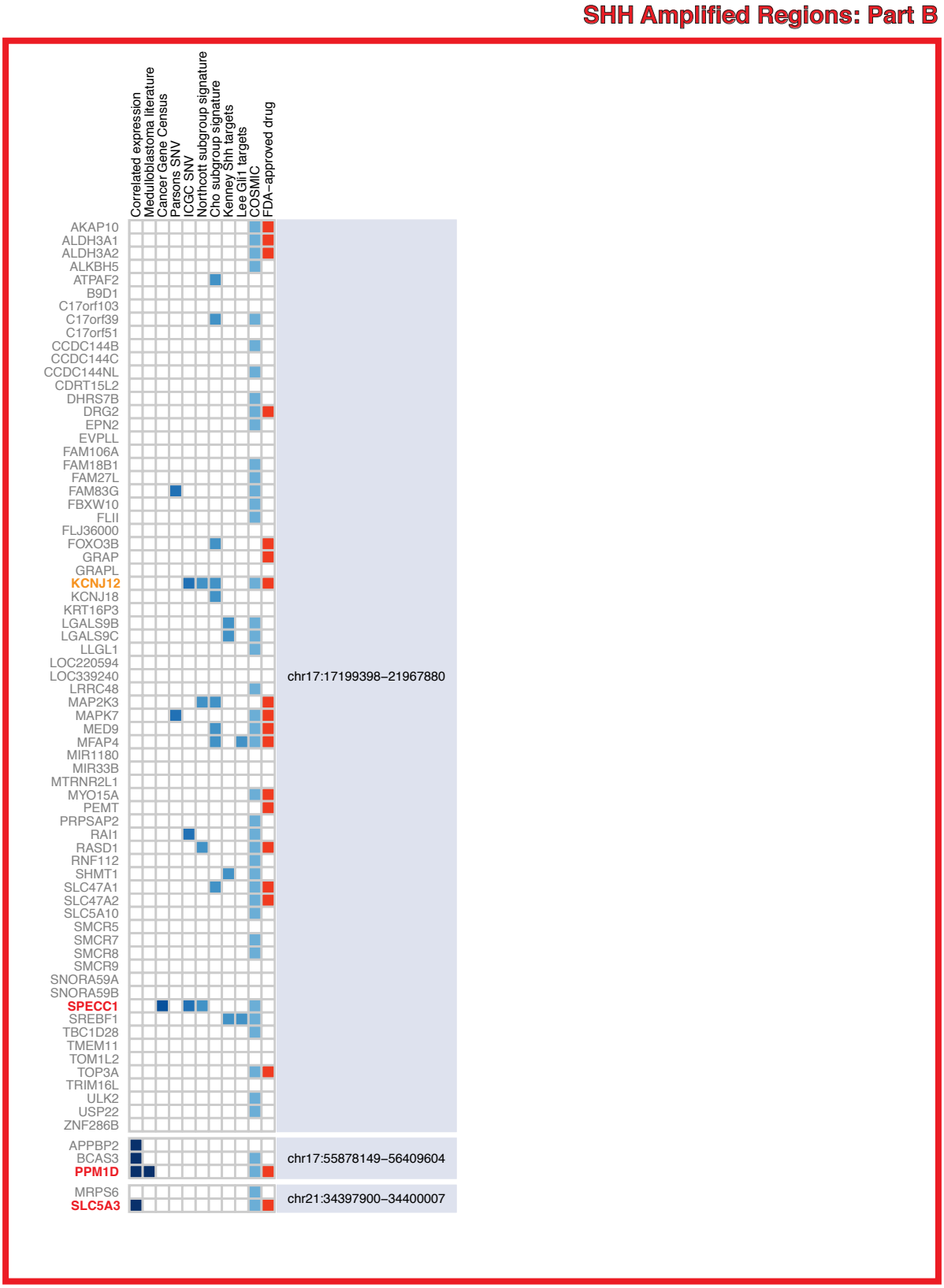


Supplementary Figure 14. Highly significant regions of focal amplification in Group 4 medulloblastoma. SNP6 copy number output for the top regions of focal amplification identified in Group 4 as determined by GISTIC2 analysis. Segmented copy number data is shown for Group 4 samples harboring amplification/gain in the regions of interest. Significance values ($-\log_{10}$ q-value) from GISTIC2 and CMDS analysis are included below the copy number profiles. Output is from the Integrative Genomics Viewer (IGV). UCSC RefSeq genes mapping to each significant region are shown below the IGV output. Regions have been sorted according to significance beginning with the most significant.

Supplementary Figure 15, Northcott, Shih et al



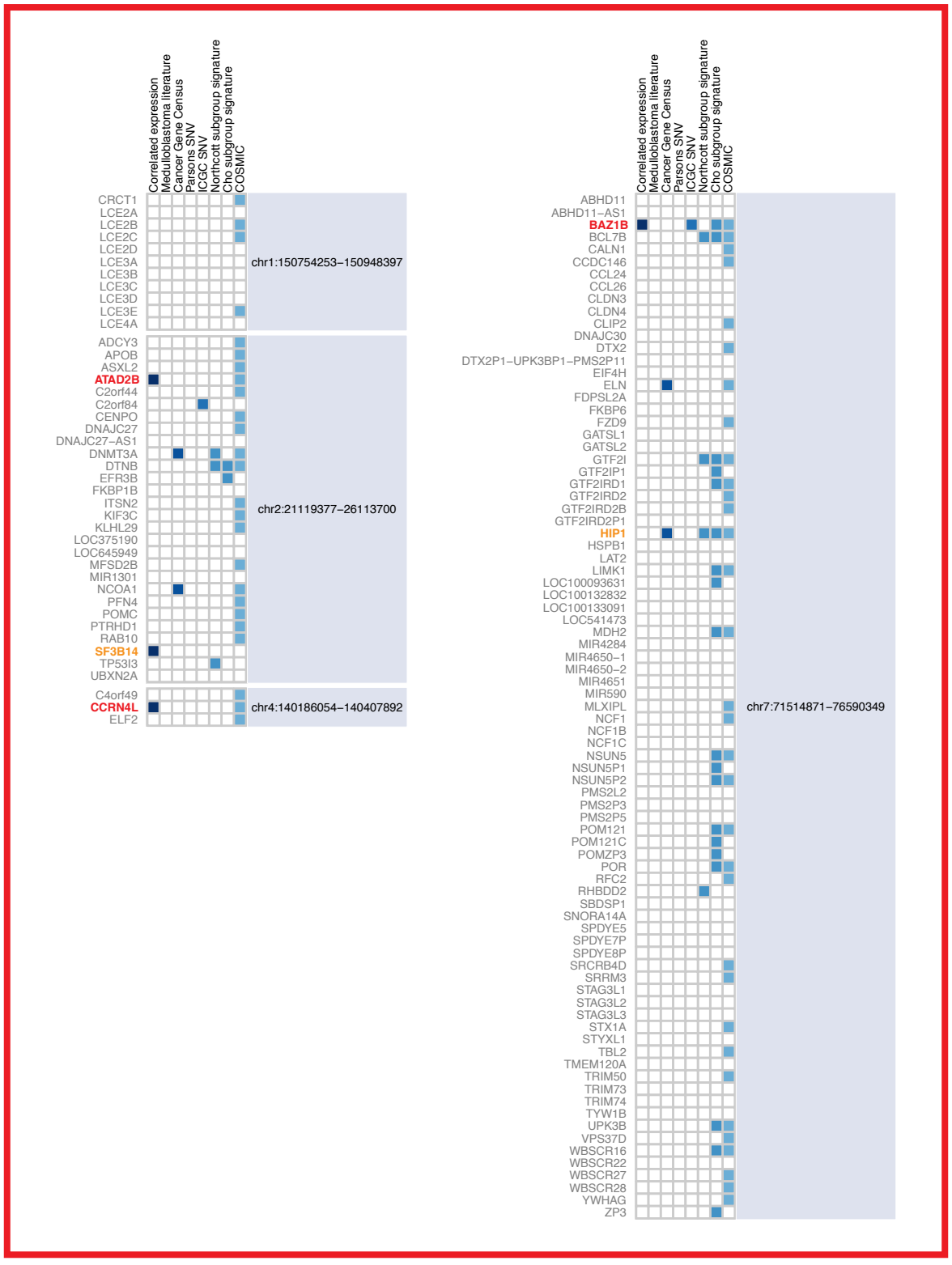
Supplementary Figure 15, Northcott, Shih et al



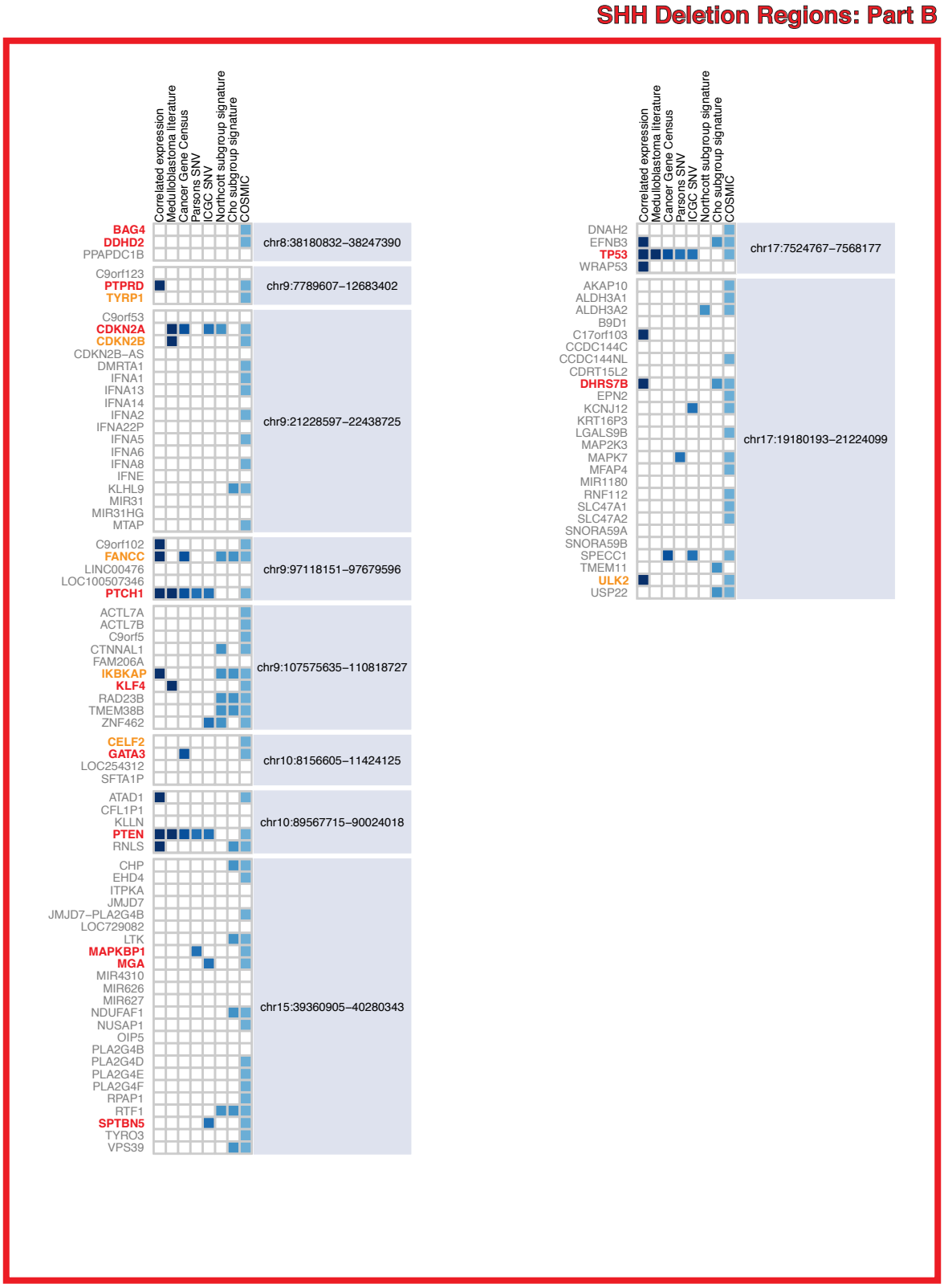
Supplementary Figure 15. Evidence-based ranking of candidate target genes in SHH amplification regions. Candidate genes from significant GISTIC2 amplification regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, Northcott subgroup signature, Cho subgroup signature, Kenney Shh targets, Lee Gli1 targets; V, COSMIC. The availability of FDA-approved drugs is shown in orange, but it is not included in the ranking. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

Supplementary Figure 16, Northcott, Shih et al

SHH Deletion Regions: Part A

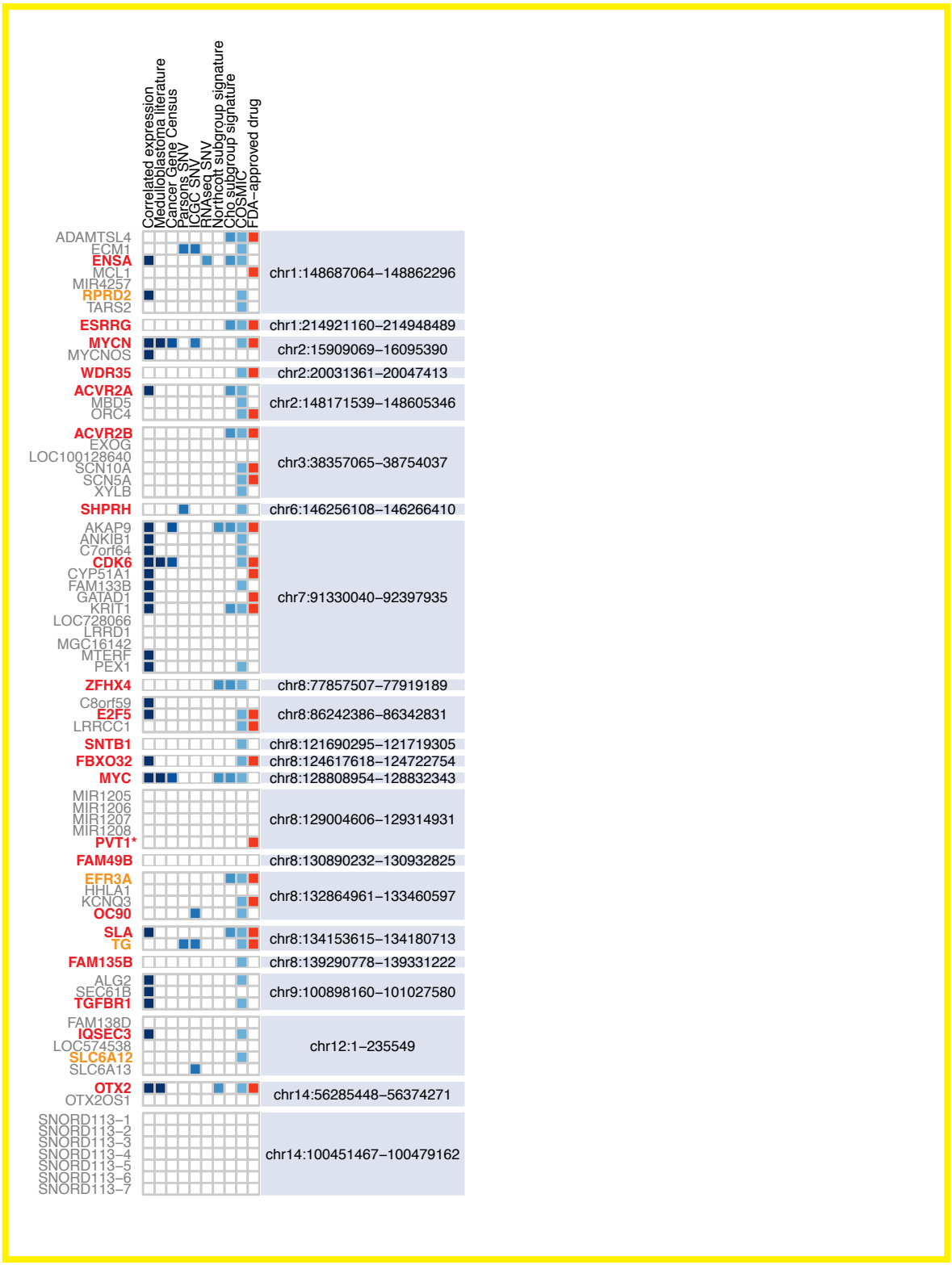


Supplementary Figure 16, Northcott, Shih et al



Supplementary Figure 16. Evidence-based ranking of candidate target genes in SHH deletion regions. Candidate genes from significant GISTIC2 deletion regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, Northcott subgroup signature, Cho subgroup signature; V, COSMIC. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

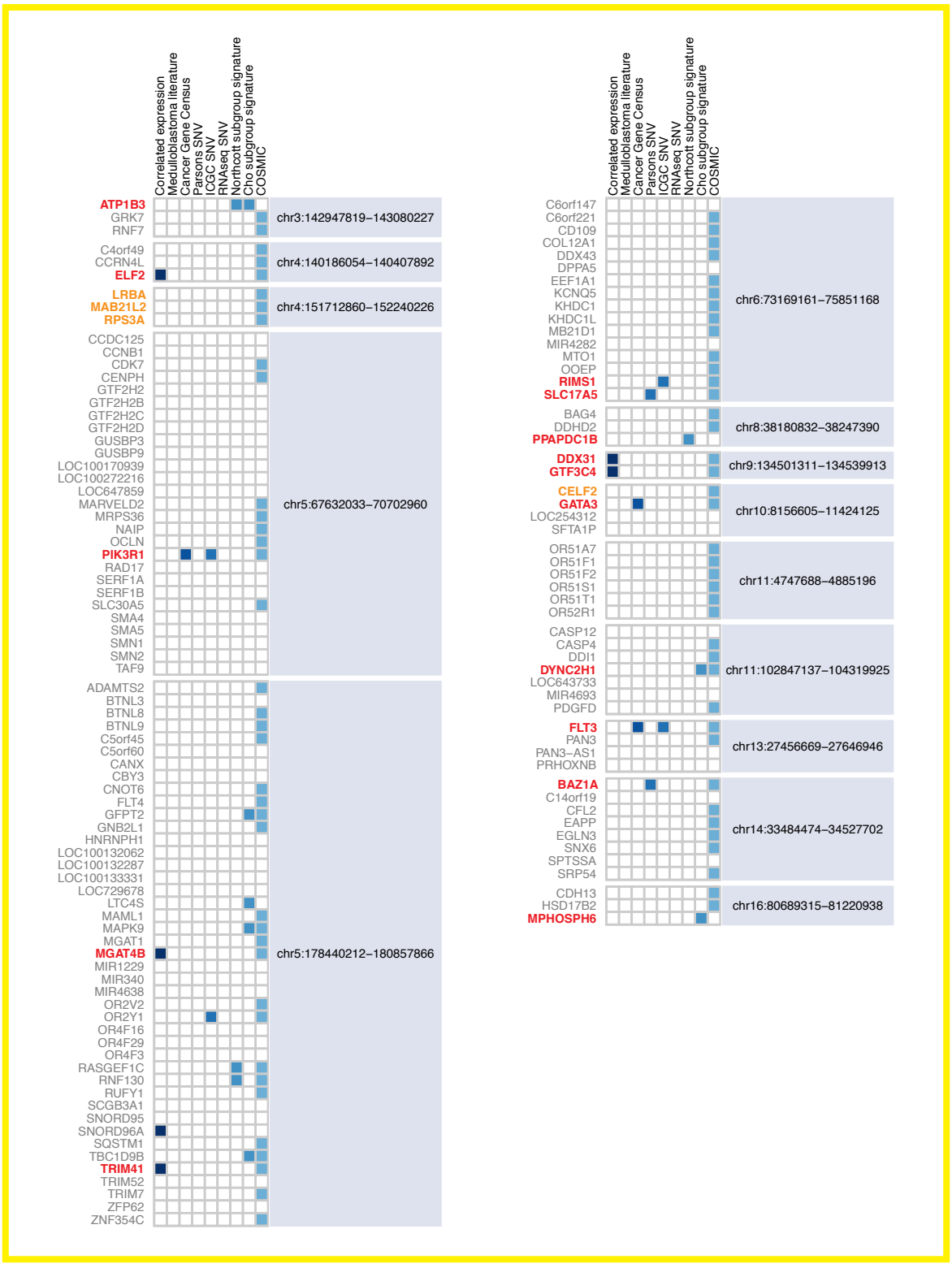
Supplementary Figure 17, Northcott, Shih et al



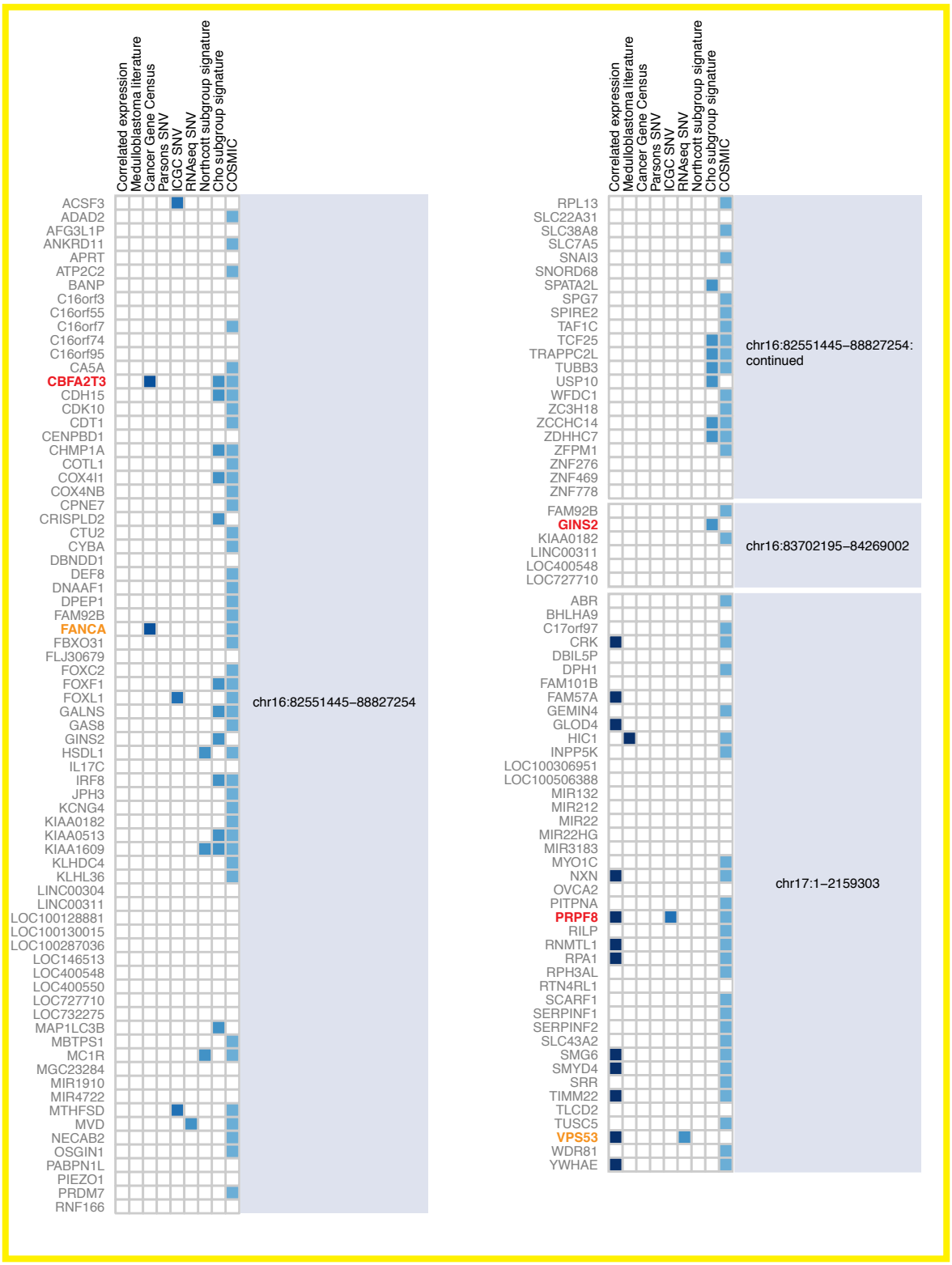
Supplementary Figure 17. Evidence-based ranking of candidate target genes in Group 3 amplification regions. Candidate genes from significant GISTIC2 amplification regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, RNAseq SNV, Northcott subgroup signature, Cho subgroup signature; V, COSMIC. The availability of FDA-approved drugs is shown in orange, but it is not included in the ranking. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

