The American Journal of Human Genetics, Volume *108*

Supplemental information

Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping

Kornelia Neveling, Tuomo Mantere, Susan Vermeulen, Michiel Oorsprong, Ronald van Beek, Ellen Kater-Baats, Marc Pauper, Guillaume van der Zande, Dominique Smeets, Daniel Olde Weghuis, Marian J.P.L. Stevens-Kroef, and Alexander Hoischen

Supplemental Data

Supplemental Data include 10 figures and 16 tables.

Figure S1. Flow diagram for filtering somatic SVs and CNVs

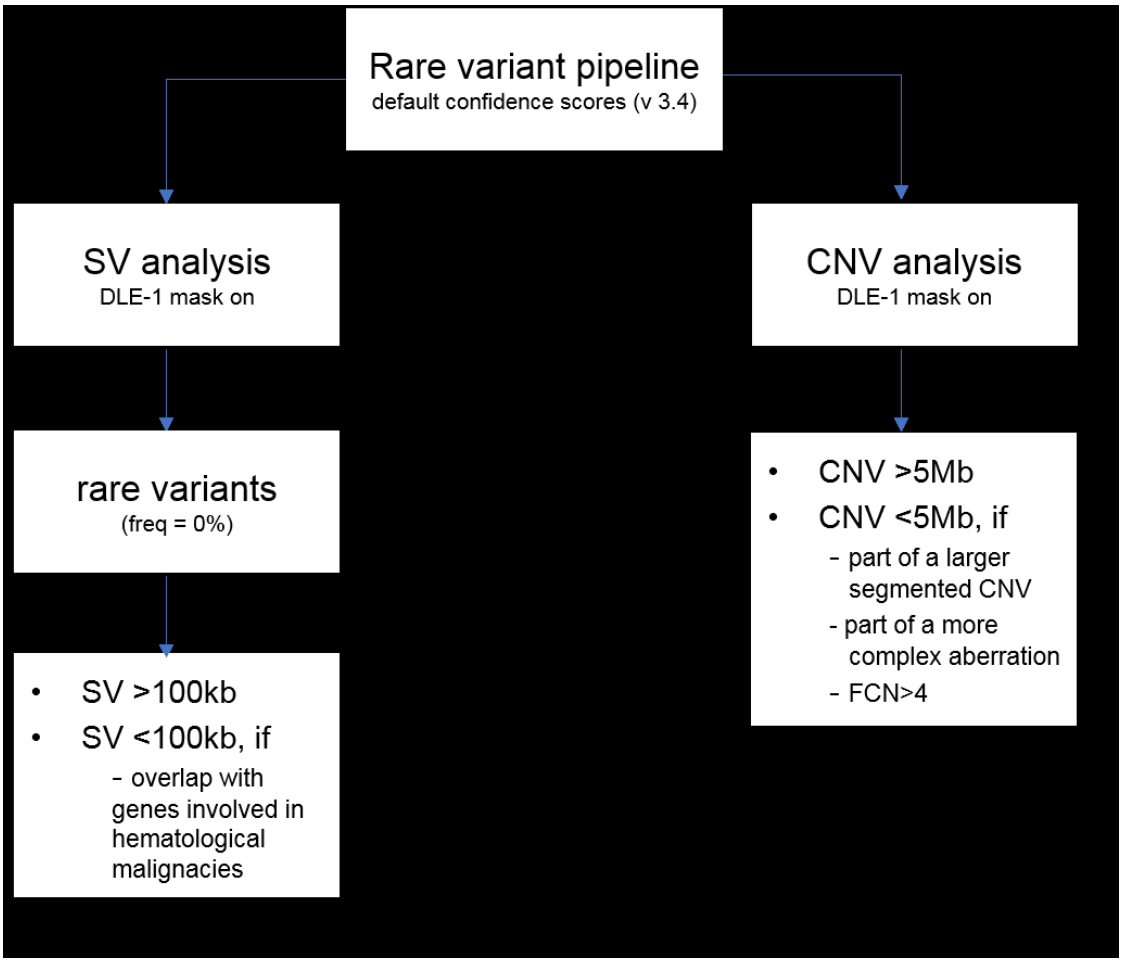

Figure S1. Flow diagram for filtering somatic SVs and CNVs. All data in this study were analyzed using the Rare Variant Pipeline from Bionano Genomics (v.3.4). Data were prefiltered using confidence scores recommended by Bionano Genomics, and variants overlapping the DLE-1 mask region were excluded (DLE-1 mask on). For SVs, also all variants with a frequency >0% in the control cohort were filtered out. For the remaining SVs, all SVs were analyzed when they were >100kb in size. SVs <100kb were analyzed when they overlapped with a gene previously implicated in hematological malignancies (Table S4). For CNVs, the current software version does not allow to filter for frequency. Therefore, only a size threshold was applied, meaning that all CNVs >5Mb were analyzed. Smaller CNVs were only taken into account if they were either part of a fragmented larger CNV, were part of a more complex aberration, for example an unbalanced translocation, or had a fractional copy number (FCN) >4.

