# Supplementary information

[Supplementary information 1](#_Toc27885938)

[Supplementary Materials 1: Tumor Volume Data Analysis 2](#_Toc27885939)

[Tumor Volume Data Analysis Metrics 2](#_Toc27885940)

[Percent Change in tumor volume () 2](#_Toc27885941)

[Area under the tumor growth curve up to time t () 2](#_Toc27885942)

[Adjusted area under the curve () 3](#_Toc27885943)

[RECIST criteria () 3](#_Toc27885944)

[Tumor Growth Inhibition () 3](#_Toc27885945)

[Progression-free Survival () 4](#_Toc27885946)

[Tumor volume data Analysis using R 4](#_Toc27885947)

[Input Data 5](#_Toc27885948)

[Instructions for running tumor volume analysis R scripts 7](#_Toc27885949)

[Supplementary Materials 2: Comparison of Growth kinetics of PDX models at each site 8](#_Toc27885950)

[Supplementary Materials 3: Mouse Read Disambiguation 18](#_Toc27885951)

[Methods 18](#_Toc27885952)

[Simulated Data 18](#_Toc27885953)

[Mouse Read Deconvolution Tools 18](#_Toc27885954)

[Read Alignment 19](#_Toc27885955)

[Results 19](#_Toc27885956)

[Benchmarking of human-mouse read disambiguation 19](#_Toc27885957)

[Supplementary Materials 4: Tumor Normal WES Variant Calling 21](#_Toc27885958)

[Benchmark Datasets 21](#_Toc27885959)

[Workflow testing 21](#_Toc27885960)

[Supplementary Materials 5: Comparison of Tumor-normal whole exome workflow performance for each of the submitted workflows 23](#_Toc27885961)

[Supplementary Materials 6: Workflows selected to process PDX whole exome and RNA-seq data 29](#_Toc27885962)

[Supplementary Materials 7: Gene expression analysis using RNA-Seq workflow 31](#_Toc27885963)

[Description 31](#_Toc27885964)

[Essential Requirements 31](#_Toc27885965)

[Reference Files and Workflow Details 32](#_Toc27885966)

[Workflow Steps and Notable Parameters 32](#_Toc27885967)

[Supplementary Materials 8: Comparisons of xenograft sequence data across PDTCs 34](#_Toc27885968)

# Supplementary Materials 1: Tumor Volume Data Analysis

## Tumor Volume Data Analysis Metrics

Computational definitions of metrics used in this paper follow.

### Percent Change in tumor volume ()

For an individual mouse, the response was determined by computing percent change in tumor volume from baseline to time as follows: % tumor volume change = , where is the tumor volume at time t and is the tumor volume at baseline. For animals for which there is no tumor volume measurement at time t but which have flanking volume measurements at time and such that , then we use linear interpolation to compute the measurement. That is, we compute

Where .

All responses defined below were based on the interpolated tumor volume changes.

### Area under the tumor growth curve up to time t ()

For this measure, we computed the area under the tumor growth curve from baseline up to time t, normalized by dividing by t. With this normalization factor, the interpretation of this measure is the average percent change in tumor size from baseline to time. If there was no tumor measurement at time t but measurements at flanking measurements at time and such that , then we used linear interpolation to estimate the tumor volume at t.

### Adjusted area under the curve ()

We also computed the area under the curve from baseline to the last measurement time point, which for animal i is given by notation . We adjusted the AUC by dividing by the length of the interval between the baseline and the last time point for which a tumor volume was computed for each mouse,, which makes the interpretation of this measure the average percent change in tumor size from baseline to last measurement in the study.

### RECIST criteria ()

This outcome mimics the typical RECIST response criteria commonly used in clinical trials for solid tumors to classify each patient as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). At prespecified time t, each animal was classified into one of these four categories based on their percent tumor volume change from baseline based on cut points , with complete response (CR) when ; partial response (PR) when , stable disease (SD) when , and progressive disease (PD) when .

We considered various sets of cut points (), including (-95,-50,10), (-95,-30,20), (-95,-30,50), (-95,-30,100), (-95,-50,50) and (-95,-50,100), and for our analyses in this paper we computed these at time t=21 days.

### Tumor Growth Inhibition ()

To measure antitumor activity of the treatment group compared to the control group, we considered the tumor growth treatment-to-control ratio (gamma\_t) estimated by one way ANOVA focused on time t. Let be, for animal i, the ratio of tumor size from baseline to time t, i.e. where and are the tumor volumes for animal i at baseline and time t, respectively. Let be an indicator of whether animal i was given the active treatment (=1) or control (=0). After log transforming the ratios, we fit the following linear model:

with

If we define as the ratio of mean tumor/mean control, called the tumor growth inhibition it can be shown that

,

The degree to which this measure is less than 1 indicates the degree of growth inhibition of the treatment. Thus, to test for antitumor activity of the treatment group at time t, we can test the null hypothesis of no treatment effect by comparing

vs. .

### Progression-free Survival ()

Progression-free survival was defined as the time until the tumor volume increases by a multiple of , (e.g. corresponds to time until tumor quadruples in size). This measure will be censored at the last measurement value for animals whose tumors never increased by that multiple**.**

## Tumor volume data Analysis using R

Our analysis scripts for the tumor volume data assume data with a specific format, as outlined here. Given data in this format, our scripts will automatically generate the analyses presented in the methods section as well as certain plots, as outlined below.

### Input Data

1. Model name
2. Treatment  
   : “Control” indicates the control group
3. Mice ID
4. Date  
   : “mm/dd/yyyy” format  
   : The earliest data per animal is the treatment start date, and we assume the tumor volume value is available at this date
5. On Treatment  
   : include two categories (Yes or No)

: “Yes” means “receiving treatment” including treatment start

1. Tumor Volume  
   : Baseline values should be included.
2. Body Weight  
   : Plan for Toxicity analysis if applicable

Note:

* The column names need to match those in the following example table
* **Baseline tumor volume** value is essential per animal.
* For the statistical tests, each treatment group should at least three samples, and the last Treatment Date at least extend to 21 days (or whatever timepoint is chosen) later than the initial Treatment Date per mouse
* Although we perform the analysis for each PDX model, multiple models can be included into the same data file (stacked by row)
* Optional: for displaying the treatment schedule with labeling of the drug names, drug table needs to be provided

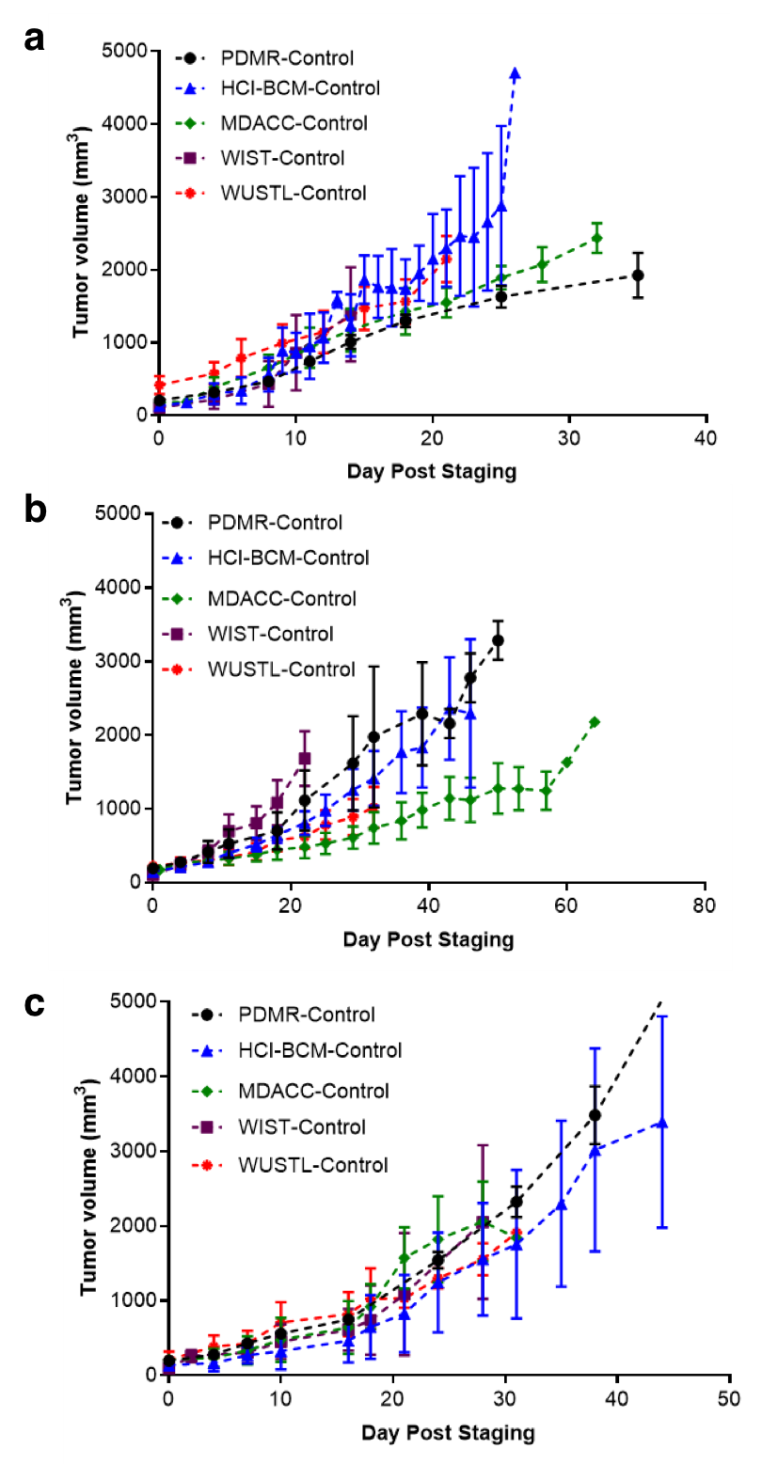
**Supplementary Table 1**. Example data in the specific input format for tumor volume data analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Treatment** | **Mice ID** | **Date** | **On Treatment** | **Tumor Volume** | **Body Weight** |
| 172845-121-T | Control | 104 | 9/28/2015 | Yes | 215 | 21.8 |
| 172845-121-T | Control | 104 | 10/1/2015 | Yes | 308 | 20.3 |
| 172845-121-T | Control | 104 | 10/5/2015 |  | 420 | 19 |
| 172845-121-T | Control | 104 | 10/9/2015 |  | 488 | 19.2 |
| 172845-121-T | Control | 104 | 10/14/2015 |  | 546 | 19.1 |
| 172845-121-T | Control | 104 | 10/20/2015 |  | 842 | 19.2 |
| 172845-121-T | Control | 104 | 10/26/2015 | Yes | 1165 | 19.2 |
| 172845-121-T | Control | 104 | 10/31/2015 |  | 1213 | 19.5 |
| 172845-121-T | Temozolomide | 301 | 9/28/2015 | Yes | 215 | 21.9 |
| 172845-121-T | Temozolomide | 301 | 10/1/2015 | Yes | 309 | 20.7 |
| 172845-121-T | Temozolomide | 301 | 10/5/2015 |  | 337 | 19.5 |
| 172845-121-T | Temozolomide | 301 | 10/9/2015 |  | 307 | 19.7 |
| 172845-121-T | Temozolomide | 301 | 10/14/2015 |  | 252 | 20.1 |
| 172845-121-T | Temozolomide | 301 | 10/20/2015 |  | 237 | 20.1 |
| 172845-121-T | Temozolomide | 301 | 10/26/2015 | Yes | 341 | 20.9 |
| 172845-121-T | Temozolomide | 301 | 10/31/2015 |  | 483 | 21.1 |

