Optical Genome Mapping for detecting Homologous Recombination Deficiency (HRD) in human breast cancers.

Sandra Vanhuele[1], Youlia Kirova[2,3], Anne-Sophie Hamy-Petit[4], Audrey Rapinat[5], David Gentien[5], Fabien Reyal[4,6], Anne Vincent-Salomon[7], Alexandre Eeckhoutte[1], Manuel Rodrigues[1,8], Tatiana Popova[1], Marc-Henri Stern[1,*]

Affiliations:

[1]Inserm U830, DNA Repair and Uveal Melanoma (D.R.U.M.) team, Institut Curie, PSL Research University, Paris, France

[2]Department of Radiation Oncology, Institut Curie, Paris, France

[3]UVSQ, Paris-Saclay University, Saint Cloud, Paris, France

[4]Department of Surgery, Institut Curie, Paris, France

[5]Genomics Platform, Translational Research Department, Research Center, Institut Curie, PSL Research University, Paris, France

[6]Residual Tumor & Response to Treatment Laboratory (RT2Lab), Translational Research Department, INSERM U932, Institut Curie, Paris, France

[7]Department of Pathology, Institut Curie, Paris, France

[8]Department of Medical Oncology, Institut Curie, Paris, France

[1]Department of Diagnostic and Theranostic Medicine, Institut Curie, Paris, France

Running title: optical genome mapping for HRD detection

Keywords (5) Optical genome mapping, HRD, homologous recombination deficiency, BRCA1, BRCA2, triple negative breast cancer

*Corresponding author: Marc-Henri Stern, Inserm U830, Institut Curie, 26 rue d'Ulm, 75248 Paris cedex 05, France. marc-henri.stern@curie.fr

Abstract

Homologous recombination deficiency (HRD) leads to genomic instability that marks HRD tumor genome with a specific genomic scar. Present in many cancers, HRD is important to be detected as it is associated with a hyper-sensitivity to some classes of drugs, in particular the PARP inhibitors. Here, we investigate the use of structural variants detected by the Optical Genome Mapping (OGM) technology to identify HRD tumors. We first compared the performance of OGM and whole genome sequencing (WGS) in an HRD triple negative breast carcinoma (TNBC) carrying a germline *BRCA2* deleterious mutation. We showed the excellent performance of OGM and its ability to recognize subclonal events not detected by WGS. We then analyzed by RVA OGM data from fifteen TNBC samples from the clinical trial RadioPARP. We defined two features characteristic of HRD. Tandem duplication (TD) and insertion events were found increased in HRD tumors. We showed that insertion calls were probably mostly TD too small to be called as TD by RVA. The insertion/TD feature fully discriminated HRD from all homologous recombination proficient (HRP) TNBC but one. This outlier carried a *CCNE1* amplicon probably explaining the excess of insertion/TD. Total numbers of translocations were similar in HRP and HRD TNBC. We suggested a novel feature, translocations and intra-fusions isolated from another event by 3 megabases. Isolated translocations and intra-fusions perfectly discriminated HRD from HRP TNBC.

Our results demonstrate that the OGM technology is an affordable way of getting an insight of the structural variants present in solid tumors, even with low tumoral cellularity. It represents an alternative technology for HRD diagnosis, which should now be evaluated in independent series of tumors of different tissue origins.

Introduction

Our cells are constantly exposed to endogenous and exogenous damage that alters DNA. Double Strand Breaks (DSB) of the DNA are the most toxic type of damage as they will lead to aneuploidy if left unrepaired before cell division. This repair can be performed using several different pathways (Ceccaldi et al. 2016; Scully et al. 2019), but the Homologous Recombination (HR) pathway is the only one able to repair *ad integrum* the genome, since it utilizes an intact copy of the broken genome region as a template, *i.e.* the sister chromatid after replication or the homologous chromosome during meiosis. Key actors in the HR pathway include BRCA1, BRCA2 and PALB2. These large proteins bind to the ends of the broken DNA to deposit on the single stranded DNA a microfilament made up of RAD51 and RAD51 paralogues. This microfilament will recognize and invade the homologous sequence to copy it. The continuity of the broken DNA will be eventually restored after resolution of the chromatid or chromosome bridges (Holliday junctions). Three other main mechanisms of repair of DSBs are the Non-Homologous End Joining (NHEJ), a process that is tolerant to nucleotide changes and insertion or deletion of bases (indel); Alternative-End Joining (Alt-EJ) / Microhomology-mediated end joining (MMEJ), a process that induces systematic small deletions with microhomologies: and Single Strand Annealing (SSA), which induces large deletions.

Homologous recombination deficiency (HRD) is an important feature to be recognized in cancer (Groelly et al. 2022). Many possible mechanisms may explain such deficiency. The main causes are mutations of key HR actors, most frequently *BRCA1* or *BRCA2*, but *PALB2, RAD51* paralogs (*RAD51C, RAD51B* and possibly *RAD51D*) mutations also play a significant role. These mutations could be inherited in hereditary breast and ovarian cancer syndrome (HBOC) or somatic. The second wild-type allele of the respective HR gene is inactivated in HRD tumors by deletion of the region, or more rarely by somatic mutations. Epigenetic inactivation of HR genes, mainly *BRCA1* and *RAD51C,* by hypermethylation of the promoter region of these genes is also a frequent mechanism of HRD. HRD is found in approximately half of triple negative (hormone receptors negative, no over-expression of HER2) breast

carcinomas (TNBC) and of high-grade serous ovarian carcinomas, and in many other solid tumors at lower frequencies (Riaz et al. 2017; Knijnenburg et al. 2018).