Figure S2. Examples of filter adjustment for CNV calling

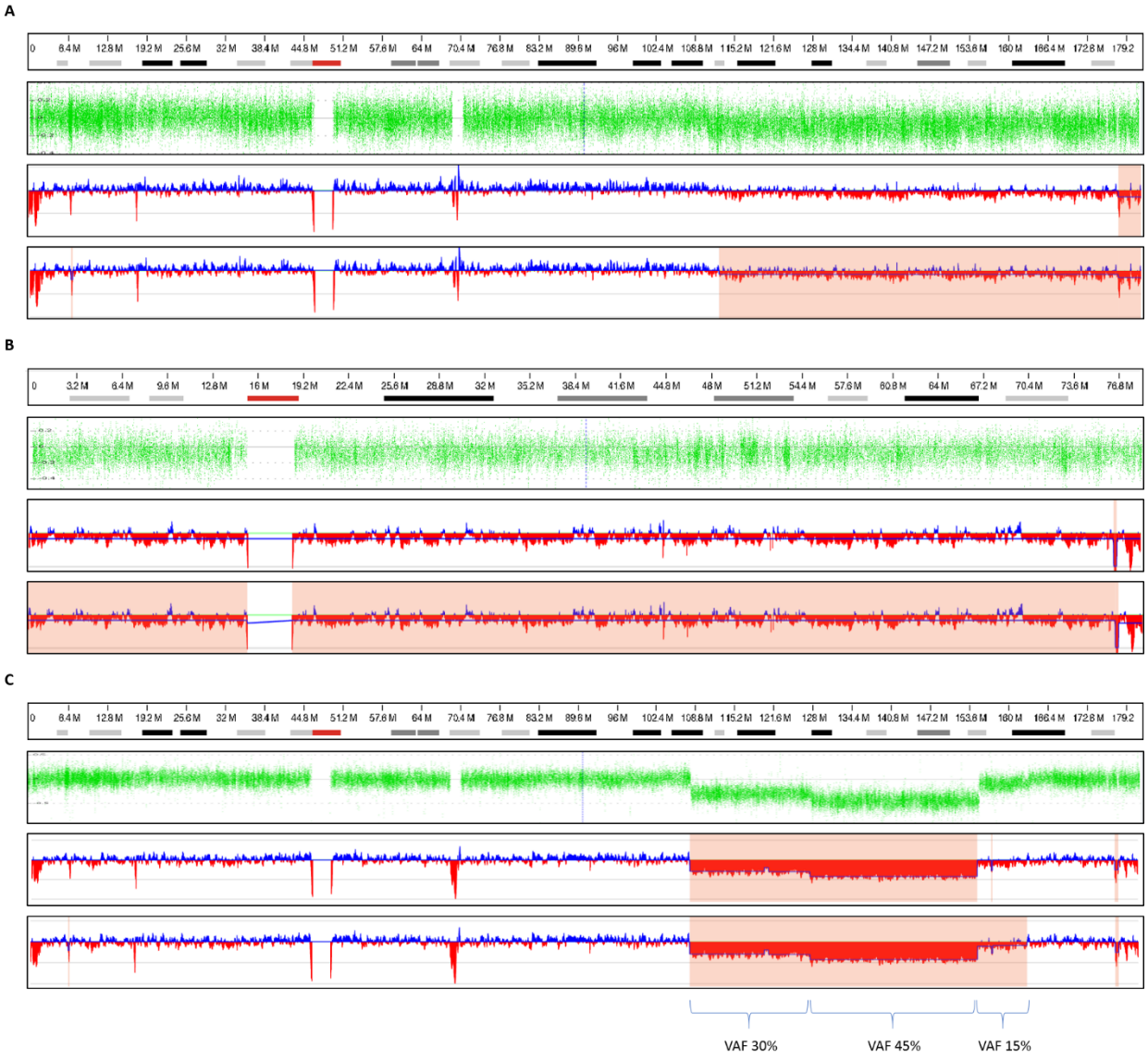

Figure S2. Examples of filter adjustment for CNV calling. For some samples with low VAF, a lower stringency threshold for CNV calling was required. For example A) 5q22.1qter loss (VAF 10%) in sample 47 (complex, MDS): lower stringency threshold shows larger CNV region than higher stringency threshold (red bars). B) Monosomy 18 (VAF 10%) in sample 47 (complex, MDS): whole chromosome 18 loss only seen with lower stringency threshold (red bars). C) Loss of 5q21.3q34 (sample 49, MDS): two different VAFs% (30%,45%) detected by high confidence threshold, VAF of 15% only detected with lower stringency threshold.