Supplementary Figure 18, Northcott, Shih et al

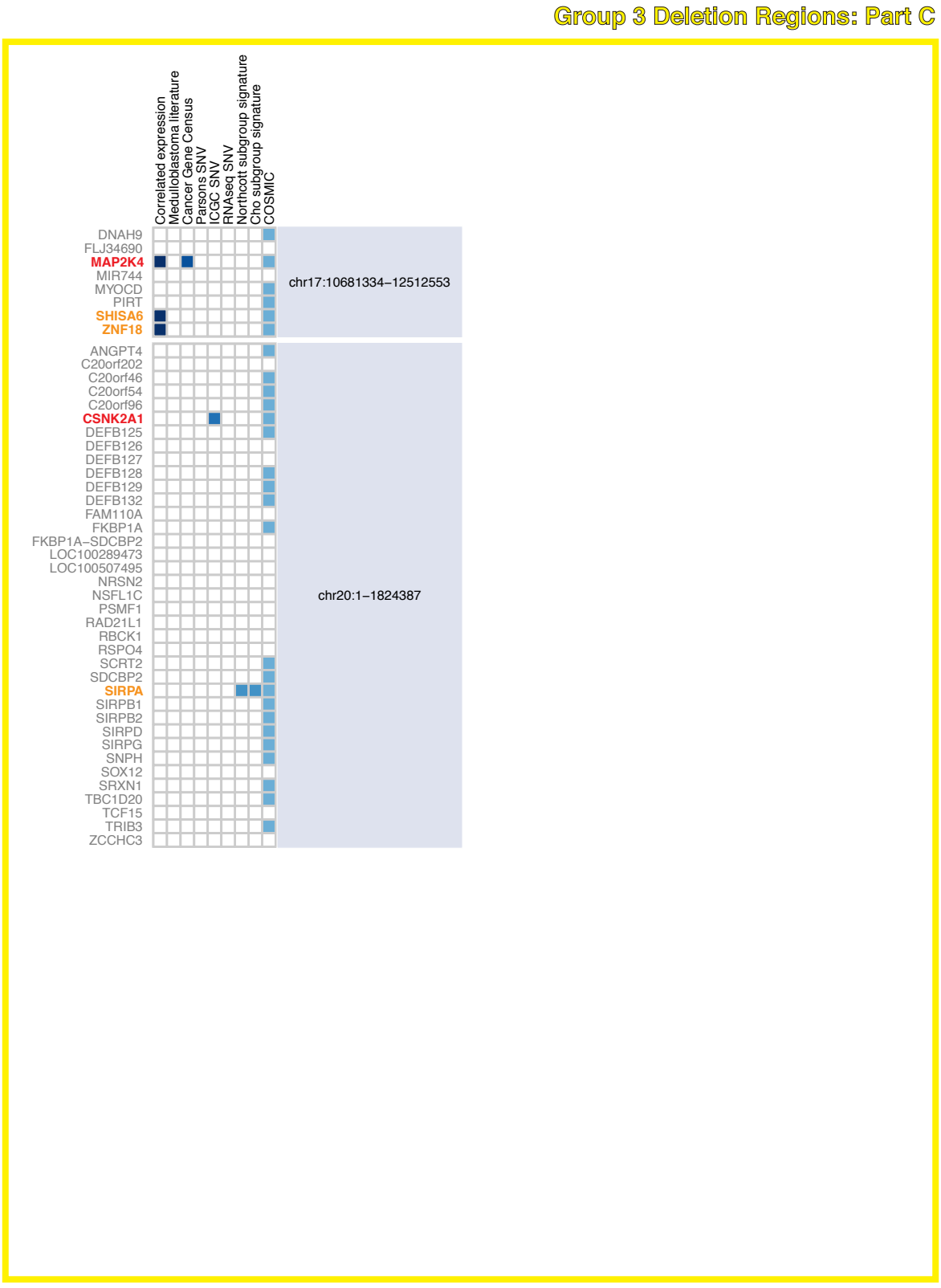
Group 3 Deletion Regions: Part A



Supplementary Figure 18, Northcott, Shih et al

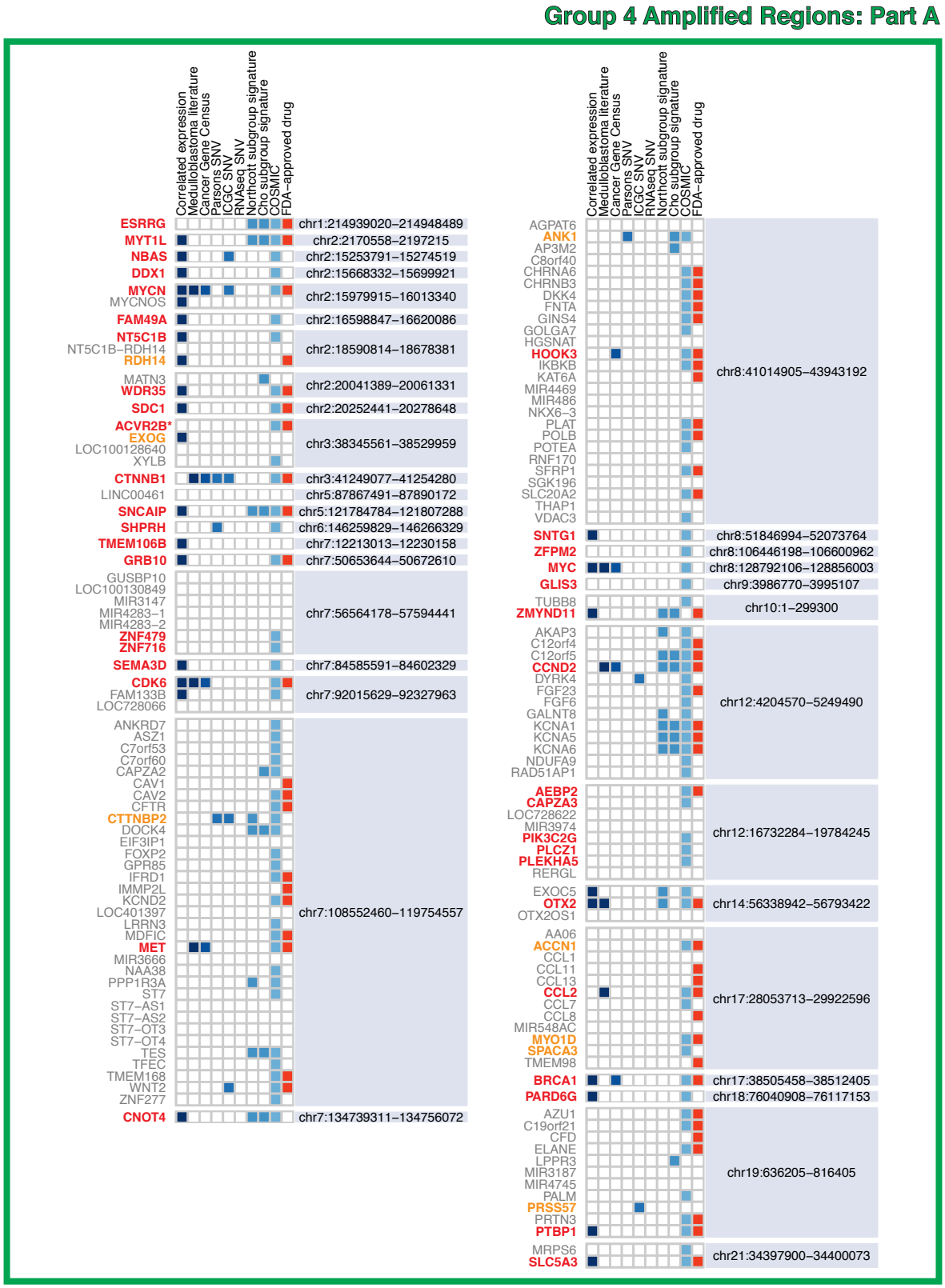


Supplementary Figure 18, Northcott, Shih et al



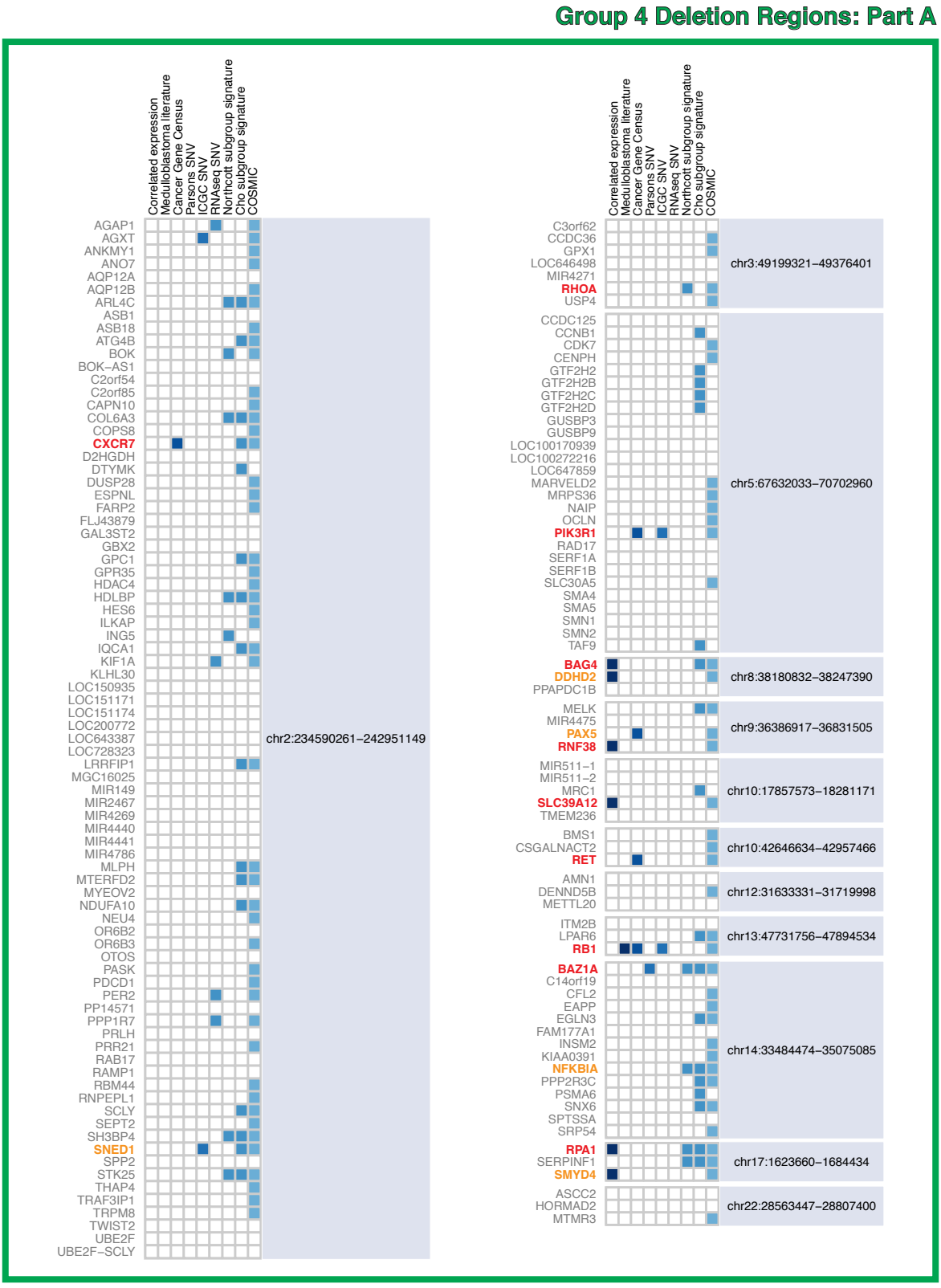
Supplementary Figure 18. Evidence-based ranking of candidate target genes in Group 3 deletion regions. Candidate genes from significant GISTIC2 deletion regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, RNAseq SNV, Northcott subgroup signature, Cho subgroup signature; V, COSMIC. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

Supplementary Figure 19, Northcott, Shih et al



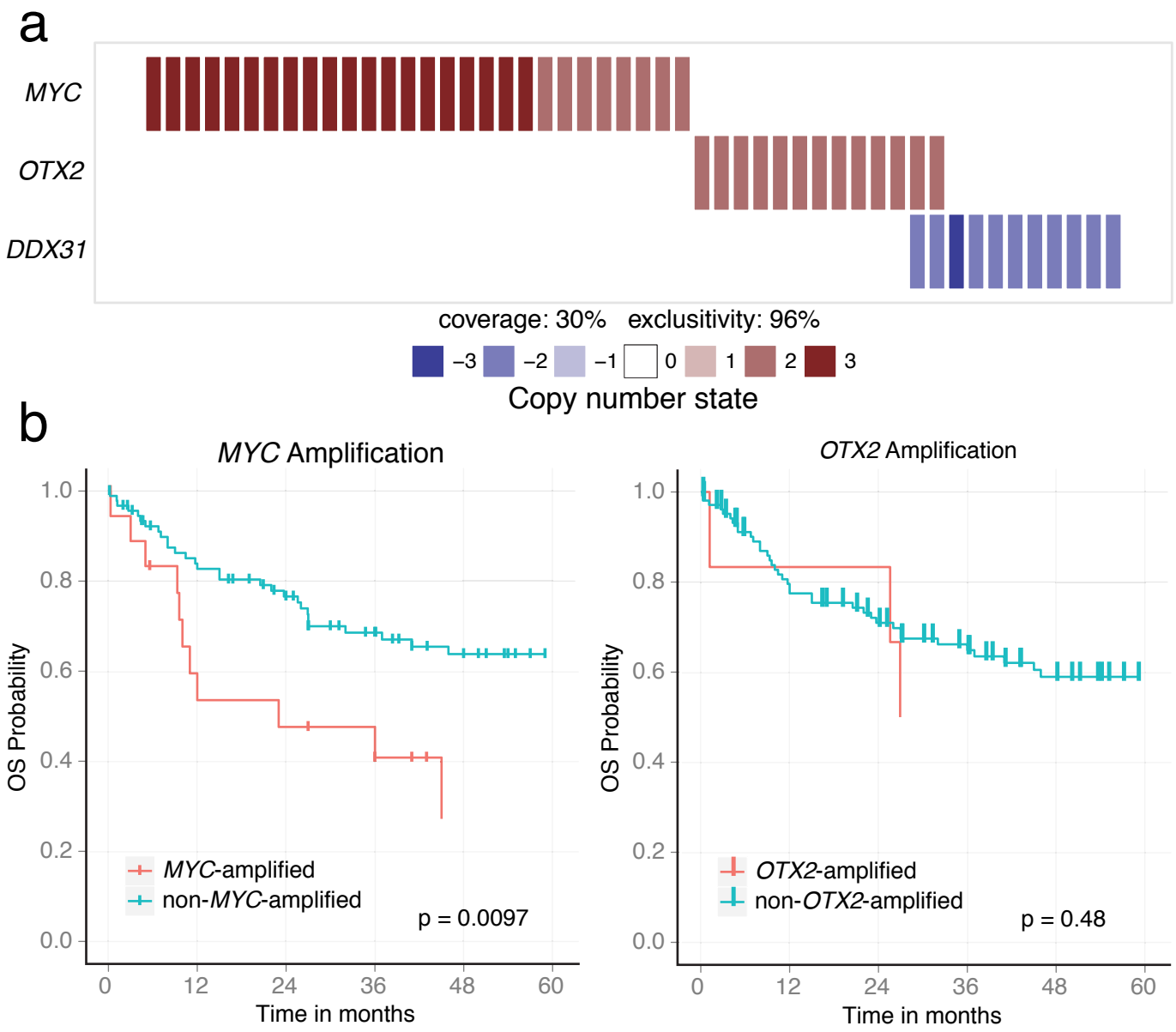
Supplementary Figure 19. Evidence-based ranking of candidate target genes in Group 4 amplification regions. Candidate genes from significant GISTIC2 amplification regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, RNAseq SNV, Northcott subgroup signature, Cho subgroup signature; V, COSMIC. The availability of FDA-approved drugs is shown in orange, but it is not included in the ranking. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

Supplementary Figure 20, Northcott, Shih et al



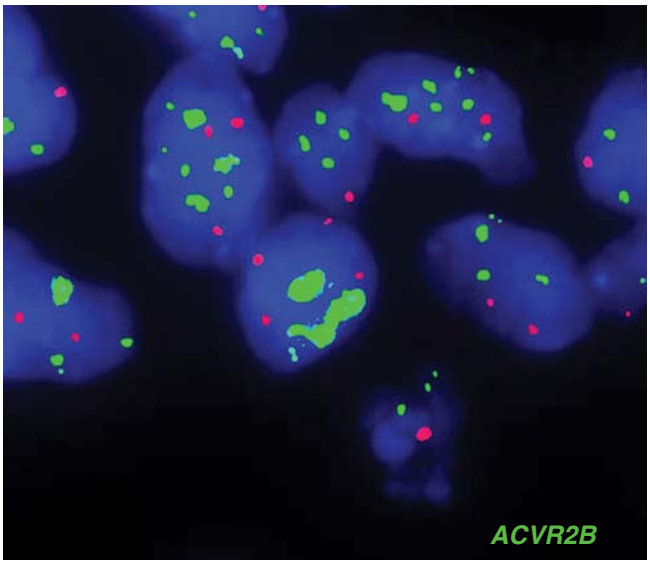
Supplementary Figure 20. Evidence-based ranking of candidate target genes in Group 4 deletion regions. Candidate genes from significant GISTIC2 deletion regions are depicted by evidence ranking grids. The top candidates appear in bold (red: first-rank candidates, orange: second-rank candidates). Within each region, genes are ranked based on availability of supporting evidence from the listed dataset. Shades of blue reflect the weight of each dataset, which are prioritized by tiers: I, Correlated expression, Medulloblastoma literature; II, Cancer Gene Census; III, Parsons SNV, ICGC SNV; IV, RNAseq SNV, Northcott subgroup signature, Cho subgroup signature; V, COSMIC. Candidate driver genes were inferred by considering evidence of each tier in increasing order.

Supplementary Figure 21, Northcott, Shih et al



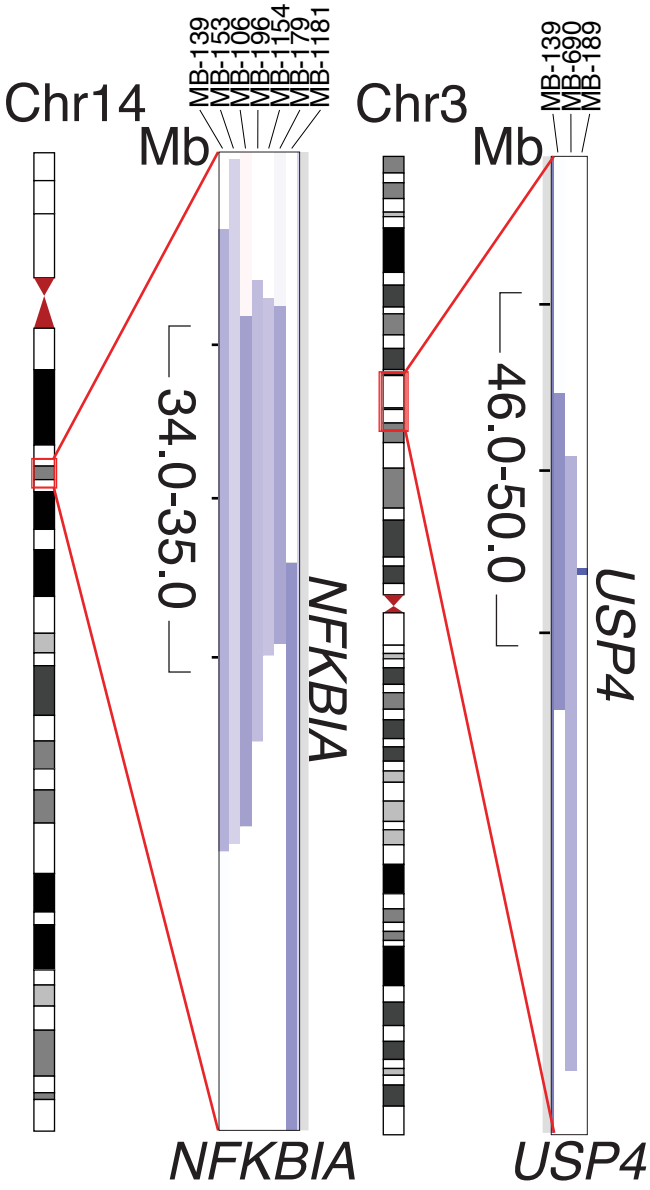
Supplementary Figure 21. Significance of *MYC* amplification in Group 3 medulloblastoma. a, Mutual exclusivity analysis of top focal SCNAs in Group 3 shows that *MYC* and *OTX2* aberrations are completely exclusive in this cohort. **b,** Prognostic significance of *MYC* and *OTX2* aberrations in Group 3 (log-rank tests).

