WisecondorX: improved copy number detection for routine shallow whole-genome sequencing (supplementary tables and figures)

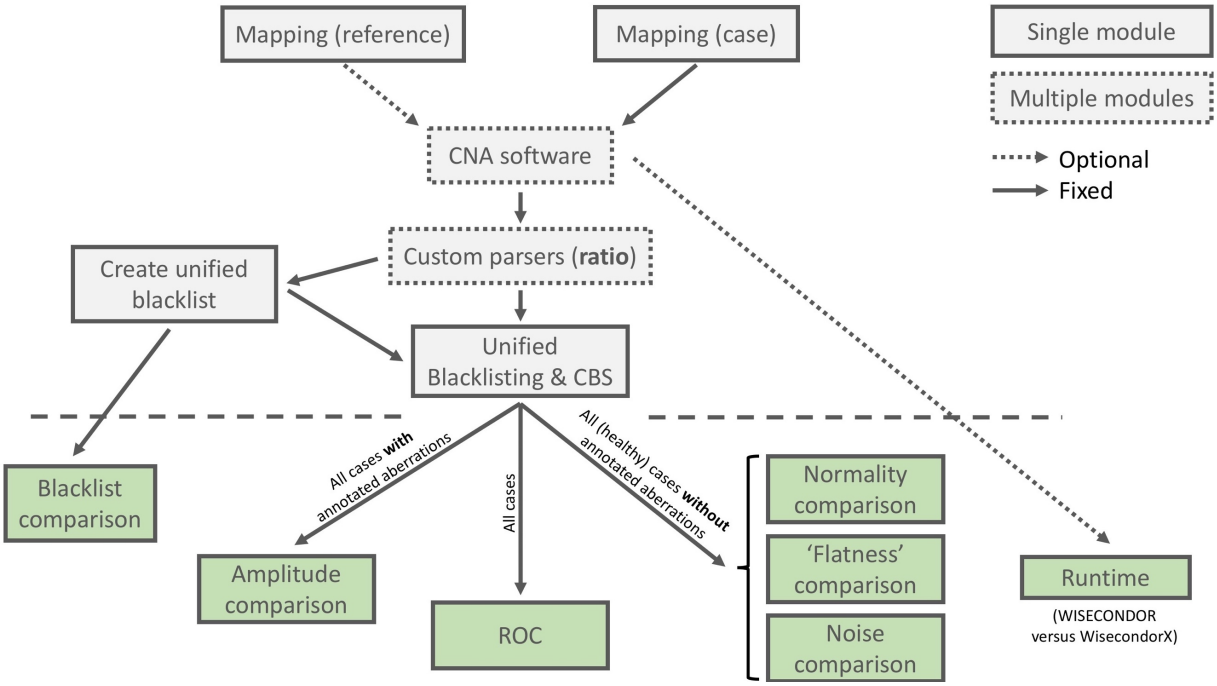

Figure S1. **Schematic overview of the computational phase in the comparative study**.
To avoid bias at any level, every sample is processed in an identical way. Finally, each case is assigned to a or multiple particular study component(s), being the part(s) for which it is most qualified.

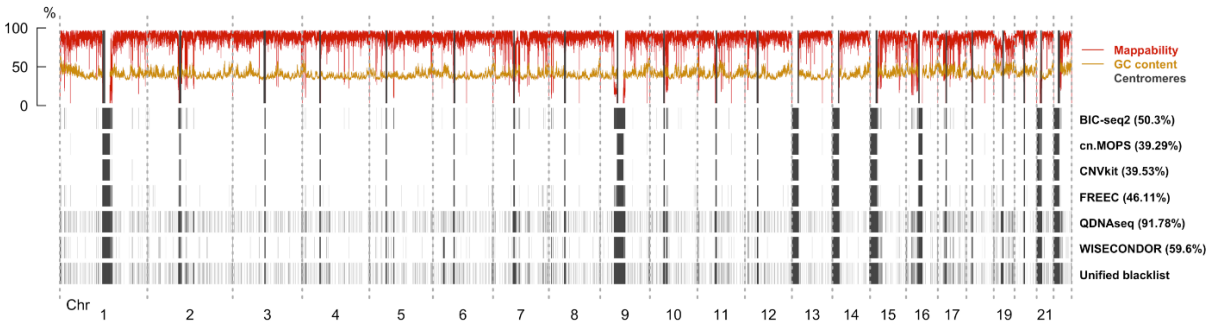

Figure S2. **Representation of the blacklists across the considered tools at 100 kb**.
Above, mappability, GC content and centromere locations are visualized. Blacklisted positions are shown by black bars, where opacity indicates width. Percentages show each tool's contribution to the unified blacklist, which by itself covers 16.04% of the human genome.

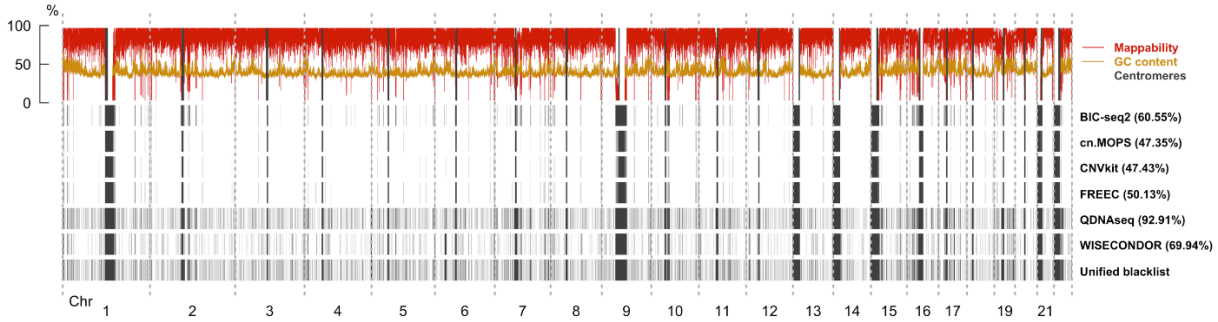

Figure S3. **Representation of the blacklists across the considered tools at 30 kb**.

