# INNOVATIONS IN LARGE SCALE LIQUID BIOPSY SEQUENCING

BROAD INSTITUTE
GENOMICS

CARRIE CIBULSKIS, MATTHEW DEFELICE[1], MARK FLEHARTY[1], BRENDAN BLUMENSTIEL[1], JUSTIN ABREU, MADELEINE DURAN[1], SHELI DOOKRAN[1], GADDY GETZ[1], VIKTOR ADALSTEINSSON[1], STACEY GABRIEL[1], NIALL LENNON[1]

[1]Broad Institute of MIT and Harvard

## Introduction

**Assay Characteristics:**
The Broad Institute Genomics Platform aims to support liquid biopsy clinical research activities by offering a comprehensive data generation and analytical platform that is:

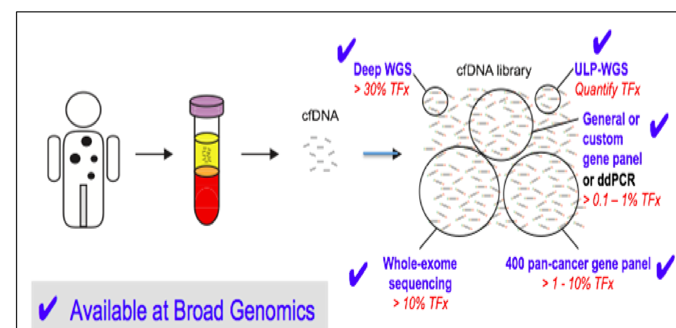
- High quality
- Flexible and scalable
- Fast enough for clinical use (< 14 days)
- Permits multiple assays to be run on the same input material
- Offers return of raw and interpreted data

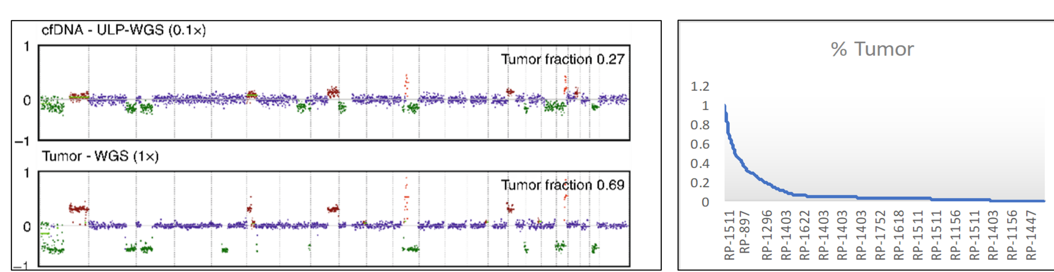
**Target Research and Clinical Data Uses:**

- Sensitive genomic profiling of cfDNA samples (exomes or large gene panels) for discovery of treatment resistance mechanisms
- Longitudinal patient disease monitoring
- Detection of minimal residual disease in patients post treatment
- Early cancer detection

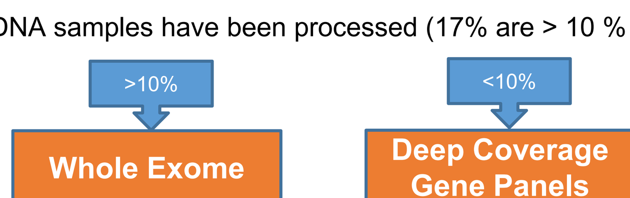
## Liquid Biopsy Approach:

Samples are collected directly from patients/hospitals (blood) or biobanks (plasma, cfDNA). Libraries are constructed and sequenced at ultra low depth whole genome (0.1x) and then used to measure the tumor content of specimens. Samples > 10% tumor fraction can be assayed with whole exome analysis and < 10% with targeted gene panels.

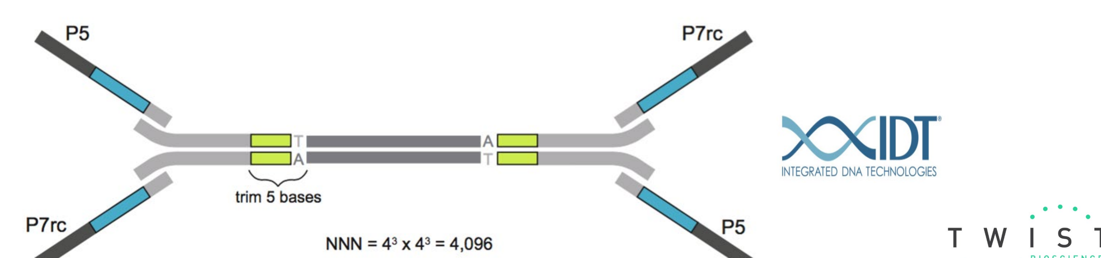



6500 cfDNA samples have been processed (17% are > 10 % Tumor Fraction)