Supplementary Figure 22, Northcott, Shih et al



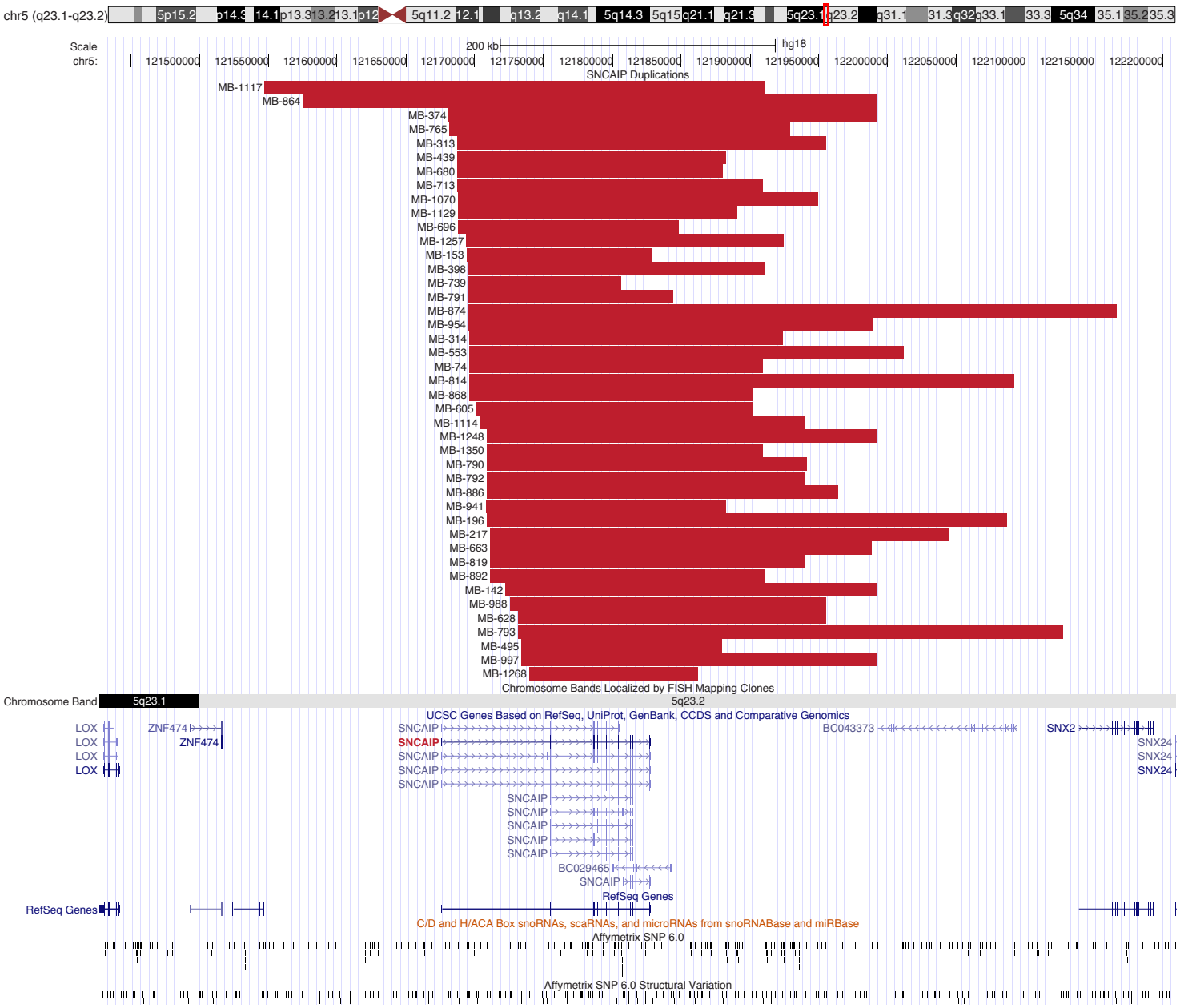
Supplementary Figure 22. Validation of *ACVR2B* amplification in medulloblastoma. FISH validation of *ACVR2B* amplification in a non-overlapping series of medulloblastomas present on a TMA.

Supplementary Figure 23, Northcott, Shih et al



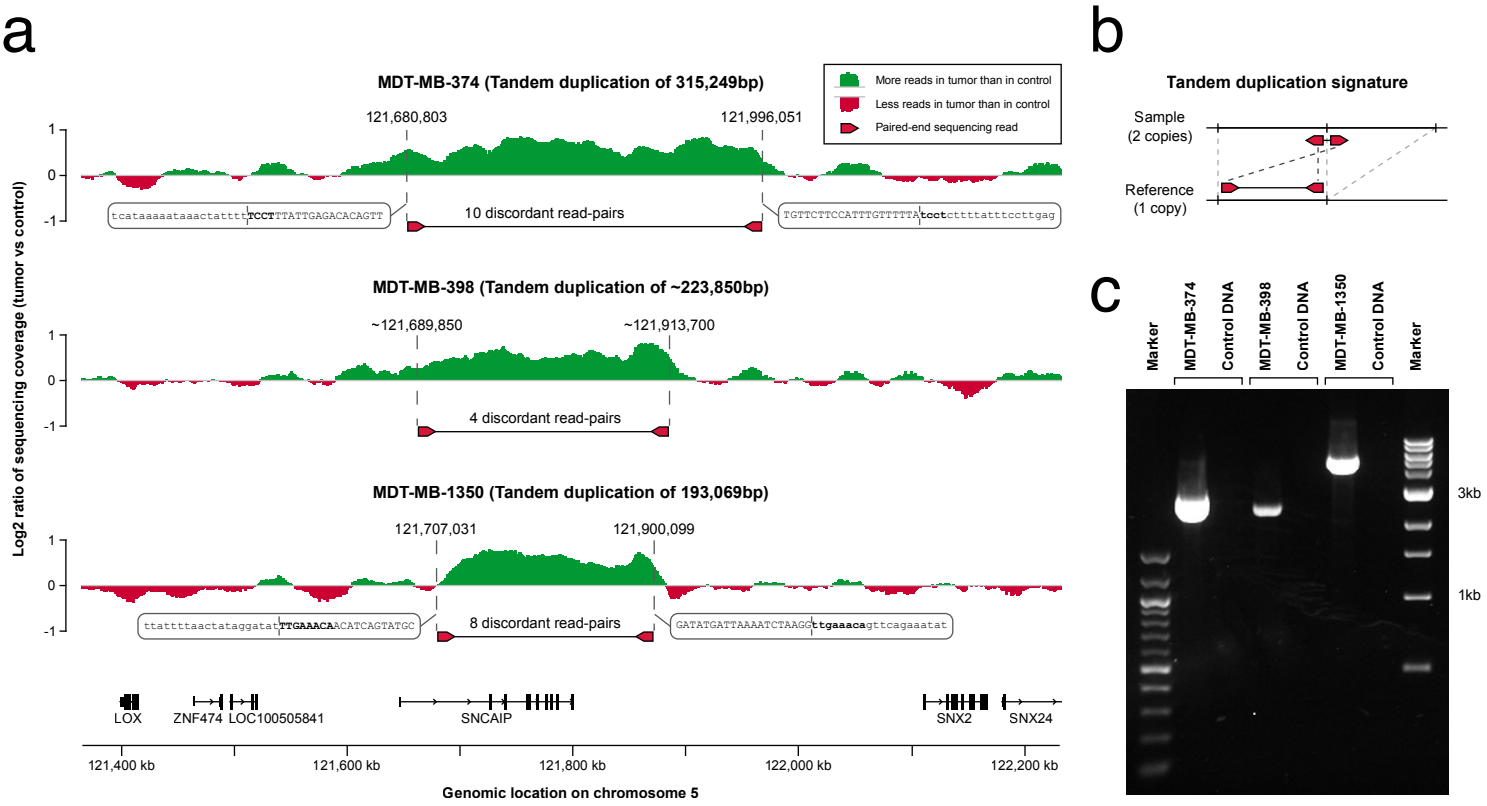
Supplementary Figure 23. Recurrent focal deletions affecting negative regulators of the NF-κB pathway in Group 4. Copy number heatmaps demonstrate recurrent focal deletions of *NFKBIA* and *USP4* in Group 4.

Supplementary Figure 24, Northcott, Shih et al



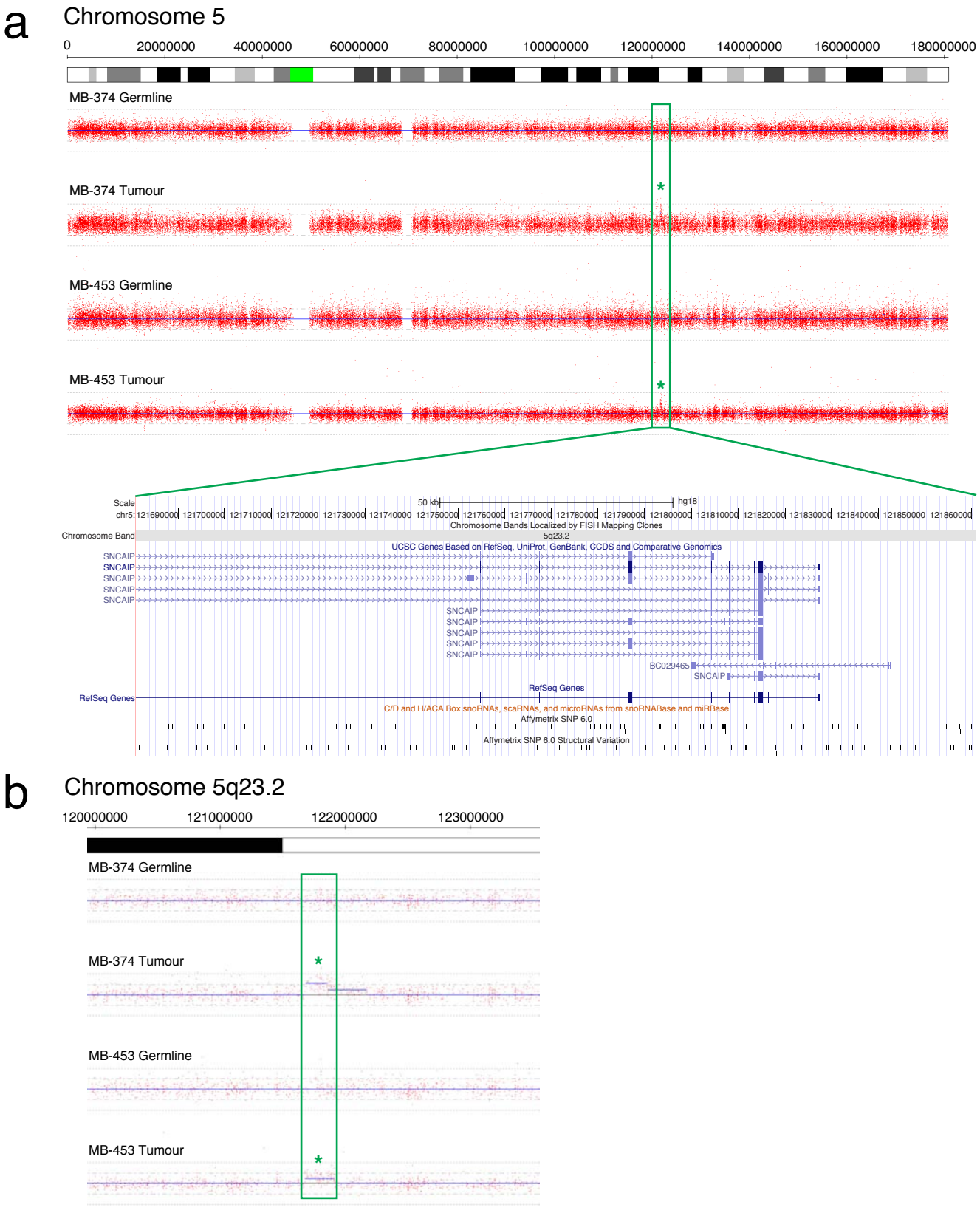
Supplementary Figure 24. Tandem duplication of *SNCAIP* on chr5q23.2 in Group 4 medulloblastoma. UCSC Genome Browser output (hg18) of the *SNCAIP* locus and neighboring genes on chromosome 5q23.2. A custom track ('SNCAIP Duplications') showing copy number segments (gains) in 43 medulloblastoma samples with *SNCAIP* duplication (highlighted red) has been included above the gene content window. All SCNAs shown above directly target *SNCAIP*; in 42/43 cases, *SNCAIP* is the only RefSeq gene affected.

Supplementary Figure 25, Northcott, Shih et al



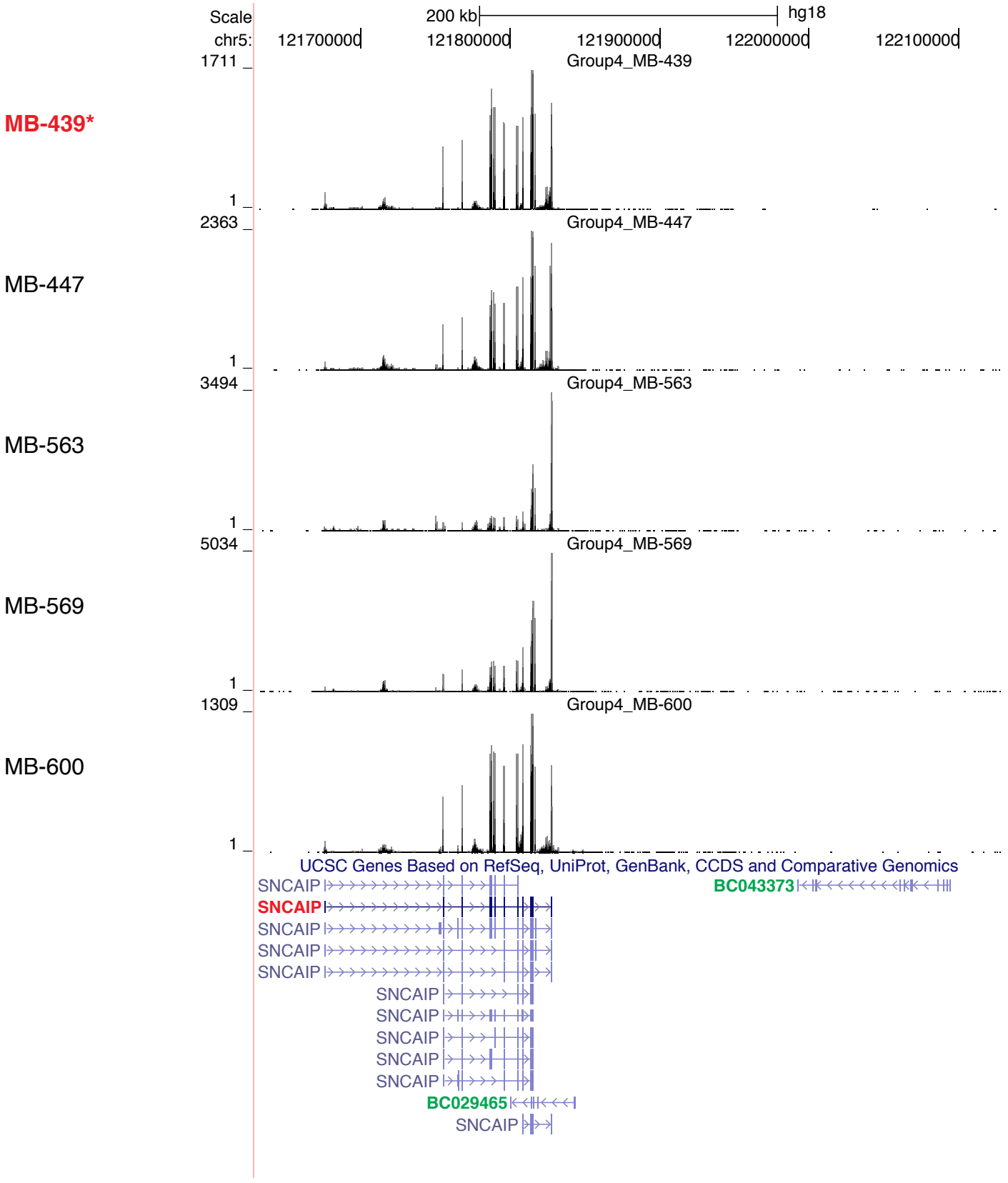
Supplementary Figure 25. Paired-end sequencing and reconstruction of tandem duplications overlapping the *SNCAIP* gene. **a,** Log_2 ratio of sequencing coverage (tumor vs control) of three Group 4 samples with SNP6-inferred *SNCAIP* copy number gains. Positive values (shown in green) indicate regions with more sequenced reads in the tumour compared to the control sample. Negative values (shown in red) indicate regions with less sequenced reads in the tumour compared to the control sample. The dashed lines mark the start and end of the duplications. Below each plot, the number of discordantly mapping read-pairs, identified by high-resolution and massive paired-end mapping, are shown. For MDT-MB-374 and MDT-MB-1350, the exact breakpoint position of the respective tandem duplication was determined by Sanger sequencing. DNA sequences corresponding to the breakpoint junctions are indicated. Uppercase sequences are within the tandem duplication. Nucleotides in bold highlight microhomologies. **b,** Scheme showing the mapping of read-pairs overlapping a tandem duplication breakpoint. The read orientation applies to long-range paired-end libraries (~4.5kb insert size) sequenced with an Illumina HiSeq2000 instrument. Inward pointing read-pairs with an enlarged insert size indicate the presence of a tandem duplication. **c,** PCR validation of the three tandem duplications. Primers were placed near both inferred breakpoint locations, in such a way that the PCR amplicon would span the breakpoint junctions. As control, commercially available human DNA was used.

Supplementary Figure 26, Northcott, Shih et al



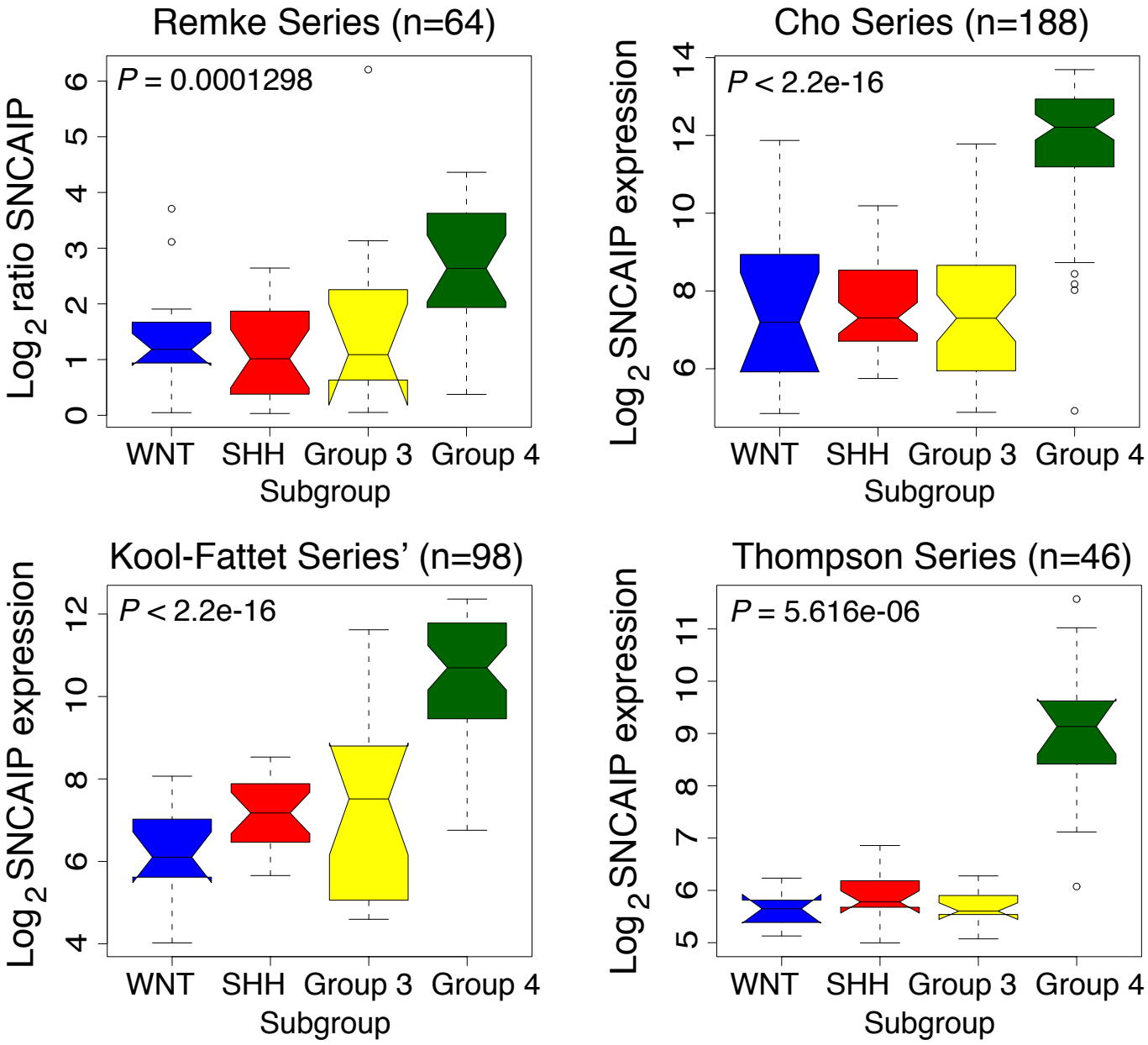
Supplementary Figure 26. Somatic copy number gain of *SNCAIP* in Group 4 medulloblastoma. **a,** SNP6 copy number plots of chromosome 5 for two matched tumour/normal pairs, MB-374 and MB-453, are shown. Copy number gains affecting *SNCAIP* (marked with an asterisk) are found in both tumour samples but not in the corresponding germline cases. **b,** Zoomed-in view of copy number profiles in (**a**), highlighting somatic gain of the *SNCAIP* locus on chr5q23.2.

Supplementary Figure 27, Northcott, Shih et al



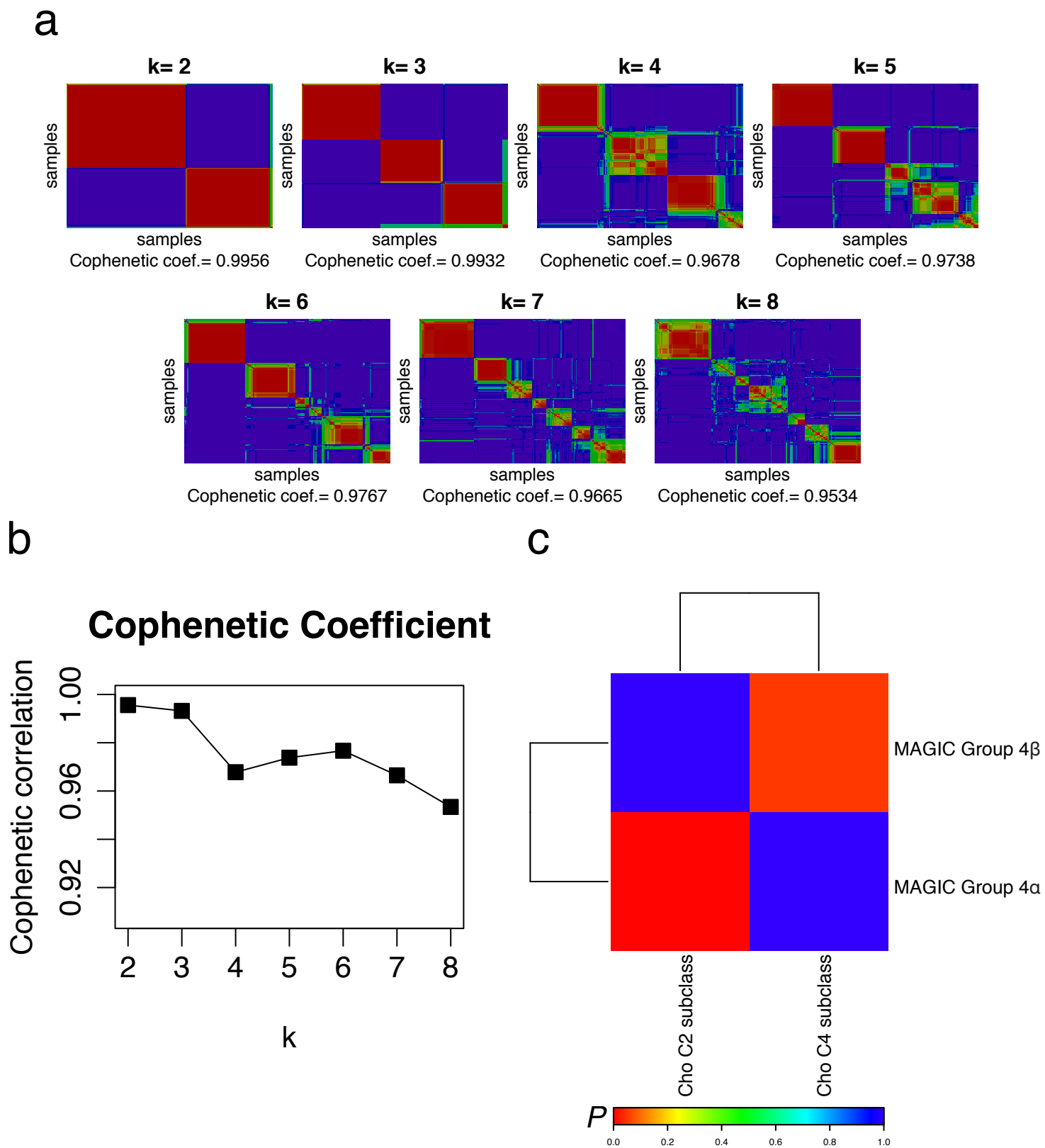
Supplementary Figure 27. SNCAIP expression in Group 4 medulloblastomas as measured by RNASeq. Read depth plots of RNASeq data for five Group 4 medulloblastomas provides strong evidence of high SNCAIP expression in these samples whereas there is no support for expression of the overlapping (BCO29465) and downstream (BCO43373) genes. MB-439 (marked in red with an asterisk) harbours the typical *SNCAIP* duplication observed in Group 4.