Discovered in cellular models (Bryant et al. 2005; Farmer et al. 2005), the hypersensitivity of HRD tumors to PARP inhibitors was demonstrated in clinical trials (Mirza et al. 2016; Ray-Coquard et al. 2019). Thus, establishing the HR status of tumors, especially for high-grade serous ovarian carcinomas, is highly clinically relevant. Different methods have already been developed to diagnose HRD. Since this deficiency is due to the inactivation of HR genes target-sequencing of these genes is the most direct approach. Despite its large use in clinics, some limitations were identified, such as interpretation of variants of unknown significance and of variants in genes with more distant role in HR (for example *ATM*, *CDK12*…), the lack of detection of *BRCA1* and *RAD51C* methylation, which are major players of HRD, and of course the inability to detect HRD of unknown origin. The current strategy consists in measuring patterns of somatic alterations of the cancer genome (mutations or structural rearrangements) directly due to the DNA repair defect. The first generation of HRD signatures, extracted from SNP-arrays, includes large genomic changes (LST; Large Scale Transition), the amount of Loss of Heterozygosity (LOH) as well as the Telomeric Allelic Imbalance (TAI) (Abkevich et al. 2012; Birkbak et al. 2012; Popova et al. 2012). Those signatures are presently the FDA-approved tests used in clinics. Large sequencing panels or Whole Genome Sequencing (WGS) allow the measure of the Single Base Substitution (SBS) signature 3 shown to be associated with *BRCA1/2* deficient tumors (Polak et al. 2017). WGS allows not only the recognition of SBS3, but also the exhaustive description of rearrangement signatures, some of them being strongly associated with HRD, such as RefSig R5 characterized by deletions smaller than 100kb. RefSig R3, characterized by a number of tandem duplications (TD) smaller than 100kb has also been shown to be associated specifically with *BRCA1*-mutated tumors (Nik-Zainal et al. 2016; Davies et al. 2017). This WGS approach and the associated diagnostic pipeline HRDetect are highly performant but remain costly in terms of sequencing and data storage, and demanding in terms of bioinformatics.

We here explore Optical Genome Mapping (OGM) as an alternative from WGS for its ability to detect HRD. OGM is an affordable genome-wide visualization technique able to detect structural rearrangements. It is non sequencing based and does not employ mechanical forces to destroy the DNA. Molecules larger than 150kb are directly extracted and labeled with fluorophore tags approximately 15 times every 100 kb. Labelled DNA is then linearized on a chip where nano channels are engraved in order to image unique large DNA molecules on the Saphyr genome imager. Molecules are then aligned to gather to create a consensus optical map that is automatically compared to a reference map in a genome wide fashion and for which any deviation in the labelling pattern or molecule coverage indicates the presence of a structural variant (SV) or a copy number variant (CNV). The entire genome has a high coverage, usually beyond 300X, allowing to detect even rare variants. Evaluation of patterns compared to a reference is then used to detect structural variants.

We first compared OGM and high-coverage WGS in one *BCRA2*-inactivated Triple Negative Breast Cancer (TNBC; negative for estrogen and progesterone receptors and not overexpressing HER2). We then studied a TNBC clinical cohort for which the HRD status was known.

Materials and Methods

Whole Genome Sequencing

Whole Genome Sequencing (WGS) was performed with Illumina NovaSeq 6000 paired-end technology and selected read length was 2 x 150bp. Resulting fastq files were aligned to the reference genome hg38 using Burrows-Wheeler Aligner v0.7.15. Duplicates were marked with Sambamba and removed with Samtools. Resulting bam files were analyzed together for structural variants using Delly v0.9.1 and Manta v1.6(Rausch et al. 2012; Chen et al. 2016). Both pipelines used a somatic filtering to filter out germline variants present in the germline. Resulting vcf files were analyzed using R and the library tidyverse.

Optical Genome Mapping (OGM)

Optical genome mapping (OGM) requires at least 5mg of frozen tissue in order to extract at least 750ng of genomic DNA (gDNA). High-molecular weight DNA was extracted by the Bionano Prep SP Tisue Kit following the manufacturer' instructions. DNA labeling was performed according to the DLS protocol with the DLE1 enzyme (CTTAAG sequence). The Saphyr Chip linearizes the labelled molecules and guides them into nanochannels to be imaged. Three samples were analyzed simultaneously for 3 days in order to obtain the highest possible coverage (Supp Table). Rare Variant Analysis (RVA) was performed using the Access software version 1.7 and solve version 3.7. The limits of detection of RVA are SVs > 5kb down to 5% allele fraction). A copy number variation tool running in parallel is able to detect CNVs > 500kb down to 10% allele fraction. Smap files generated by Bionano's pipelines were downloaded and data analyses were performed with R using the tidyverse library.

Data and statistical analyses

To compare breakpoints of structural variants between two different methods we used the following distance:

Dist = |Astart – Bstart| + |Aend– Bend|

where Astart is the starting chromosomal position of the variant detected in method A and Aend the ending chromosomal position of the same variant, and the same annotation is used for method B.

We filtered out all the variants detected by Delly or Manta supported by less than 10 reads, as well as deletions, duplications, insertions and inversions smaller than 1kb. Structural variants of the same type with distance between their breakpoints of less than 6kb (below the resolution of OGM) were combined as a same event. For Delly and Manta comparison, SVs were considered equivalent if they have the same type, and if their distance is smaller than 10 for precise variants supported by discordant split reads, and smaller than 510bp for imprecise variants supported only by discordant paired reads.

Variants detected by OGM and present in the control database of Bionano were filtered out. When comparing OGM and WGS, the threshold of maximum distance to consider an event equivalent used is 50kb.

The Wilcoxon Mann-Whitney nonparametric test was used in order to assess significant differences between the two groups of samples. P values < 0.05 were considered significant. The significance of the p values is annotated on graphs as * if $p < 0.05$, ** if $p < 0.005$ and *** if $p < 0.001$.

