

## BRIEF COMMUNICATION OPEN

## Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer

Zsafia Sztupinszki<sup>1</sup>, Miklos Diossy<sup>1</sup>, Marcin Krzystanek<sup>1</sup>, Lilla Reiniger<sup>2,3</sup>, István Csabai<sup>4</sup>, Francesco Favero<sup>5</sup>, Nicolai J. Birkbak<sup>6,7</sup>, Aron C. Eklund<sup>1</sup>, Ali Syed<sup>8</sup> and Zoltan Szallasi<sup>1,3,9,10</sup>

The first genomic scar-based homologous recombination deficiency (HRD) measures were produced using SNP arrays. As array-based technology has been largely replaced by next generation sequencing approaches, it has become important to develop algorithms that derive the same type of genomic scar scores from next generation sequencing (whole exome “WXS”, whole genome “WGS”) data. In order to perform this analysis, we introduce here the scarHRD R package and show that using this method the SNP array-based and next generation sequencing-based derivation of HRD scores show good correlation (Pearson correlation between 0.73 and 0.87 depending on the actual HRD measure) and that the NGS-based HRD scores distinguish similarly well between BRCA mutant and BRCA wild-type cases in a cohort of triple-negative breast cancer patients of the TCGA data set.

*npj Breast Cancer* (2018)4:16; doi:10.1038/s41523-018-0066-6

## INTRODUCTION

Reliable quantification of homologous recombination deficiency of human tumor biopsies, especially in the case of ovarian and breast cancer, is expected to identify patients that are particularly sensitive to platinum or PARP inhibitor-based therapy.<sup>1</sup> Before the widespread introduction of next generation sequencing (NGS) to characterize tumor biopsies, SNP arrays were used to identify large-scale genomic aberrations associated with homologous recombination deficiency, often induced by the loss of BRCA1 or BRCA2 function. Three such measures were identified: telomeric allelic imbalance (HRD-TAI score),<sup>2</sup> loss of heterozygosity profiles (HRD-LOH score),<sup>3</sup> and large-scale state transitions (HRD-LST score).<sup>4</sup> These three measures have also been combined into a single summary measure of HR deficiency.<sup>5</sup> The HRD-LOH score has also become an integral part of a recently published, whole-genome sequencing-based measure of homologous recombination deficiency, HRDetect.<sup>6</sup> These measures, along with functional assays,<sup>7</sup> showed promise to identify HR-deficient cases and thus predict response to platinum or PARP inhibitor therapy.<sup>2,8,9</sup> Since NGS has become the main genomic characterization method of cancer biopsies, it has become essential to migrate the SNP array-based methodology to NGS-based platforms.

TCGA breast cancer biopsies have been both SNP array profiled and subjected to NGS allowing a direct comparison.<sup>8</sup>

## RESULTS AND DISCUSSION

We found good correlation between the SNP array-based and NGS-based HRD scores (Fig. 1). When comparing the results of the scarHRD R package to SNP array-based measurements, we found the following Pearson correlation coefficients: number of telomeric allelic imbalances (NtAI):  $r = 0.84$  ( $R^2 = 0.70$ , adjusted  $R^2 = 0.70$ ,  $p < 2.2e-16$ ), large-scale transition (LST)  $r = 0.79$  ( $R^2 = 0.62$ , adjusted  $R^2 = 0.62$ ,  $p < 2.2e-16$ ) loss of heterozygosity (HRD-LOH)  $r = 0.73$  ( $R^2 = 0.53$ , adjusted  $R^2 = 0.52$ ,  $p < 2.2e-16$ ). These three measures are often combined for diagnostic purposes<sup>5</sup> and in HRDetect.<sup>6</sup> Therefore, we also compared the sum of the three scores across the two platforms (HRD sum):  $r = 0.87$  ( $R^2 = 0.75$ , adjusted  $R^2 = 0.75$ ,  $p < 2.2e-16$ ) (Fig. 1). The artificial reduction of coverage to 30× did not affect this correlation (Supplementary Material, Figure S7-S8). The BRCA1/2-mutated samples showed significantly higher NGS-based HRD-sum values (Fig. 2, Supplementary Figure S6). The predictive value of HRD-sum, measured as AUC value of the corresponding ROC curve, was 80.8% (Supplementary Figure S2).