Supplementary Figure 28, Northcott, Shih et al



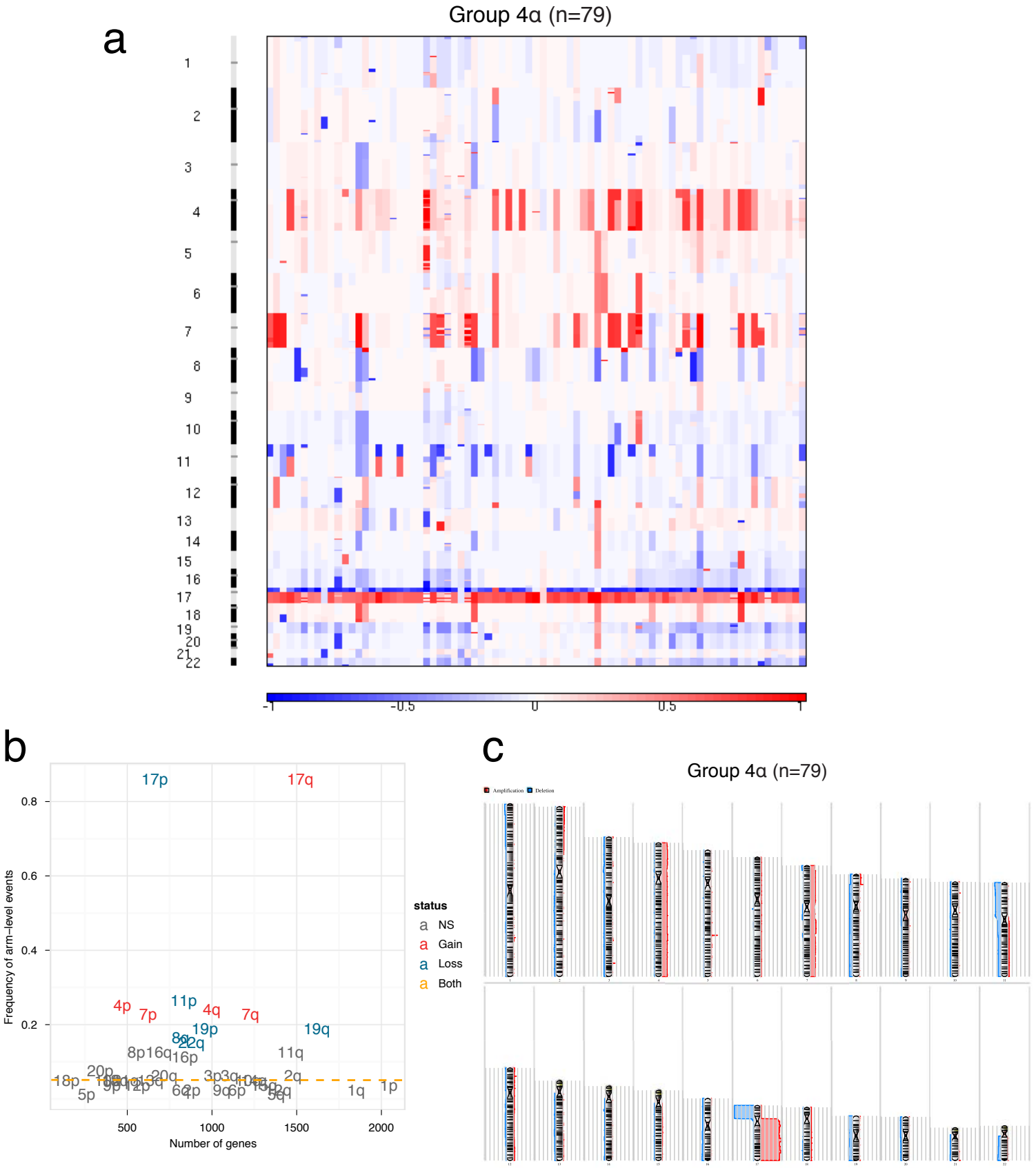
Supplementary Figure 28. Validation of SNCAIP as a Group 4 signature gene across five published medulloblastoma expression datasets. SNCAIP expression was analyzed by subgroup using expression array data published in recent studies by Thompson, Kool, Fattet, Cho, and Remke, totaling 396 cases profiled on four different array platforms. In all datasets, SNCAIP exhibited highest expression in Group 4.

Supplementary Figure 29, Northcott, Shih et al



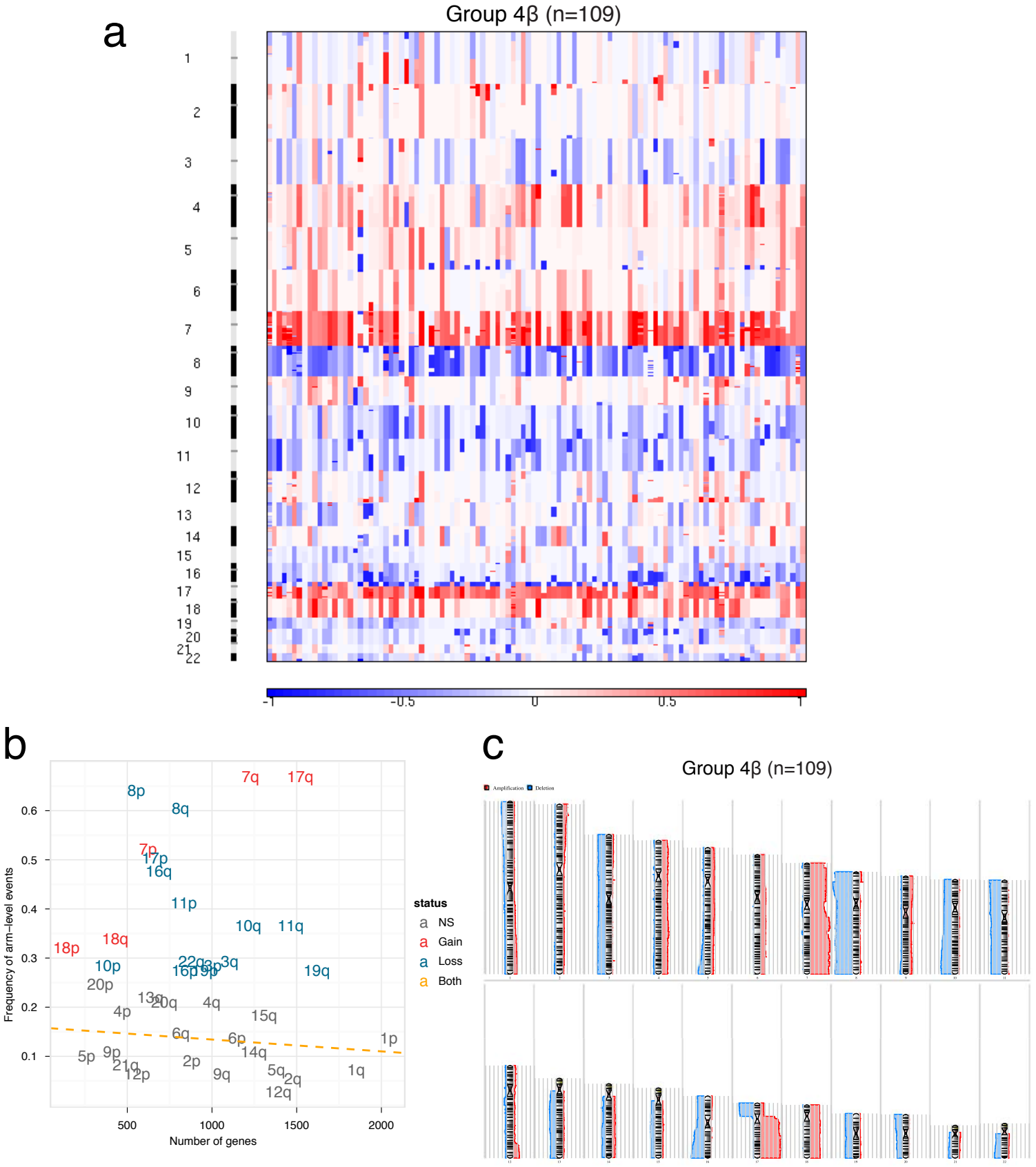
Supplementary Figure 29. Identification of Group 4 medulloblastoma subtypes. a, Consensus NMF clustering of Group 4 gene expression array data. 188 Group 4 medulloblastomas were profiled by Affymetrix Gene Array 1.1 and subjected to consensus clustering in GenePattern. Heatmaps for k=2-8 are shown. **b,** Cophenetic Coefficient plot for the clustering results in (**a**) demonstrates the greatest support for k=2 subtypes of Group 4. **c,** SubClass mapping results for Group 4 expression data. The two Group 4 subtypes (designated MAGIC Group 4α and 4β) identified in the current study were compared with the subtypes reported by Cho et al (C2 and C4). FDR matrix output by SubMap indicates Group 4α=C2 and 4β=C4.

Supplementary Figure 30, Northcott, Shih et al



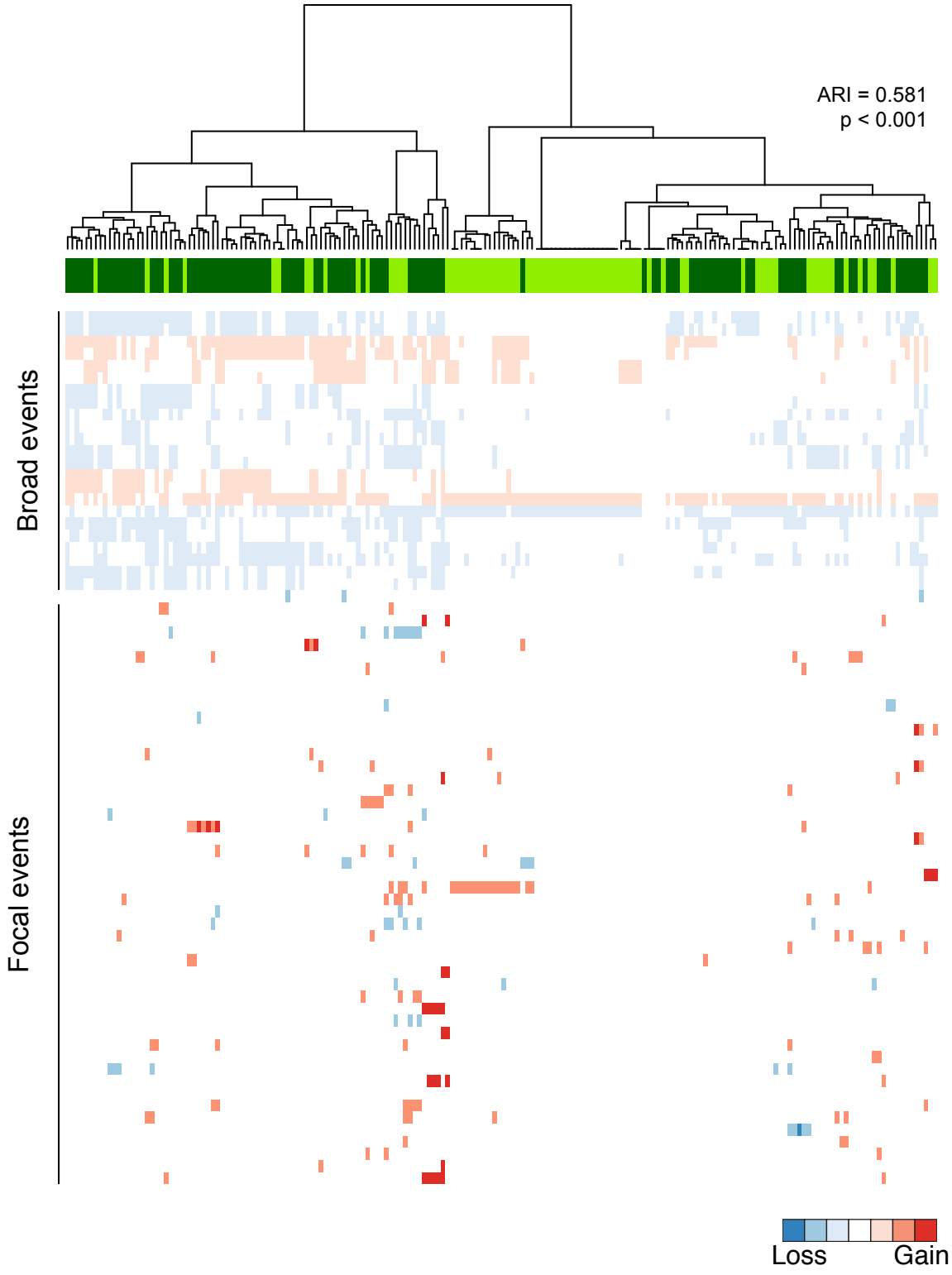
Supplementary Figure 30. SNP6 copy number profile of Group 4α medulloblastoma. a, Copy number heatmap for 79 Group 4α medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in Group 4α tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in Group 4α medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 31, Northcott, Shih et al



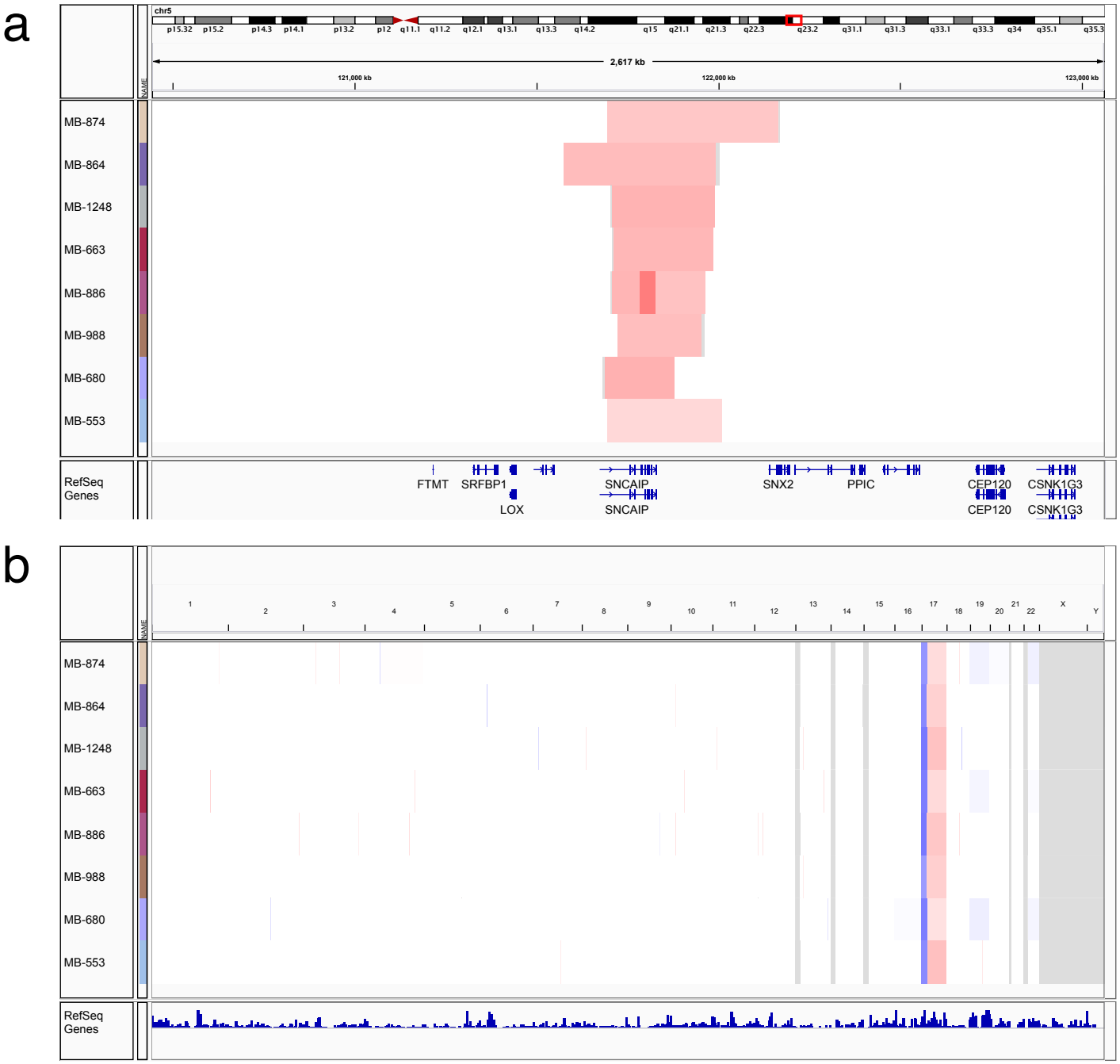
Supplementary Figure 31. SNP6 copy number profile of Group 4β medulloblastoma. **a,** Copy number heatmap for 109 Group 4β medulloblastomas profiled by SNP6. Amplifications are shown in red, deletions in blue. **b,** Broad cytogenetic events in Group 4β tumours. Significant chromosome arm-level gains are depicted in red, arm-level losses in blue. **c,** Summary plot of gains and losses in Group 4β medulloblastoma. Ideogram shows the relative distribution of broad gains (red) and losses (blue) across the 22 autosomes. Output from Partek Genomics Suite.

Supplementary Figure 32, Northcott, Shih et al



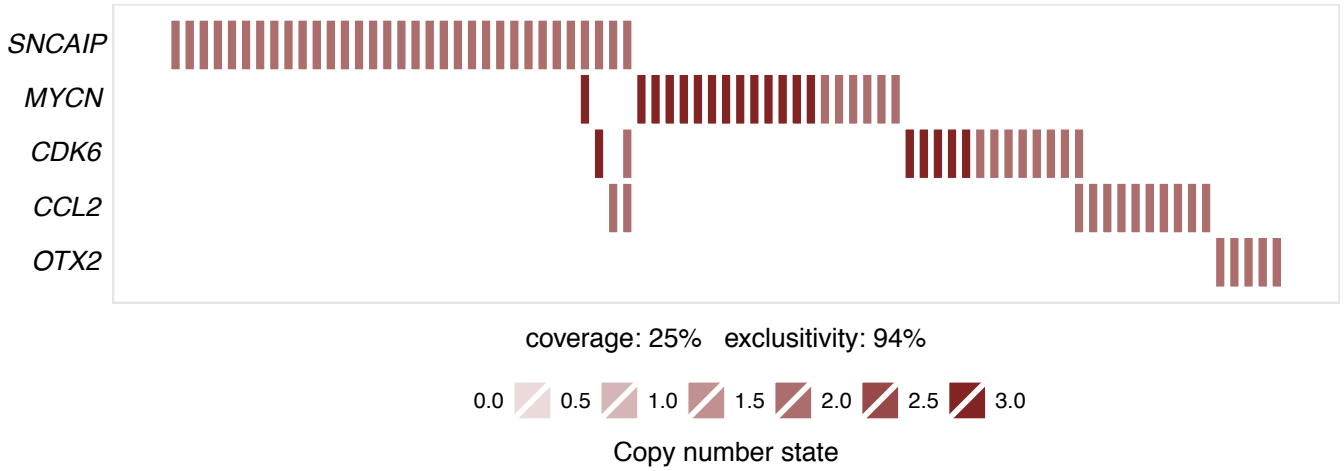
Supplementary Figure 32. Unsupervised hierarchical clustering of broad and focal SCNAs in Group 4 medulloblastoma. Coloured side bar indicates Group 4 subtypes (Group 4α: light green, Group 4β: dark green). Only focal events identified in the pan-cohort GISTIC2 analysis were included in the clustering. Resulting clusters show significant agreement with known Group 4 expression subtypes ($p < 0.001$, Chi-squared test). ARI, Adjusted Rand Index.

Supplementary Figure 33, Northcott, Shih et al

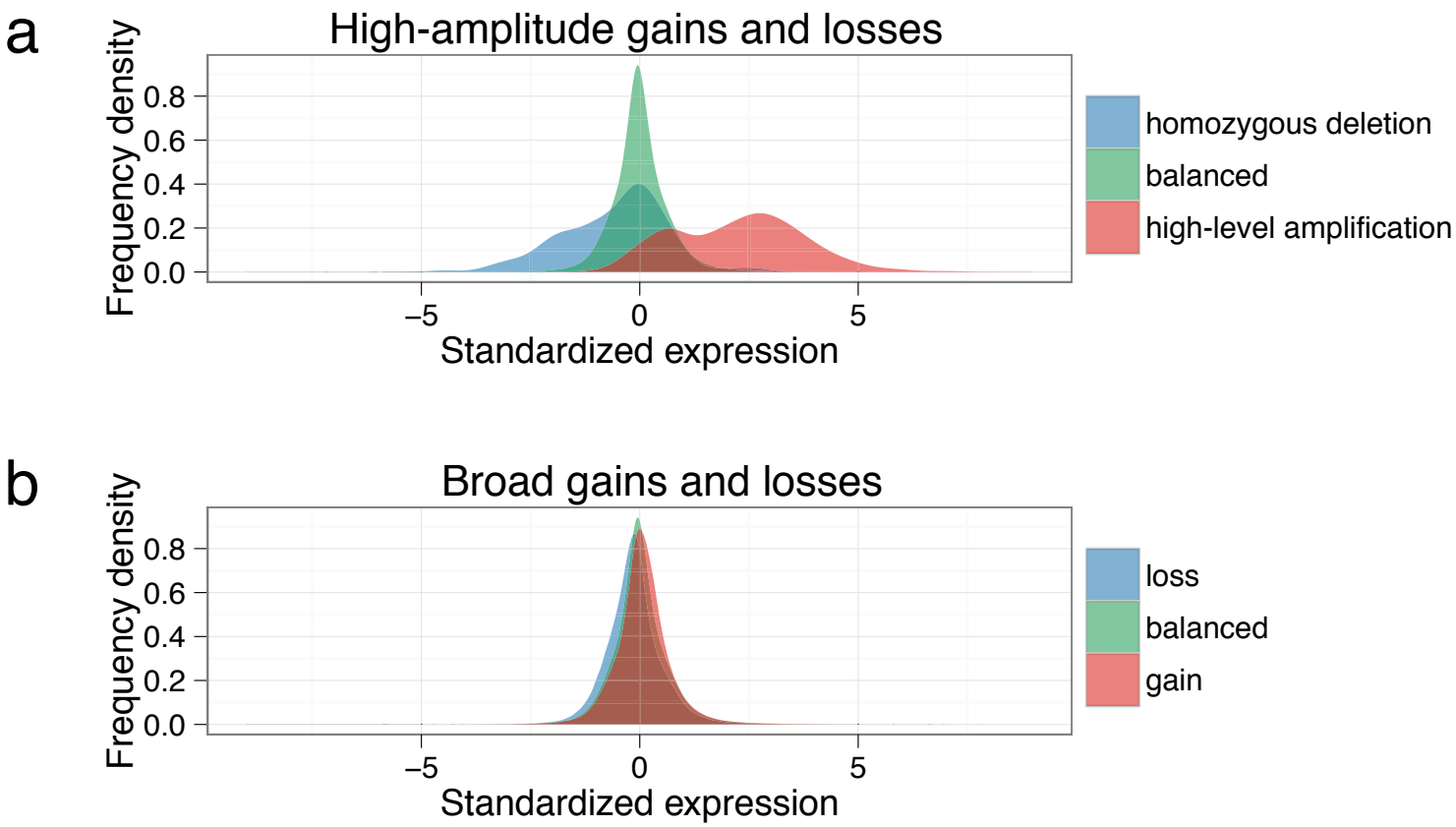


Supplementary Figure 33. Cytogenetic profile of *SNCAIP* duplicated Group 4α medulloblastomas. a, Copy number heatmap for a subset of Group 4α medulloblastomas (n=8) showing single copy gain of *SNCAIP* on chr5q23.2. Output from IGV. **b,** Genome-wide copy number heatmap for the same Group 4α cases shown in (**a**). Broad cytogenetic aberrations are exceedingly rare in these cases, with the exception of isochromosome 17q (i{17}q), which is common to all of them. Output from IGV.