Results and Discussion

Comparison of OGM and WGS to detect structural variants in tumor DNA

In order to evaluate OGM results we compared the structural variants detected by the RVA to those detected by WGS in a TNBC occurring in a *BRCA2*-mutation carrier. This sample was first analyzed with OGM, without its paired control at a coverage of 984X. Using the VAF of the detected variants, we estimated the tumor cellularity of this sample to be around 60%. Using RVA method we detected 46 deletions, 9 duplications (3 tandem, 2 inverted and 4 split duplications), 9 insertions, 5 inversions, 49 translocations et 27 intra-fusions. This tumor sample was also sequenced with WGS along with its paired blood sample. We obtained a mean coverage of 83X for the tumor sample and 50X for the germline sample. Using the VAF of the detected variants, we estimated the tumor cellularity of the sample around 40%. We used two methods for calling structural variants (SVs): Delly and Manta. Delly detected 24 translocations, 27 deletions, 2 duplications, 10 inversions and no insertion, whereas Manta detected 75 translocations, 58 deletions, 11 duplications, 34 inversions and no insertion. All the SVs detected by Delly were found by Manta, except for one inversion.

We compared the SVs detected by OGM using RVA with those detected by WGS using Delly or Manta (Figure 1). Delly detected 13 out of 49 translocations, 17 out of 46 deletions,1 out of 9 duplications, none of 9 insertions and none of 5 inversions detected by OGM. Manta alone found 32 out of 49 translocations, 25 out of 46 deletions, 3 out of 9 duplications but none of 9 insertions and none of 5 inversions detected by OGM. To be noticed, two translocations and

two deletions detected by OGM were actually identical within the resolution of the technique, and both events corresponded to a single event detected by Delly and Manta. Translocations detected by OGM but not by WGS had less molecules supporting them, indicating a lower VAF and probably a sub-clonal event. Thus, OGM was more sensitive than WGS to detect sub-clonal variants, but we cannot exclude (i) that the two tumor sample extractions contained different subclonal structure; (ii) that these lower VAF events detected by OGM corresponded to noise. For the other SVs detected only by OGM, 4 deletions and 2 duplications had a VAF ≥ 0.45 and were probably germline variants. Since the intra-fusions are not detectable by Delly or Manta, we compared OGM intra-fusions with deletions, duplications and inversions detected by WGS and confirmed 10 intra-fusions with Delly and 17 with Manta. For the events detected only by WGS, 20 deletions and 3 duplications detected by Manta were smaller than 5kbp and therefore not detectable by OGM.
The correspondence between OGM and WGS results was thus largely validated, confirming the reliability of the OGM technique.

Structural variants that discriminate HRD and HRP TNBC

We then analyzed by OGM a cohort of TNBC with a known status for *BRCA* genes and HRD status determined by shallowHRD (Eeckhoutte et al. 2020). Out of the 24 patients included in the RadioPARP clinical trial (ClinicalTrials.gov Identifier: NCT03109080) (Loap et al. 2022), frozen archived material was available for 17 cases (Supp Table 1). Two of the 17 retrieved samples did not pass the quality control. Overall, we analyzed by RVA 15 high quality samples, of which 8 were HRD and 7 HRP at a mean coverage of 517X. All SVs detected by RVA were visualized on circos plots (Figure 2; Supp Figure 1). A visual inspection of these results revealed that the HRD tumors appeared to be carrying more SVs than the HRP tumors. Filtered SVs were then reclassified by type and size for duplications (splitted, inverted and tandem), deletions, insertions, and inversions (Table 1). We found 3 categories of SVs reaching significance when comparing counts in HRD and HRP tumors: tandem duplications (TD) between 10 and 100kb, insertions between 1 and 10kb and between 10

and 100kb (Wilcoxon rank test; p-values $1.9x10^{-2}$, $3.1x10^{-4}$ and $1.9x10^{-4}$, respectively)(Figure 3A). The high number of TD in *BRCA1* tumors was anticipated, as they have been described as a key feature of these tumors. However, the detection of small insertions was surprising in such a tumor context (Nik-Zainal et al. 2016). We thus hypothesized that these small insertions could actually be TD. This is anticipated as OGM will re-categorize duplication as insertion when low label signals prevent from identifying a repeated pattern. Indeed, when plotting the sizes of the insertions and TDs detected by OGM in each sample, insertions were found rarely exceeding 30kb, whereas TDs were rarely smaller than 30kb (Figure 3B). Also supporting this hypothesis, manual inspection of ten of the largest insertions (≥ 30 kb) allowed to reclassify all of them but one as TD (one was reclassified as an inverted duplication)(Figure 3C).

Therefore, we subsequently combined insertions and duplications according to sizes as a same feature of HRD (Figure 3D). This feature fully separates HRD from HRP tumors but one. Visual inspection of the circos plot of this tumor revealed that it carried an amplification of *CCNE1* (Figure 3E). Interestingly, over-expression of *CCNE1* is known to induce over-replication, and has been associated with a TD phenotype of various sizes (Menghi et al. 2018). Having identified and explained this outlier case, the TD/insertion feature then fully discriminated HRP and HRD TNBC (Figure 3G). However, previous works have demonstrated that this TD phenotype less than 100 kb is very specific of *BRCA1*-inactivated but not of *BRCA2*-inactivated tumors (Nik-Zainal et al. 2016). These TD are most probably linked to other functions of BRCA1 in genome maintenance, independently of HR (Willis et al. 2017).

