Optical genome mapping enables constitutional chromosomal aberration detection

Tuomo Mantere,[1,2,3,12] Kornelia Neveling,[1,4,12] Céline Pebrel-Richard,[5] Marion Benoist,[6] Guillaume van der Zande,[1] Ellen Kater-Baats,[1] Imane Baatout,[6] Ronald van Beek,[1] Tony Yammine,[7,8] Michiel Oorsprong,[1] Faten Hsoumi,[6] Daniel Olde-Weghuis,[1] Wed Majdali,[6] Susan Vermeulen,[1] Marc Pauper,[1] Aziza Lebbar,[6] Marian Stevens-Kroef,[1] Damien Sanlaville,[7,9] Jean Michel Dupont,[6,10] Dominique Smeets,[1] Alexander Hoischen,[1,2,11,12,*] Caroline Schluth-Bolard,[7,9,12] and Laïla El Khattabi[6,10,12,*]

Summary

Chromosomal aberrations including structural variations (SVs) are a major cause of human genetic diseases. Their detection in clinical routine still relies on standard cytogenetics. Drawbacks of these tests are a very low resolution (karyotyping) and the inability to detect balanced SVs or indicate the genomic localization and orientation of duplicated segments or insertions (copy number variant [CNV] microarrays). Here, we investigated the ability of optical genome mapping (OGM) to detect known constitutional chromosomal aberrations. Ultra-high-molecular-weight DNA was isolated from 85 blood or cultured cells and processed via OGM. A *de novo* genome assembly was performed followed by structural variant and CNV calling and annotation, and results were compared to known aberrations from standard-of-care tests (karyotype, FISH, and/or CNV microarray). In total, we analyzed 99 chromosomal aberrations, including seven aneuploidies, 19 deletions, 20 duplications, 34 translocations, six inversions, two insertions, six isochromosomes, one ring chromosome, and four complex rearrangements. Several of these variants encompass complex regions of the human genome involved in repeat-mediated microdeletion/microduplication syndromes. High-resolution OGM reached 100% concordance compared to standard assays for all aberrations with non-centromeric breakpoints. This proof-of-principle study demonstrates the ability of OGM to detect nearly all types of chromosomal aberrations. We also suggest suited filtering strategies to prioritize clinically relevant aberrations and discuss future improvements. These results highlight the potential for OGM to provide a cost-effective and easy-to-use alternative that would allow comprehensive detection of chromosomal aberrations and structural variants, which could give rise to an era of "next-generation cytogenetics."

Introduction

Structural variants (SVs) play an important role in human diversity and diseases. The emergence of cytogenetic tools, beginning with karyotyping followed by fluorescence *in situ* hybridization (FISH) and copy number variant (CNV) microarrays, allowed for SV detection and thereby significantly contributed to the discovery of disease-associated genes.[1–3] Despite their significant limitations, these techniques remain major components of the routine genetic investigation tools portfolio for constitutional and somatic diseases.

Karyotyping is indicated for diseases where numerical and structural balanced aberrations are highly represented, such as in reproductive disorders; its overall diagnostic rate, however, is well below 10%.[4–7] Indeed, karyotyping has a very low resolution (estimated to be 5–10 Mb on

average), it shows highly variable quality between samples and laboratories, and it is highly dependent on the expertise of technicians and cytogeneticists. This expertise has been decreasing in recent years because of lack of training. Hence, there is strong need for a more robust, high-resolution, and automatable method.

CNV microarrays represent one such robust routine tool that enables the diagnosis of sub-chromosomal CNVs, including clinically relevant microdeletions/microduplications.[1,2,8] Today, this analysis is recommended as a first-tier test for developmental disorders (DDs) with or without multiple congenital anomalies[9,10] because it allows for an improved diagnostic rate reaching 15% to 20% compared to less than 5% with karyotyping.[11,12] However, CNV microarrays are not able to detect mosaicism lower than 5%–20% or balanced chromosomal aberrations, nor are they able to decipher the orientation of duplicated

[1]Department of Human Genetics, Radboud University Medical Center, 6500HB Nijmegen, the Netherlands; [2]Radboud Institute of Medical Life Sciences, Radboud University Medical Center, 6500HB Nijmegen, the Netherlands; [3]Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit and Biocenter Oulu, University of Oulu, 90220 Oulu, Finland; [4]Radboud Institute of Health Sciences, Radboud University Medical Center, 6500HB Nijmegen, the Netherlands; [5]Department of Chromosomal and Molecular Genetics, University Hospital of Clermont-Ferrand, 63003 Clermont-Ferrand, France; [6]Department of Cytogenetics, APHP.centre – Université de Paris, Hôpital Cochin, 75014 Paris, France; [7]Institut Neuromyogène, CNRS UMR 5310, INSERM U1217, Lyon 1 University, 69008 Lyon, France; [8]Unit of Medical Genetics, Saint-Joseph University, 1107 2180 Beyrouth, Lebanon; [9]Department of Genetics, Hospices Civils de Lyon, 69677 Bron, France; [10]Université de Paris, Cochin Institute U1016, INSERM, 75014 Paris, France; [11]Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, 6500HB Nijmegen, the Netherlands
[12]These authors contributed equally
*Correspondence: laila.el-khattabi@aphp.fr (L.E.K.), alexander.hoischen@radboudumc.nl (A.H.)
https://doi.org/10.1016/j.ajhg.2021.05.012.

segments or the location of inserted ones, and their resolution remains restricted to a few kilobases.

Recent breakthroughs in sequencing technologies raised great interest in complementing or replacing cytogenetic tools with an all-in-one genetic test allowing for the detection of both nucleotide variants and structural variants.[13–15] Moreover, short-read sequencing reached reasonable costs and is versatile in terms of protocols (gene panel, whole-exome sequencing, and whole-genome sequencing [WGS]). Although SV detection from exome or genome sequencing continues to improve,[16,17] most comprehensive detection requires a combination of multiple computational analysis tools[18–21] as also established by the 1000 Genomes Project SV Consortium.[22] Indeed, the sequencing-based detection of some SVs remains challenging because of the relatively limited read length and the repetitive nature of sequences at some SV breakpoints because many of them are mediated by non-allelic homologous recombination of repeats. It is expected that long-read sequencing could eventually enable near perfect variant assessment of an individual's genome; so far, technologies and analyses, as well as throughput and prices, prohibit its routine clinical use.[23]

To this end, a tool complementary to sequencing that may replace standard cytogenetics would offer great additional value. Optical genome mapping (OGM) consists of imaging very long linear single DNA molecules (median size > 250 kb) that have been labeled at specific sites. Since its first description,[24] this formerly tedious technique has been updated by Bionano Genomics and is now marketed as "optical mapping for structural variation analysis by whole-genome imaging." This now combines microfluidics, high-resolution microscopy, and automated image analysis to allow for high-throughput whole-genome imaging and its *de novo* assembly.[25,26] Such assemblies were so far mainly used as a scaffold to guide the assembly of next-generation sequencing (NGS) contigs to build reference genomes of several plant and animal species.[27–29] More recently, methods dedicated to the detection of SVs in humans have been developed. Data analysis thereby includes two distinct pipelines: a CNV pipeline that allows for the detection of large unbalanced aberrations (usually > 5 Mb), including aneuploidies, based on normalized molecule coverage, and an SV pipeline that compares the labeling patterns and distances between the constructed genome maps of the studied sample and a given reference. The latter allows for the genome-wide detection of SVs, including insertions, deletions, and duplications as well as inversions and translocations, as small as a few hundred base pairs.

OGM recently proved to allow for efficient detection of a wide range of chromosomal anomalies in leukemia.[30] It has also been used to detect germline SVs in individual research participants[31,32] and samples from the 1000 Genomes Consortium,[22] as well as to unravel population-specific SVs.[33]

The aim of the current proof-of-principle study was to evaluate the ability of OGM to detect simple and complex constitutional chromosomal aberrations of clinical relevance, which had been previously identified by standard-of-care approaches (karyotype, FISH, and/or CNV microarray).

Subjects and methods

Individual selection and sample collection

This multicenter study involved a total of 85 samples from four genetic academic centers from the Netherlands (Radboud University Medical Center [RUMC]) and France (Cochin Hospital in Paris, Hospices Civils in Lyon, and the University Hospital of Clermont-Ferrand). Individuals were referred to one of the inclusion centers for developmental or reproductive disorders. Recommended chromosomal investigations were performed according to the clinical indications. Karyotyping was performed in case of reproductive disorders or family history of a balanced chromosomal anomaly, as well as prenatal diagnosis. CNV microarray, and karyotyping for some samples, was done in case of DDs. In some cases, we performed additional investigations including FISH to characterize an identified anomaly. All samples with a cytogenetically abnormal result were anonymized and processed for OGM according to consent practices, local ethical guidelines, and institutional review board that allows de-identified sample use.

Individuals for whom (1) a chromosomal anomaly was identified by karyotyping, CNV microarray, or FISH and (2) for which there was enough residual blood (EDTA [ethylene-diamine-tetra-acetic acid] or heparin) or cultured cells available after routine testing were included after anonymization. Blood samples for ultra-high-molecular-weight (UHMW) genomic DNA (gDNA) extraction were stored at −20°C for a maximum of 1 month and at −80°C for longer-term storage. In addition, several individuals with known aberrations had residual material other than blood, including eight amniotic fluid cell lines, four chorionic villi cell lines, and eight lymphoblastoid cell lines, which were all generated from primary cultures according to standard diagnostic procedures.

Classical cytogenetics

Karyotyping, FISH, and CNV microarrays were all performed prior to this study according to standard procedures of the diagnostic laboratories.

In brief, karyotyping was performed according to previously described standard protocols.[34] Chromosomal abnormalities were described according to the International System for Human Cytogenetic Nomenclature (ISCN, 2020). FISH was performed on standard chromosome slides according to the manufacturer's instructions (Vysis, Abbott, USA) or with isolated bacterial artificial chromosome (BAC) clones as FISH probes following standard procedures. CNV microarray was performed with the Agilent SurePrint G3 ISCA v2 comparative genomic hybridization (CGH) 8x60K or SurePrint G3 Human CGH Microarray 4x180K (Agilent Technologies, Santa Clara, CA, USA) or with the Affymetrix Cytoscan HD Array (Thermo Fisher Scientific, Waltham, MA, USA). Genome coordinates were provided according to hg19/GRCh37 human reference genome.

Ultra-high-molecular-weight DNA isolation, DNA labeling, and data collection for optical genome mapping

For each individual, UHMW gDNA was isolated from 400 μL of whole peripheral blood (EDTA or heparin) or 1–1.5 million cultured cells (lymphoblastoid cells, amniotic cells, or chorionic villi cells) via the SP Blood and Cell Culture DNA Isolation Kit

according to manufacturers' instructions (Bionano Genomics, San Diego, CA, USA). Briefly, we treated cells with lysis-and-binding buffer (LBB) to release gDNA that was bound to a nanobind disk, washed, and eluted in the provided elution buffer.

UHMW gDNA molecules were labeled with the DLS (Direct Label and Stain) DNA Labeling Kit (Bionano Genomics, San Diego, CA, USA). We used Direct Label Enzyme (DLE-1) and DL-green fluorophores to label 750 ng of gDNA. After a wash-out of the DL-green fluorophores excess, DNA backbone was counterstained overnight before quantitation and visualization on a Saphyr instrument.

Labeled UHMW gDNA was loaded on a Saphyr chip for linearization and imaging on the Saphyr instrument (Bionano Genomics, San Diego, CA, USA) (Figure S1).

De novo assembly and structural variant calling

The *de novo* assembly and variant annotation pipeline were executed with Bionano Solve software v.3.4 or v.3.5. Results were analyzed through two distinct pipelines: a CNV pipeline that allows for the detection of large unbalanced aberrations based on normalized molecule coverage and an SV pipeline that compares the labeling patterns between the constructed sample genome maps and a reference genome map. Reporting and direct visualization of structural variants were performed with Bionano Access software v.1.4.3 or v.1.5.1. The following filtering thresholds were applied: hg19 DLE-1 SV mask was turned on (this filter is intended to mask common SV regions and highly repetitive parts of the human genome such as segmental duplications), confidence values for insertion/deletion = 0, inversion = 0.01, duplications = −1, translocation = 0, and CNV = 0.99. These filter settings have been optimized with data from samples 1–42 following a first analysis of OGM results with prior knowledge of the searched aberrations.

SV calls were compared to an OGM dataset of 204 human population control samples from apparently healthy individuals (provided by Bionano Genomics and previously published)[33] to filter out common SVs and potential artifacts (both technical and reference-genome related) (Figure S1). We retained only rare SVs that were not detected in any of the population control samples.

The software represents the results from both pipelines in a circos plot, which allows for an easy overview of the detected variants at a glance (Figure S2). Of note, the software calls duplications that are smaller than 30 kb "insertions" because the label density may not be informative enough to exactly determine the origin of the inserted material. Inversions involving segments of 5 Mb or larger are called "intra-chromosomal translocations."

Data analysis, variant filtering, and comparisons

All OGM results were analyzed genome wide for all samples irrespective of the individual's chromosomal status (Table S1). After optimization of filter settings with samples 1–42, results from samples 43–85 were analyzed in a blinded fashion. We subsequently compared SVs and CNVs of all cases (samples 1–85) detected by OGM to the clinically relevant aberrations previously identified by standard-of-care techniques (karyotype, FISH, and/or CNV microarray). In the few samples that initially showed negative results, OGM data were re-analyzed with less stringent filter settings (lowering the CNV confidence score to 0.95 or turning off the DLE-1 mask), as marked in Table S1.

In order to optimize the number of calls per sample that would require individual and clinical interpretation, we applied additional filtering steps after analyses with both SV and CNV tools: (1) a size cutoff of 20 kb was applied for calls derived from the SV tool (similar to the commonly applied size cutoff used for clinical CNV microarrays), and (2) a more stringent fractional copy number (FCN)-based filtering (including only deletions with FCN < 1.2 and duplications with FCN > 2.8) was applied for calls obtained with the CNV tool (Figure S1).

The list of aberrations could also be checked for the overlap with a coding gene. This overlap with genes was defined following the default settings as given by the manufacturer, which requires 1 bp overlap after applying a 12 kb buffer corresponding to average label distance ± standard deviation around each gene. The gene list derives from the UCSC gene track of known canonical transcripts.

A result was considered concordant when at least one of the Bionano Solve pipelines, SV or CNV, either after the primary analysis or a re-analysis, correctly detected the aberration. Aberrations with breakpoints in the (peri-)centromeric regions of any chromosome or p arm of acrocentric chromosomes were beyond the scope of this study because of the lack of a reference map in those regions.

Results

Population description

All 85 samples included in this study were previously analyzed by karyotyping, FISH, and/or CNV microarray according to the reason for referral and the respective international recommendations (Figure 1, Table S1). Reasons for referral included developmental delay, including autism spectrum disorders or intellectual disability, associated or not with congenital malformations (49 individuals, 57.6%); reproductive disorders (15 individuals, 17.6%); familial history of chromosomal aberration (12 individuals, 14.1%); and abnormal prenatal screening or ultrasound results (nine individuals, 10.6%). These samples exhibited a total of 99 chromosomal aberrations with 11 different types of aberrations from the previous standard diagnostics tests, summarized in Figure 1. Additionally, nine known aberrations in this cohort were beyond the scope of this study and were therefore excluded from further analyses, as explained in the subjects and methods section.

Results of optical genome mapping with Bionano genome imaging

Bionano genome imaging generated on average 655 Gbp of data per sample (853 Gbp for samples processed in Nijmegen, aiming at ~200× genome coverage, and 463 Gbp for samples processed in France, aiming at ≥80× genome coverage per sample). The average N50 molecule length (≥150 kbp) was 267 kbp, meaning that at least half of the genome was covered by molecules larger than 267 kbp on average after discarding molecules smaller than 150 kbp, following manufacturers' recommendations. Label density was 15.1 labels/100 kbp. This resulted in an average map rate of 76.8% and an effective coverage of 152× (192× for Radboudumc samples and 114× for French samples) (Table S2). The difference in coverage between the two cohorts was expected because samples from the Radboudumc were run with a coverage beyond the
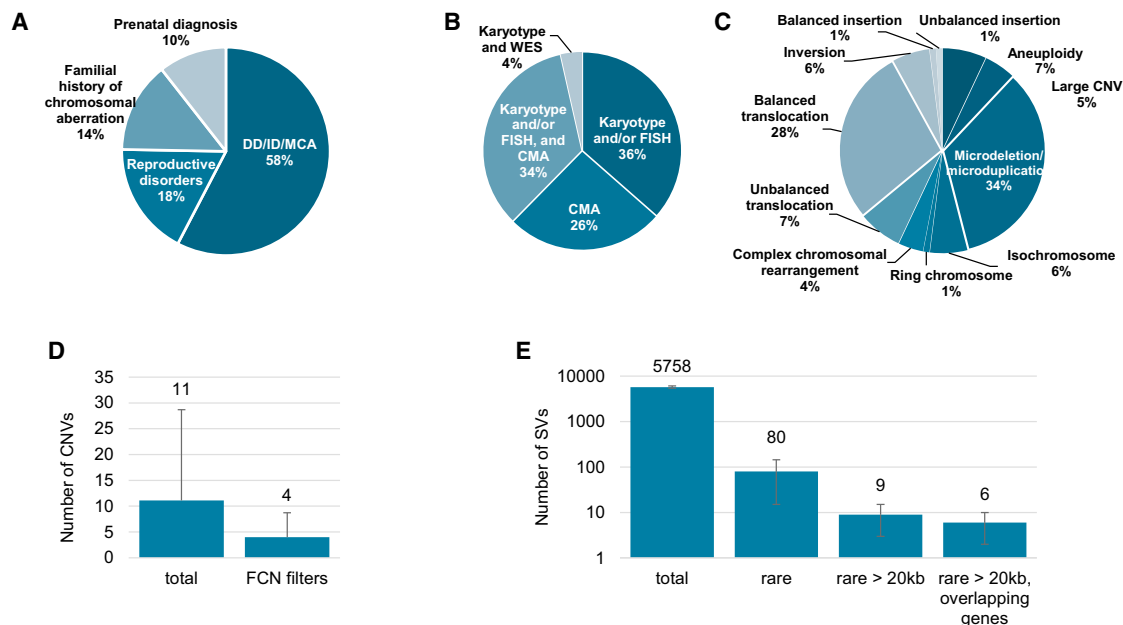
Figure 1. Description of the study population and the detected aberrations
(A) Main reason for referral.
(B) Distribution of the different cytogenetic and molecular tests used for diagnosis.
(C) Distribution of chromosomal aberrations as assessed by standard of care genetic investigations.
(D) CNV detection and filtering. Average number of CNVs detected per sample before and after applying an FCN filtering (<1.2 for deletions and >2.8 for duplications).
(E) SV detection and filtering. Average number of SVs detected per sample before and after applying sequential filtering steps: rare variants only (not found in a control database including 204 human samples), rare variants larger than 20 kb, and rare variants larger than 20 kb and that overlap with genes. Error bars in (D) and (E) represent standard deviations.

recommended minimum. However, this did not affect the molecule size nor labeling density (Table S2), and importantly, it did not affect the overall number of called SVs or CNVs per sample (Table S3).