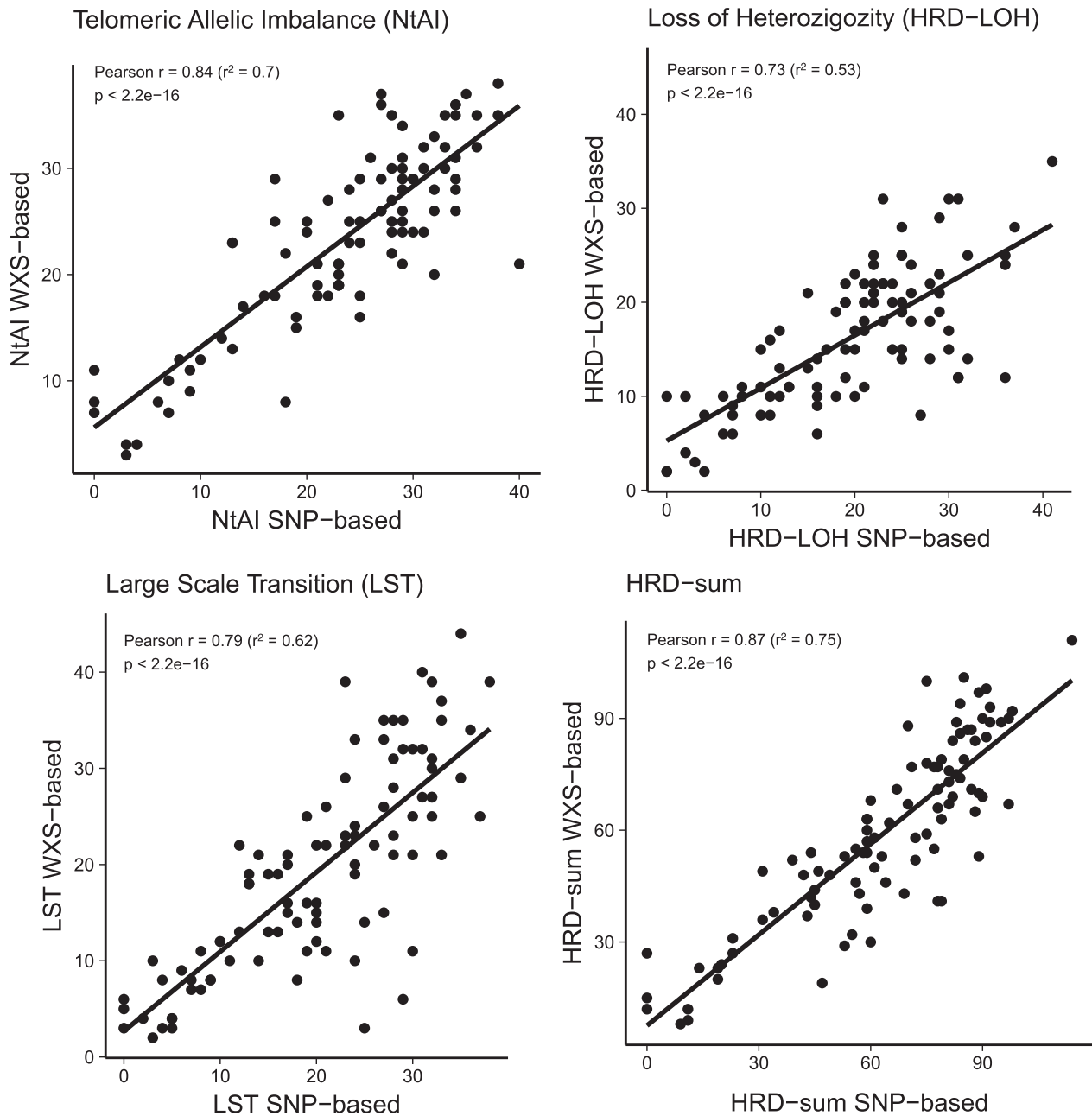
There was no significant difference in SNP versus WXS-based estimation of tAI, LST, and HRD-sum, but the number of LOH events were significantly lower in the WXS-based estimation ( $p = 0.012$ , Kolmogorov-Smirnov test). This could be attributed to differences in segmentation algorithm (the more segmented the WXS data is the lower number of LOHs that are called) or to low sample quality, coverage. However, when comparing the ROC curves for BRCA1/2 status of the SNP-based and WXS-based HRD-

