A subset of lung cancer cases shows robust signs of homologous recombination deficiency associated genomic aberration based molecular signatures.

Miklos Diossy[1*], Zsofia Sztupinszki[2*], Judit Borcsok[2*], Marcin Krzystanek[2], Viktoria Tisza[3], Sandor Spisak[4], Orsolya Rusz[5], István Csabai[6], Judit Moldvay[5,7], Anders Gorm Pedersen[1], David Szuts[8], Zoltan Szallasi [2,3,5,9]

[1] Department of Health Technology, Technical University of Denmark, Kemitorvet 208, 2800 Lyngby, Denmark

[2] Danish Cancer Society Research Center, Copenhagen, Denmark

[3] Computational Health Informatics Program, Boston Children's Hospital, USA,

[4] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts

[5] 2nd Department of Pathology, MTA-SE NAP, Brain Metastasis Research Group, Hungarian Academy of Sciences, Semmelweis University, Budapest, Hungary

[6] Department of Physics of Complex Systems, Eötvös Loránd University, Budapest, Hungary

[7]Department of Tumor Biology, National Korányi Institute of Pulmonology-Semmelweis University, Budapest, Hungary.

[8] Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary.

*These authors contributed equally

[9] Correspondence should be sent to:
Zoltan Szallasi, Computational Health Informatics Program (CHIP) Boston Children's Hospital, Harvard Medical School, 300 Longwood Ave., Boston Massachusetts, USA, 02215, e-mail: Zoltan.szallasi@childrens.harvard.edu, +1-617-355-2179

Abstract

Consistent with their assumed mechanism of action, PARP inhibitors show significant therapeutic efficacy in breast, ovarian and prostate cancer, which are the solid tumor types most often associated with loss of function of key homologous recombination genes. It is not known, however, how often other solid tumor types may be homologous recombination deficient. Specific DNA aberration profiles, genomic scars are induced by homologous recombination deficiency (HRD) and those could be used to assess the presence or absence of this DNA repair pathway aberration in a given tumor biopsy. We analyzed whether the various HRD associated mutational signatures are present in the whole genome and whole exome sequencing data of lung cancer in the TCGA cohorts and found evidence that a subset of those cases shows robust signs of HR deficiency. These clinical cases could be candidates for PARP inhibitor treatment and their prioritization for clinical trials could be achieved using next generation sequencing based mutational signatures.

PARP inhibitors are a new class of cancer therapeutic agents that are most effective in tumors lacking the homologous recombination mediated DNA repair pathway[1]. They are approved either in BRCA1/2 mutant tumors or in solid tumor types most often associated with lack of HR in the form of BRCA1/2 function deficiency[1]. It is possible, however, that other tumor types not associated with germline BRCA1/2 mutations may also be HR deficient. Non-small cell lung cancers, for example, show somatic mutations in the BRCA1/2 genes in 5-10% of the cases[2], but it is not known how often those mutations are associated with loss of function of the gene and/or loss

of heterozygosity or other mechanisms suppressing BRCA1/2 function that lead to a bona fide HR deficient phenotype. We were seeking such tumor cases by investigating the next generation sequencing based DNA aberrations profiles of TCGA cohorts.

Loss of function of the key homologous recombination genes BRCA1 and BRCA2 is associated with a range of distinct mutational signatures that include: 1) A single nucleotide variation based mutational signature ("COSMIC signature 3" or "BRCA signature" as labeled in the original publication[3]), 2) a short insertions/deletions based mutational profile, often dominated by deletions with microhomology, a sign of alternative repair mechanisms joining double strand breaks in the absence of homologous recombination[4,5] 3) large scale rearrangements such as non-clustered tandem duplications of a given size range (mainly associated with BRCA1 loss of function) or deletions in the range of 1-10kb (mainly associated with BRCA2 loss of function)[6]. We have recently shown that several of these DNA aberration profiles are in fact directly induced by the loss of BRCA1 or BRCA2 function[5].

Therefore, we analyzed all available whole genome sequencing data from the TCGA lung adenocarcinoma (LUAD) and squamous lung cancer (LUSC) cohorts and determined which of the above listed mutational signatures are present in these cases. Based on analyzing whole genome (n=42 and n=48, respectively) and whole exome (n=553 and n=489 samples) data we compared their utility for estimating HRD.

Results

Loss of function mutations of HR genes in lung cancer

Detailed analysis on the germline and somatic mutations of DNA repair genes were performed. We identified loss of function mutations for the BRCA2 gene in three LUSC and two LUAD cases (all three in LUSC and one in LUAD were coupled with LOH), and loss of function mutation for BRCA1 for one case in LUSC (Supplementary Figures 1-4). A LUAD case was also identified with a RAD51B germline mutation that was recently shown to be associated with HR deficiency[7] accompanied with an LOH in the tumor. We hypothesized that some of these cases may exhibit robust signs of homologous recombination deficiency induced mutational signatures.

HR deficiency associated mutational signatures in lung squamous carcinoma

The three BRCA2 mutant LUSC samples showed an elevated short deletion/insertion ratio of >2 with two of the cases having the highest such ratio in the cohort (Figure 1B). Increased deletion/insertion ratios were described previously for BRCA2 deficient cancers using whole genome sequencing data[8]. Two of the three BRCA2 deficient cases also showed the highest proportion (>0.1) of larger than 2 bp long microhomology mediated deletions and the same two cases had the highest proportion of larger than 9 bp long short deletions (Supplementary Figure 9). These two indel-patterns have also been described previously in BRCA2 deficient human cancer biopsies[4,5,8].

We did not detect an increased SNV signature 3 (originally described in BRCA1/2 deficient tumors[3]) in these particular cases, probably because the high level of smoking induced mutational signatures would mask them even if they were present

(Figure 1C and D). (For a detailed distribution of SNV based signatures see Suppl. Fig 8).

On the other hand, both of the above mentioned BRCA2 deficient cases showed the highest number of RS5 rearrangement signatures (Figure 1D), which were previously described in breast cancer to be strongly associated with loss of function of BRCA2[6].

Taken together, two of the three likely BRCA2 deficient LUSC specimens showed clear signs of BRCA2 deficiency associated mutational signatures and thus those cases are likely homologous recombination deficient.

HR deficiency associated mutational signatures in lung adenocarcinoma

In the case of LUAD, consistent with the lower number of smokers in this tumor type, in about half of the cases (20 out of 42) the smoking signature did not dominate the SNV signatures and the contribution of other mutational processes could be clearly detected (Supplementary Figure 8). More prominently, the BRCA2 mutant case with LOH (TCGA-78-7143), along with six other cases, showed a strong presence of signature 3. The same BRCA2 mutant sample had a high proportion of microhomology mediated deletions and four of the other six samples showed high deletion/insertion ratios along with high proportions of microhomology mediated deletions (Figure 1A). The RAD51B case (TCGA-64-1680) showed both the signs of the HR deficiency associated indel patterns and a high signature 3 ratio.

The BRCA2 mutant case and the four other cases showing HRD-like SNV and indel patterns also showed presence of the rearrangement signatures associated with

BRCA function loss, although to a lesser extent than that seen in TCGA-64-1680 and in BRCA1/2 mutant breast cancer in general.

HRDetect scores in the LUAD and LUSC WGS cohorts

Considering the different types of mutational signatures induced by the loss of function of HR genes, and that in a given tumor the loss of that gene may have a different impact on those mutational signatures, it was suggested recently that the SNV, short indel and large-scale rearrangement signatures along with a CNV-derived genomic scar score[9] be combined into a single HRD quantifier, HRDetect[10]. This complex HR deficiency measure was trained on the number and relative distribution of HRD induced DNA aberration profiles in breast cancer.

We calculated the breast cancer trained HRDetect values for all WGS cases by standardizing the lung predictors combined with the original breast cancer dataset, and found that the two above described, likely BRCA2 deficient LUSC cases;TCGA-66-2766 and TCGA-21-5782 have the highest HRDetect values, the former of which even exceeded 0.7, which was proposed to be the threshold value for bona fide HR deficient cases in breast cancer (Supplementary Figure 11).

We also calculated the HRDetect values when the predictors were standardized on the lung cancer cases alone (Figure 2B). These values are in general higher since, as we pointed out, some of the HRD suspect cases showed strong signs of some (e.g. SNV and short indel based) HRD signatures but not others (large rearrangement based signatures) (Figure 1C and D). In other words, it is possible that the individual parameters in a lung cancer specific HRDetect model will be significantly different

from those in breast cancer. Both in the case of LUAD and LUSC, eight of the analyzed cases showed a >0.7 lung cancer normalized HRDetect value (Figure 2B).

HR deficiency associated biomarkers in whole exome sequencing data

While whole genome sequencing data carry the most information about HRD induced mutational processes, we previously showed that whole exome sequencing data can also be used for these purposes under certain conditions, albeit using only 1 % of all aberrations that are present in the whole genome data[11]. While WES data contain a significantly reduced number of mutations, we extended our analysis to lung cancer WES data to allow for a more comprehensive assessment of the possible number of HR deficient lung cancer cases. (There are about ten times as many WES than WGS covered cases in the TCGA).

We started with the comparative analysis of those cases that had both WGS and WES data available. (Suppl Figure 14). For the LUSC cases, the HRD-LOH score showed strong (0.83), and the ratio of signature 3 showed reasonable (0.41) correlation across the WES and WGS data . Due to the lower number of detectable deletions in whole exomes in general, all microhomology mediated deletions were considered in WES if their size exceeded 1bp, and this number was compared to the >2bp microhomologies in whole genomes. While on average there was a two order of magnitudes difference in the absolute number of deletions between the corresponding pairs, they exhibited a strong correlation (0.79). Of the two most likely HR deficient LUSC cases, TCGA-21-5782 showed good correlation of all three measures across the WES and WGS data. In the other case (TCGA-66-2766), however,

the WES data did not recapture the same mutational signatures as the WGS data (Supplementary Figure 13).

For LUAD all three measures showed correlations between 0.55 and 0.84 across the WGS and WES data, and the cases with high HR deficiency associated attributes, like signature 3 or the HR deficiency-like insertion/deletion pattern in the WGS data had showed the same tendency in WES data as well.

For the distribution of the various HR deficiency mutational signatures across the entire WES based cohorts see Supplementary Figures 12 and 13.

HRDetect scores in the LUAD and LUSC whole exome sequencing data

Finally, we calculated the HRDetect values based on the standardized and log-transformed attributes of the WES data (further details of the whole exome model are available in the Supplementary Notes, Section 3.5). For breast cancer, there is a reasonably good correlation between the WES and WGS HRDetect values[11], and we found moderate correlations (~0.5) between the corresponding pairs in both of the analyzed lung cancer cohorts as well (Supplementary Figure 14).

The two LUSC patients showing signs of HR-deficiency based on whole genome sequencing, also had high HRDetect values based on whole exome sequencing analysis (TCGA-21-5782: 0.80, TCGA-66-2766: 0.66). Since the HR deficiency status of these two cases are supported by WGS data, we used the lower HRDetect value of these two cases as a putative threshold for HR deficiency in the WES characterized LUSC cohort. In the LUSC WES cohort 16% of the patients had higher than 0.66 HRDetect scores (Figure 3B), while in the case of the LUAD cohort (Figure 3A) 3.8%

of the patients had at least as high HRDetect score as the RAD51B-mutated sample (TCGA-64-1680).

Since high HRDetect scores were reported to be associated with better clinical outcome in platinum treated breast cancer[12], we were wondering whether lung cancer cases below and above the HRDetect thresholds that we determined in the WES data have significantly different outcome when treated with platinum containing therapy. However, higher HRDetect-scores were not associated with better outcome in these cohorts. (Supplementary Figures 15 and 16).

Discussion

PARP inhibitors show significant clinical efficacy in tumor types that are often associated with BRCA1/2 mutations, such as breast, ovarian, and prostate cancer. In order to further expand this clinical benefit, there are several ongoing clinical trials evaluating the efficacy of PARP inhibitors in non-small cell lung cancer, such as the PIPSeN (NCT02679963) and Jasper (NCT03308942) trials. If, however, the clinical benefit is strongly associated with HRD in this tumor type and only a minority of lung cancer cases harbor this DNA repair pathway aberration, then the success of those clinical trials will greatly depend on our ability to identify and prioritize the HRD cases.

In order to develop such a diagnostic method, we first analyzed the BRCA1/2 mutant lung cancer cases. Lung cancer is usually not associated with germline BRCA1/2 mutations, although a few sporadic cases have been reported[13]. Nevertheless, due to

e.g. smoking, about 5-10% of non-small cell lung cancer cases show somatic mutations in either the BRCA1 or BRCA2 genes. Some of those are likely to be pathogenic and associated with LOH as well. In our analysis, these cases clearly showed the mutational signatures usually associated with HRD. This strongly suggests that there are some bona fide HRD cases amongst lung cancer as well. Beyond mutations in BRCA1/2, HRD can be induced by a variety of mechanisms, such as suppression of expression of BRCA1 by promoter methylation. This is reflected by the fact that a significant number of ovarian and breast cancer cases show clear patterns of HRD associated mutational signatures in the absence of mutations of BRCA1/2 or other key HR genes[10]. Furthermore, BRCA1 mutant cases can be rendered HR proficient and thus PARP inhibitor resistant by the loss of other genes such as 53BP1 or REV7 etc[14,15]. Therefore, downstream mutational signatures, such as those investigated in our analysis, could be more accurate measures of HRD than the mutational status or expression change of individual genes. In fact, we identified several lung cancer WGS cases with high HRD induced mutational signatures that were not associated with BRCA1/2 mutations and it is reasonable to assume that those signatures were also induced by HRD. It is important to estimate the proportion of potentially HRD non-small lung cancer cases to optimize PARP inhibitor trials. Since, we had only a limited number of WGS covered cases (less than one hundred in total), we extended our analysis to WES as well. Previously we found, that while WES based HRD estimates are less accurate and less sensitive than WGS based estimates, they still provide a clinically informative measure of HRD[11]. The large number of WES covered cases in TCGA allowed us to make a reasonable first estimate at least on the

upper bound of the proportion of HRD cases. Based on the HRD associated mutational profiles it is likely that less than 20% of lung cancer cases are associated with HRD and thus likely be responding to PARP inhibitors.

