Comprehensive genomic profiling using VARIANT*Plex*™ and FUSION*Plex*™ panels along with Archer's new tumor-only homologous recombination deficiency algorithm

4615

Glen Bjerke, Taylor R. Patterson, Devin Tauber, Allison Hadjis, Ryan Rogge, Dhruti Legare, Mark Rogers, Laura Johnson, Brent Lutz

Research & Development, Integrated DNA Technologies, Boulder, CO

* Corresponding author: gbjerke@idtdna.com

Introduction

Comprehensive Genomic Profiling (CGP) for tumor profiling involves running one all-inclusive Next Generation Sequencing (NGS) assay that profiles all necessary indications of tumor biology. These tests typically include a DNA assay, a RNA assay and genome wide biomarkers. Herein, we detail IDT's Archer CGP suite of assays, offering cutting-edge precision oncology research solutions designed to provide a thorough and accurate molecular characterization of tumors. This modular set of assays includes VARIANT*Plex* (DNA) and FUSION*Plex* (RNA) profiling with a full range of biomarkers that offer high concordance with established orthogonal methods, ensuring both reliability and reproducibility in genomic profiling. The newest component of Archer's CGP assays is Homologous Recombination Deficiency (HRD) analysis, which is based on Archer's tumor-only Copy Number Variation 2.0 (CNV2.0) and Allele Specific Copy Number (ASCN) calling along with an independent measure of Non-homologous End Joining (NHEJ). Archer's HRD solution comprises a modular set of chemistry targets that can be added to any VARIANT*Plex* panel and a HRD bioinformatics workflow within Archer Analysis software. Archer's HRD pipeline shows high concordance with orthogonal Genomic Instability Score (GIS) methods and consistent calling across different targeted panels and input masses. Further CGP biomarkers for DNA analyses include Microsatellite Instability (MSI) and Tumor Mutational Burden (TMB) calling. In addition to genome-wide biomarker analyses, all VARIANT*Plex* Solid Tumor (ST) panels include Single Nucleotide Variant (SNV), Insertion/Deletion (InDel), and Internal Tandem Duplication (ITD) calling with panel sizes of up to 460 genes. The RNA components of Archer's CGP offering are FUSION*Plex* panels that allow fusion calling from low purity (>= 10% tumor cellularity) samples along with RNA variant calling, which we show to be consistent with DNA allele frequencies. Furthermore, all VARIANT*Plex* and FUSION*Plex* panels can be customized to allow users the flexibility to make a CGP solution according to their individual needs. Archer's CGP panels offer a complete end-to-end solution including automatable chemistry workflows, Archer Analysis bioinformatics software, integration with Molecular Health's Guide Case Annotation Solution (MH Guide CAS) tertiary analysis platform and now include tumor-only HRD calling alongside the full suite of biomarker testing for RNA and DNA panels.

Methods

IDT's Archer HRD module consists of two pools of Anchored Multiplex PCR (AMP™) primers: (1) a "SNP array" that targets likely heterozygous SNPs spaced approximately every 500 kb across the genome and (2) primers that target inverted repeats that are expected to be more susceptible to double stranded breaks and therefore enriched for genomic scars such as indels. These result in metrics for ASCN entropy and percent of indels repaired by non-homologous end joining (NHEJ) (Figure 1), which combined give an HRD score. ASCN also gives information on Loss Of Heterozygosity (LOH) and produces an estimate of tumor purity and tumor ploidy allowing total and minor absolute copy number estimates from CNV fold changes and SNP allele frequencies. Inverted repeat primers add distinct information about which DNA repair mechanisms were functioning within the sample as inverted repeats can form secondary structures that frequently result in double stranded breaks. When indels are observed at these sites, the length of the indel and the length of flanking microhomology can inform which DNA repair pathway was active, which we summarize as the percent of indels repaired by NHEJ. Percent NHEJ tends to be lower in HRD (Homologous Recombination Deficiency) samples than in HRR (Homologous Recombination Repair capable) samples because a greater proportion of breaks that would otherwise be perfectly repaired by homologous recombination are instead repaired by other, more error prone pathways. HRD scores are calculated by Archer Analysis from the percent of indels repaired by NHEJ and the ASCN entropy score using a multiple linear regression model. Scores are scaled similarly to LOH+TAI+LST methods. A HRD classification is then reported based on the HRD score using user-defined thresholds. By default, Archer Analysis uses two thresholds, <33 for HRD Negative and ≥42 for HRD Positive, allowing for an optional intermediate range between the two thresholds. The Archer Analysis MSI module is a set of primers designed to target 114 MSI loci across the genome. Each MSI site is assayed using the entropy of the site length and is classified as stable or unstable. The MSI algorithm outputs the overall percentage of unstable loci and classifies each sample as MSS (microsatellite stable), intermediate or MSI-H (MSI high). Archer TMB calling uses DNA variants called across the panel Regions of Interest (ROI). Detected variants are classified as somatic or germline using a proprietary machine learning algorithm. A linear model is then run using somatic variant counts and panel size to predict a TMB score that mimics the traditional WES based TMB score. All sequencing data were generated using Archer Analysis bioinformatic software.