Above, mappability, GC content and centromere locations are visualized. Blacklisted positions are shown by black bars, where opacity indicates width. Percentages show each tool's contribution to the unified blacklist, which by itself covers 13.94% of the human genome.

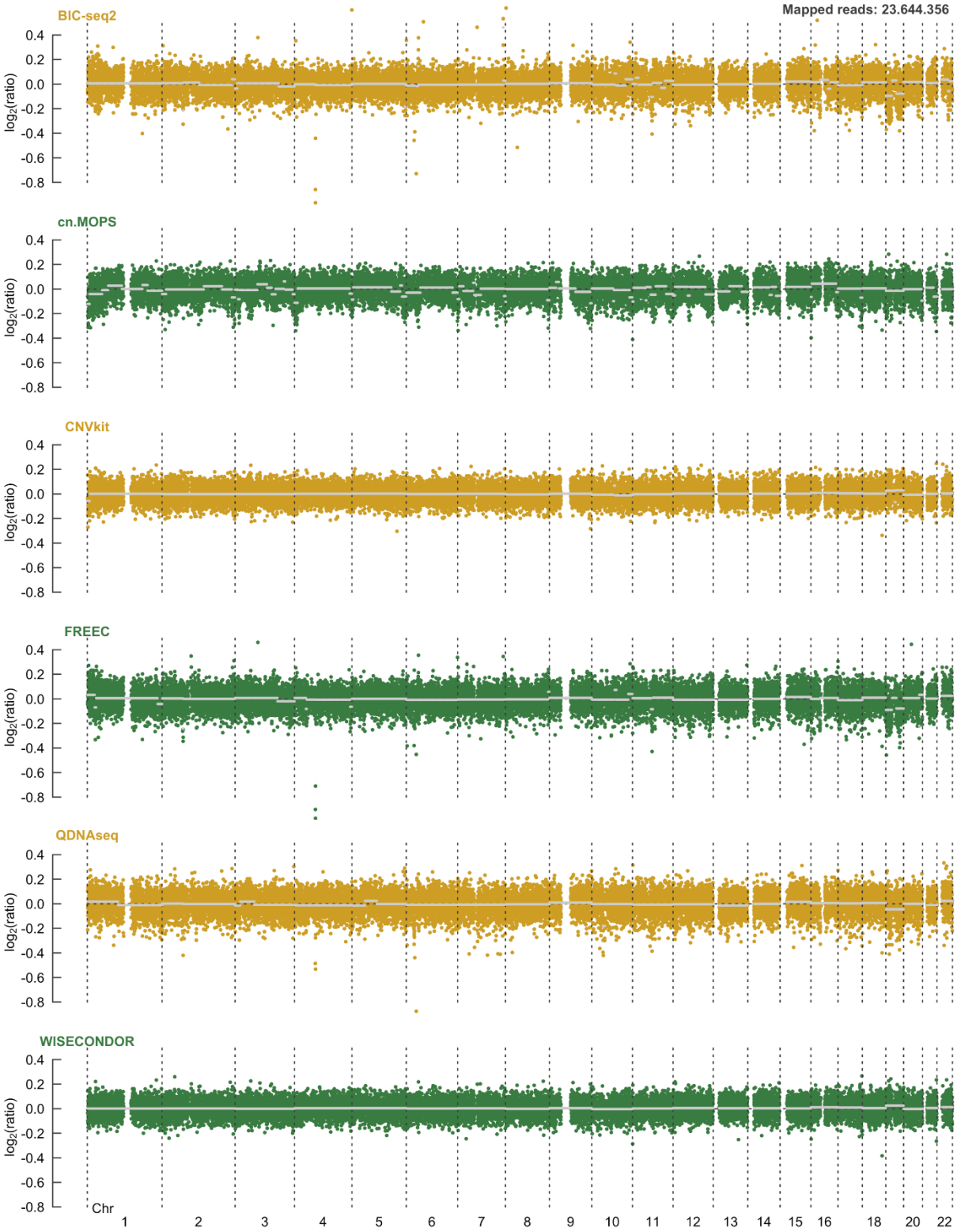

Figure S4. **Typical autosome-wide profiles for a healthy NIPT sample**.

NIPT-1 is depicted.