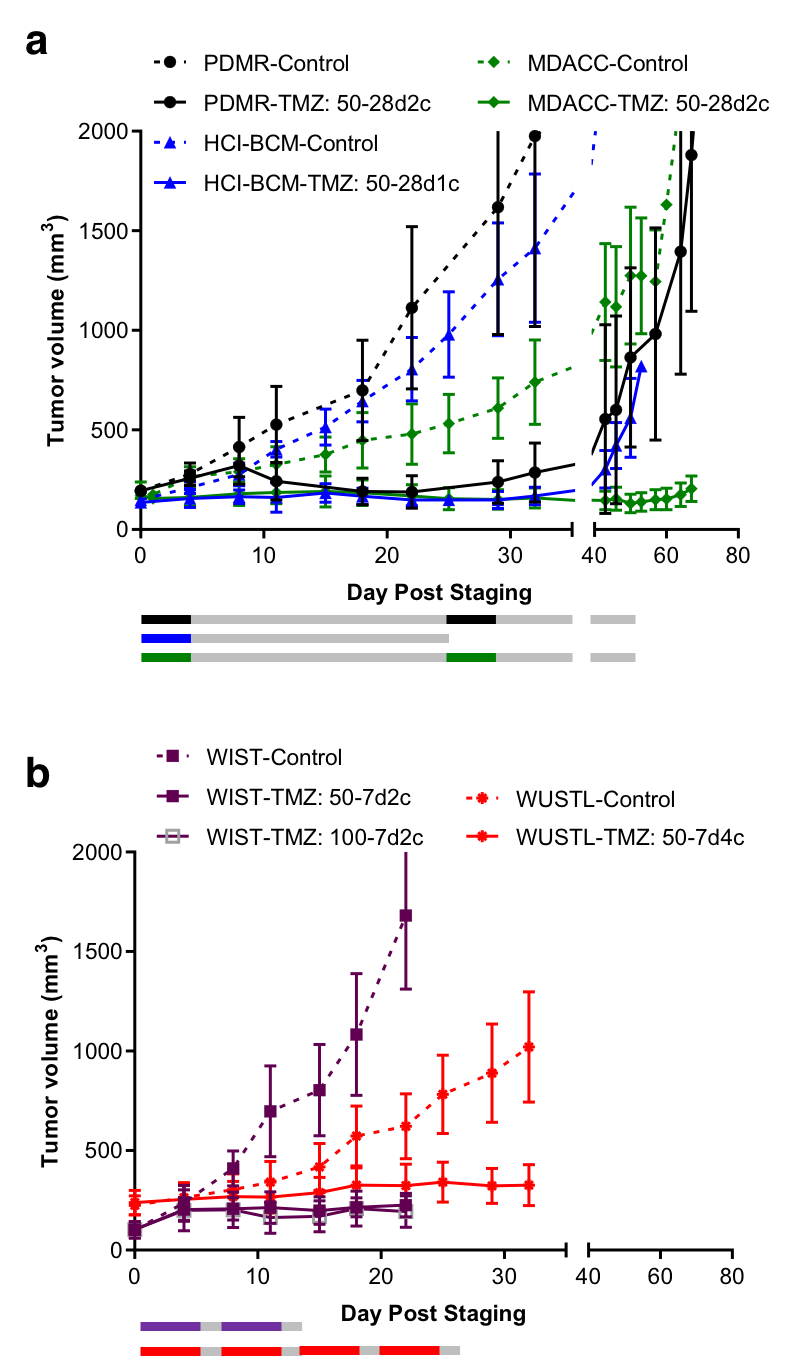
### Instructions for running tumor volume analysis R scripts

A user can execute the tumor volume analysis script within R studio after setting the project title. project path, data file, and the number of days to include in the analysis. The tumor volume analysis script is available from the CGC ( Email cgc@ sbgenomics.com to request access).

# Supplementary Materials 2: Comparison of Growth kinetics of PDX models at each site

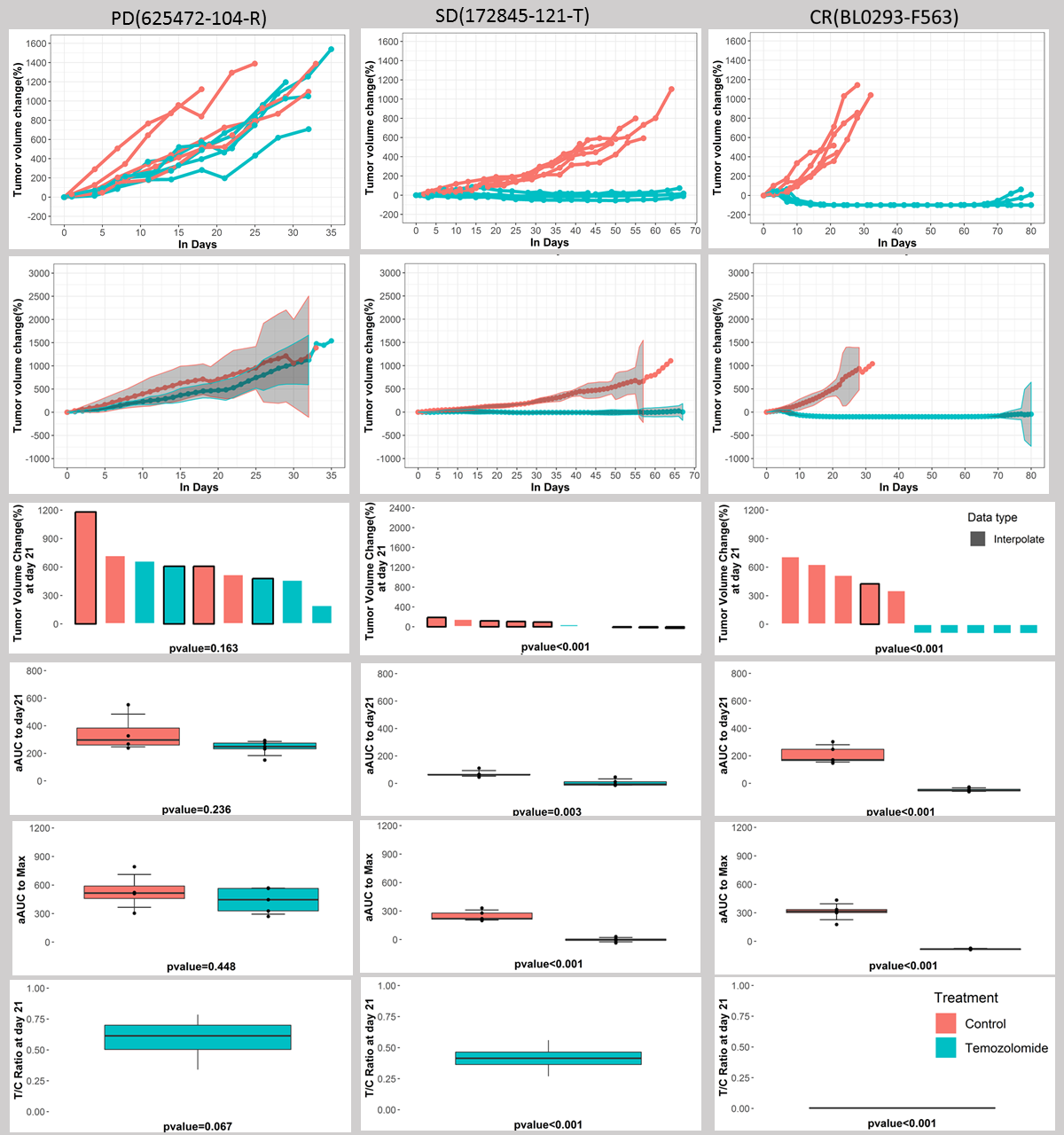
****

**Supplementary Figure 1**: Vehicle control arms at the PDMR and each PDTC for PDX models 625472-104-R **a)**, 172845-121-T **b)**, and BL0293-F563 **c)**. Data represent median tumor volume ± SD from staging of the drug study (day 0) until end-of-study. Tumor volume is not normalized. For statistical assessments, see Figure 3 and Table 2.



Supplementary Figure 2: Comparison of 28-day and 7-day cycle treatment in the intermediary responsive model 172845-121-T. Comparison of tumor volumes observed in a) cohorts treated with temozolomide daily for 5 days on a 28-day cycle (1 or 2 cycles) compared to b) cohorts treated with temozolomide daily for 5 days on a 7-day cycle (2 or 4 cycles). Colored bars represent treatment cycles at each site. All studies achieved tumor growth inhibition irrespective of cycle length. Dashed lines, vehicle control groups, Solid lines, temozolomide treatment groups. Median ± SD.

**Supplementary Materials 3: Drug response assessment of each PDX model at different sites**



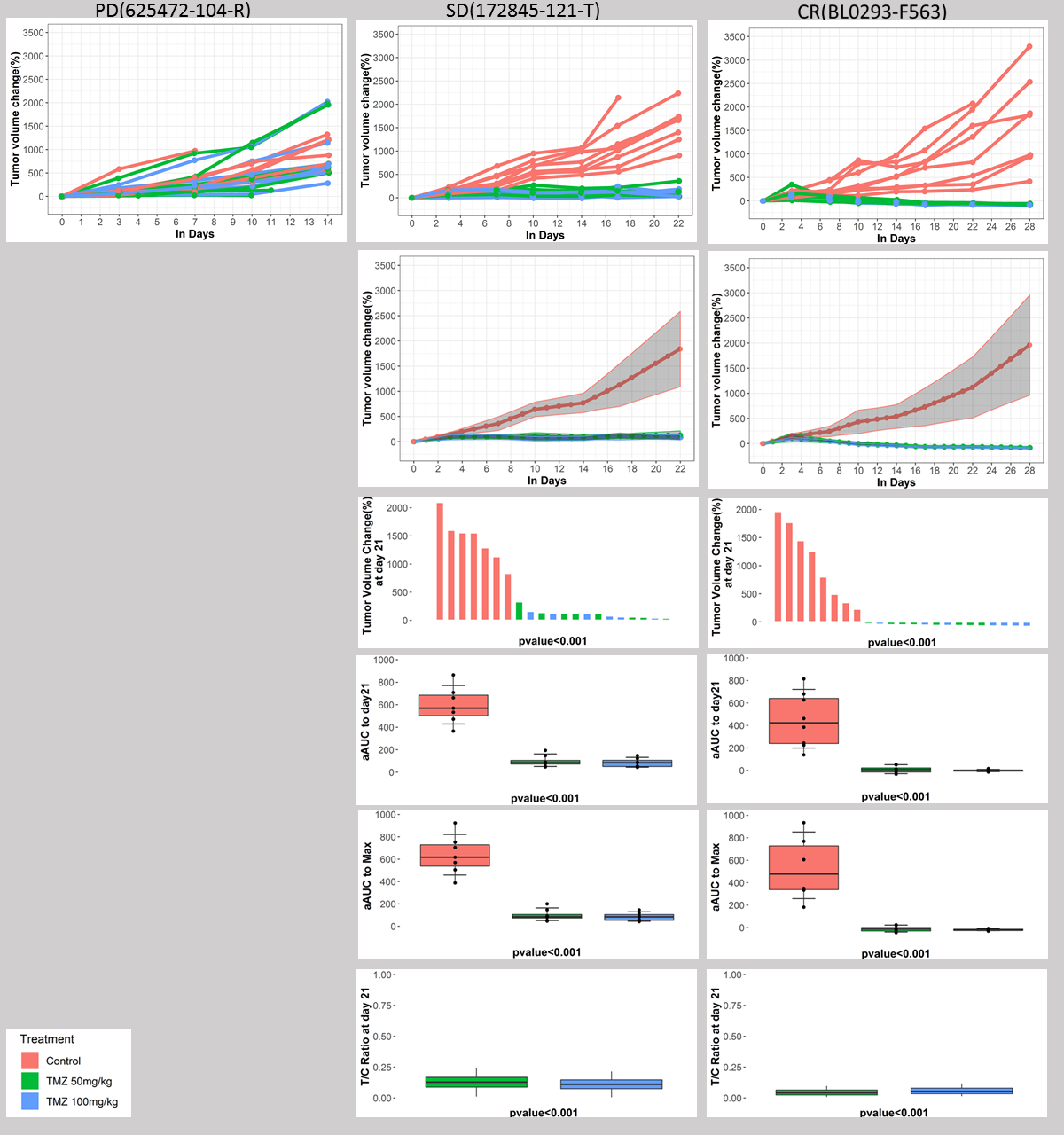
Supplementary Figure 3. Analytical Summaries, MDACC Study. Analytical results from the MDACC study for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively), with interpolated individual curves (row 1), mean curves for treatment and control with 95% confidence bands (row 2), waterfall plots demonstrating (row 3), boxplots of (row 4) and (row 5) for treatment and control, and a boxplot of (row 6), along with p-values comparing treatment to control for each measure.



Supplementary Figure 4. Analytical Summaries, WUSTL Study. Analytical results from the WUSTL study for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively), with interpolated individual curves (row 1), mean curves for treatment and control with 95% confidence bands (row 2), waterfall plots demonstrating (row 3), boxplots of (row 4) and (row 5) for treatment and control, and a boxplot of (row 6), along with p-values comparing treatment to control for each measure.