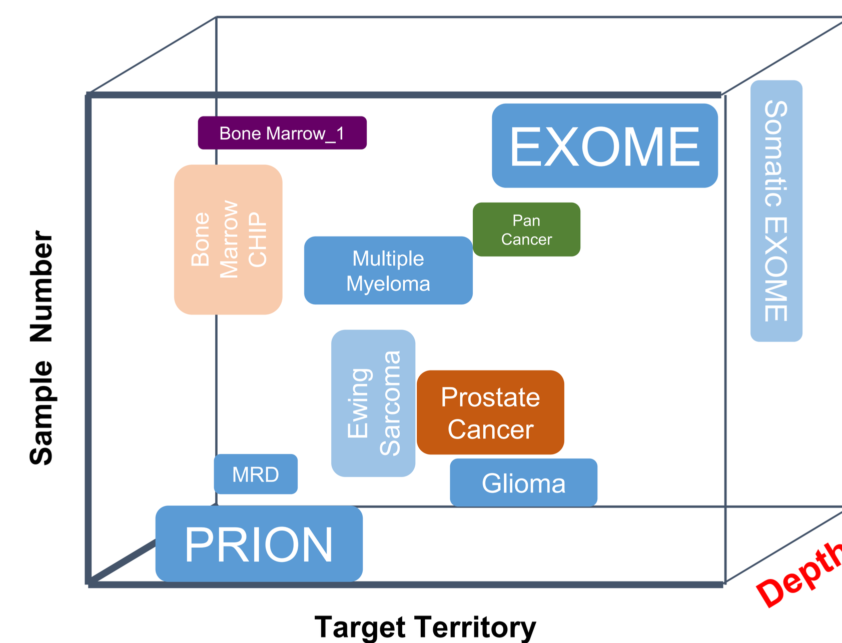

## Duplex UMI Adapter Library Construction




Sensitive and accurate mutation discovery requires ultra deep sequencing and error correction. Our platform utilizes IDT's "Stubby-Y" duplex adapters incorporate in-line double stranded UMIs allowing for assembly of duplex consensus reads from both top and bottom strands of original molecule. Enrichment for exome or targeted gene panel regions is performed using TWIST biosciences reagents.

## Custom & Pan-Cancer Targeted Panels

Our process is designed to easily generate customized (project specific) targeted panels for sequencing the most high priority genomic loci specific to research goals. A range of project design options can be accommodated including variable amounts of samples, target territory and sensitivity goals (depth of sequencing).

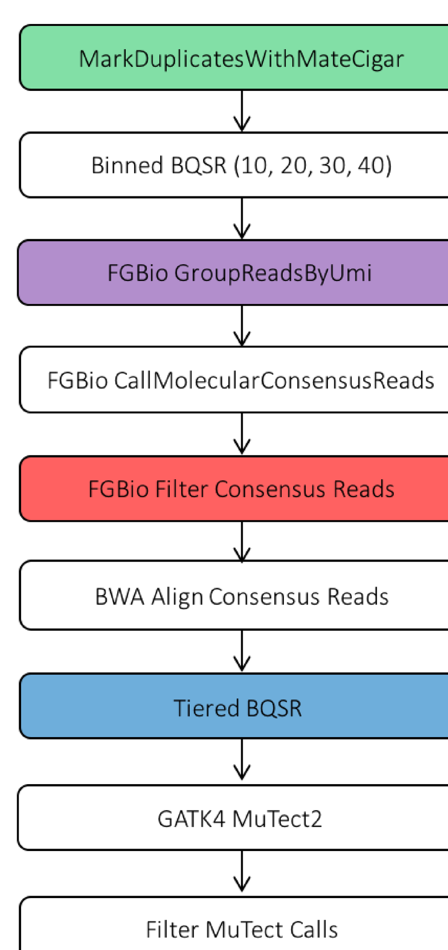



A standard pan-cancer targeted panel is also available for projects not requiring customization.