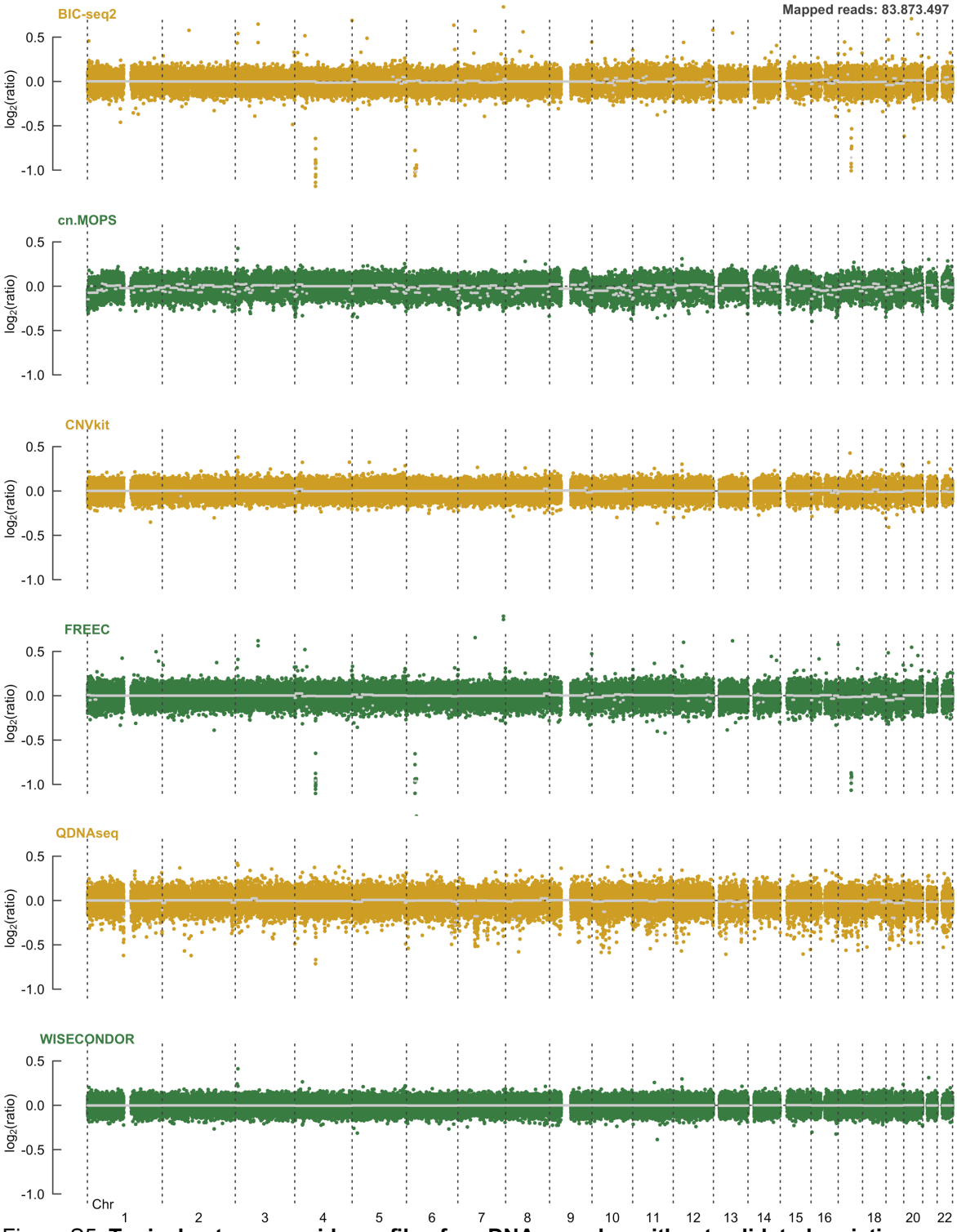

Figure S5. **Typical autosome-wide profiles for gDNA samples without validated variations**. gDNA-5 is depicted. Note that narrow deviations seem to be present according to the reciprocity of some dots across all of the tools (e.g. the gain near the p-telomere of chromosome 3), as small aberrations are found naturally in every healthy individual. The reference-free tools hint towards novel larger aberrations (e.g. at chromosome 4, 6 and 17), which are likely false results, due to incorrect normalization.