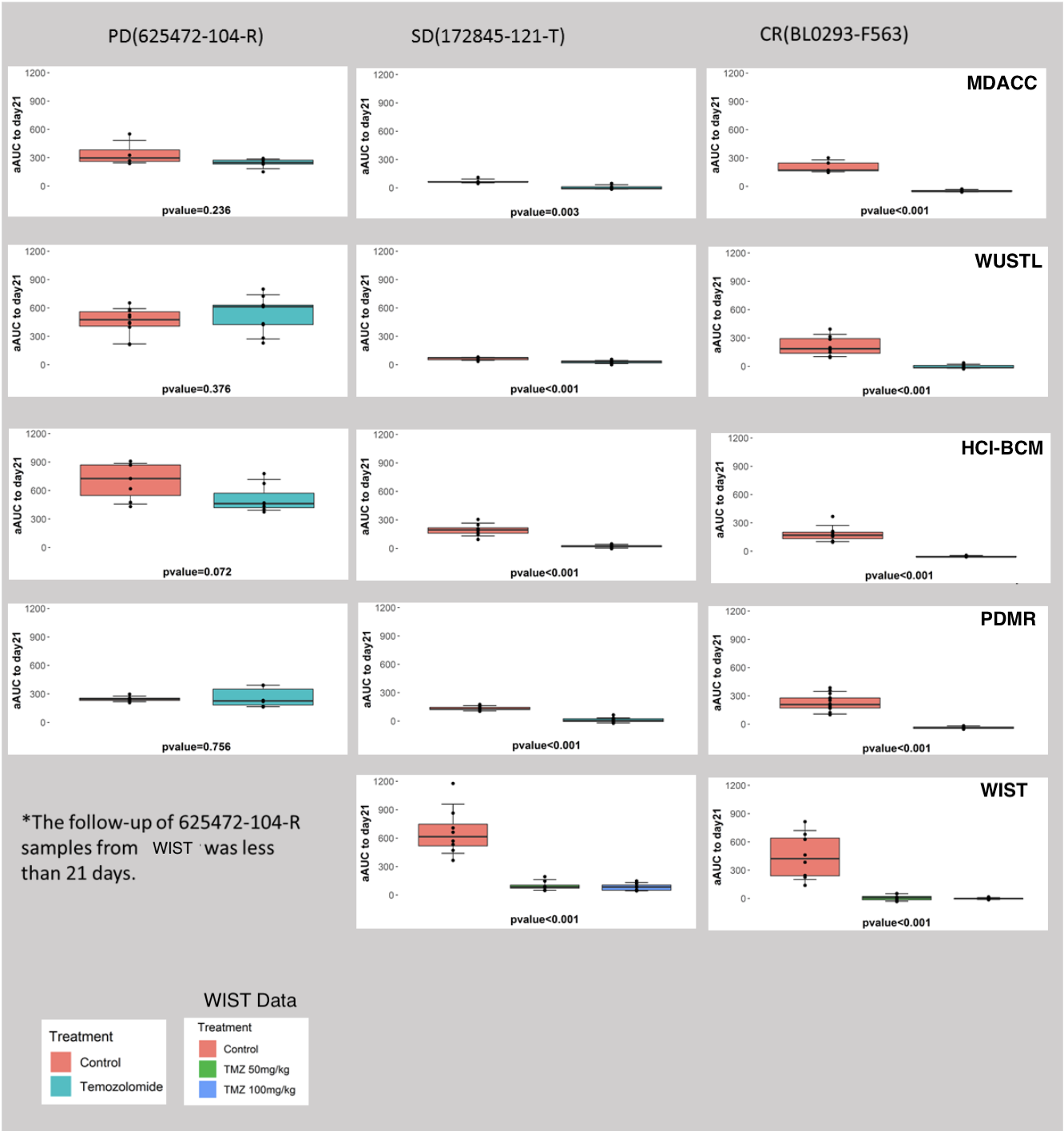
Supplementary Figure 5. Analytical Summaries, PDMR Study. Analytical results from the PDMR study for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively), with interpolated individual curves (row 1), mean curves for treatment and control with 95% confidence bands (row 2), waterfall plots demonstrating (row 3), boxplots of (row 4) and (row 5) for treatment and control, and a boxplot of (row 6), along with p-values comparing treatment to control for each measure.



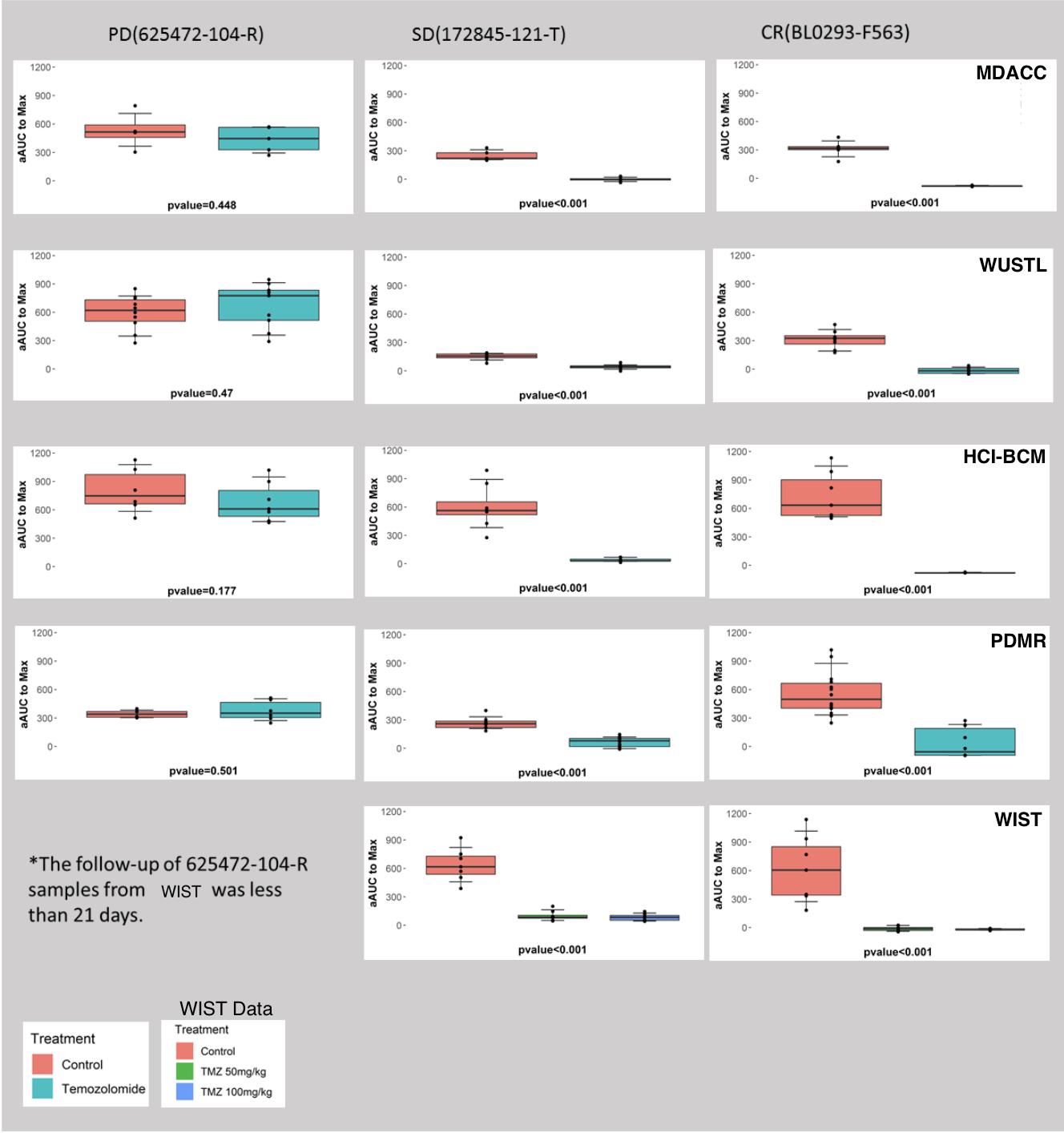
Supplementary Figure 6. Analytical Summaries, WIST Study. Analytical results from the WIST study for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively), with interpolated individual curves (row 1), mean curves for treatment and control with 95% confidence bands (row 2), waterfall plots demonstrating (row 3), boxplots of (row 4) and (row 5) for treatment and control, and a boxplot of (row 6), along with p-values comparing treatment to control for each measure. The 625472-104-R was stopped after 14 days because of the lack of treatment response.



Supplementary Figure 7. Results, Analysis of Change in Tumor Volume . Analytical results of the change in tumor volume from baseline to day 21 from various studies for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively).



Supplementary Figure 8. Results, adjusted area under the curve to 21 days . Analytical results of the adjusted area under the tumor growth curve from baseline to day 21 from various studies for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively). (PDMR - NCI Patient-Derived Models Repository, HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC.)



Supplementary Figure 9. Results, adjusted area under the curve to end date . Analytical results of the adjusted area under the tumor growth curve from baseline to day of last measurement from various studies for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively). (PDMR - NCI Patient-Derived Models Repository, HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC.)



Supplementary Figure 10. Results, tumor growth inhibition to day 21 . Analytical results of the tumor growth inhibition to day 21 measuring the mean ratio between treatment and baseline in terms of ratio of tumor size from baseline to day 21, from various studies for progressive model (625472-104-R), stable disease model (172845-121-T) and complete response model (BL0293-F563) (columns 1-3, respectively). (PDMR - NCI Patient-Derived Models Repository, HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC.)

# Supplementary Materials 3: Mouse Read Disambiguation

## Methods

### Simulated Data

For the WES data benchmark, an experimental WES series of simulated human-mouse mixtures (90:10, 80:20, 60:40, 50:50, 40:60, 20:80) was created by mixing two 100 % pure WES human and mouse samples (sample ids 14311X2 and 14311X8, respectively). A Python script based on HTseq (v0.6.1) was used to create the mixtures by randomly subsampling input pairs of human and mouse FASTQ.GZ files to a predefined fraction to obtain 56-57 million read pairs per data point. We also tested RNA-seq data using a simulated human-mouse mixture series (90:10, 80:20, 60:40, 50:50, 40:60, 20:80) based on a pair of human and mouse lung tissue RNA-seq samples (ENCSR129KCJ and ENCSR870AQU, respectively) (16). The RNA-seq datasets were prepared with the same Python script used for mixing WES data.

### Mouse Read Deconvolution Tools

We compared several tools for mouse-human read deconvolution. These were Xenome (v1.0.0) (17), BBSplit (v37.93) (18), Disambiguate (v1.0; commit c52402a) (18), ICRG (Callari et al., 2018) (19), and XenofilteR (v1.5) (20). BBtools Seal tool was also initially considered; however, it was excluded because of extremely high default RAM requirements. These methods followed three main approaches: BBSplit and Xenome operate on raw FASTQ data, with BBSplit aligning whole reads and Xenome using k-mers to classify reads. Disambiguate and XenofilteR require pre-aligned inputs (i.e. reads aligned to both host and graft reference genomes) and use alignment quality scores for classification. ICRG relies on aligning reads to a combined host-graft reference sequence.

### Read Alignment

For tools requiring aligned data inputs (BAM files), BWA-MEM (20) was used for alignment. For the mouse-human disambiguation step, only reads unambiguously classified as human by a tool were labeled “human.” All other reads were considered “not human” for the true/false positive/negative calling. This classification scheme provided a simplified common framework for evaluating all benchmarked tools. All metrics were calculated via a Python script based on HTseq.

## Results

### Benchmarking of human-mouse read disambiguation

We first compared the efficacy of the five pipelines (**Supplementary Table 2**) for human-mouse read disambiguation using a series of simulated benchmark WES and RNA-Seq datasets. The WES benchmark dataset consisted of a paired end exome-seq data series with human-mouse ratios of 90:10, 80:20, 60:40, 50:50, 40:60, and 20:80 created by computationally mixing two 100% pure WES human and mouse samples (sample ids 14311X2 and 14311X8, respectively). Similarly, the RNA-Seq benchmark dataset consisted of a simulated human-mouse mixture series (90:10, 80:20, 60:40, 50:50, 40:60, 20:80) created from human and mouse lung RNA-seq samples (ENCSR129KCJ and ENCSR870AQU, respectively). The above mentioned simulated WES and RNA-Seq datasets were used to test the five commonly used human-mouse read deconvolution tools: BBSplit, Xenome, Disambiguate, Xenofilter, and ICRG. All tools achieved >99 % precision for both WES and RNA-Seq benchmarks (**Figure 4**). Xenofilter showed the lowest recall (96.60 % and 89.63 % recall in WES and RNA-seq benchmarks, respectively), whereas BBSplit showed the best overall performance i.e. highest precision without any loss in recall (99.87 % and 99.64 % precision in WES and RNA-seq benchmarks, respectively), followed closely by Disambiguate on WES data (99.77 % precision) and Xenome on RNA-seq data (99.77 % precision).

# Supplementary Materials 4: Tumor Normal WES Variant Calling

## Benchmark Datasets

Two simulated whole exome-seq datasets were used in the benchmark for the tumor-normal variant calling workflow. The first dataset (DN) was prepared by researchers from HCI-BCM and consisted of 100x data based on two normal samples, spiked with 30,466 SNPs, 1,723 insertions, and 4,192 deletions sampled from ClinVar, at 0.025, 0.05, 0.1, 0.2, and 0.3 simulated variant allele frequency (VAF) and 10 and 50 % mouse contamination. Germline variants were called with HaplotypeCaller and blacklisted before analysis. The second dataset (BS) was NA12878 WES data (~250x coverage; with 10 % mouse reads contamination) which was spiked with BamSurgeon [i] (default parameters, haplosize 151) at 0.05, 0.1, 0.2, and 0.3 VAF using both the ClinVar variant set used for the other simulated dataset (BS-DN) and 30000 TCGA BRCA SNPs combined with indels from the ClinVar set) (BS-BRCA). For both variant sets, 0.05 and 0.3 VAF samples were also downsampled to 130x and 65x to analyze coverage effects (the experimental coverage of the datasets was 250x).