We made every effort to detect a likely explanation for the cases with significant HRD associated mutational signatures but TCGA profiles have significant limitations due to e.g. normal tissue contamination. For example, significant expression deficiency or LOH of the BRCA1 or BRCA2 genes can often be masked by the presence of these genes in the normal cells in the tumor biopsy.

We did not find a correlation between the likely presence HRD and better survival upon treatment in lung cancer, which is probably due to the fact that these patients were treated in addition to platinum with other agents as well. Furthermore, sensitivity or resistance to platinum treatment is also associated with several other mechanisms in addition to HRD [16,17].

Acknowledgment

This work was supported by the Research and Technology Innovation Fund (KTIA_NAP_13-2014-0021 and NAP2-2017-1.2.1-NKP-0002 to Z.S.); Breast Cancer Research Foundation (BCRF-17-156 to Z.S.) and the Novo Nordisk Foundation Interdisciplinary Synergy Programme Grant (NNF15OC0016584 to Z.S. and I.C.), The Danish Cancer Society grant (R90-A6213 to MK). The results shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Conflict of Interest

Z. Szallasi is an inventor on a patent used in the myChoice HRD assay.

Authors contribution

MD, ZS and JB designed the analysis and performed all computational analysis and participated in preparation of the manuscript. MK, VT, SS, OR, IC, JM, AGP, and DS participated in designing the analysis and prepared the manuscript. ZS designed the analysis, had supervisory role and prepared the manuscript.

List of Figures:

Figure 1: Summary of the HRD-related predictors in the LUAD and LUSC whole genome datasets.

A: Fraction of microhomology mediated deletions with larger or equal to 3 bp in length, versus deletions/insertions ratio in the TCGA LUAD whole genome dataset.

B: Fraction of microhomology mediated deletions with larger or equal to 3 bp in length, versus deletions/insertions ratio in the TCGA LUSC whole genome dataset.

C-D: Unsupervised (k-means) clusters of the HRD-related genomic biomarkers in the LUAD (C) and LUSC (D) cohorts. The considered attributes (vertical axes) are the following:

HRD_sum: the sum of the three allele-specific CNV-derived genomic scars (HRD-LOH + LST + ntAI); RS3_ratio: relative ratio of structural variants originating from rearrangement signatures 3; RS5_ratio: relative ratio of structural variants originating from rearrangement signatures 5; mhm_ratio: ratio of microhomology-mediated deletions; Sig3_ratio: relative ratio of mutations originating from point-mutation signature 3.

The number k was set to 4 in both cases, and the samples had been arranged according to their respective clusters. Dark and light gray tiles on the top panels indicate that the sample they belong to (horizontal axis) had one of the highest values in the corresponding attribute. The two shades of grey separate the top tercile range into two sextiles ranges. The darker color indicates that the value lies in the [5/6,1] quantile range, the ligter color indicates that it lies in the [4/6,5/6] quantile range. Below the summary of the biomarkers, the smoking history of the participants is

shown (years: years of active smoking, cig.per.day: average number of cigarettes smoked per day, ts_history: tobacco smoking history). Empty tiles stand for NAs in the dataset, i.e. they do not necessarily translate to a non-smoking history.

Figure 2: HRD scar scores and HRDetect scores of the LUAD and LUSC WGS datasets

In both panels, the sample names are colored according to their genotypes: yellow – BRCA2 heterozygote mutant, red – BRCA2 homozygote mutant, blue – BRCA1 heterozygote mutant.

A, The total sum of the genomic scar scores (HRD-LOH, LST, and ntAI) determined from the LUAD and LUSC whole cancer genome`s allele specific copy number profiles.

B, HRDetect scores calculated using the original, breast cancer whole genome-based HRDetect weights by following the original article`s standardization and attribute-transformation strategies (Davies et al. 2017).

Figure 3: Distribution of the exonic HRDetect scores of the TCGA LUAD (A) and LUSC (B) whole exome samples.

The bars are colored according to the mutational status of the two BRCA genes. Furthermore, the exonic version of a single LUAD sample (TCGA-64-1680) with a likely pathogenic RAD51B mutation and high HRD-related biomarkers in the whole genome dataset is highlighted in pink. The two likely biallelic BRCA2 mutant LUSC

samples from the whole genome analysis (TCGA-21-5782 and TCGA-66-2766) are highlighted in orange.

Online Methods

The normal and tumor BAM files were downloaded for the whole exome sequencing (WXS) samples from TCGA. There were n=489 lung squamous carcinoma (LUSC) and n=553 lung adenocarcinoma (LUAD) samples available with both normal and tumor samples. The Mutect2 vcf files and the clinical data were downloaded from the TCGA data portal (portal.gdc.cancer.gov).

The BAM files for the whole genome sequenced samples were download via the ICGC data portal (dcc.icgc.org). (LUSC: n=48, LUAD: n= 42 patients.)

Mutation, Copy number, and Structural Variant Calling

Germline single nucleotide mutations were specifically called at and around the key HR-related genes for genotyping purposes using HaplotypeCaller, while somatic point-mutations and indels had been called using Mutect2 (GATK 3.8). In order to ensure the high fidelity of the reported SNVs, additional hard-filters had been applied to the resulting variants. In the germline case, the minimum mapping quality (PHRED) was set to 50, variant quality to 20 and a minimum coverage of 15 was ensured, while in case of the somatic SNVs and indels, the minimum tumor LOD (logarithm of odds) was set to 6, the normal to 4, the normal depth to 15, the tumor depth to 20 and the minimally allowed tumor allele frequency to 0.05.

Copy number profiles were determined using Sequenza[18], with fitted models in the ploidy range of [1,5] and cellularity range [0,1]. When a fitted model's predictions significantly differed from the expected ploidy-cellularity values, an alternative solution was selected manually.

Structural variants were detected via BRASS (v6.0.0 - https://github.com/cancerit/BRASS). Through additional hard filters, the minimum amount of variant-supporting read-pairs was set to 6, and a successful local de novo assembly of the reads by velvet was demanded.

Genotyping

The genotypes of the key homologous recombination related genes was determined via annotating the small-scale variant files using intervar[19]. Variants predicted as pathogenic or likely pathogenic were considered deleterious, while variants with unknown significance were treated with greater care but kept as wild-type.

Mutational Signatures

Somatic point-mutational signatures were determined with the deconstructSigs R package[20], by using the cosmic signatures as a mutational-process matrix (Supplementary Figures 5-8).

The extraction of rearrangement signatures was executed according to the following strategy: first, the reported structural variants were mapped to the alphabet of the 32-dimensional structural variant-affecting mutational alphabet, and stored into the matrices $\mathbf{M}_{\mathbf{LUSC}}$ and $\mathbf{M}_{\mathbf{LUAD}}$. Due to the low number of samples in the two WGS cohorts, the extraction of de novo rearrangement signatures was not achievable. Instead, a breast cancer-based, previously described matrix of mutational signatures ($\mathbf{P}$) was used[6]. From these matrices, the signature composition ($\mathbf{E}$) was estimated by solving

the non-negative least squares problem $||\mathbf{P}*\mathbf{E} - \mathbf{M}||_2$, subject to $E_{ij} > 0$, for all i and j (Supplementary Figure 10).

Genomic scar scores

The calculation of the genomics scar scores (loss-of-heterozygosity: LOH, large scale transitions: LST and number of telomeric allelic imbalances: ntAI) were determined using the scarHRD R package.

HRDetect:

Since our cohorts did not have enough clearly HR-deficient cases, the derivation of two LUAD and LUSC specific HRDetect models was not achievable. Instead, on the whole genomes, the scores were calculated using the original, breast cancer-derived model, while the whole exomes relied on an alternative, whole exome-based, but also breast cancer-specific model (further details are available in the Supplementary Notes). In order to get to the results of Figure 2B, the variables were standardized within their respective cohorts (i.e. N_LUSC = 48, N_LUAD = 42). As an alternative way of interpreting the attributes, both sample sets had been appended to the breast cancer predictors, and the standardization step was executed on the resulting larger datasets as well. The distribution of the resulting alternative HRDetect scores are available in Supplementary Figure 11, and both the scores and sample-attributes are available in Supplementary Tables 1 and 2.

References:

1. Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355,** 1152–1158 (2017).
2. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
3. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
4. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
5. Zámborszky, J. *et al.* Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene* **36,** 5085–5086 (2017).
6. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534,** 47–54 (2016).
7. Tumiati, M. *et al.* A Functional Homologous Recombination Assay Predicts Primary Chemotherapy Response and Long-Term Survival in Ovarian Cancer Patients. *Clin. Cancer Res.* **24,** 4482–4493 (2018).
8. Decker, B. *et al.* Biallelic BRCA2 Mutations Shape the Somatic Mutational Landscape of Aggressive Prostate Tumors. *Am. J. Hum. Genet.* **98,** 818–829 (2016).
9. 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).
10. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23,** 517–525 (2017).
11. Diossy, M. *et al.* Breast cancer brain metastases show increased levels of genomic aberration based homologous recombination deficiency scores relative to their corresponding primary tumors. *Ann. Oncol.* **22,** 3608 (2018).
12. Zhao, E. Y. *et al.* Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer. *Clin. Cancer Res.* **23,** 7521–7530 (2017).
13. Donner, I. *et al.* Germline mutations in young non-smoking women with lung adenocarcinoma. *Lung Cancer* **122,** 76–82 (2018).
14. Bunting, S. F. *et al.* BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. *Mol. Cell* **46,** 125–135 (2012).
15. Xu, G. *et al.* REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* **521,** 541–544 (2015).
16. Bruno, P. M. *et al.* A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. *Nat. Med.* **23,** 461–471 (2017).

17. Wang, D. & Lippard, S. J. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov* **4,** 307–320 (2005).
18. Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26,** 64–70 (2015).
19. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* **100,** 267–280 (2017).
20. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17,** 31 (2016).

