Report CoriandR: ChrOmosomal abeRration Identifier AND Reporter in R

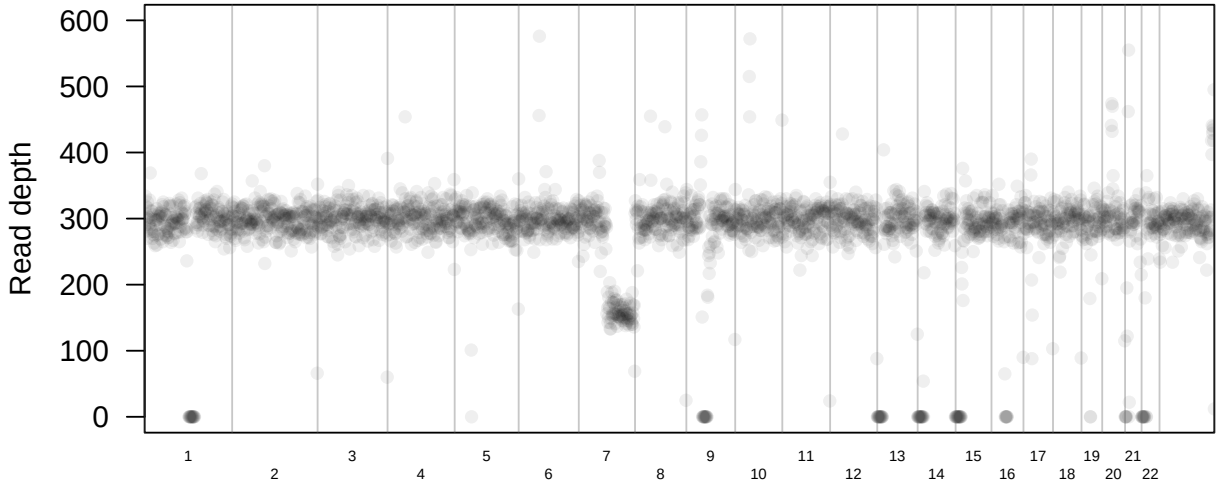
23-10-2024, 16:35

Calculated Karyotyping

This report documents the calculated karyotyping and estimation of CNVs from sequencing data of hamatological malignances and solid tumours in research with `coriandR`.

Absolute sequencing depth of sequencing data in sample

The following plot visualises the absolute sequencing depth in analysed sample. The grey lines represent the boundaries between the chromosomes. chromosomes, the chromosome names are shown in the lower part of the figure.



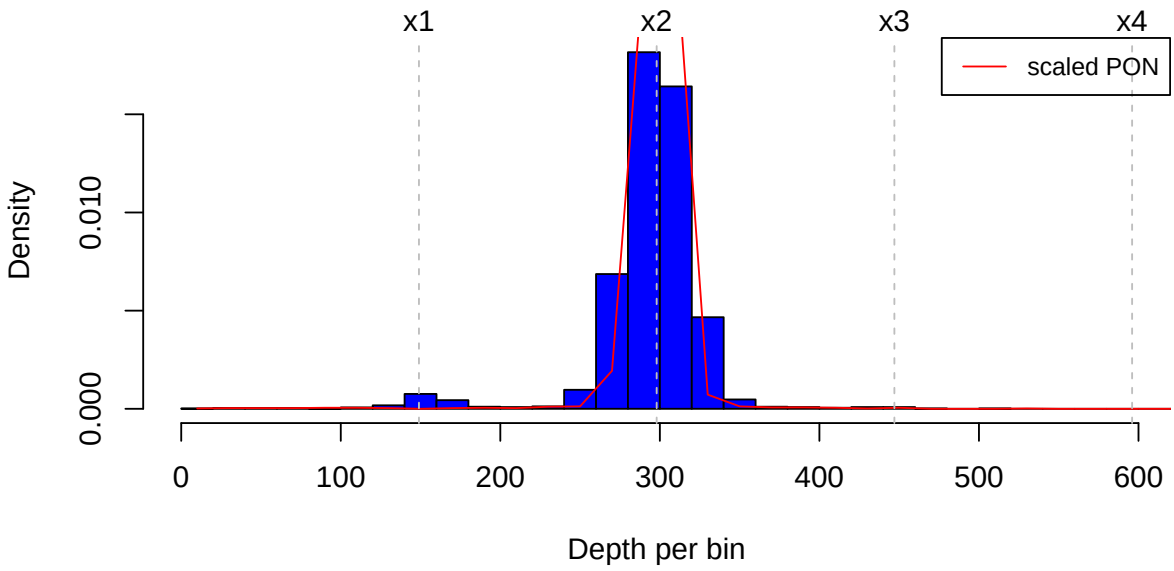
Sample sequencing characteristics and mapping statistics