## Workflow testing

Five tumor-normal WES data analysis workflows from PDXNet research groups were tested on the benchmark sets, as detailed in **Supplementary Table 2**, with the goal of evaluating the accuracy in the presence of variable mouse contamination, coverage, and VAF. Starting from FASTQ data the workflows performed mouse-human disambiguation, alignment, and variant calling with one or more somatic variant callers. For the variant calling step, Mutect2 (21,22), VarScan2 (23), and Strelka2 (9) each featured in two workflows. Manta (24) and Pindel (25) structural variant callers were also used, but were not evaluated, as the benchmark focused on small variants, i.e. SNPs and indels with length <50 base pairs.

Precision/recall and pseudo-ROC curves were used to evaluate the detection of SNPs, insertions and deletions. Pseudo-ROC curves are plots of descending false discovery rate (FDR) vs. true positive rate (recall). To produce the curves, VCF calls are ranked using a caller-provided quantitative score. These were TLOD, SSC, QSS/QSI for MuTect2, VarScan2 and Strelka2, respectively. These rankings allowed us to order the calls for the FDR and TPR calculations. Python scripts were used to calculate the relevant metrics.

For all the submitted workflows, default parameters were used as specified by the workflow authors. Details are provided in the workflows that are accessible through the CGC upon request.

# Supplementary Materials 5: Comparison of Tumor-normal whole exome workflow performance for each of the submitted workflows

All reads were required to meet the quality control cutoff that at least half of the nt positions have >20 phred base quality. We removed adaptors using Trimmomatic v 0.36 [(Bolger et al. 2014)](https://paperpile.com/c/x2k6wS/eXsW). Trimmed reads were aligned to the human genome (build GRCh38.p5) using bwakit v0.7.15 [(Li and Durbin 2009)](https://paperpile.com/c/x2k6wS/4T2O). Mouse reads were removed with xenome v 1.0.0 [(Conway et al. 2012)](https://paperpile.com/c/x2k6wS/os1D) at default parameters. The alignments were converted to BAM format using Picard SortSam v 1.140 (<https://broadinstitute.github.io/picard/>), and duplicates were removed by the Picard MarkDuplicates utility. BaseRecalibrator from the Genome Analysis Tool Kit (GATK) v4.0.5.1 (26,27) was used to adjust the quality of raw reads. Training files for the base quality scale recalibration were Mills\_and\_1000G\_gold\_standard.indels.hg38.vcf.gz, Homo\_sapiens\_assembly38.known\_indels.vcf.gz, and dbSNP v151. Mean target coverage was determined for each sample by Picard CollectHsMetrics, and a MultiQC (28) report was generated. Aligned BAM files were indexed by GATK and passed to GATK Mutect2 v 4.0.5.1. Variants were called in Mutect2 using the Exome Aggregation Consortium (29) database lifted over to GRCh38 as a germline reference with the allele frequency of samples not in reference set to 0.0000082364. Variant calls were then filtered using GATK FilterMutectCalls v 4.0.5.1. Calls were segregated by chromosome with SnpSift Split v 4.3 (30) and annotated by SnpEff v 4.3 (31) using the snpEff v4.3 GRCh38 database. Two additional annotations of variant calls were done with SnpSift dbNSFP using database dbNSFP3.2a, and with SnpSift Annotate using the catalogue of somatic mutations in cancer (COSMIC) v80 database (32). A reference implementation of this workflow is developed and deployed on the CGC.

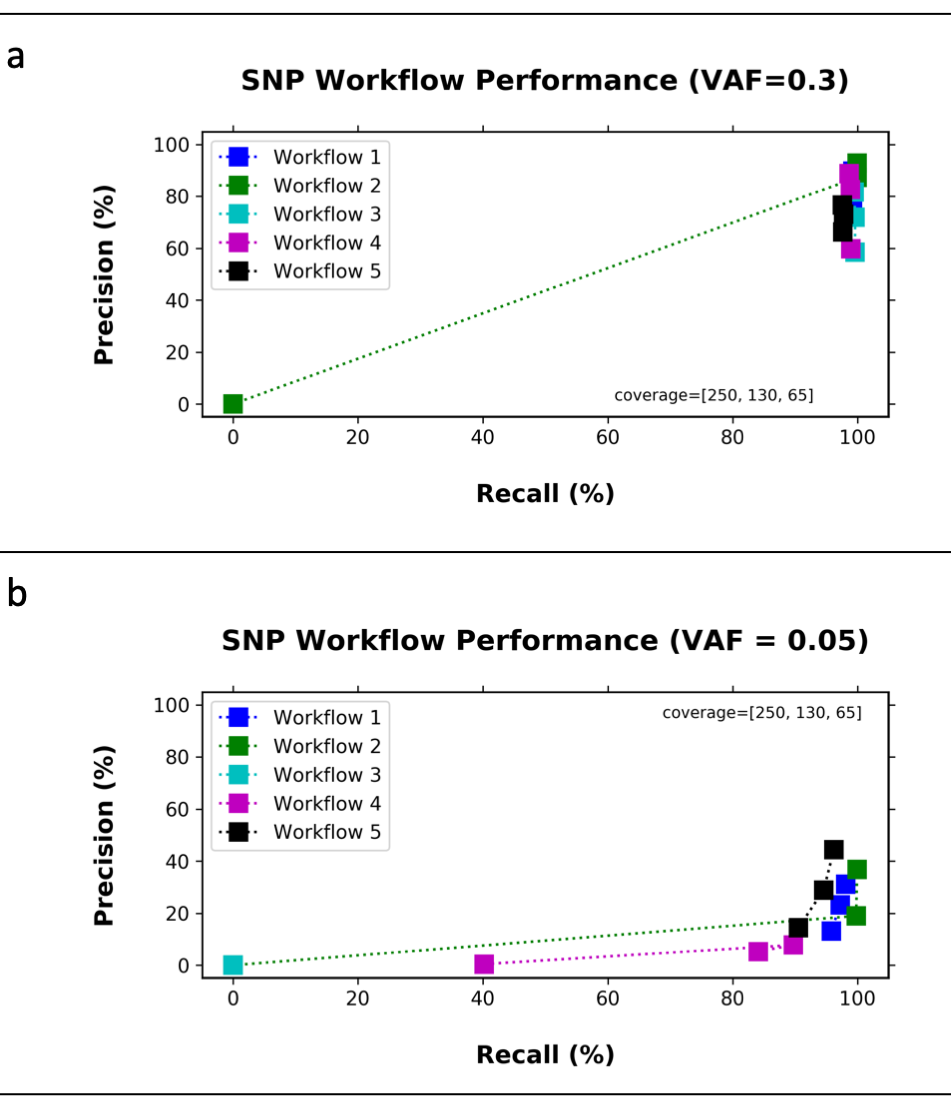
The PDTCs submitted a total of five tumor-normal whole exome variant calling workflows. A summary of the tools used in each of the workflows is described in the **Supplementary Table 2** below. Wiring schematics for each of the submitted workflows are shown in **Supplementary Figures 12 and 13.** Detailed command lines that were used to execute each workflow can be made available upon request. Detailed description about tools that each workflow comprises of can also be made available upon request via access to the workflow(s) on the CGC.

Supplementary Table 2. Main components of Tumor-Normal Whole Exome Workflows submitted by PDTCs

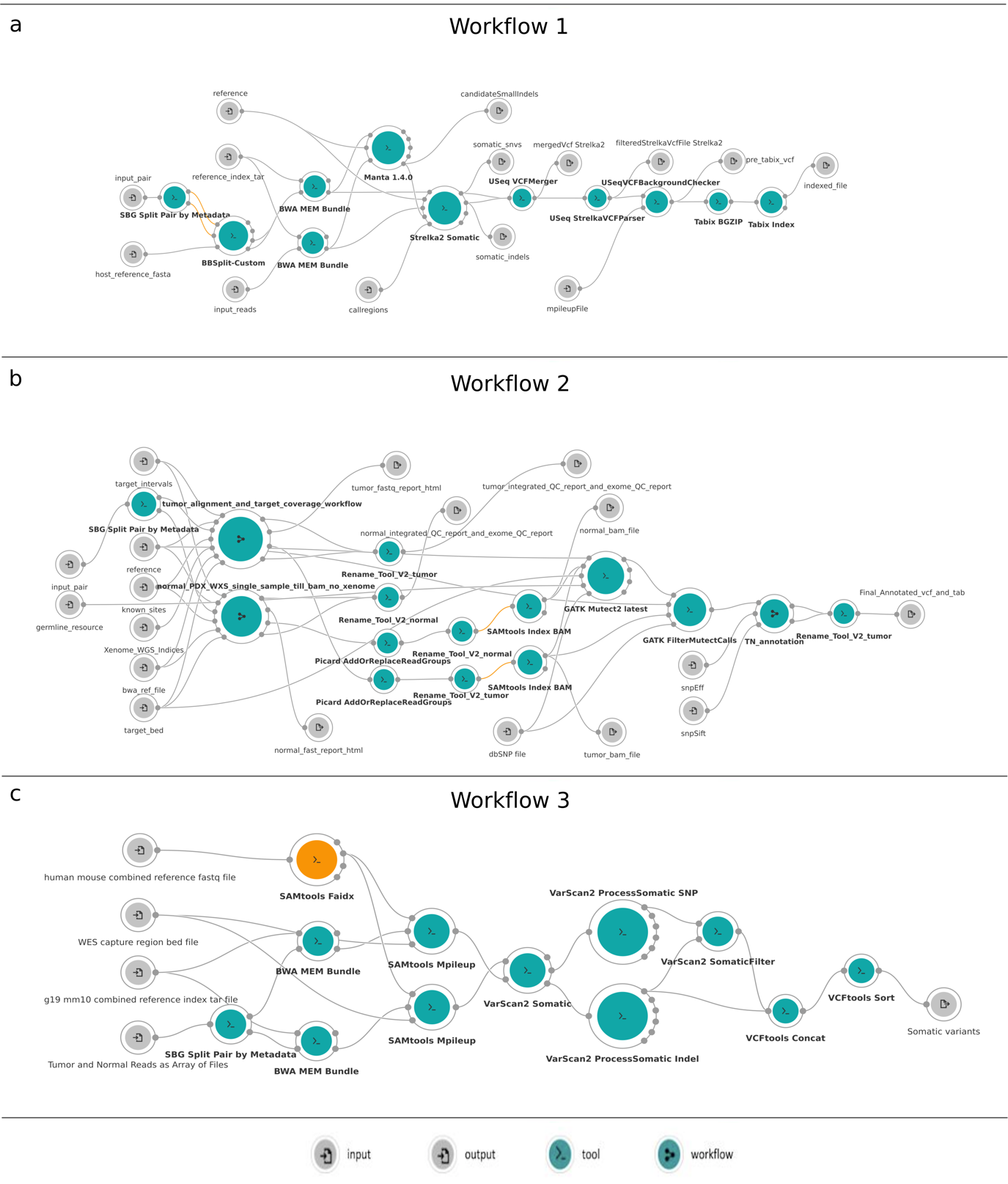
|  |  |
| --- | --- |
| **Workflows submitted by PDTCs** | **Main Components** |
| Workflow 1 | (BBSplit, BWA-MEM), **Manta**, **Strelka2**, USeq, Tabix |
| Workflow 2 | Trimmomatic, Xenome, BWA-alt-aware, **MuTect2**, snpEff, snpSift |
| Workflow 3 | combined human-mouse reference; BWA, samtools, **VarScan2** |
| Workflow 4 | BWA, custom mouse reads filtering; samtools, **MuTect2**, **VarScan2** |
| Workflow 5 | GATK, **VarScan2**, **Pindel**, **Strelka2**, samtools, VEP, snpSift, Picard |

Supplementary Table 3. Performance of the five PDTC workflows across SNPs at 10% and 50% mouse reads contamination.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Workflow 1** | **Workflow 2** | **Workflow 3** | **Workflow 4** | **Workflow 5** |
|  |  | % | % | % | % | % |
| **SNP Precision** |  |  |  |  |  |  |
| 10% Mouse |  | 96.4 | 94.0 | N/A | 2.2 | 56.5 |
| Contamination |  | 98.8 | 97.7 | 1.7 | 12.4 | 71.5 |
|  |  | 99.0 | 98.5 | 39.1 | 13.3 | 75.5 |
|  |  | 99.1 | 98.7 | 98.4 | 33.0 | 76.0 |
|  |  | 99.1 | 98.7 | 99.1 | 45.0 | 76.0 |
| 50% Mouse |  | 96.3 | 94.0 | N/A | 2.2 | 56.3 |
| Contamination |  | 98.7 | 97.7 | 1.6 | 12.4 | 71.3 |
|  |  | 99.0 | 98.5 | 38.3 | 13.3 | 75.3 |
|  |  | 99.1 | 98.7 | 98.4 | 33.0 | 75.9 |
|  |  | 99.1 | 98.7 | 99.1 | 45.0 | 75.8 |
| **SNP Recall** |  |  |  |  |  |  |
| 10% Mouse |  | 23.9 | 20.7 | N/A | 2.6 | 31.7 |
| Contamination |  | 70.9 | 57.2 | 0.0 | 16.9 | 61.3 |
|  |  | 91.3 | 87.4 | 0.5 | 18.4 | 75.2 |
|  |  | 96.6 | 97.2 | 52.2 | 59.0 | 77.7 |
|  |  | 97.3 | 98.6 | 97.2 | 97.9 | 78.0 |
| 50% Mouse |  | 23.9 | 20.7 | N/A | 2.6 | 31.7 |
| Contamination |  | 70.9 | 57.2 | 0.0 | 16.9 | 61.3 |
|  |  | 91.3 | 87.4 | 0.5 | 18.4 | 75.2 |
|  |  | 96.6 | 97.2 | 52.2 | 59.0 | 77.7 |
|  |  | 97.3 | 98.6 | 97.2 | 97.9 | 78.0 |

****

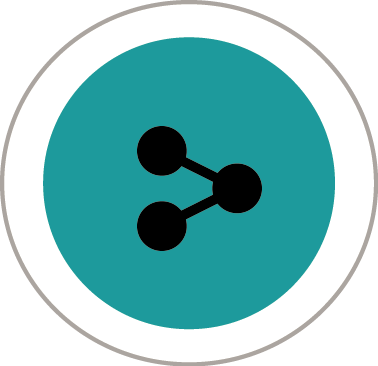
Supplementary Figure 11: Performance of the five PDTC workflows across SNPs at 250x, 130x and 65x coverage at 0.05 or 0.3 variant allele frequency (VAF). Precision and recall across SNPs of a simulated dataset (BS-DN) at 250, 130, and 65 coverage values and at two variant allele frequencies: a) 0.05 VAF and b) 0.3 VAF. As expected, workflow performance improved at higher VAFs and coverage. Workflow 2 that contains Mutect2 as the variant caller performs consistently the best across all samples in SNP calling at both 0.05 and 0.3 VAF.