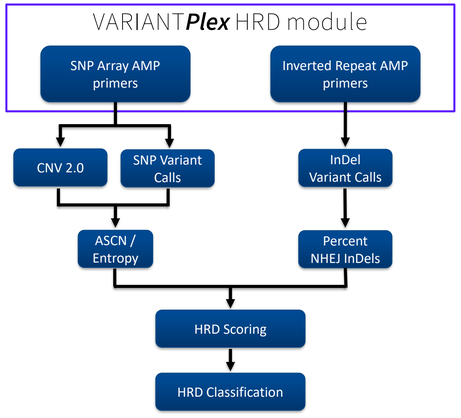


Figure 1. Flow chart of the Archer Analysis HRD algorithm. The HRD module contains two pools of AMP primer sets that give distinct information about either (1) the allele specific copy number state across the genome or (2) the status of the cellular double strand break repair pathway. Each primer set results in an independent measure of HRD which are then combined during HRD scoring and classification.

Results

Archer HRD concordance with orthogonal GIS

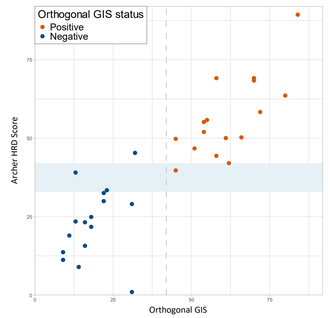


Figure 2. Archer HRD concordance with orthogonal GIS. 32 samples that were previously characterized by an orthogonal assay using LOH+LST+TAI were run using VARIANT*Plex* Complete ST v2 or VARIANT*Plex* Pan ST v2 panels using 50 ng of input. Samples are colored by their GIS positive or negative status from the orthogonal assay. The GIS cutoff for HRD low vs high is 42 (denoted by the grey dotted vertical line). The Archer HRD assay classifies samples < 33 as HRD low, 33-42 as HRD intermediate (shown as a grey box) and > 42 as HRD high. Using a cutoff of 42 for both assays results in one false positive and one false negative for the Archer HRD assay. Positive Percent Agreement (PPA) is 93.75% (15/16 libraries), Negative Percent Agreement (NPA) is 93.75% (15/16 libraries), and the overall accuracy is also 93.75% (30/32 libraries). The sample pool contained 28 FFPE tumor samples, 1 blood sample and 3 FFPE cell line reference samples

Archer HRD consistency between panels

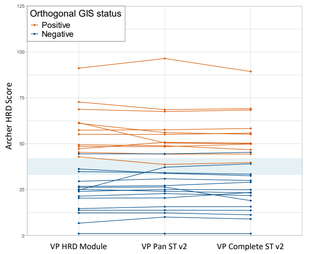


Figure 3. HRD score is consistent across two panels and the HRD module alone. 28 samples prepped with VARIANT*Plex* Complete ST v2 (as in Figure 2) were reanalyzed using the panel files for VARIANT*Plex* Pan ST v2 or the VARIANT*Plex* HRD module. Samples are colored by the orthogonal GIS score (positive / negative with a cutoff of 42). The Archer HRD intermediate zone is shown as 33-42.

Archer HRD consistency across input masses

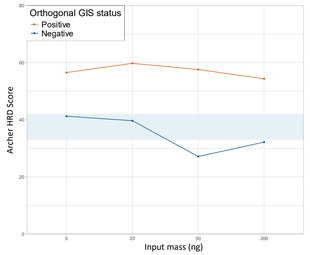


Figure 4. HRD score is consistent across input masses. Seraseq FFPE HRD High-Positive (orange) and Seraseq FFPE HRD Negative (blue) reference materials were prepped using 5, 20, 50 or 200 ng of input mass and the VARIANT*Plex* Complete ST v2 panel, sequenced and analyzed using Archer Analysis. The HRD High-Positive and HRD negative materials have GIS scores of 72 and 31, respectively.

