Cas9-directed long-read sequencing to increase the resolution of optical genome mapping to single basepair level

Eddy N. de Boer, Vincent Vroom, Arjen J. Scheper, Lennart F. Johansson, Laura Bosscher, Nettie Rietema, Sabrina Z. Commandeur, Nine V.A.M. Knoers, Birgit Sikkema-Raddatz, Cleo C. van Diemen and Eva van den Berg

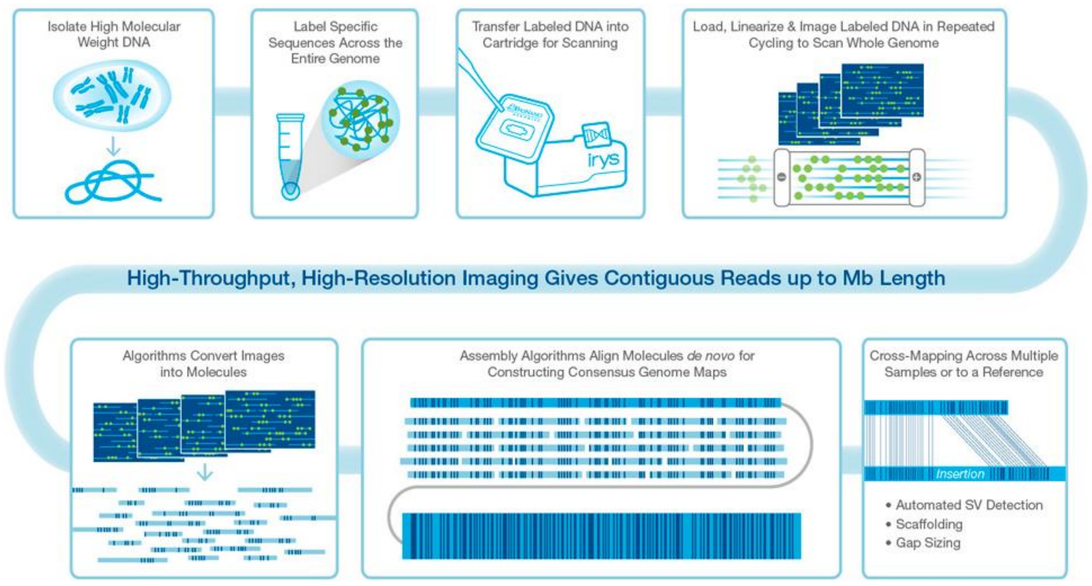
Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Introduction

Innovations in leukemia diagnostics

Leukemias are clinically and genetically heterogeneous. The type of genetic aberration is important for progression prognosis and choice of treatment. Karyotyping, SNP-Array and FISH are the standard-of-care (SOC) methods in leukemia diagnostics. Optical genome mapping (OGM) is a technique that can detect different types of structural aberrations and is able to replace the SOC methods in one technique.

Bionano Optical Genome Mapping



OGM can detect much smaller aberrations than the SOC methods (OGM 500bp vs. Karyotyping ~10Mb). However, the OGM resolution may be too low in specific situations:

- OGM is not able to define the location of deletions and duplications between two labels.
- OGM breakpoints are disputable when the labels are not distinctive enough.

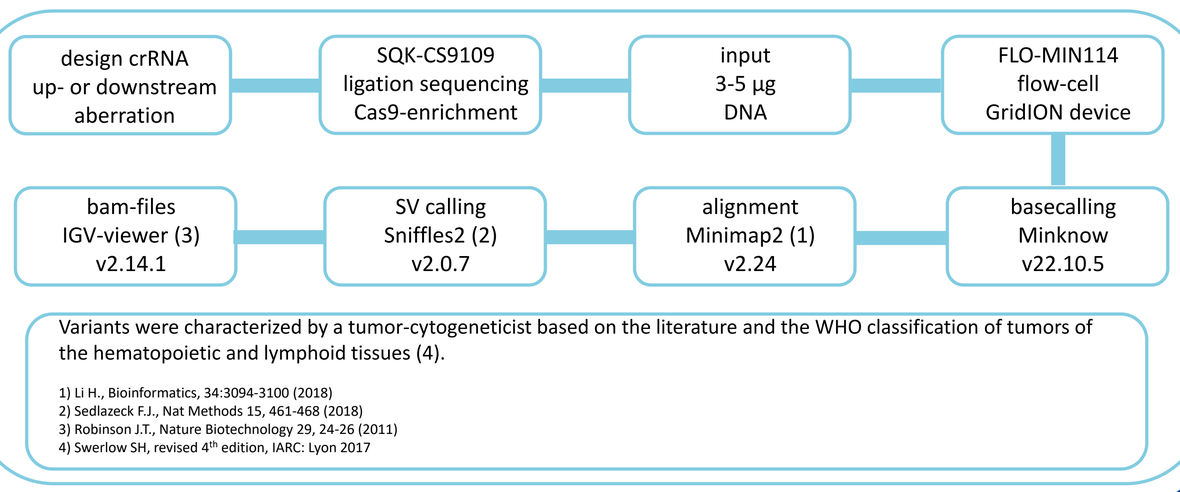
Aim: In this proof of principle study, we tested whether Cas9-directed Nanopore sequencing can increase the OGM resolution to single basepair level in order to improve classification

Method Cas9-directed nanopore sequencing

In an OGM validation study using a selected group of leukemia (bone-marrow) samples with different types of aberrations included, most of the OGM detected aberrations were confirmed with the SOC methods. However, some aberrations could not be confirmed with the SOC methods due to a lack of resolution. These aberrations were used for this study (see table below).