Structural variant calling identified on average 5,758 (±344) SVs per sample, the vast majority of which corresponded to insertions and deletions (with an average of 4,127 [±239] and 1,549 [±108], respectively). Filtering out events that were present in a database comprising 204 population control samples resulted in an average of 80 (±65) rare SVs per sample (Figure 1, Table S3). These numbers were further reduced to 9 (±6) rare SVs per sample when we applied a lower size cutoff of 20 kb, similar to common practice in diagnostic CNV microarray workflows,[35] in order to reduce the "search space" per sample. Only 6 (±4) of those rare SVs that were >20 kb were overlapping with genes. Events that were prone to artifacts, e.g., due to mis-assemblies at low-copy repeat regions, are mostly masked so that they do not contribute to the call set (this filtering is part of the "mask filter," which we apply systematically in the primary analysis of a sample, described in the subjects and methods section). A few SV calls may erroneously not be filtered out in spite of applying the "mask filter" (Figure S3). However, these events can be easily identified and excluded after a rapid visual inspection because they are systematic and often involve heterochromatic regions of, e.g., chromosome 1 and chromosome 9 (see Figure S3).

The detection of large CNVs was performed with a separate coverage-depth-based algorithm that is included in the *de novo* assembly and variant calling pipeline (the CNV pipeline).[36,37] This analysis resulted in an average of 11 (±18) CNVs per sample without applying any threshold cutoffs (Table S3). We have tested different thresholds of size and FCN. When applying an FCN threshold of >1.2 and <2.8 for autosomal deletions and duplications, respectively, these numbers drop to 4 (±5) CNVs per sample without losing sensitivity because all known aberrations with clinical relevance were still detected (Tables S3 and S4). Of note, large CNVs are often segmented into multiple calls (Figure S4), which artificially increases the average number of CNVs. This segmentation may be due to gaps in the reference genome and to repeat structures, such as segmental duplications, that disrupt contiguous calls. This is also observed with other CNV calling tools in CNV microarrays and WGS results.[38]

Detection of diagnostically reported aberrations with optical genome imaging

All diagnostically reported aberrations in our study cohort were detected correctly either by SV or CNV calling and several aberrations were identified by both algorithms, reaching a 100% concordance for OGM with the previous diagnostic test results (Table S1). The contribution of each pipeline to the detection of aberrations according to their
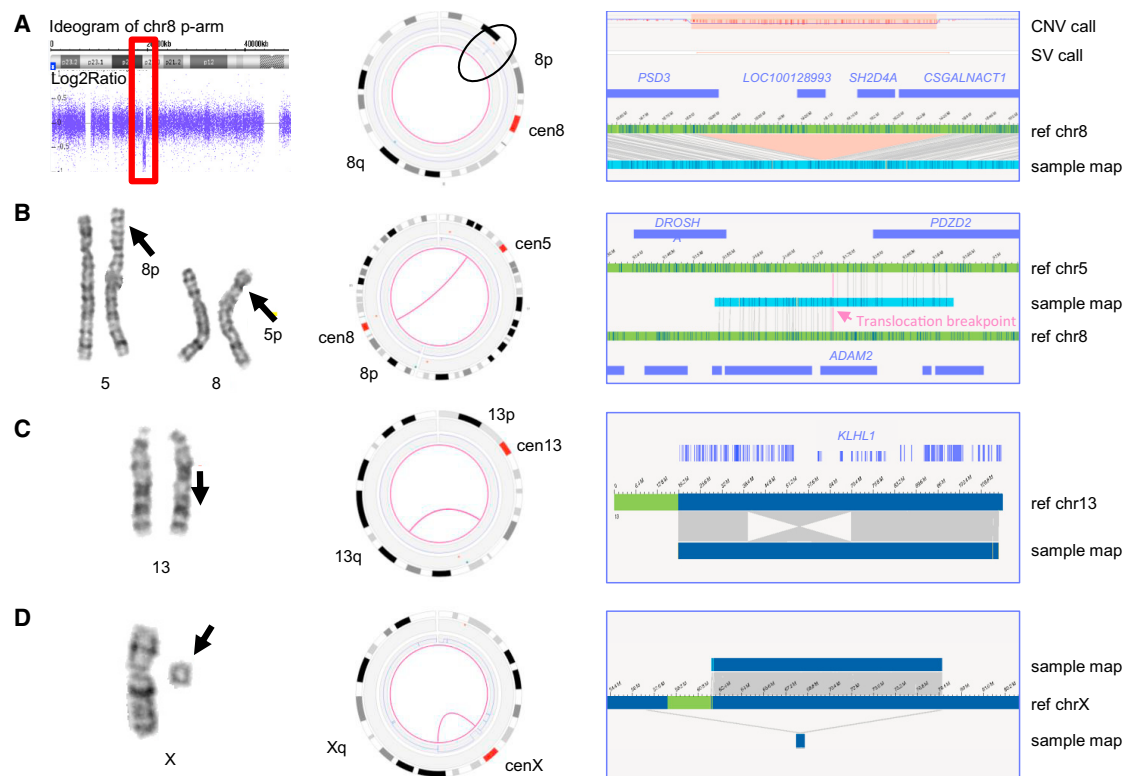
Figure 2. Results from representative examples of different chromosomal aberrations

(A) Sample 1. Left: CNV microarray data showing an 8p22p21.3 deletion (18,825,888–19,364,764 bp). Middle: genome-wide circos plot showing chromosome 8. The CNV profile represented by the blue line shows the deletion (black circle). Right: sample genome map against reference, showing the deletion and affected genes (hg19).

(B) Sample 18. Left: karyogram showing a reciprocal translocation between chromosomes 5 and 8, t(5;8)(p13.1;p11.2). Middle: circos plot showing chromosomes 5 and 8 connected with a pink line, representing the translocation. Right: genome map, of which the left part maps to chromosome 8 and the right part to chromosome 5.

(C) Sample 15. Left: karyogram showing an inversion on chromosome 13 (red arrow). Middle: circos plot showing the inversion represented by a pink line connecting two distal regions on chromosome 13. Right: genome map that is partly inverted when compared to the reference. One of the breakpoints is most likely interrupting *KLHL1*.

(D) Sample 39. Left: karyogram showing one chromosome X and a small marker. Middle: circos plot of chromosome X. The CNV plot in blue shows a large loss of material on most of the p- and q-arms, and the pink line connecting the distal regions of the non-deleted segment indicates the presence of a ring chromosome. Right: different genome maps (dark blue bars on top and below the reference) indicating the presence of the ring chromosome. The assembled genome map below the reference shows that the left part (light green bar) maps to a region upstream of the centromere, whereas the right part (light blue bar) maps to a region downstream of the centromere.

type is summarized in Table S5. Both pipelines complement each other and allow reaching 100% detection rate. For 5/85 samples (34, 42, 50, 76, and 81), however, we needed to adapt filter settings to detect the expected aberrations (see Table S1). Adaptation included setting the confidence value for CNVs to 0.95 (samples 34, 42, and 76) and turning off the SV DLE-1 mask filter (samples 50 and 81). Because some SVs may be mediated by repeat sequences, this masking filter may require refinement. These less stringent filters yielded two additional SV calls for samples 50 and 81 and, respectively, 15, 34, and 4 CNV calls for samples 34, 42, and 76 prior to any additional filtering steps (e.g., size cutoff, overlapping with genes, visual inspection of artifacts).

The 99 identified aberrations included seven aneuploidies, 19 deletions, 20 duplications, 34 translocations, six inversions, two insertions, six isochromosomes, and one ring chromosome (Figure 1). In addition, four of our

samples showed complex chromosomal rearrangements, defined as cases where aberrations involve three or more chromosomes or when at least four SVs are detected on the same chromosome. Graphical representations of different types of chromosomal aberrations are depicted in Figure 2, Figure S5, and Figure S6. Detection capabilities and limitations of OGM compared to standard-of-care cytogenetics (CNV microarray and karyotyping) are summarized in Table S6.

Aneuploidies, partial aneuploidies, and large CNVs

Our study cohort included seven full aneuploidy samples, including three XXY, two monosomy X, one trisomy 14, and one trisomy 21 (the two latter ones were detected in prenatal samples and were mediated by Robertsonian translocations). In addition, four mosaic monosomy X samples were included (Table S1). All aneuploidies of the autosomes were called correctly with the applied algorithms, whereas
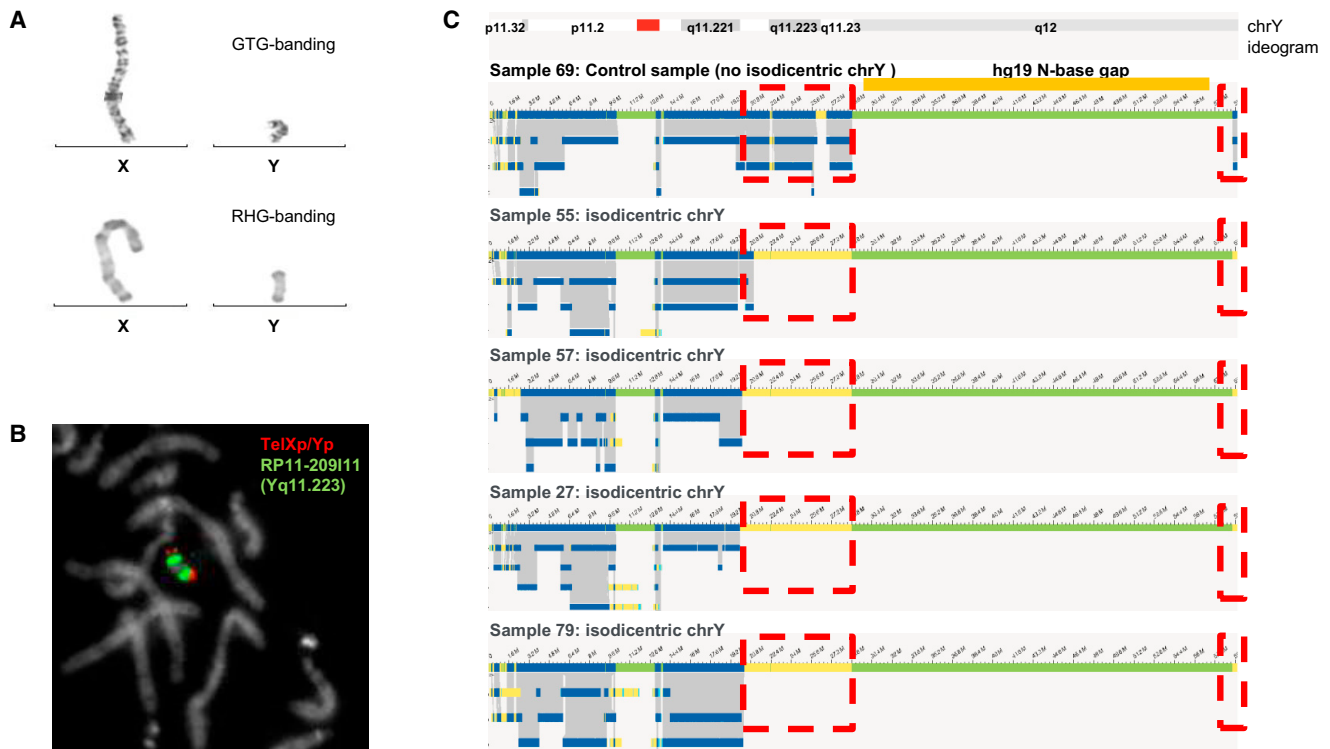
Figure 3. Isodicentric Y chromosomes show specific assembly map patterns
(A) G-banding with trypsin-Giemsa (GTG) and reverse banding using heat and Giemsa (RHG) of X and Y chromosomes of sample 57.
(B) FISH for sample 57 via fluorescently labeled probes TelXp/Yp (green) and RP11-209I11 (red, located at Yq11.223).
(C) Genome maps of Y chromosomes of samples 69 (normal chromosome Y), 55, 57, 27, and 79. Dotted red boxes indicate where isodicentric Y chromosomes have no coverage when compared to non-isodicentric Y chromosomes (the top genome map) corresponding to the deleted segment. Most likely the genome maps of the four isochromosomes by optical genome mapping point to three different breakpoints (deletion starts), (sample 55, 21.0 Mb; samples 57 and 79, 20.1 Mb; and sample 27, 19.9 Mb). The first three breakpoints coincide with palindrome 5 (hg19, chrY: 19,567,684–20,063,140 bp), which may be involved in the mediation of the isochromosomes or cause issues in a more precise assembly with the current optical genome mapping. The deletion in sample 55 locates in direct proximity of palindrome 4 (hg19, chrY: 20,612,099–20,802,247 bp).

the aneuploidies of the sex chromosomes had to be manually inferred from the visualized data of the CNV plot (Figure S2). This manual inference is no longer required with the future versions of Bionano Solve. In addition to whole-chromosome aneuploidies, five large CNVs ranging in size between 6.6 and 14 Mb and seven large aberrations corresponding to derivative chromosomes from unbalanced translocations detected by karyotyping were included and detected correctly.