Figure S3. Circos plots of all complex cases

Figure S3. Circos plots of all complex cases. A) Explanation of visualization of the circos plot. B) Overview of all circos plots of the complex cases in the study (without size cut-offs).

Figure S4. Venn diagrams: technical comparison between OGM and standard-of-care tests

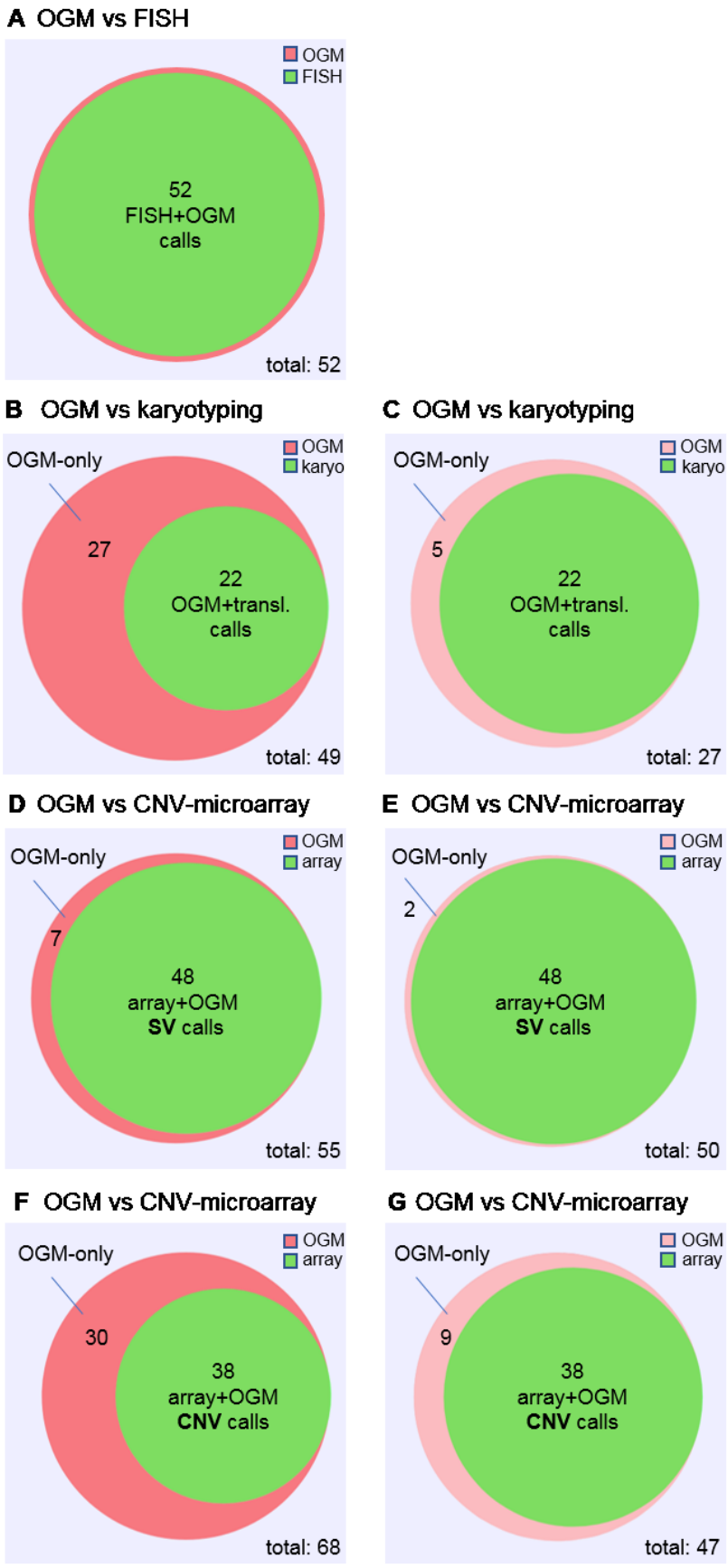

Figure S4. Venn diagrams: technical comparison between OGM and standard-of-care tests
A) Comparison between OGM and FISH. 16 FISH probes were used to analyze 52 loci in N=25 cases (simple and complex). All results were identical between OGM and FISH. B+C) Comparison between OGM and karyotyping (translocations) for N=25 simple cases. B) From a total of 49 OGM translocation calls, karyotyping confirmed 22. Twenty-seven translocation calls were not confirmed by karyotyping. C) Of the 27 not-confirmed translocation calls, 22 were overlapping a DLE-1 mask region and should have been filtered out (Figure S5). Correcting for this leaves five residual unconfirmed calls, of which at least 2 represent likely true positive calls (Figure S6). D+E) Comparison of SV calls provided by OGM with CNV-microarray data in N=21 simple cases. D) Of 55 SVs detected by OGM, 48 matched the CNV-microarray data. The remaining seven SVs that were not detected by CNV-microarray are shown in Figure S7. E) Five of the 7 SVs that were not detected by CNV-microarray overlap with DLE-1 mask regions and should have been filtered out. This leaves only 2 SVs that are not confirmed by CNV-microarray. F+G) Comparison of CNV calls provided by OGM with CNV-microarray data in N=21 simple cases. F) Of 68 CNVs detected by OGM, 38 were also seen in the CNV-microarray data. The additional 30 CNV calls were not detected by CNV-microarray. G) Of those 30 OGM CNVs calls that were not detected by CNV-microarray, 21 originated from 3 low coverage samples (samples 11, 13, 24, ~200x coverage, Table S1). Excluding these CNV calls leaves nine CNV calls that were not seen by CNV-microarray.