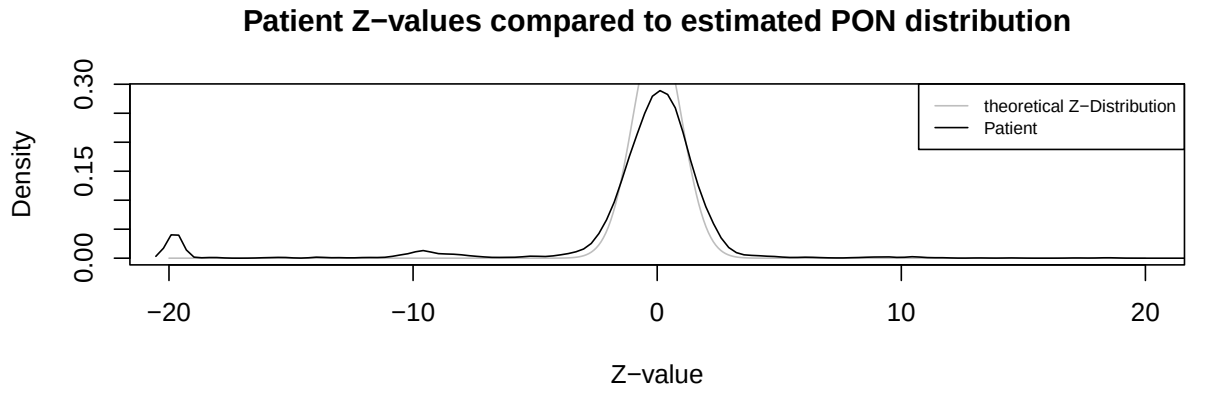
Sample characterisic	Value
Sample name	**98217**
Sample gender	**f**
Raw read pairs	897.816
Average read length	149,281
Unique mapping pairs	438.243

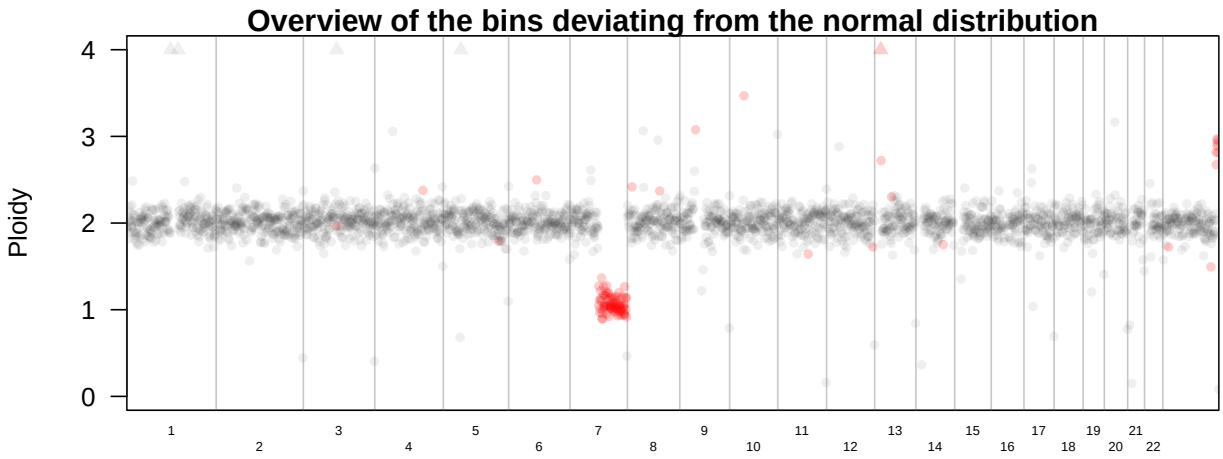
Distribution of sequencing depth per bin in sample

This picture shows the distribution of bins (with density on the y-axis) according to the number of reads per bin (x-axis). The red curve represents the distribution of the PON scaled to the same depth.

Ideally, the histogram and the red curve match completely. If not, this can indicate larger deletions (the first peak is smaller then the PON) or larger amplifications (the third peak becomes higher). The labels `x1`, `x2`, `x3`, `x4` indicate the ploidy of the bins.