Isochromosomes

Six of our samples contained isochromosomes. Four of those were isodicentric Y chromosomes, one sample contained an isodicentric chromosome 15, and another one had an isodicentric chromosome X. Strikingly, CNV plots from OGM match perfectly those from CNV microarray, reproducing patterns that suggest mosaic aberrations (Figure S7 and Figure S8). The four isodicentric Y chromosomes all showed a similar genome map pattern characterized by a complete absence of genome maps at the q arm starting from q11.221 compared to other XY samples (Figure 3 and Figure S7). The largest part of q12, except the pseudo-autosomal region 2 in Yqter, has no coverage

in any of the samples, including controls, because this part of the chromosome represents a repeat-rich gap in the reference genome (hg19, N-base gap). Interestingly, whereas samples 27, 57, and 79 had a nearly identical coverage pattern, only sample 55 showed a slightly different breakpoint: a part of q11.222 was still covered. While the coverage pattern strongly suggests the presence of isochromosomes in all samples, it should be noted that the SV pipeline was not able to detect the fusion event at the q arm of idic(Y). This is most likely due to the high complexity of the region that is extremely rich in inverted segmental duplications[39] and the incomplete reference genome for the Y chromosome. The comparison between putative breakpoints for each sample and the localization of the palindromes as characterized by Lange et al.[40] and Skov et al.[41] shows that, for all samples, the breakpoint and presumably the fusion point falls in one of those complex sequences (Figure 3), which most likely explains why the SV algorithm did not detect the SV event. This should be improved in the future with (1) the improved versions of the reference genome for chromosome Y and (2) the upgrades of OGM software via the addition of the possibility of subassemblies using the longest molecules only. Finally,
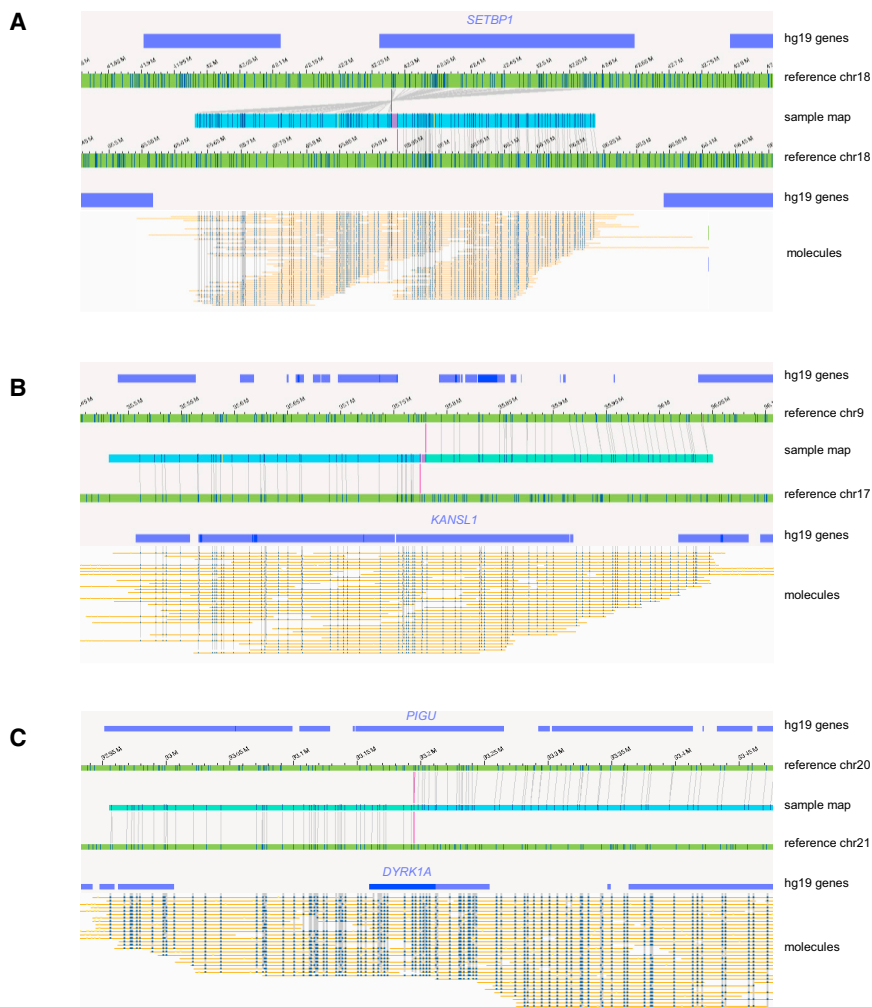
Figure 4. Examples of inversions and translocations interrupting well-known disease-causing genes

(A) Inversion inv(18)(q22.1q12.3), disrupting *SETBP1* in sample 47.
(B) Translocation t(9;17)(p13.3;q21.31), interrupting *KANSL1* in sample 49.
(C) Translocation t(20;21)(q11.22;q22.13), interrupting *DYRK1A* and *PIGU* in sample 54.

tected, connecting positions chrX: g.57,009,891 (p11.21) and chrX: g.78,599,384 (q21.1), confirming and refining the positions previously detected by karyotyping. The FCN of 1.6 for this region (versus expected CN 2) is consistent with the mosaic state of this ring chromosome (Figure S9).

Translocations and inversions

Thirty-four of the investigated samples carried previously identified balanced (n = 27) and unbalanced (n = 7) translocations, and six others displayed inversions, which were all detected by OGM. As expected, unbalanced translocations were detectable by both SV calling and CNV calling, whereas balanced translocations and inversions were only detected by SV calling (Table S5).

Traditionally, balanced translocations can be detected via karyotyping but not via CNV microarray. OGM is able to refine translocation breakpoints for such cases. Accordingly, several balanced translocations and inversions were shown to most likely disrupt protein-coding genes, including the well-described *SETBP1* (MIM: 611060), *KANSL1* (MIM: 612452), *DYRK1A* (MIM: 600855), and *PIGU* (MIM: 608528); the latter two are disrupted by the same translocation (Figure 4). The breakpoints for *KANSL1* (sample 49) had previously been validated with FISH and WGS,[43] whereas the others are newly uncovered and still need to be confirmed. In all cases, the individual's phenotype matches the expected phenotype for the dominant diseases associated with the respective genes (*DYRK1A*, *KANSL1*, and *SETBP1*). The detection of these breakpoints with OGM was much more accurate than with karyotyping. For the few breakpoints for which WGS data were available for comparison,[43] the breakpoint accuracy was within 5 kb (Figure S10).

Microdeletions and microduplications

In addition to large chromosomal aberrations (aneuploidies, large CNVs, and translocations), our cohort included 34 microdeletions/microduplications (<5 Mb).
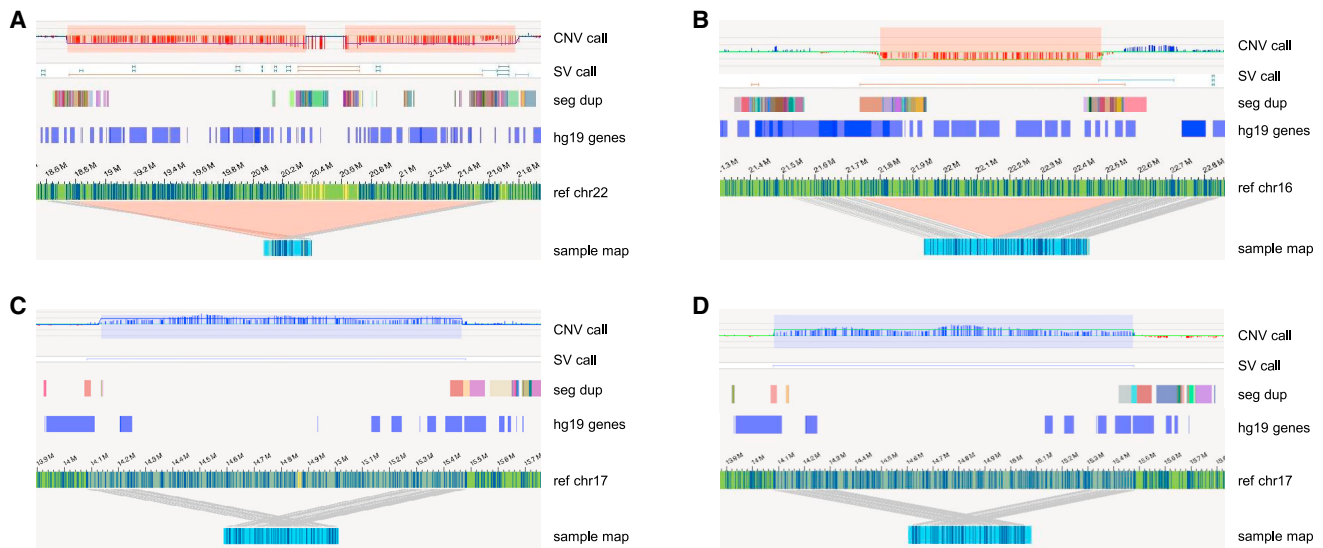
centromeres currently cannot be detected, and hence the distinction between dicentric versus monocentric status may remain uncertain in some cases where only the p arm is involved. However, it is worth noting that these limitations do not change the diagnosis nor the genetic counseling for the individual.

Regarding isochromosomes 15 and X, their detection resulted in complex coverage patterns. For chromosome 15, the FCNs of the affected regions differed (the proximal segment had an FCN = 4 and the distal segment had an FCN = 3), which was consistent with CNV microarray data suggesting an asymmetric idic(15), as previously described.[42] The isodicentric chromosome X was present in low mosaic state as shown by routine methods as well as OGM (see Figure S8).

Ring chromosome

One of the samples analyzed contained a mosaic ring chromosome X, as previously detected by karyotyping (Figure 2). The karyotype reported was 45,X[14]/46,X,r(X)(p11.21q21.1)[21]. The individual presented with growth retardation and development delay. Following OGM, an intrachromosomal translocation on chromosome X was de-

Figure 5. Examples of recurrent CNVs involved in well-recognized microdeletion/microduplication syndromes
(A) A 22q11.21 deletion from sample 2.
(B) A 16p12.2 deletion from sample 76.
(C and D) Two distinct cases of 17p12 duplication from samples 8 (C) and 43 (D) in which the SV call reveals identical breakpoint locations. Sample maps also identify the location and orientation of the duplicated segments as tandem duplications.

These ranged in size from 34 kb (sample 84) to 4.2 Mb (sample 44) and included some of the well-known microdeletion/microduplication syndromes, such as DiGeorge syndrome (22q11.2 deletion syndrome [MIM: 188400]), Williams-Beuren syndrome (deletion 7q11.23 [MIM: 194050]), Charcot-Marie-Tooth syndrome type 1A (CMT1A, duplication 17p12 [MIM: 118220]), and 1q21.1 susceptibility locus for Thrombocytopenia-Absent Radius (TAR) syndrome (MIM: 274000). Although the presence of segmental duplications (SegDups) for several of these microdeletions/microduplications often leads to breaking of the genome maps, all microdeletions/microduplications were correctly called by either the SV or CNV algorithms or both (Tables S1 and S5). SegDups often mediate non-homologous allelic recombination (NAHR), which results in recurrent CNVs; some of these respective regions belong to the most complex sequences of the human genome.[44] The ultra-long reads produced by OGM allow assemblies even spanning several SegDups as shown for the 22q11.2 microdeletion causing DiGeorge syndrome and the 16p12.2 microdeletion syndrome and in two cases with 17p12 tandem duplication causing Charcot-Marie-Tooth syndrome (Figure 5).

Complex cases
Finally, four of the samples included in this study presented with complex rearrangements (samples 28, 52, 55, and 66), all of which were from individuals with developmental delay and/or intellectual disability (Table S1). In general, OGM allowed to resolve precise breakpoints in contrast to karyotyping. For example, the karyotype of sample 28 (Figure 6) showed a translocation t(3;6)(q1?2;p2?2), a derivative chromosome 4 (?der(4)(:p1?2->q1?2:)), and a derivative chromosome 5 (der(5)(4pter->4p1?2::4q1?2->4q34.2::5p14.2->5qter)) in different clones. CNV microarray showed losses on 4q34 (4q34.2q34.3(176587929_190957474)x1 dn) and 5p15 (5p15.33p14.2(113577_24449849)x1 dn). Following OGM, the translocation t(3;6)(q1?2;p2?2) was identified as t(3;6)(q13.12;q24.3). In addition, a translocation t(4;5)(q34.2;p14.2), a loss of 4q34.2q34.3, and a loss of 5p15.33p14.2 were detected, concordant with previous results. In the same sample, OGM also revealed putative additional translocations t(3;4)(q13.11;q12), t(3;4)(q13.11;p11), and t(4;6)(q12;p22.3) and an inversion inv(13)(q31.2;q33.3) (Figure 6).

Another sample (66) showed a three-way translocation t(3;13;5)(p11.1;p12;p14) after karyotyping and four losses on chromosome 3 (3p14.1(65238298_68667113)x1,3p13(70127345_73724765)x1,3p12.1(83784489_85467284)x1,3q11.2(97180779_97270083)x1) following CNV microarray (Figure 6). In addition to confirming these aberrations, except the breakpoint on the p arm of chr13, OGM unraveled additional complex rearrangements on chromosome 3, leading to the identification of a chromoanagenesis. For all residual samples with complex rearrangements, see Table S1 and Figures S11 and S12.

Comparison of CNV detection by optical genome mapping and CNV microarrays
The main purpose of this study was to assess the ability of OGM to detect clinically relevant aberrations that were previously identified by standard-of-care tests (karyotyping, FISH, and/or CNV microarrays). Full assessment of the true positive rate for all SVs, regardless of their clinical relevance, is beyond the scope of the present study.
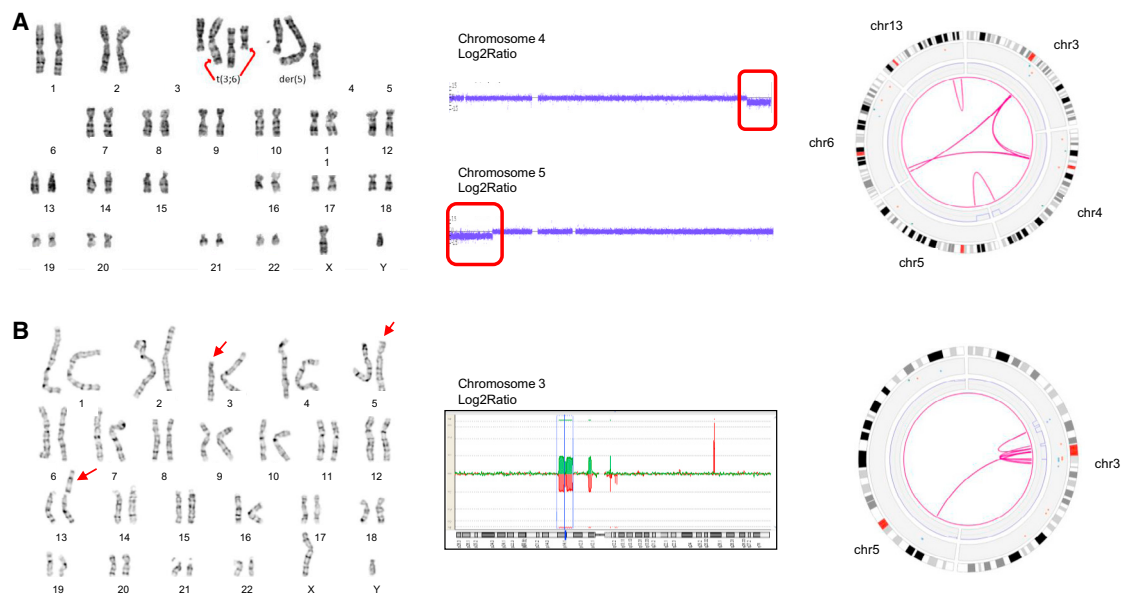
Figure 6. Complex rearrangements

(A) Sample 28: (from left to right) karyogram showing the translocation t(3;6) and the derivative chromosomes 4 and 5; CNV-microarray data showing two *de novo* deletions on chromosomes 4 and 5; and optical genome mapping circos plot of chromosomes 3, 4, 5, 6, and 13. Optical genome mapping data confirm all known aberrations and show the presence of additional translocations t(3;4) and t(4;6) plus an inversion on chromosome 13.

(B) Sample 66: (from left to right) karyogram showing a three-way translocation 46,XY,t(3;13;5)(p11.1;p12;p14); CGH pane showing multiple deletions on chromosome 3; circos plot of chromosomes 3 and 5 (chr13 not represented in this plot because of the centromeric localization of the breakpoint) showing multiple intrachromosomal translocations on chromosome 3 and a translocation between chromosomes 3 and 5; and genome maps of chromosome 3 showing multiple rearrangements on chromosome 3, suggesting a chromoanagenesis.

However, for a few samples (56, 62, 64, 77, 78, and 79), complete CNV microarray data were available, allowing for a comparison to CNVs detected with OGM (after filtering steps).

All clinically significant events were successfully detected by OGM. In addition, 36 CNVs were still called by OGM after our proposed filter settings. Verification of CNV microarray data revealed that four of these were called by CNV microarray as well but were not reported because they were considered benign. Small events (<100–200 kb) that were not called by CNV microarray at first (n = 32) were located either in regions that were not represented in the ISCA Agilent 8x60K DNA array (n = 12) or in regions that were represented by only few probes, down to one, (n = 20) (Table S7). Fifteen out of 20 small CNVs for which at least one probe was present in the microarray had a log ratio that was consistent with the aberration (log ratio > 0.25 for duplications and < −0.25 for deletions). The five remaining calls that were not supported by CNV microarray could be either false positive calls of OGM or false negative calls of CNV microarray. In total, at least 79% (19/24) of the OGM CNV calls were supported by CNV microarray data. Overall, comparison to CNV microarray results supports low false positive and false negative rates of OGM for CNV detection.

Taken together, OGM allowed the correct unraveling and further refinement of complex karyotypes (which previously required the combination of karyotyping, FISH, and CNV microarrays) by combining the detection of balanced and unbalanced events in one assay and at an unprecedented resolution.

Discussion

In this study, we have shown that OGM is capable of comprehensively detecting all classes of chromosomal aberrations and may complement or replace current cytogenetic technologies. In summary, we identified all 99 previously reported aberrations from 85 samples, reaching 100% concordance. Our cohort is representative of what is expected in a clinical setting because it included samples from different tissues, various clinical indications, and a great variety of different chromosomal aberrations. We demonstrated that OGM enables the detection of aneuploidies, CNVs, and other structural variations, including balanced and unbalanced rearrangements at sizes ranging from a few kilobases to several megabases. The combination of two analysis pipelines, one based on coverage depth and the other one based on the comparison of a *de novo*-assembled genome map to a reference map, allows for the most complete detection of all balanced and unbalanced aberrations, as shown by our results. In fact, the first pipeline performs better for large deletions and duplications and is currently the only tool to detect terminal chromosomal losses or other events that do not create the

fusion of unique novel molecules, such as aneuploidies. The second pipeline is more sensitive to small CNVs down to a few hundred base pairs and allows for better breakpoint resolution of SVs. However, the breakpoint uncertainty can be 3.2 kb on average, which corresponds to the average distance between two labels, genome wide, and accounts for a misalignment of one label. Of note, for five out of 99 aberrations, subtle adaptation of filter setting was required to reach full concordance, including lowering the CNV confidence for three samples and turning off the SV DLE-1 mask filter for two samples. While the SV DLE-1 filter proves to be useful, as it masks the extremely copy-number-variable (and/or repeat-rich) regions of the human genome, this may mask some clinically relevant regions. This was the case for two events in our study: t(X;1)(p22.32;q21.1) (sample 50) and t(7;11)(q11.23;q13.4) (sample 81). While loosening filters did not dramatically increase the number of variants per case, we anticipate that this may not be required anymore with improved variant calling and filtering. Subsequent updates of the software should therefore contain both optimized confidence values and a more fine-scaled mask that takes into account clinically relevant regions.