Figure S5. Putative false positive translocations overlapping DLE-1 mask regions

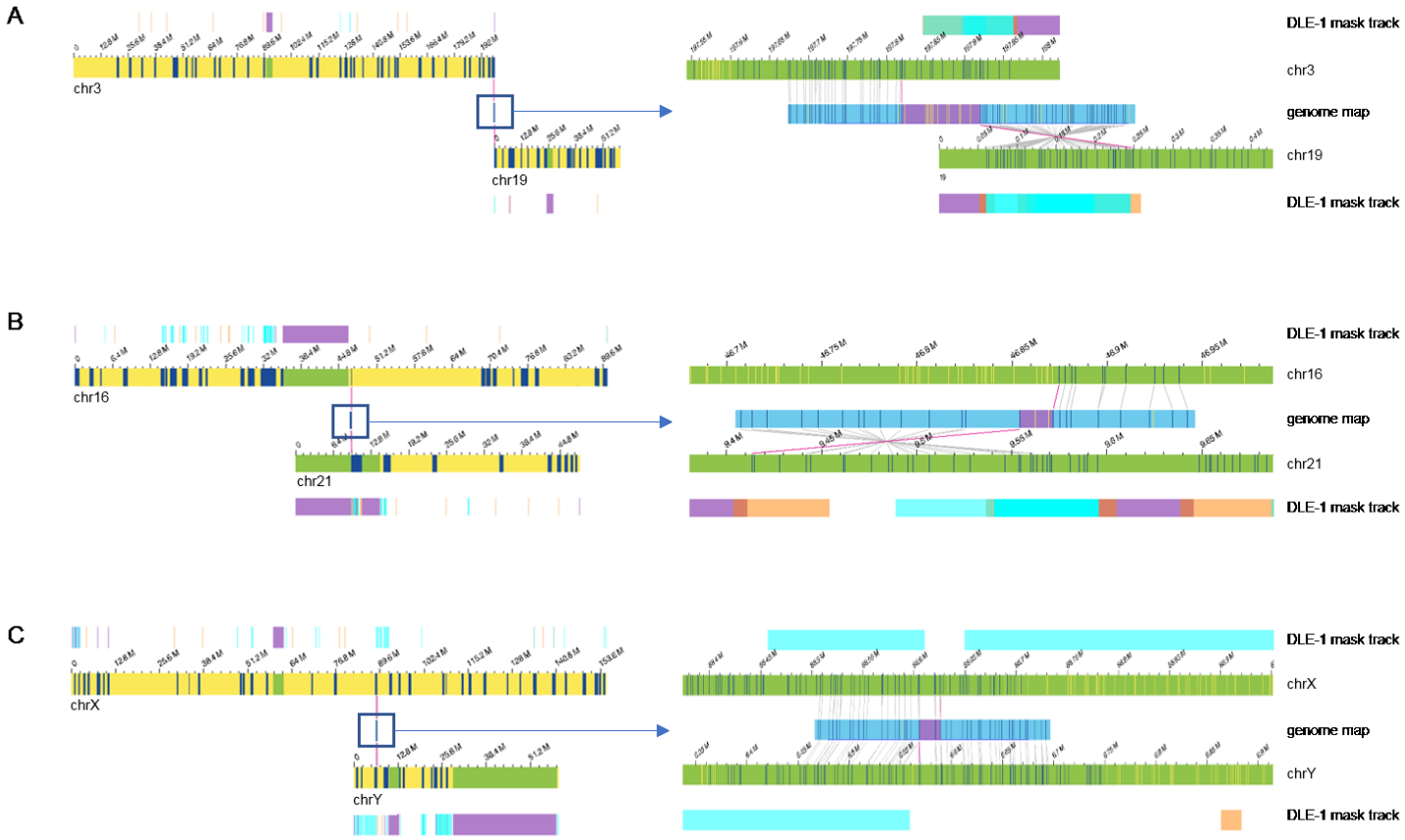

Figure S5. Putative false positive translocations overlapping DLE-1 mask regions. A). Putative false positive translocation t(3;19)(q29;p13.3) in sample 6. The breakpoints of this translocation are in the repeat-rich telomeric regions of both chromosomes. B) Putative false positive translocation t(16;21)(q11.2;p11.2) in sample 14. The breakpoints of this translocation are in the pericentromeric heterochromatin of chromosome 16 and the repeat-rich acrocentric short arm of chromosome 21. C) Putative false positive translocation t(X;Y)(q21.31;p11.2) in sample 17. The breakpoints of this translocation involve SegDup regions on both chromosomes. A-C: The DLE-1 mask track shows SegDup- (mint), common false positive- (orange) and gap- (purple) regions and is provided by Bionano Genomics.

Figure S6. Putative true positive translocations

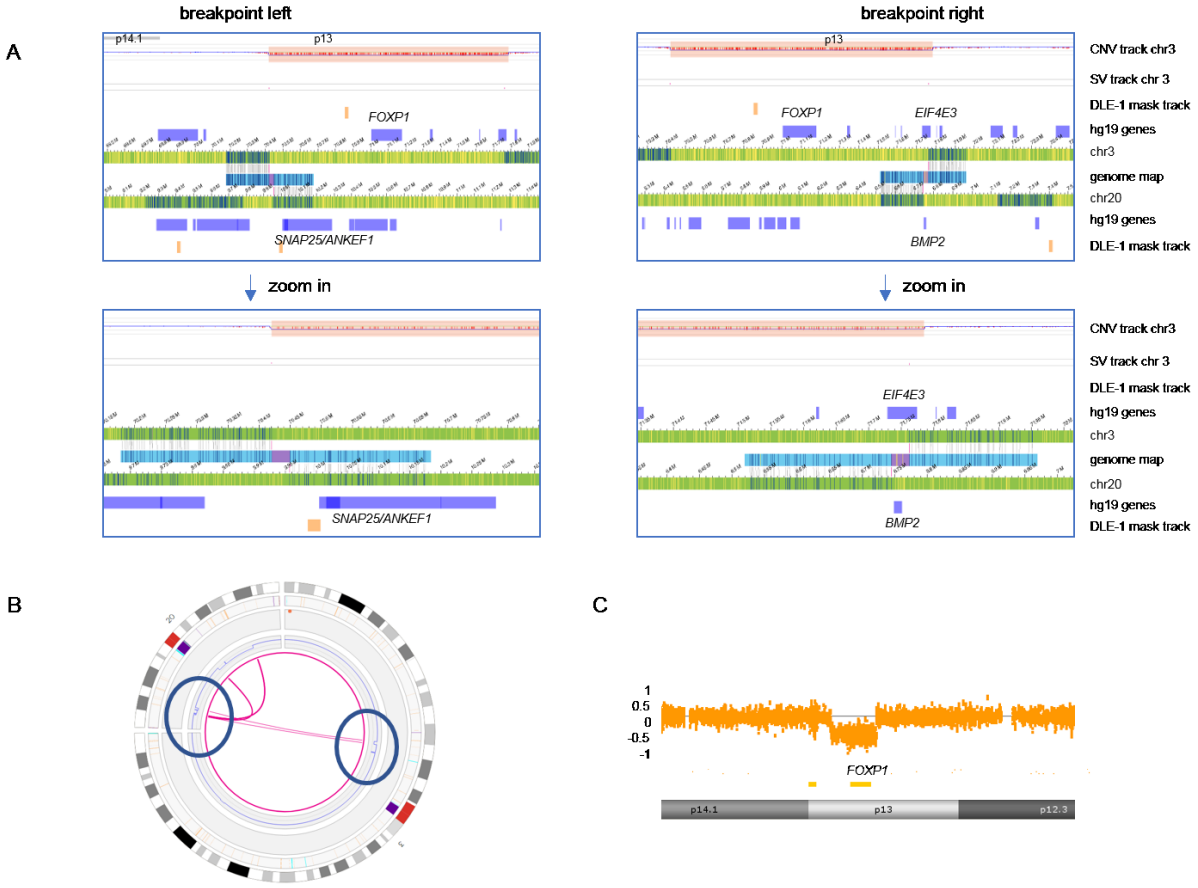

Figure S6. Putative true positive translocations. A) Two different translocations t(3;20)(p13;p12.2) and t(3;20)(p13;p12.3) were detected by OGM in sample 36 that have not been reported by karyotyping before, though only 2 metaphases were analyzed. No repeat-rich regions are involved. Both translocations are flanking a 1.3 Mb deletion on chromosome 3p13 (70429305_71751401) and a 2.5 Mb deletion on chromosome 20p12.3p12.2 (7427312_9942214), and are part of a more complex rearrangement including other parts of chromosome 20 as well. The oncogene *FOXP1* on chromosome 3p13 is deleted by these translocations, and the genes *EIF4E3* and *BMP2* are disrupted. B) Circos plot showing only chromosome 3 and 20. The circos plot shows that the breakpoints of the translocations co-localize with CNV calls of both deletions, underlining our hypothesis that these translocations are true. C) CNV microarray data of the respective sample show that the deletion on 3p13 was also visible by the CNV-microarray, although it was not clinically reported (<5Mb, no AML gene), additional evidence for a true event.

