Sequencing analysis of clinical tissue samples: Fresh Frozen vs. FFPE

Gad Getz, Ph.D.

Massachusetts General Hospital
Cancer Center and Department of Pathology

Broad Institute of Harvard and MIT

Illumina Cancer Webinar Series

May 30, 2013
What is the full set of genome alterations within the cancer (and germ-line)—mutations, copy number, translocations, epigenetic, etc? Mutational processes? Heterogeneity and tumor evolution?

(A) Discovery
Find cancer genes and pathways $\rightarrow$ cancer genome projects

(1) Which genome alterations are statistically significant in the population?
(2) In which genes and pathways do these alterations occur?
(3) What are the subtypes of the disease?
(4) How do events and subtypes correlate with clinical parameters?
(A) Discovery
Find cancer genes and pathways → cancer genome projects

**Characterization (Individual)**

- **met**
- **somatic**
- **primary**
- **germline**

**Interpretation (Population)**

The **power** comes from

- **depth of sequencing**, reduced sequencing error rates, longer reads and insert sizes (rearrangements)
- **number of patients**, improved models for background mutation processes, well controlled cohorts with clinical parameters
(B) Clinical practice
Dynamic clinical decision support and follow-up

FFPE

CancerActionDB

Interactive system

Characterization

somatic

germline

Annotation

- chr7:140099605 A>T
- BRAF V600E
- Significant in melanoma
- Activating mutation
- Sensitive to PLX4032

Clinical Decision

Treatment bases on:
- Patient data
- Tumor context
- Possible actions based genetic and clinical data
- Dynamic clinical trials
- Combination therapy
Drop in sequencing costs and ability to select parts of the genome to sequence → affordable to perform deep (100-150x) exome sequencing of many patients

Next-gen sequencing: base-level characterization

Next-gen sequencing: base-level characterization

Reference sequence
Chr 1

Point mutation
Indel

Homzygous deletion
Hemizygous deletion
Gain
Translocation breakpoint
Pathogen

Copy number alterations

muTect
Kristian Cibulskis

Indelocator
Andrey Sivachenko

SegSeq
Derek Chiang

dRanger & BreakPointer
Mike Lawrence Yotam Drier
Drier et al. Genome Res. (2013)

PathSeq
Alex Kostic C. Pedamallu
Analysis pipeline @ Broad: Reads to BAMs

Picard pipeline
Sequencing platform

Tim Fennell
Mark DePristo

Align to genome (BWA)
Mark duplicate reads
Base quality recalibration
BAM file


Mark duplicate reads: both ends need to align to the same location

Base-quality recalibration: estimate base quality based on multiple covariates.

Illumina/HiSeq 2000

Empirical Q

Machine Q

Empirical Q – Reported Q

First of pair reads

Second of pair reads

Original Recalibrated

Genome Analysis Tool Kit (GATK)

Q = -10 \log_{10}(\text{prob. of error})

DePristo et al. Nature Genetics (2011)
Visualization of BAMs using IGV

Picard pipeline
Sequencing platform

Tim Fennell
Mark DePristo

Align to genome (BWA)
Mark duplicate reads
Base quality recalibration

dbGaP
BAM file
Visualization (IGV)

Integrative Genomic Viewer (IGV) at www.broadinstitute.org/igv

Analysis pipeline @ Broad: BAMs to annotated variants

**Picard pipeline**
Sequencing platform

Tim Fennell
Mark DePristo

DNA

Align to genome (BWA)

Mark duplicate reads

Base quality recalibration

BAM file

Visualization (IGV)

dbGaP

Picard pipeline

Cancer Genome Analysis

Doug Voet

Cibulskis, McKenna et al. Bioinformatics (2011)

Characterization

Annotation (*Oncotator*) + Reports

Mutations

Indels

Purity/ploidy

Copy number

Rearrangements

Pathogens

...
What’s in a cancer sample? How did it get there?

Driver event – An event that increased the fitness of the cell when it occurred

Cancer genes / pathways -- genes / pathways that harbor driver driver events. Each is associated with a frequency of driver events in cancer

Adapted from Campbell et al. Nature (2010)
What’s the effect of having DNA come from a mixture of cells?