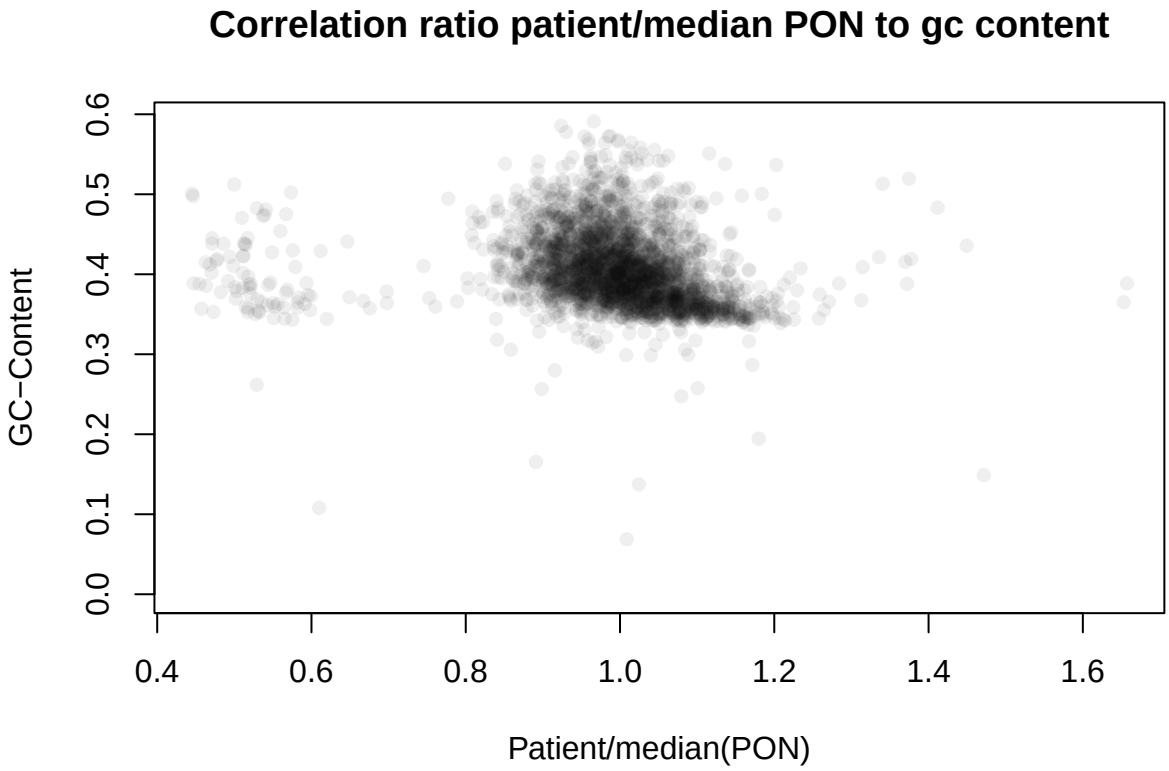






In the first step of calculated karyotyping, the sequencing data of a tumour sample are normalized by the median sequencing depth per bin. Thereafter, we used standardisation with calculation of the pseudo z-values of the distribution of the bins. A z-score represents the number of standard deviations from the standard value of the reference population (PON) for the analysed sample. The first picture visualises the z-values of the sample and PON.

In addition, we tested pseudo z-scores against a normal distribution with parameters of the PON in a two-tailed test. The obtained p-values were adjusted using the Benjamini-Hochberg method (Benjamini and Hochberg 1995) in control of the false discovery rate. In the second picture, an overview of normalised bins (data points) with expectes ploidy (y-axis) is shown across the chromosomal coordinates (x-axis). Bins with the read distribution outside the 99 % of the normal distribution are marked in red.



This picture displays the correlation of normalised median in sample / in PON to gc content in human reference genome. The gc content should be between 35 % and 60 %. High correlation between normalised median in sample / in PON to gc content can be a result of large abberations or sequencing bias.

Calculated numerical karyotype:

del(7)(q21.11q36.3)

amp(9)(p11.2p11.2)

amp(10)(q11.1q11.1)

amp(13)(p11.1q11)

amp(X)(q28q28)

The presentation of the calculated numerical karyotype is based on the ISCN nomenclature.

International Standing Committee on Human Cytogenomic Nomenclature and McGowan-Jordan, J. and Hastings, R.J. and Moore, S. (2020): ISCN 2020: An International System for Human Cytogenomic Nomenclature. In: An International system for human cytogenetic nomenclature. Karger Publishers. ISBN: 9783318067064.