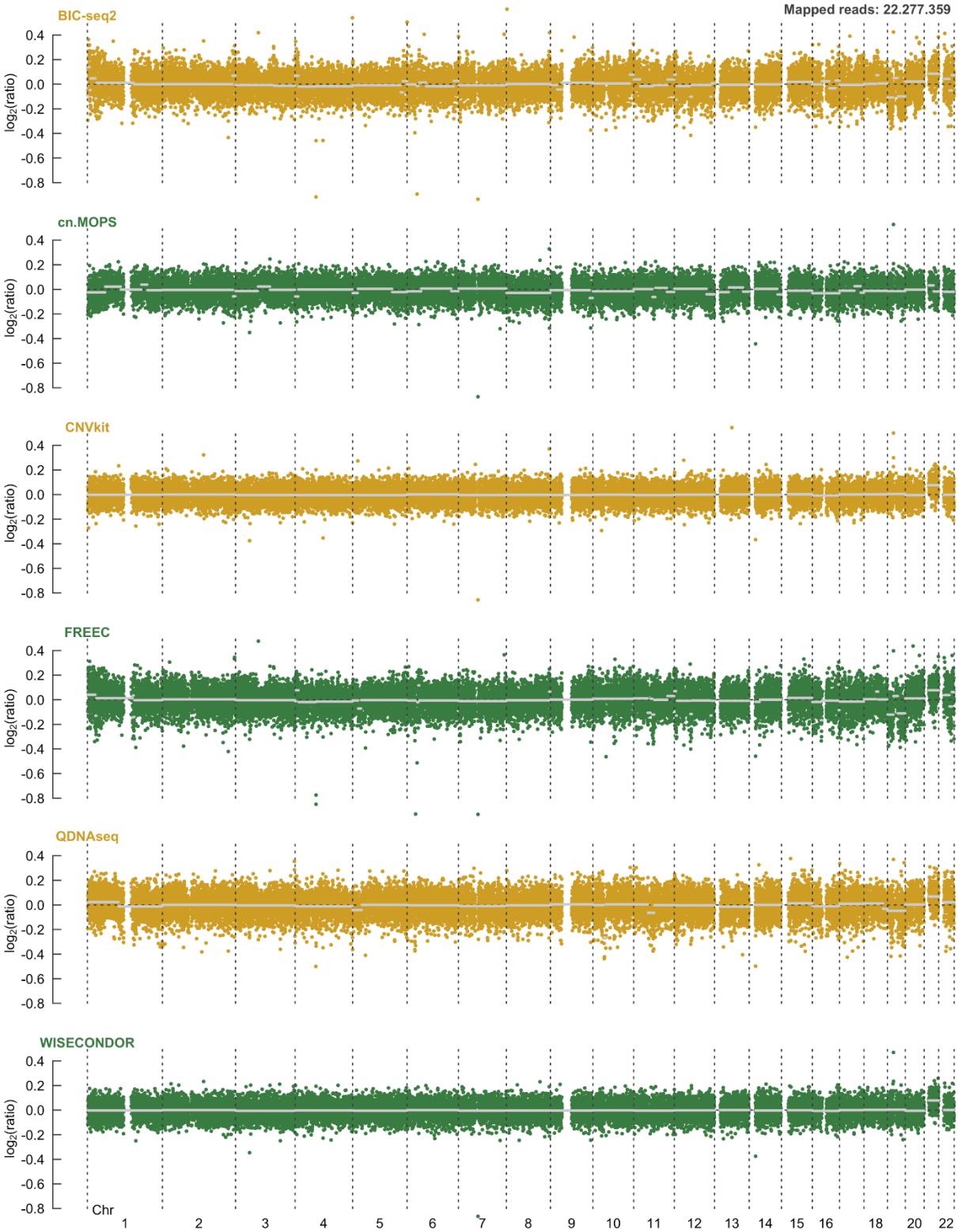

Figure S6. **Typical autosome-wide profiles for a trisomy 21 NIPT sample**.

NIPT-11 is depicted.

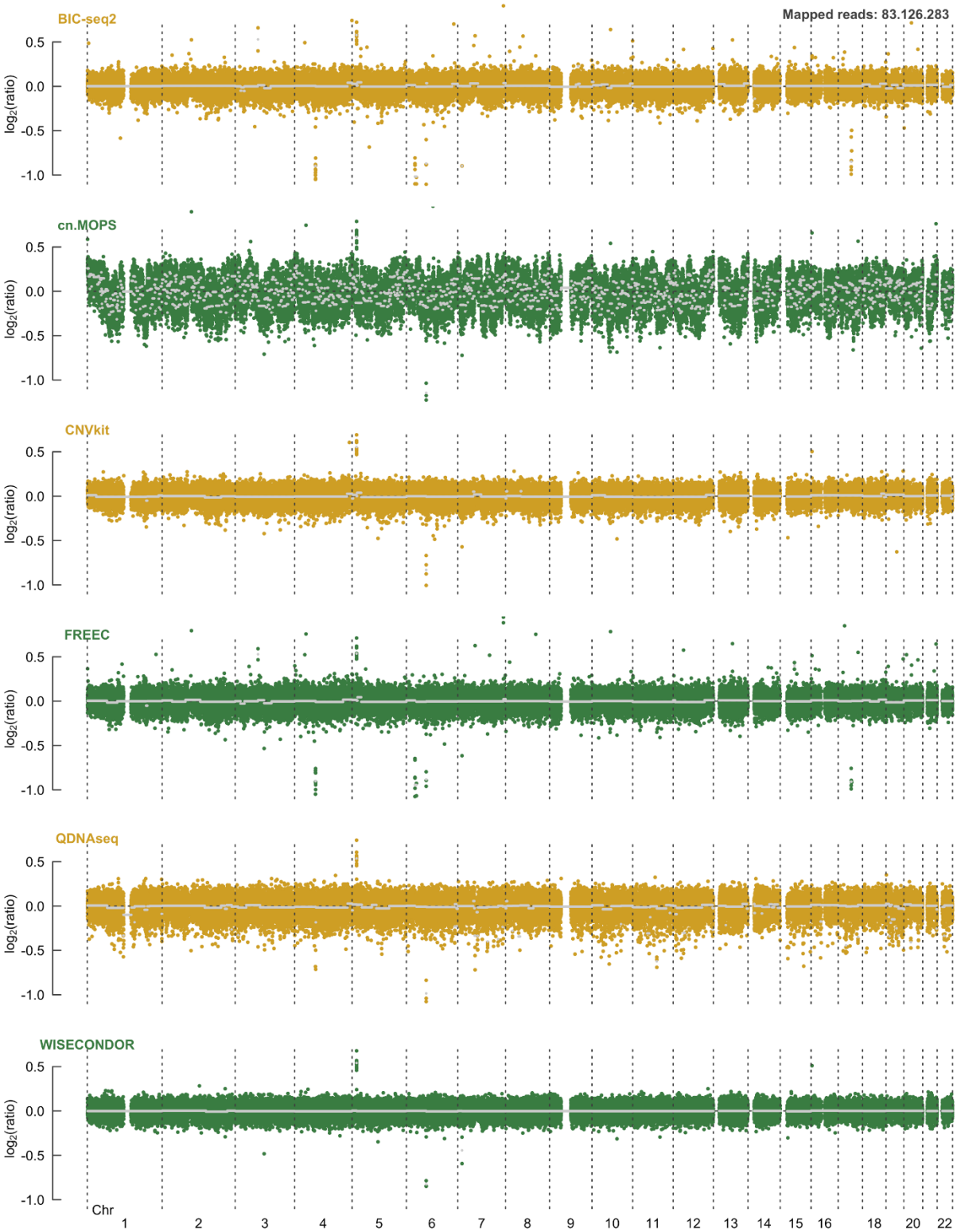

Figure S7. **Typical autosome-wide profiles for gDNA samples with validated variations.** gDNA-13 is depicted. This sample has a validated duplication and deletion at chromosomes 5 and 6, respectively (Table S3).

