QUALITY CONTROL WORKFLOWS DEVELOPED FOR THE NCI PATIENT-DERIVED MODELS REPOSITORY USING LOW PASS WHOLE GENOME SEQUENCING AND WHOLE EXOME SEQUENCING

Ting-Chia Chang[1], Li Chen[1], Biswajit Das[1], Yvonne A. Evrard[1], Chris A. Karlovich[1], Tomas Vilimas[1], Alyssa Chapman[1], Nikitha Nair[1], Luis Romero[1], Anna J. Lee Fong[1], Amanda Peach[1], Brandie Fullmer[1], Lindsay Dutko[1], Kelly Benauer[1], Gloryvee Rivera[1], Erin Cantu[1], Shahanawaz Jiwani[1], Nastaran Neishaboori[1], Tomas Forbes[1], Corinne Camalier[1], Luke Stockwin[1], Michael Mullendore[1], Michelle A. Eugeni[2], Dianne Newton[1], Melinda G. Hollingshead[2], P. Mickey Williams[1], James H. Doroshow[3]

[1]Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, [2]National Cancer Institute at Frederick, Biological Testing Branch, Developmental Therapeutics Program, Frederick, MD and [3]National Cancer Institute, Division of Cancer Treatment and Diagnosis, Bethesda, MD

NIH NATIONAL CANCER INSTITUTE

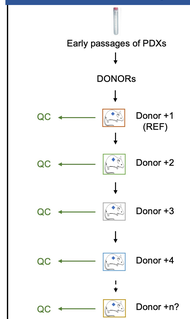
NATIONAL CANCER INSTITUTE PATIENT-DERIVED MODELS REPOSITORY

Abstract Number: 1913

Background

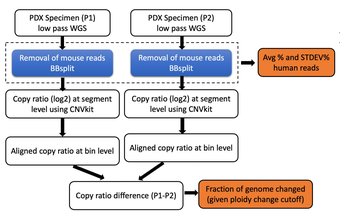
The National Cancer Institute's Patient-Derived Models Repository (NCI PDMR; https://pdmr.cancer.gov/) is developing a variety of patient-derived xenograft (PDX) models for pre-clinical drug studies. All NCI PDMR models undergo quality control (QC) processes. Two unique QC workflows were established a) to assess genomic stability across PDX model passages and b) to confirm the suitability of PDX-derived cancer associated fibroblasts (CAFs) as germline surrogates when blood was not available. Multiple bioinformatics QC assessments have been developed to measure the genomic fidelity in these PDX models using low-pass whole genome sequencing (LP-WGS) and in CAFs using whole exome sequencing (WES).

QC Workflow 1. To assess genomic stability across PDX model passages for pre-clinical drug study



- 38 PDX models derived from 19 rare cancers were established for preclinical drug study with multiple novel therapeutic combinations.
- The donor materials were prepared from the mixture of early passages of PDX tumors and serially passaged for the drug studies and QC assessment.
- To measure the model fidelity, each passage underwent several quality control assessments:
  - Pathology assessment
  - Human:mouse DNA content assessment
  - Low pass whole genome sequencing to determine the average fraction of genome changed compared to the original donor material

Workflow of CNA assessment across the whole genome by LP-WGS



➢ **Mouse read removal and % human reads estimation**

- BBSplit[1]: *k*-mer based method to map reads to human and mouse references simultaneously

➢ **Copy number detection algorithm (CNVkit[2]) based on read depth**

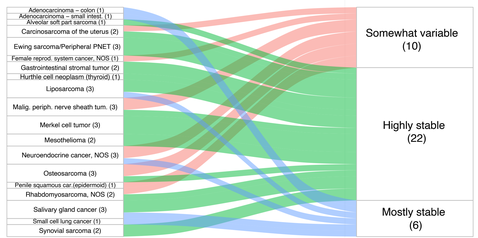
- Bin size (1Mbp) is determined by estimating the read counts/depths, required to get an average of 200-300 reads per bin when sequencing coverage is as low as 0.3x
- The fraction of genome changed between two samples is defined as the fraction of altered genomic regions with relative copy number changes >0.4(log2 ratio, 1.32 ploidy)
- The fraction of genome changed is calculated by comparing the copy numbers of each passage sample to the original donor sample (REF)

Average % human reads estimation of 38 PDX models

Passage	# of Models	Range of Actual Passage	Avg. # of PDX samples/model (Range)	Avg. % human reads	Avg. STDEV% human reads
Donor +1	38	1.5-3	4.68(2-8)	82.70	4.62
Donor +2	17	2-4	4.29(3-7)	76.93	4.73
Donor +3	25	3.5-4.5	4.72(2-10)	82.68	3.43
Donor +4	5	4-6	4.00(3-5)	84.03	2.51
Donor +5	9	5.5-6.5	4.67(2-9)	79.58	3.91
>= Donor +6	7	6.5-10	5.08(3-9)	81.17	5.36

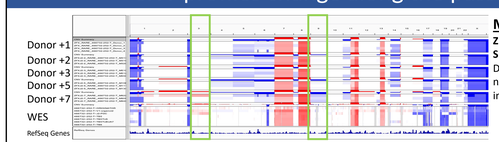
- 502 PDX samples derived from 38 models have been sequenced by LP-WGS
- The percentages of human contents were very consistent, approximately 76-84% on inter-passages

Different intra-passage heterogeneity in REF samples

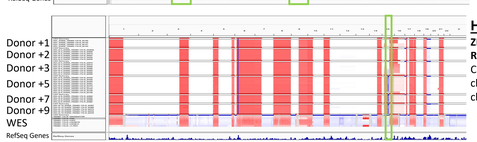


- Alluvium plot classified 38 rare tumor models derived from 19 histologies into 3 categories based on average fraction of genome changed within REF passage (> 5%: somewhat variable; 1.5%-5%: mostly stable; <1.5% highly stable models)