The current technology is not yet capable of detecting breakpoints of balanced SVs lying within large repetitive, unmappable regions such as centromeres, p arm of acrocentric chromosomes, or constitutive heterochromatin stretches. The 85 samples presented here included nine such events, none of which were unequivocally detected. These, however, were considered beyond the current technical capability and were judged to fall beyond the scope of this study prior to the analysis. Challenges to map such breakpoints are due to gaps in the human reference genome,[45] impeding the development of specific labeling strategies. Moreover, centromeric repeats are several megabases long, far larger than the longest single molecules that can be obtained with any current technology. However, in some cases, we were able to detect translocations with breakpoints that locate within pericentromeric regions and that were not detected by paired-end whole-genome sequencing (samples 50, 51, and 54; unpublished data). In our opinion, OGM may be best suited to assemble such complex regions of the genome because it is not based on sequencing, and hence cannot be biased by the sequence context itself, and it allows individual reads up to 2 Mb, which is unprecedented by most other technologies. This is supported by a most recent study in which the combination of sequencing-based technologies with OGM enabled the first full assembly of the human X chromosome from telomere-to-telomere.[45]

The strength of combining ultra-long reads (up to 2 Mb) and being agnostic to sequence context was also highlighted by the detection and full assembly of recurrent microdeletion/microduplication syndromes that are known to be mediated by NAHR of segmental duplication, i.e., low-copy repeats (Figure 5).[46] This included a 22q11.21 deletion, a 16p12.2 deletion, and a 17p12 duplication in two independent samples. However, for four out of seven classical microdeletion/microduplication syndromes, the assembly did not yet fully resolve the SV, but we expect that an assembly of all shall be feasible with subsequent improvements of the technology. This access to complex regions is a core benefit of OGM because these regions are generally considered to be among the most complex region of the human genome. A very recent study highlighted not only the assembly of the 22q11.2 microdeletion/microduplication region[46] but also allowed showing the complexity of existing human and non-human primate haplotypes. OGM was also shown to be essential for even more complex regions, e.g., the chromosome X centromere assembly.[45] The better understanding of these very complex regions is of great importance because subtle differences in breakpoints may disrupt/impact genes within the segmental duplications that may explain subtle genotype-phenotype differences, as, e.g., expected for the 3q29 deletion/duplication.[47]

The detection of rearrangements involving the Y chromosome can also be challenging. For example, an isochromosome Y with breakpoints in the long arm of chromosome Y is not detectable by sequencing technologies. In our cohort, four individuals had a Y chromosomal aberration (samples 27, 55, 57, and 79) where the coverage profile was very suggestive of an isochromosome. OGM CNV profiles matched those from CNV microarray and were consistent with the mosaic state of respective aberrations. The inability to detect the actual fusion molecules with the SV pipeline is most likely due to the highly complex nature of the sequence at this region with lots of inverted highly repetitive palindromic sequence motifs.[39,48] All four deletion breakpoints locate in immediate proximity of palindromes 4 or 5.[40,41] Improvements in assembly tools, higher coverage, and sub-assemblies of the longest molecules only may help resolve this issue in the future, as recently shown by the Telomere-to-Telomere consortium, which used OGM to assemble complex regions of chromosome X.[45]

The ability to detect not only unbalanced but also balanced SVs genome wide with high accuracy is another advantage of OGM. Some of the breakpoints of balanced translocations detected here fall into known disease-related genes that they may disrupt. This was previously unknown for individual 54, whose karyotype is 46,XY,t(20;21)(q11.2;q21). OGM refined the chromosome 21 breakpoint to 21q22.13 instead of 21q21 and showed that this balanced translocation disrupts *DYRK1A*. The individual displays autism spectrum disorder and microcephaly consistent with *DYRK1A* haploinsufficiency, which has been shown to be associated with autism spectrum disorder, intellectual disability, and microcephaly, and as such, this refines the molecular diagnosis for the respective individual.[49–51] Similarly, other cases of likely gene disruptions provided hints into the molecular diagnosis of intellectual disability. For example, individual 47 had an inversion that we found to most likely disrupt

SETBP1, and individual 49 had a translocation that may break *KANSL1* in concordance with previously published WGS results.[43] In both cases, haploinsufficiency of the respective genes is known to lead to clinical syndromes including intellectual disability[52,53] consistent with our individuals' phenotypes.

OGM was also able to detect complex rearrangements including multiple translocations or even chromoanagenesis. OGM results suggested more complex events than expected in samples 28, 47, 52, 66, 70, and 74, where the additional SV calls need to be further validated with an orthogonal method.

From a technical point of view, our results support the robustness of the technology because our samples were processed in three different laboratories or facilities. Results were highly similar in terms of quality metrics, number of variant calls, and performance, stated by the 100% concordance with conventional cytogenetic analyses. Some differences in the total number of calls are most likely due to a different version of analysis software used for a few samples (70–76) or reflect differences in the level of complexity of the aberrations. Clinically relevant results were unchanged.

Here, we suggest a suited filtering strategy that resulted in only six SV calls and four CNV calls on average per sample, which is compatible with a diagnosis activity in clinical cytogenetic laboratories. Depending on event type, the same events may be supported by both CNV and SV calls (Table S5). While detecting all previously known events, the total number of aberrations is well in line with CNV microarray data with similar genome-wide resolution and is strongly suggestive of a low false positive rate, as suggested by high concordance with CNV microarray findings (Table S7). This may be in contrast to NGS-based SV calling; several reports point out the high number of false calls with sequencing-based technologies.[54,55] A full comparison with a variety of short- and long-read sequencing technologies is beyond the scope of this manuscript, but in line with our observation, recent studies suggest a high degree of concordance between OGM and WGS.[22] One could also consider subsequent analysis filters to reduce further the number of SVs to interpret in a clinical setting, e.g., include overlap of SVs with known disease-related genes or loci. This is an important point to make genetic investigations time efficient because the higher the delay to diagnosis, the lower the chances of successful management of the disease.

The main focus here was to investigate the concordance, i.e., sensitivity, for known aberrations to explore the possibility to replace standard cytogenetic assays by optical genome mapping. In DDs, OGM could complement sequencing approaches to allow for a comprehensive genomic investigation. In reproductive disorders, we foresee a much-improved analysis with OGM. However, to prevent missing balanced Robertsonian or whole-arm chromosome-balanced translocations in a few cases, a quick count of a few metaphase spreads can be performed for specific clinical indications in which these events are frequently involved.

Optical genome mapping with Bionano genome imaging can best be compared to an ultra-high-resolution karyotype reaching approximately 10,000 times higher resolution than the conventional karyotype. It is likely that, at some stage, long-read sequencing approaches may allow fully comprehensive assessment of all SVs and chromosomal aberrations in each personal genome, possibly after *de novo* genome assembly.[45,56] Nonetheless, some benefits of OGM may prevail: (1) the ultra-long molecules reaching up to 2 Mb and a current N50 of more than 250 kb are unprecedented by any current NGS platform; (2) the analysis of native molecules without any PCR or library preparation; (3) the relative ease of analysis not requiring bioinformatics processing by the user nor significant data storage capacities; (4) the relative low costs given the level of coverage; (5) OGM can produce 300–1600× genome coverage, allowing the reliable detection of rare somatic events, with additional improvements in development;[30] (6) the absence of sequencing could be preferred in some cases to avoid undesired incidental findings, especially for some individuals referred for reproductive disorders; and (7) the ability of OGM to analyze repeat-contraction-related diseases such as facioscapulohumeral muscular dystrophy[57,58] opens up new perspectives for the detection of expansion diseases such as fragile X syndrome or Huntington disease.

To conclude, this proof-of-principle study pioneered the technical validation of optical genome mapping as a solid alternative approach to karyotyping, FISH, and CNV microarrays for the detection of clinically relevant constitutional chromosomal aberrations. We showed that OGM is capable of reaching 100% concordance while detecting all different types of chromosomal anomalies, including aneuploidies and CNVs as well as balanced chromosomal abnormalities and complex chromosomal rearrangements. Current limitations in detecting balanced aberrations involving (peri)centromeric regions may require complementing OGM analysis with a simple check for such balanced translocations by karyotyping of few metaphases in individuals with reproductive disorders. This may become obsolete with future improvements in technical and analytical aspects of OGM as well as gap filling in the human reference genome. Additional improvements such as streamlining the reporting of both SV and CNV algorithms, faster turnaround times, and allowing for ISCN-compatible nomenclature are already expected. These will allow for full prospective clinical-utility studies shortly and pave the way toward diagnostic implementation.

Data and code availability

Due to local regulation, individuals' data cannot be made publicly available. However, we can respond to individual requests that can be sent to the corresponding authors. All software is commercially available via Bionano Genomics. All filter settings suggested here can be reproduced in the available Bionano Genomics software suite.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.05.012.

Web resources

Bionano access, https://bionanogenomics.com/support/software-downloads/#bionanoaccess
OMIM, https://www.omim.org/

References

1. Vissers, L.E., Veltman, J.A., van Kessel, A.G., and Brunner, H.G. (2005). Identification of disease genes by whole genome CGH arrays. Hum. Mol. Genet. *14*, R215–R223.

2. Speicher, M.R., and Carter, N.P. (2005). The new cytogenetics: blurring the boundaries with molecular biology. Nat. Rev. Genet. *6*, 782–792.

3. Smeets, D.F. (2004). Historical prospective of human cytogenetics: from microscope to microarray. Clin. Biochem. *37*, 439–446.

4. Chantot-Bastaraud, S., Ravel, C., and Siffroi, J.P. (2008). Underlying karyotype abnormalities in IVF/ICSI patients. Reprod. Biomed. Online *16*, 514–522.

5. Hofherr, S.E., Wiktor, A.E., Kipp, B.R., Dawson, D.B., and Van Dyke, D.L. (2011). Clinical diagnostic testing for the cytogenetic and molecular causes of male infertility: the Mayo Clinic experience. J. Assist. Reprod. Genet. *28*, 1091–1098.

6. De Braekeleer, M., and Dao, T.N. (1990). Cytogenetic studies in couples experiencing repeated pregnancy losses. Hum. Reprod. *5*, 519–528.

7. De Braekeleer, M., and Dao, T.N. (1991). Cytogenetic studies in male infertility: a review. Hum. Reprod. *6*, 245–250.

8. Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. Nat. Rev. Genet. *12*, 363–376.

9. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. *86*, 749–764.

10. de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, Sv., Nillesen, W.M., Huys, E.H., Leeuw, Nd., et al. (2005). Diagnostic genome profiling in mental retardation. Am. J. Hum. Genet. *77*, 606–616.

11. Schinzel, A. (2001). Catalogue of unbalanced chromosome aberrations in man. (De Gruyter).

12. van Karnebeek, C.D., Jansweijer, M.C., Leenders, A.G., Offringa, M., and Hennekam, R.C. (2005). Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness. Eur. J. Hum. Genet. *13*, 6–25.

13. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344–347.

14. Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., et al. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet. Med. *20*, 435–443.

15. Stavropoulos, D.J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R.K.C., Szego, M.J., et al. (2016). Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. NPJ Genom. Med. *1*, 15012.

16. Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.A., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat. Genet. *49*, 36–45.

17. Dong, Z., Wang, H., Chen, H., Jiang, H., Yuan, J., Yang, Z., Wang, W.J., Xu, F., Guo, X., Cao, Y., et al. (2018). Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. Genet. Med. *20*, 697–707.

18. Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019). Comprehensive evaluation of structural

variation detection algorithms for whole genome sequencing. Genome Biol. *20*, 117.

19. Monlong, J., Cossette, P., Meloche, C., Rouleau, G., Girard, S.L., and Bourque, G. (2018). Human copy number variants are enriched in regions of low mappability. Nucleic Acids Res. *46*, 7236–7249.

20. Luo, F. (2019). A systematic evaluation of copy number alterations detection methods on real SNP array and deep sequencing data. BMC Bioinformatics 20 (*Suppl 25*), 692.

21. Zhao, L., Liu, H., Yuan, X., Gao, K., and Duan, J. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. BMC Bioinformatics *21*, 97.

22. Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*, 1784.

23. Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-Read Sequencing Emerging in Medical Genetics. Front. Genet. *10*, 426.

24. Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., and Wang, Y.K. (1993). Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science *262*, 110–114.

25. Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., and Kwok, P.Y. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol. *30*, 771–776.

26. Chan, S., Lam, E., Saghbini, M., Bocklandt, S., Hastie, A., Cao, H., Holmlin, E., and Borodkin, M. (2018). Structural Variation Detection and Analysis Using Bionano Optical Mapping. Methods Mol. Biol. *1833*, 193–203.

27. Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G., et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. Nat. Genet. *51*, 224–229.

28. Kronenberg, Z.N., Fiddes, I.T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O.S., Underwood, J.G., Nelson, B.J., Chaisson, M.J.P., Dougherty, M.L., et al. (2018). High-resolution comparative analysis of great ape genomes. Science *360*, eaar6343.

29. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55.

30. Neveling, K., Mantere, T., Vermeulen, S., Oorsprong, M., van Beek, R., Kater-Baats, E., Pauper, M., van der Zande, G., Smeets, D., Weghuis, D.O., et al. (2020). Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging. bioRxiv. https://doi.org/10.1101/2020.02.06.935742.

31. Barseghyan, H., Délot, E.C., and Vilain, E. (2018). New technologies to uncover the molecular basis of disorders of sex development. Mol. Cell. Endocrinol. *468*, 60–69.

32. Du, C., Mark, D., Wappenschmidt, B., Böckmann, B., Pabst, B., Chan, S., Cao, H., Morlot, S., Scholz, C., Auber, B., et al. (2018). A tandem duplication of BRCA1 exons 1-19 through DHX8 exon 2 in four families with hereditary breast and ovarian cancer syndrome. Breast Cancer Res. Treat. *172*, 561–569.

33. Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A.K.Y., McCaffrey, J., Young, E., Lam, E.T., Hastie, A.R., Wong, K.H.Y., et al. (2019). Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nat. Commun. *10*, 1025.

34. Bates, S.E. (2011). Classical cytogenetics: karyotyping techniques. Methods Mol. Biol. *767*, 177–190.

35. Vermeesch, J.R., Brady, P.D., Sanlaville, D., Kok, K., and Hastings, R.J. (2012). Genome-wide arrays: quality criteria and platforms to be used in routine diagnostics. Hum. Mutat. *33*, 906–915.

36. Bionano Genomics. Bionano Solve Theory of Operation: Structural Variant Calling. https://bionanogenomics.com/wp-content/uploads/2018/04/30110-Bionano-Solve-Theory-of-Operation-Structural-Variant-Calling.pdf.

37. Bionano Genomics. Introduction to Copy Number Analysis. https://bionanogenomics.com/wp-content/uploads/2018/04/30210-Introduction-to-Copy-Number-Analysis.pdf.

38. Gross, A.M., Ajay, S.S., Rajan, V., Brown, C., Bluske, K., Burns, N.J., Chawla, A., Coffey, A.J., Malhotra, A., Scocchia, A., et al. (2019). Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. Genet. Med. *21*, 1121–1130.

39. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

40. Lange, J., Skaletsky, H., van Daalen, S.K., Embry, S.L., Korver, C.M., Brown, L.G., Oates, R.D., Silber, S., Repping, S., and Page, D.C. (2009). Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell *138*, 855–869.

41. Skov, L., Schierup, M.H.; and Danish Pan Genome Consortium (2017). Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. PLoS Genet. *13*, e1006834.

42. Wang, N.J., Parokonny, A.S., Thatcher, K.N., Driscoll, J., Malone, B.M., Dorrani, N., Sigman, M., LaSalle, J.M., and Schanen, N.C. (2008). Multiple forms of atypical rearrangements generating supernumerary derivative chromosome 15. BMC Genet. *9*, 2.

43. Schluth-Bolard, C., Diguet, F., Chatron, N., Rollat-Farnier, P.A., Bardel, C., Afenjar, A., Amblard, F., Amiel, J., Blesson, S., Callier, P., et al. (2019). Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders. J. Med. Genet. *56*, 526–535.

44. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. Science *297*, 1003–1007.

45. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79–84.

46. Demaerel, W., Mostovoy, Y., Yilmaz, F., Vervoort, L., Pastor, S., Hestand, M.S., Swillen, A., Vergaelen, E., Geiger, E.A., Coughlin, C.R., et al. (2019). The 22q11 low copy repeats are characterized by unprecedented size and structural variability. Genome Res. *29*, 1389–1401.

47. Pollak, R.M., Zinsmeister, M.C., Murphy, M.M., Zwick, M.E., Mulle, J.G.; and Emory 3q29 Project (2020). New phenotypes associated with 3q29 duplication syndrome: Results from the 3q29 registry. Am. J. Med. Genet. A. *182*, 1152–1166.

48. Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature *423*, 873–876.

49. Lee, K.S., Choi, M., Kwon, D.W., Kim, D., Choi, J.M., Kim, A.K., Ham, Y., Han, S.B., Cho, S., and Cheon, C.K. (2020). A novel de novo heterozygous DYRK1A mutation causes complete loss of DYRK1A function and developmental delay. Sci. Rep. *10*, 9849.