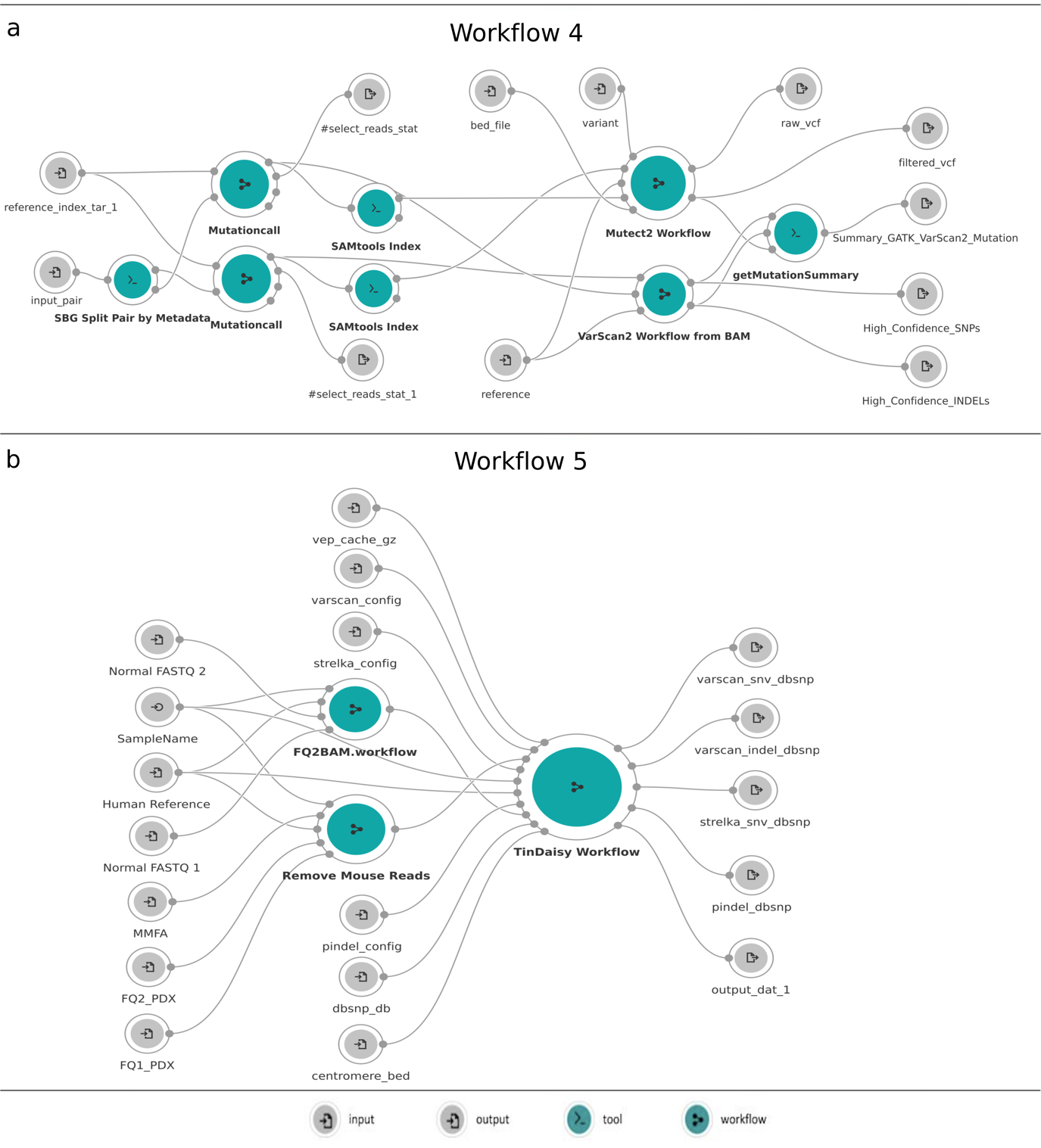


Supplementary Figure 12: Wiring diagrams for submitted whole exome workflows submitted by the PDTCs. Wiring diagrams include nodes and connections. Nodes depict inputs - A picture containing object, clock

Description automatically generated, outputs - A picture containing object, clock

Description automatically generated, tools - A picture containing object, clock

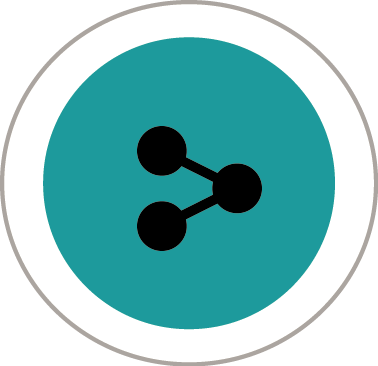
Description automatically generated, and workflows - . Connections between nodes depict that input to a node is from the output of another node. Orange nodes - identify a tool or a workflow with an available update. a) Workflow submitted by PDTC 1, b) Workflow submitted by PDTC 2, c) Workflow submitted by PDTC 3.



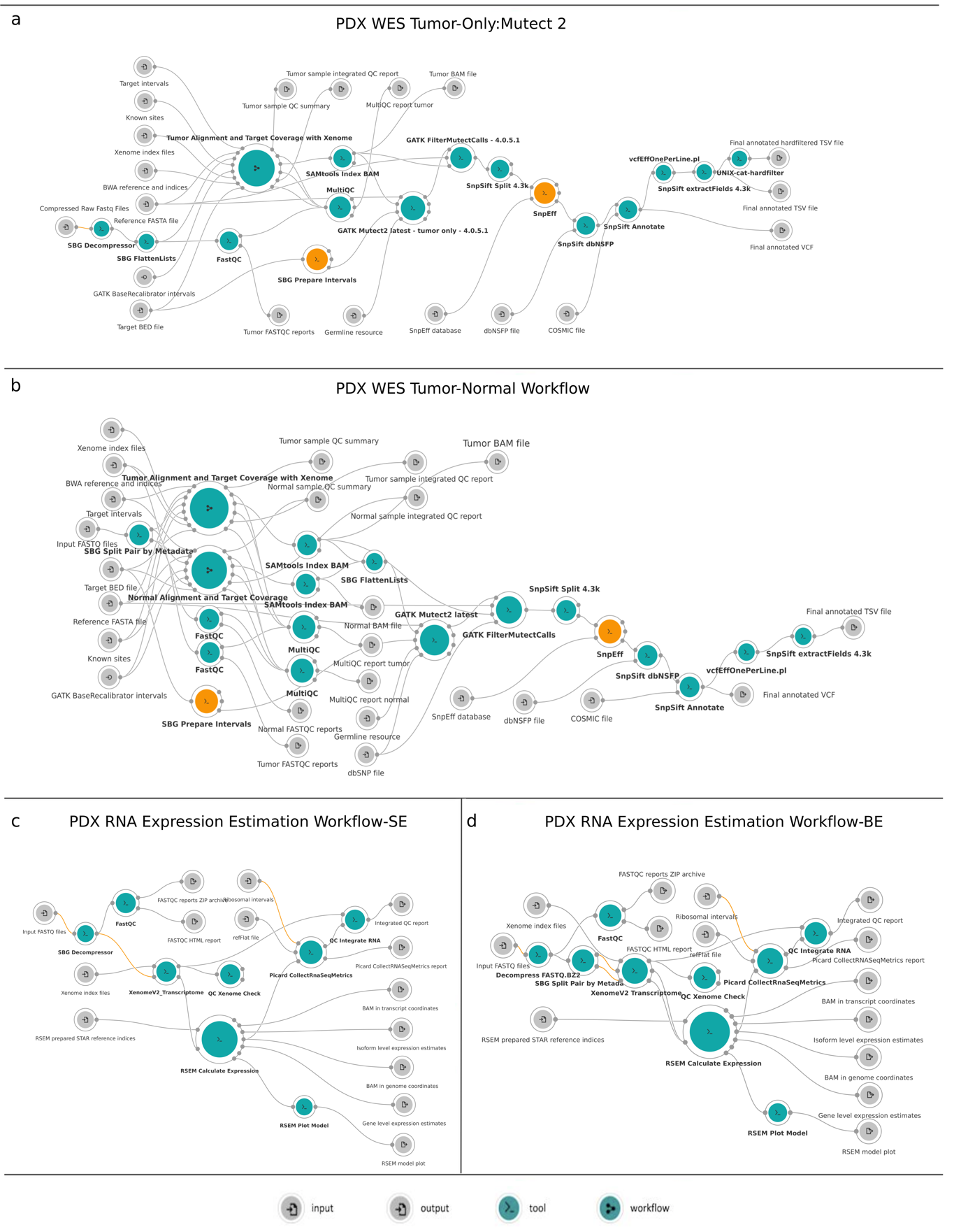
Supplementary Figure 13: Wiring diagrams for submitted whole exome workflows submitted by the PDTCs. Wiring diagrams include nodes and connections. Nodes depict inputs - A picture containing object, clock

Description automatically generated, outputs - A picture containing object, clock

Description automatically generated, tools - A picture containing object, clock

Description automatically generated, and workflows - . Connections between nodes depict that input to a node is from the output of another node. Orange nodes - identify a tool or a workflow with an available update. a) Workflow submitted by PDTC 4, b) Workflow submitted by PDTC 5

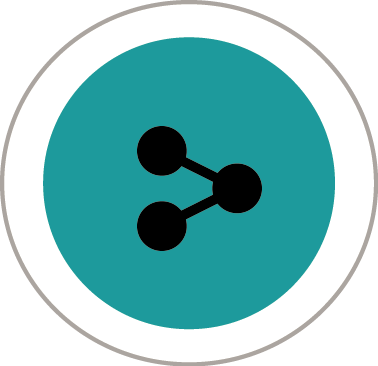
# Supplementary Materials 6: Workflows selected to process PDX whole exome and RNA-seq data



Supplementary Figure 14: Workflow selected to process PDX Data. The panel figure depicts whole exome and RNA-Seq workflows wiring diagram selected by the PDCCC to process PDX data generated for this study. Wiring diagrams include nodes and connections. Nodes depict inputs - A picture containing object, clock

Description automatically generated, outputs - A picture containing object, clock

Description automatically generated, tools - A picture containing object, clock

Description automatically generated, and workflows - . Connections between nodes depict that input to a node is from the output of another node. Orange nodes - identify a tool or a workflow with an available update. a) PDX whole exome tumor only workflow. The PDCCC selected the whole exome workflow after evaluating 5 workflows submitted by the PDTCs. b) PDX whole exome tumor-normal workflow. The PDCCC generated by the whole exome tumor-normal workflow by adapting the selected workflow described in a). c) PDX RNA Expression Estimation Workflow – Single Execution (SE). The PDCCC selected the RNA-seq workflow from established best practices. d) PDX RNA Expression Estimation Workflow – Batch Execution (BE). The PDCCC adapted the PDX RNA Expression Estimation Workflow shown in c) to ­accept batch inputs.

# Supplementary Materials 7: Gene expression analysis using RNA-Seq workflow

## Description

This RSEM workflow (RSEM 1.2.31) for quantifying gene expression uses the STAR aligner and is optimized to work with FASTQ input files.

To process multiple samples, please consider running batch tasks with this workflow and aggregating the results using Prepare Multisample Data [workflow](https://cgc.sbgenomics.com/u/pdxnet/pdxnet-jax-rna-seq-workflow/apps/#pdxnet/pdxnet-jax-rna-seq-workflow/prepare-multisample-data).