Figure 1

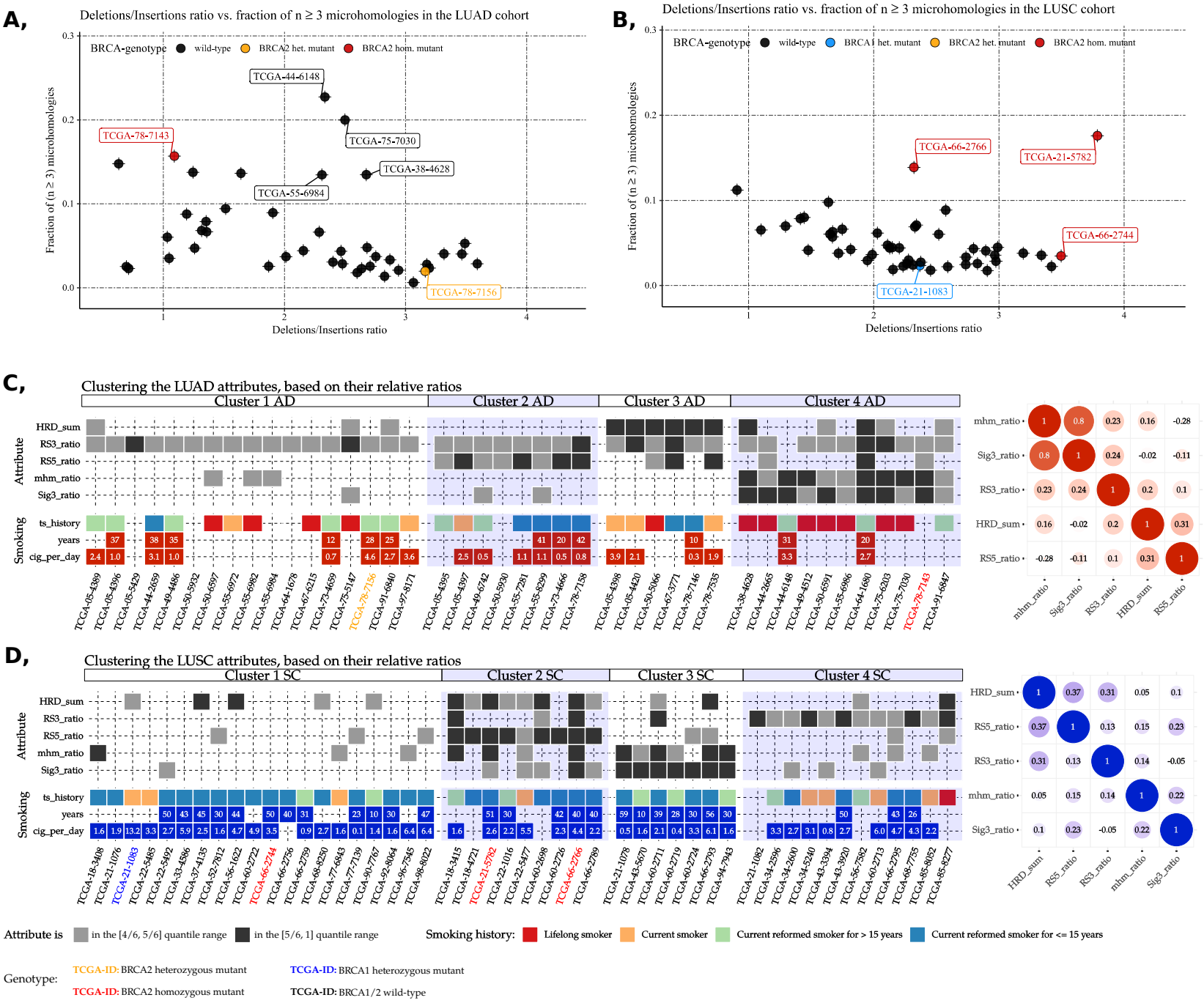

A,
Deletions/Insertions ratio vs. fraction of n ≥ 3 microhomologies in the LUAD cohort
BRCA-genotype
wild-type
BRCA2 het. mutant
BRCA2 hom. mutant
TCGA-44-6148
TCGA-78-7143
TCGA-75-7030
TCGA-38-4628
TCGA-55-6984
TCGA-78-7156
Fraction of (n ≥ 3) microhomologies
Deletions/Insertions ratio
B,
Deletions/Insertions ratio vs. fraction of n ≥ 3 microhomologies in the LUSC cohort
BRCA-genotype
wild-type
BRCA1 het. mutant
BRCA2 het. mutant
BRCA2 hom. mutant
TCGA-66-2766
TCGA-21-5782
TCGA-66-2744
TCGA-21-1083
Fraction of (n ≥ 3) microhomologies
Deletions/Insertions ratio
C,
Clustering the LUAD attributes, based on their relative ratios
Cluster 1 AD
Cluster 2 AD
Cluster 3 AD
Cluster 4 AD
Attribute
HRD_sum
RS3_ratio
RS5_ratio
mhm_ratio
Sig3_ratio
Smoking
ts_history
years
cig_per_day
mhm_ratio
Sig3_ratio
RS3_ratio
HRD_sum
RS5_ratio
D,
Clustering the LUSC attributes, based on their relative ratios
Cluster 1 SC
Cluster 2 SC
Cluster 3 SC
Cluster 4 SC
Attribute
HRD_sum
RS3_ratio
RS5_ratio
mhm_ratio
Sig3_ratio
Smoking
ts_history
years
cig_per_day
HRD_sum
RS5_ratio
RS3_ratio
mhm_ratio
Sig3_ratio
Attribute is
in the [4/6, 5/6] quantile range
in the [5/6, 1] quantile range
Smoking history:
Lifelong smoker
Current smoker
Current reformed smoker for > 15 years
Current reformed smoker for <= 15 years
Genotype:
TCGA-ID: BRCA2 heterozygous mutant
TCGA-ID: BRCA2 homozygous mutant
TCGA-ID: BRCA1 heterozygous mutant
TCGA-ID: BRCA1/2 wild-type

Figure 1 caption:

Figure 1: Summary of the HRD-related predictors in the LUAD and LUSC whole genome datasets.
A: Fraction of microhomology mediated deletions with larger or equal to 3 bp in length, versus deletions/insertions ratio in the TCGA LUAD whole genome dataset.
B: Fraction of microhomology mediated deletions with larger or equal to 3 bp in length, versus deletions/insertions ratio in the TCGA LUSC whole genome dataset.
C-D: Unsupervised (k-means) clusters of the HRD-related genomic biomarkers in the LUAD (C) and LUSC (D) cohorts. The considered attributes (vertical axes) are the following:
HRD_sum: the sum of the three allele-specific CNV-derived genomic scars (HRD-LOH + LST + ntAI);
RS3_ratio: relative ratio of structural variants originating from rearrangement signatures 3;
RS5_ratio: relative ratio of structural variants originating from rearrangement signatures 5;
mhm_ratio: ratio of microhomology-mediated deletions; Sig3_ratio: relative ratio of mutations originating from point- mutation signature 3.
The number k was set to 4 in both cases, and the samples had been arranged according to their respective clusters. Dark and light gray tiles on the top panels indicate that the sample they belong to (horizontal axis) had one of the highest values in the corresponding attribute. The two shades of grey separate the top tercile range into two sextiles ranges. The darker color indicates that the value lies in the [5/6,1] quantile range, the ligter color indicates that it lies in the [4/6,5/6] quantile range. Below the summary of the biomarkers, the smoking history of the participants is shown (years: years of active smoking, cig.per.day: average number of cigarettes smoked per day, ts_history: tobacco smoking history). Empty tiles stand for NAs in the dataset, i.e. they do not necessarily translate to a non-smoking history.

Figure 2

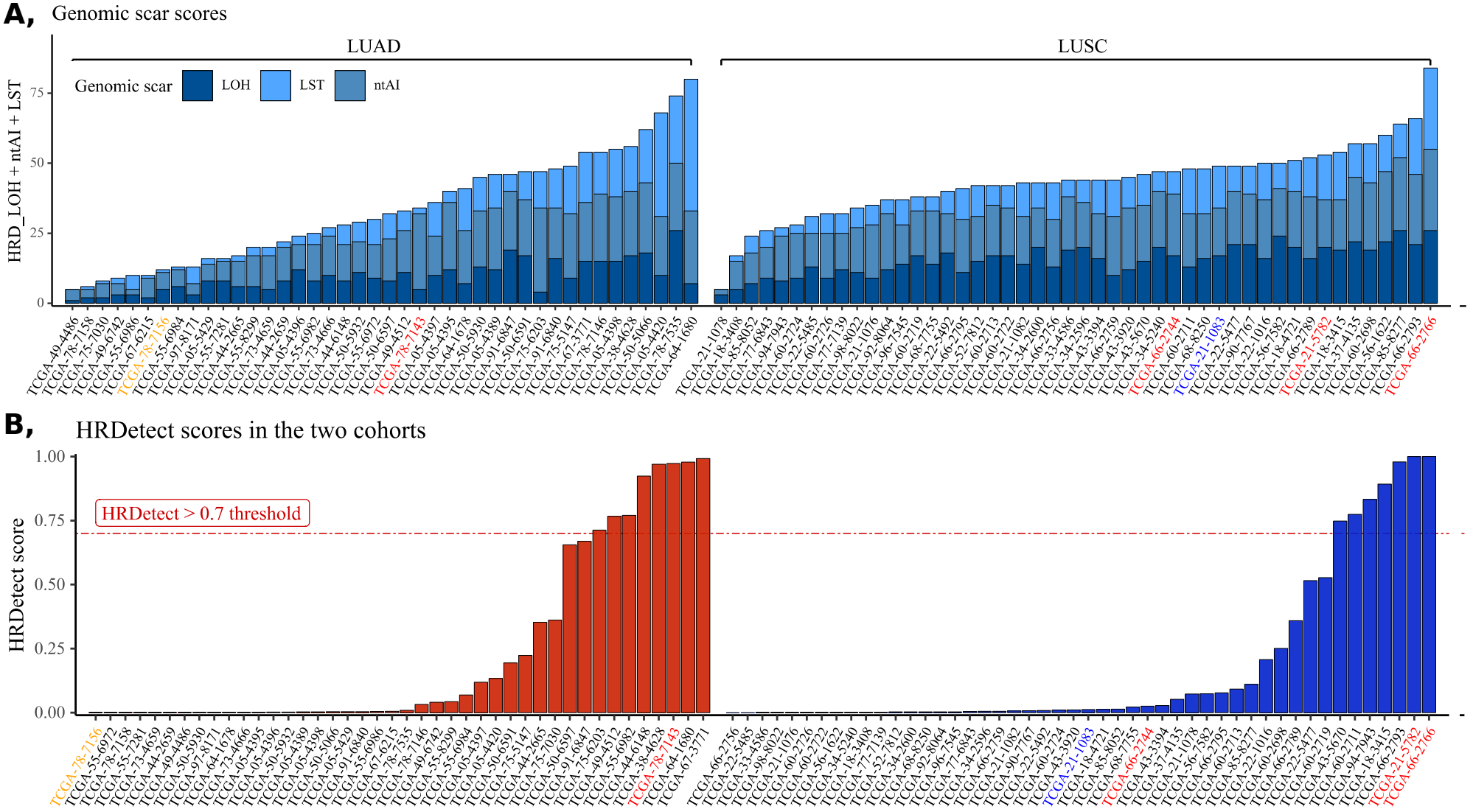

A,
Genomic scar scores
LUAD
LUSC
Genomic scar
LOH
LST
ntAI
HRD_LOH + ntAI + LST
0
25
50
75
B,
HRDetect scores in the two cohorts
HRDetect > 0.7 threshold
HRDetect score
0.00
0.25
0.50
0.75
1.00

Figure 2 caption:

Figure 2: HRD scar scores and HRDetect scores of the LUAD and LUSC WGS datasets
In both panels, the sample names are colored according to their genotypes: yellow – BRCA2 heterozygote mutant, red – BRCA2 homozygote mutant, blue – BRCA1 heterozygote mutant.
A, The total sum of the genomic scar scores (HRD-LOH, LST, and ntAI) determined from the LUAD and LUSC whole cancer genome`s allele specific copy number profiles.
B, HRDetect scores calculated using the original, breast cancer whole genome-based HRDetect weights by following the original article`s standardization and attribute-transformation strategies (Davies et al. 2017).

Figure 3

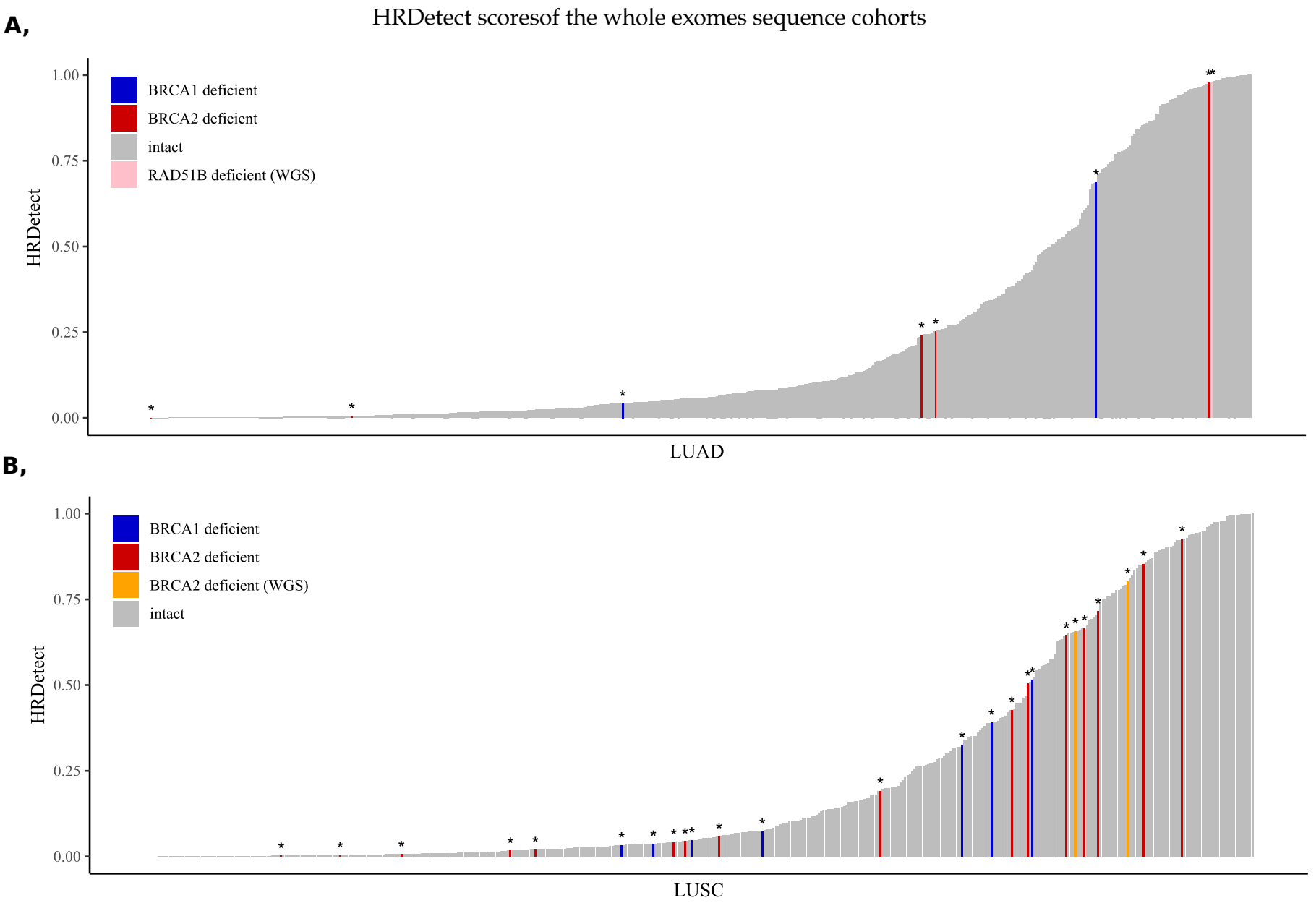

A,
HRDetect scoresof the whole exomes sequence cohorts
BRCA1 deficient
BRCA2 deficient
intact
RAD51B deficient (WGS)
HRDetect
1.00
0.75
0.50
0.25
0.00
LUAD
B,
BRCA1 deficient
BRCA2 deficient
BRCA2 deficient (WGS)
intact
HRDetect
1.00
0.75
0.50
0.25
0.00
LUSC

Figure 3 caption

Figure 3: Distribution of the exonic HRDetect scores of the TCGA LUAD (A) and LUSC (B) whole exome samples.
The bars are colored according to the mutational status of the two BRCA genes. Furthermore, the exonic version of a single LUAD sample (TCGA-64-1680) with a likely pathogenic RAD51B mutation and high HRD-related biomarkers in the whole genome dataset is highlighted in pink. The two likely biallelic BRCA2 mutant LUSC samples from the whole genome analysis (TCGA-21-5782 and TCGA-66-2766) are highlighted in orange.