Surprisingly, we found no difference in number of detected translocations between HRD and HRP tumors (Table 2, Figure 2). This contrasts with the strong association of this feature defined by WGS with HRD (Nik-Zainal et al. 2016), and with their clear contrast between HRD and HRP tumors by visual inspection of the circus plot (Figure 2). The translocations detected in the HRD tumors appeared randomly distributed across the genome, which is not the case of the HRP tumors. We previously described tumor genome of TNBC and have found the rearrangements following two density distributions with a cut-off of around 3 Mb (Popova et al.

2012). We thus suggested a new feature of isolated event when a free interval of a given translocation (or intra-fusion) breakpoint from another event was ≥3 Mb. We found that isolated translocation, isolated intra-fusion and isolated translocation+intra-fusion features perfectly discriminated HRD and HRP tumors, and even the *CCNE1* amplified tumor (Figure 4). We further tested all possible free intervals from 0 to 5Mb and evaluated the number of isolated translocations+intra-fusions for each interval according to the HR status. Actually, a perfect discrimination was observed from 0.8 to 4.5 Mb free intervals, demonstrating the robustness of this new feature of isolated events. As a cross-validation, we then compared the number of isolated events detected by OGM and of the large-scale genomic alterations (LGA) feature identified by the shallowHRD pipeline previously developed (Eeckhoutte et al. 2020). These two features correlated well (r^2 = 0.78; p_value = 1.3 $x10^{-5}$), suggesting that the two independent methods are measuring a similar genomic feature (Figure 4). The only outlier sample being of low tumor content, which may explain an underestimated LGA score by shallowHRD.

In conclusion, OGM analyses of TNBC allowed a cost effective and readily interpretable genome profiling, even in samples with low tumor content. Two robust and important features were identified. Firstly, tandem duplications, which were often recognized as insertions by the present Bionano pipeline when less than 30 kb, and which high number (insertions+duplications) almost perfectly identified *BRCA1* tumors. However, we also found that *CCNE1*-amplified tumors could present a similar TD phenotype. Secondly, the number of isolated inter- and intra-chromosomal translocation events (translocations and intra-fusions, respectively), as defined by a free interval of another event of around 3 Mb, represented a robust feature of HRD.

OGM is thus an interesting and promising approach for HRD determination. Validation of the features we are proposing are ongoing in independent TNBC cohorts, in different breast cancer subtypes, and in other types of cancers, in particular high-grade ovarian carcinomas. For future application of OGM in clinics, as pros, this method is cost-effective and highly efficient even with samples of low tumor content, it does not require complex analyses and high data storage.

As cons, it requires a high quantity and quality of frozen tumor samples and an OGM dedicated platform. More technical developments and optimization are thus mandatory before realistic implementation of OGM in routine diagnosis of HRD.

Acknowledgments

This work was supported by a Sponsored Research Agreement with Bionano Genomics and S.V. was supported by this grant. The authors thank François-Clément Bidard for his scientific and clinical inputs, the center of biological resources (CRB) of Institut Curie for providing frozen tumors fragments, in accordance with the ethical standards of the institutional and national research committees. DNA optical mapping was supported with the grant SESAME 2019 from the Region Ile France. The authors thank the NGS platform of the research center of Institut Curie. High-throughput sequencing was performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the *Agence Nationale de la Recherche* ("*Investissements d'Avenir*" program), by the ITMO-Cancer Aviesan (*Plan Cancer* III), and by the SiRIC-Curie program (SiRIC grant INCa-DGOS-4654).

Disclosures

This work was supported by a Sponsored Research Agreement with Bionano Genomics and S.V. was supported by this grant. T.P. and M.-H.S. are co-inventors of the LST method (US20170260588, US20150140122 and exclusive license to Myriad Genetics).

References

Ceccaldi R, Rondinelli B, D'Andrea AD. 2016. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26**: 52-64.

Scully R, Panday A, Elango R, Willis NA. 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**: 698-714.

Groelly FJ, Fawkes M, Dagg RA, Blackford AN, Tarsounas M. 2022. Targeting DNA damage response pathways in cancer. *Nat Rev Cancer*.

Riaz N, Blecua P, Lim RS, Shen R, Higginson DS, Weinhold N, Norton L, Weigelt B, Powell SN, Reis-Filho JS. 2017. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nature Communications* **8**: 857.

Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS, Liu Y, Akbani R, Feng B, Donehower LA, Miller C, Shen Y, Karimi M, Chen H, Kim P, Jia P, Shinbrot E, Zhang S, Liu J, Hu H, Bailey MH, Yau C, Wolf D, Zhao Z, Weinstein JN, Li L, Ding L, Mills GB, Laird PW, Wheeler DA, Shmulevich I, Monnat RJ, Jr., Xiao Y, Wang C. 2018. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep* **23**: 239-254.e236.

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**: 913-917.

Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**: 917-921.

Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, Fabbro M, Ledermann JA, Lorusso D, Vergote I, Ben-Baruch NE, Marth C, Madry R, Christensen RD, Berek JS, Dorum A, Tinker AV, du Bois A, Gonzalez-Martin A, Follana P, Benigno B, Rosenberg P, Gilbert L, Rimel BJ, Buscema J, Balser JP, Agarwal S, Matulonis UA, Investigators E-ON. 2016. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* **375**: 2154-2164.

Ray-Coquard I, Pautier P, Pignata S, Perol D, Gonzalez-Martin A, Berger R, Fujiwara K, Vergote I, Colombo N, Maenpaa J, Selle F, Sehouli J, Lorusso D, Guerra Alia EM, Reinthaller A, Nagao S, Lefeuvre-Plesse C, Canzler U, Scambia G, Lortholary A, Marme F, Combe P, de Gregorio N, Rodrigues M, Buderath P, Dubot C, Burges A, You B, Pujade-Lauraine E, Harter P, Investigators P-. 2019. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. *N Engl J Med* **381**: 2416-2428.

Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, Tran TV, Williams D, Iliev D, Jammulapati S, FitzGerald LM, Krivak T, DeLoia JA, Gutin A, Mills GB, Lanchbury JS. 2012. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer* **107**: 1776-1782.

Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, Bowman-Colin C, Li Y, Greene-Colozzi A, Iglehart JD, Tung N, Ryan PD, Garber JE, Silver DP, Szallasi Z, Richardson AL. 2012. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov* **2**: 366-375.

Popova T, Manie E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-Zafrani B, Bollet M, Longy M, Houdayer C, Sastre-Garau X, Vincent-Salomon A, Stoppa-Lyonnet D, Stern MH. 2012. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* **72**: 5454-5462.

Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, Rosebrock D, Livitz D, Kubler K, Mouw KW, Kamburov A, Maruvka YE, Leshchiner I, Lander ES, Golub TR, Zick A, Orthwein A, Lawrence MS, Batra RN, Caldas C, Haber DA, Laird PW, Shen H, Ellisen LW, D'Andrea AD, Chanock SJ, Foulkes WD, Getz G. 2017. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet* **49**: 1476-1486.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjaerde OC, Langerod A, Ringner M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporte I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodriguez-Gonzalez FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, van't Veer L, Tutt A, Knappskog S, Tan BK, Jonkers J, Borg A, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JW, Borresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR. 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**: 47-54.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, Ramakrishna M, Martin S, Boyault S, Sieuwerts AM, Simpson PT, King TA, Raine K, Eyfjord JE, Kong G, Borg A, Birney E, Stunnenberg HG, van de Vijver MJ, Borresen-Dale AL, Martens JW, Span PN, Lakhani SR, Vincent-Salomon A, Sotiriou C, Tutt A, Thompson AM, Van Laere S, Richardson AL, Viari A, Campbell PJ, Stratton MR, Nik-Zainal S. 2017. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**: 517-525.

Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333-i339.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, Cox AJ, Kruglyak S, Saunders CT. 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**: 1220-1222.

Eeckhoutte A, Houy A, Manie E, Reverdy M, Bieche I, Marangoni E, Goundiam O, Vincent-Salomon A, Stoppa-Lyonnet D, Bidard FC, Stern MH, Popova T. 2020. ShallowHRD: detection of homologous recombination deficiency from shallow whole genome sequencing. *Bioinformatics* **36**: 3888-3889.

Loap P, Loirat D, Berger F, Rodrigues M, Bazire L, Pierga JY, Vincent-Salomon A, Laki F, Boudali L, Raizonville L, Mosseri V, Jochem A, Eeckhoutte A, Diallo M, Stern MH, Fourquet A, Kirova Y. 2022. Concurrent Olaparib and Radiotherapy in Patients With Triple-Negative Breast Cancer: The Phase 1 Olaparib and Radiation Therapy for Triple-Negative Breast Cancer Trial. *JAMA Oncol*.

Menghi F, Barthel FP, Yadav V, Tang M, Ji B, Tang Z, Carter GW, Ruan Y, Scully R, Verhaak RGW, Jonkers J, Liu ET. 2018. The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations. *Cancer Cell* **34**: 197-210 e195.

Willis NA, Frock RL, Menghi F, Duffey EE, Panday A, Camacho V, Hasty EP, Liu ET, Alt FW, Scully R. 2017. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**: 590-595.

Figure Legends

Figure 1. Comparison of Whole Genome Sequencing (WGS) and Optical Genome Mapping (OGM) of a BRCA2-mutated triple-negative breast carcinoma (TNBC).
Upper panel: Circos plots representing WGS Structural Variant (SV) analyses with Delly and Manta bioinformatics pipelines and OGM RVA analyses. Del: deletions, dup: duplications, ins: insertions, inv: inversions. Lower panel: Venn diagrams representing RV comparisons with Delly, Manta and RCA analyses of translocations, deletions and duplications.

Figure 2. Circos plots of Optical Genome Mapping (OGM) of four representative examples of triple-negative breast carcinomas (TNBC) analyzed by RVA.
Left panel: two examples of HRD TNBC. Right panel: two examples of HRP TNBC.

Figure 3. Analysis of insertions (Ins) and tandem duplications (TD) in the series of triple-negative breast carcinomas (TNBC).
A. Box plot representing the number of insertions (Ins) of two size classes and of tandem duplication (TD) in HRD (orange) and HRP (blue) TNBC. B. Representation of number of INS (orange) and TD (blue) according to size classes in the series of TNBC. C. Manual re-alignment showing that this rearrangement called INS by RVA could be re-classified as DUP. D. Box plot representing the number of insertions plus duplication (Ins+Dup) of two size classes in HRD (orange) and HRP (blue) TNBC. E. Circos plot of the outlier HRP case with enlargement of the *CCNE1* region showing its amplification. * and *** indicate p_value < 0.05 and < 0.001, respectively; Wilcoxon Mann-Whitney nonparametric test.

Figure 4. Analysis of translocation events in the series of triple-negative breast carcinomas (TNBC).
A. Box plot representing the number of total translocation events (TRA), isolated translocations (Isolated TRA), Intra-fusions (INTRA), isolated intra-fusions (Isolated INTRA) and isolated TRA plus INTRA in HRD (orange) and HRP (blue) TNBC. B. 2D plot representing number of isolated TRA plus INTRA (x-axis) and number of Large Genomic Rearrangements (LGA) detected by shallowHRD (y-axis). ** and *** indicate p_value < 0.005 and < 0.001, respectively; Wilcoxon Mann-Whitney nonparametric test.