Supplementary Figure 34, Northcott, Shih et al

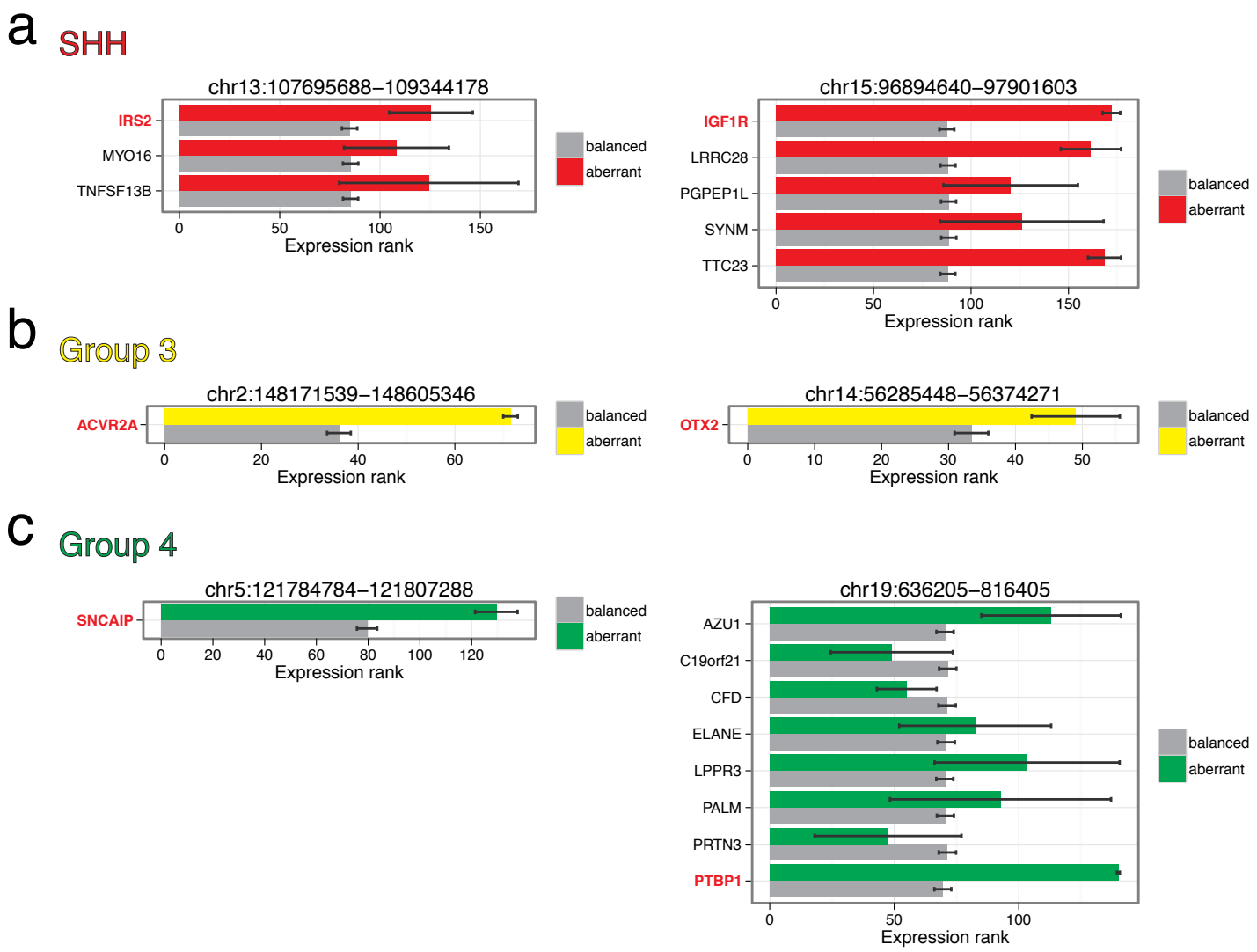


Supplementary Figure 34. Mutual exclusivity analysis of focal SCNAs in Group 4 medulloblastoma. Sample coverage plot shows the extent of overlap and/or exclusivity for top focal gains/amplifications in Group 4: *SNCAIP*, *MYCN*, *CDK6*, *CCL2*, and *OTX2*.



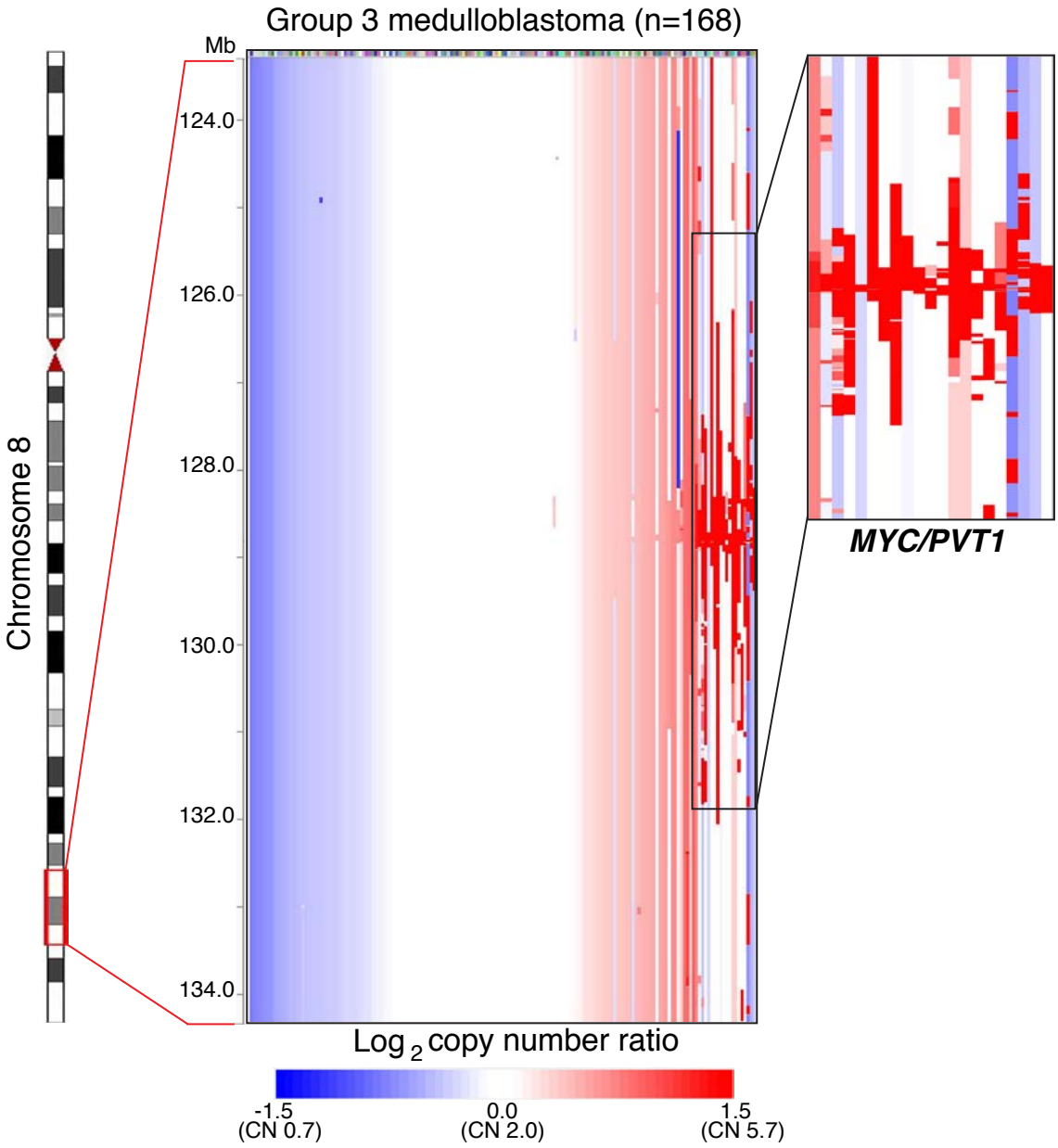
Supplementary Figure 35. Global association of somatic copy number aberrations and gene expression. a, Frequency density plot of gene expression within homozygous deletion regions, high-level amplification regions, or balanced regions. High-level amplified genes are significantly upregulated ($p < 2.2E-16$, Wilcoxon rank-sum test) and homozygously deleted genes are significantly downregulated ($p < 2.2E-16$). **b,** Frequency density plot of gene expression within chromosome arm level gains or losses, or balanced regions. Genes within broad gains are significantly upregulated ($p = 0.00012$) and genes within broad losses are significantly downregulated ($p = 1.6E-08$).

Supplementary Figure 36, Northcott, Shih et al



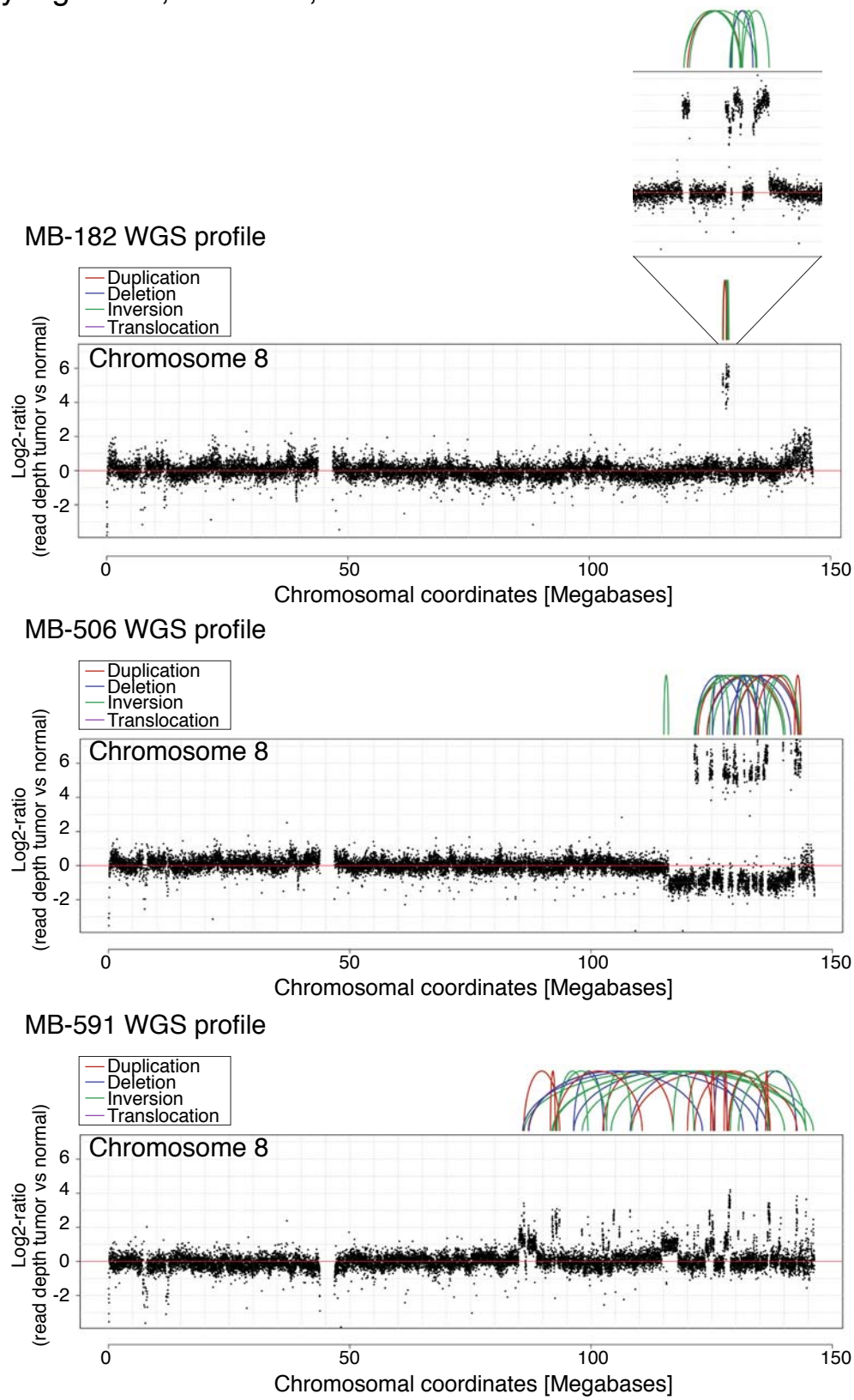
Supplementary Figure 36. Expression of candidate target genes is driven by copy-number changes. a-c, Bar plots depict expression of candidate genes from significant GISTIC2-amplified regions for (**a**) SHH, (**b**) Group 3, and (**c**) Group 4 medulloblastomas. Expression values of genes across samples were converted to ranks (increasing order) for the purpose of comparison by non-parametric tests between samples with balanced (gray) or aberrant (colours) copy number at each indicated locus. Error bars represent standard errors and reflect sample-to-sample variability in gene expression. Amplified genes exhibiting significantly (and consistently) upregulated expression are depicted in red (Mann-Whitney test with false discovery rate correction). False discovery rates were tuned by adaptive thresholds for each locus so that no false positives are expected by chance.

Supplementary Figure 37, Northcott, Shih et al

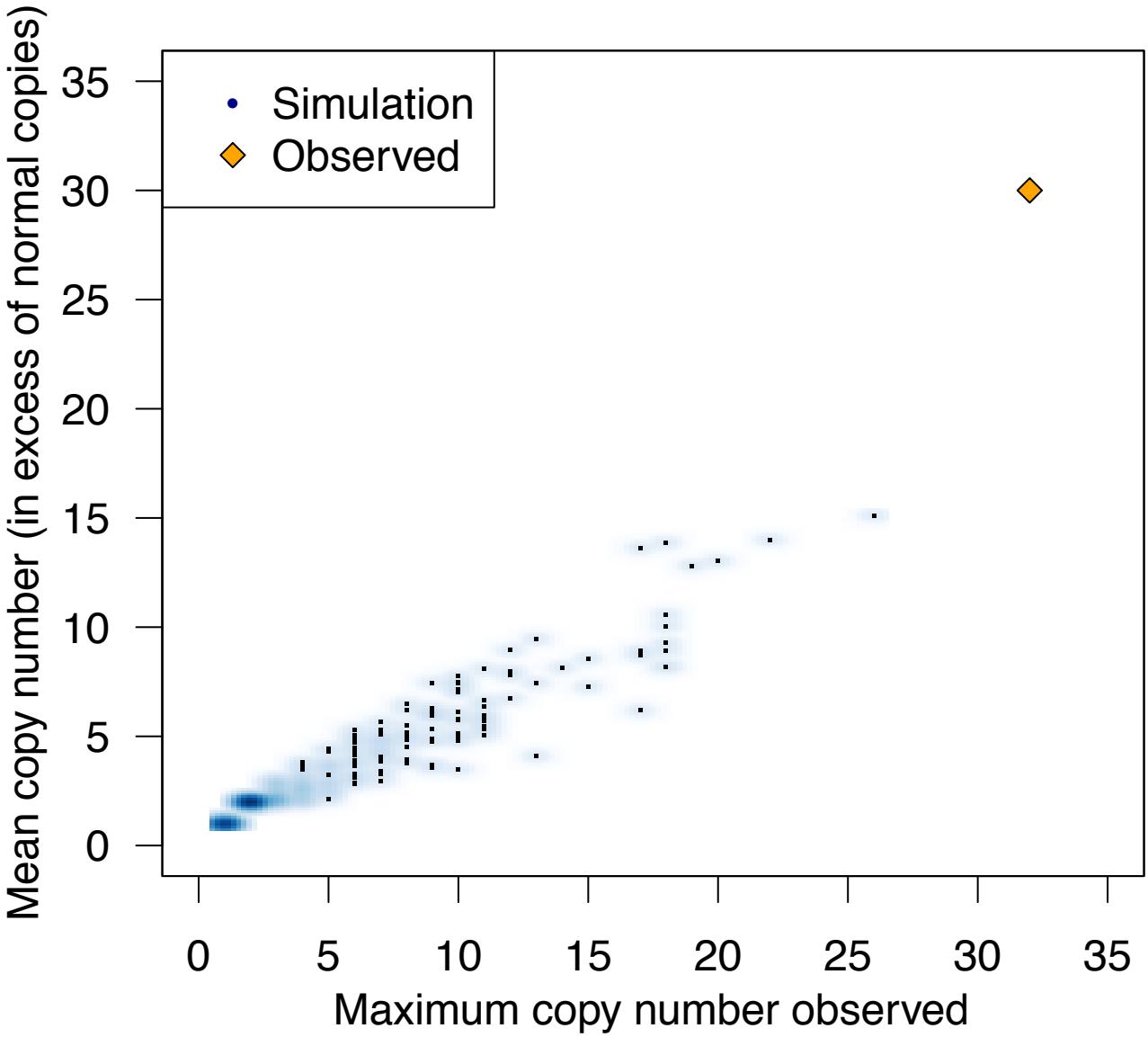


Supplementary Figure 37. Co-amplification of *MYC* and *PVT1* in Group 3 medulloblastoma. Copy number heatmap of 168 Group 3 cases demonstrating frequent high-level amplification of the *MYC* locus on chromosome 8q24.21. *MYC* is focally amplified in 28/168 (16.7%) Group 3 tumours. Notably, 20/28 (71.4%) *MYC*-amplified cases co-amplify the adjacent *PVT1* gene, either in part or in full.

Supplementary Figure 38, Northcott, Shih et al

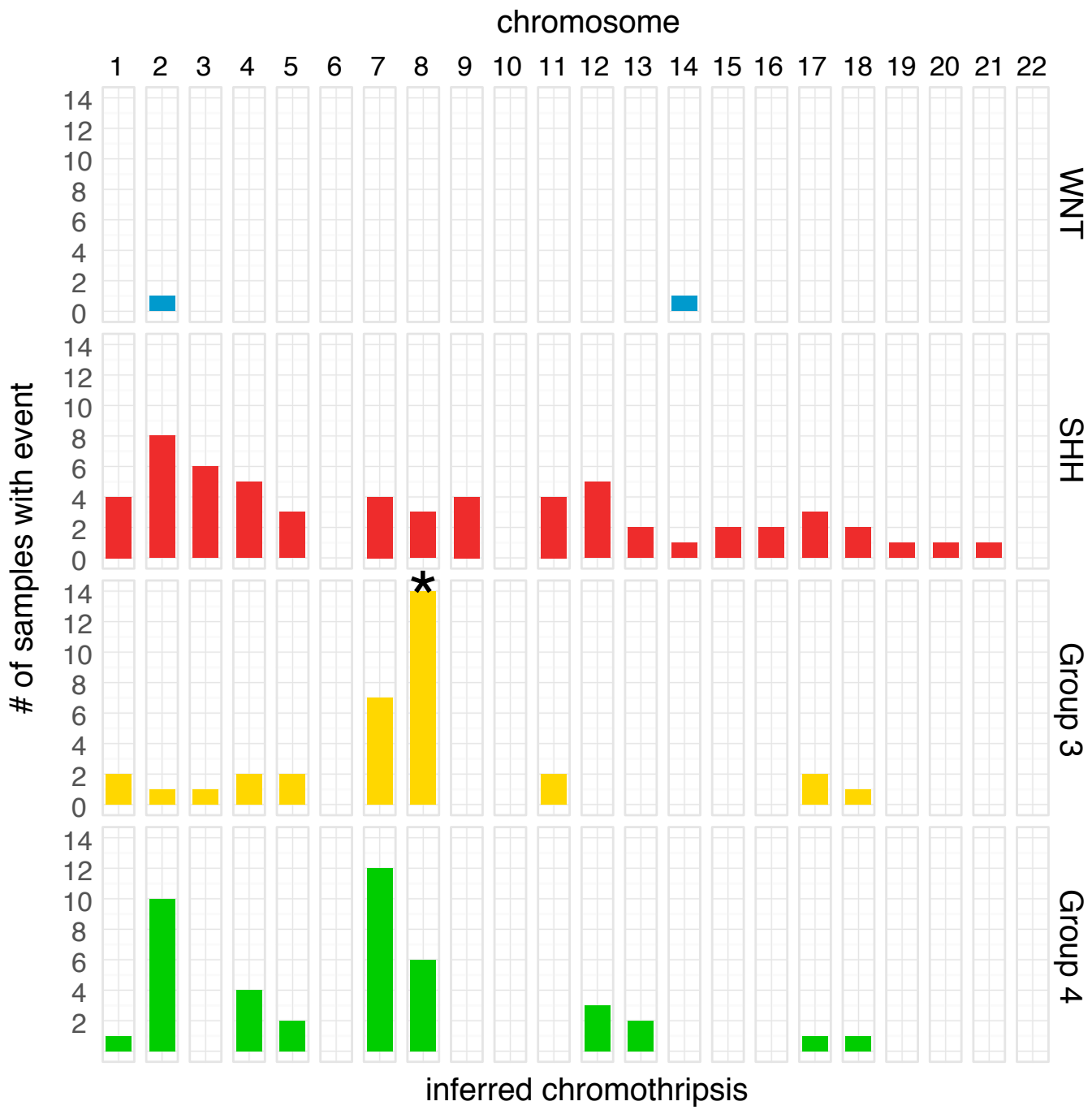


Supplementary Figure 38. Recurrent chromothripsis on chromosome 8 in Group 3 medulloblastoma. WGS confirms complex patterns of rearrangement on chromosome 8q24 in PVT1 fusion(+) Group 3 samples. Chromosome 8 copy number plots as estimated by read depth of tumour *vs.* a non-neoplastic control show extensive rearrangements consistent with chromothripsis in the highly amplified 8q24 region.