**Broad Pan-Cancer Standard Panel:**
- 2 Mb of target territory
- Exons of 396 cancer related genes
- Selected introns of 40 genes
- MSI flanking loci
- 2000 polymorphic sites for CNV analysis exploration

## Analysis Pipeline

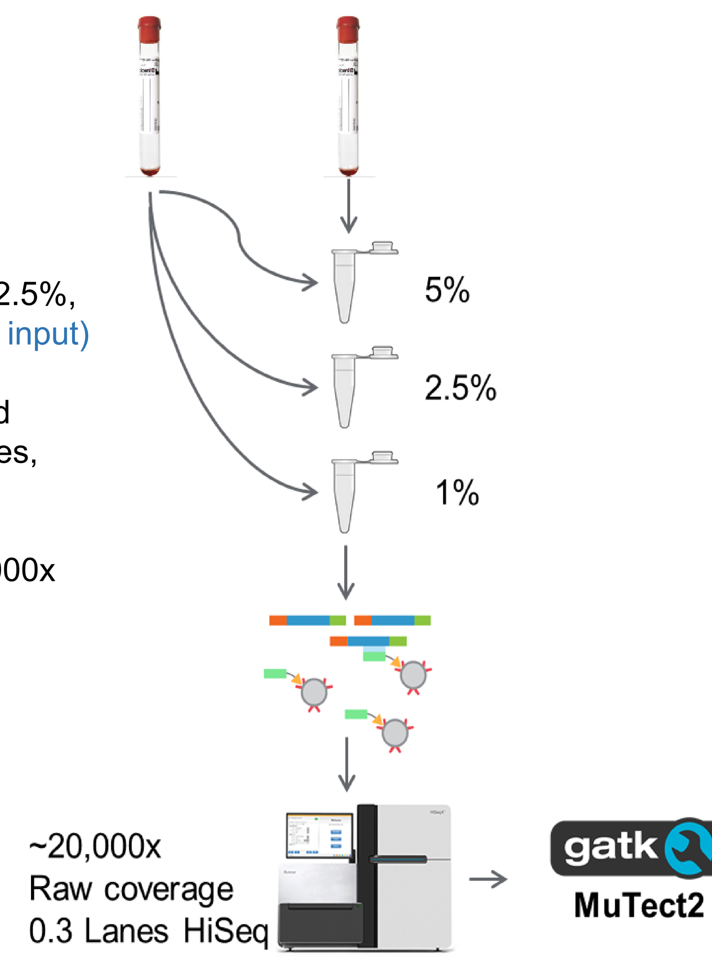
The analysis pipeline processes a cfDNA tumor paired with a genomic DNA (gDNA) normal. Reads are grouped together if they appear to have come from the same original molecule (UMI tags). CallMolecularConsensus reads joins duplicate reads into a single synthetic read which improves the quality but filters singleton read. BQSR is used to generate accurate base quality scores and variants are called using GATK4 MuTect2. Filtering is also performed using GATK4 FilterMutectCalls.




**Data Deliverables (Terra Workspace)**
Raw BAM or CRAM file
Duplex BAM or CRAM file
Annotated VCF or MAF file (variants)

## Measuring Assay Performance

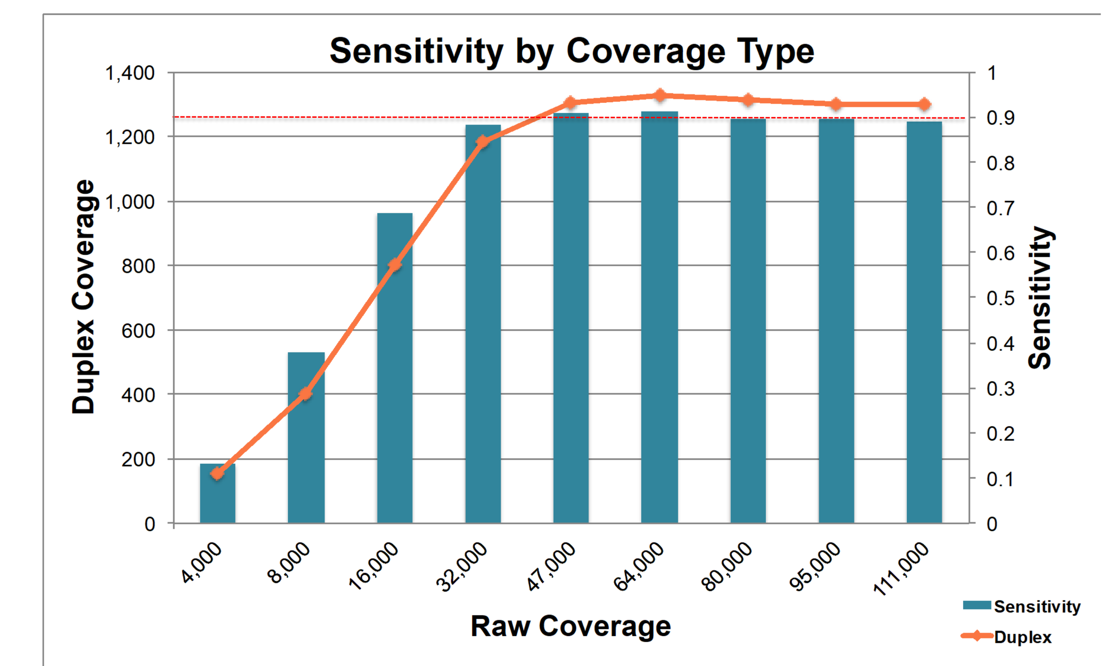
- cfDNA extracted from two healthy individuals
- Sequenced individually to generate variant truth set
- "A" spiked into "B" at 5%, 2.5%, and 1% (20ng cfDNA total input)
- Selection with Broad Blood Biopsy V1 panel (402 genes, ~2Mb territory
- Deep sequencing to ~20,000x raw coverage
- De-duplicated by UMI
- Generate consensus calls