<sup>1</sup>Department of Bio and Health Informatics, Technical University of Denmark, Kemitorvet 208, Lyngby 2800, Denmark; <sup>2</sup>1st Department of Pathology and Experimental Research, Semmelweis University, Budapest, Hungary; <sup>3</sup>2nd Department of Pathology, MTA-SE NAP, Brain Metastasis Research Group, Hungarian Academy of Sciences, Semmelweis University, Budapest, Hungary; <sup>4</sup>Department of Physics of Complex Systems, Eötvös Loránd University, Budapest, Hungary; <sup>5</sup>Biotech Research & Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen DK-2200, Denmark; <sup>6</sup>The Francis Crick Institute, London, UK; <sup>7</sup>University College London Cancer Institute, London, UK; <sup>8</sup>Danish National Life Science Supercomputing Center, Copenhagen, Denmark; <sup>9</sup>Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA and <sup>10</sup>Harvard Medical School, Boston, MA, USA

Correspondence: Zoltan Szallasi (zoltan.szallasi@childrens.harvard.edu)

Received: 6 February 2018 Revised: 16 May 2018 Accepted: 17 May 2018

Published online: 02 July 2018



**Fig. 1** Correlation between Affymetrix SNP 6.0 array-based and whole exome sequencing-based measurements of homologous recombination deficiency (telomeric allelic imbalance, loss of heterozygosity, large-scale transitions, and the sum of these estimates)

score, there was no significant difference between the SNP array-based and NGS-based methods. (Supplementary Figure S3).

According to our expectations and previous results the BRCA1/2-deficient cases showed higher values for each of the four scores (Supplementary Figure S4-S5).

The sum of the three HRD scores showed good correlation across the two platforms. Thus in more advanced NGS-based HRD deficiency measures such as HRDetect, the SNP array-based step could be replaced by an NGS-based estimate of the HR deficiency scores.