A subset of lung cancer cases shows robust signs of homologous recombination deficiency associated genomic aberration based molecular signatures.

Supplementary Material

Date of the latest update:
February 26, 2019

Contents

List of Figures

Supplementary tables not included in the Supplementary Text

Supplementary Table 1-2 are available separately, in csv format.

- **Supplementary Table 1**: List of the LUAD whole genomes with their HRD-related attributes
- **Supplementary Table 2**: List of the LUSC whole genomes with their HRD-related attributes

1 Analyzed cohorts

1.1 Whole Genomes

42 LUAD and 48 LUSC WGScohorts had been downloaded from the icgc data portal:

- LUAD cohort: https://icgc.org/ZV9
- LUSC cohort: https://icgc.org/ZVC

1.2 Whole Exomes

Both binary alignment files and MuTect2 vcfs had been downloaded from the gdc data portal. Altogether 553 LUAD and 489 LUSC whole exomes were considered.

1.3 Donors with both WGS and WES samples

The majority of the patients who had whole genome data available, had whole exome sequences as well. All the 48 patients in the LUSC WGS cohort had at least 1 corresponding whole exome, however out of the 42 LUAD WGS patients only 39 had whole exomes. The WGS samples without pairs:

- TCGA-05-5429
- TCGA-64-1678
- **TCGA-78-7143**

Since **TCGA-78-7143** had a likely pathogenic germline BRCA2 mutation coupled with an LOH, we have created an exonic bam-slice using the reads that cover the exome from the WGS bam, in order to check whether the BRCAness phenotype is detectable in the exonic version.

2 Genotyping

Genotypes were determined according to the following scheme; we have called germline variants via GATK (v3.8) HaplotypeCaller, and on whole genomes somatic variants with GATK (v3.8) MuTect2 (for whole exomes, MuTect2-derived vcfs were already available from the gdc data portal). The pathogenecity of these variants was assessed using Intervar (v2.0). From the resulting variant files only the exonic and the +/- 10 nucleotide regions around the exons (in order to account for the possible splice-variants) were considered. From these mutations only those were kept, that were predicted as "Likely Pathogenic", "Pathogenic" or "Uncertain" according to ClinVar (20170905). At last, a threshold on the depth of these variants were set to 20.

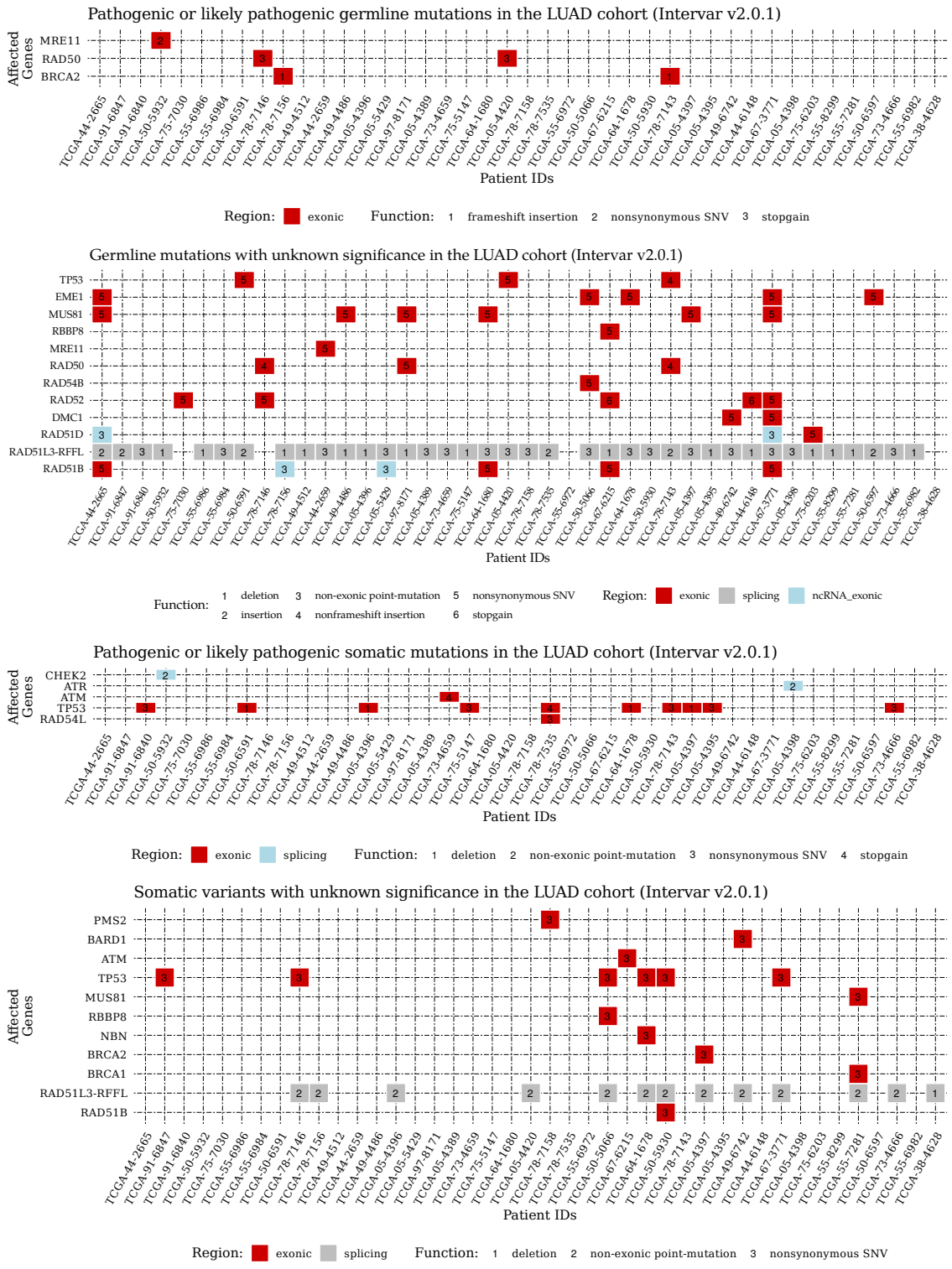
If a variant had been characterized as pathogenic or likely pathogenic by intervar, the corresponding sample was considered mutant, assuming that at least a heterozygous mutation is present in the sample. Variants with unknown significance were collected separately, but they did not affect the genotyping scheme.

2.1 Lung adenocarcinoma WGS samples - mutations in HR-relevant genes

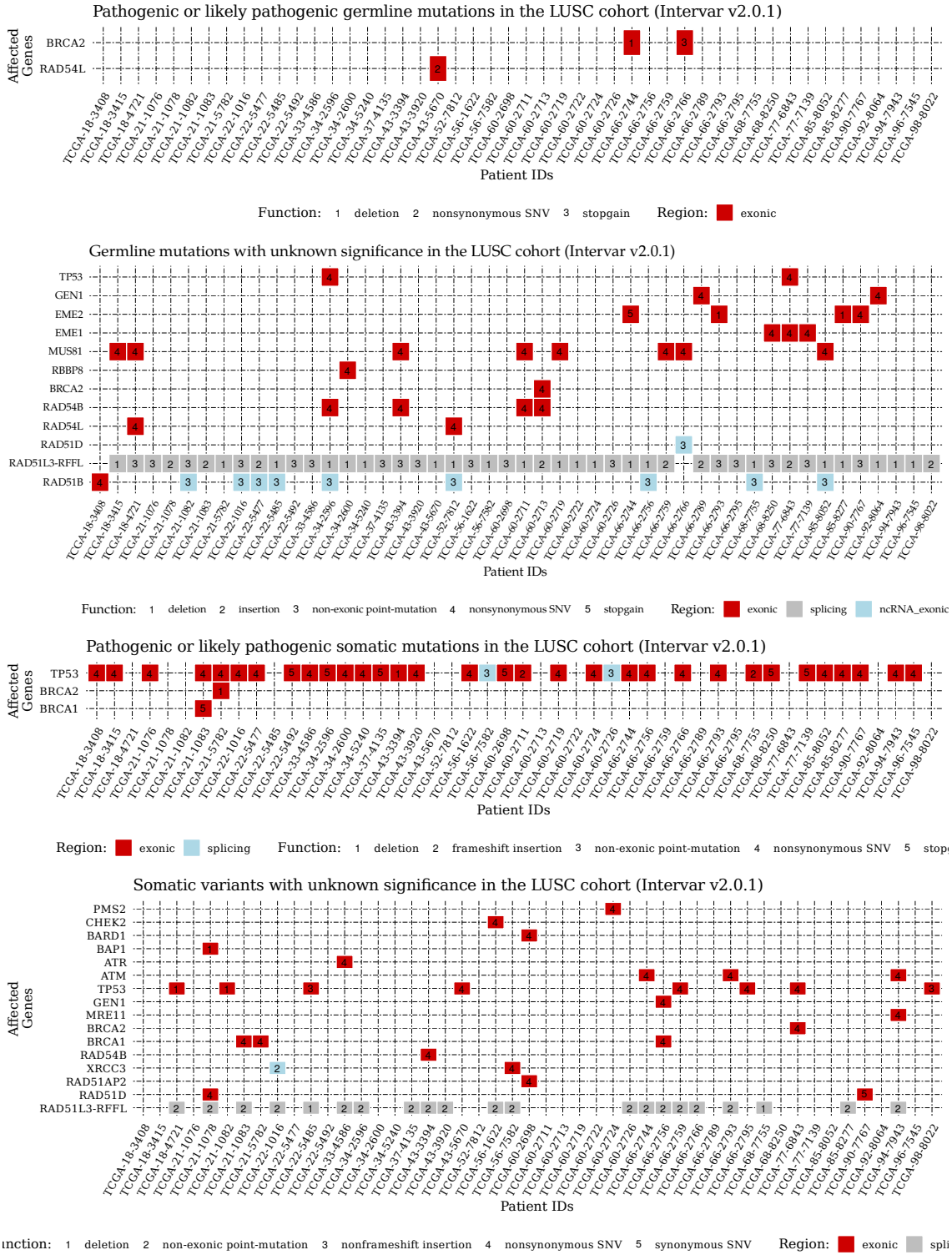
Mutations found in the LUAD WGS cohort are summarized in Suppl.Fig. 1.

2.2 Lung squamous carcinoma WGS samples - mutations in HR-relevant genes

Mutations found in the LUSC WGS cohort are summarized in Suppl.Fig. 2.



Suppl.Fig. 1: *First two panel from the top: Pathogenic or likely pathogenic and UNK germline mutations in the LUAD WGS cohort. Bottom two panel: Somatic variants in the LUAD WGS cohort*



Suppl.Fig. 2: *First two panel from the top: Pathogenic or likely pathogenic and UNK germline mutations in the LUSC WGS cohort. Bottom two panel: Somatic variants in the LUSC WGS cohort*

Suppl.Fig. 3: *Estimated occurrences of LOH events in the analyzed genes. Segment means are estimated using the sequenza and copynumber R packages.*

2.3 Loss of Heterozygosity

The occurrance of Loss of heterozygosity was estimated using the samples' sequenza-derived copy-number segments. If the copy-numbers of either the A or B alleles dropped to zero within the coordinates of a gene, then the LOH event was registered (Suppl. Fig. 3).

2.4 Methylation

Since the majority of the samples had only HumanMethylation 27k data available or didn't have methylation info at all, the genotyping scheme did not consider the methylation status of the gene-specific probes.