Sample No.	OGM aberration	Detection problem
Sample1	OGM[GRCh37]17p13.1(7,545,377_7,588,071)x1	aberration located in between 2 OGM labels
Sample2	OGM[GRCh37]17p13.1(7,545,377_7,588,071)x1	aberration located in between 2 OGM labels
Sample3	OGM[GRCh37]17p13.1(7,545,377_7,588,071)x1	aberration located in between 2 OGM labels
Sample4	OGM[GRCh37]inv(3)(q25.33;q26.2)(160,014,744_168,882,939)	resolution SOC too low
	OGM[GRCh37]3q25.33(159,902,689-160,014,744)x1	resolution SOC too low
	OGM[GRCh37]3q26.2(168,882,939-168,907,480)x1	resolution SOC too low
Sample5	OGM[GRCh37]inv(15)(q24.1;q24.1)(72,959,741_74,362,190)	resolution SOC too low
	OGM[GRCh37]3p13(71,086,423_71,375,386)x1 (3% allele frequency)	resolution SOC too low
	OGM[GRCh37]t(14;17)(q32.33;q25.3) (106,249,815;80,915,618)	disputable OGM breakpoint
Sample6	OGM[GRCh37]ins(12p13.2)(11,889,007_11,895,594)	aberration located in between 2 OGM labels

Workflow Cas9-directed Nanopore sequencing.



Variants were characterized by a tumor-cytogeneticist based on the literature and the WHO classification of tumors of the hematopoietic and lymphoid tissues (4).

1) Li H., Bioinformatics, 34:3094-3100 (2018)
2) Sedlazeck F.J., Nat Methods 15, 461-468 (2018)
3) Robinson J.T., Nature Biotechnology 29, 24-26 (2011)
4) Swerlow SH, revised 4th edition, IARC: Lyon 2017

Conclusion & Discussion

- All 10 OGM aberrations were confirmed with Cas9-directed Nanopore sequencing.
- The estimated OGM breakpoints are localized 0.2-5.5Kb of the redefined breakpoints.
- Cas9-directed Nanopore sequencing can fill the gap of low-resolution areas of OGM.
- Increasing the resolution to single basepair level improves the classification of variants.
- The turn-around-time of Cas9-directed Nanopore sequencing can be limited to 1 or 2 working day(s) by designing and validating guide-RNAs for known hotspots in advance.

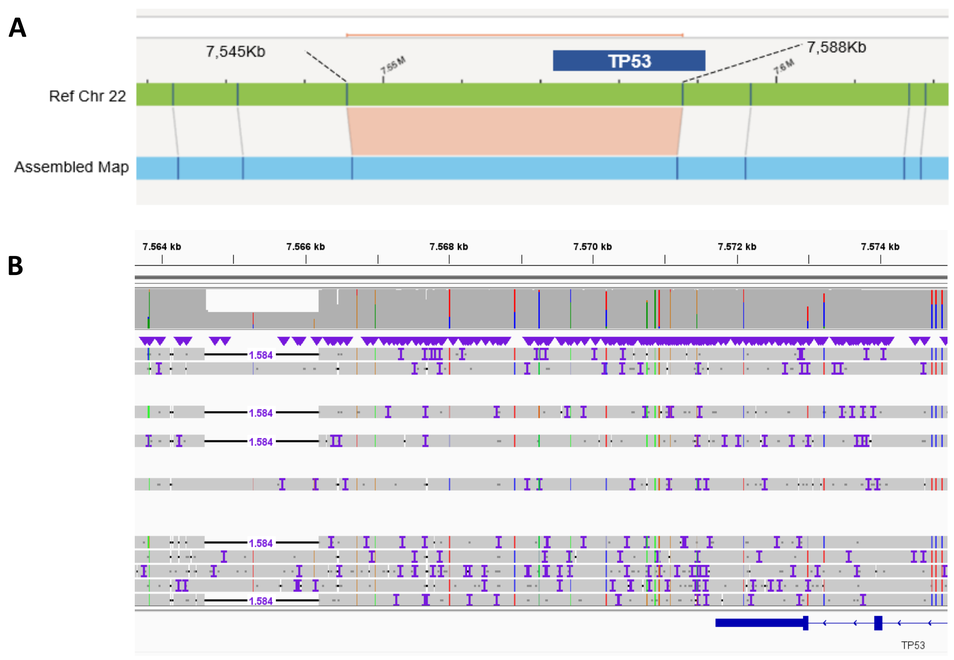
Results Cas9-directed Nanopore sequencing

Representative example of finding where SOC resolution is too low.