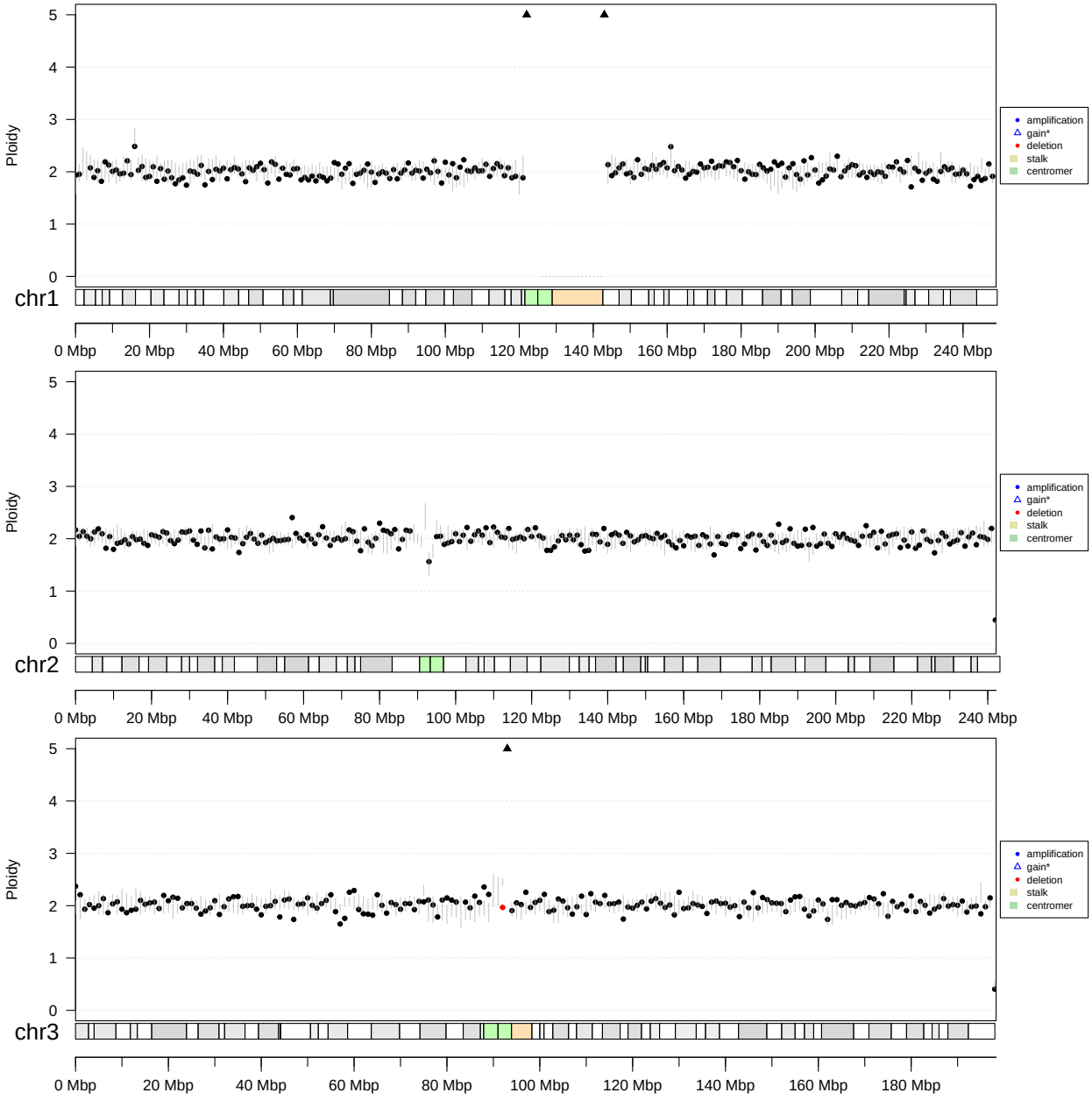
Genes affected by CNVs (copy number variations)*:

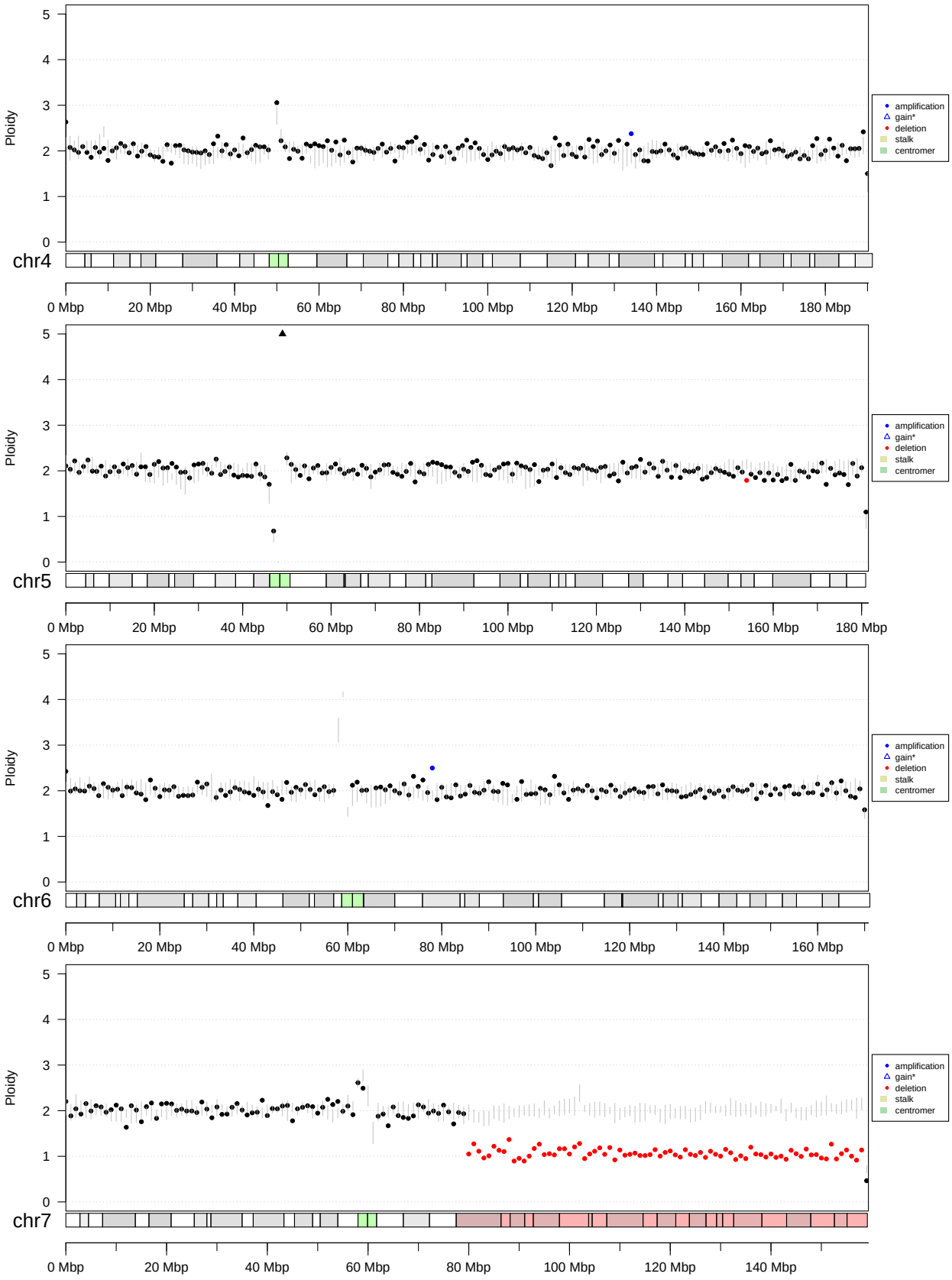
	id	symbol	chr	start	end	aberration
23074	ENSG00000019991	HGF	chr7	81.699.010	81.770.438	del
23552	ENSG00000105851	PIK3CG	chr7	106.865.278	106.908.980	del
23659	ENSG00000105976	MET	chr7	116.672.196	116.798.386	del
23865	ENSG00000128602	SMO	chr7	129.188.633	129.213.545	del
24082	ENSG00000157764	BRAF	chr7	140.719.327	140.924.928	del
24217	ENSG00000197993	KEL	chr7	142.941.114	142.962.363	del
24328	ENSG00000055130	CUL1	chr7	148.697.914	148.801.110	del
24329	ENSG00000106462	EZH2	chr7	148.807.383	148.884.321	del
24436	ENSG00000106615	RHEB	chr7	151.466.012	151.520.120	del
24449	ENSG00000055609	KMT2C	chr7	152.134.922	152.436.005	del
265031	ENSG00000156531	PHF6	chrX	134.373.253	134.428.791	del
268991	ENSG00000196924	FLNA	chrX	154.348.524	154.374.638	amp

