Small Panel Detection of Double-Strand Break Repair Pathways to Inform Homologous Recombination Deficiency using Anchored Multiplex PCR and Next Generation Sequencing

ST140

Devin Tauber, Ryan Rogge | Archer, Boulder, CO. USA.

Introduction

Homologous Recombination Deficiency (HRD) is a tumor phenotype stemming from mutations in the homologous recombination (HR) pathway. With HRD, double-strand breaks (DSBs) are repaired through non-homologous, error prone pathways. This can result in structural variants, CNVs, InDels, and genomic instability (GI). Current HRD detection assays require tens of thousands of genomic targets to inform on HRD status. We sought to produce a targeted Anchored Multiplex PCR (AMP™) panel that maintains HRD detection accuracy, while minimizing the genomic space required to do so. Using a publicly available dataset of cancer WGS, we found that certain kinds of InDels were a significant indicator of HRD status. We used this information to develop a 131 primer AMP panel by targeting regions of the genome that are prone to double-strand breakage. Using InDels called from these regions, we have developed a novel method for making HRD calls at a fraction of the size of comparator assays.

Here we introduce two methods which rely on data from the relative activity of double-strand break repair (DSBr) pathways and CNV information to make determinations about HRD status. These methods (see poster ST105) are designed to work for IDT Archer VARIANT*Plex*™ panels and do not require a paired normal sample.

HRD_IR Method Description

Raw reads from a sequenced AMP library are deduplicated using molecular barcodes incorporated during AMP library preparation. Reads are aligned then InDels and CNVs are called by Archer Analysis. InDels called from our HRD_IR primers are then binned by size and length of flanking microhomology into one of three non-homologous DSBr pathways (Non-homologous end joining (NHEJ), Microhomology-mediated end joining (MMEJ), and Single-strand annealing (SSA)).

Using the number of InDels assigned to each repair pathway, we calculate the % contribution of each pathway to the sample. We combine this data with output from our cohortless (no panel of normals required) CNV 2.0 pipeline or ASCN pipeline to produce an HRD status determination. This approach is modular and can be applied to several IDT Archer VARIANT*Plex* panels with increased accuracy when using the ASCN pipeline (see poster #ST105).

Motivation

HRD is a phenotype characterized by a shift in the DSBr pathway equilibrium. When HR is impaired, this produces an upregulation of non-homologous modes of DSBr. We sought to characterize this phenotype by analyzing 577 breast cancer samples (203 HRD+, 371 HRD-) from a publicly available cohort that underwent whole genome sequencing (WGS). HRD status was determined using HRDetect; a publicly available classifier (Davies et al., 2017). We found that among many features analyzed, InDels binned by length and degree of flanking microhomology gave the most significant separation between HRD+ and HRD- distributions (Figure 1). Using this data, we were able to train a classifier with similar performance to HRDetect (Figure 1). We sought to re-create these results with a targeted AMP panel that could supplement IDT Archer VARIANT*Plex* panels with the ability for HRD classification.

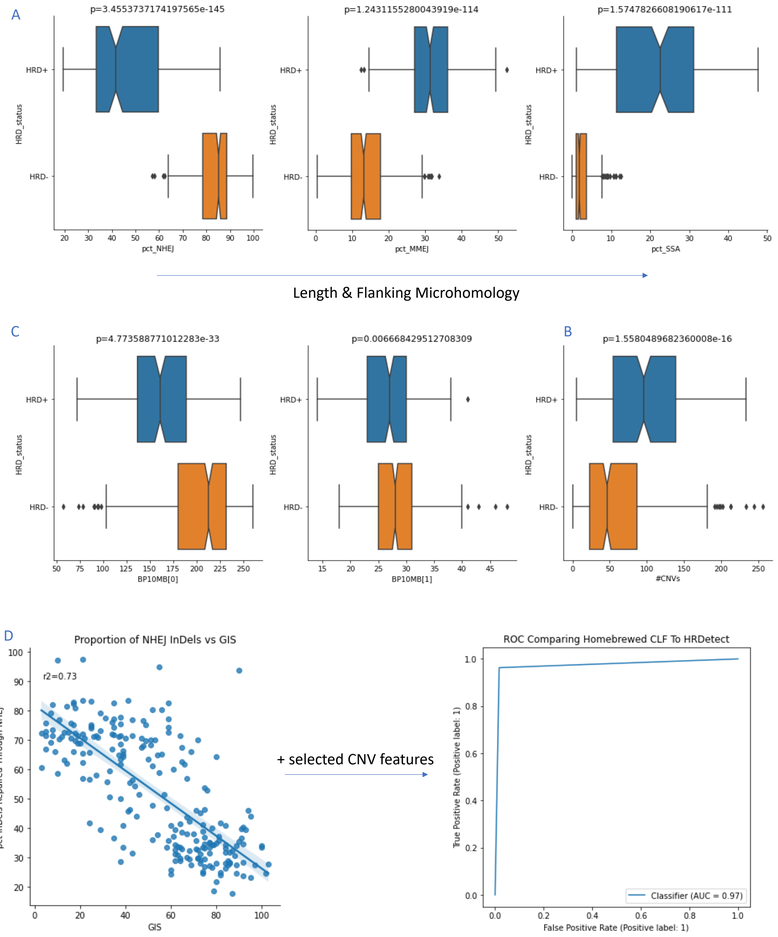

Figure 1. Results of WGS analysis on publicly available dataset characterized by HRDetect.