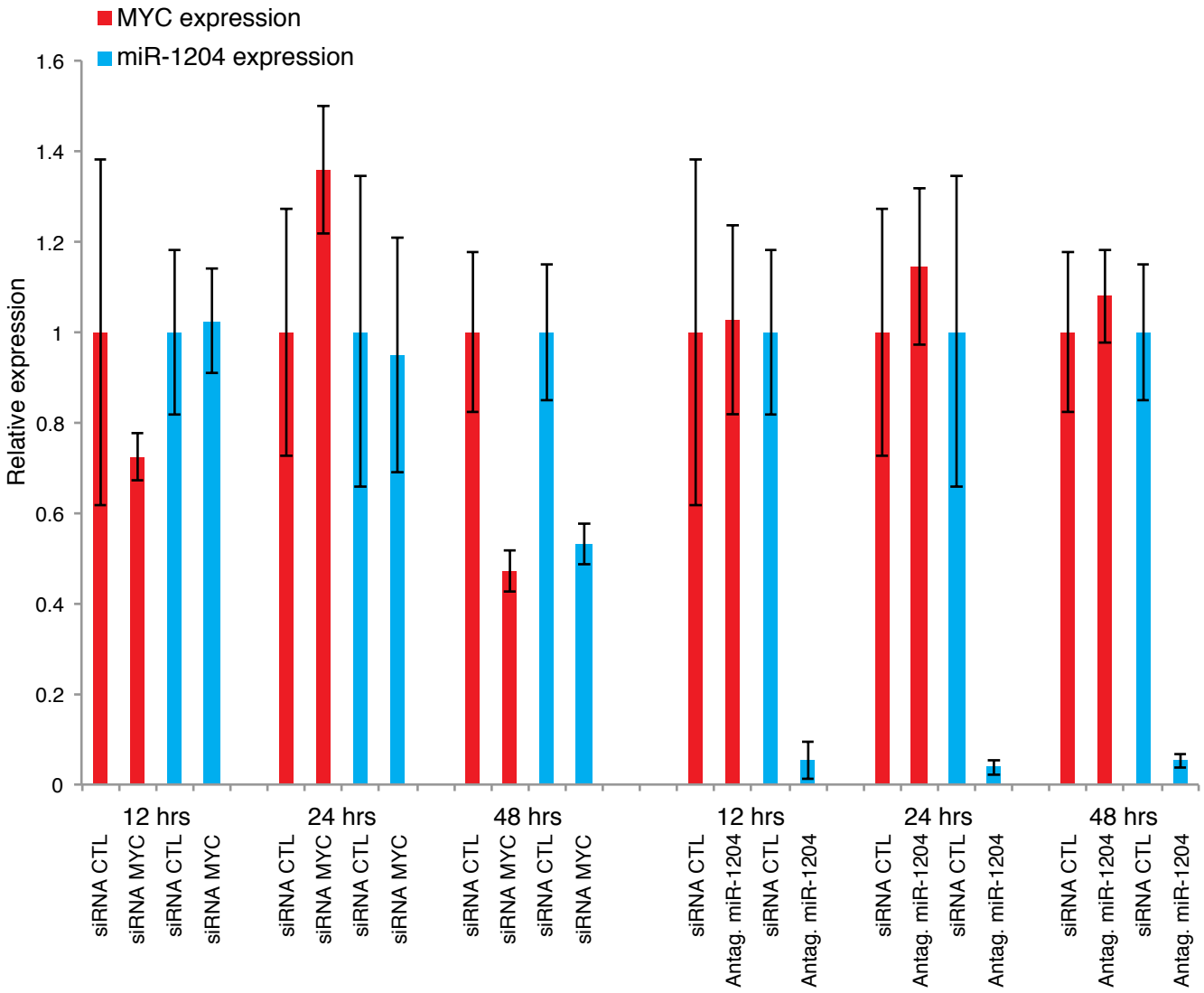
Supplementary Figure 39. Simulation of progressive genomic rearrangements of chromosome 8. Progressive rearrangements of chromosome 8 were simulated using the observed breakpoints. The mean copy number of amplified segments (y-axis) as well as the maximum copy number estimated in the simulations (x-axis) are distinct from the observed data.

Supplementary Figure 40, Northcott, Shih et al



Supplementary Figure 40. Chromosomal distribution of chromothripsis across medulloblastoma subgroups. Quantification of chromothripsis by chromosome across medulloblastoma subgroups confirms a significant enrichment of chromothripsis on chromosome 8 in Group 3 (q=0.0004). Samples showing at least 10 copy number state changes on a single chromosome were considered to have undergone chromothripsis.

Supplementary Figure 41, Northcott, Shih et al



Supplementary Figure 41. Verification of MYC and miR-1204 knockdown in MED8A cells. Quantitative RT-PCR of MED8A cells transfected with either siRNA control, siRNA-MYC, or antagomiR against miR-1204 verifies inhibition of the genes of interest. MYC knockdown results in inhibition of both MYC and miR-1204. Error bars represent SEM of duplicate experiments.

Supplementary Table 1. Summary of MAGIC SNP6 array profiling

I) SNP6 hybridizations	**1239**
i) Presumed primary medulloblastomas	**1228**
ii) Presumed medulloblastoma cell lines	**11**
II) Quality control	
Average SNP QC-call-rate (%)	97.54
Median SNP QC-call-rate (%)	97.92
SNP QC-call-rate range (%)	85.67–99.40
Average contrast-QC	2.664
Median contrast-QC	2.698
Contrast-qc range	-0.096–3.854
Average MAPD score	0.252
Median MAPD score	0.248
MAPD score range	0.189–0.406
Samples that failed QC (SNP qc-call-rate, contrast-qc, MAPD score)	**12**
III) Genotype concordance	
Samples with duplicate genotypes	**95**
i) Samples hybridized 2x	81
ii) Samples hybridized 3x	7
iii) Relapses/metastases profiled with matched primary	22
iv) Xenografts/cell lines profiled with matched primary	5
Unique samples available for analysis post-QC	**1132**
i) Presumed primary medulloblastomas	1123
ii) Presumed medulloblastoma cell lines	9
IV) Integration of clinical details	
Samples eliminated via clinical details/genetics	**36**
i) PNET	8
ii) Ependymoma	6
iii) Medullomyoblastoma	6
iv) ATRT	5
v) Melanotic medulloblastoma	2
vi) Medulloepithelioma	1
vii) Glioblastoma multiforme	1
viii) ETANTR (chr19q amplicon)	3
ix) Relapsed/metastatic medulloblastomas (no primary available)	3
x) Cell line	1
V) Final set of unique MB samples for MAGIC	**1097**
i) Final set of unique primary medulloblastomas for MAGIC	**1087**
ii) Final set of unique medulloblastoma cell lines for MAGIC	**10**

Supplementary Table 2. Clinical features of the MAGIC cohort

Variable	WNT	SHH	Group 3	Group 4	All MAGIC
Number of patients	76	266	168	317	1087
Frequency in cohort (%)	9.19	32.16	20.31	38.33	N/A
Gender (count)					
Male	34	146	105	224	665
Female	41	106	56	82	374
Male : female ratio	0.83 : 1	1.38 : 1	1.88 : 1	2.73 : 1	1.78 : 1
Gender (%)					
Male	45.33	57.94	65.22	73.20	64.00
Female	54.67	42.06	34.78	26.80	36.00
Age (years)					
Average	11.7	13.2	6.4	9.2	9.7
Median	10.0	7.6	5.6	8.4	8.0
Range	2.0–56.3	0.1–52.0	0.6–29	1.0–36.0	0.1–56.3
Age category (%)					
Infants (≤3)	2.67	36.08	24.53	6.84	20.35
Children (4–15)	78.67	33.33	72.33	85.99	66.06
Adults (≥16)	18.67	30.59	3.14	7.17	13.60
Histological subtype (%)					
Classic	86.96	51.40	68.38	84.31	71.68
Desmoplastic/MBEN	5.80	35.51	11.03	7.66	16.92
LCA	7.25	13.08	20.59	8.03	11.39
Metastatic stage (%)					
M0	88.46	80.90	57.27	59.49	68.27
M+	11.54	19.10	42.73	40.51	31.73
5-year overall survival (%)					
All ages	97.22	70.58	59.99	69.42	69.67
Infants (≤3)	100.00	71.71	38.70	38.03	60.35
Children (4-15)	100.00	65.07	65.65	71.10	71.28
Adults (≥16)	90.00	75.73	60.00	74.29	75.22

Supplementary Table 6. Driver genes in medulloblastoma subgroups

Cytoband	Peak limits	# genes in peak	Candidate driver(s)	*q* value	Subgroup incidence	High-level amplification?	Subgroup enriched?
SHH							
2p24.3	15.965-16.011	2	*MYCN*	4.51E-28	8.2%	Yes	No
2q14.2	121.169-121.502	1	*GLI2*	3.05E-13	5.2%	Yes	Yes
12p13.32	4.088-4.301	2	*CCND2*	1.27E-06	3.4%	Yes	No
6q24.3	146.260-146.263	1	*SHPRH*	3.00E-02	3.4%	No	No
17q23.2	55.878-56.410	3	*PPM1D*	9.56E-06	2.2%	Yes	Yes
1p34.2	39.994-40.275	6	*MYCL1*	1.25E-04	2.2%	Yes	Yes
15q26.3	96.895-97.902	6	*IGF1R*	3.07E-02	2.2%	Yes	Yes
1q32.1	202.018-202.796	14	*MDM4, PIK3C2B*	7.37E-03	1.9%	Yes	Yes
13q33.3	107.696-109.344	3	*IRS2*	2.08E-02	1.9%	Yes	Yes
3p22.2	37.234-38.566	18	*ACVR2B*	8.37E-03	1.9%	Yes	No
11q22.1	101.471-101.502	1	*YAP1*	3.23E-04	1.5%	Yes	Yes
12p12.1	21.334-21.515	4	*IAPP*	9.84E-02	1.5%	Yes	Yes
7q21.2	91.870-92.443	7	*CDK6*	7.37E-03	1.5%	Yes	No
17p11.2	17.199-21.968	69	*Unknown*	3.95E-02	1.1%	Yes	Yes
6p21.2	37.296-37.405	2	*TBC1D22B, TMEM217*	8.26E-02	1.1%	Yes	Yes
1p22.3	86.700-87.635	9	*LMO4*	9.39E-02	1.1%	Yes	Yes
Group 3							
8q24.21	128.809-128.832	1	*MYC*	3.54E-33	16.7%	Yes	Yes
8q24.21	129.005-129.315	5	*PVT1*	2.78E-04	11.9%	Yes	Yes
14q23.1	56.285-56.374	2	*OTX2*	7.31E-03	7.7%	No	Yes
6q24.3	146.256-146.266	1	*SHPRH*	8.66E-02	6.0%	No	Yes
1q41	214.921-214.948	1	*ESRRG*	9.51E-02	4.2%	No	Yes
14q32.31	100.451-100.479	7	*SNORD113-1/7*	8.07E-02	3.6%	No	Yes
7q21.2	91.330-92.398	13	*CDK6*	2.40E-03	3.0%	Yes	No
2p24.3	15.909-16.095	2	*MYCN*	2.30E-05	2.4%	Yes	No
2q22.3	148.172-148.605	3	*ACVR2A*	8.68E-05	2.4%	Yes	Yes
9q22.33	100.898-101.028	3	*TGFBR1*	7.28E-02	2.4%	Yes	Yes
3p22.2	38.357-38.754	6	*ACVR2B*	2.44E-02	1.2%	Yes	No
8q21.2	86.242-86.343	3	*E2F5*	9.51E-02	1.2%	Yes	Yes
Group 4							
5q23.2	121.785-121.807	1	*SNCAIP*	1.10E-14	10.4%	No	Yes
2p24.3	15.980-16.013	2	*MYCN*	4.14E-32	6.3%	Yes	No
7q21.2	92.016-92.328	3	*CDK6*	1.41E-22	4.7%	Yes	Yes
3p22.1	41.249-41.254	1	*CTNNB1*	2.66E-02	2.8%	No	Yes
6q24.3	146.260-146.266	1	*SHPRH*	1.89E-04	2.8%	No	No
12p13.32	4.205-5.249	13	*CCND2*	2.54E-05	2.8%	Yes	No
14q23.1	56.339-56.793	3	*OTX2*	7.00E-02	2.2%	No	No
3p22.2	38.346-38.530	4	*ACVR2B*	1.45E-03	1.9%	Yes	No
7q21.11	84.586-84.602	1	*SEMA3D*	7.68E-04	1.9%	Yes	Yes
1q41	214.939-214.948	1	*ESRRG*	6.66E-02	1.3%	No	No
8q24.21	128.792-128.856	1	*MYC*	2.49E-03	1.3%	Yes	No
19p13.3	0.636-0.816	11	*Unknown*	2.85E-04	0.9%	Yes	Yes
8p11.21	41.015-43.943	26	*KAT6A*	3.05E-02	0.6%	Yes	Yes

Supplementary Table 8. WNT GISTIC Regions: Amplifications

Cytoband	Peak Limits	Size (Kbp)	GISTIC *q*-value	CMDS	# genes	Key candidate(s)
14q23.1	56.244-56.372	128.20	0.043587	No	2	*OTX2*
18q23	76.064-76.117	52.76	0.094307	No	1	*PARD6G*
21q22.11	34.398-34.401	3.49	0.0081927	No	2	*MRPS6, SLC5A3*

Supplementary Table 9: *CTNNB1* mutation and chr6 loss in WNT medulloblastoma

Sample ID	Exon 3 Mutation		Chromosome 6 Status	
	Codon	Amino Acid	*p* arm	*q* arm
MB-15	GGA → GAA	G34E	loss	loss
MB-18	GAC → AAC	D32N	loss	loss
MB-128	GGA → GAA	G34E	loss	loss
MB-145	GAC → TAC	D32Y	loss	loss
MB-147	TCT → TGT	S33C	loss	loss
MB-178	GGA → AGA	G34R	loss	loss
MB-190	GAC → TAC	D32Y	balanced	balanced
MB-191	GGA → GAA	G34E	loss	loss
MB-301	GAC → AAC	D32N	loss	loss
MB-302	GAC → AAC	D32N	loss	loss
MB-303	TCT → CCT	S37P	loss	loss
MB-307	GAC → AAC	D32N	loss	loss
MB-328	TCT → TTT	S37F	loss	loss
MB-336	GGA → GAA	G34E	loss	loss
MB-361	GAC → GTC	D32V	loss	loss
MB-401	GAC → AAC	D32N	loss	loss
MB-407	TCT → TTT	S45F	balanced	balanced
MB-408	TCT → TAT	S37Y	balanced	balanced
MB-416	GAC → TAC	D32Y	loss	loss
MB-418	TCT → TGT	S33C	loss	loss
MB-449	TCT → TGT	S37C	balanced	balanced
MB-455	GAC → TAC	D32Y	loss	loss
MB-496	GGA → AGA	G34R	loss	loss
MB-498	TCT → TGT	S33C	loss	loss
MB-502	GGA → AGA	G34R	loss	loss
MB-505	TCT → TGT	S33C	loss	loss
MB-520	TCT → TTT	S33F	balanced	balanced
MB-521	TCT → TGT	S33C	balanced	loss
MB-662	GGA → GAA	G34E	loss	loss
MB-670	GGA → GAA	G34E	loss	loss
MB-695	TCT → TAT	S37Y	loss	loss
MB-701	GAC → GGC	D32G	loss	loss
MB-719	TCT → TAT	S37Y	balanced	loss
MB-788	TCT → TAT	S33Y	loss	loss
MB-812	GGA → AGA	G34R	loss	loss
MB-828	GGA → AGA	G34R	loss	loss
MB-873	TCT → TTT	S37F	loss	loss
MB-883	GAC → AAC	D32N	loss	loss
MB-891	TCT → TTT	S33F	loss	loss
MB-905	GAC → TAC	D32Y	loss	loss

Supplementary Table 9: *CTNNB1* mutation and chr6 loss in WNT medulloblastoma

Sample ID	Exon 3 Mutation		Chromosome 6 Status	
	Codon	Amino Acid	*p* arm	*q* arm
MB-956	GAC → TAC	D32Y	loss	loss
MB-972	TCT → TTT	S33F	loss	loss
MB-1055	TCT → TTT	S33F	balanced	balanced
MB-1082	GAC → TAC	D32Y	loss	loss
MB-1089	TCT → TTT	S33F	balanced	balanced
MB-1109	GAC → GTC	D32V	loss	loss
MB-1130	TCT → CCT	S33P	loss	loss
MB-1169	TCT → TGT	S37C	loss	loss
MB-1171	TCT → TTT	S37F	loss	loss
MB-1198	TCT → TGT	S33C	loss	loss
MB-1235	TCT → TTT	S37F	balanced	balanced
MB-1237	TCT → GCT	S37A	loss	loss
MB-1255	TCT → TGT	S33C	loss	loss
MB-1259	GAC → TAC	D32Y	balanced	loss
MB-1275	GAC → AAC	D32N	balanced	balanced
MB-1299	TCT → TGT	S33C	loss	loss
MB-1321	GAC → AAC	D32N	balanced	balanced
MB-1331	GAC → TAC	D32Y	loss	loss
MB-1335	TCT → CCT	S33P	loss	loss
MB-1343	TCT → TAT	S33Y	loss	loss
MB-1355	TCT → TGT	S33C	loss	loss
MB-1368	TCT → GCT	S33A	loss	loss
MB-1380	GGA → AGA	G34R	loss	loss
MB-389	no mutation	no mutation	balanced	balanced
MB-566	no mutation	no mutation	balanced	balanced
MB-687	no mutation	no mutation	loss	loss
MB-811	no mutation	no mutation	loss	loss
MB-865	no mutation	no mutation	loss	loss
MB-1026	no mutation	no mutation	balanced	balanced
MB-1029	no mutation	no mutation	loss	loss
MB-1220	no mutation	no mutation	balanced	Balanced

Supplementary Table 10. SHH GISTIC Regions: Amplifications

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
1p34.2	39.994-40.275	280.64	0.00012472	Yes	6	*MYCL1*
1p22.3	86.700-87.635	935.38	0.093888	Yes	9	*LMO4*
1q32.1	202.018-202.796	778.48	0.0073744	Yes	14	*MDM4, SOX13*
2p24.3	15.965-16.011	45.93	4.51E-28	Yes	2	*MYCN*
2q14.2	121.169-121.502	333.24	3.05E-13	Yes	1	*GLI2*
3p22.2	37.234-38.566	1,331.71	0.0083719	Yes	18	*MYD88, ACVR2B*
3q25.1	152.175-152.643	468.30	0.11644	Yes	8	*MED12L, P2RY14*
4q12	55.270-55.277	7.06	1.05E-06	No	1	*KIT*
5q32	146.422-146.446	23.97	0.12294	No	1	*PPP2R2B*
6p21.2	37.296-37.405	108.63	0.082636	No	2	*TBC1D22B, TMEM217*
6q24.3	146.260-146.263	2.95	0.030014	No	1	*SHPRH*
7q21.2	91.870-92.443	573.20	0.0073744	Yes	7	*CDK6*
8q22.1	97.375-97.460	85.53	0.19704	No	1	*PTDSS1*
8q22.1	98.146-98.188	42.29	0.19704	No	1	*PGCP*
11q22.1	101.471-101.502	30.65	0.00032282	Yes	1	*YAP1*
12p13.33	0.000-4.096	4,095.66	0.14874	Yes	35	*FOXM1, KDM5A*
12p13.32	4.088-4.301	213.23	1.27E-06	Yes	2	*CCND2*
12p12.1	21.334-21.515	181.07	0.098396	Yes	4	*IAPP, RECQL*
12q15	67.385-67.414	29.61	0.24019	No	1	*NUP107*
13q33.3	107.696-109.344	1,648.49	0.020779	No	3	*IRS2*
15q26.3	96.895-97.902	1,006.96	0.030691	Yes	6	*IGF1R*
16p13.2	0.000-7.446	7,445.89	0.11705	No	223	*Unknown*
17p11.2	17.199-21.968	4,768.48	0.039471	Yes	69	*Unknown*
17q23.2	55.878-56.410	531.46	9.56E-06	Yes	3	*PPM1D*
21q22.11	34.398-34.400	2.108	0.021451	No	2	*MRPS6, SLC5A3*

Supplementary Table 11. SHH GISTIC Regions: Deletions

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
1q21.3	150.754-150.948	194.15	0.17418	No	11	*CRCT1, LCE2B*
2p23.3	21.119-26.114	4994.32	0.11198	No	28	*DNMT3A, NCOA1*
4p16.3	0.000-9.394	9394.17	0.078486	Yes	126	*Unknown*
4q31.1	140.186-140.408	221.84	0.0025458	No	3	*CCRN4L, ELF2*
5q35.3	175.326-180.858	5531.52	0.0036882	Yes	105	*Unknown*
7p22.3	0.000-7.245	7245.34	0.093194	No	82	*Unknown*
7q11.23	71.515-76.590	5075.48	0.054351	Yes	83	*Unknown*
8p12	38.181-38.247	66.56	0.0096674	No	3	*BAG4, DDHD2*
9p23	7.790-12.683	4893.80	0.14285	No	3	*PTPRD, TYRP1*
9p21.3	21.229-22.439	1210.13	0.052058	Yes	18	*CDKN2A, CDKN2B*
9q22.32	97.118-97.680	561.45	1.31E-09	Yes	5	*PTCH1, FANCC*
9q31.3	107.576-110.819	3243.09	0.099337	No	10	*KLF4, IKBKAP*
10p14	8.157-11.424	3267.52	0.045888	No	4	*GATA3, CELF2*
10q23.31	89.568-90.024	456.30	1.24E-09	Yes	5	*PTEN*
15q15.1	39.361-40.280	919.44	0.0026476	Yes	24	*MAPKBP1, RTF1*
17p13.1	7.525-7.568	43.41	0.018874	No	4	*TP53*
17p11.2	19.180-21.224	2043.91	0.063384	Yes	26	*SPECC1, MAPK7*
20p13	0.000-5.941	5940.89	0.054351	Yes	109	*Unknown*

Supplementary Table 12. Group 3 GISTIC Regions: Amplifications

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
1q21.2	148.687-148.862	175.23	0.095091	Yes	7	*ECM1, MCL1*
1q41	214.921-214.948	27.33	0.095091	No	1	*ESRRG*
2p24.3	15.909-16.095	186.32	2.30E-05	No	2	*MYCN*
2p24.1	20.031-20.047	16.05	0.19014	No	1	*WDR35*
2q22.3	148.172-148.605	433.81	8.68E-05	No	3	*ACVR2A*
3p22.2	38.357-38.754	396.97	0.024444	Yes	6	*ACVR2B*
6q24.3	146.256-146.266	10.30	0.086557	No	1	*SHPRH*
7q21.2	91.330-92.398	1067.90	0.0024037	Yes	13	*CDK6*
8q21.11	77.858-77.919	61.68	0.15945	No	1	*ZFHX4*
8q21.2	86.242-86.343	100.45	0.095091	No	3	*E2F5, LRRCC1*
8q24.12	121.690-121.719	29.01	0.016081	No	1	*SNTB1*
8q24.13	124.618-124.723	105.14	0.0052581	Yes	1	*FBXO32*
8q24.21	128.809-128.832	23.39	3.54E-33	Yes	1	*MYC*
8q24.21	129.005-129.315	310.33	0.00027813	Yes	5	*PVT1*
8q24.21	130.890-130.933	42.59	0.0030156	Yes	1	*FAM49B*
8q24.22	132.865-133.461	595.64	0.18746	Yes	4	*EFR3A, KCNQ3*
8q24.22	134.154-134.181	27.10	0.0035215	Yes	2	*TG, SLA*
8q24.23	139.291-139.331	40.45	0.095091	No	1	*FAM135B*
9q22.33	100.898-101.028	129.42	0.07282	No	3	*TGFBR1*
12p13.33	0.000-0.236	235.55	0.036881	No	5	*IQSEC3, SLC6A12*
14q23.1	56.285-56.374	88.82	0.0073102	No	2	*OTX2*
14q32.31	100.451-100.479	27.70	0.080686	No	7	*SNORD113-1*