## Essential Requirements

The following metadata fields are essential and should be assigned to input FASTQ files:

1. Sample ID: Any string. The identifier should be identical for both paired-end FASTQ files.
2. Paired-end: 1 or 2

This workflow will process both uncompressed and compressed FASTQ files (FASTQ.GZ, FASTQ.BZ2) and has been designed for paired-end data. By default, the workflow assumes unstranded data (Forward probability input parameter set to 0.5). Please adjust the value of this parameter (0.0 or 1.0) based on the library prep of your data.

#### The following output files will be generated:

* Gene level expression estimates
* Isoform level expression estimates
* RSEM model plot
* BAM in transcript coordinates
* BAM in genome coordinates
* FASTQC reports ZIP archive
* FASTQC HTML report
* Integrated QC report
* Picard CollectRNASeqMetrics report

## Reference Files and Workflow Details

Required reference input files:

1. Xenome is used to classify reads as human or mouse. Xenome indices were built on hg38 and pseudoNOD transcriptome (based on SNP incorporation into mm10 genome from Sanger [ftp://ftp-mouse.sanger.ac.uk/REL-1505-SNPs\_Indels/]). The default value of k=25 is used during the indices preparation. Default file input: Xenome\_transcriptome\_indices\_GRCh38\_91\_pseudoNod\_mm10.tar.gz
2. STAR indices archive prepared by RSEM Prepare Reference (v.1.2.31). The default input file (GRCh38.91.chr\_patch\_hapl\_scaf\_rsem-1.2.31.star-index-archive.tar) was built using a GRCh38 FASTA file (primary assembly, EBV, alt contigs, decoys, and HLA contigs) and an annotation GRCh38 GTF file from Ensembl (release 91) (ftp://ftp.ensembl.org/pub/release-91/gtf/homo\_sapiens/Homo\_sapiens.GRCh38.91.chr\_patch\_hapl\_scaff.gtf.gz)
3. refFlat file (hg38) used by Picard CollectRnaSeqMetrics tool. Downloaded from: http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/refFlat.txt.gz Default file input: refFlat.ucsc\_hg38.txt
4. Ribosomal intervals (hg38) used by Picard CollectRnaSeqMetrics tool. Default file input: rRNA\_hg38.interval

## Workflow Steps and Notable Parameters

#### Step 1: Optional input preprocessing

If FASTQ.BZ2 files are provided as inputs, the files will be decompressed before further analysis (as Xenome will only accept FASTQ and FASTQ.GZ files). FASTQ.GZ and uncompressed FASTQ input files will be passed on to other tools in the workflow.

#### Step 2: FASTQC analysis

Quality of the input FASTQ files is checked with FASTQC.

#### Step 3: Xenome classification of reads

FASTQ pairs are split (SBG Split Pair by Metadata) based on the appropriate paired\_end metadata field values and classified by Xenome as mouse or human. QC Xenome Check tool checks that a sufficient number of reads have been classified as human. By default, minimum number of human reads required is set to 1000000, however this parameter is exposed (Minimum number of human-specific reads) and can be adjusted by the user. *Note*: If the Minimum number of human-specific reads cutoff is not met, the tasks will fail. If you expect <1000000 human reads in your input data, or are testing the workflow with subsetted files, please adjust this parameter accordingly.

#### Step 4: RSEM expression estimation

Expression is estimated using RSEM Calculate Expression tool (RSEM 1.2.31), with STAR as the aligner. Please ensure that the reference indices archive supplied to the tool has been prepared accordingly. RSEM Plot Model tool is used to generate RSEM plots.

Please note that by default, the workflow is setup to process unstranded data (Forward probability input parameter set to 0.5). Please make sure to adjust the value of this parameter (0.0, 0.5 or 1.0) based on the library-prep used.

#### Step 5: Additional QC

Additional QC reports are collected from Picard CollectRnaSeqMetrics tool and Xenome.

# Supplementary Materials 8: Comparisons of xenograft sequence data across PDTCs

Supplementary Table 4. Total genomic region (in megabases [Mb]) covered by each capture

|  |  |
| --- | --- |
| **Center** | **Capture Array Genome Coverage (Mb)** |
| HCI-BCM | 115.23 |
| MDACC | 63.50 |
| WUSTL | 39.11 |
| WIST | 45.31 |
| HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC. | |

Supplementary Table 5. Mean number of calls after restricting to intersected array and applying allele frequency-based threshold

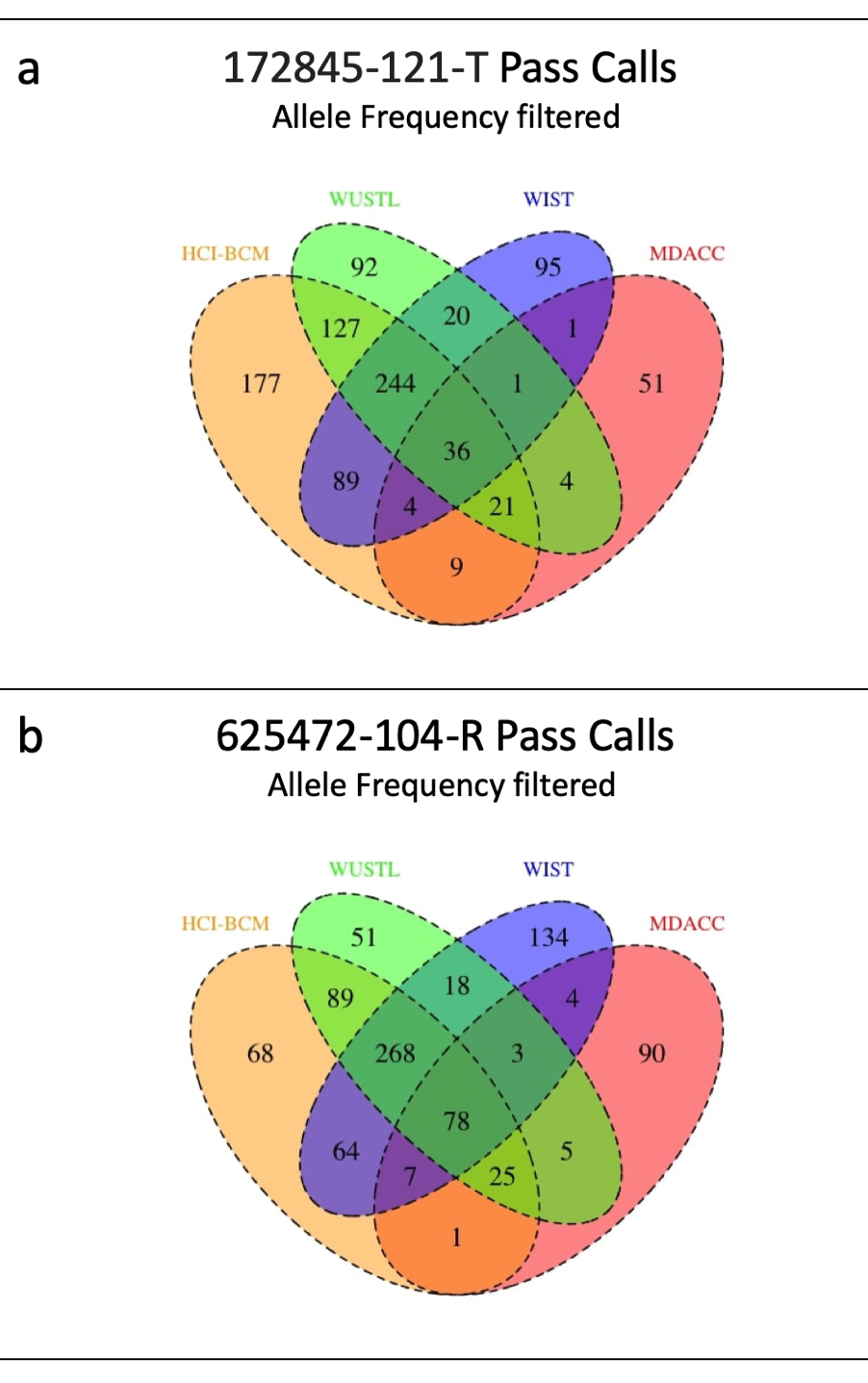
|  |  |  |  |
| --- | --- | --- | --- |
| **Center** | **Average Number Variant Calls Center Specific Arrays** | **Average Number Variant Calls Intersected Arrays** | **Average Number Variant Calls AF Filtered Intersected Arrays** |
| **HCI-BCM** | 77,885.33 | 5,333.00 | 3,286.00 |
| **MDACC** | 15,614.33 | 6,253.33 | 1,340.67 |
| **WUSTL** | 14,569.67 | 8,103.67 | 3,274.00 |
| **WIST** | 11,618.00 | 5,406.00 | 3,114.67 |
| HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC. | | | |

Supplementary Table 6. RNA-Seq library-prep, sequencing type and depth for each sample

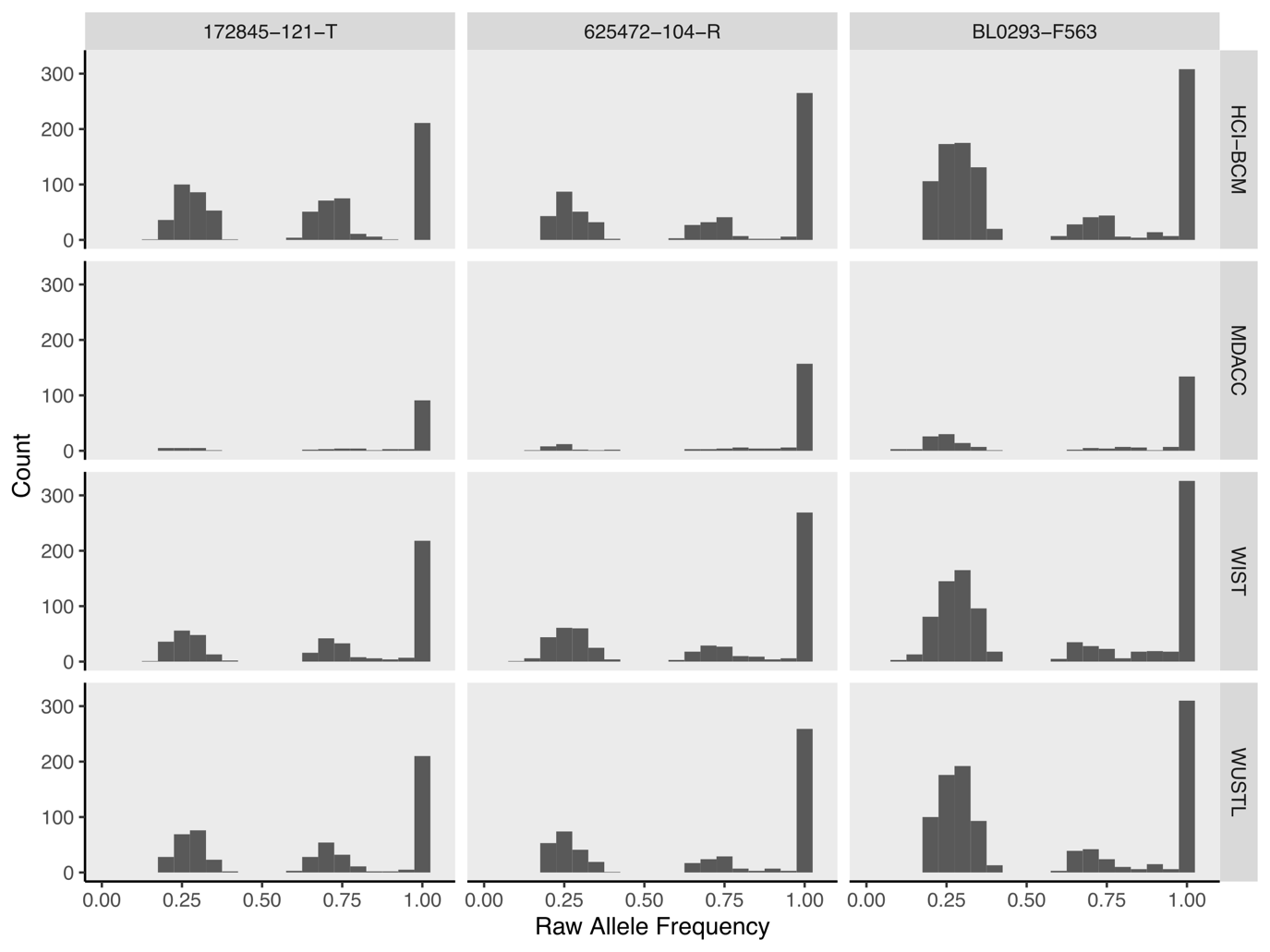
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Center** | **CaseID** | **PrepKit** | **Sequencing** | **total-reads** |
| **HCI-BCM** | BL0293-F563 | IlluminaTruSeqStandardRNAkitwithRiboZeroGold | Illumina HiSeq SE | 88,858,611 |
|  | 172845-121-T | IlluminaTruSeqStandardRNAkitwithRiboZeroGold | Illumina HiSeq SE | 83,090,055 |
|  | 625472-104-R | IlluminaTruSeqStandardRNAkitwithRiboZeroGold | Illumina HiSeq SE | 99,052,483 |
| **MDACC** | BL0293-F563 | Kapa stranded mRNA-Seq Kit | Illumina HiSeq PE | 86,514,781 |
|  | 172845-121-T | Kapa stranded mRNA-Seq Kit | Illumina HiSeq PE | 71,158,865 |
|  | 625472-104-R | Kapa stranded mRNA-Seq Kit | Illumina HiSeq PE | 67,009,735 |
| **WUSTL** | BL0293-F563 | TruSeq Stranded Total RNA | NovaSeq S4 PE | 44,188,536 |
|  | 172845-121-T | TruSeq Stranded Total RNA | NovaSeq S4 PE | 56,610,898 |
|  | 625472-104-R | TruSeq Stranded Total RNA | NovaSeq S4 PE | 43,067,125 |
| **WIST** | BL0293-F563 | Lexogen mRNA | Illumina PE | 51,503,760 |
|  | 172845-121-T | Lexogen mRNA | Illumina PE | 53,648,665 |
|  | 625472-104-R | Lexogen mRNA | Illumina PE | 52,142,953 |
| HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC. | | | | | |