2.5 Final genotypes

The final genotypes are summarized in Suppl.Fig. 4. Both cohorts had likely pathogenic heterozygous or homozygous BRCA1/2 mutants among their samples:

LUAD:

- TCGA-75-7156 (likely pathogenic BRCA2 germline mutation)
  /frameshift insertion at chr13:32912949,T>TTGTGC/
- TCGA-78-7143 (likely pathogenic BRCA2 germline mutation + LOH)
  /frameshift insertion at chr13:32906473, A>ACCTAATCTTACTATAT/

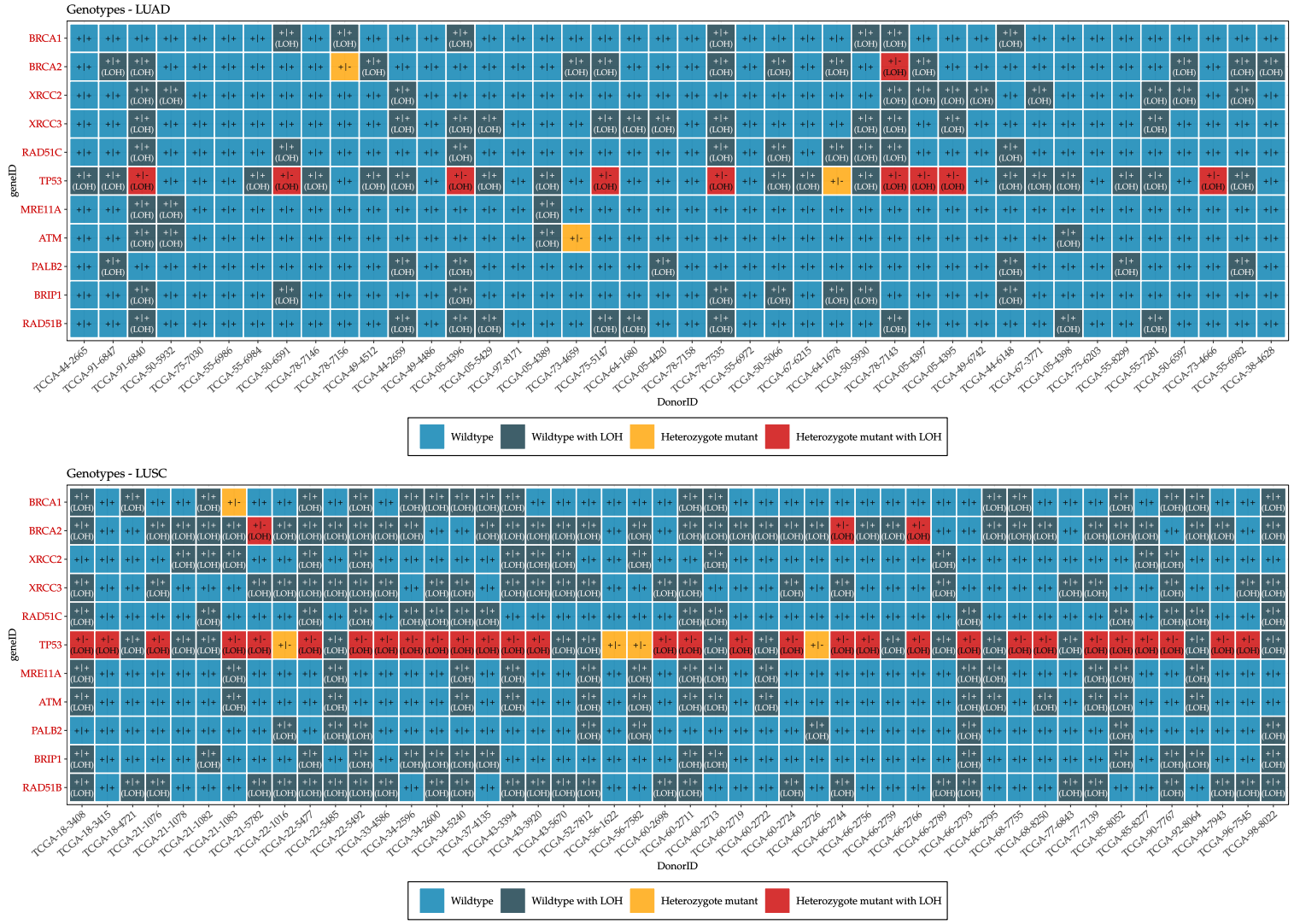
LUSC:

- TCGA-21-1083 (likely pathogenic BRCA1 somatic mutation)
  /stopgain SNV at chr17:41244585, G>C/
- TCGA-21-5782 (likely pathogenic BRCA2 somatic mutation + LOH)
  /frameshift deletion at chr13:32930627, AG>A/
- TCGA-66-2744 (likely pathogenic BRCA2 germline mutation + LOH)
  /frameshift deletion at chr13:32912337, CTG>C/
- TCGA-66-2766 (likely pathogenic BRCA2 germline mutation + LOH)
  /stopgain SNV at chr13:32914349, G>T/

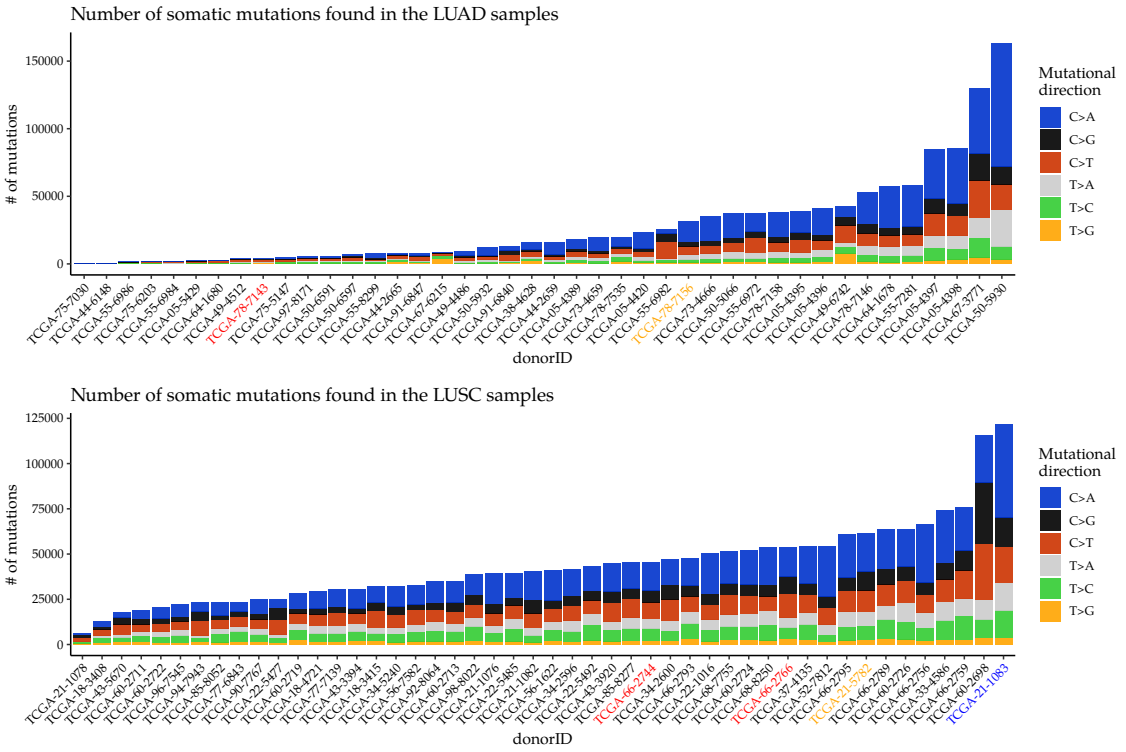
In addition, TCGA-64-1680 – a LUAD sample with high HRD-related genomic aberration scores – had a UNK germline mutation in RAD51B /nonsynonymous SNV at chr14:68352672, A>G/.

2.6 Whole Exomes

Genotyping of the whole exomes followed a similar strategy to the whole genomes'.



Suppl.Fig. 4: *Final genotypes of the LUAD and LUSC WGS cohorts. Genotyping is based on the presence of a pathogenic or likely pathogenic somatic/germline mutation in the gene and whether a loss of heterozygosity event accompanies them. - Heterozygote: at least a germline/somatic mutation present, but no LOH, homozygote: at least a germline/somatic mutation present AND an LOH.*



Suppl.Fig. 5: *Summary of the somatic substitutions detected in the whole genome cohorts. The vertical axis contains overall numbers, the colors on the bars indicate the relative composition of the mutational directions. Most samples are dominated by C>A mutations. Top panel: LUAD, Bottom panel: LUSC.*

3 HRD-related biomarkers

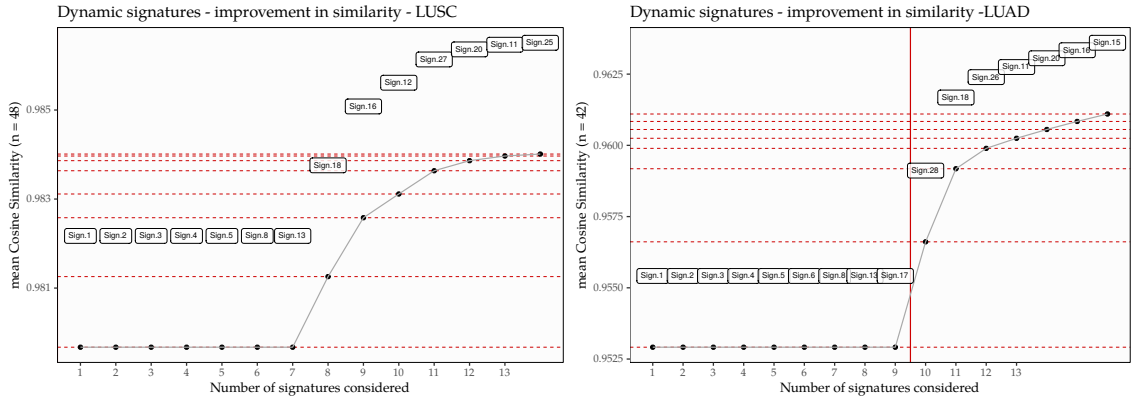
The HRD-induced genomic fingerprints analyzed in this study were the following:

1. Somatic Substitution Signatures [1]
2. Microhomology-mediated deletion ratio, and insertion/deletion ratio [3, 4]
3. Genomic scar scores [6, 7, 8]
4. Rearrangement Signatures [10]

3.1 Somatic Substitution Signatures

Somatic variants returned by MuTect2 had to PASS the following criteria as well, on top of the default filters of MuTect2

- TLOD ≥ 6
- NLOD ≥ 3
- Normal depth ≥ 15
- Tumor depth ≥ 20
- Alt. allele supporting reads in the tumor ≥ 5



Suppl.Fig. 6: *A graphical representation of the dynamic signature addition process. In each iteration a signature that would improve the most on the cosine similarities were considered, however no additional signatures had been added to the list in either cohort. Left: LUSC, Right: LUAD*

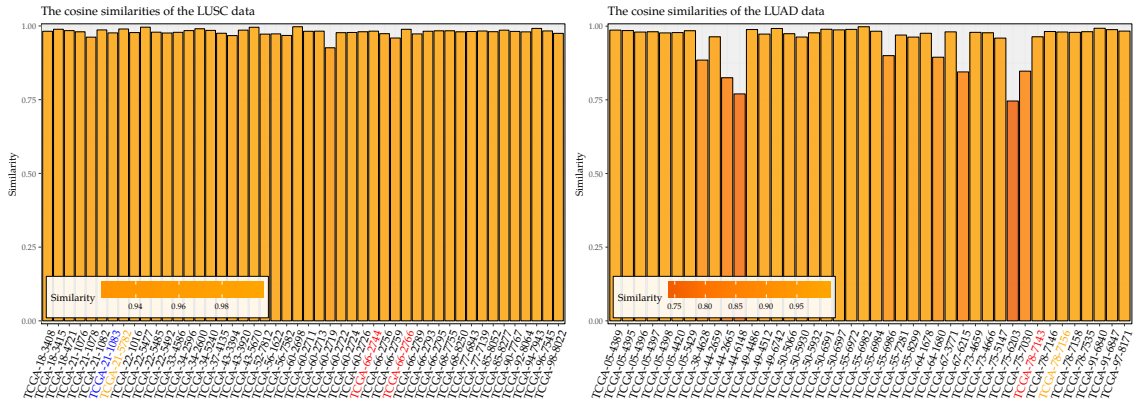
- Alt. allele supporting reads in the normal = 0
- Alt. allele frequency ≥ 0.05
- FILTER field = "PASS"

The resulting distribution of SNV is displayed in Suppl. Fig. 5.