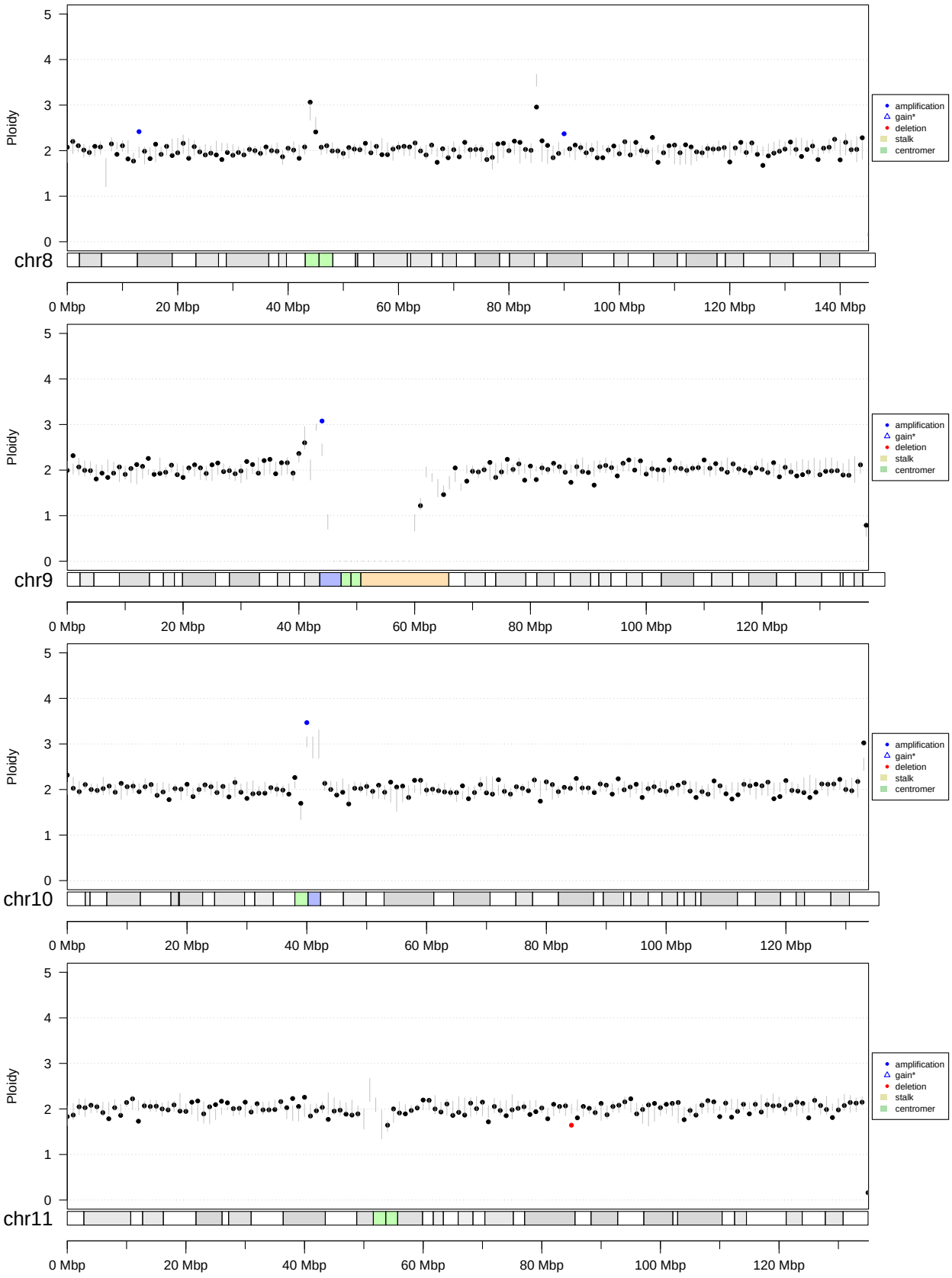
Bailey et al. (2018): Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 173 (2), 371-385.e18. DOI: 10.1016/j.cell.2018.02.060.;