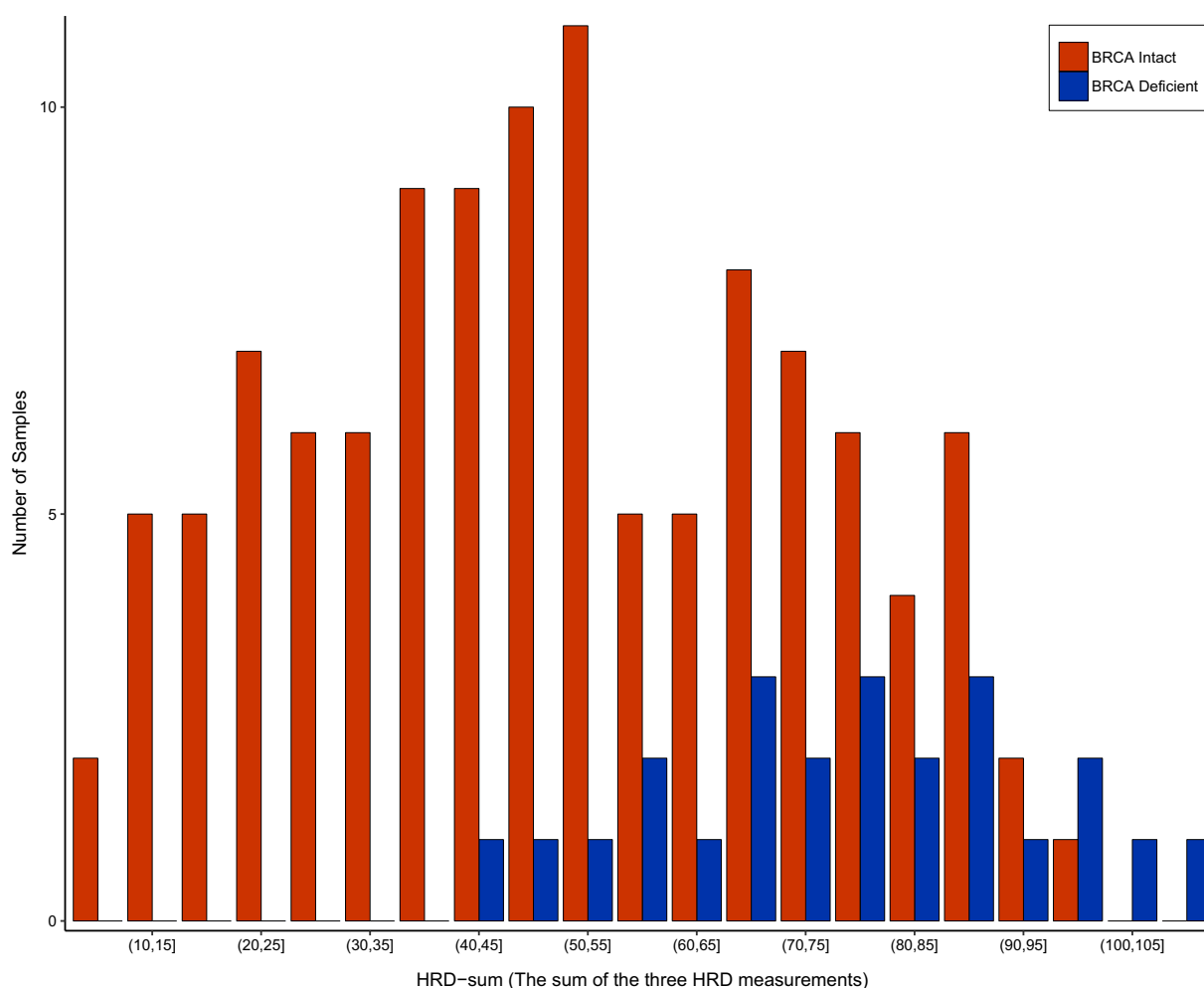
#### Brief description of the methods

Based on receptor status determined by immunohistochemistry, 139 paired tumor and normal samples of the TCGA breast cancer cohort could be classified as triple-negative breast cancer. From

these patients 95 had Affymetrix SNP 6.0 array-based HRD estimates (LOH, TAI, LST), previously published by our group.<sup>10</sup> In this publication we present the scarHRD R package (<https://github.com/sztup/scarHRD>) which estimates the level of the three HR deficiency measures using NGS data.

A sample's LOH score is the total number of LOH regions across the entire genome that are larger than 15 Mb but do not cover whole chromosomes. In the original publication this 15 Mb lower limit for LOH was determined by comparing SNP array profiles between BRCA mutant and BRCA wild-type cases.<sup>3</sup> We performed a similar analysis using NGS data and found that the original 15 Mb cutoff performed best in this case as well (Supplementary Figure S1).

The LST is defined as a chromosomal break between adjacent regions of at least 10 Mb, with a distance between them not larger than 3 Mb.



**Fig. 2** Distribution of HRD-sum values in BRCA1/2 deficient and in BRCA1/2 intact triple-negative breast cancer samples from TCGA. HRD-sum values were determined with the scarHRD R package

The number of telomeric allelic imbalances is the number of AIs (the unequal contribution of parental allele sequences with or without changes in the overall copy number of the region) that extend to the telomeric end of a chromosome.