ASCN visualization examples

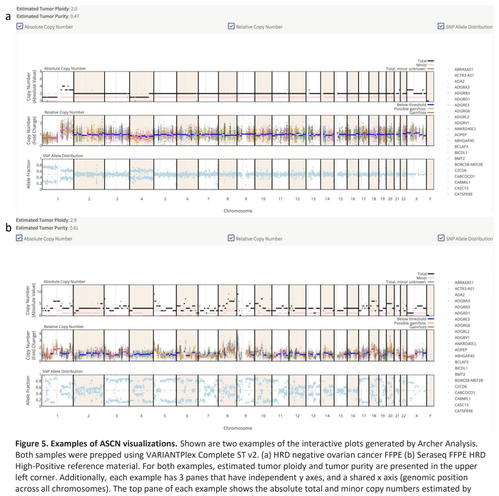


Figure 5. Examples of ASCN visualizations. Shown are two examples of the interactive plots generated by Archer Analysis. Both samples were prepped using VARIANTPlex Complete ST v2. (a) HRD negative ovarian cancer FFPE (b) Seraseq FFPE HRD High-Positive reference material. For both examples, estimated tumor ploidy and tumor purity are presented in the upper left corner. Additionally, each example has 3 panes that have independent y axes, and a shared x axis (genomic position across all chromosomes). The top pane of each example shows the absolute total and minor copy numbers estimated by combining the information of the SNP allele frequencies, CNV fold changes, estimated tumor purity, and estimated tumor ploidy. The middle pane shows the fold change for each primer (as colored points) and the lines denote segments of primers of similar fold change as determined by the tumor-only CNV 2.0 segmentation process. The distinction between "Possible gain/loss" and "gain/loss" in the figure legend and segment coloring is the significance level based on user defined alpha thresholds. In the bottom pane each point is a SNV, and the y-axis is the allele frequency (AF). A dashed line is drawn at 0.5, where one would expect heterozygous SNVs to be. The HRD negative sample (a) shows relatively few CNVs, as most of the genome has two total copies and one minor copy. However, the HRD positive sample is very unstable, indicated by the many different copy number states, CNVs, and allele frequency splits away from 0.5. The HRD score for the HRD negative ovarian cancer FFPE (a) is about 1. The HRD score for the Seraseq FFPE HRD High-Positive reference material (b) is about 58.

Microsatellite instability

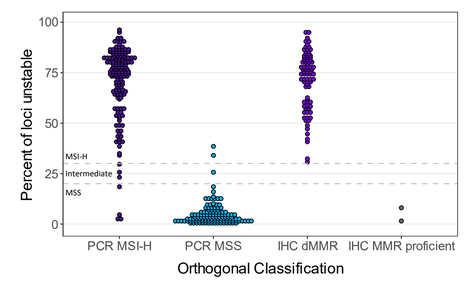
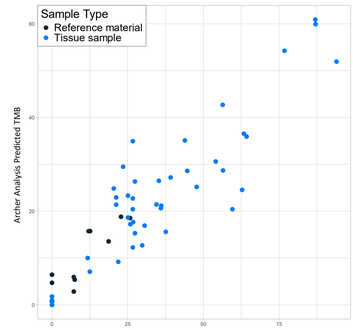


Figure 6. Archer Analysis Microsatellite Instability (MSI) calling. Shown above on the y-axis are Archer MSI scores from 362 de-identified FFPE samples, prepared using 15-250 ng input mass per library and sequenced to 750k reads each. Samples are grouped on the x-axis according to their MSI status from orthogonal assays. The two sets on the left were characterized by a MSI PCR test and the two sets on the right were characterized by Immunohistochemistry (IHC). PCR characterized samples are split into MSI high (MSI-H) and MSI stable (MSS) groups. IHC characterized samples are split into mismatch repair deficient (dMMR) and mismatch repair (MMR) proficient groups. Positive Percent Agreement = 98.3% and Negative Percent Agreement = 98.4%.

Tumor mutational burden



	10 Mutations/Mbp	(-) TMB-Intermediate: 5-20 Mutations/Mbp
Concordance	95.0%	100%
Positive percent agreement	93.2%	100%
Negative percent agreement	100%	100%

Figure 7. Tumor Mutational Burden. 60 samples were prepared using VARIANT*Plex* Pan ST with 50-100 ng of input per sample and sequenced using Illumina® NextSeq500/550™ , NextSeq2000® or NovaSeq® instruments. Performance metrics below the graph are shown using a strict cutoff of 10 mutations per megabase (left) or excluding any sample in the Archer intermediate zone for either NGS assay evaluated (right, 5-20 mutations per megabase for VARIANT*Plex* Pan ST). Reference materials used: 11 SeraCare® TMB Standards, 4 WES-characterized reference cell lines (from ATCC or GIAB). Tissue sample inputs used: 39 de-identified FFPE tumor tissue samples characterized using a 500+ gene hybrid capture NGS assay, 2 de-identified FFPE normal adjacent tissue samples and 4 de-identified peripheral blood samples. Normal samples were assumed to have a TMB of 0.