We analyzed 577 samples from the ICGC data portal that had HRD status annotations provided from HRDetect in order to determine what features were important for HRD classification.

In (a), we found that binning InDels by their length and degree of flanking microhomology gave the greatest degree of separation between HRD+ and HRD- distributions. We characterized InDels by the relative %Activity of three non-homologous DSBr pathways. In concordance with literature, we see an increase in MMEJ and SSA with HRD+ samples.

In (b) we saw an increase in the #CNVs in HRD+ samples. (c) When binning CNVs by the #breakpoints/10MB genome, we also saw an increase with HRD+ status. Albeit not the same degree of significance as our DSBr pathway metrics fueled by InDel information.

In (d) we saw that %NHEJ had a negative correlation to GIS. When %NHEJ is combined with select CNV features, we can train a classifier to detect HRD with comparable accuracy to HRDetect on WGS samples.

Performance

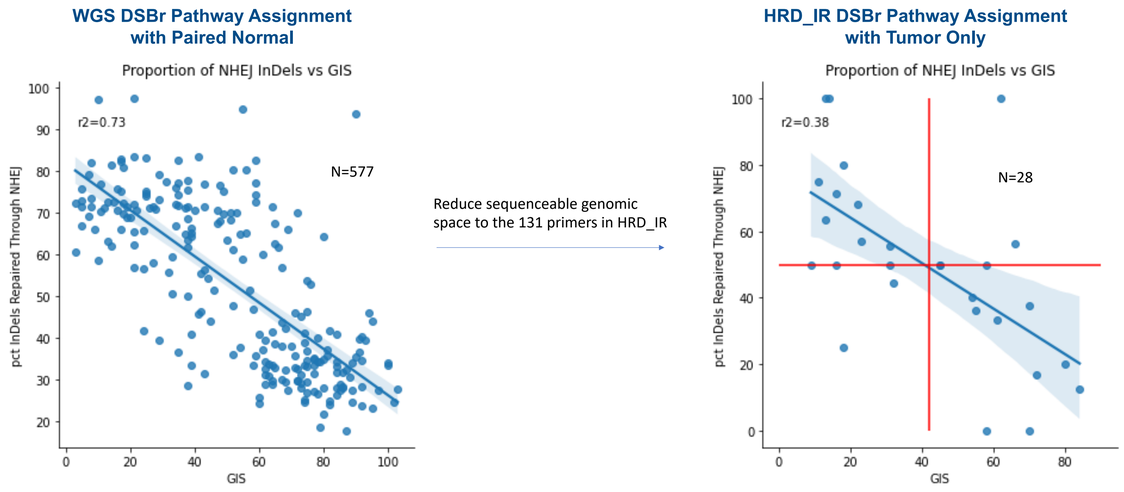

Figure 2. Comparison of %NHEJ Metric from WGS to HRD_IR AMP panel

VARIANT*Plex* Complete Solid Tumor + HRD_IR libraries were prepped from 28 samples of Breast, Ovarian, and Seracare HRD reference standard FFPE DNA with 50 ng input and without paired normal. InDels from HRD_IR primers were passed through a proprietary DSBr pathway assignment program. GIS scores were provided with the samples and were quantified from an orthogonal method. Four samples did not have InDels and therefore could not be quantified. However, this can be overcome by sequencing to the recommended read depth for the paired VARIANT*Plex*™ panel. %NHEJ provided a negative correlation to GIS, comparable to results from WGS (n=577).

Figure 3. CNV Features Detected Using the CNV 2.0 Pipeline in Archer Analysis

The results of CNV detection using VARIANT*Plex* Complete Solid Tumor + HRD_IR. CNVs were detected from 28 breast, ovarian, and Seracare samples without a paired normal. These features are included in the HRD classifier in combination with %NHEJ. (a) Scatterplot of the number of CNV gains detected per sample against an orthogonal GIS. In concordance with WGS, there is a positive correlation between the number of CNV gains and HRD status. (b) Scatterplot of 10MB genomic segments that had a single CNV breakpoint, plotted against an orthogonal GIS. In concordance with WGS, there is a positive correlation between BP10MB[1] and HRD status.