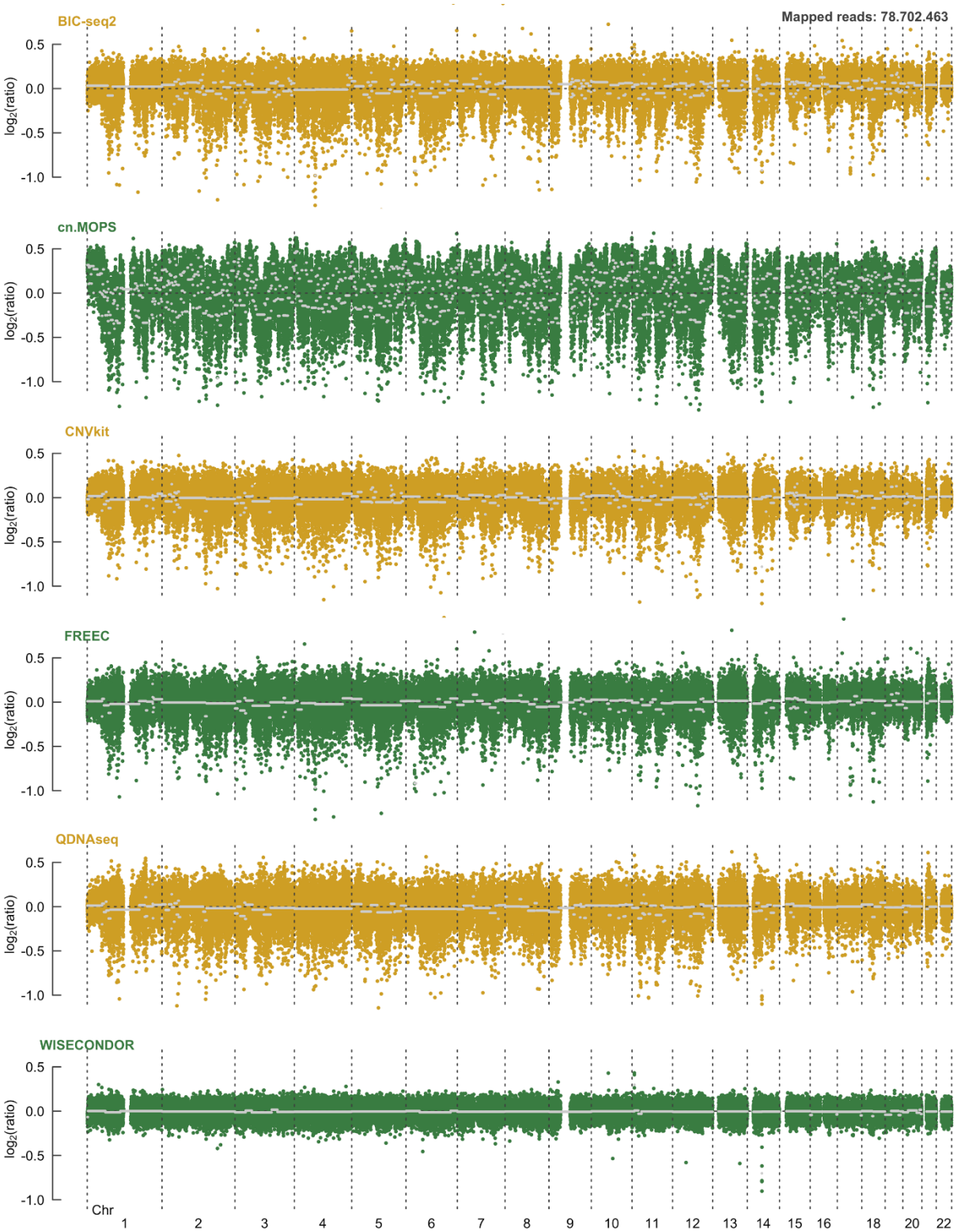

Figure S8. **Autosome-wide profile comparison of problematic sample gDNA-3**.

WISECONDOR is the only tool that seems to correctly normalize this sample.