Fusion calling by percent tumor cellularity

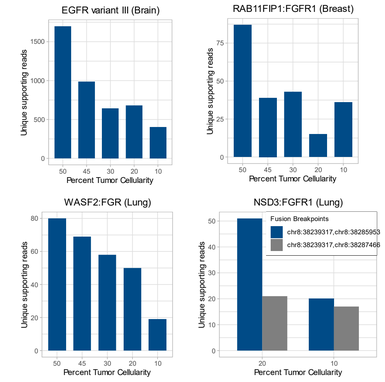


Figure 8. RNA Fusion calling by percent tumor cellularity. Shown are the number of unique reads detected (y-axis) for four oncogenic fusion positive materials across various tumor percentages (x-axis). Each sample was prepared using FUSION*Plex* Comprehensive and 50 ng FFPE input diluted into NAT (normal adjacent to tumor) FFPE material to give the indicated tumor cellularity. Libraries were sequenced on Illumina® NextSeq500/550® instruments and down sampled to between 700 K – 3.5 M reads.

Comparing DNA and RNA variant allele frequencies

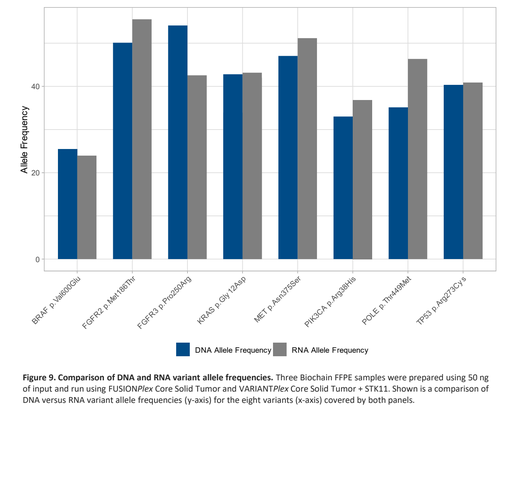


Figure 9. Comparison of DNA and RNA variant allele frequencies. Three Biochain FFPE samples were prepared using 50 ng of input and run using FUSION*Plex* Core Solid Tumor and VARIANT*Plex* Core Solid Tumor + STK11. Shown is a comparison of DNA versus RNA variant allele frequencies (y-axis) for the eight variants (x-axis) covered by both panels.

Archer Panel Options for Comprehensive Genomic Profiling

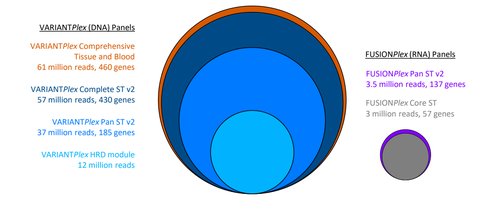


Figure 10. IDT/Archer Solid Tumor panel options for DNA and RNA. Shown is a comparison of panel size for the four VARIANT*Plex* (left) and two FUSION*Plex* (right) catalog panel options for comprehensive genomic profiling. Additionally, all Archer panels can be customized to meet customer needs. The circle size for each panel correlates with the suggested read depth for that panel. All DNA panels shown contain the HRD module and SNV, InDel, ITD, CNV, MSI & TMB calling (except for the HRD module alone). Additionally, the HRD module can be added to any VARIANT*Plex* panel.

Conclusions

Comprehensive genomic profiling is an emerging technology allowing NGS profiling of broad tumor types using one set of assays. Herein, we detail IDT's Archer set of CGP assays that now include HRD, ASCN and CNV 2.0 capabilities in addition to the previously included SNV, InDel, ITD, CNV, MSI, TMB and RNA fusion calling. Archer's new tumor only HRD algorithm results in HRD calls with high concordance to orthogonal methods and consistency between panels and input masses. Archer offers four VARIANT*Plex* and two FUSION*Plex* catalog panels that can be used in any combination or customized to meet any customer's needs. Archer's NGS panels offer a comprehensive solution and incorporate automatable chemistry workflows, Archer Analysis bioinformatics software, integration with Molecular Health's Guide Case Annotation Solution (MH Guide CAS) tertiary analysis platform and now encompass tumor-only HRD calling among a complete set of biomarker assays for RNA and DNA panels.

Glen Bjerke, Ph.D.
Integrated DNA Technologies
gbjerke@idtdna.com