A spike in series of healthy donor cfDNA was used to measure assay sensitivity at variable allele fraction. False positives are assayed using replicate runs of healthy donor 100% samples.

## Sensitivity as a function of sequencing depth

*What depth of sequencing is required?*




To explore the relationship between coverage and sensitivity at higher sequencing depths one sample was sequenced to 4x the target depth of the assay.

20ng of cfDNA for the 2.5% healthy donor spike-in targeted with the 2Mb Pan-Cancer panel. Ultra deep sequencing (>100,000x raw coverage) was performed and down sampled. Sensitivity as a function of raw and duplex consensus coverage is assessed. Observed minor allele fraction = 0.9%.

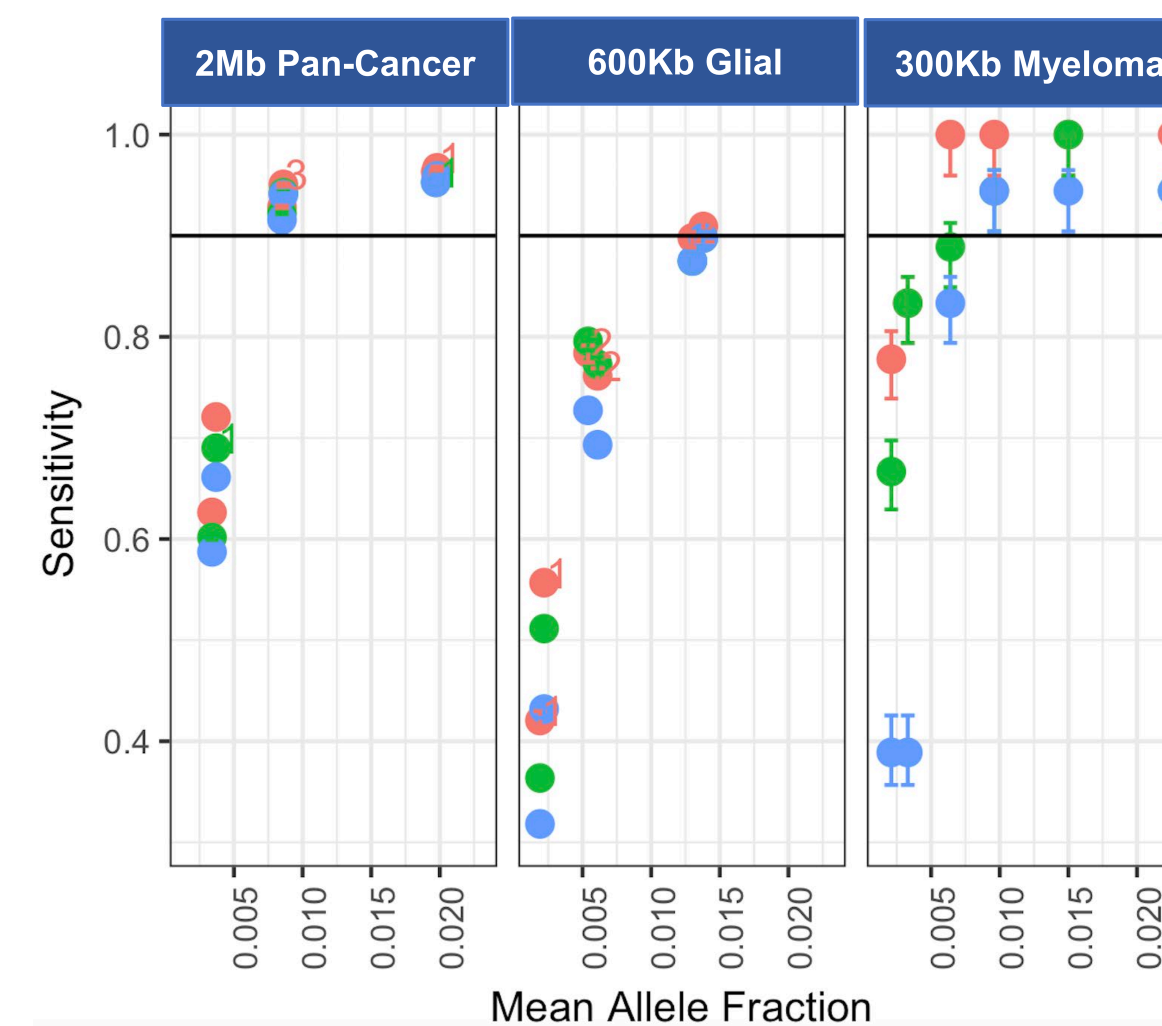
## Panel Benchmark Performance - Sensitivity and False Positive Rate