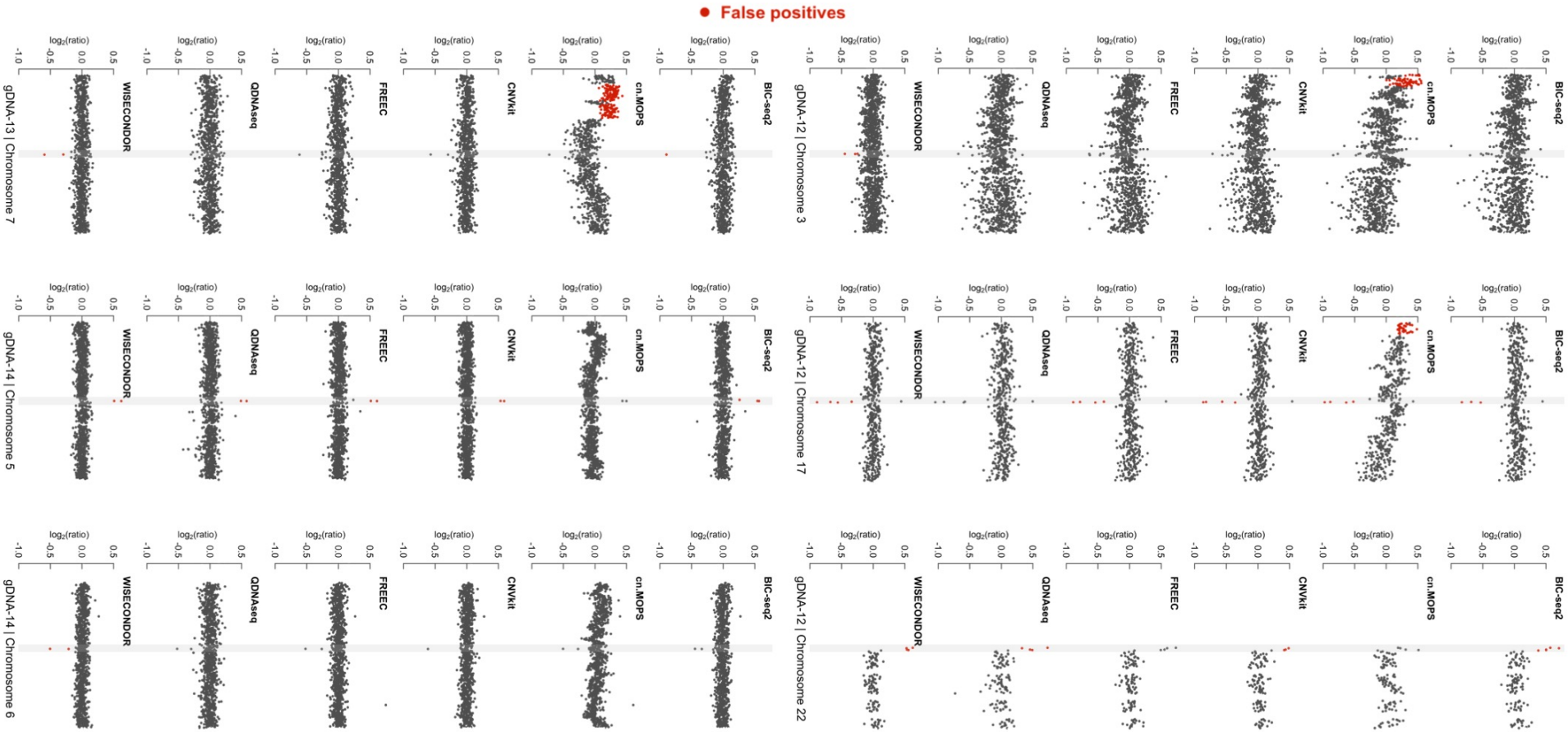

Figure S9. **Claimed false positives by WISECONDOR**.

The chromosomes are only partly visualized. WISECONDOR's false positives (6 in total) are positioned in the middle of each plot, additionally marked by transparent grey bars running across the tools. The reciprocity within these bars hints towards unannotated (mosaic) events rather than false positives. This was not observed in the far majority of false positives returned by other tools, unless for some 'deviations' exclusively returned by the reference-free tools.