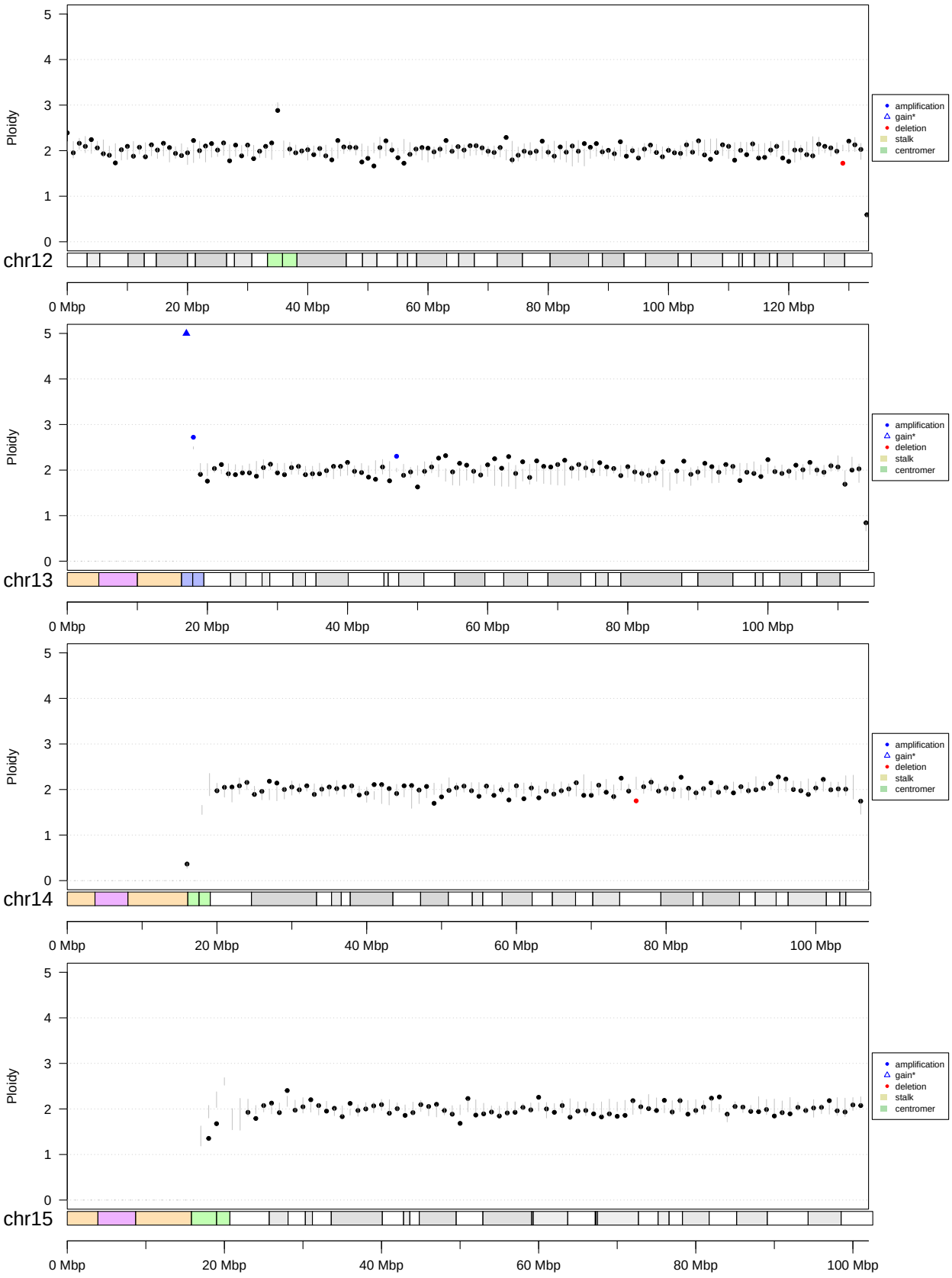
Papaemmanuil et al. (2016): Genomic Classification and Prognosis in Acute Myeloid Leukemia. The New England journal of medicine 374 (23), S. 2209–2221. DOI:10.1056/NEJMoa1516192.