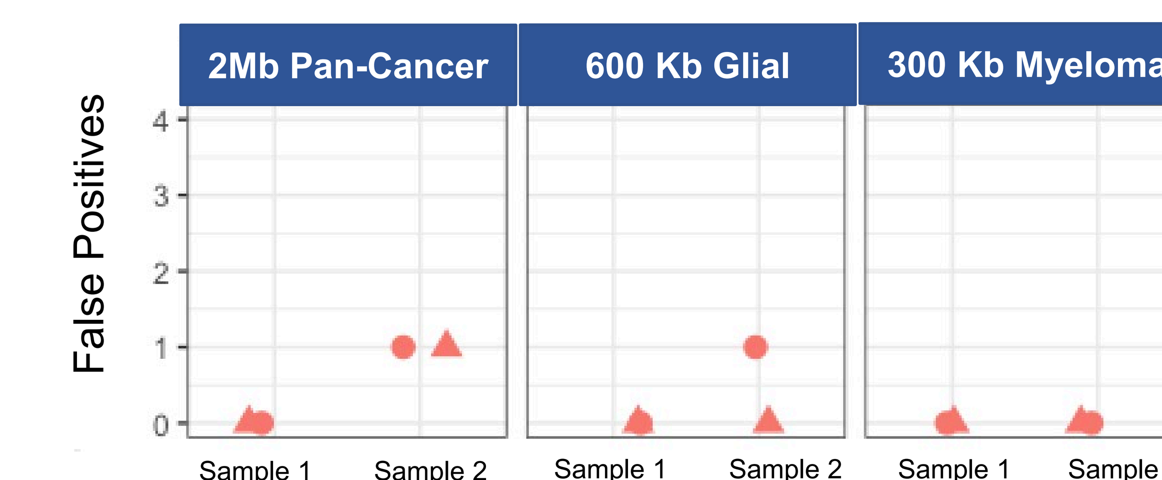
### 2Mb Pan-Cancer panel

- 20ng cfDNA input
- 2Mb pan-cancer panel
- ~25,000x raw coverage
- Sensitivity for 470 truth sites
- **>90%** sensitivity for <1% MAF
- **0.38 FP/ Mb** sequenced territory

### 317Kb Disease specific panel

- 20ng cfDNA input
- 317Kb Multiple Myeloma panel
- ~95,000x raw coverage
- **>90%** sensitivity for <1% MAF
- **0 FP/ Mb** sequenced territory




## Conclusion and Future Directions

In order to support impactful clinical research across a wide breadth of study designs, consideration for sensitivity and accuracy is paramount. By leveraging duplex UMI adapters with strong error suppression and filtering, we have built a unified workflow to support somatic and germline studies requiring sensitive detection of low allele fraction variants down to <1% MAF.
We have sought to understand and demonstrate the critical relationship between sample input amount, raw sequencing coverage and consensus duplex recovery in the context of sensitivity for detecting low allele fraction variants in cfDNA. This assay is currently undergoing clinical validation and will be offered in our Clinical Research Sequencing Platform (CRSP) this summer.
Expanding on this capacity and technology, we aim to further develop targeted assays that may lead to highly sensitive detection of minimal residual disease in cancer patients.

## Acknowledgments

Data used in this poster was generated at the Broad Institute, for more information please visit: http://genomics.broadinstitute.org/