Table 1. Analysis of differential structural variants (SVs) between homologous recombination proficient (HRP) and deficient (HRD) triple negative breast cancers (TNBC)

Type of SVs	Size	P_values
DEL	<10kb	0.16
	<100kb	0.46
	<1Mb	0.32
	<10Mb	0.51
	>10Mb	1.00
TD	<10kb	1.00
	<100kb	**0.019**
	<1Mb	0.71
	<10Mb	1.00
	>10Mb	1.00
DUP split	<10kb	1.00
	<100kb	0.68
	<1Mb	0.75
	<10Mb	0.76
	>10Mb	1.00
DUP inv	<10kb	1.00
	<100kb	0.056
	<1Mb	0.16
	<10Mb	1.00
	>10Mb	1.00
INS	**<10kb**	**0.00031**
	<100kb	**0.00093**
	<1Mb	0.28
	<10Mb	1.00
	>10Mb	1.00
INV	<10kb	1.00
	<100kb	0.079
	<1Mb	0.084
	<10Mb	1.00
	>10Mb	1.00
TRA	TRA	0.32
Intra-F	Intra-F	0.46

SV : structural variant, DEL : deletion, TD : tandem duplication, DUP : duplication, INS : insertion, INV : inversion, TRA : translocation, Intra-F : intrafusion. P_variant is given comparing HRD and HR TNBC.

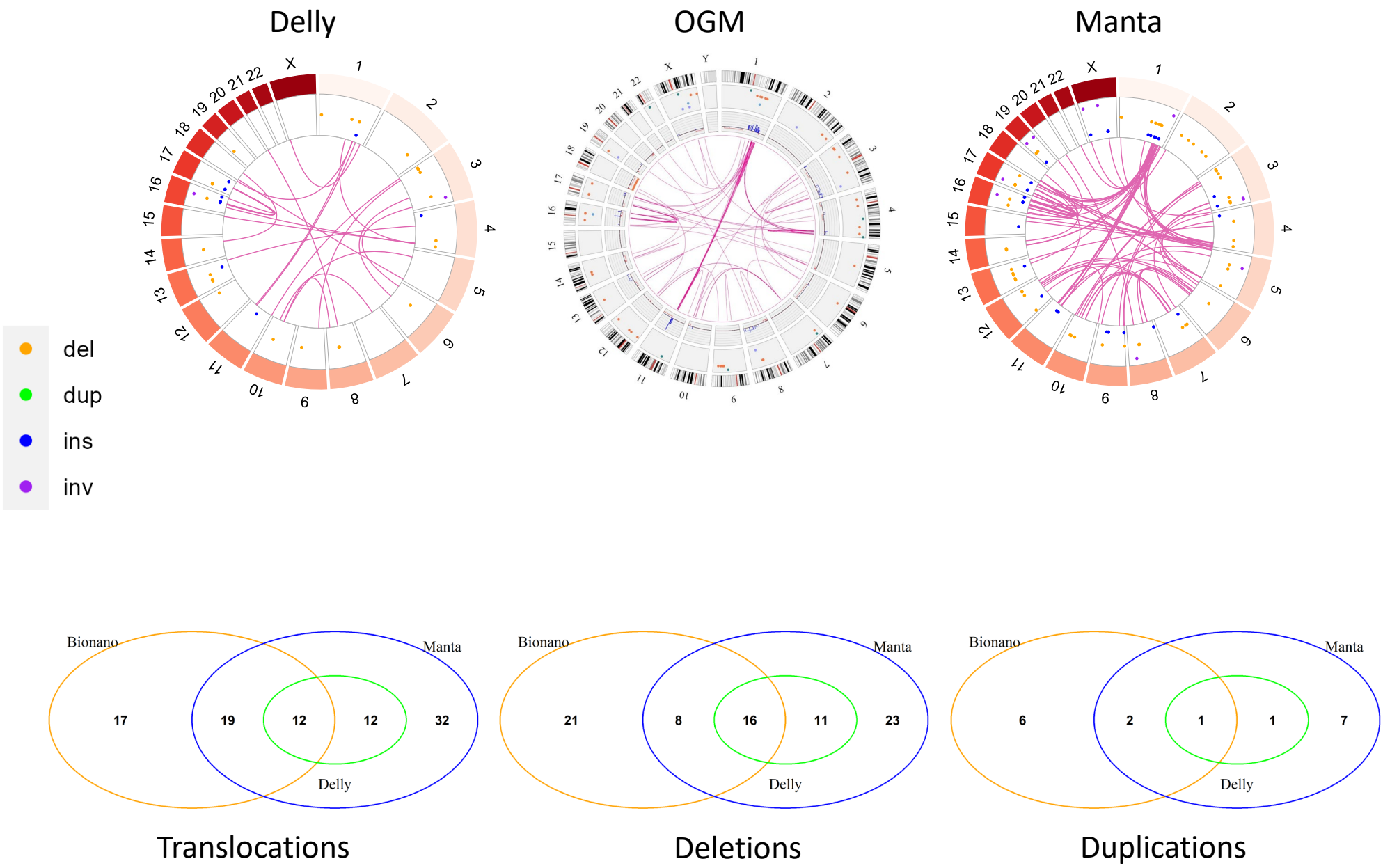
Delly
OGM
Manta
del
dup
ins
inv
Bionano
Manta
Delly
17
19
12
12
32
Translocations
Bionano
Manta
Delly
21
8
16
11
23
Deletions
Bionano
Manta
Delly
6
2
1
1
7
Duplications