50. van Bon, B.W., Coe, B.P., Bernier, R., Green, C., Gerdts, J., Witherspoon, K., Kleefstra, T., Willemsen, M.H., Kumar, R., Bosco, P., et al. (2016). Disruptive de novo mutations of DYRK1A lead to a syndromic form of autism and ID. Mol. Psychiatry *21*, 126–132.

51. Ji, J., Lee, H., Argiropoulos, B., Dorrani, N., Mann, J., Martinez-Agosto, J.A., Gomez-Ospina, N., Gallant, N., Bernstein, J.A., Hudgins, L., et al. (2015). DYRK1A haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies. Eur. J. Hum. Genet. *23*, 1473–1481.

52. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. *46*, 1063–1071.

53. Koolen, D.A., Kramer, J.M., Neveling, K., Nillesen, W.M., Moore-Barton, H.L., Elmslie, F.V., Toutain, A., Amiel, J., Malan, V., Tsai, A.C., et al. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. Nat. Genet. *44*, 639–641.

54. Cameron, D.L., Di Stefano, L., and Papenfuss, A.T. (2019). Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat. Commun. *10*, 3240.

55. Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. Genome Biol. *20*, 246.

56. Chaisson, M.J., Wilson, R.K., and Eichler, E.E. (2015). Genetic variation and the de novo assembly of human genomes. Nat. Rev. Genet. *16*, 627–640.

57. Zheng, Y., Kong, L., Xu, H., Lu, Y., Zhao, X., Yang, Y., Yu, G., Li, P., Liang, F., Jin, H., and Kong, X. (2020). Rapid prenatal diagnosis of Facioscapulohumeral Muscular Dystrophy 1 by combined Bionano optical mapping and karyomapping. Prenat. Diagn. *40*, 317–323.

58. Dai, Y., Li, P., Wang, Z., Liang, F., Yang, F., Fang, L., Huang, Y., Huang, S., Zhou, J., Wang, D., et al. (2020). Single-molecule optical mapping enables quantitative measurement of D4Z4 repeats in facioscapulohumeral muscular dystrophy (FSHD). J. Med. Genet. *57*, 109–120.

Supplemental information

Optical genome mapping enables constitutional

chromosomal aberration detection

Tuomo Mantere, Kornelia Neveling, Céline Pebrel-Richard, Marion Benoist, Guillaume van der Zande, Ellen Kater-Baats, Imane Baatout, Ronald van Beek, Tony Yammine, Michiel Oorsprong, Faten Hsoumi, Daniel Olde-Weghuis, Wed Majdali, Susan Vermeulen, Marc Pauper, Aziza Lebbar, Marian Stevens-Kroef, Damien Sanlaville, Jean Michel Dupont, Dominique Smeets, Alexander Hoischen, Caroline Schluth-Bolard, and Laïla El Khattabi

Supplemental Data Description

Supplemental Data includes 12 figures and 7 tables.

Figure S1. Workflow of optical genome mapping technique.

Figure S2. Visual representation of optical genome mapping data.

Figure S3. Potential mapping artifacts.

Figure S4. Representative example of fragmented CNVs.

Figure S5. Representation of different chromosomal abnormalities.

Figure S6. Representative CNV profiles obtained with optical genome mapping, for different samples.

Figure S7. Isodicentric Y-chromosomes.

Figure S8. Optical genome mapping and CNV-microarray profiles for isochromosomes 15 and X.

Figure S9. Small ring chromosome X.

Figure S10. Optical genome mapping breakpoint detection for translocation t(9;17)(p13;q21), disrupting the gene *KANSL1*.

Figure S11. Complex sample 52.

Figure S12. Complex sample 55.

Table S1. Comparison of previous diagnostic findings with optical genome mapping results.

Table S2. Technical performance of optical genome mapping.

Table S3. Overall numbers of variants per sample.

Table S4. SV and CNV statistics per size range.

Table S5. Detection rate of chromosomal aberrations, by each pipeline, according to their type.

Table S6. Comparison of capabilities and limitations of optical genome mapping versus conventional cytogenetic methods, karyotyping and CNV-microarray, in detecting different types of chromosomal aberrations.

Table S7. Concordance of all filtered SVs and CNVs between optical genome mapping and CNV-microarray
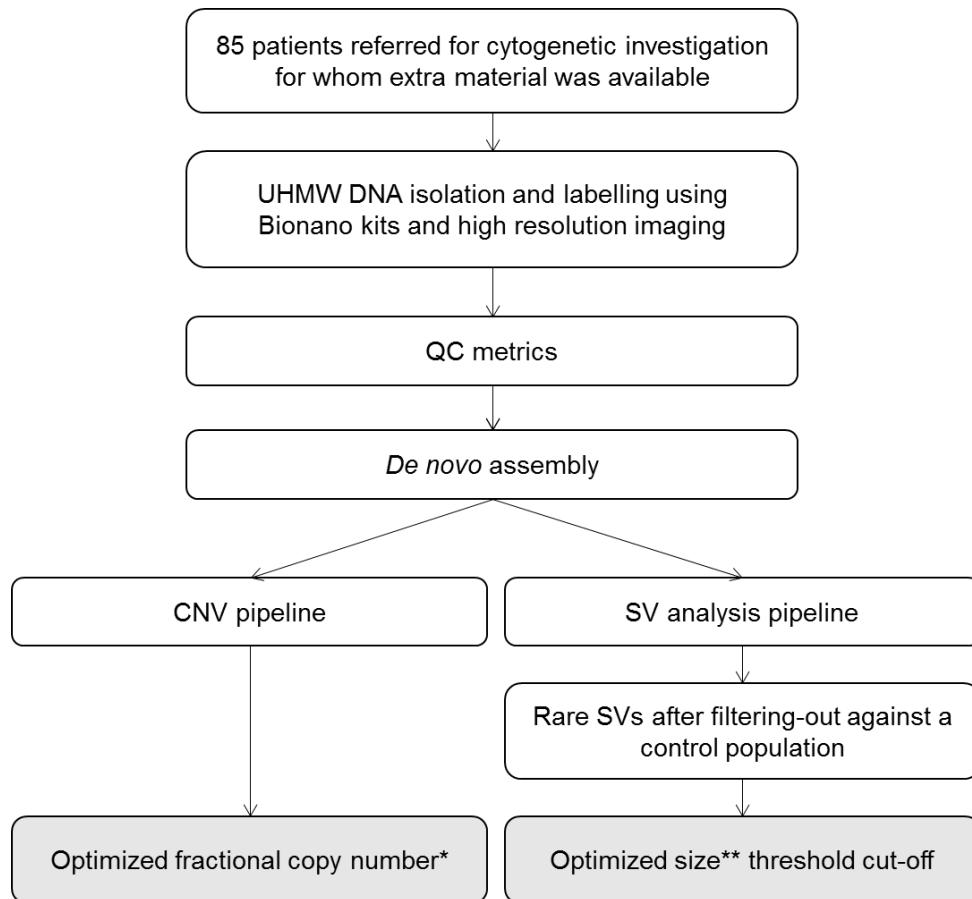
Figure S1. Workflow of optical genome mapping technique.

For this study, 85 samples for whom extra material was available were included. Ultra-high molecular weight DNA was extracted using the Bionano solution phase DNA isolation method. Labeling was done using the DLE-1 chemistry. High resolution imaging of DNA molecules was done on Bionano Saphyr instruments. As different centers were included, different amount of data was produced (~800Gbp for Radboud UMC, ~300Gbp for the French centers), and samples were analyzed using different software versions (3.4.1 and 3.5). A *de novo* assembly was performed, and both SVs and CNVs were called.

Grey boxes are optional steps to reduce the number of calls without losing sensitivity for clinically relevant aberrations. *FCN ≤1.2 for losses and ≥2.8 for gains, **recommended SV pipeline size cut-off = 20 Kb
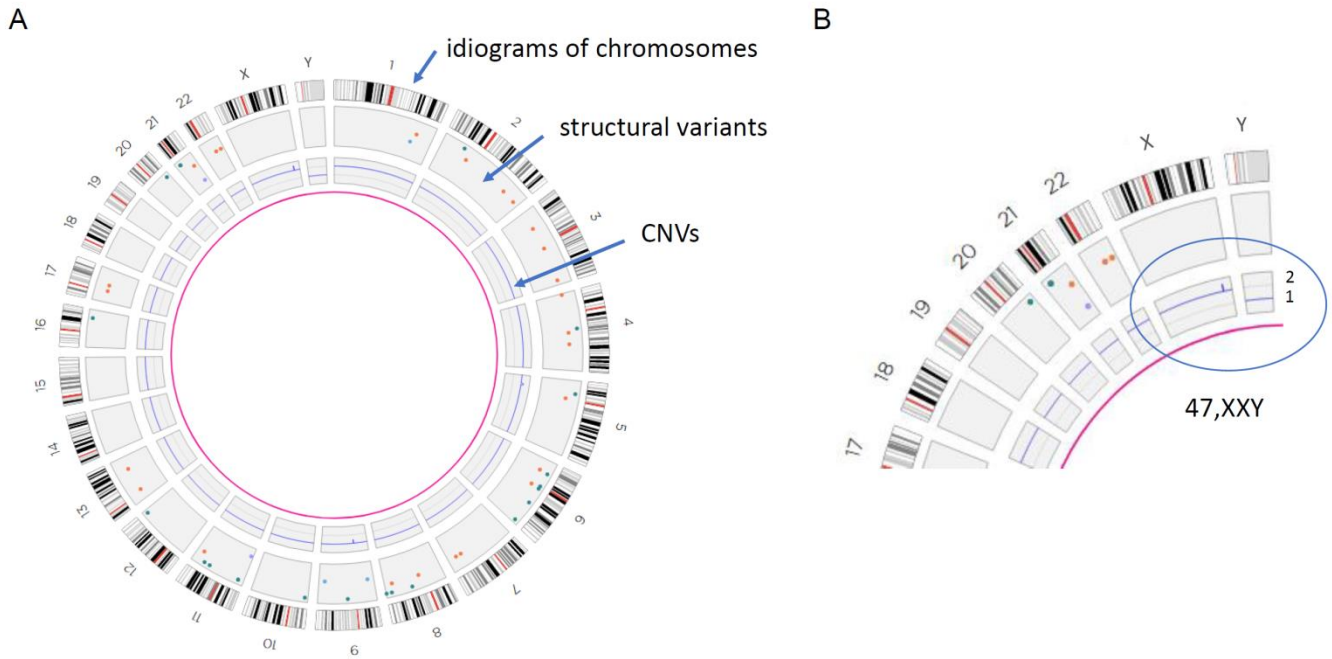
Figure S2. Visual representation of optical genome mapping data.
A) Genome-wide circos plot showing all 24 chromosomes in a circular way. For each chromosome, the idiograms are shown at the outside of the circosplot, with ideogram-style chromosomal banding and the centromeres in red. Different colored dots in the boxes underneath represent different called SVs. The blue line in the box underneath represents the CNV profile, with each peak representing a CNV call. B) Part of a circos plot, showing the sex chromosomes. The blue CNV line shows two copies of chromosome X, as for autosomes, and one copy of chromosome Y consistent with a sex chromosome aneuploidy (47,XXY, resulting in Klinefelter syndrome).

A) Intrachromosomal rearrangement chr 1 (sample 22)
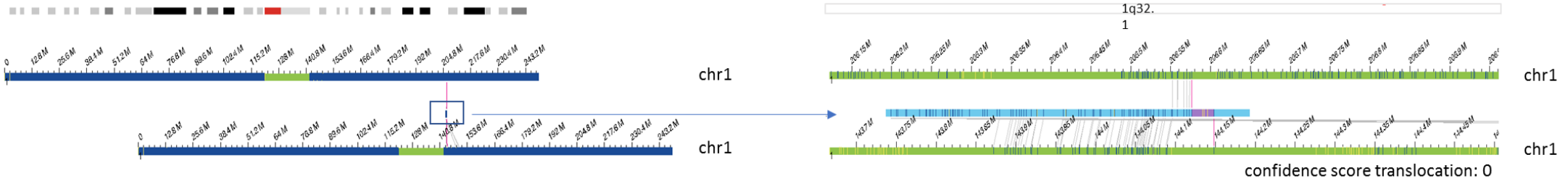
B) Intrachromosomal rearrangement chr 9 (sample 11)
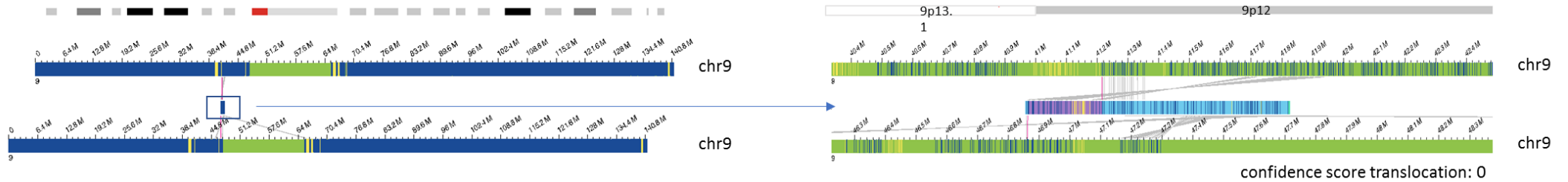
Figure S3 Potential mapping artifacts.

This figure shows two different rearrangements called in two different samples that likely derive from mapping difficulties. A) Intrachromosomal rearrangement in chr1, B) Intrachromosomal rearrangement in chr9. In both samples, the rearrangement breakpoints occur in highly repetitive regions, often in close proximity to the centromere. The left side of each subfigure shows the whole chromosome view of the respective translocation, whereas the right side shows the zoom-in of the corresponding molecule map. The many different connecting lines between the labels in the reference genome map and the samples' genome map let assume that mapping of highly repetitive regions leads to false positive calls.
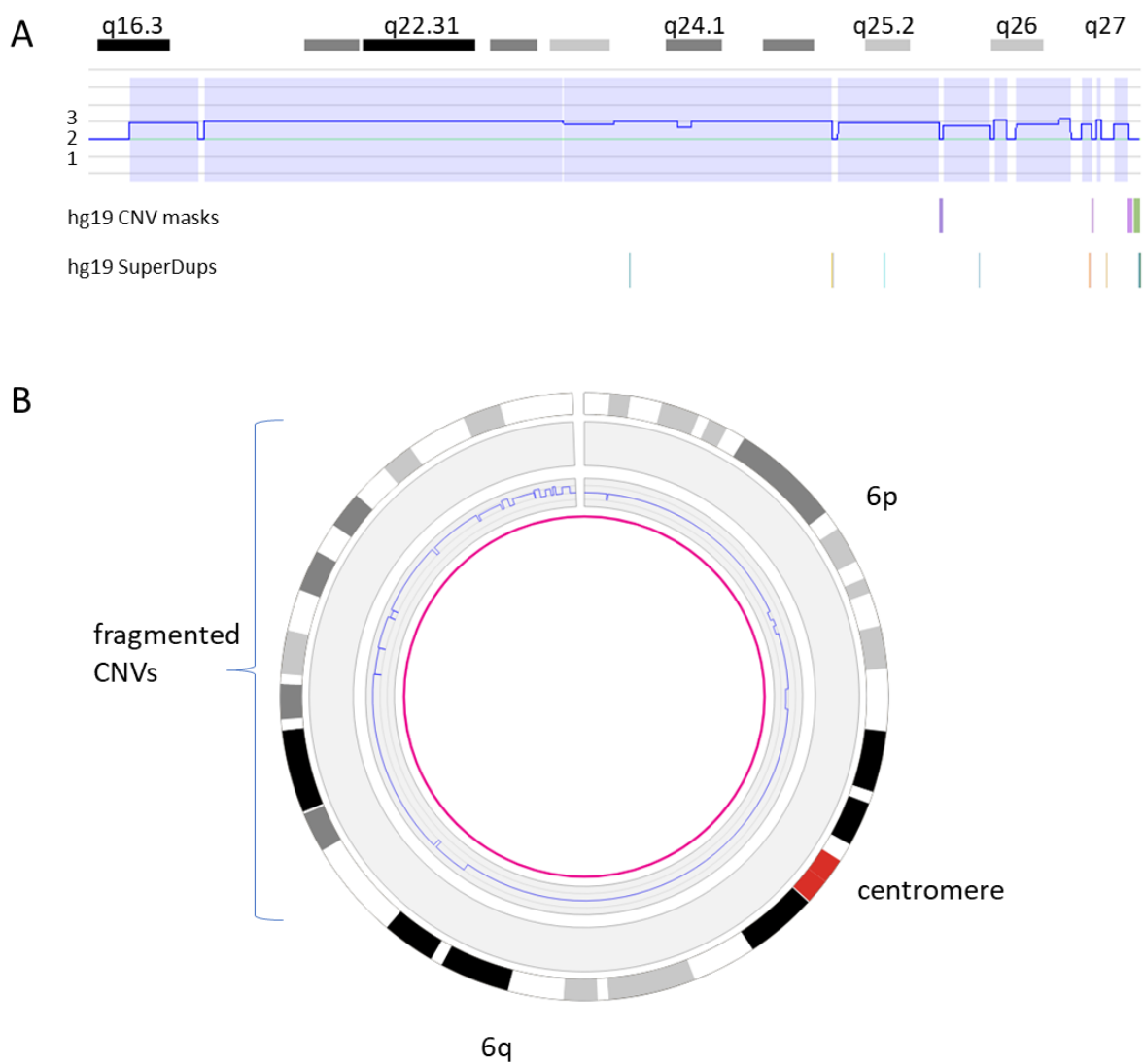
Figure S4. Representative example of fragmented CNVs.