CNV profiles change along the passages



Mostly stable model:
ZFX (6466732-252-R): 49 yo Male, SMALL INTESTINE ADENO
Donor+7 showed inconsistent copy number profiles along the passages in chromosomes 3, 9

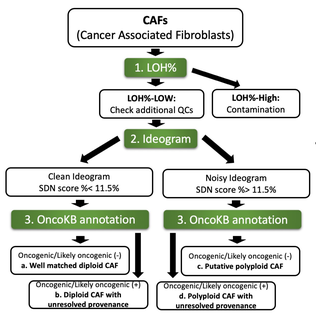


Highly stable model:
ZDL (755483-174-R): 48 yo Male, RHABDOMYOSARCOMA
Copy number profiles slightly changed along the passages in chromosome 15q

- **Most models showed no significant evolution in the limited number of passages sequenced**
- **Some models' copy number profiles were altered slightly when the passage numbers increased**
- The potential reasons for the inconsistent copy number profiles along the passages:
  - Possible sampling bias as the different lineages were established by sampling from different regions of the original tumor, potentially capturing different elements of heterogeneity
  - The low resolution in LP-WGS

QC Workflow 2. To confirm the suitability of PDX-derived CAFs as germline surrogates

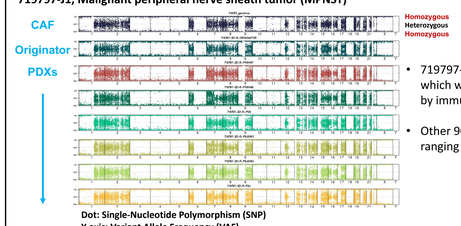
Workflow of CAF QC assessment



- In PDX research, the matched normal sample for somatic variant and copy number calling algorithms is extremely crucial. However, blood specimens are not always available for all PDX models. 91 PDX-derived CAFs, from 32 histologies, were used as germline surrogate material when blood was not available.
- To evaluate purity of CAFs, three QC steps were constructed based on WES data:
  Step 1: Percentage of genomic loss of heterozygosity (LOH) based on a set of ~800k heterozygous Single-Nucleotide Polymorphism (SNPs) from a population-level genomic database (gnomAD)[3]
  Step 2: Ideogram: plot of SNPs variant allele frequencies; SNP Distribution Noise (SDN)
  Step 3: Oncogenic/likely oncogenic variants annotated by OncoKB[4]

Step 1: LOH% Estimate

719797-J1, Malignant peripheral nerve sheath tumor (MPNST)



- 719797-J1 indicated high LOH% (42%), which was confirmed as tumor cell line by immunohistochemistry (IHC).
- Other 90 CAFs showed little variance, ranging from 0.6% - 2.3% LOH

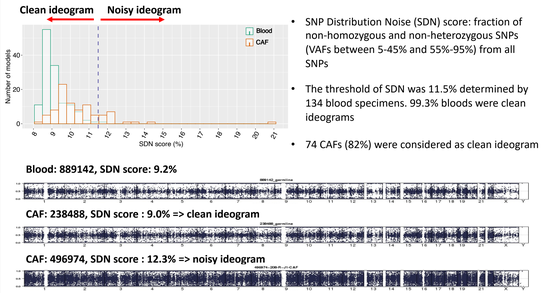
Summary

- We developed standard QC workflows to evaluate genomic stability of PDX models during pre-clinical study for model validation and qualify CAFs as germline surrogates.

Reference

1. Bishnell et al. BBMerge – Accurate paired shotgun read merging via overlap (2017), 12(10): e0185056.
2. Talevich et al. CNVkit: Genome-wide copy number detection and visualization from targeted sequencing (2014), 12(4):e1004873
3. Karczewski et al. The mutational constraint spectrum quantified from variation in 141,456 humans (2020), 581, 434-443
4. Chakravarty et al. OncoKB: A Precision Oncology Knowledge Base (2017), 10.1200/PO.17.00011

Step 2: SNPs Distribution Ideogram



- SNP Distribution Noise (SDN) score: fraction of non-homozygous and non-heterozygous SNPs (VAFs between 5-45% and 55%-95%) from all SNPs
- The threshold of SDN was 11.5% determined by 134 blood specimens. 99.3% bloods were clean ideograms
- 74 CAFs (82%) were considered as clean ideogram

Step 3: OncoKB Annotation and Categorization

- OncoKB annotation was used for identifying the oncogenic and likely oncogenic variants in CAFs

Ideogram/OncoKB	Clean ideogram	Noisy ideogram
Oncogenic/likely oncogenic (-)	a. Well matched diploid CAF- 75.6% (68/90)	c. Putative polyploid CAF- 10% (9/90)
Oncogenic/likely oncogenic (+)	b. Diploid CAF with unresolved provenance- 6.6% (6/90)	d. Polyploid CAF with unresolved provenance- 7.7% (7/90)

- Majority of CAFs were well matched diploid
- 6.6% diploid CAF with unresolved provenance had ≥ 1 germline oncogenic variants confirmed with PDX samples
  - **CAFs in category a and b are suitable as germline surrogates**
- 10% of CAFs showed putative polyploidy on SNP ideograms with no oncogenic variants
  - **This category is suitable for somatic variant calling**
- 7.7% of CAFs had polyploidy and oncogenic variants present
  - **Further derived PDX samples to confirm or IHC evaluation are needed**
  - **Perform WES with any leftover originator specimen for CAF unresolved provenance to resolve some of these cases**

Frederick National Laboratory for Cancer Research

sponsored by the National Cancer Institute

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.