Supplementary Table 13. Group 3 GISTIC Regions: Deletions

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
3q23	142.948-143.080	132.41	0.008657	No	3	*ATP1B3, GRK7*
4p16.1	0.000-9.394	9394.17	0.19331	No	126	*Unknown*
4q31.1	140.186-140.408	221.84	0.11826	No	3	*CCRN4L, ELF2*
4q31.3	151.713-152.240	527.37	0.060434	No	3	*MAB21L2, RPS3A*
5q13.2	67.632-70.703	3070.93	0.20612	No	27	*PIK3R1, CDK7*
5q35.3	178.440-180.858	2417.66	4.43E-10	Yes	43	*MAPK9, RASGEF1C*
6q13	73.169-75.851	2682.01	0.13269	No	16	*SLC17A5, CD109*
8p12	38.181-38.247	66.56	0.1406	No	3	*PPAPDC1B, BAG4*
9q34.13	134.501-134.540	38.60	2.01E-08	No	2	*DDX31, GTF3C4*
10p14	8.157-11.424	3267.52	0.042514	No	4	*GATA3, CELF2*
11p15.4	4.748-4.885	137.510	0.0052643	Yes	6	*OR51A7*
11q22.3	102.847-104.320	1472.79	0.077979	No	7	*DYNC2H1, CASP4*
13q12.2	27.457-27.647	190.28	0.052909	No	4	*FLT3, PAN3*
14q13.1	33.484-34.528	1043.23	0.077979	Yes	8	*BAZ1A, SNX6*
16q23.3	80.689-81.221	531.62	0.0913	Yes	3	*MPHOSPH6, CDH13*
16q23.3	82.551-88.827	6275.81	0.12178	Yes	96	*Unknown*
16q24.1	83.702-84.269	566.81	0.024343	Yes	6	*GINS2, FAM92B*
17p13.3	0.000-2.159	2159.30	0.20612	No	42	*HIC1*
17p12	10.681-12.513	1831.22	0.030517	Yes	8	*MAP2K4, DNAH9*
20p13	0.000-1.824	1824.39	0.017067	Yes	37	*SIRPA*

Supplementary Table 14. Group 4 GISTIC Regions: Amplifications

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
1q41	214.939-214.948	9.47	0.066553	No	1	*ESRRG*
2p25.3	2.171-2.197	26.66	0.19858	Yes	1	*MYT1L*
2p24.3	15.254-15.275	20.73	0.00013144	Yes	1	*NBAS*
2p24.3	15.668-15.700	31.59	0.00014206	Yes	1	*DDX1*
2p24.3	15.980-16.013	33.43	4.14E-32	Yes	2	*MYCN*
2p24.3	16.599-16.620	21.24	0.10051	Yes	1	*FAM49A*
2p24.2	18.591-18.678	87.57	0.0058887	Yes	3	*NT5C1B*
2p24.1	20.041-20.061	19.94	0.022966	No	2	*MATN3, WDR35*
2p24.1	20.252-20.279	26.21	0.01601	No	1	*SDC1*
3p22.2	38.346-38.530	184.40	0.0014461	Yes	4	*ACVR2B*
3p22.1	41.249-41.254	5.20	0.026638	No	1	*CTNNB1*
5q14.3	87.867-87.890	22.68	0.23651	No	1	*LINC00461*
5q23.2	121.785-121.807	22.51	1.10E-14	Yes	1	*SNCAIP*
6q24.3	146.260-146.266	6.50	0.00018853	No	1	*SHPRH*
7p21.3	12.213-12.230	17.15	0.1597	No	1	*TMEM106B*
7p12.2	50.654-50.673	18.97	0.0066882	No	1	*GRB10*
7p11.1	56.564-57.594	1030.26	0.0075151	No	7	*ZNF479, ZNF716*
7q21.11	84.586-84.602	16.74	0.00076778	Yes	1	*SEMA3D*
7q21.2	92.016-92.328	312.34	1.41E-22	Yes	3	*CDK6*
7q31.1	108.552-119.755	11202.10	0.060782	No	33	*MET*
7q33	134.739-134.756	16.76	0.089932	No	1	*CNOT4*
8p11.21	41.015-43.943	2928.29	0.030523	No	26	*HOOK3, ANK1*
8q11.22	51.847-52.074	226.77	0.02536	No	1	*SNTG1*
8q23.1	106.446-106.601	154.77	0.15106	No	1	*ZFPM2*
8q24.21	128.792-128.856	63.90	0.0024916	No	1	*MYC*
9p24.2	3.987-3.995	8.34	0.0059279	No	1	*GLIS3*
10p15.3	0.000-0.299	299.30	0.034517	No	2	*ZMYND11*
12p13.32	4.205-5.249	1044.92	2.54E-05	Yes	13	*CCND2, KCNA1*
12p12.3	16.732-19.784	3051.96	0.1247	No	8	*PIK3C2G, AEBP2*
14q23.1	56.339-56.793	454.48	0.069953	No	3	*OTX2*
17q12	28.054-29.923	1868.88	0.053268	No	12	*CCL2, MYO1D*
17q21.31	38.505-38.512	6.95	0.15071	No	1	*BRCA1*
18q23	76.041-76.117	76.25	0.05313	No	1	*PARD6G*
19p13.3	0.636-0.816	180.20	0.00028473	No	11	*LPPR3, PALM*
21q22.11	34.398-34.400	2.17	0.00060755	No	2	*MRPS6, SLC5A3*

Supplementary Table 15. Group 4 GISTIC Regions: Deletions

Cytoband	Peak Limits	Size (Kbp)	GISTIC q-value	CMDS	# genes	Key candidate(s)
2q37.3	234.590-242.951	8360.89	0.00029198	No	81	*Unknown*
3p21.31	49.199-49.376	177.08	0.01734	Yes	7	*RHOA, USP4*
5q13.2	67.632-70.703	3070.93	0.095745	No	27	*PIK3R1, CCNB1*
5q35.3	175.243-180.858	5614.77	8.26E-14	Yes	106	*Unknown*
8p12	38.181-38.247	66.56	0.020097	No	3	*BAG4, DDHD2*
9p13.2	36.387-36.832	444.59	0.15499	No	4	*PAX5, MELK*
10p12.33	17.858-18.281	423.60	0.057012	No	5	*MRC1, SLC39A12*
10q11.21	42.647-42.957	310.83	0.0018621	No	3	*RET*
12p11.21	31.633-31.720	86.67	0.23467	No	3	*DENND5B*
13q14.2	47.732-47.895	162.78	0.00024593	No	3	*RB1*
14q13.2	33.484-35.075	1590.61	0.039558	Yes	14	*BAZ1A, NFKBIA*
17p13.3	1.624-1.684	60.78	0.014875	No	3	*SMYD4, RPA1*
22q12.2	28.563-28.807	243.95	0.25114	No	3	*MTMR3*
22q13.31	41.580-49.691	8111.72	8.45E-05	Yes	103	*Unknown*

Supplementary Table 17. Summary of RNASeq data

Sample	Subgroup	Total Sequenced Reads (millions)	Gb (Total)	Chastity Passed (%)	Aligned (%)*	Uniquely Aligned (%)*
MB-182	Group 3	171.37	17.14	94.87	82.72	77.60
MB-412	Group 3	143.15	14.31	94.78	79.77	73.90
MB-549	Group 3	161.70	16.17	94.90	82.59	77.21
MB-591	Group 3	143.07	14.31	95.45	83.19	79.59
MB-759	Group 3	168.60	16.86	94.20	78.87	75.29
MB-522	Group 3	117.44	11.74	86.36	71.28	63.46
MB-586	Group 3	209.59	20.96	83.66	81.22	75.56
MB-329	Group 3	200.46	20.05	94.00	82.14	78.65
MB-333	Group 3	158.10	15.81	94.76	81.67	78.29
MB-511	Group 3	214.46	21.45	93.23	80.50	77.21
MB-532	Group 3	216.24	21.62	93.08	78.62	73.81
MB-594	Group 3	146.55	14.65	95.31	81.91	76.78
MB-607	Group 3	186.69	18.67	94.27	82.27	79.00
MB-439	Group 4	137.00	13.70	95.44	84.82	81.45
MB-447	Group 4	184.16	18.42	94.03	83.65	79.92
MB-563	Group 4	219.11	21.91	91.70	77.49	73.58
MB-569	Group 4	188.17	18.82	93.20	78.83	74.19
MB-600	Group 4	117.02	11.70	96.44	86.09	82.52

* percent of Chastity-passed reads

Supplementary Table 18. PVT1 fusions identified by RNASeq

Sample	Subgroup	Read Pairs	Spanning Reads	Type	Transcript Breakpoint	Genes
MB-182	Group 3	17	525	duplication	chr8:128819674 | chr8:128972426	*MYC/PVT1*
MB-586	Group 3	106	121	inversion	chr8:128898314 | chr8:134332102	*PVT1/NDRG1*
MB-182	Group 3	7	158	inversion	chr8:128598808 | chr8:128972426	*NA/PVT1*
MB-0527	Group 4	4	67	duplication	chr12:3888813 | chr12:6955557	*NA/NA*
MB-0532	Group 3	6	74	deletion	chr3:117646494 | chr3:119198718	*LSAMP/NA*
MB-0437	Group 4	6	93	deletion	chr3:117646494 | chr3:119198718	*LSAMP/NA*

Supplementary Table 19: PVT1-MYC fusion transcripts identified by RT-PCR/sequencing

Sample ID	Chromosome	Start	Stop
D458	Chr8	128876115	128876162
	Chr8	128877322	128877499
	Chr8	128972017	128972426
	Chr8	128819674	128819720
MB-524	Chr8	128876115	128876162
	Chr8	128972017	128972426
	Chr8	128819677	128819720
MB-1003	Chr8	128876115	128876163
	Chr8	128672133	128672284
	Chr8	128819679	128819720
MB-1377	Chr8	128876115	128876162
	Chr8	128972017	128972426
	Chr8	129070590	129070719
	Chr8	129091018	129091146
	Chr8	129273260	129273360
	Chr8	129274916	129275110
	Chr8	128819674	128819720
MB-1240	Chr8	128876115	128876162
	Chr8	128480320	128480445
	Chr8	128496315	128496449
	Chr8	128538807	128538950
	Chr8	128672133	128672284
	Chr8	128819679	128819720
MB-182	Chr8	128972185	128972426
	Chr8	128819677	128820004
D425	Chr8	128876128	128876162
	Chr8	128797600	128797375
	Chr8	128770850	128770742
	Chr8	128819676	128819720
MB-1338	Chr8	128876115	128876162
	Chr8	128420708	128420862
	Chr8	128480320	128480445
	Chr8	128496315	128496449
	Chr8	128672133	128672284
	Chr8	128819679	128819720

Supplementary Table 19: PVT1-MYC fusion transcripts identified by RT-PCR/sequencing

Sample ID	Chromosome	Start	Stop
MED8A	Chr8	128876115	128876162
	Chr8	127959101	127959217
	Chr8	128087982	128088046
	Chr8	127437224	127436950
	Chr8	127435173	127435121
	Chr8	127390550	127390441
	Chr8	128819678	128819720
MB-586	Chr8	128876090	128876162
	Chr8	134332039	134331865
	Chr8	134330196	134330140
	Chr8	134329351	134329298
MB-591	Chr8	128876115	128876162
	Chr8	136770849	136771024
	Chr8	128819675	128819720
MB-548	Chr8	128876115	128876162
	Chr8	127959101	127959217
	Chr8	128087982	128088045
	Chr8	128819675	128819720
MB-506	Chr8	128876115	128876162
	Chr8	128972017	128972427
	Chr8	128770852	128770742
	Chr8	128819676	128819720

Supplementary Table 20. Summary of WGS Data

Sample	Tissue Source	Total Sequenced Reads (million)	Gb (Total)	Chastity Passed (%)	Aligned (%)*	Uniquely Aligned (%)*	Estimated Coverage
MB-586	Normal	1431.97	143.20	0.96	0.97	0.95	28.6
MB-586	Tumour	877.44	87.74	0.97	0.98	0.95	23.7
MB-182	Tumour	720.27	72.03	0.96	0.96	0.93	21.3
MB-506	Tumour	1225.80	122.58	0.90	0.97	0.94	29.76
MB-591	Tumour	1123.53	112.35	0.92	0.98	0.96	26.32

* percent of Chastity-passed reads

Supplementary Table 21. Summary of genomic rearrangements identified by WGS

Tumor ID	Read Pairs	Spanning Reads	Rearrangement Type	Breakpoint
MB-586	522	916	duplication	chr8:128358681 | chr8:131003895
	230	300	inversion	chr8:128898613 | chr8:134334795
	226	284	duplication	chr8:138151282 | chr8:144720390
	72	107	deletion	chr8:134190644 | chr8:144695698
	134	101	inversion	chr8:130010482 | chr8:144575540
	141	82	duplication	chr8:133975475 | chr8:138476809
	44	18	deletion	chr8:130933695 | chr8:138413808
	*	*	inversion	chr8:130933766 | chr8:138409988
	8	50	translocation	chr11:38769234 | chr8:52894035
	4	28	translocation	chr11:38769245 | chr8:52892696
MB-182	*	*	inversion	chr8:128376622|chr8:128490341
	*	*	inversion	chr8:128511892|chr8:128724152
	*	*	inversion	chr8:127665816|chr8:128730482
	*	*	inversion	chr8:128511892|chr8:128724152
	*	*	inversion	chr8:127749539|chr8:128496176
	*	*	deletion	chr8:128358025|chr8:128681850
	*	*	duplication	chr8:127721581|chr8:128509496
	*	*	inversion	chr8:128340130|chr8:128739091
	*	*	inversion	chr8:127667933|chr8:128495165
	*	*	duplication	chr8:128699377|chr8:128703999
	647	189	inversion	chr8:128529954|chr8:128920150
	156	119	inversion	chr8:128483915|chr8:128503517
MB-506	1261	4545	inversion	chr8:124550752|chr8:128732360
	800	1780	inversion	chr8:121393886|chr8:139855703
	978	1479	inversion	chr8:121788137|chr8:135083440
	1689	1296	deletion	chr8:121600073|chr8:131692073
	1402	1167	inversion	chr8:135633998|chr8:143227809
	895	1110	duplication	chr8:122179234|chr8:134837168
	614	1114	duplication	chr8:124037949|chr8:139855860
	716	889	duplication	chr8:142179088|chr8:143486891
	246	825	deletion	chr8:134095104|chr8:136240715
	755	717	inversion	chr8:129966255|chr8:143089147
	734	342	deletion	chr8:129595164|chr8:134567530
	685	247	inversion	chr8:125207698|chr8:140187756
	982	233	inversion	chr8:136440723|chr8:143016798
	400	244	deletion	chr8:125138504|chr8:127431174
	999	192	inversion	chr8:136440934|chr8:143016798
	1871	122	deletion	chr8:124080899|chr8:141423353
	389	166	duplication	chr8:134094959|chr8:142857977
	729	114	deletion	chr8:128239838|chr8:142857248
	1379	24	inversion	chr8:135633999|chr8:143227812
	224	34	inversion	chr8:130378666|chr8:134510306
	18	27	inversion	chr8:115084238|chr8:116047840
	31	24	inversion	chr8:121443194|chr8:136331286
	15	17	duplication	chr8:129800778|chr8:143021850
	29	16	deletion	chr8:129789133|chr8:133020748

Supplementary Table 21. Summary of genomic rearrangements identified by WGS

Tumor ID	Read Pairs	Spanning Reads	Rearrangement Type	Breakpoint
MB-591	64	1259	duplication	chr8:136150104|chr8:136676402
	73	626	deletion	chr8:85820837|chr8:134456611
	265	211	inversion	chr8:128840459|chr8:136683965
	93	199	duplication	chr8:86058578|chr8:93489364
	181	189	deletion	chr8:108117419|chr8:131531652
	187	177	inversion	chr8:92987896|chr8:99358152
	36	129	inversion	chr8:92164234|chr8:144528605
	141	123	duplication	chr8:124800334|chr8:125631338
	5	127	deletion	chr8:128840459|chr8:136683970
	167	118	inversion	chr8:104131902|chr8:140105117
	141	113	duplication	chr8:128495369|chr8:137052439
	300	106	duplication	chr8:121461272|chr8:136985098
	60	100	inversion	chr8:86151543|chr8:144603161
	132	93	duplication	chr8:92588680|chr8:110612811
	78	74	inversion	chr8:91933728|chr8:144594818
	90	59	inversion	chr8:98152159|chr8:146077903
	71	55	duplication	chr8:128411232|chr8:137092060
	44	55	duplication	chr8:119968554|chr8:125269158
	63	50	inversion	chr8:103167217|chr8:108419056
	51	43	deletion	chr8:125496202|chr8:128154508
	277	31	duplication	chr8:91585131|chr8:92589079
	21	41	deletion	chr8:96391716|chr8:102656636
	27	40	inversion	chr8:92738707|chr8:103228839
	37	36	inversion	chr8:130545868|chr8:144503944
	65	33	duplication	chr8:127594366|chr8:128155000
	24	34	duplication	chr8:102468947|chr8:117016078
	72	32	duplication	chr8:108403007|chr8:142698277
	74	26	deletion	chr8:96408243|chr8:136575345
	24	24	deletion	chr8:87109716|chr8:123090840
	12	23	duplication	chr8:86993707|chr8:123091034
	13	15	deletion	chr8:134311536|chr8:142529042
	28	8	inversion	chr8:117012059|chr8:121316990

* found manually by visual inspection

Supplementary Methods

Sections

Patient samples and nucleic acid extraction

All patient samples were procured in accordance with the Research Ethics Board at The Hospital for Sick Children (Toronto, Canada). Samples were obtained as frozen tissue biopsies at the time of diagnosis and stored at -80°C until processed for purification of nucleic acids. Frozen tissue was available for ~75-80% of cases included in the study; the remaining cases were shipped as pre-isolated DNA and/or RNA.

Whenever possible, tumour isolates were partitioned for both standard DNA and RNA extraction. Tissues were either manually homogenized using a mortar and pestle in the presence of liquid nitrogen or in an automated manner using a Precellys 24 tissue homogenizer (Bertin Technologies, France), according to the manufacturer's instructions. High molecular weight DNA was extracted by SDS/Proteinase K digestion followed by 2-3 phenol extractions and ethanol precipitation. Total RNA was isolated using the Trizol method (Invitrogen, USA) using standard protocols. DNA and RNA were quantified using a NanoDrop 1000 instrument (Thermo Scientific, USA) and integrity assessed either by agarose gel electrophoresis (DNA) or Agilent 2100 Bioanalyzer (RNA; Agilent, USA) at The Centre for Applied Genomics (TCAG, Toronto, Canada). RNA with an RNA Integrity Number (RIN) of ≥7.0 was required for analysis by either Affymetrix Gene array or RNASeq.

SNP6 copy number analysis

SNP array processing and quality control

For genotyping of DNA samples we used the Affymetrix Genome-Wide Human SNP Array 6.0, which includes more than 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation. All 1.8 million probes can be used to detect CNVs making the median inter-marker distance less than 700 bases across the genome.

DNA was independently prepared, labeled, and hybridized to the Affymetrix SNP 6.0 arrays according to standard protocols as provided by the manufacturer and described elsewhere[1-3]. Briefly, 500ng of total genomic DNA was digested with NspI and StyI restriction enzymes (New England Biolabs, Boston, MA) to prepare a reduced representation of the genome that was ligated to universal adaptors and amplified by PCR. PCR conditions were optimized to preferentially amplify fragments ranging from 200-1,100bp. PCR amplification products for each restriction enzyme digest were combined and purified using polystyrene beads. The amplified DNA was then fragmented with DNaseI, biotin-labeled, and hybridized to the array. After hybridization, arrays were washed on Affymetrix fluidics stations and scanned using the Gene Chip Scanner 3000 7G.

For initial quality control, array files were loaded into Affymetrix Genotyping Console (GTC). Those meeting Affymetrix's recommended quality control guideline of contrast QC>0.4 were used for further analyses. Following copy number analysis in GTC, we used the MAPD score (i.e. Median of the Absolute values of all Pairwise Differences between log2 ratios for a given chip) to identify possible outlier arrays, using the Affymetrix recommended threshold of MAPD=0.35 as an upper limit for acceptable hybridizations.

To check for duplicate samples and cryptic relatedness between cases, pair-wise identity by descent (IBD) was calculated for every pair of cases from the genotypes of 934,968 SNP probes on the SNP6 platform, using the PLINK toolset[4]. Those pairs of samples with PI_HAT values (defined as P(IBD=2)+0.5*P(IBD=1)) greater than 0.2 were further investigated for relatedness and only one sample from each pair was used in the final dataset.

Generation and normalization of copy number profiles

Affymetrix SNP6 CEL files were processed in dChip[5] to obtain raw copy number estimates. Arrays were normalized by quantile normalization and signal intensities computed using the MBEI method (PM-only) inherent to the dChip program. To generate a diploid reference baseline for copy number analysis of medulloblastoma samples, we used Affymetrix SNP6 data from 132 individuals from the Ontario Population Genomics Platform (OPGP) epidemiological project and the HapMap PROJECT. Germline DNA from these samples was genotyped in the same microarray facility as the tumour samples, using identical experimental protocols as described above.

The normalized copy number estimates from dChip were subsequently imported into the R environment, and the copy number profiles were segmented using the Circular Binary Segmentation (CBS) algorithm from the DNAcopy (v1.24)[6] package, with the undo split option enabled (method = sdundo, undo.SD = 1). The resulting segmentation profiles were further processed to reduce artificial segments. Segments with fewer than 10 markers were removed. Adjacent segments whose copy number states differed by less than 0.25 were merged together using their size-weighted mean. This merging step is repeated iteratively until no more segments can be merged. The segmentation profile of each sample was then median-centered. Further, normal CNVs reported in the Database of Genomic Variants (DGV, v10)[7] were filtered from the segmentation profiles. Segments that reciprocally overlapped (Dice coefficient > 0.5) with normal CNVs were removed. CNVs reported on BAC End Sequencing, BAC Array CGH, ROMA, and FISH were excluded from this filter. Upon removing a segment, the upstream segment was merged to the downstream segment using a size-weighted mean. The aforementioned merging and CNV-filtering steps helped reduce the occurrence of broad segments being broken into non-contiguous pieces by artificial segments or normal CNVs (which influences the downstream GISTIC2 analysis and segment classification).

The copy number segments were classified as balanced or one of 6 copy number aberrations (CNAs) based on the following criteria:

Class	Log R ratio (r)	Size (s)
Balanced	$\lvert r \rvert \leq 0.2$	---
Gain	$r > 0.2$	---
Loss	$r < -0.2$	---
Focal gain	$r > 0.2$	$s < 12$ Mbp
Focal loss	$r < -0.2$	$s < 12$ Mbp
High level amplification	$r > \log_2(5/2)$	$s < 12$ Mbp
Homozygous deletion	$r < -\log_2(0.7/2)$	$s < 12$ Mbp

Identification of recurrent copy number aberrations

The post-processed segmentation files profiles were analyzed using two algorithms, GISTIC2 and modified CMDS, to identify recurrent copy number events. GISTIC2 requires prior single sample segmentation, and hence may be affected by the presence of segmentation artifacts. In contrast, CMDS requires no single sample segmentation and works with the raw copy number profiles

directly. Many post-processing steps were lacking from the distributed CMDS program (e.g. multiple hypothesis correction, peak calling, CNV filtering); therefore, a number of post-processing steps were added.