Supplementary Table 7. Sequencing depth and Mean Target coverage per sample across each center

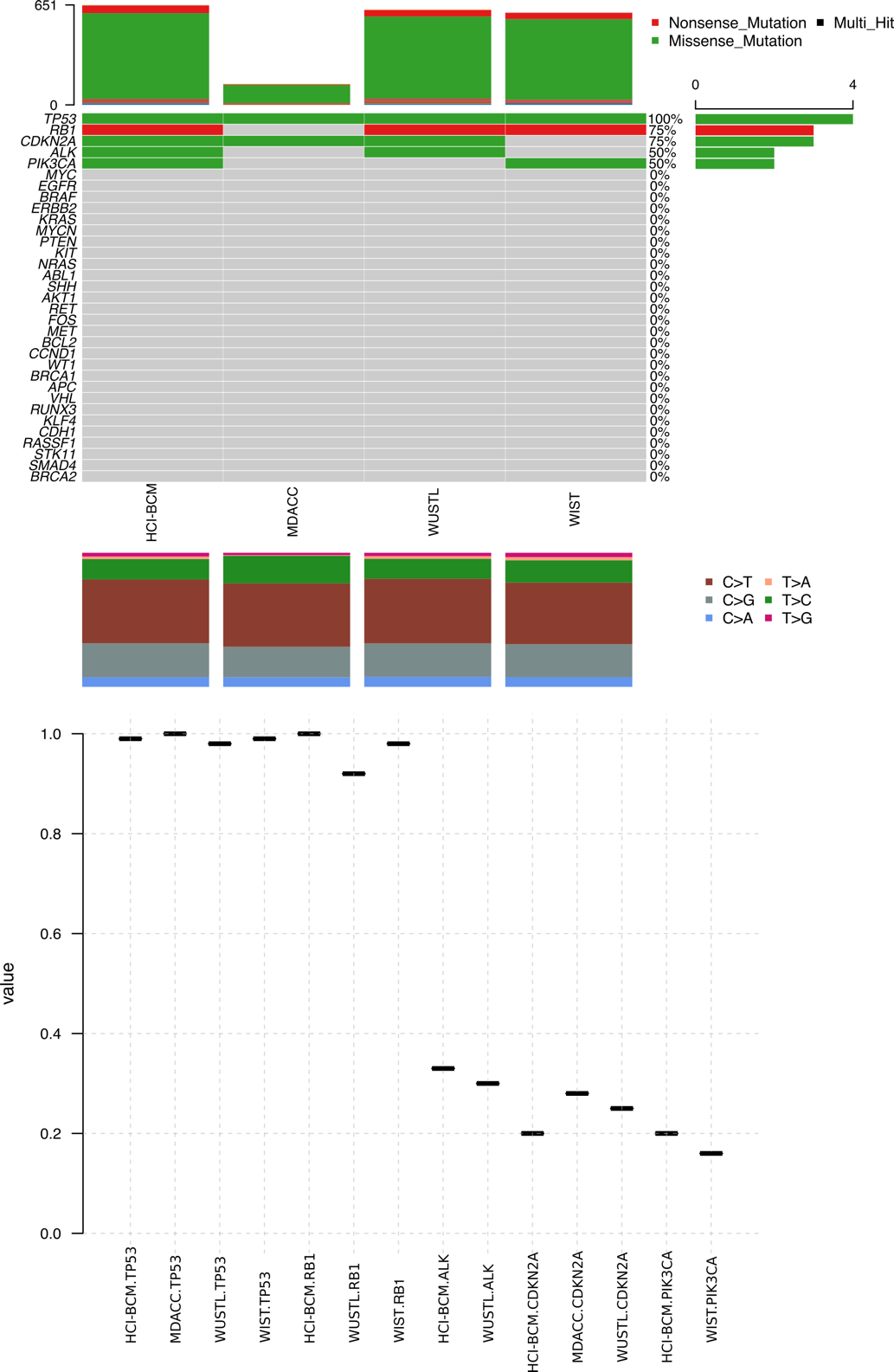
|  |  |  |  |
| --- | --- | --- | --- |
| **Center** | **CaseID** | **Total Reads** | **Mean Target  Coverage** |
| **HCI-BCM** | BL0293-F563 | 84306497 | 182.489987 |
|  | 172845-121-T | 84500785 | 171.117572 |
|  | 625472-104-R | 69122847 | 144.39062 |
| **MDACC** | BL0293-F563 | 24840423 | 34.378847 |
|  | 172845-121-T | 22722611 | 25.75321 |
|  | 625472-104-R | 23350341 | 28.91726 |
| **WUSTL** | BL0293-F563 | 63360889 | 225.221598 |
| 172845-121-T | 47171801 | 163.915966 |
|  | 625472-104-R | 36508439 | 131.624613 |
| **WIST** | BL0293-F563 | 249671426 | 192.639997 |
|  | 172845-121-T | 114733973 | 122.825594 |
|  | 625472-104-R | 129198866 | 133.174483 |
| HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC. | | | |



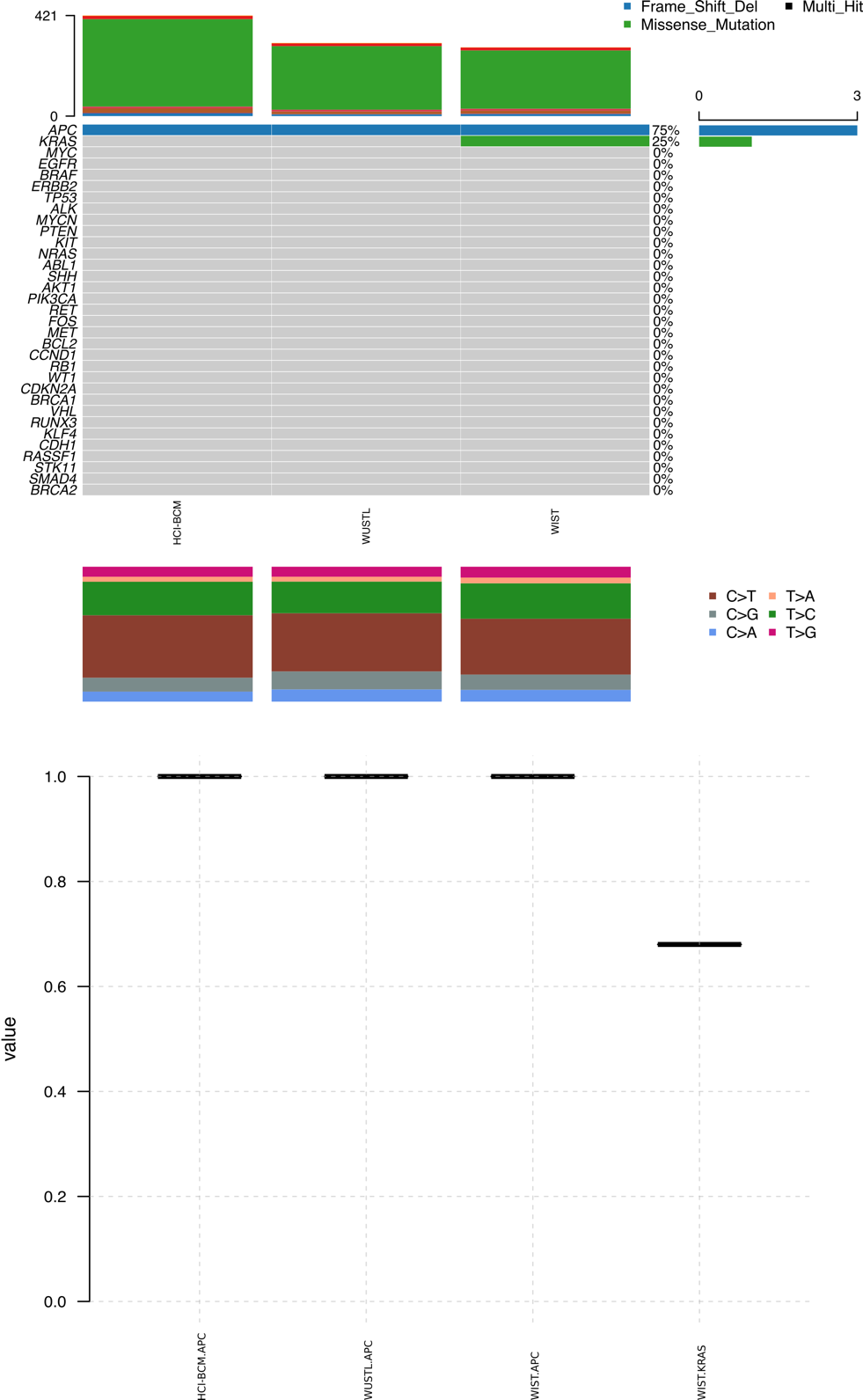
Supplementary Figure 15. A Venn diagram showing the overlap in high-quality variant calls among centers by model using intersected array and removing lower allele frequency (AF <5% calls. a) Pass variant calls for 172845-121-T, b) Pass variant calls for 625472-104-R. (HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC).



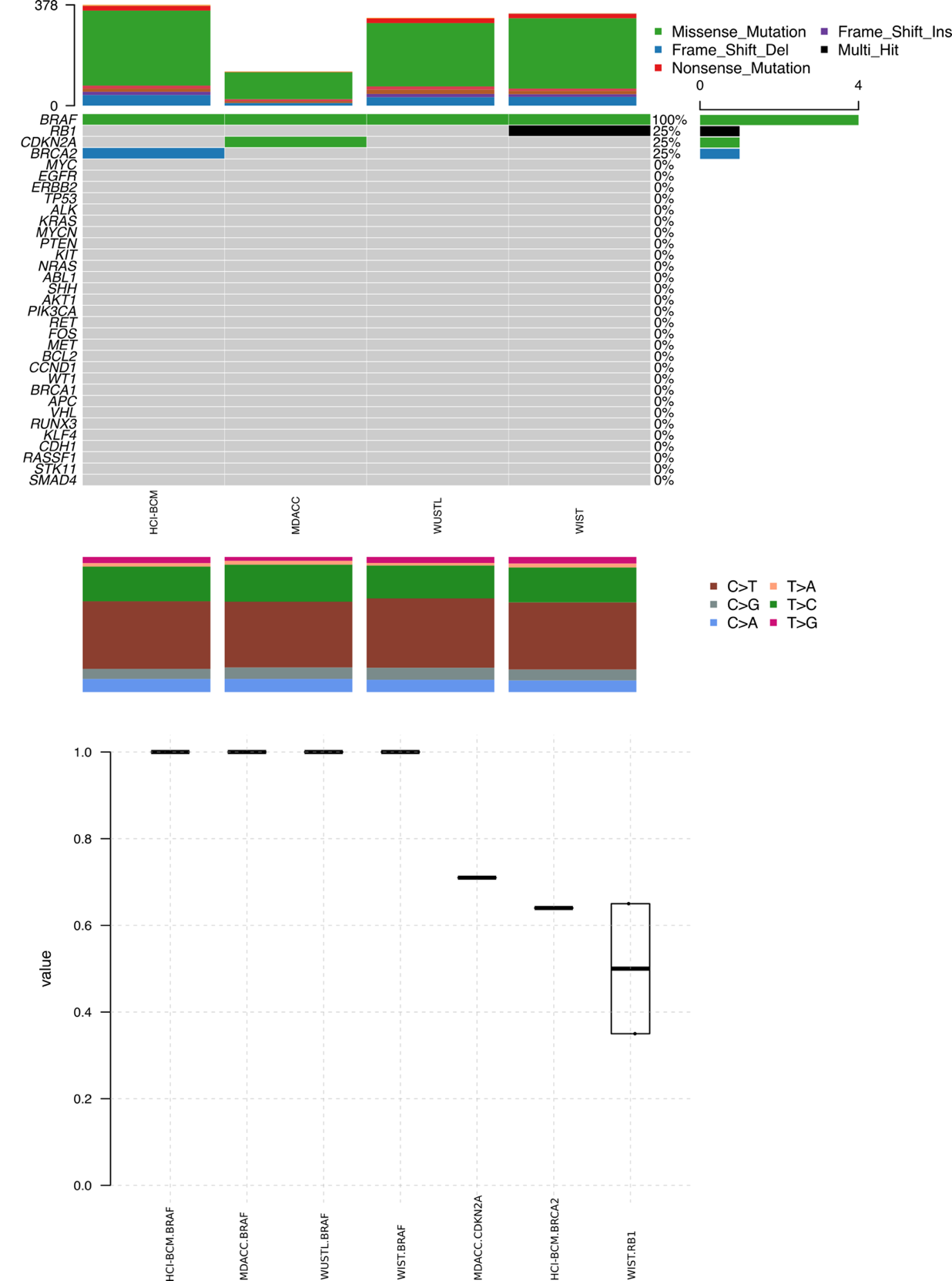
Supplementary Figure 16. Raw allele frequency for high-quality variants filtered to intersected array at allele frequency ≥ 0.2. (HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC.)