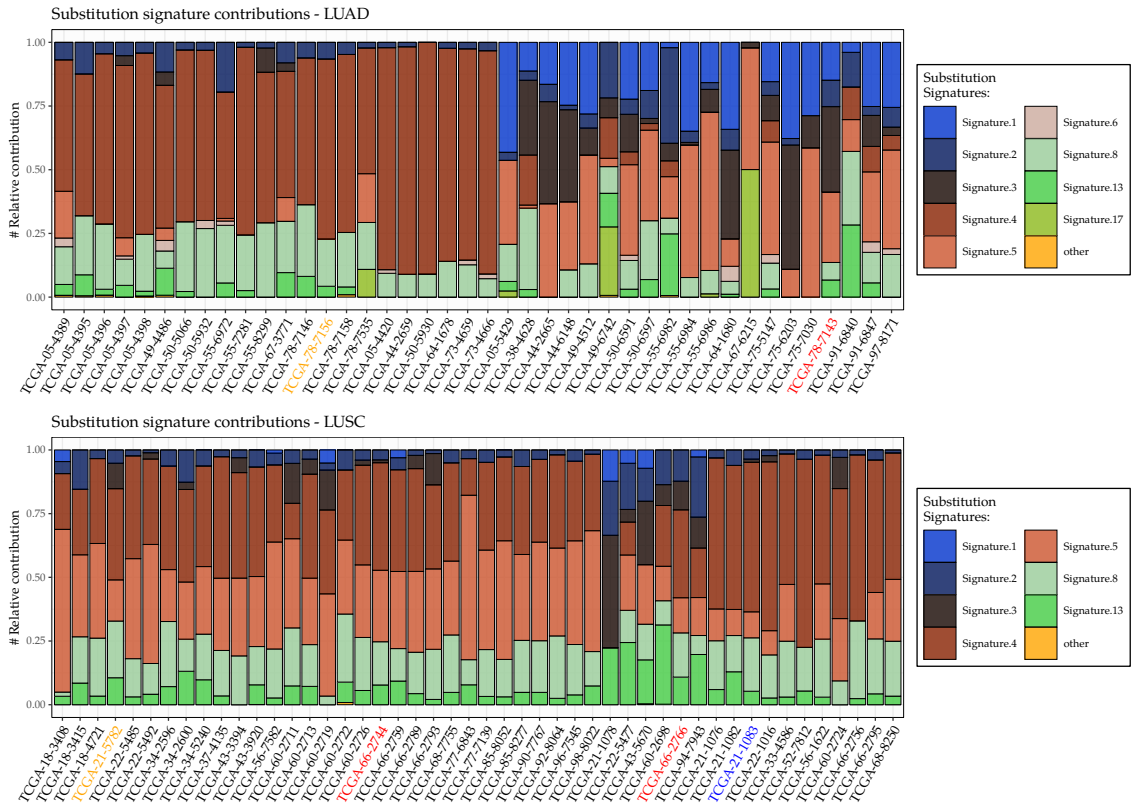
Somatic signatures were extracted with the help of the `deconstructSigs` R package [2]. The list of possible mutational processes whose signatures' linear combination could lead to the final mutational catalogs (a.k.a. mutational spectra) was confined to those, that were reportedly present in lung adenocarcinomas and squamous carcinomas according to the COSMIC database (i.e. in LUAD: Signatures 1, 2, 4, 5, 6, 13, and 17, in LUSC: 1, 2, 4, 5, and 13). Furthermore, since we were primarily interested in their HR-related signature composition, we have added Signature 3 and 8 to the lists. After the evaluation of their signature compositions, the mutational catalogs of the samples were reconstructed, and the cosine of the angle between the 96-dimensional original and reconstructed vectors were measured (cosine similarity). Using this technique, we have also checked whether the incorporation of any additional signatures would improve the mean reconstruction similarities significantly, but the improvement was negligible in both WGS cohorts (Suppl. Fig. 6). In general, the final cosine similarities were adequately high, especially between the original and reconstructed squamous carcinoma whole genomes (Suppl. Fig. 7).
The final mutational signatures can be observed in Suppl. Fig. 8.

3.2 Classification of deletions

It has been shown recently, that cancer cells that exhibit homologous recombination deficiency, have unique characteristics in their indel profiles. Specimens with biallelic BRCA1/2 mutations have significantly more deletions that are longer than 10 bp than BRCA1/2 wild-type tumors, and they also tend to have more deletions than insertions [4]. It has also been found, that these deletions mostly arise due to the activity of the Microhomology Mediated End Joining (MMEJ) or the Single Strand Annealing (SSA) DNA repair pathways, and thus the relative ratio of microhomology mediated (mhm) deletions among them is significantly higher than in HR-competent cases [3]. Since the HR and MMEJ pathways differentiate at the point when RPA binds to the ssDNA overhangs, a dysfunctional BRCA2 protein involuntarily gives rise to an increased MMEJ/SSA activity. Non-surprisingly, the aforementioned increase in the mhm-deletion ratio is much more obvious in samples with BRCA2-/- mutations than in BRCA1-/- tumors.



Suppl.Fig. 7: *Cosine similarities between the original and reconstructed mutational alphabets. Left: LUSC, Right: LUAD*



Suppl.Fig. 8: *Somatic signature composition of the LUAD and LUSC whole genomes. Top panel: LUAD, Bottom panel: LUSC.*

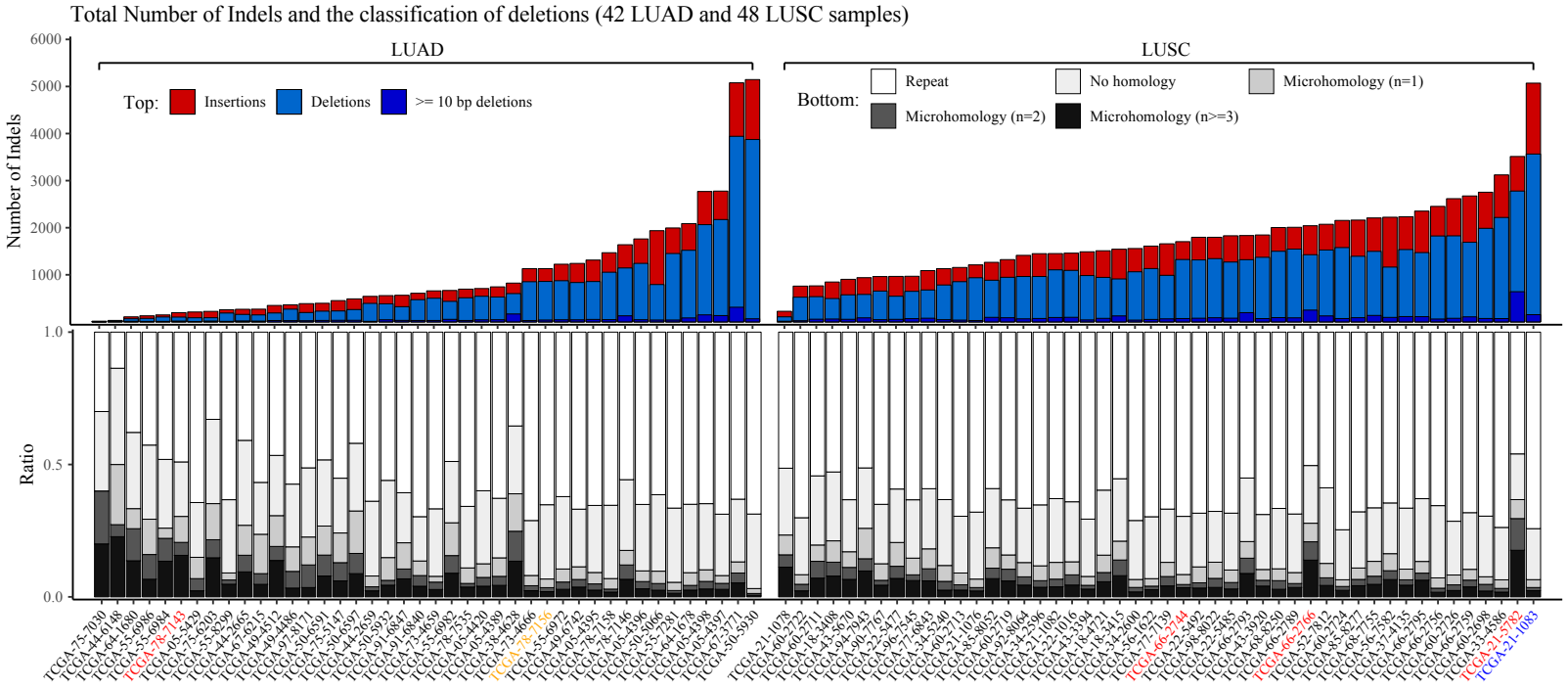
In general, deletions were classified into three sets: (1) complete repetitions; when the complete deleted sequence is repeated after the deletion in the reference genome, (2) microhomologies; when only the first n nucleotides of the deleted sequence is repeated after the deletion and (3) unique deletions, when the sequence following the deletion has no resemblance to the deleted series of nucleotides. However, since the repetition of the first 1-2 nucleotides could occur by pure chance (With 0.25 and 0.0625 probabilities respectively - assuming

that all 4 nucleotides can occur with the same probability), when investigating the effects of the MMEJ/SSA pathway, it is considered a good practice to work with the $n \geq 3$ microhomologies only. Suppl.Fig. 9 provides a summary of this analysis.

3.3 Rearrangement Signatures

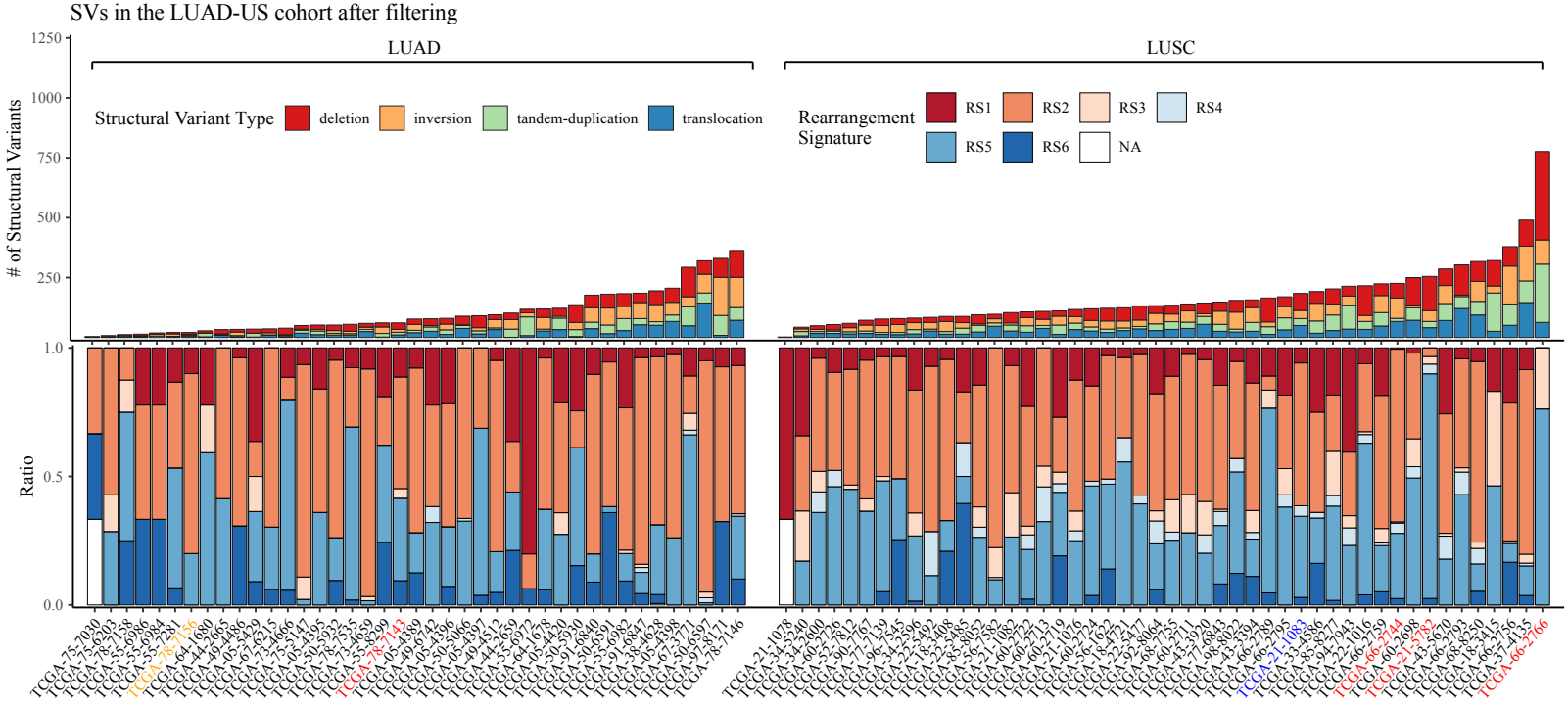
Structural Variants had been called using BRASS (v5.4.1). In the analysis only those variants were considered, whose reads could be denovo-assembled by velvet, and at least 6 read-pairs had supported them.

The resulting structural variants then were mapped to the rearrangement-signature alphabet [10], and a non-negative least-squares strategy was executed to extract their signature composition. Similarly to the substitution signatures, the similarity between the reconstructed and original spectra was quantified using the cosine between their two 32-dimensional vectors. Since the currently available list of rearrangement signatures is based on breast cancer whole genomes, it isn't surprising, that the cosine similarities of the reconstructions were generally low, especially in the LUAD cohort: mean(cosSim(LUAD)) = 0.77 ± 0.25, mean(cosSim(LUSC)) = 0.89 ± 0.07. A summary of the structural variants and their rearrangement signature composition is displayed in Suppl.Fig. 10.



Suppl.Fig. 9: *Top panel: Total number of indels in the LUAD and LUSC whole genomes. The horizontal axis contains the absolute number of hard-filtration-passing indels, the bars are colored according to their insertion/deletion content. Deletions are futher divided into <10 and ≥ 10 bp deletions.*

Bottom panel: Relative constituents of deletions according to the deletion-classification scheme. The three major categories are repeats, microhomologies and deletions without homologous recurrences. Microhomologies are further divided into n=1 bp, n=2 bp, and n≥3 bp variants.



Suppl.Fig. 10: *Top panel: Hard-filter-passing structural variants present in the LUAD and LUSC chohorts. The vertical axis shows the total number of structural variants in each sample, on the horizontal axis samples are sorted according to this number. The colors of the bars indicate the relative ratio of each type of structural variants.*

Bottom panel: Rearrangement signatures in the two lung cancer cohorts. The bars only show the relative compositions, the order of the samples follow the order of the structural variant (top) plot.