GISTIC2 (v2.0.12)[8] was run with default parameters unless specified otherwise (brlen = 0.5, conf = 0.9), on the segmentation profiles of the entire cohort and of each subgroup separately. The significant peaks were filtered based on the gene content (>= 1 gene spanned), size of the wide peak (< 12 Mbp), and additionally based on containment within a normal CNV region (one-way overlap > 90%) reported in the filtered DGV database. This second CNV-filtering step ensures that amalgamated regions reported in DGV are considered, as these regions would often have poor reciprocal overlap with individual query segments and would be missed by the previous CNV-filtering procedure. It is not possible to apply this secondary CNV-filtering procedure directly on the segmentation file, as the CNV regions are often very large and this filtering can cause the loss of many informative CNA segments.

CMDS (v 1.0)[9] was run using default parameters unless specified otherwise (w = 40, s = 1), on the unstratified and subgroup-stratified segmentation profiles, for each chromosome separately. Since the raw outputs had a high false positive rate, the outputs were processed further. The z-scores were recalculated (separately for each chromosome) from the Fisher's z-transformed of the Pearson correlations, using the means of the dominant components from Gaussian mixture models (k = 3). This recalculation ensured that the z-scores were not skewed due to the presence of multiple modes in the raw output. The z-scores were further detrended using Empirical Mode Decomposition (EMD)[10], which corrects for trends due to recurring broad events. The p-values derived from the z-scores were then corrected for multiple hypothesis correction using the qvalue R package(v1.1)[11]. Finally, the peak regions were identified using a simple q-value threshold (FDR = 0.05).

Estimation of the copy number states of genes

The copy number states of all RefSeq genes in each sample were inferred from the respective post-processed segmentation profiles. The state for a given gene was determined by the copy number segment (classified as described earlier) that spans the greatest proportion of the gene (if multiple segments span the gene). Further, a gene is considered lost/deleted only if any portion of its coding region is spanned by a loss/deletion segment; a gene is considered gained/amplified only if at least 50% of the coding region is spanned by a gained/amplified segment.

Identification of recurrent broad events

Identification of recurrent CNAs above explicitly excluded broad events (based on a broad length cutoff in GISTIC and by a detrending procedure after CMDS). Therefore, broad events in the segmentation profiles were analyzed separately, using an approach similar to GISTIC's broad event analysis.[8]

The $\log_2$ R ratio (LRR) of each chromosome was calculated using a size-weighted mean of all segments mapping to the chromosome. A chromosome was declared gained if its LRR was greater than 0.2, lost if the LRR was less than -0.2, and balanced otherwise. Unlike GISTIC, gained and lost broad events were analyzed together. The significance of the frequency of each broad event was tested using the exact binomial test. Each broad event frequency was compared to the background frequency, which was determined from a robust regression of the observed

frequencies with respect to gene content (i.e. number of RefSeq genes) across all chromosomes. (This approach was motivated by GISTIC's broad event analysis.)

Identification of chromothripsis

The occurrence of chromothripsis was identified using tumour copy number profiles as previously described.[12] Using the post-processed segmentation profiles, chromothripsis was identified on a chromosome based on the presence of greater than 10 copy number state changes. The enrichment of chromothripsis on a chromosome for a subgroup was determined using the hypergeometric test, comparing the observed incidence to the background incidence (tallied across all samples). Select samples inferred to have chromothripsis were confirmed by WGS to identify rearrangements.

Subgroup enrichment analysis of recurrent copy number aberrations

Recurrent copy number aberrations (CNAs) identified by GISTIC2 in the unstratified and subgroup-stratified analysis were tested for subgroup enrichment. In the case of stratified analysis, regions identified in each subgroup were combined together. Since the reported region coordinates of common CNA events can differ among the strata, regions that reciprocally overlap (Dice coefficient > 0.2) were merged together using their union. In the enrichment analysis, the frequency of recurrent CNAs in each subgroup was compared against the remaining subgroups, and odds ratios were estimated using Fisher's conditional maximum likelihood estimate. CNAs with odds ratios greater than 2.0 were considered subgroup-enriched. A similar enrichment analysis was repeated for comparing combinations of two subgroups against the remaining, in order to identify CNAs that are enriched in the ensemble of two subgroups (and are not otherwise enriched in a single subgroup).

Integration of gene expression and copy number aberrations

To directly assess the correlation between significant CNA and gene expression, expression profiles were generated on 285 medulloblastomas from our study, including samples from SHH (n=51), Group 3 (n=46), and Group 4 (n=188), using the Affymetrix Gene 1.1 ST platform. Global integration of expression data was performed by comparing expression levels of amplified or deleted genes relative to genes in balanced regions (Mann-Whitney tests). Further, specific integration of expression data at each significant CNA locus were done by comparing expression of genes in samples with the aberration against those that do not, within in medulloblastoma subgroup. Multiple hypothesis correction by the false discovery rate method was applied to each locus independently, and the false discovery rate threshold was adaptively tuned for each locus so that the no false positives are expected. The resulting lists of genes with copy-number driven expression were used in candidate identification.

Identification of candidate driver genes in each significant region

Many of the significant regions identified using GISTIC and CMDS span multiple genes. Therefore, multiple lines of evidence were used to prioritize putative target genes within each region. Evidences collected from integrated expression data, the literature, and multiple datasets were classified into the following tiers:

Type	Description	Tier
Correlated expression	Gene expression is driven by SCNA in the integrated analysis	I
Medulloblastoma literature	Implicated in medulloblastoma from the literature (PubMed)	I
Cancer Gene Census	Documented in the Cancer Gene Census	II
Parsons	Reported to be somatically mutated in the Parsons *et al.*[13] study on SNVs identified in medulloblastoma	III
ICGC	Reported to be somatically mutated in the ICGC medulloblastoma WGS study (Nature, in review)	III
Northcott signature gene	Reported in the Northcott *et al.*[14] medulloblastoma expression subgroup study	IV
Cho signature gene	Reported in the Cho *et al.*[15] medulloblastoma expression subgroup study	IV
RNASeq SNV	Identified in the pilot RNASeq screen for SNV (Group 3 and 4 only)	IV
Gli1 target	Identified as a Gli1 target in Lee *et al.*[16] (SHH subgroup only)	IV
Shh-inducible target	Identified as Shh-inducible gene by a screen in cerebellar granule neuron precursors (SHH subgroup only)	IV
COSMIC	Documented in the Catalogue of Somatic Mutations in Cancer	V

The priority of each gene in a region was ranked based on the total score of the above lines of evidence, weighted by their respective tiers. A higher tier (lower number) is assigned a higher weight. The weights were assigned so that supporting evidence from the next tier is only considered for breaking ties at previous tiers. At most two genes from each region were selected for network analysis using this evidence-driven ranking.

Mutual exclusivity analysis

The significant gene lists were analyzed to detect mutual exclusive relationships by iterating through all possible combination of genes in the list (for combination sizes from 2 to 6). Within each subgroup, combinations of genes with the highest exclusivity scores were identified. The exclusivity score was defined as the number of samples that harbour exactly one aberration among the genes in the combination; in other words, it is the product of exclusivity and coverage as defined by Miller *et al.*[17]

$$coverage = \frac{\text{number of samples with any aberration of genes in the combination}}{\text{total number of samples}}$$

$$exclusivity = \frac{\text{number of samples with exactly 1 aberration of genes in the combination}}{\text{number of samples with any aberration of genes in the combination}}$$

$$coverage = coverage \times exclusitivity$$

Network analysis

Pathway enrichment analysis of copy number aberrations was carried out using g:Profiler web server[18] and visualized in Cyotscape software as an Enrichment Map network[19]. Candidate driver genes from selected GISTIC2 regions were compiled for SHH, Group 3, and Group 4, ranked by frequency, and subsequently queried for significantly enriched functional categories with g:Profiler using the ordered list algorithm (FDR-corrected cutoff p=0.05, hypergeometric test).

Detected categories were filtered with a custom R script to only include Gene Ontology (GO) terms, and Reactome and KEGG pathways, using an upper limit of 500 genes per gene set and the requirement that at least two putative driver genes intersect with the gene set. A small number of non-informative KEGG pathways were removed from the final list. The overlap coefficient value 0.6 was used in Enrichment map visualization. Enrichment maps were manually adjusted to highlight the most significant themes for visualization purposes.

Unsupervised clustering analysis of copy number events

All significant broad events (spanning chromosome arms) and focal events (identified as described earlier) identified in the pan-cohort analysis were used in the unsupervised clustering of medulloblastoma samples, by the Ward linkage method and the Euclidean distance metric, as implemented in the R environment. The copy number states were converted to absolute values and samples with unknown subgroup affiliation were removed prior to clustering. The agreement between the observed clusters and the medulloblastoma subgroups were assessed by the Adjusted Rand Index and tested by the χ^2 test.

Expression array processing and data analysis

For gene expression array profiling, 400ng total RNA was processed and hybridized to the Affymetrix Gene 1.1 ST array at TCAG according to the manufacturer's instructions. The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm. Prior to clustering analysis of Group 4 medulloblastomas, 500-1000 high standard deviation (SD) genes were selected and the expression signals unlogged. Unsupervised clustering was carried out using the NMFConsensus module available on the GenePattern public server (Broad Institute, USA) with default parameters. The cophenetic coefficient metric was used to assess stability of the tested sample clusters.

nanoString CodeSets and data analysis

To determine subgroup affiliation of the MAGIC cohort, a custom nanoString CodeSet was designed to assess the expression status of 22 medulloblastoma signature genes and samples processed as described in detail in a recent publication[20]. Samples were processed as recommended by nanoString at the University Health Network (UHN) Microarray Facility using an input of 100ng total RNA. Raw nanoString counts for each gene within each experiment were subjected to a technical normalization using the counts obtained for positive control probe sets prior to a biological normalization using the three housekeeping genes included in the CodeSet. Normalized data was $\log_2$-transformed and then used as input for class prediction analysis using the PAM method. A series of 101 medulloblastomas with known subgroup affiliation were used as a training dataset for class prediction[14].

RNASeq library generation and data analysis

All RNASeq was performed at the BC Cancer Agency Genome Sciences Centre (Vancouver, Canada), according to established protocols[21]. Samples selected for RNASeq exhibited RIN values of >7.0 as determined by Agilent bioanalyzer. RNA sequencing libraries were constructed from

2ug of DNAseI treated mRNA as previously described[21]. Paired-end libraries were sequenced on an Illumina Genome Analyzer II, generating 100bp sequence reads.

Reads were aligned to the human reference genome (National Center for Biotechnology Information Build 36.1; hg18) and a database of exon junctions[22], using the Burrows-Wheeler aligner[23].

Identification of gene fusions

To identify fusion genes, we conducted *de novo* assembly of each library using Trans-ABySS[24]. First, using ABySS[25], reads were assembled into contigs. Since transcripts can differ widely in expression, multiple assemblies were performed in parallel using different starting *k*-mer values (i.e. substrings of *k* bp). Contigs from different assemblies were merged into a smaller non-redundant set. Paired-end information was then used to estimate inter-contig distances, and to unambiguously merge contigs into a final assembly. Final contigs were aligned to known transcripts (Ensembl, UCSC genes, RefSeq, and Aceview), and to the reference human genome. Alignments that did not match any known annotations, and those that matched multiple annotations were shortlisted as candidate large-scale rearrangements and fusion genes, respectively.

Whole genome sequencing (WGS)

WGS was performed at the BC Cancer Agency Genome Sciences Centre (Vancouver, Canada), according to established protocols[21]. WGS sequence data was aligned to the reference human genome using the Burrows-Wheeler algorithm[23]. MB-586 was aligned to hg18, while the later samples (MB-182, MB-506, and MB-591) were aligned to hg19. Copy-number state was inferred from read alignment data using CnaSeq (http://www.bcgsc.ca/platform/bioinfo/software/cnaseq). CnaSeq compares read coverage between matched samples (i.e. tumor vs. normal), and reports the relative $\log_2$ ratio of read depth in bins of 200bp. The analysis was repeated using a custom program that compares median read depths between matched samples in 1000kbp windows. ABySS was run to identify regions participating in large-scale rearrangements[25], confirming the events identified using Trans-ABySS, and providing accurate, base-pair resolution rearrangement breakpoints.

Large insert, paired-end WGS was also performed at the EMBL, as previously described[26].

Simulation of Progressive Genomic Rearrangements

A Monte Carlo strategy was employed to simulate the step-wise generation of the observed breakpoints. Briefly, rearrangements were randomly sampled without replacement from the observed events and successively applied to the chromosome 8 sequence. Starting with the wild-type chromosome 8 sequence, each rearrangement was applied to the current chromosomal configuration at each step. When more than one copy of the rearrangement breakpoint existed in the current configuration (e.g. because it had been duplicated), one of the possible locations was randomly picked. When a rearrangement breakpoint was no longer available (e.g. due to a previous deletion event), it was discarded and another selected. Simulations were repeated 1000 times, and for each instance we noted the maximum copy number achieved, as well as the mean copy number in excess of unamplified segments (e.g. wild-type copy number). None of the simulations reached the high-level copy number (either mean or max) of the observed amplicon.

Circos plot and double-minute chromosome structure

Trans-ABySS output was used to determine the physical connections between amplified chromosome 8 segments. Briefly, read pairs that flanked the breakpoints were used along with spanning reads (reading through the breakpoint) to link disparate amplified segments (these correspond to assembled Trans-ABySS contigs). This process revealed in MB-586 that the 5 amplified regions are fused together in non-canonical order and orientation, and form a 2.36Mb double minute chromosome structure.

Identification of actionable targets

For genes displaying significant (GISTIC) copy number gain, we searched for antagonists using various drug compound databases. The methodology is as follows:

Using gene symbols, we searched initially in the GeneCards Database (1), which provided compound information from several other sources including DrugBank (2), Human Metabolome Database (3), Tocris Bioscience (4), Novoseek, STITCH (5), Pubchem (6), and Pharmacogenomics Knowledge Base. We next queried the following databases: SuperTarget (7), Matador (8), Comparative Toxicogenomics Database (CTD, 9), Promiscuous (10) and Potential Drug Target Database (PDTD, 11). In the event of multiple drug targets, a separate entry was listed. Direct targets were defined as drug compounds with a reported physical interaction with the protein of interest, while indirect targets were compounds, which were reported to decrease gene transcript or protein expression. PubMed ID references were also provided for gene/antagonist relationships. Lastly, we determined whether identified compounds were being used in US clinical trials (12), and/or FDA approved (13).

Gene Cards: http://www.genecards.org/

Drug Bank: http://www.drugbank.ca/

HMDB: http://www.hmdb.ca/

Tocris Bioscience: http://www.tocris.com/

STITCH: http://stitch.embl.de/

Pubchem: http://pubchem.ncbi.nlm.nih.gov/

SuperTarget: http://insilico.charite.de/supertarget/index.php?site=home

Matador: http://matador.embl.de/

CTD: http://ctdbase.org/

Promiscuous: http://bioinformatics.charite.de/promiscuous/index.php?site=home

PTDTD: http://www.dddc.ac.cn/pdtd/index.php

US Clinical Trials: http://clinicaltrials.gov/

FDA: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm

PVT1-MYC fusion detection

RNA was converted to cDNA using SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, U.S.A.) following the manufacturer's instructions. To validate the PVT1-MYC and PVT1-NDRG1 fusions identified by RNASeq, PCR primers were used under the following amplification conditions: [94°C 3 min, (94°C 30 sec, 58°C 30 sec, 72°C 1 min) x 30 cycles, 72°C 5 min, 4°C hold]. PCR amplicons were cut and purified from 1% agarose gels using the QIAquick gel extraction kit (Qiagen, U.S.A.). Purified amplicon was subcloned into the TOPO TA Cloning vector pCR2.1-TOPO (Invitrogen, U.S.A,), and the plasmid was purified using the Purelink quick plasmid miniprep kit following the manufacturer's instructions, prior to Sanger sequencing.

To detect other PVT1-MYC fusions, cDNA from *MYC* amplified tumours was screened using PCR primers that anneal to PVT1 exon 1 and downstream of the MYC exon 2 alternative translation start site. The target locus was amplified with primers (PVT1-Ex1-F2, MYC-Ex2-Rev2) and Taq polymerase (Invitrogen, U.S.A.) [94°C 3 min, (94°C 30 sec, 58°C 30 sec, 72°C 1 min) x 30 cycles, 72°C 5 min, 4°C hold] or Phire II (NEB, U.S.A.) [98°C 30 sec, (98°C 5 sec, 63°C 5 sec, 72°C 1 min) x 40 cycles, 72°C 5 min, 4°C hold]. PCR amplicons were cut and purified from 1% agarose gels using the QIAquick gel extraction kit (Qiagen, U.S.A.). Purified amplicon was subcloned into the TOPO TA Cloning vector pCR2.1-TOPO (Invitrogen, U.S.A,), and the plasmid was purified using the Purelink quick plasmid miniprep kit following the manufacturer's instructions, prior to Sanger sequencing.

Primer sequences:

PVT1-F1	AGCTGGCTGAGAGGGTTGAG
MYC-Exon2-R1	TTCCAGATATCCTCGCTGGG
PVT1-NDRG1-fus-Fb	CAGTGGATTTCCTTGCGGA
PVT1-NDRG1-fus-Rb	GAGTTGCACTCCACCACGG
PVT1-Ex1-F2	GTTGGCGGTCCCTGTGAC
MYC-Ex2-Rev2	CCTGTTGGTGAAGCTAACGTTG

TaqMan qPCR of miR-1204, 1205, 1207

Total RNA (2ug) was reverse transcribed to cDNA using the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, USA). The miRNA specific primers for hsa-miR-1204 (002872), hsa-miR-1205 (002778), hsa-miR-1207-3P (002826) and RNU6B (001093) were multiplexed in the reverse transcription reactions: 16°C 30 min, 42°C 60 min, 85°C 5 min. Taqman qPCR reactions were run in triplicates and contained 154 ng/well of converted total RNA. Preliminary results showed hsa-miR-1206 (002878) was not detectable and was excluded from the experiment.

The expression level of target miRNAs was determined using the ddCT method,[27] after adjusting the cycle threshold (CT) for plate-to-plate variation with a calibrator sample. The CT values were subsequently normalized against the endogenous control (RNU6B), and the resulting dCT values for each target were expressed relative to the mean of the reference samples. Tumour samples that were balanced at the miRNA locus served as the references.

CTNNB1 mutation analysis

Seventy-one medulloblastoma samples were analyzed for mutations of the *CTNNB1* gene by direct sequencing. Genomic DNA was subjected to PCR for the amplification of *CTNNB1* from exon 2 to exon 4 using primers BCATEx2F (GAAAATCCAGCGTGGACAATG) and BCATEx4R (TCGAGTCATTGCATACTGTCC) with Platinum Pfx DNA Polymerase (Invitrogen) under the following conditions: 94°C for 3 min; 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 68°C for 2 min; and then a final extension step of 68°C for 7 min. The 630bp PCR product was then purified using PureLinkTM PCR Purification Kit (Invitrogen), followed by sequencing using primers BCATEx3F (GCTGATTTGATGGAGTTGGAC) and BCATEx3R (AAGGACTGAGAAAATCCCTGTTC).

Cell lines, cell culture and transfections

The ONS76 cell line was obtained from the Institute for Fermentation (Osaka, Japan). The MED8A cell line was obtained from Dr. Richard J. Gilbertson (St Jude Children's Research Hospital, Memphis, TN). ONS76 was maintained in culture in Dulbecco modified Eagle medium (DMEM) supplemented with 10% FBS and 1x antibiotic/antiMYCotic solution. MED8A was cultured in DMEM supplemented with 20% FBS and 1x antibiotic/antiMYCotic solution. All cell lines were kept at 37°C, in 5% CO2. Media and reagents for cell culture were purchased from Wisent, Inc (St Bruno, Quebec, Canada). Anti hsa-miR1204 (TAATGGAGACCAGGCCACG) and Negative CTL A (GTGTAACACGTCTATACGCCCA) were purchased from Exiqon Life Science (miRCURY LNA™ microRNA Inhibitor cat: 411266-04 and 199004-04 respectively). MYC silencing was performed with a siRNA: CGAUUCCUUCUAACAGAAAUU (Dharmacon).

MTS Cell Proliferation Assay

The Promega CellTiter 96 AQueous One Solution Cell Proliferation Assay was used, as per the manufacturer's instructions (Promega, Madison, WI). Low confluence ONS76 and MED8A were transfected with 10pmol of oligonucleotides with X-tremeGENE siRNA Transfection Reagent (Roche) as recommended by the manufacturer. Proliferation was measured starting from 18h after transfection (t=0) for a 72-hour period. Average O.D. at 490nm of two different experiments (each conducted in triplicate) is shown; statistical significance at 72h post transfection was assessed by T-Test.

Statistical and bioinformatic analyses

Statistical and bioinformatic analyses were performed in the R statistical environment (v2.13) or using custom programs/scripts written in Python, C++, or Go. Enrichment analyses were done using the hypergeometric test. The significance of chromosome arm frequencies were done using the exact binomial test, comparing the observed frequency to the expected frequency derived from a robust regression of event frequency and gene content, in a similar manner to the 'broad analysis' in GISTIC2. Comparisons of event frequencies across medulloblastoma subgroups were performed using Fisher's exact test. Expressions of genes across samples were compared using the Mann-Whitney test (and confirmed with the Student's independent t-test). Where applicable, multiple hypothesis corrections were performed using the false discovery rate method[11]. Univariate survival analyses were done using the log-rank test, as implemented in the survival R package (v2.36).

References

1. Kennedy, G.C. et al. Large-scale genotyping of complex DNA. *Nature biotechnology* **21**, 1233-7 (2003).
2. Korn, J.M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* **40**, 1253-60 (2008).
3. McCarroll, S.A. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature genetics* **40**, 1166-74 (2008).
4. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-75 (2007).
5. Lin, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-40 (2004).
6. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-63 (2007).
7. Iafrate, A.J. et al. Detection of large-scale variation in the human genome. *Nat Genet* **36**, 949-51 (2004).
8. Mermel, C.H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
9. Zhang, Q. et al. CMDS: a population-based method for identifying recurrent DNA copy number aberrations in cancer from high-resolution data. *Bioinformatics* **26**, 464-9 (2010).
10. Wu, Z., Huang, N.E., Long, S.R. & Peng, C.K. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc Natl Acad Sci U S A* **104**, 14889-94 (2007).
11. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-5 (2003).
12. Rausch, T. et al. Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations. *Cell* **148**, 59-71 (2012).
13. Parsons, D.W. et al. The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435-9 (2011).
14. Northcott, P.A. et al. Medulloblastoma comprises four distinct molecular variants. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 1408-14 (2011).
15. Cho, Y.J. et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol* **29**, 1424-30 (2011).
16. Lee, E.Y. et al. Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis. *Proc Natl Acad Sci U S A* **107**, 9736-41 (2010).
17. Miller, C.A., Settle, S.H., Sulman, E.P., Aldape, K.D. & Milosavljevic, A. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Med Genomics* **4**, 34 (2011).
18. Reimand, J., Arak, T. & Vilo, J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* **39**, W307-15 (2011).

19. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one* **5**, e13984 (2010).

20. Northcott, P.A. et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta neuropathologica* **123**, 615-26 (2012).

21. Shah, S.P. et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809-13 (2009).

22. Morin, R. et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques* **45**, 81-94 (2008).

23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).

24. Robertson, G. et al. De novo assembly and analysis of RNA-seq data. *Nature methods* **7**, 909-12 (2010).

25. Simpson, J.T. et al. ABySS: a parallel assembler for short read sequence data. *Genome research* **19**, 1117-23 (2009).

26. Korbel, J.O. et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420-6 (2007).

27. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).