This figure shows an example of a large CNV which is a consequence of a Robertsonian translocation. However, instead of having one call of a large CNV, several smaller CNVs are called. A) CNV coverage plot of 6q16.3q27 region. B) Circos plot of the entire chromosome 6.

Deletion
Sample 1: 8p22p21.3(18825888_19364764)x1
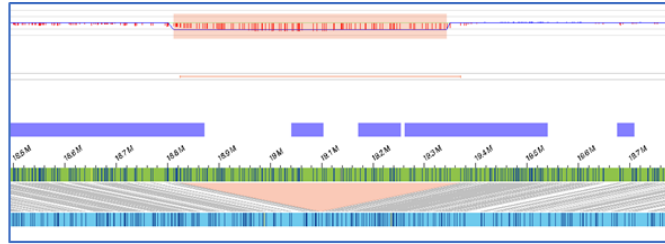
Duplication
Sample 7: 3q29(194349341_194988874)x3
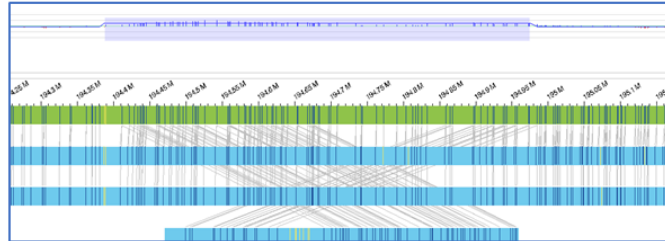
Inversion
Sample 15:
inv(13)(q12q22)
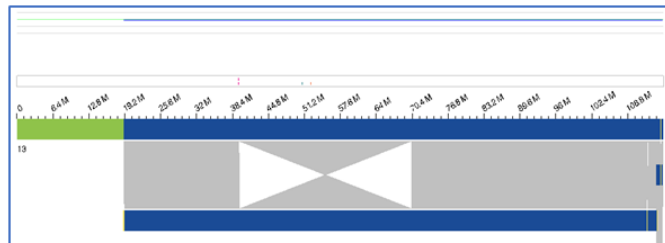
Translocation balanced
Sample 14:
t(3;6)(q11.2;q13)
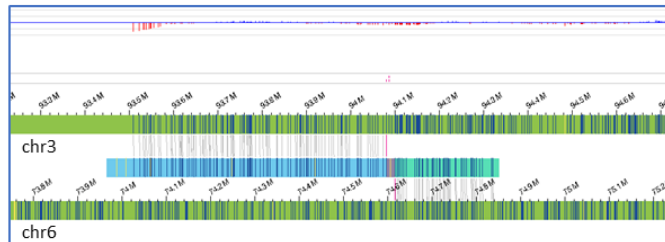
Translocation unbalanced
Sample 26:
t(12;20)(p13.32;p12.3),
12p13.33p13.32 loss, 20p13p12.3 gain
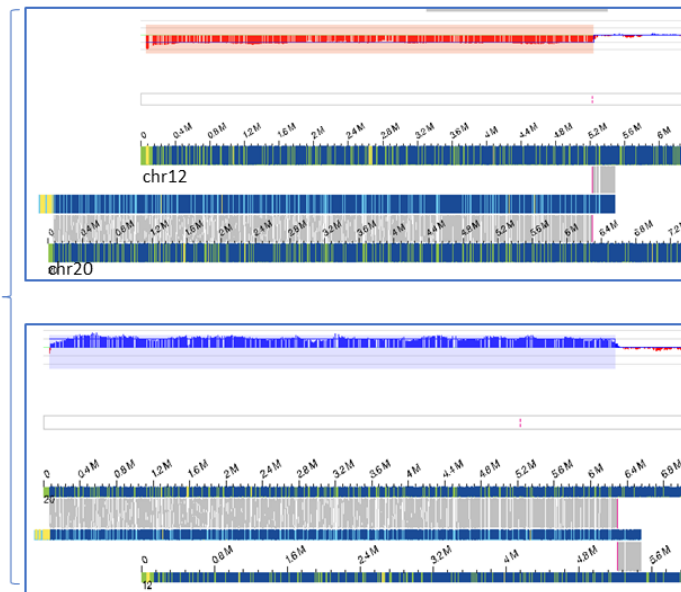
Figure S5 (1/2). Representation of different chromosomal abnormalities.
This figure shows an example of each type of chromosomal aberration that was detected in this study. All aberrations shown are described in detail in Table S1.

Isochromosome

Sample 77:
15q11.2q13.2 gain (4x)
15q13.3 gain (3x)
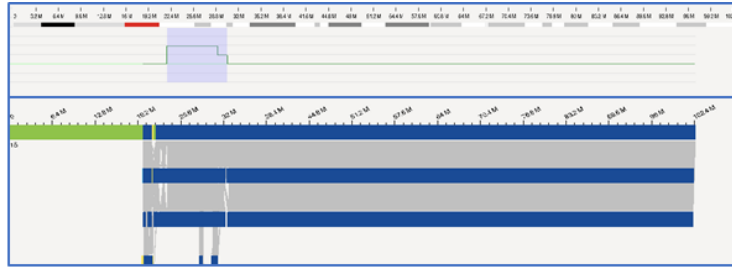
Ring chromosome

Sample 39:
r(X)(p11.21q21.1)
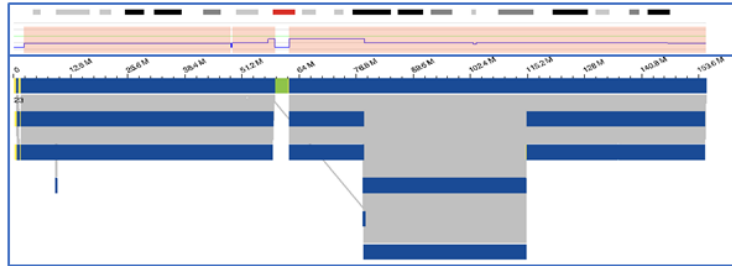
Complex rearrangement

Sample 66:
3p14.1(65238298_68667113)x1
3p13(70127345_73724765)x1
3p12.1(83784489_85467284)x1
3q11.2(97180779_97270083)x1
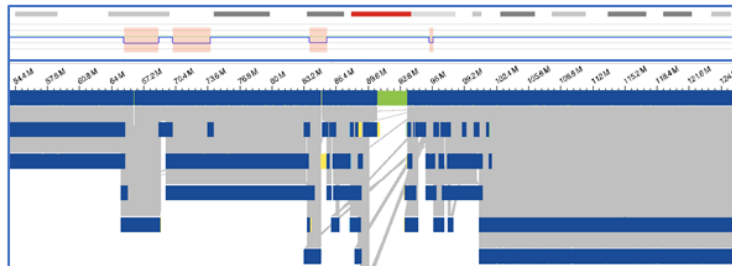
Aneuploidy

Sample 24:
trisomy 21 via Robertsonian translocation

Large terminal deletion

Sample 41:
del(9)(p23pter)
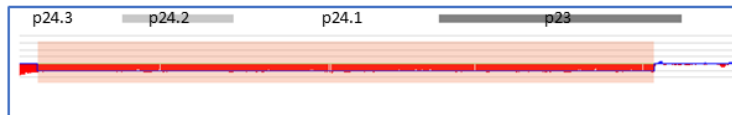
Large duplication

Sample 60:
8p23.1p12 gain
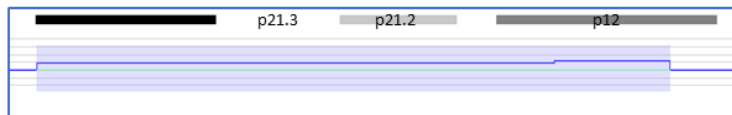
Figure S5 (continued 2/2). Representation of different chromosomal abnormalities.
This figure shows an example of each type of chromosomal aberration that was detected in this study. All aberrations shown are described in detail in Table S1.
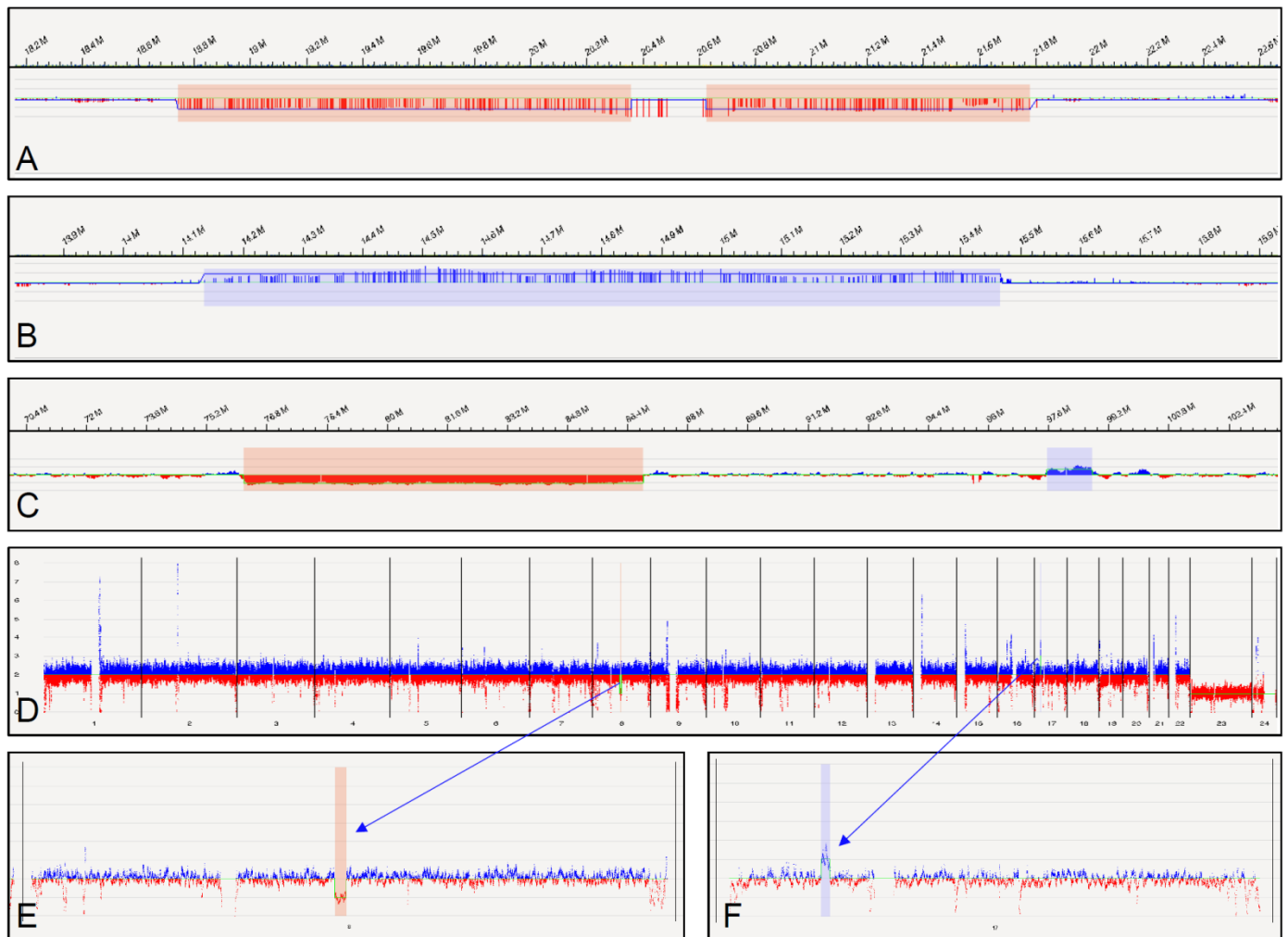
Figure S6. Representative CNV profiles obtained with optical genome mapping, for different samples.

A) Sample 2. Loss of 22q11.21(18645354_21465660). B) Sample 8. Gain of 17p12(14087934_15436895). C) Sample 70. Loss of 6q14.1q14.3(76385698_86884355), and gain of 6q16.1(97661978_98726638). D) Genome-wide CNV view (available in Bionano Solve v1.5) of sample 73 with E) chromosome 8 highlighted (showing a deletion) and F) chromosome 17 highlighted (showing a duplication). Blue: gains, Red: losses.
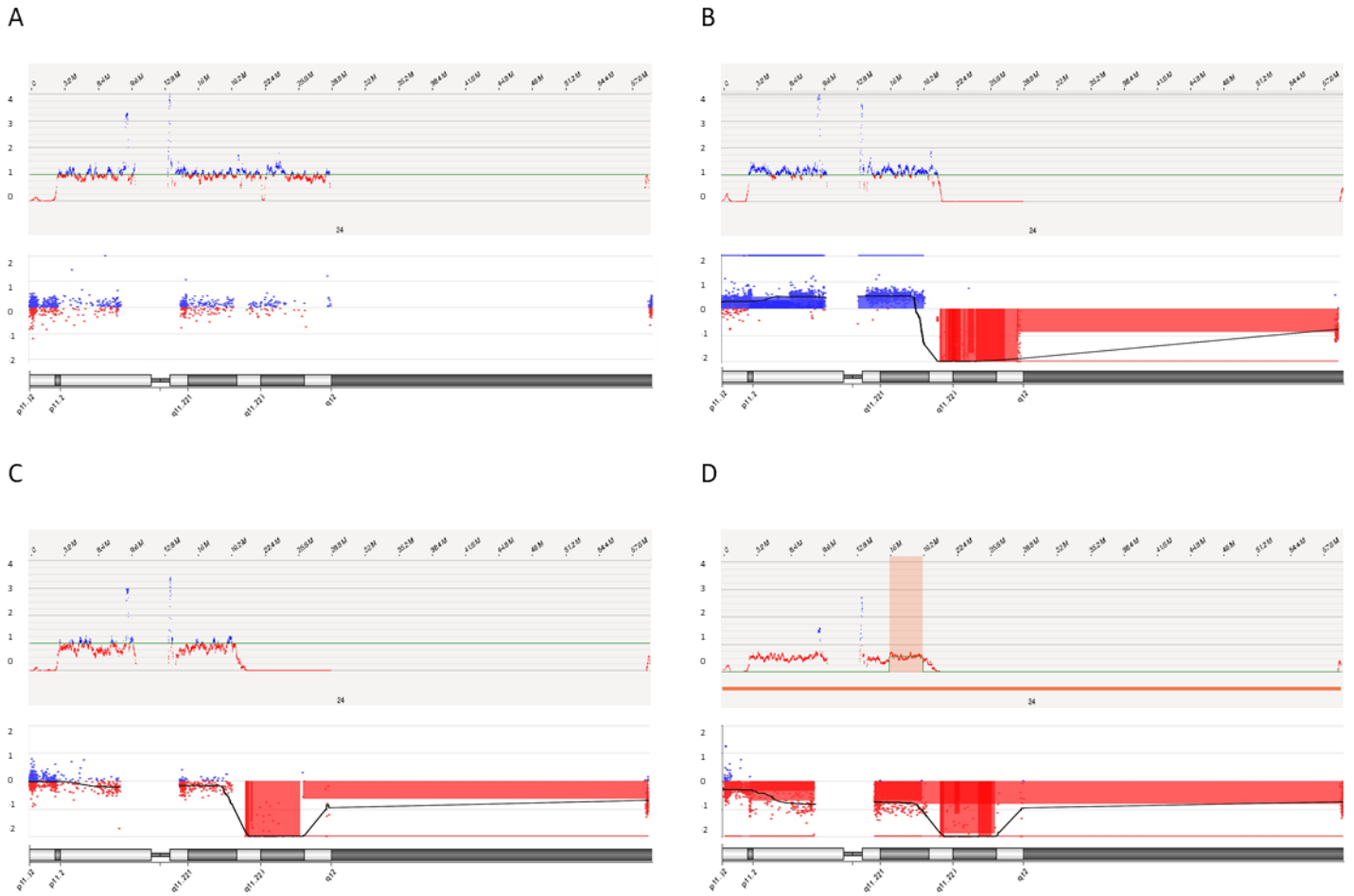
Figure S7: Isodicentric Y-chromosomes.