Figure S7. SVs not confirmed by CNV-microarray

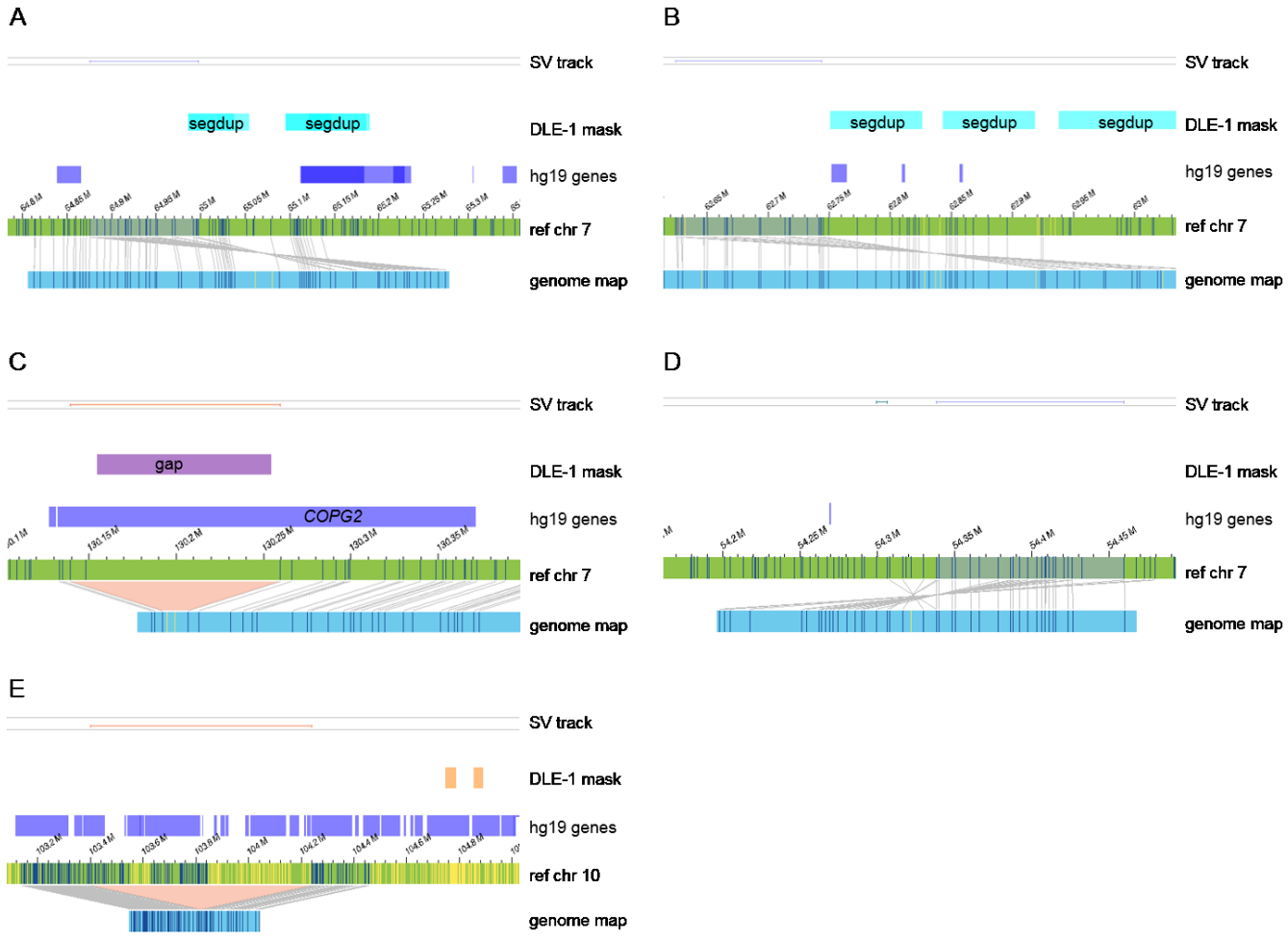

Figure S7: SVs not confirmed by CNV-microarray. A-C) SVs overlapping with DLE-1 mask regions. A) An inverted duplication at 7q11.21 was reported for sample 12. The genome map for this translocations is located in a region with SeqDup. B) Another inverted duplication at 7q11.21 was detected in sample 19, in close proximity to the translocation in sample 12. Also this translocations is located in a region with SegDup C) A recurrent deletion at 7q32.2 was detected in samples 22, 23 and 29. This deletion overlaps with a gap in the reference. D-E) SVs not overlapping with DLE-1 mask regions. D) An inverted duplication at 7p11.2 was detected in sample 27. No repeat-rich regions are involved. E) A deletion at 10q24.32 was detected in sample 12. No repeat-rich regions are involved.