3.4 Genomic scar scores

We have used the `sequenza` R-package [5] to estimate the copy number profiles of the non-small cell lung cancer cohorts. Sequenza can utilize the whole context of whole exome and whole genome sequences, and as such requires the original normal and tumor binary alignment files (BAMs) along with the reference fasta (grch37 in the whole genome and grch38 in the whole exome cases) file that was used for the alignment to do its analysis. When it is done, it provides an estimated allele specific copy number profile for the sample, with the segments corresponding to the parental alleles stored in a data frame. The three genomic scar scores had been calculated from these data frames [6, 7, 8]. The scores had been determined using the scarHRD [9] R package.

3.5 HRDetect scores

Since the whole genome cohort was too small, and the whole exome cohort didn't have enough BRCA mutants, we could not train a new logistic regression model. Instead, we used the original, breast-cancer-specific weights. For the whole genomes:

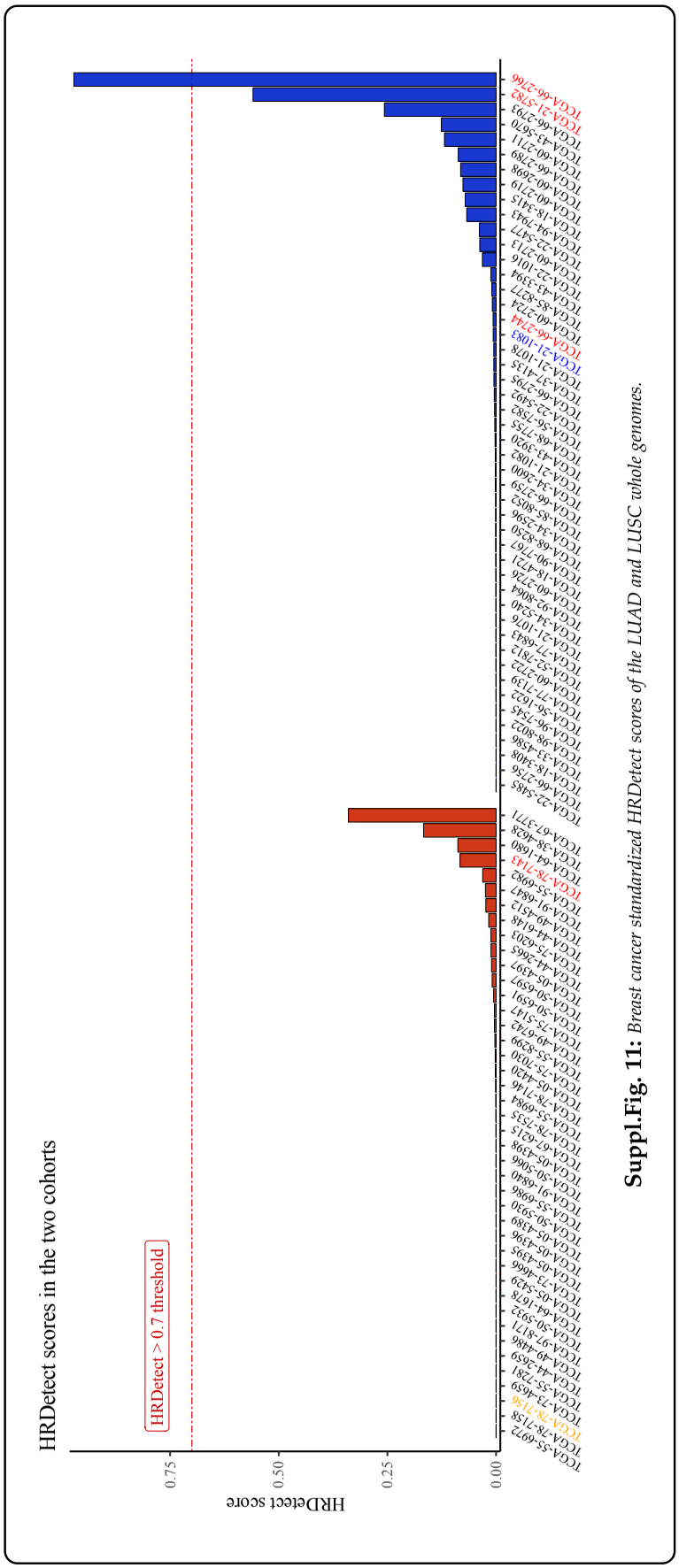
$$\begin{aligned} \text{intercept} &= -3.3642 \\ \text{Signature.8} &= 0.09062 \\ \text{HRD-LOH} &= 0.6666 \\ \text{RS5} &= 0.8467 \\ \text{RS3} &= 1.1532 \\ \text{Signature.3} &= 1.6114 \\ \text{mhm.del.ratio} &= 2.3977 \end{aligned}$$

For the whole exomes, we have used a different model, trained on 560 artificially derived (from whole genomes) breast cancer whole exomes [11]:

$$\begin{aligned} \text{intercept} &= -2.6192939 \\ \text{Signature.17} &= 0.067098 \\ \text{Signature.20} &= 0.09409 \\ \text{Signature.26} &= 0.16166 \\ \text{Signature.6} &= 0.310146 \\ \text{Signature.18} &= 0.31205 \\ \text{mhm.del.ratio} &= 0.314225 \\ \text{Signature.8} &= 0.61474 \\ \text{Signature.13} &= 0.83017 \\ \text{Signature.3} &= 2.00757 \\ \text{HRD-LOH} &= 2.3865 \end{aligned}$$

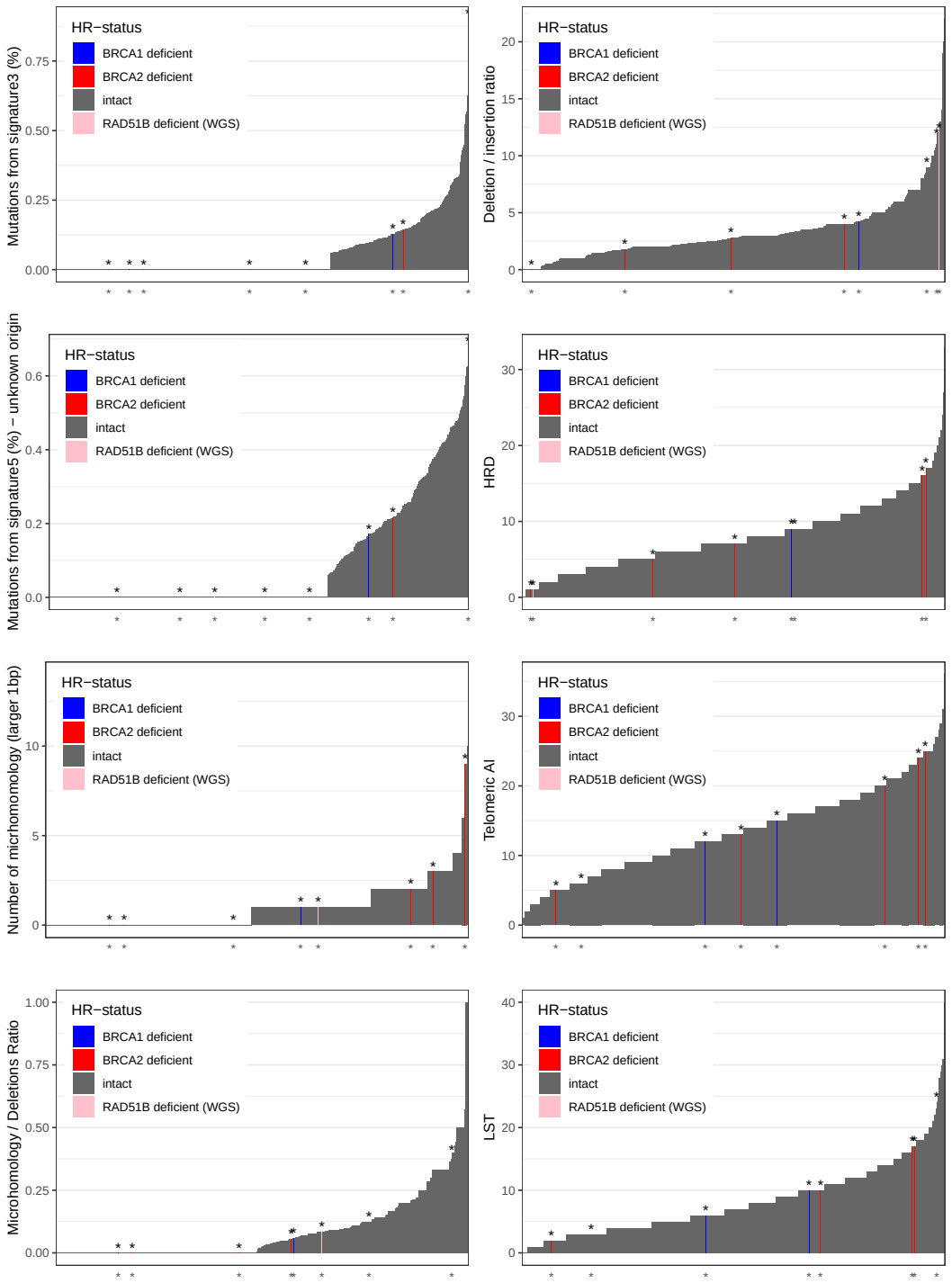
Before they were seeded to the logistic models, sample attributes had been standardized and log-transformed, just as they were in the original paper [12], however this could had been done in two ways. The first, which is reported in the main article is when the standardization step contains the lung cohorts (LUAD or LUSC separately) only. Since the distributions of the HRD-related attributes is most likely different than the distributions present among breast cancer samples, this form of standardization makes more biological sense.

However, we argued, that it is worth to check what would be the HRDetect status of these samples, if we would treat them as breast cancers. In order to check this, the two lung cohorts had been appended to the 560 breast WGS dataset, and the standardization was executed on the resulting 602 (breast + LUAD) and 608 (breast + LUSC) specimens. The resulting distribution of "breast-standardized" HRDetect scores is displayed on Suppl.Fig. 11. In this scenario only a single LUSC sample (TCGA-66-2766) exceeds the 0.7 HRDetect score threshold.

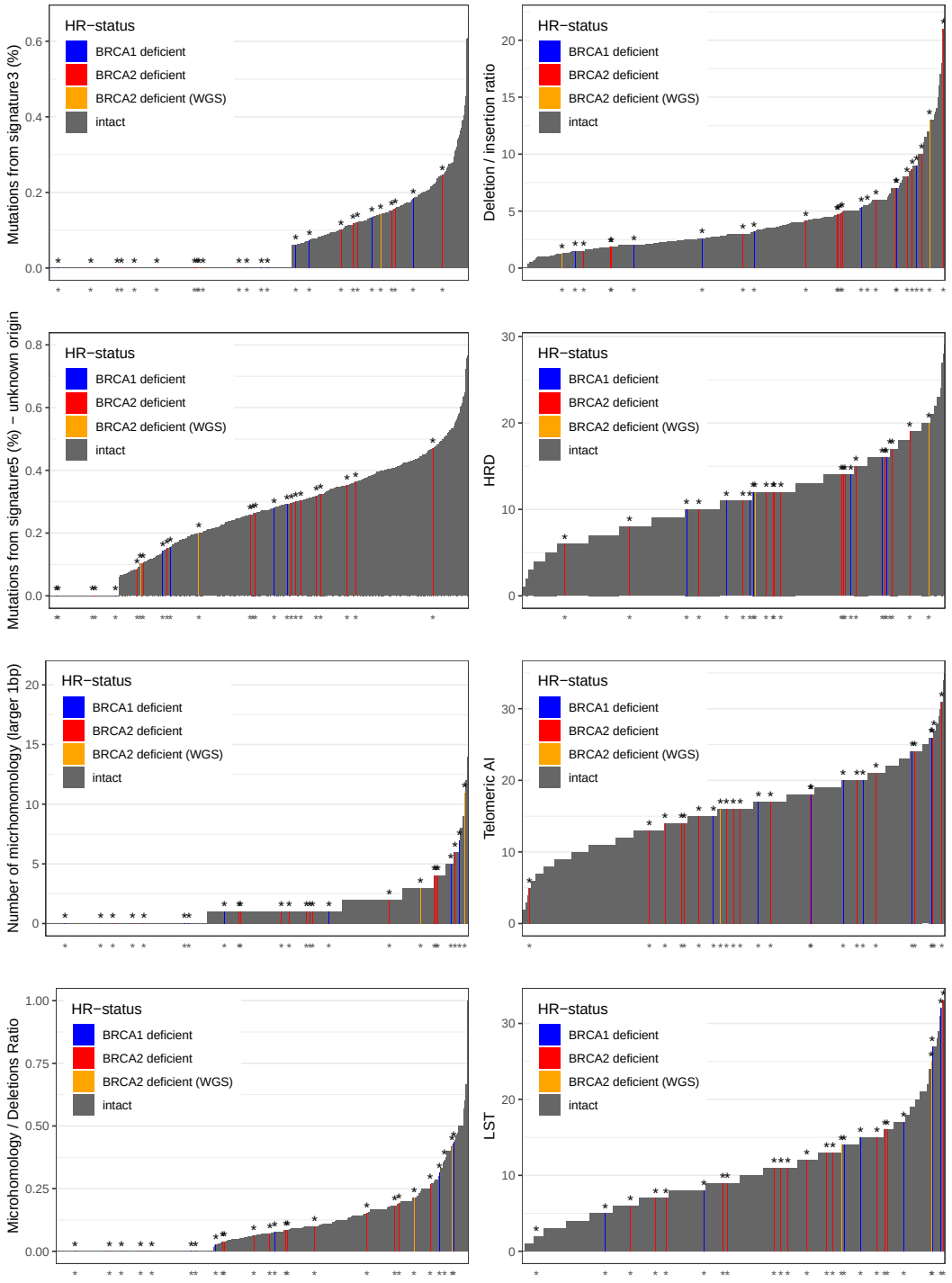


Suppl.Fig. 11: *Breast cancer standardized HRDetect scores of the LUAD and LUSC whole genomes.*