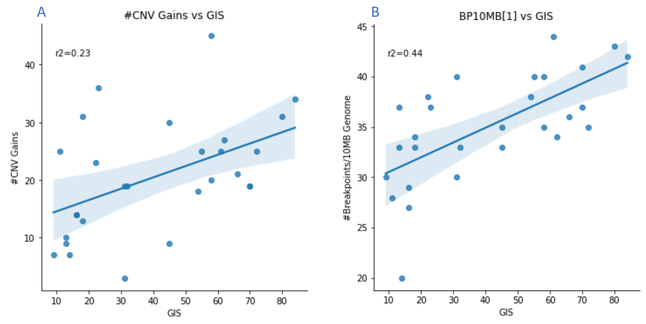

Figure 4. Results from HRD callers on VARIANT*Plex* Data

The results of the new HRD methods for 28 inputs from Breast, Ovarian, and Seracare solid tumor samples. The HRD classifier was trained on 577 samples total, with 206 being HRD+ and 371 being HRD-. The accuracy on WGS validation data was 97%. When applied to VARIANT*Plex* panel data, it achieves an 86% accuracy with the 28 samples. 1 HRD- sample and 3 HRD+ samples were misclassified out of the 28 samples total (sensitivity = 79%, specificity = 93%). (a) displays a scatterplot of Model Probability (>0.7 = HRD+, ≤ 0.7 = HRD-) against an orthogonal GIS, with higher model probabilities being indicative of HRD+. (b) Results from Archer's newly released VARIANT*Plex* HRD Module, which combines data from HRD_IR primers and ASCN output from a 5,000 primer SNP array. Using this method, accuracy increases to 93.75% (PPA = 93.75%, NPA = 93.75%). For more information, see poster #ST105.

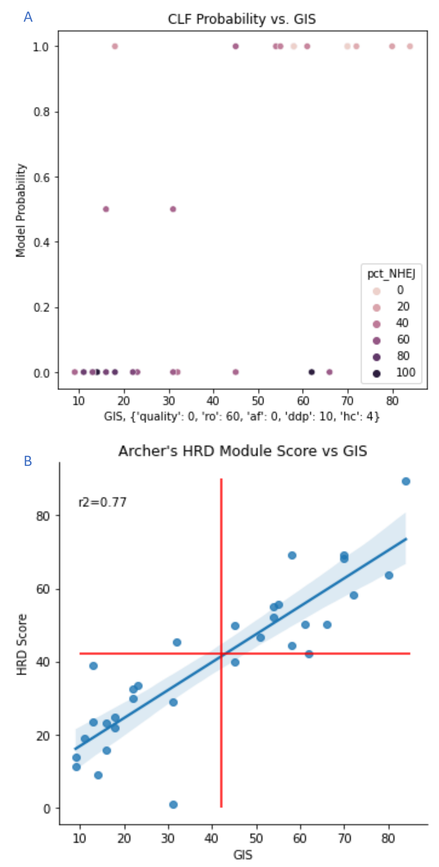

Table 1. Comparison of HRD Module Add-on Size for Common HRD Detection Assays.

Table 1 displays a comparison of Archer's HRD_IR primer panel to other common HRD module add-ons for competitor assays. In general, HRD modules require a large sequenceable genomic space to make determinations about HRD status. Archer's HRD_IR panel is ~0.5% the size of common HRD modules, thereby reducing read requirements and price per sample.

Assay	HRD Module Size	Required Reads
HRDetect	WGS	WGS
Targeted Sequencing	+ 20-25k Probes	~50M
VARIANTPlex™ + HRD Module	+ 5,131 Primers	+ 12M
VARIANTPlex™ + HRD_IR Only	+ 131 Primers	+ 1M

Conclusions

To ascertain features of importance for HRD classification, we analyzed a cohort of 577 HRD annotated WGS cancer samples from a publicly available dataset. We found that a striking feature in determining HRD status came from InDels binned by length and flanking microhomology, with HRD+ samples generally having longer InDels and higher degrees of flanking microhomology. By assigning InDels into specific DSBr pathways, we found a negative correlation between %NHEJ and GIS (Figure 1). We sought to develop a targeted panel to detect this information, resulting in HRD_IR; a small panel composed of 131 primers that mimics results from WGS (Figure 2). HRD_IR, in combination with CNVs from an Archer VARIANT*Plex* panel such as Complete Solid Tumor, can be used to determine HRD status with 86% accuracy (Figure 3). Additionally, HRD_IR primers are available in Archer's newly released HRD module (see poster #ST105), with an accuracy of 93.75%. HRD_IR is only 0.5% the size of other common NGS based HRD modules and does not require a normal cohort, thereby greatly reducing genomic footprint and read requirements to inform on HRD (Table 1).

Devin A. Tauber
Assay Development R&D, Archer (Integrated DNA Technologies)
dtauber@idtdna.com