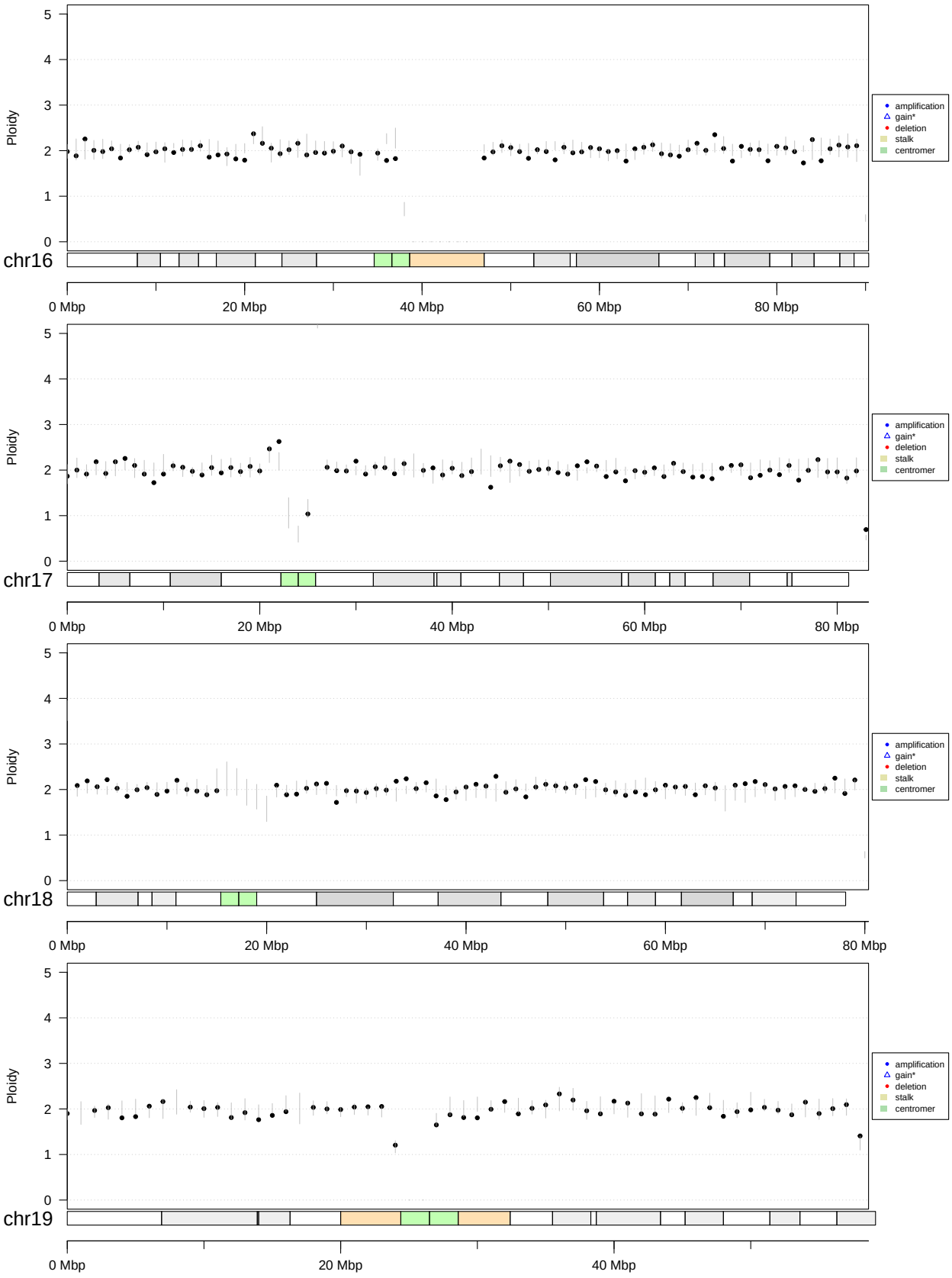
Chromosome overview plots:

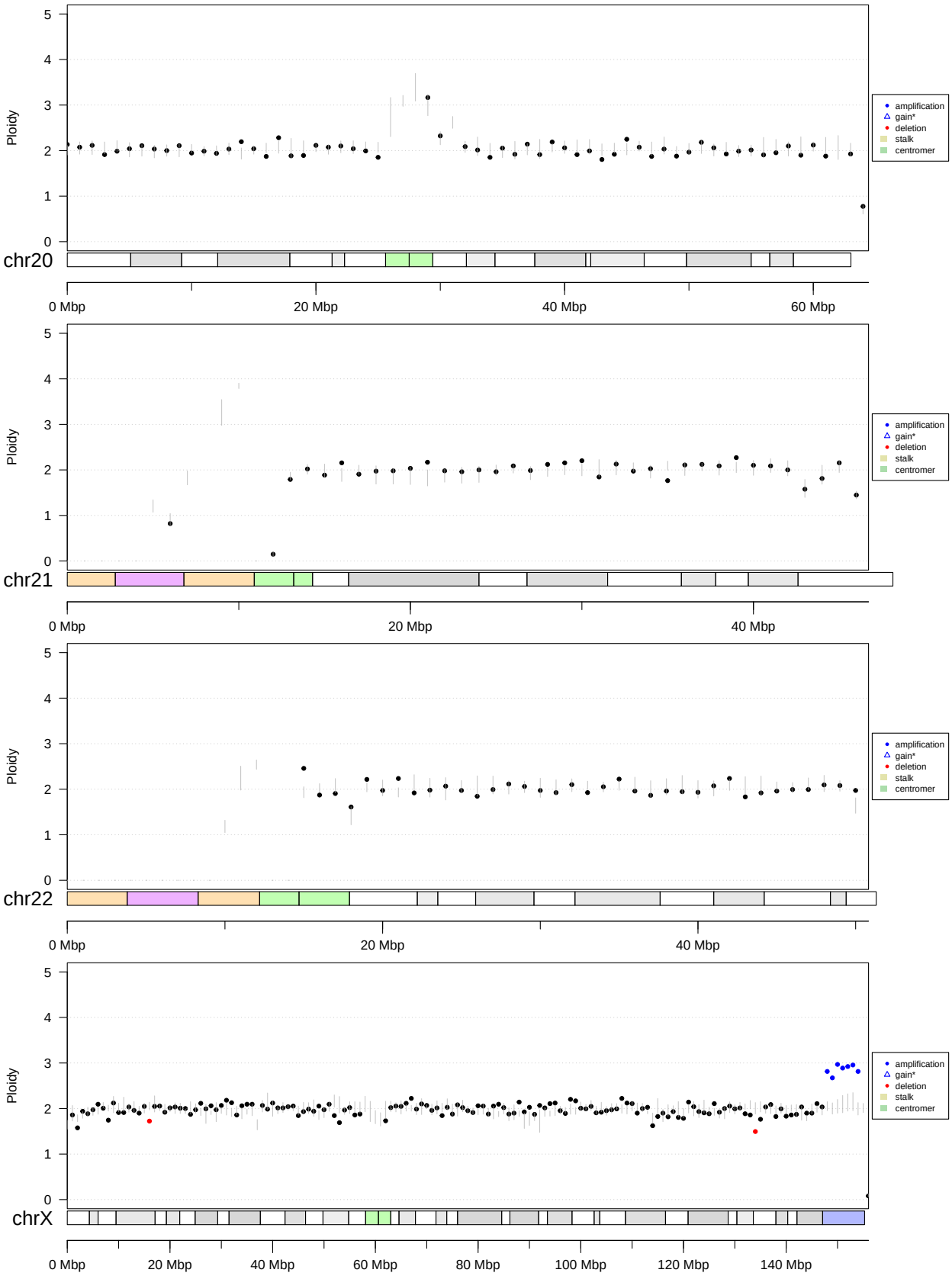



Ploidy
amplification
gain*
deletion
stalk
centromer
chr4
chr5
chr6
chr7
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
120 Mbp
140 Mbp
160 Mbp
180 Mbp


Ploidy
0
1
2
3
4
5
chr8
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
120 Mbp
140 Mbp
amplification
gain*
deletion
stalk
centromer
chr9
chr10
chr11


Ploidy
0
1
2
3
4
5
chr12
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
120 Mbp
amplification
gain*
deletion
stalk
centromer
Ploidy
chr13
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
amplification
gain*
deletion
stalk
centromer
Ploidy
chr14
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
amplification
gain*
deletion
stalk
centromer
Ploidy
chr15
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
amplification
gain*
deletion
stalk
centromer


Ploidy
0
1
2
3
4
5
chr16
chr17
chr18
chr19
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
amplification
gain*
deletion
stalk
centromer


Ploidy
0
1
2
3
4
5
chr20
0 Mbp
20 Mbp
40 Mbp
60 Mbp
amplification
gain*
deletion
stalk
centromer
chr21
0 Mbp
20 Mbp
40 Mbp
chr22
0 Mbp
20 Mbp
40 Mbp
chrX
0 Mbp
20 Mbp
40 Mbp
60 Mbp
80 Mbp
100 Mbp
120 Mbp
140 Mbp

* The reads that are marked with triangles have a particularly deep sequencing depth, i.e. there are more than 4 copies.

** The legend contains characteristics to Giemsa stain results (Cheung 2001). Recognised stain values: gneg, gpos50, gpos75, gpos25, gpos100, acen, gvar, stalk.

Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, Furey TS, Kim UJ, Kuo WL, Olivier M et al. Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature. 2001 Feb 15;409(6822):953-8. PMID: 11237021