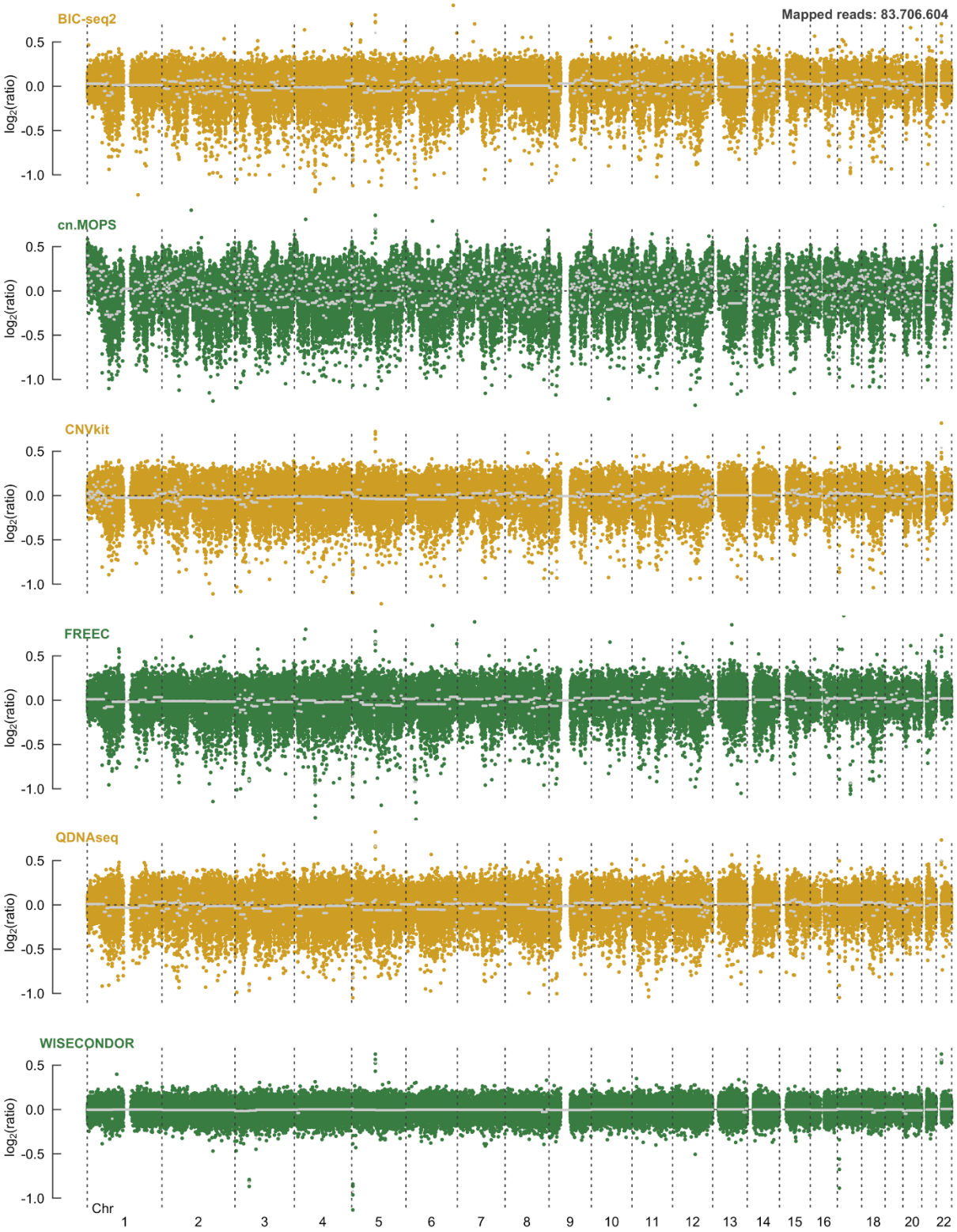

Figure S10. **Autosome-wide comparison of problematic sample gDNA-12.**

This sample has three validated aberrations in chromosomes 3 and 5 (Table S3). WISECONDOR is the only tool that seems to correctly normalize this sample.

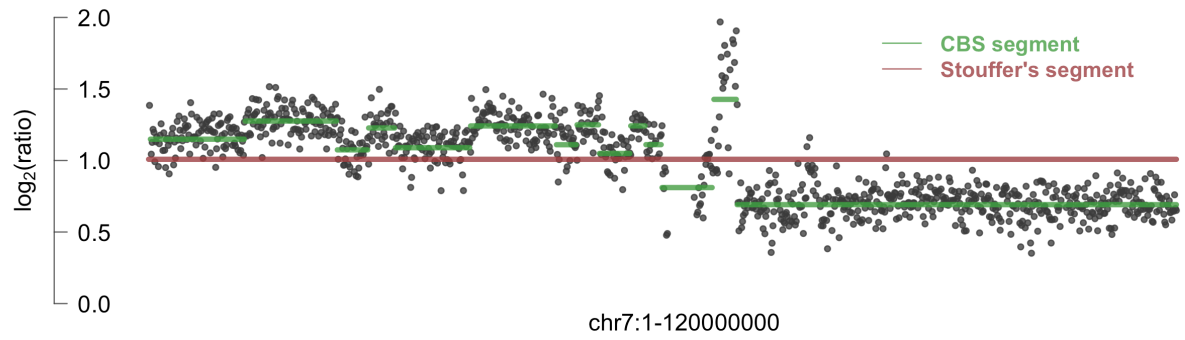

Figure S11. **Segmentation comparison between WISECONDOR and WisecondorX.**
This sample was not included in the validation set. It concerns a liquid biopsy from a patient with phase IV lung carcinoma. Note that chromosome 7, only partly represented, was subjected to numerous consecutive deletions and duplications finally resulting in a large gain with $\log_2$ ratios centered around 1. The WISECONDOR algorithm (red) solely recognizes this segment as an amplification, yet does not further deal with the complex pattern. WisecondorX's CBS approach (green) however does not suffer this shortcoming.

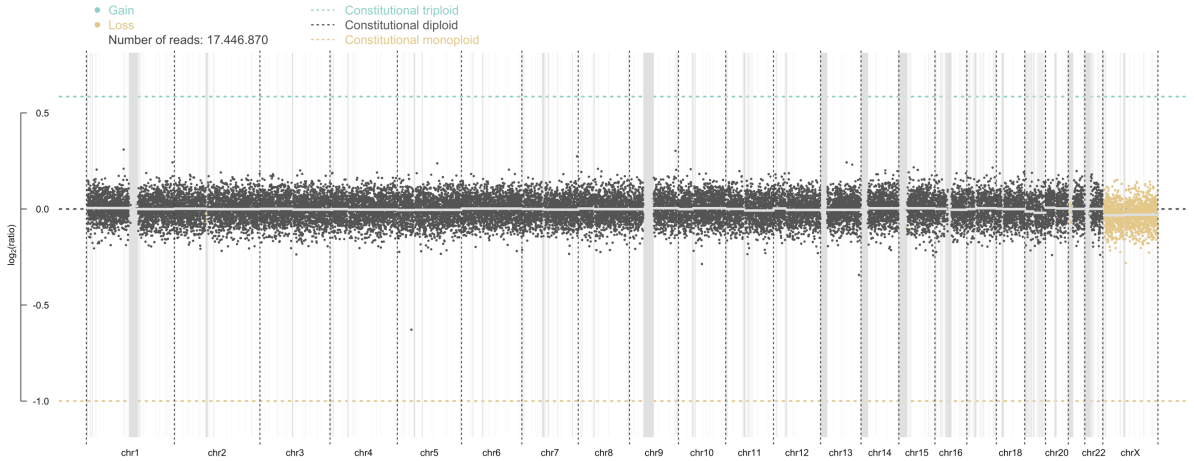

Figure S12. **Part of WisecondorX's genome-wide output for a fetus with monosomy X**.
WisecondorX shows a decreased value for the X-chromosome (NIPT-21), whilst few reads mapped the Y-chromosome (considered noise). Monosomy X was confirmed by chorionic villus sampling.