Figure 1

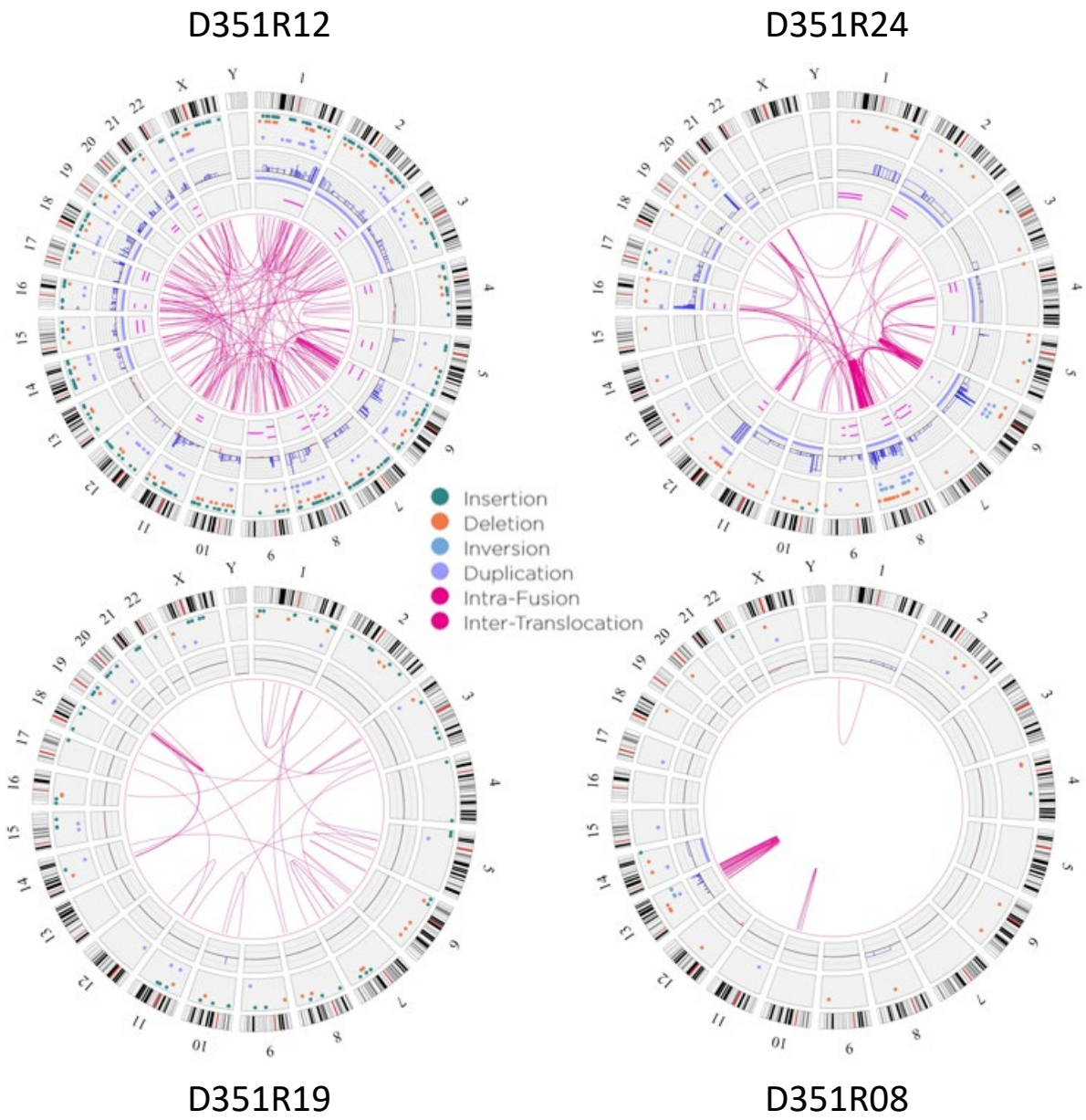
D351R12
D351R24
D351R19
D351R08
Insertion
Deletion
Inversion
Duplication
Intra-Fusion
Inter-Translocation