(A) In sample4 OGM detected an 8.9Mb *MECOM* included inversion of chromosome 3 (q25.33;q26.2) while two adjacent 112 kb (q25.33) and 25Kb (q26.2) regions were deleted.
(B) Cas9-directed Nanopore sequencing redefined the breakpoints to 168,888,415 (159,905,686 (mis-matches)) and 160014965 (168907304 (mis-matches)).
The inversion was redefined to 8.9Mb (160,014,965-168,888,415) and the deletions to 109Kb (159,905,686-160,014,965) and 19kb (168,888,415-168,907,304). The 8.9Mb inversion includes the complete *MECOM* region but does not reposition *MECOM* (q26.2) towards the *GATA2* enhancer (3q21) which drives the *MECOM* expression.
This aberration is characterized as clinically not relevant.

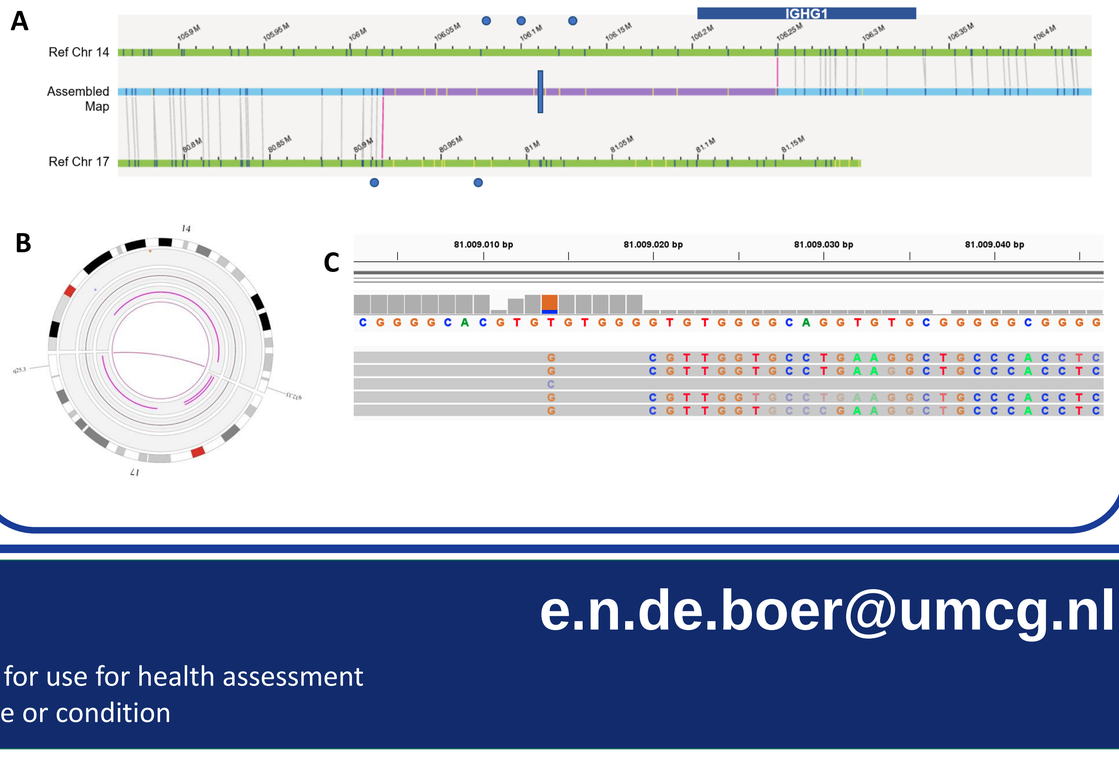
Representative example of an aberration located between two OGM-labels.

(A) In sample 1-3 OGM detected a heterozygous 1.2Kb deletion in a 42Kb region (17p13.1) between two labels including *TP53*.
(B) Cas9-directed Nanopore sequencing redefined the aberration to a 1.6Kb deletion upstream of *TP53*.
The aberration, a known CNV (1000 genomes, Nature 526, 68-74 (2015)), is characterized as clinically not relevant.



Representative example of an OGM aberration with a disputable breakpoint.

(A&B) In sample 5 OGM detected an *IGHG1* included translocation between chromosome 14 and 17 (q32.33;q25.3) as shown in the circos plot and the alignment of the assembly. We evaluated the uncertain OGM breaking-region (purple area, A) and designed crisprs (blue dots, A). Cas9-directed Nanopore sequencing confirmed the aberration and redefined the breakpoints (blue vertical line, A).
(C) The chromosome 14 breakpoint is located 88Kb upstream of *IGHG1*. *IGHG*1 is not affected.
This aberration is characterized as clinically not relevant.



university of groningen
umcg
e.n.de.boer@umcg.nl