Tool	Publication	Latest update	Main language	Version	Uses non-paired reference set
BIC-seq2	Xi et al, 2016	?	Perl	v0.7 (norm v0.2.4)	✗
cn.MOPS	Klambauer et al, 2012	2016	R	v1.24.0	✓
CNAnorm	Gusnanto et al, 2012	2017	R		
CNAseg	Ivakhno et al, 2010	?	R		
cnD	Simpson et al, 2010	2012	D		
CNV-seq	Xie et al, 2009	2014	Perl		
CNVeM	Zwang et al, 2013	?	C++		
CNVer	Medvedev et al, 2010	2012	C++		
CNVkit	Talevich et al, 2016	2018	Python	v0.9.3	✓
CNVnator	Abyzov et al, 2011	2017	C++		
ERDS	Zhu et al, 2012	?	Perl		
FREEC	Boeva et al, 2012	2017	R	v11	✗
JointSLM	Magi et al, 2011	2011	R		
QDNAseq	Scheinin et al, 2014	2017	R	v1.14.0	✗
RDXplorer	Zhao et al, 2009	2011	Python		
readDepth	Miller et al, 2011	2016	R		
seqCNA	Mosen-Ansorena et al, 2014	2017	R		
SegSeq	Chiang et al, 2009	2009	MATLAB		
WISECONDOR	Straver et al, 2014	2017	Python	Github master branch	✓

Table S1**. Depth of coverage WGS tools selected for the comparative analysis**.

The final selection, marked in bold, was based on a combination of a tool's latest update, assumed diagnostic ability, popularity and distinct properties. This resulted in three software that require healthy reference samples, and three that do not.

Type	Reference		Cases		
	Samples (n)	Mapped reads (M)	Without known aberrations (n)	With known Aberrations (n)	Analysis at bin size (kb)
NIPT	50	22.8 ± 1.7	10	15	100
gDNA	50	86.5 ± 15.7	10	5	30

Table S2. **Samples used in the comparative study**.

Since some tools require healthy reference samples, fifty were selected per type of analysis. Furthermore, a set of twenty test cases with proven aberrations (Table S3) and a second equally-sized set without variations were selected for the actual comparative analysis, where the bin size was set according to the number of mapped reads in the reference sets – this to obtain similar Gaussian noise levels.

Sample ID	Source	FF (%)	Aberration	Validation
Chromosomal aberrations				
NIPT-11	Fetal	8.2%	Trisomy 21	Confirmed by AC
NIPT-12	Fetal	7.2%	Trisomy 21	Blood redraw followed by NIPT sWGS
NIPT-13	Fetal	12.2%	Trisomy 21	Confirmed on fetal biopsy by sWGS
NIPT-14	Fetal	14.6%	Trisomy 21	Confirmed by AC
NIPT-15	Placental	8.1%	Trisomy 9	Not confirmed by AC
NIPT-15	Fetal	8.1%	Trisomy 21	Confirmed by AC
NIPT-16	Fetal	18.8%	Trisomy 21	Confirmed by CVS & ultrasound irregularity
NIPT-17	Placental	9.5%	Trisomy 14	Not confirmed by AC
NIPT-18	Placental	24.3%	Trisomy 20	Not confirmed by AC
NIPT-19	Placental	9.6%	Trisomy 16	Confirmed mosaic T16 on placental biopsy by sWGS
NIPT-20	Placental	18.1%	Trisomy 3	Not confirmed by AC
NIPT-20	Placental	18.1%	Trisomy 18	Not confirmed by AC
NIPT-21	Fetal	5.1%	Monosomy X	Confirmed by CVS & ultrasound irregularity
Large subchromosomal aberrations				
NIPT-22	Maternal	7.1%	18q11.2 (+)	Confirmed by sWGS on maternal lymphocyte DNA
NIPT-23	Fetal	12%	18p11 (+); isochromosome	Confirmed by AC & postnatal sWGS on lymphocyte DNA
NIPT-24	Fetal	15.8%	5p15.5-5p13.2 (+)	Confirmed by AC
NIPT-25	Fetal	23.4%	22q11.21 (-)	Confirmed by AC
Small subchromosomal aberrations				
gDNA-11	Constitutional	N/A	chr9:26265001-26430000 (-)	Trio analysis – paternal origin
gDNA-11	Constitutional	N/A	chr17:3510001-3555000 (-)	Trio analysis – maternal origin
gDNA-12	Constitutional	N/A	chr3:47550001-47640000 (-)	Trio analysis – *de novo* origin
gDNA-12	Constitutional	N/A	chr5:3675001-3885000 (-)	Trio analysis – maternal origin
gDNA-12	Constitutional	N/A	chr5:78750001-78855000 (+)	Trio analysis – maternal origin
gDNA-13	Constitutional	N/A	chr5:14880001-15195000 (+)	Trio analysis – paternal origin
gDNA-13	Constitutional	N/A	chr6:65805001-65925000 (-)	Trio analysis – maternal origin
gDNA-14	Constitutional	N/A	chr2:44505001-44580000 (-)	Trio analysis – maternal origin
gDNA-14	Constitutional	N/A	chr7:146250001-146685000 (-)	Trio analysis – paternal origin
gDNA-14	Constitutional	N/A	chr12:85455001-85545000 (+)	Trio analysis – paternal origin
gDNA-15	Constitutional	N/A	chr2:112650001-112740000 (-)	Trio analysis – maternal origin

Table S3. **Used case samples with validated aberrations**.

In total, thirteen chromosomal, four large and eleven small subchromosomal variations were included across fifteen NIPT and five gDNA samples with validated aberrations. NIPT validation was mostly performed by amniocentesis and chorionic villus sampling whilst gDNA aberrations were discovered and confirmed by trio analysis. *De novo* annotations could be assigned if they were supported by phenotype. Abbreviations: sWGS = shallow whole-genome sequencing; FF = fetal fraction; AC = amniocentesis; CVS = chorionic villus sampling; (+) = gain; (-) = loss.