Figure 2

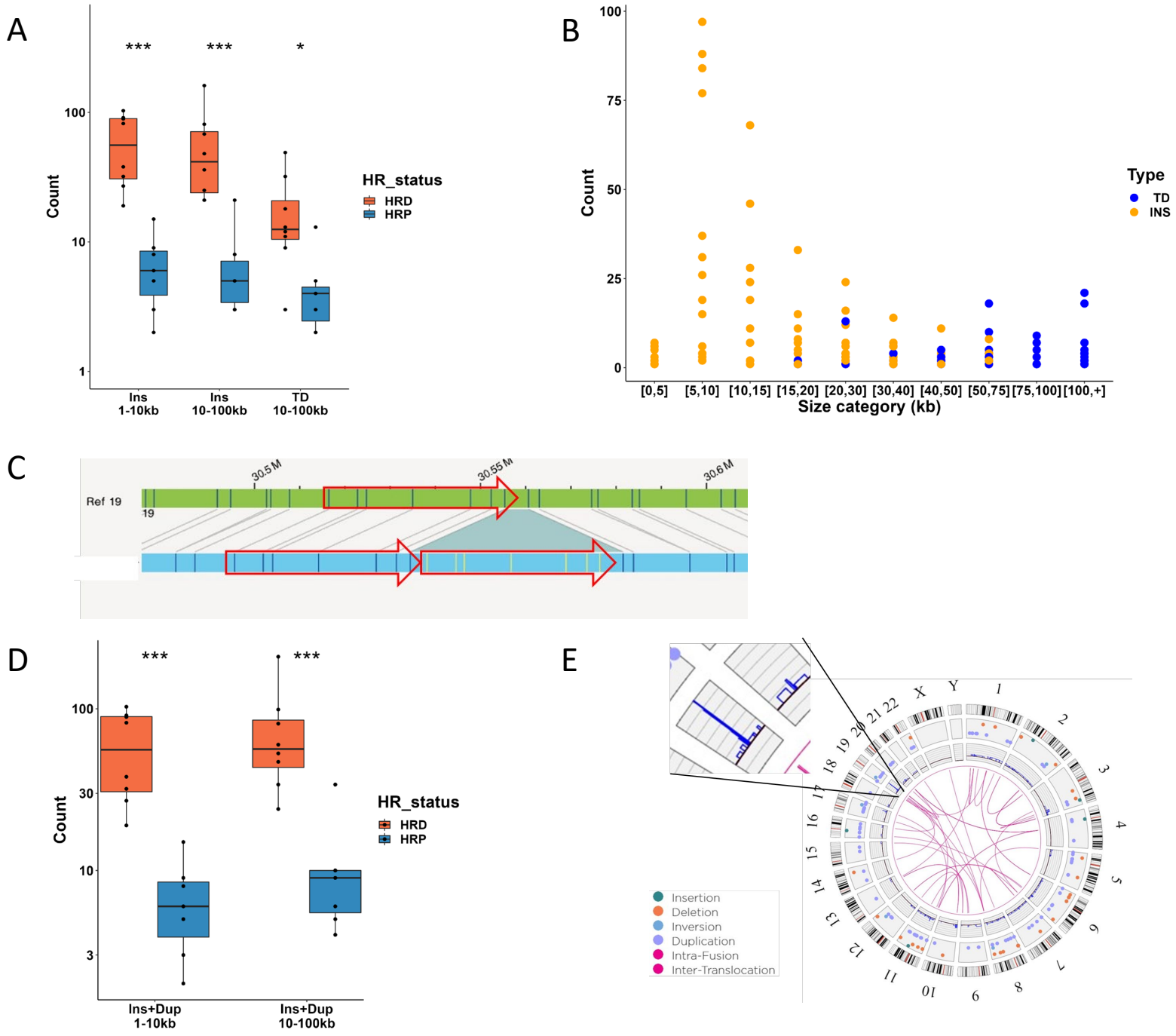
A
***
***
*
100
10
1
Count
Ins
1-10kb
Ins
10-100kb
TD
10-100kb
HR_status
HRD
HRP
B
100
75
50
25
0
Count
[0,5]
[5,10]
[10,15]
[15,20]
[20,30]
[30,40]
[40,50]
[50,75]
[75,100]
[100,+]
Size category (kb)
Type
TD
INS
C
Ref 19
19
30.5 M
30.55 M
30.6 M
D
***
***
100
30
10
3
Count
Ins+Dup
1-10kb
Ins+Dup
10-100kb
HR_status
HRD
HRP
E
Insertion
Deletion
Inversion
Duplication
Intra-Fusion
Inter-Translocation

Figure 3

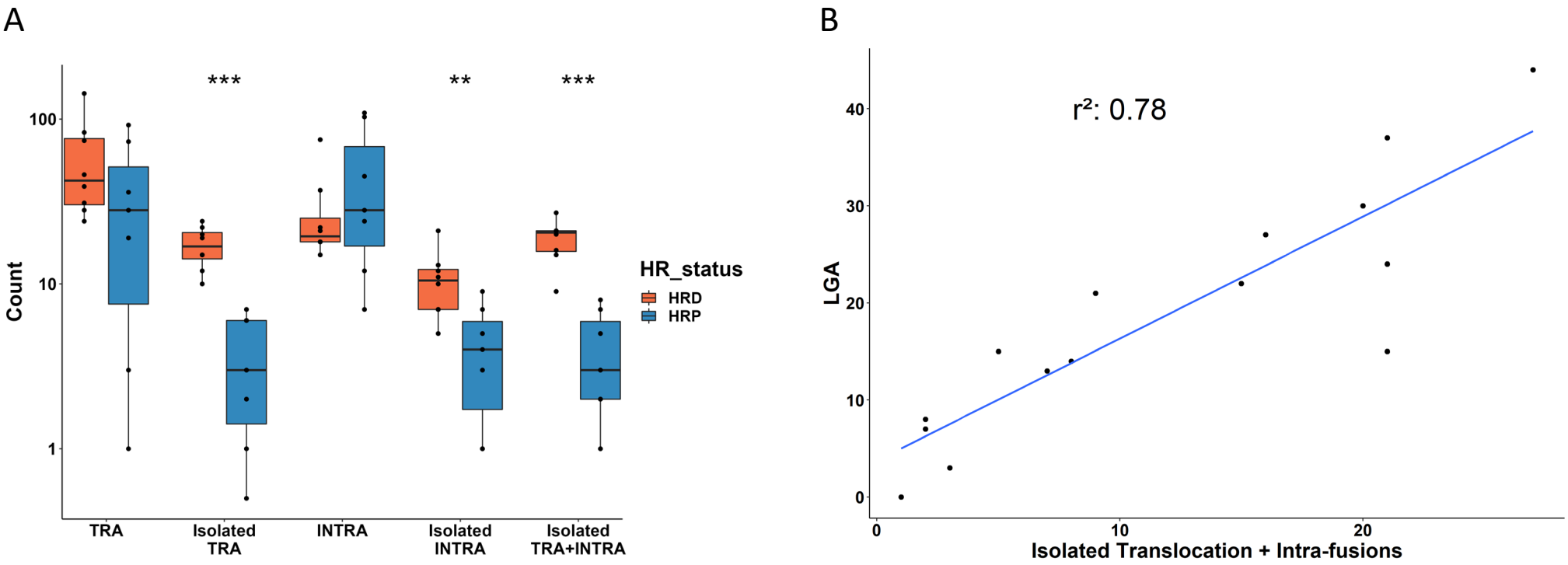
A
***
**
***
100
10
1
Count
TRA
Isolated TRA
INTRA
Isolated INTRA
Isolated TRA+INTRA
HR_status
HRD
HRP
B
r²: 0.78
40
30
20
10
0
LGA
0
10
20
Isolated Translocation + Intra-fusions

Figure 4