Allele-specific copy number estimation is a crucial part of estimating HR deficiency. As previously shown, allele-specific copy number estimation from NGS data performed using the Sequenza R package show high agreement with SNP array-based copy number profiles.<sup>11</sup> The scarHRD package is, therefore, able to use Sequenza preprocessed files as well as other allele-specific segmentation files in the same format.

As it has been previously shown that in ovarian cancer the sum of the genomic scar scores is elevated in BRCA-deficient cancers,<sup>5</sup> an additional aim of our study was to compare the unweighted numeric sum of LOH, tAI, and LST, called here HRD-sum, to the BRCA1/2 status of the patients. A sample was classified as BRCA-deficient if (1) there was a deep deletion of BRCA1/2, (2) a germline and a somatic mutation in BRCA1/2 with LOH, or (3) if LOH had co-occurred with promoter methylation in one of the *BRCA1/2* genes. The somatic mutation status (mutations with likely pathogenic function) and methylation data was acquired from the TCGA data portal. The germline mutation status was determined using HaplotypeCaller, and was annotated with Intervar,<sup>12</sup> likely pathogenic mutations and frameshift insertion/deletion with unknown significance were used in our analysis. LOH was determined using Sequenza's allele-specific segmentation results (Supplementary Table S1).

#### Data availability

The data sets generated during the current study are available from the corresponding author on reasonable request.

#### Code availability

The code/algorithm for performing the experiments is available for download at <https://github.com/sztup/scarHRD>.

#### ACKNOWLEDGEMENTS

This work was supported by the Research and Technology Innovation Fund (KTIA\_NAP\_13-2014-0021 to Z.S.); Breast Cancer Research Foundation and the Novo Nordisk Foundation Interdisciplinary Synergy Programme Grant (NNF15OC0016584 to I.C. and Z.S.), by the ÚNKP-17-4-III-SE-63 New National Excellence Program of the Ministry of Human Capacities to L.R. and by Tesaro Inc. The results shown here are based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

#### AUTHOR CONTRIBUTIONS

Conception, design, writing, and review of the manuscript: Zs.S., M.D., M.K., L.R., I.C., F. F., N.J.B., A.C.E., A.S., and Zo.S. Development of methodology: Zs.S., M.D., M.K. Analysis and interpretation of data: Zs.S., M.D., M.K., Zo.S.

## ADDITIONAL INFORMATION

**Supplementary information** accompanies the paper on the *npj Breast Cancer* website (<https://doi.org/10.1038/s41523-018-0066-6>).

**Competing interests:** N.J.B., A.C.E., and Zo.S are listed as co-inventors on a patent on telomeric allelic imbalance, which is owned by Children's Hospital Boston and licensed to Myriad Genetics. The remaining authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Lord, C. J. & Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **16**, 110–120 (2016).
2. Birkbak, N. J. et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov.* **2**, 366–375 (2012).
3. Abkevich, V. et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **107**, 1776–1782 (2012).
4. Popova, T. et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **72**, 5454–5462 (2012).
5. Telli, M. L. et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin. Cancer Res.* **22**, 3764–3773 (2016).
6. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
7. Mutter, R. W. et al. Bi-allelic alterations in DNA repair genes underpin homologous recombination DNA repair defects in breast cancer. *J. Pathol.* **242**, 165–177 (2017).
8. Mirza, M. R. et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N. Engl. J. Med.* **375**, 2154–2164 (2016).
9. Zhao, E. Y. et al. Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer. *Clin. Cancer Res.* **23**, 7521–7530 (2017).
10. Marquard, A. M. et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. *Biomark. Res.* **3**, 9 (2015).
11. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
12. Li, Q. & Wang, K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am. J. Hum. Genet.* **100**, 267–280 (2017).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018