4 HRD-related genomic features extracted from the whole exomes

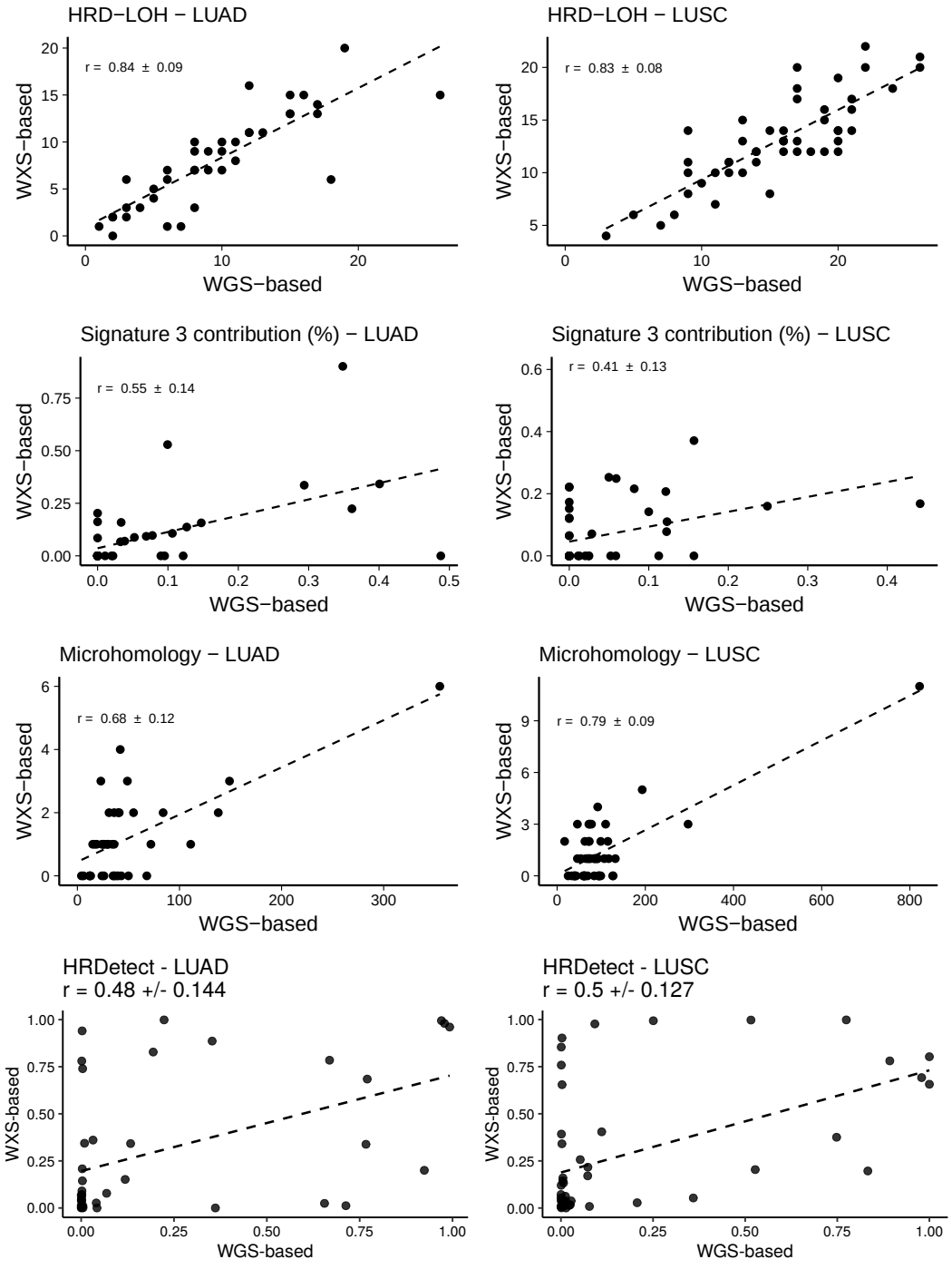


Suppl.Fig. 12: *Distribution of genomic scar scores (HRD-LOH, Telomeric Allelic Imbalance, Large-scale Transitions), homologous-recombination deficiency related mutational signatures (Signature 3, 5), number of micrhomomology-mediated deletions, microhomology / deletions ratio, deletion / insertion ratio and BRCA1/2-status in whole exome sequnced lung adenocarcinoma samples (n=553).*



Suppl.Fig. 13: *Distribution of genomic scar scores (HRD-LOH, Telomeric Allelic Imbalance, Large-scale Transitions), homologous-recombination deficiency related mutational signatures (Signature 3, 5), number of micrhomomology-mediated deletions, microhomology / deletions ratio, deletion / insertion ratio and BRCA1/2-status in whole exome sequnced lung squamous carcinoma samples (n=489).*

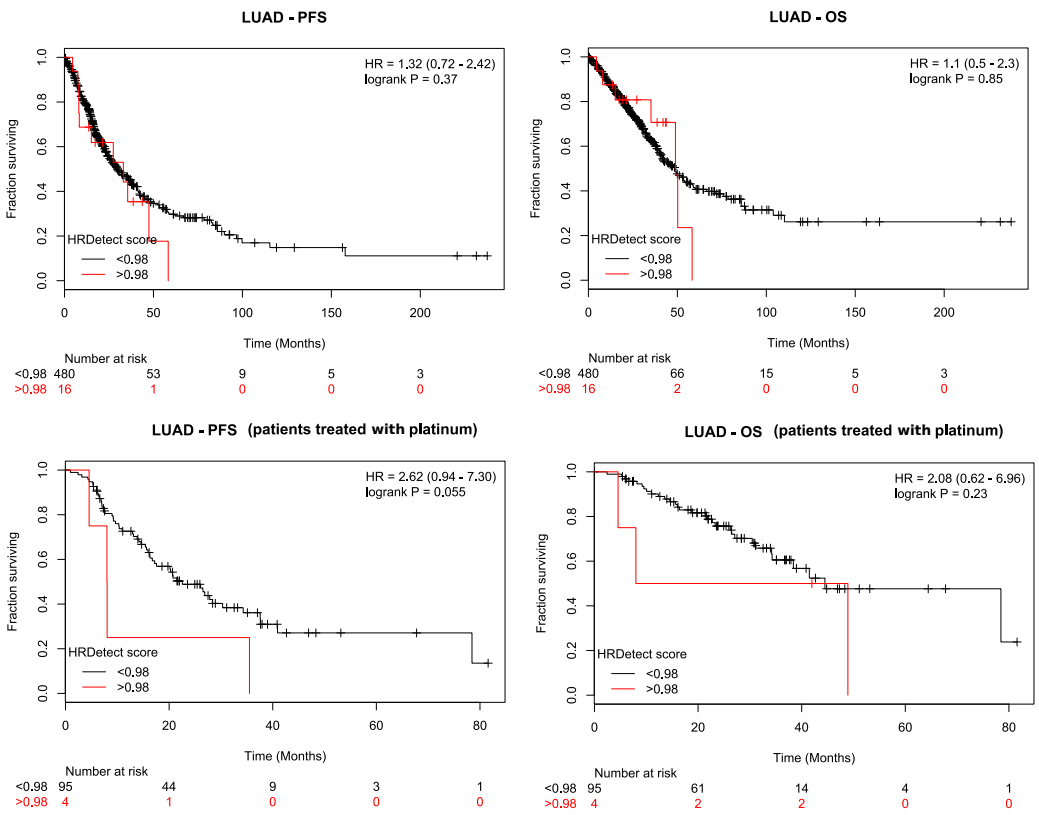
5 Correlations between the WES and WGS HRD-predictors



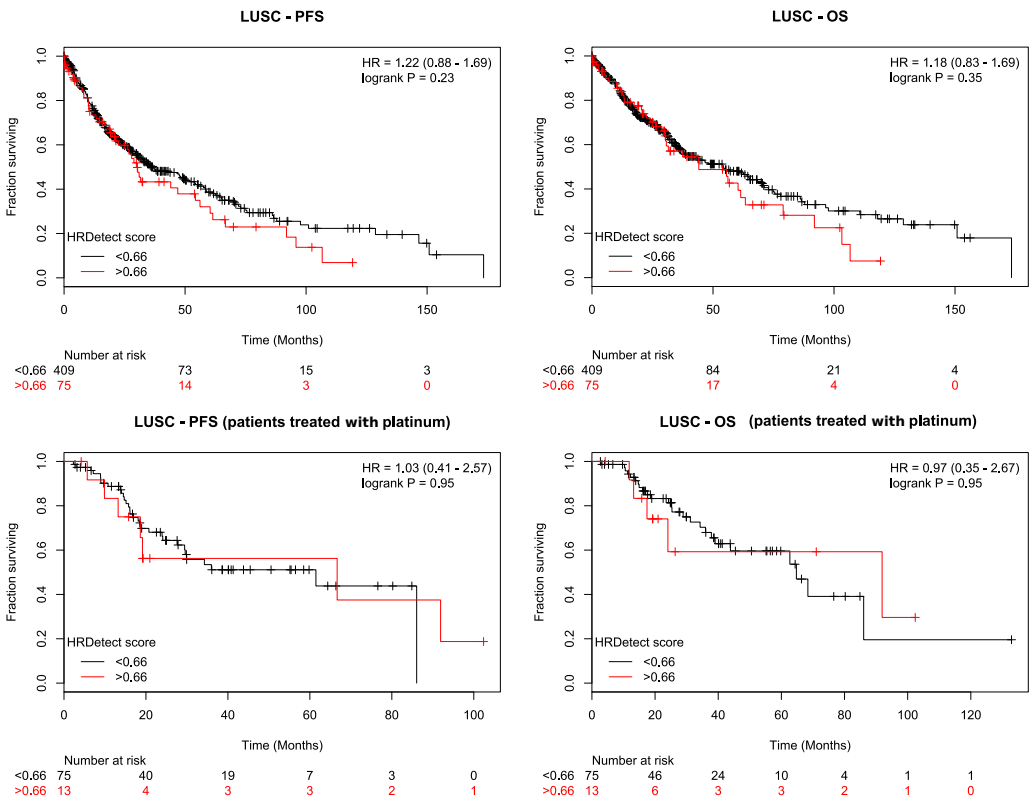
Suppl.Fig. 14: *Correlation of the three main components (number of HRD-LOH events, number of microhomology-mediated deletions, and contribution of Signature 3 to the mutational profile) of HRDetect between paired whole exome and whole genome sequenced samples*

6 Survival analysis

Higher WXS-based HRDetect-score was not associated with better progression free survival (PFS) or overall survival (OS) in LUAD and LUSC patients in the TCGA dataset. There was also no significant difference among the subset of patients who received platinum treatment.



Suppl.Fig. 15: *Survival curves - LUAD*



Suppl.Fig. 16: *Survival curves - LUSC*

References

[1] Alexandrov LB, Nik-Zainal S, Wedge DC, et al. *Signatures of mutational processes in human cancer* Nature. 2013 Aug 22; 500(7463): 415–421

[2] Rosenthal R, McGranahan N, Herrero J et al. *DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution.* Genome Biol. 2016;17:31.

[3] Zámborszky J, Szikriszt B, Gervai JZ, et al. *Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions.* Oncogene 2017;36(6):746-755.

[4] Decker B, Karyadi DM, Davis BW, el al. *Biallelic BRCA2 Mutations Shape the Somatic Mutational Landscape of Aggressive Prostate Tumors.* Am J Hum Genet. 2016; 98(5):818-829

[5] F. Favero, T. Joshi, A. M. Marquard et al. *Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data* Ann Oncol. 2015;26(1):64–70.

[6] Abkevich V , Timms KM , Hennessy BT, et al. *Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer.* Br J Cancer. 2012;107(10):1776-82

[7] Popova T1, Manié E, Rieunier G, Caux-Moncoutier V, et al. *Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation.* Cancer Res. 2012;72(21):5454-62

[8] Birkbak NJ, Wang ZC, Kim JY, et al. *Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents.* Cancer Discov. 2012;2(4):366-375

[9] Sztupinszki ZS, Diossy M, Krzystanek M, Reiniger L, et al. *Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer.* npj Breast Cancervolume 4, Article number: 16 (2018)

[10] Nik-Zainal S, Davies H, et al. *Landscape of somatic mutations in 560 breast cancer whole-genome sequences* Nature volume 534, pages 47–54 (02 June 2016)

[11] Diossy M, Reiniger L, Sztupinszki Z, Krzystanek M, et al. *Breast cancer brain metastases show increased levels of genomic aberration-based homologous recombination deficiency scores relative to their corresponding primary tumors.* Ann Oncol. 2018 Sep 1;29(9):1948-1954.

[12] Davies H, Glodzik D, Morganella S et al. *HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures.* Nat. Med. 2017;23(4):517–525.