****

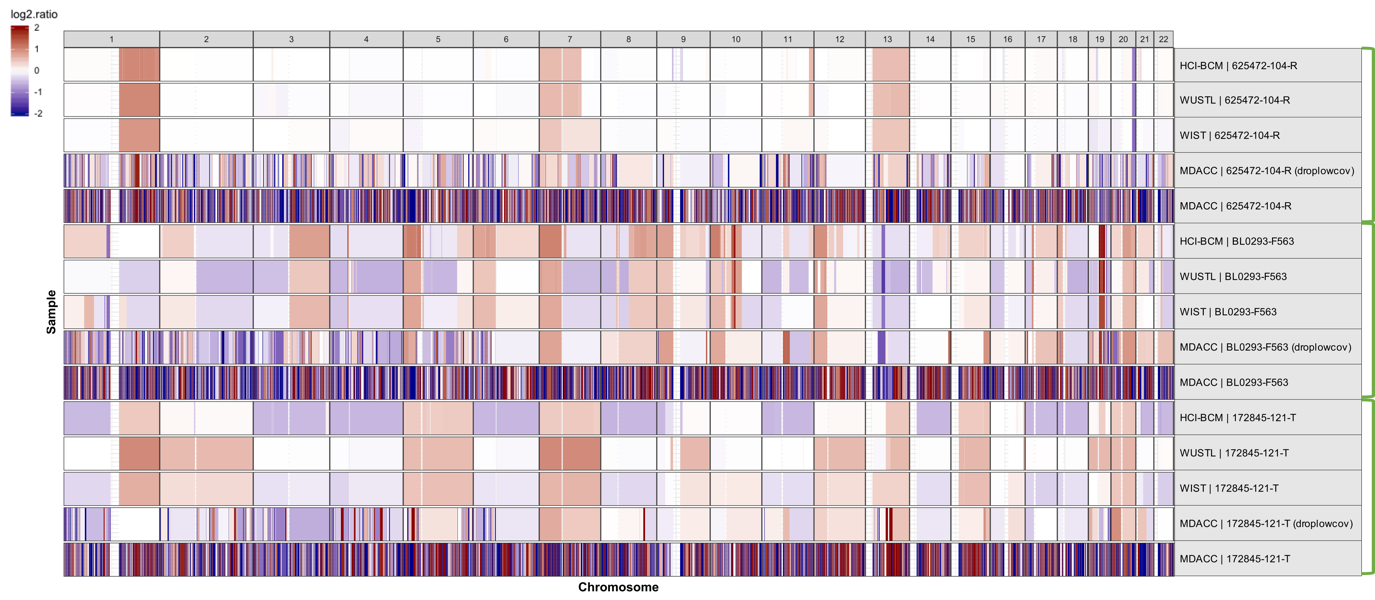
Supplementary Figure 17. Comparison of mutations in cancer-related genes for JAX-BL029. (left) Mutations called in cancer genes at each center. Colors visually indicate the fraction of centers observing each mutation. (right). Allele frequencies observed for called cancer gene mutations at each center. (HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)



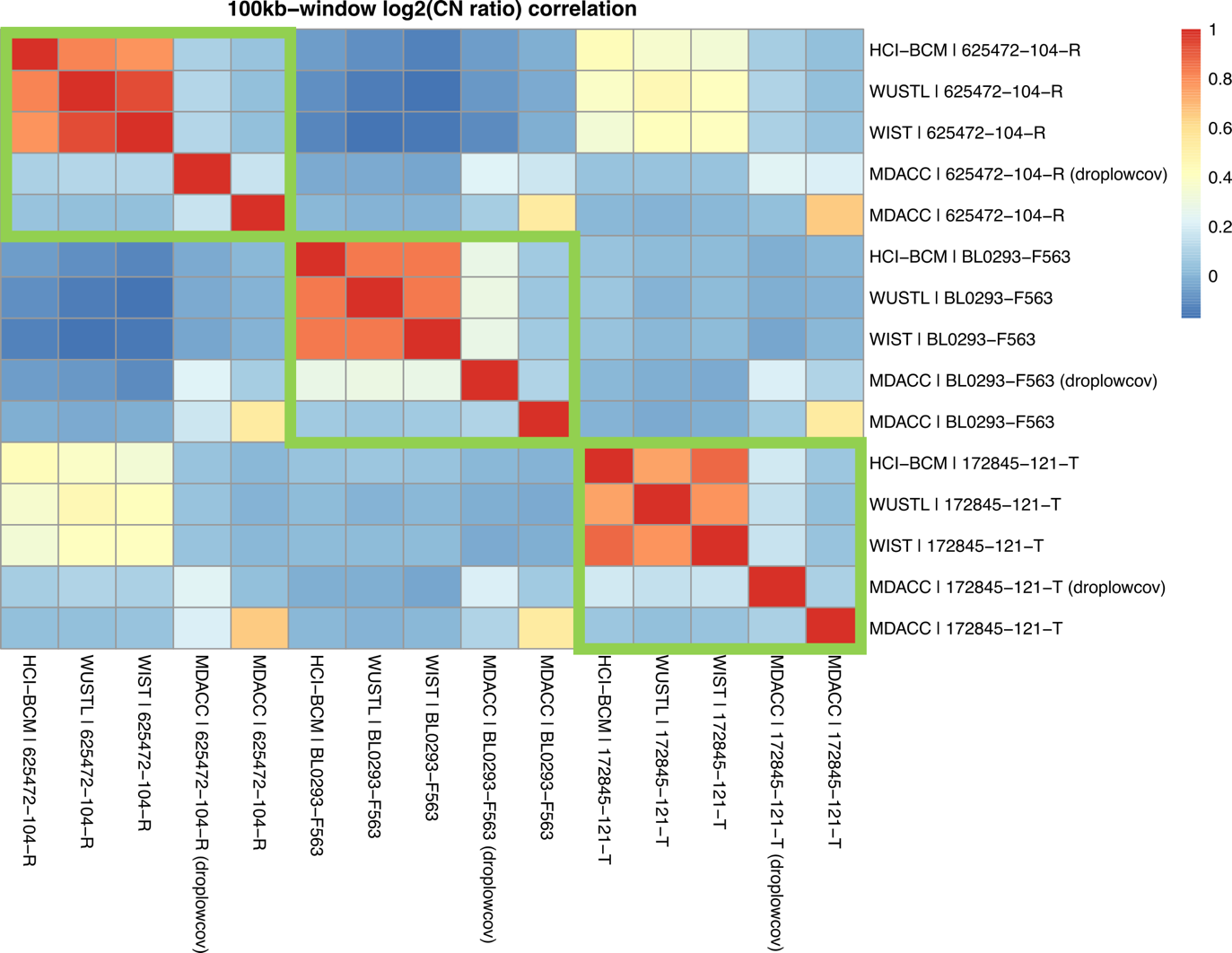
Supplementary Figure 18. Comparison of mutations in cancer-related genes for PDMR-172845. (left) Mutations called in cancer genes at each center. Colors visually indicate the fraction of centers observing each mutation. (right). Allele frequencies observed for called cancer gene mutations at each center. (HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)



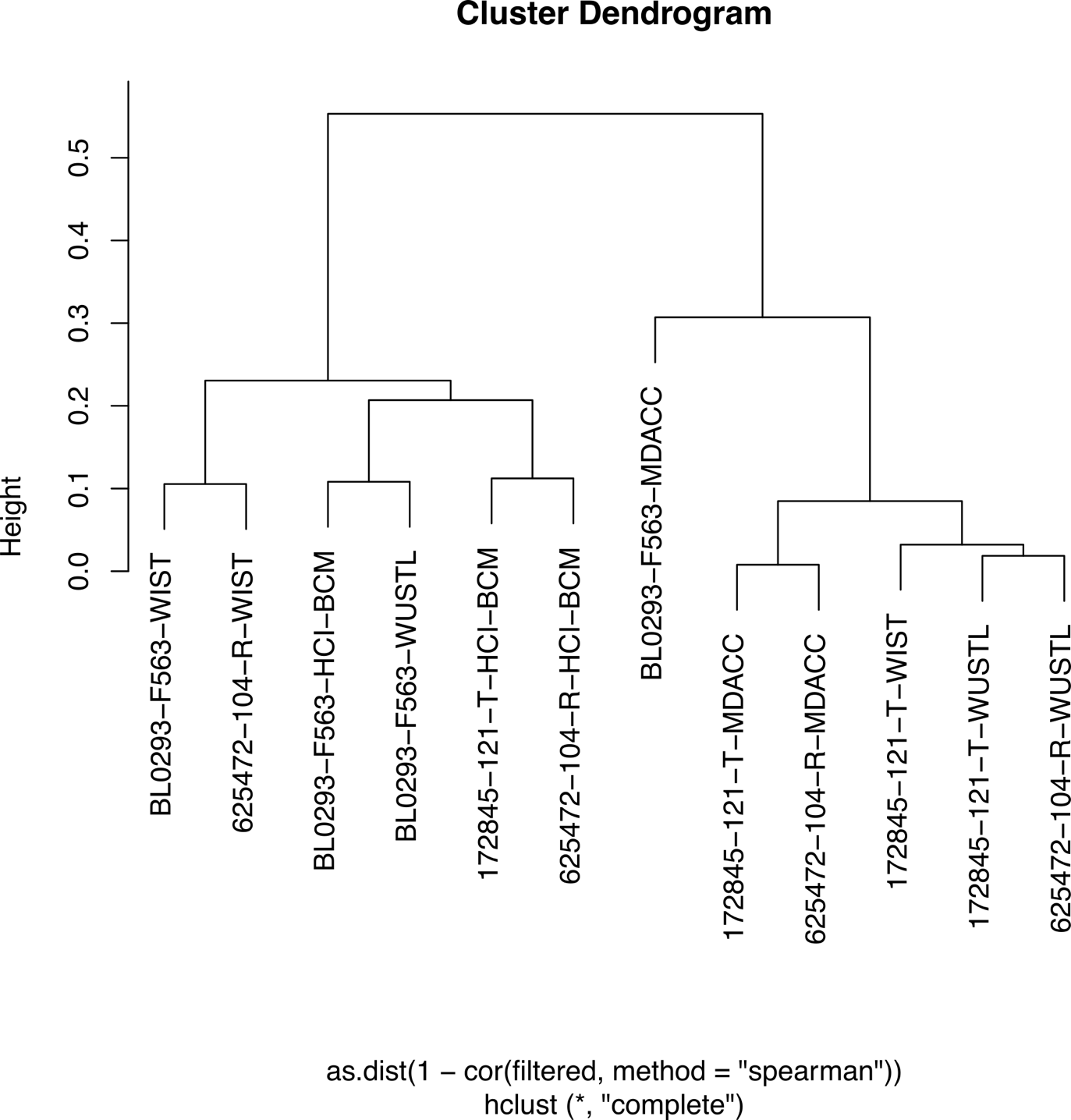
Supplementary Figure 19. Comparison of mutations in cancer-related genes for PDMR-625473. (top) Mutations called in cancer genes at each center. Colors visually indicate the fraction of centers observing each mutation. (bottom). Allele frequencies observed for called cancer gene mutations at each center. (PDMR - NCI Patient-Derived Models Repository, HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)

****

**Supplementary Figure 20.** A genome-wide view of median-centered copy number alteration (CNA) segments. The samples are labelled as “Center | CaseID” as shown in Supplementary Table 7. “droplowcov” indicates that the “drop low coverage” option in CNVKIT was used. The green brackets group the samples from the same model (CaseID) generated by the different centers.(HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)



Supplementary Figure 21. A Pearson correlation coefficient based heatmap for the segmented log2 copy number ratio binned across 100kb-windows for each pair of samples. The samples are labelled as “Center | CaseID” as shown in Supplementary Table 7. “droplowcov” indicates that the “drop low coverage” option in CNVKIT was used. The green boxes group the samples from the same model (CaseID) generated by the different centers. (HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)



Supplementary Figure 22. Dendrogram of TMM normalized count per million values. (PDMR - NCI Patient-Derived Models Repository, HCI-BCM - Huntsman Cancer Institute/Baylor College of Medicine, MDACC - MD Anderson Cancer Center, WUSTL - Washington University-St. Louis, and WIST - The Wistar Institute/University of Pennsylvania/MDACC)