Optical genome mapping CNV profile (top) and CNV-microarray CNV profile (bottom) of Y-chromosomes of samples 80 (normal chrY, A), sample 55 (B), sample 57 (C) and sample 79 (D). The chromosomal position is given on top, the corresponding chromosome banding below each CNV-microarray profile. Sample 27 is not shown here because it was analyzed with a different CNV-microarray platform.
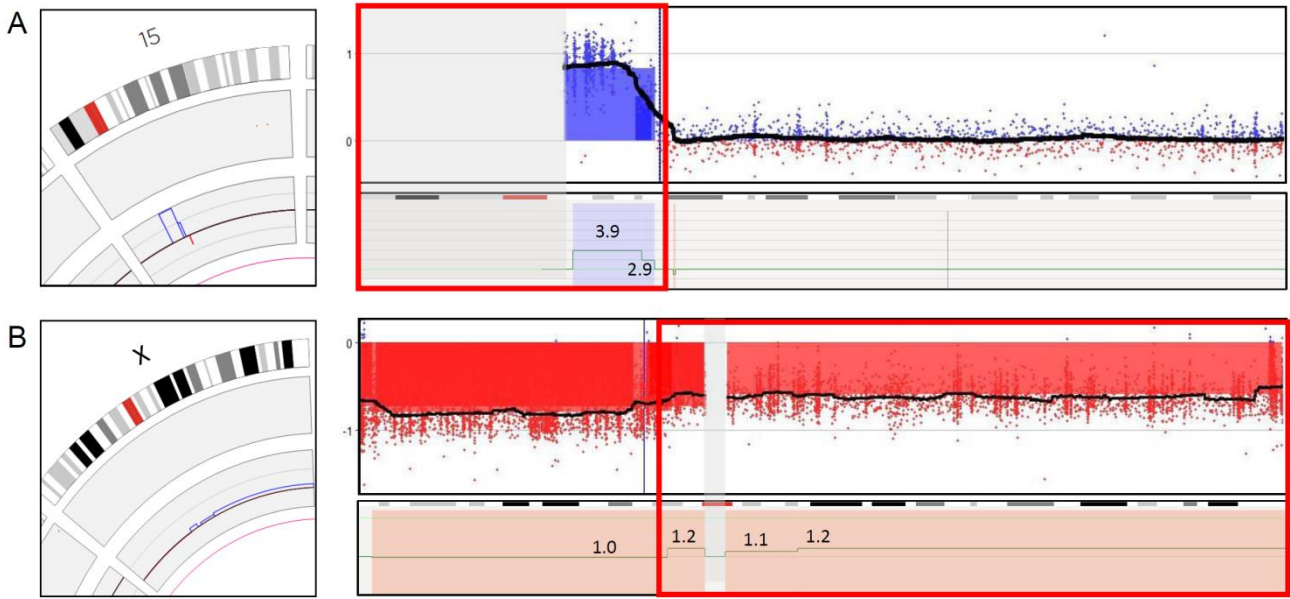
Figure S8. Optical genome mapping and CNV-microarray profiles for isochromosomes 15 and X.

A) Sample 77 (ish idic(15)(D15Z1+,SNRPN++,D15Z1+)). Left: Circos plot showing an abnormal CNV profile on chromosome 15. Top right: CNV-microarray data showing a gain on chr15. Bottom right: optical mapping data, showing a CNV profile that is nearly identical to the CNV-microarray profile. Numbers present fractional copy numbers. B) Sample 78 (46,X,idic(X)(p11.21)). Left: Circos plot showing a CNV baseline suggesting one copy of chromosome X (compared to the CNV line of chr22 partially shown on the left side). Additionally, the CNV profile shows a mosaic "gain" (compared to the baseline) on part of the chrX p-arm and the whole q-arm. Top right: CNV-microarray data showing a global loss on chrX (compared to a 46,XX control sample). However, the degree of loss varies within the chromosome consistent with a mosaic 45,X/46,X,idic(X)(p11.21) karyotype. Bottom right: optical mapping data showing a CNV profile that is nearly identical to the CNV-microarray profile. Numbers present fractional copy numbers. Red box shows parts of the chromosome 15 and X respectively that make up the iso-chromosomes. Grey box indicates the centromere (15 and X) and/or acrocentric p-arm (15).
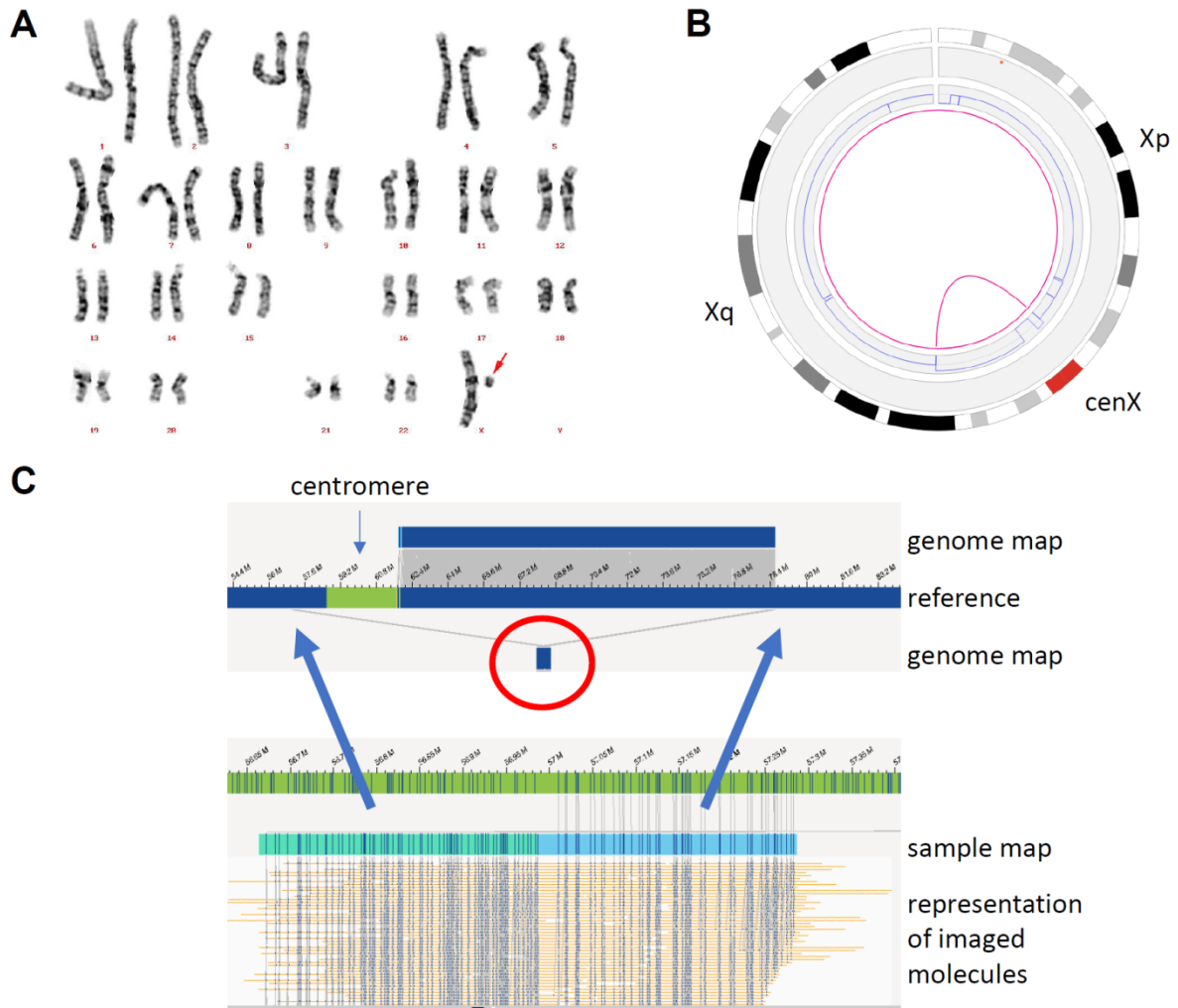
Figure S9. Small ring chromosome X.

A) Karyogram of sample 39. The red arrow is pointing towards the small X ring chromosome. B) Circos-plot (of chromosome X only) of sample 39. The pink line in the center of the circosplot is indicating the presence of the ring chromosome (called as an intrachromosomal translocation). C) Different genome maps (dark blue bars on top and below the reference) indicating the presence of the ring chromosome. The individual molecules for the genome map below the reference (highlighted by a red circle) are shown at the bottom of this figure. The left part of these molecules (light green bar) map to a region upstream of the centromere, whereas the right part of the same molecules (light blue bar) map to a region downstream of the centromere.
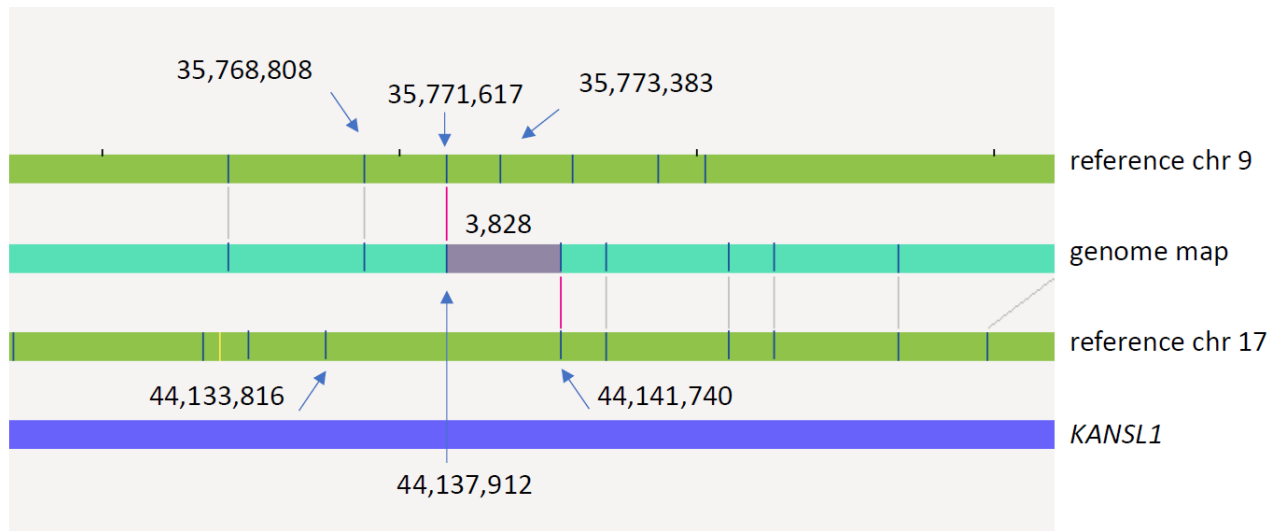
Figure S10. Optical genome mapping breakpoint detection for translocation t(9;17)(p13;q21), disrupting the gene *KANSL1*.

The two green bars represent the references of chromosomes 9 and 17, respectively. The mint bar in between represents the genome map of the translocation. The blue bar underneath represents the *KANSL1* gene. Small vertical black lines represent identified labels, and the red vertical lines indicate the translocation breakpoints, with an uncertain region of 3,828 bp in between shown in purple. The breakpoints are located between basepair-positions 35,771,617 and 35,773,383 on chromosome 9, and between 44,137,912 and 44,141,740 on chromosome 17.
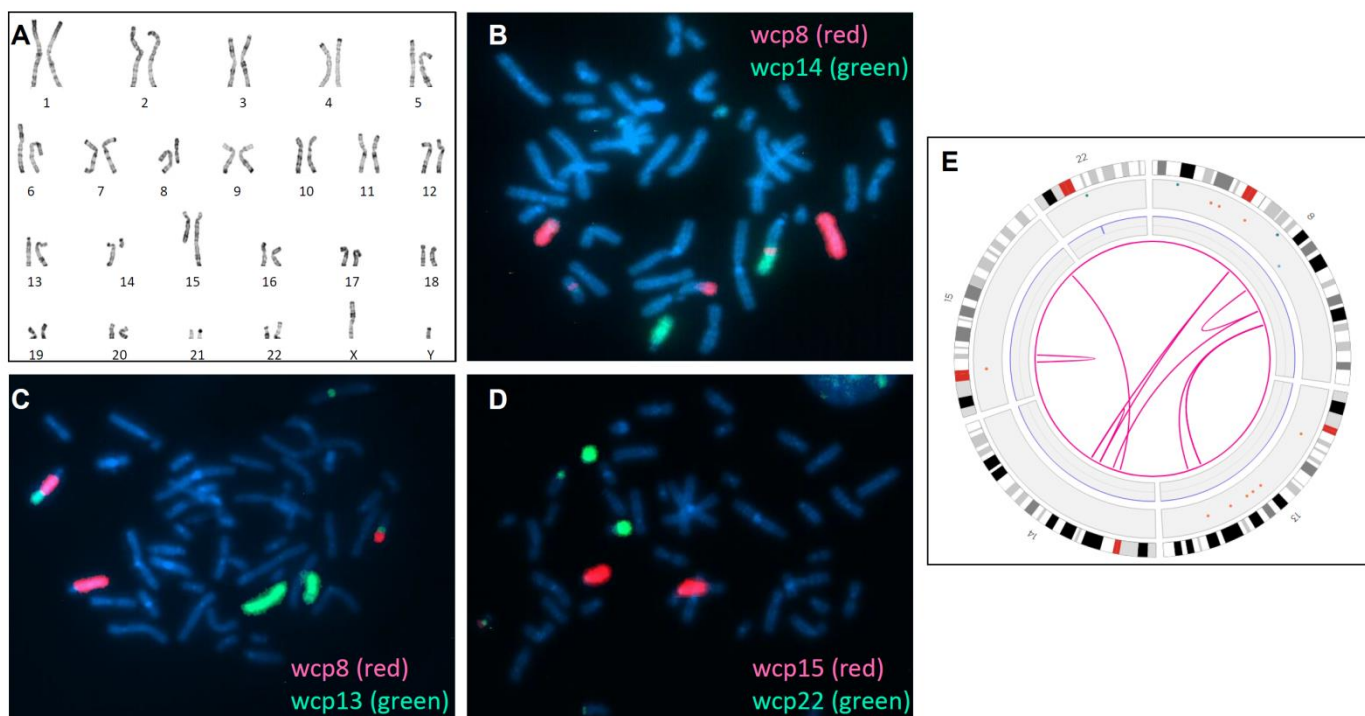
Figure S11. Complex sample 52.

A) Karyotype of sample 52, interpreted as 46,XY,der(8)t(8;22)(q12;q12),der(13)t(8;13)(q31;q23),der(14)t(14;15)(q11.2;q25),der(15)t(14;15)(q21;q24),der(22)t(13;22)(q31.1;p11.2). B) FISH of sample 52, using FISH probes wcp8 (red), wcp14 (green). C) FISH of sample 52, using FISH probes wcp8 (green), wcp13 (red). D) FISH of sample 52, using FISH probes wcp15 (green), wcp22 (red). E) Optical mapping circos plot, showing different translocations t(8;13), t(8;14), t(14,15), and intrachromosomal translocations on chr 8 and chr 15.
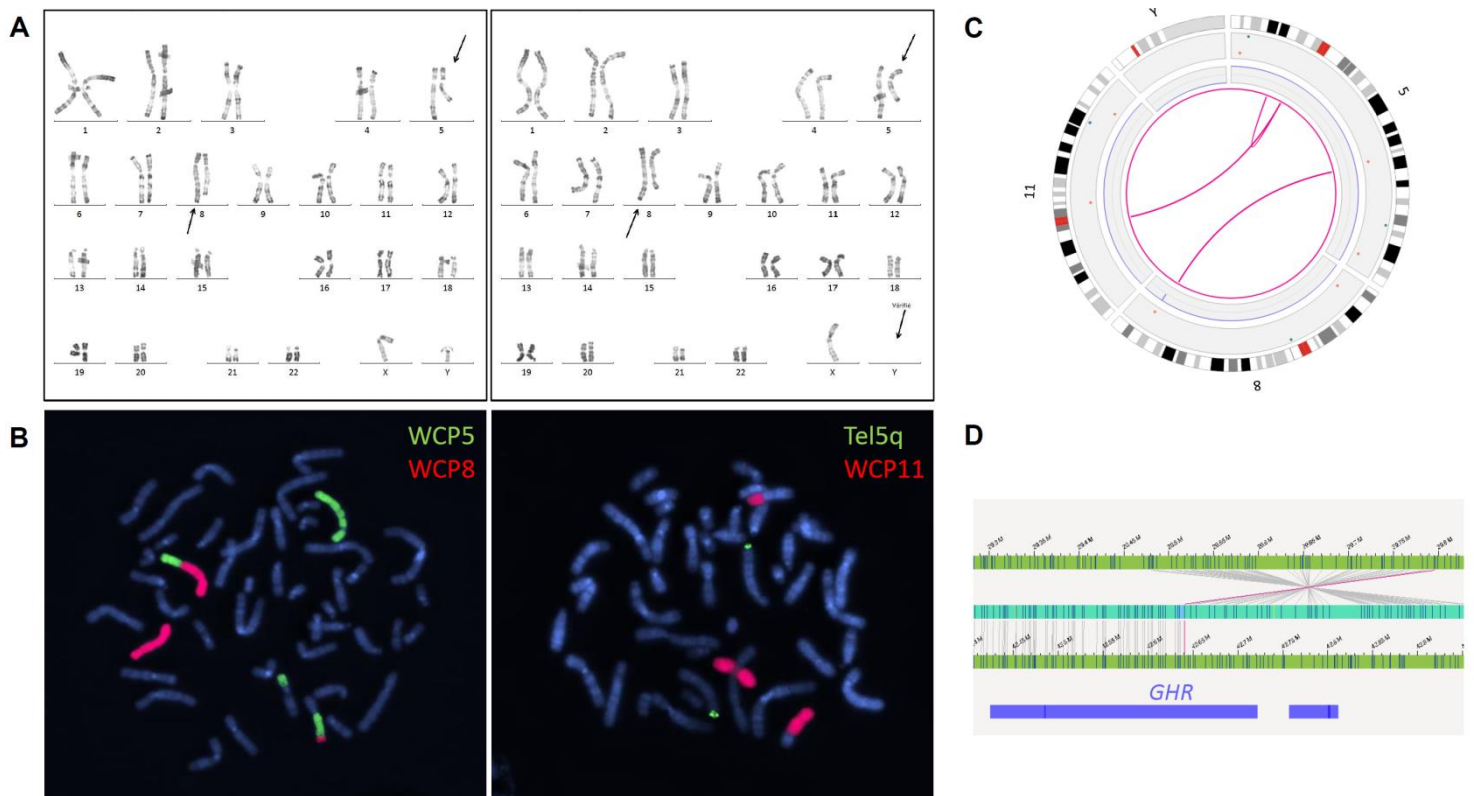
Figure S12: Complex sample 55.