Figure S8. Balanced translocations leading to potential gene disruptions or fusion genes

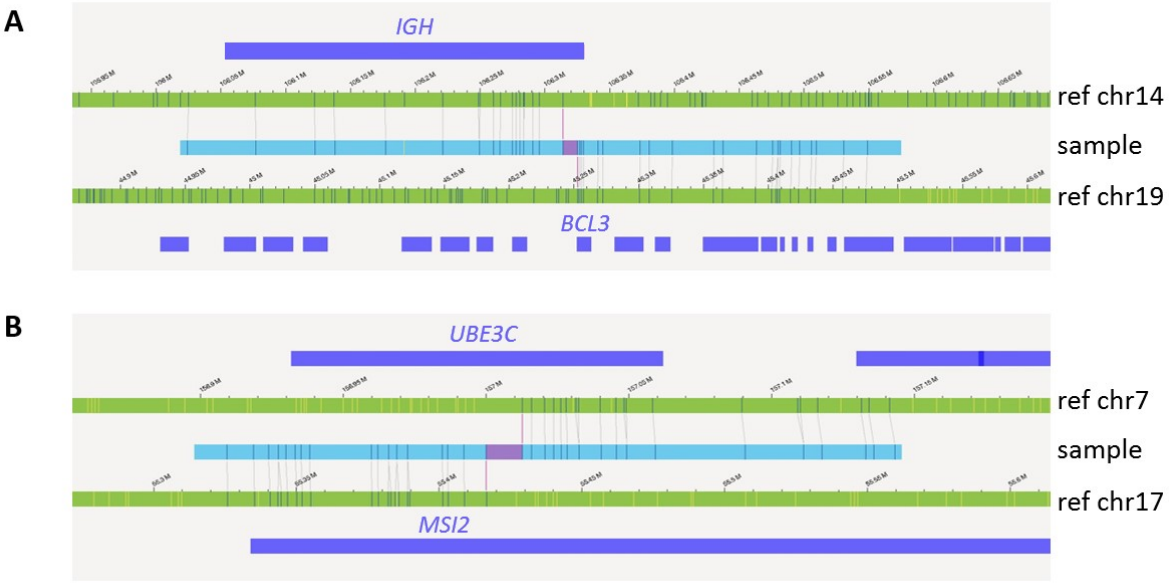

Figure S8. Balanced translocations leading to potential gene disruptions or fusion genes. A) Balanced translocation between *IGH* and *BCL3* that was not reported previously in our CLL sample, as only CNV-microarray data was available (sample 43). B) Balanced translocation between *UBE3C* and *MSI2* in an AML sample (sample 3). Interestingly, this translocation one was the only reported aberration for this AML case, emphasizing the potential importance of the event.