A) Karyotype of sample 55, interpreted as 46,X,idic(Y)(q11.22),t(5;8)(q23;q24),t(5;11)(p12;p13)[32/50] /45,X,t(5;8)(q23;q24),t(5;11)(p12;p13)[10/50]/47,XY,idic(Y)(q11.22),t(5;8)(q23;q24),t(5;11)(p12;p13) [8/50]. B) FISH of sample 55, showing the translocations t(5;8) (left) and t(5;11) (right). C) Optical mapping circos plot of sample 55, showing the translocations t(5;8), t(5;11) and an intrachromosomal translocations 5. D) Bionano genome maps, showing the intrachromosomal translocation on chromosome 5, which is disrupting the gene *GHR*.

Table S1. Comparison of previous diagnostic findings with optical genome mapping results.

Provided in Excel sheet format

Table S2. Technical performance of optical genome mapping.

Sample	Starting material	Total DNA >=150 kbp (Gbp)	Avg N50 >=150kbp (Mbp)	Avg N50 >=20kbp (Mbp)	Map rate (>= 150 kbp)	Effective coverage (x)	Avg label density (per 100kbp)
1	EDTA blood	1 320	0,24	0,12	69%	286	15,69
2	EDTA blood	1 332	0,26	0,15	78%	323	15,19
3	EDTA blood	1 302	0,26	0,13	57%	233	17,88
4	EDTA blood	1 305	0,24	0,13	58%	243	17,14
5	EDTA blood	1 347	0,26	0,21	87%	368	15,20
6	EDTA blood	811	0,21	0,08	43%	110	15,28
7	EDTA blood	807	0,21	0,10	55%	140	15,34
8	EDTA blood	867	0,24	0,12	60%	163	15,19
9	EDTA blood	816	0,25	0,17	67%	167	15,38
10	EDTA blood	824	0,22	0,11	68%	175	15,25
11	EDTA blood	840	0,24	0,13	67%	176	15,40
12	EDTA blood	534	0,33	0,33	40%	69	16,16
13	EDTA blood	814	0,24	0,15	79%	200	15,19
14	Hep. Blood	801	0,24	0,14	66%	162	15,25
15	Hep. Blood	655	0,24	0,15	63%	124	14,40
16	Hep. Blood	821	0,22	0,14	72%	175	13,76
17	Hep. Blood	813	0,20	0,12	63%	149	13,27
18	Hep. Blood	823	0,20	0,12	64%	157	14,28
19	Hep. Blood	825	0,22	0,13	69%	169	14,21
20	Hep. Blood	817	0,19	0,12	65%	150	12,97
21	Hep. Blood	832	0,21	0,13	67%	162	13,83
22	Amniotic fluid	200	0,30	0,30	60%	38	15,40
23	Amniotic fluid	812	0,27	0,21	75%	177	13,47
24	Chorionic villi	826	0,25	0,18	73%	176	13,74
25	Amniotic fluid	850	0,26	0,20	72%	177	13,26
26	Amniotic fluid	802	0,24	0,18	88%	220	15,33
27	Amniotic fluid	801	0,30	0,22	85%	214	15,39
28	Amniotic fluid	849	0,26	0,17	78%	186	12,77
29	Chorionic villi	617	0,28	0,21	90%	173	14,96
30	Amniotic fluid	842	0,33	0,27	89%	242	14,72
31	Chorionic villi	820	0,27	0,21	88%	223	14,40
32	Amniotic fluid	817	0,33	0,26	90%	228	14,97
33	Chorionic villi	802	0,29	0,20	84%	219	14,97
34	Lymphoblastoid cell line	839	0,32	0,21	86%	224	15,15
35	Hep. Blood	814	0,27	0,21	87%	218	14,37
36	Hep. Blood	845	0,30	0,25	89%	233	14,91
37	Hep. Blood	809	0,25	0,10	71%	178	14,74
38	Hep. Blood	857	0,27	0,22	89%	233	14,37
39	Hep. Blood	855	0,25	0,17	86%	228	14,01
40	Hep. Blood	810	0,24	0,17	77%	189	12,92
41	Hep. Blood	811	0,21	0,13	73%	184	14,23
42	Hep. Blood	823	0,24	0,18	82%	208	14,19
43	EDTA blood	499	0,26	0,17	87%	135	14,50
44	EDTA blood	506	0,26	0,12	82%	130	15,11
45	EDTA blood	503	0,25	0,17	85%	134	15,27
46	EDTA blood	495	0,24	0,16	78%	122	14,96
47	EDTA blood	488	0,23	0,10	70%	107	15,58
48	EDTA blood	486	0,30	0,25	90%	137	14,60
49	Lymphoblastoid cell line	376	0,28	0,20	79%	95	15,79
50	Lymphoblastoid cell line	328	0,28	0,20	70%	73	16,38
51	Lymphoblastoid cell line	484	0,29	0,20	64%	98	15,74
52	Lymphoblastoid cell line	348	0,29	0,20	78%	87	15,71
53	Lymphoblastoid cell line	495	0,28	0,19	73%	115	15,42
54	Lymphoblastoid cell line	345	0,27	0,16	75%	82	15,25
55	EDTA blood	541	0,29	0,20	85%	144	14,47
56	EDTA blood	519	0,26	0,17	88%	142	14,51
57	EDTA blood	324	0,22	0,13	75%	75	14,78
58	EDTA blood	328	0,23	0,14	84%	85	15,10
59	EDTA blood	324	0,26	0,17	74%	75	15,17
60	EDTA blood	481	0,29	0,22	75%	113	15,27
61	EDTA blood	333	0,22	0,10	68%	71	15,05
62	EDTA blood	492	0,29	0,13	74%	113	15,90
63	EDTA blood	496	0,25	0,16	84%	129	15,04
64	EDTA blood	504	0,28	0,20	90%	143	15,44
65	EDTA blood	346	0,27	0,17	89%	95	15,13
66	EDTA blood	380	0,30	0,19	83%	102	16,28
67	EDTA blood	353	0,31	0,25	90%	100	16,83
68	EDTA blood	371	0,35	0,27	81%	96	17,17
69	EDTA blood	480	0,32	0,25	85%	129	16,46
70	EDTA blood	400	0,36	0,36	89%	111	14,24
71	EDTA blood	400	0,31	0,31	87%	109	14,47
72	EDTA blood	1 311	0,22	0,12	50%	208	14,88
73	EDTA blood	400	0,32	0,32	84%	105	15,75
74	EDTA blood	911	0,25	0,15	38%	109	22,36
75	EDTA blood	400	0,28	0,28	86%	108	15,45
76	EDTA blood	400	0,30	0,30	81%	102	15,78
77	EDTA blood	488	0,34	0,28	87%	134	14,82
78	EDTA blood	436	0,27	0,18	91%	122	15,06
79	EDTA blood	483	0,32	0,24	87%	132	14,92
80	EDTA blood	424	0,31	0,25	92%	120	15,06
81	EDTA blood	483	0,31	0,24	93%	140	15,62
82	EDTA blood	449	0,29	0,20	91%	127	15,56
83	EDTA blood	411	0,27	0,17	86%	110	15,94
84	EDTA blood	406	0,27	0,13	88%	110	15,29
85	Lymphoblastoid cell line	481	0,29	0,11	77%	117	15,30
Average		655	0,27	0,18	77%	152	15,13
Max value		1 347	0,36	0,36	93%	368	22,36
Min value		200	0,19	0,08	38%	38	12,77
Standard deviation		270	0,04	0,06	12%	59	1,21

Table S3. Overall numbers of variants per sample.

Provided in Excel sheet format

Table S4. SV and CNV statistics per size range.

a. Size distribution of SV calls for the cohort

		Number of rare SVs in each size category			
	Total rare SVs	0-20kb	20-100kb	100-500kb	>500kb
All chromosomes	6785 (100%)	5996 (88.4%)	423 (6.2%)	160 (2.4%)	206 (3.0%)

b. CNV size distribution in the total cohort (initial filter-setups: non-masked CNVs only and excluding those with CN=2)

		Number of CNV calls in each size category			
	Total	< 0.5Mb	0.5-1.0Mb	1-5Mb	>5Mb
All chromosomes	942 (100%)	471/942 (50%)	246/942 (26%)	174/942 (18%)	51/942 (6%)
Autosomes only	895 (100%)	463/895 (52%)	241/895 (27%)	162/895 (18%)	29/895 (3%)

c. CNV size distribution in the total cohort (additional filtering: excluding deletions with fractional CN >1.2 and duplications with CN < 2.8)

		Number of CNV calls in each size category			
	Total	< 0.5Mb	0.5-1.0Mb	1-5Mb	>5Mb
All chromosomes	366 (100%)	224/366 (61.2%)	61/366 (16.7%)	37/366 (10.1%)	44/366 (12%)
Autosomes only	324 (100%)	218/324 (67.3%)	57/324 (17.6%)	25/324 (7.7%)	24/324 (7.4%)

Table S5. Detection rate of chromosomal aberrations, by each pipeline, according to their type.

Aberration type		SV tool call	CNV tool call	Only called by one tool	Called by both tools	Global rate of detection
Aneuploidies (n=7)		NA	7/7 (100%)	7/7 (100%)	NA	100%
Unbalanced structural aberrations	Microdeletions and duplications (n=34)	27/34 (79%)	27/34 (79%)	14/34 (41%)	20/34 (59%)	100%
	Large CNV losses and gains (n=5)*	2/5 (40%)	5/5 (100%)	3/5 (60%)	2/5 (40%)	100%
	Unbalanced translocation (n=7)	4/7 (57%)	5/7 (51%)	3/7 (43%)	4/7 (57%)	100%
	Unbalanced insertions (n=2)	2/2 (100%)	2/2 (100%)	0/2 (0%)	2/2 (100%)	100%
	Ring chromosomes (n=1)	1/1 (100%)	1/1 (100%)	0/1 (0%)	1/1 (100%)	100%
	Isochromosomes (n=6)	0/6 (0%)	6/6 (100%)	6/6 (100%)	0/6 (0%)	100%
	complex chromosomal rearrangements (n=4)	4/4 (100%)	2/4 (50%)	2/4 (50%)	2/4 (50%)	100%
Balanced structural aberrations	Balanced translocation (n=27)	27/27 (100%)	NA	27/27 (100%)	NA	100%
	Inversions (n=6)	6/6 (100%)	NA	6/6 (100%)	NA	100%

*this includes terminal aberrations that are currently not called by the SV pipeline

NA: not applicable

Table S6. Comparison of capabilities and limitations of optical genome mapping versus conventional cytogenetic methods, karyotyping and CNV-microarray, in detecting different types of chromosomal aberrations.

Aberration type	Detection of aberration			Localization of aberration		
	Standard of care genome-wide tests		Optical genome mapping	Standard of care genome-wide tests		Optical genome mapping
	CNV-microarray	Karyotyping		CNV-microarray	Karyotyping	
	aberrations larger than 20-200 kb	aberrations larger than 7-10Mb	aberrations >few kb	aberrations larger than 20-200 kb	aberrations larger than 7-10Mb	aberrations >few kb
Aneuploidy	✓	✓	✓	✗	✓	✓
Unbalanced Robertsonian translocation	✓	✓	✓	✗	✓	✗
Balanced Robertsonian translocation	✗	✓	✗	✗	✓	✗
Unbalanced reciprocal translocation	✓	✓	✓	✗	✓	✓
Balanced reciprocal translocation	✗	✓	✓	✗	✓	✓
Deletion	✓	✓	✓	✗	✓	✓
Insertion	unbalanced only	✓	✓	✗	✓	✓
Duplication	✓	✓	✓	✗	✓	✓
Inversion	✗	✓	✓	✗	✓	✓
Isochromosome	✓	✓	✓	✗	✓	✓
Ring chromosome	unbalanced only	✓	✓	✗	✓	✓

Table S7. Concordance of all filtered SVs and CNVs between optical genome mapping and CNV-microarray.

Sample ID	Bionano call						CNV-microarray call					Result[a]
	Bionano tool	Chr	Start	End	Size (bp)	Type	Chr	Start	End	Call size or # of probes	LogR	
56	SV	4	161863865	161923311	59447	duplication	4	161869551	161869610	1 probe	0,5	Supported
	SV	10	81438916	81594303	137161	deletion	-	-	-	No probes	-	Inconclusive
	SV	16	34462037	34556454	94418	duplication_inverted	16	34482042	34482101	2 probes	0,5	Supported
	SV	16	34462037	34578881	116845	duplication_inverted	16	34565215	34565274		0,5	Supported
	SV	16	79404969	79430008	20599	deletion	16	79420905	79420964	1 probe	-0,76	Supported
62	SV	1	49908206	50001818	84129	deletion	1	49959780	49959839	1 probe	-0,92	Supported
	SV	1	246313442	246520313	206871	deletion	1	246321447	246501692	180245 bp	mean='-1.12	Confirmed
	SV	2	233239437	233288334	27988	deletion	2	233270613	233270672	1 probe	-0,77	Supported
	SV	5	120421707	120472623	43718	deletion	-	-	-	No probes	-	Inconclusive
	SV	9	71845510	71961535	116026	duplication	9	71846678	71846737	3 probes	0,55	Supported
							9	71917699	71917758		0,39	
							9	71957664	71957723		0,57	
	SV	18	65845905	65907488	49589	deletion	18	65891840	65891899	1 probe	-0,54	Supported
64	CNV	7	76109311	76623844	514533	deletion	7	76139282	76558024	418743 bp	mean=-0.85	Confirmed
	CNV	8	2120483	2349321	228838	deletion	8	2206158	2206217	2 probes	0,16	Not supported
							8	2308926	2308985		-0,07	
	SV	4	138246546	138337696	91151	duplication	4	138259451	138259510	2 probes	0,42	Supported
							4	138324113	138324172		0,59	
	SV	9	32737531,7	32787627,8	36735	deletion	9	32759696	32759755	1 probe	-1,02	Supported
77	CNV	X	6503942	8131734	1627792	duplication	X	6488721	8097511	1608791 bp	mean=0.51	Confirmed
	SV	4	139888286	139919269	23983	deletion	-	-	-	No probes	-	Inconclusive
	SV	11	85146480	85175179	23022	deletion	-	-	-	No probes	-	Inconclusive
	SV	14	86454708	86498388	31877	deletion	-	-	-	No probes	-	Inconclusive
78	SV	1	22308530	22333055	23125	deletion	-	-	-	No probes	-	Inconclusive
	SV	2	55920504	55948466	27963	duplication	-	-	-	No probes	-	Inconclusive
	SV	2	131031140	131977013	74897	deletion	-	-	-	10 probes	mean=-0.01	Not supported
	SV	2	241651052	241705013	53962	duplication	2	241687451	241687509	1 probe	0,42	Supported
	SV	3	100340351	100451628	111278	duplication	3	100360692	100360751	2 probes	0,54	Supported
							3	100419565	100419624		0,61	
	SV	3	165014055	165083895	42602	deletion	3	165065420	165065479	1 probe	0,17	Not supported
	SV	16	78037966	78081307	36813	deletion	-	-	-	No probes	-	Inconclusive
79	SV	2	130790519	131181302	390783	deletion	2	130928976	130929035	3 probes	0,08	Not supported
							2	131015497	131015556		-0,03	
							2	131084449	131084508		-0,05	
	SV	7	6118254	6146732	21897	deletion	7	6146033	6146092	1 probe	-0,06	Not supported
	SV	9	105811280	105874071	55238	deletion	-	-	-	No probes	-	Inconclusive
	SV	10	5664747	5716793	52047	duplication	10	5684186	5684245	1 probe	0,71	Supported
	SV	10	12102683	12140526	37844	duplication	-	-	-	No probes	-	Inconclusive
	SV	11	49055393	49105125	26589	deletion	-	-	-	No probes	-	Inconclusive
	SV	13	38066960	38129134	50292	deletion	13	38106845	38106904	1 probe	-1,03	Supported
	SV	14	74571941	74597810	25870	duplication	-	-	-	No probes	-	Inconclusive
	SV	14	106601449	106784215	182767	duplication	14	106665119	106665178	3 probes	0,5	Supported
							14	106716345	106716404		0,59	
							14	106774613	106774672		0,67	
	SV	15	86274566	86553109	278544	duplication_split	15	86288646	86542460	253814 bp	mean=0.55	Confirmed
	SV	X	23907510	23934260	26751	duplication	X	23914694	23914753	1 probe	0,72	Supported

[a] confirmed= the event is called by CNV-microarray (>=5 consecutive deviating probes with a LogR>0.25 for duplication and <-0.25 for deletion), supported= less than 5 probes but consistent LogR deviation (threshold>0,25 for dup <-0,25 for del), inconclusive = uncovered region (no probes in the microarray at the respective region), Not supported = CNV-microarray and OM results are not matching