Figure S9. Loss-of-heterozygosity calling with OGM

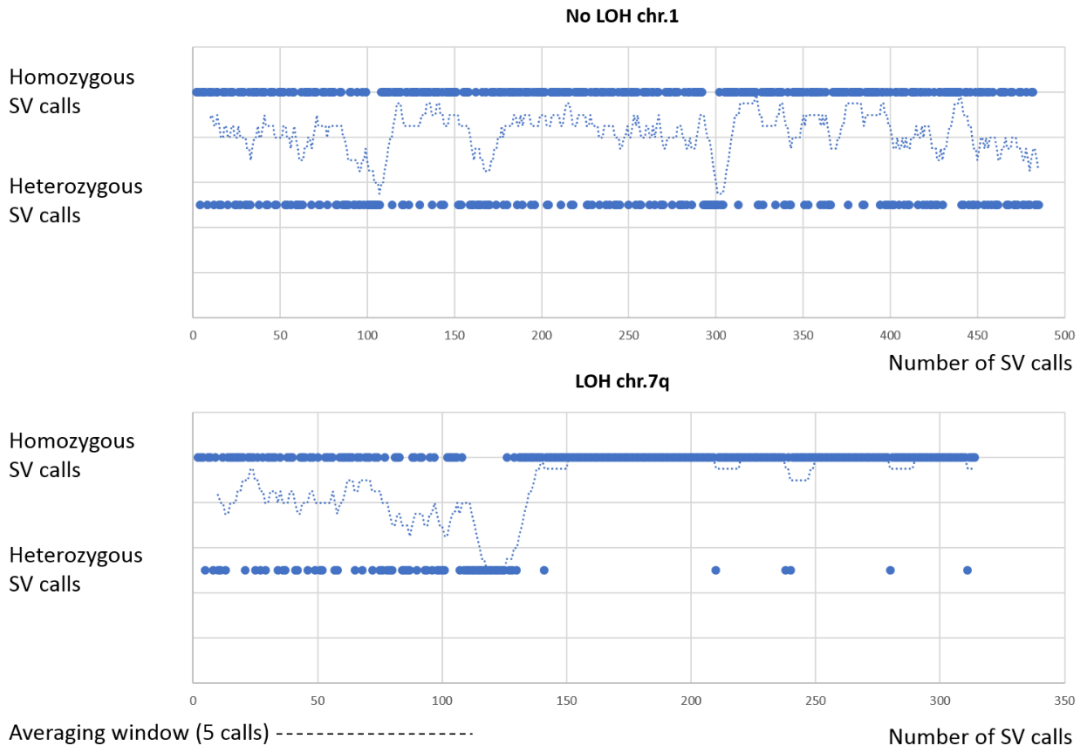

Figure S9. Loss-of-heterozygosity calling with OGM. Example of a previously identified large LOH region (87 Mb) in sample 25 (bottom). Within the homozygous region 88% of SV calls were called as homozygous by the de novo assembly algorithm, in contrast to chromosome 1 without LOH of the same sample (top).

Figure S10. Comparison of CNV-microarray data and OGM

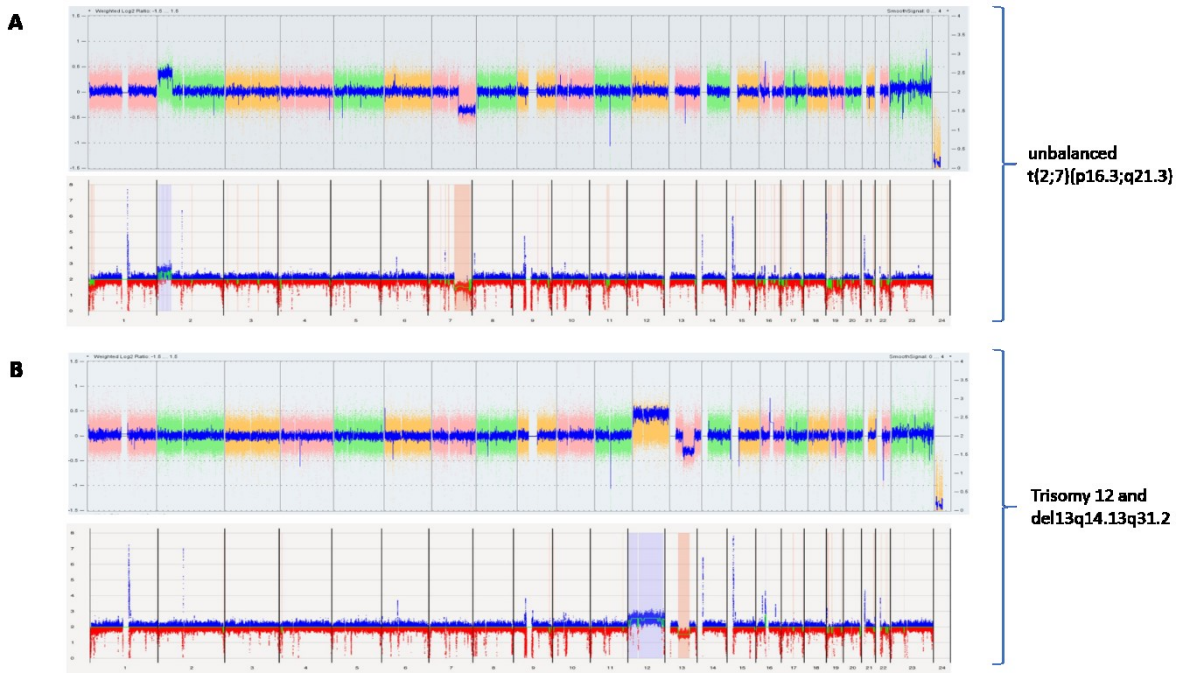

Figure S10. Comparison of CNV-microarray data and OGM. Comparison of CNV-microarray results with the new 'whole genome CNV' visualization that is enabled in the latest Bionano Access software v1.5, for two cases. A) Unbalanced translocation t(2;7)(p16.3;q21.3) in sample 2, shown with array (upper panel) and OGM data (lower panel), B) Trisomy 12 and 13q14.13q31.2 loss in sample 9, shown with array (upper panel) and OGM data (lower panel).