Mutation allelic fraction depends on purity, local absolute copy number and multiplicity

The **mutation allelic fraction** is the fraction of alleles (DNA molecules) from a locus that carry a mutation ➔ Also the expected fraction of supporting reads

DNA from tumors is a mix of normal genomes and complex tumor genomes

- **Purity**: % tumor cells
- **Ploidy**: mass of DNA in tumor cells

Purity = 67%

Absolute copy number in tumor = 4

Mutation multiplicity = 1 ➔ Allelic fraction = 2/10 = 0.2
Mutation allelic fraction depends on purity, local absolute copy number and multiplicity and cancer cell fraction (CCF).

but what if the mutation is only in a fraction of cancer cells?

**CCF = 1**

Purity = 67%
Absolute copy number in tumor = 4
Mutation multiplicity = 1
**Cancer cell fraction (CCF) = 1**
⇒ Allelic fraction = 2/10 = 0.2

**CCF = 0.5**

Purity = 67%
Absolute copy number in tumor = 4
Mutation multiplicity = 1
**Cancer cell fraction (CCF) = 0.5**
⇒ Allelic fraction = 1/10 = 0.1

Detecting Somatic Single Nucleotide Variations (SSNVs)

MuTect: A Highly Sensitive Somatic Mutation Caller


Available at: [www.broadinstitute.org/cancer/cga/mutect](http://www.broadinstitute.org/cancer/cga/mutect)
(1) Need a highly sensitive and specific mutation caller. Benchmarking is critical to evaluate tools and monitor progression.

A mutation caller is a classifier asking at every genomic locus “Is there a mutation here?” Classifiers are evaluated using the Receiver Operators Characteristic (ROC) curve.

In fact, a mutation caller is characterized by many ROC curves depending on the
(i) allele fraction
(ii) coverage of tumor and normal
(iii) sequencing/alignment noise

adapted from http://en.wikipedia.org/wiki/Receiver_operating_curve
What is the signal that we are looking for? How sensitive can we be?

**Frequency:** Somatic single nucleotide variations occur ~ 1 mutation / Mb

**Number of alternate reads:** Depends on sequencing coverage and allelic fraction

**Allelic fraction** = 0.1

**33% N**

**67% T**

**30x coverage**

- $\text{Prob (0 alt reads)} = 4.2\%$
- $\text{Prob (1 alt read)} = 14.1\%$
- $\text{Prob (2 alt reads)} = 22.8\%$
- $\text{Prob (3 alt reads)} = 23.6\%$
- $\text{Prob (4 alt reads)} = 17.7\%$

- Theoretical sensitivity
  - $\geq 1$ 95.8%
  - $\geq 2$ 81.6%
  - $\geq 3$ 58.9%
  - $\geq 4$ 35.2%
  - $\geq 5$ 17.5%

**Binomial distribution:** 30 trials with $p=0.1$
What is the noise? two types of false positives

1. NO EVENT
   - At risk: Every base
   - Source: Misread bases, Sequencing artifacts, Misaligned reads

2. GERMLINE EVENT (in T+N)
   - At risk: ~1000 germline / Mb (known)
     10-20 rare germline / Mb (novel)
   - Source: Low coverage in normal

⇒ Need to measure the two types of false positives (at dbSNP and non-dbSNP sites)
Back of the envelope calculation of false positive rate

1. NO EVENT

At risk: Every base
Source: Misread bases, Sequencing artifacts, Misaligned reads

Assumptions:
- 1 incorrect base called in ~3000 sequenced bases (typical value)
- Independent errors
- Sequence depth of 30x

99.7% of sites will have no errors
0.3% of sites will have 1 error
5 x 10^-6 of sites will have 2 errors
0.005 x 10^-6 of sites will have 3 errors

We must require at least 3 alternate reads for the false positive rate < 0.05 x 10^-6

In reality, since the sequencing errors are not independent the fractions are much higher but still (for 30x coverage) 3 alternate reads together with appropriate filters reach the desired false positive rate.
‘Virtual tumor’ approach to measure specificity and sensitivity (ROC curve)
MuTect: higher sensitivity with similar false positive rate

- Sensitivity calculations
- ROC curve at 30x, allele fraction=0.1 using real data

MuTect can detecting low allele fraction mutations, hence suited for studying impure and heterogeneous tumors

ABSOLUTE: Estimating purity, absolute copy number and cancer cell fraction (CCF)


Available at: www.broadinstitute.org/cancer/cga/ABSOLUTE
ABSOLUTE: Inference of purity and average ploidy

Cancer cell fractions may have large uncertainties

Distribution of CCF of point mutations (purple) and copy-number alterations (cyan).

Assuming there are distinct sub-populations of cancer cells, subclonal mutations and copy-number alterations should cluster in a small unknown number of discrete values ➔

A Bayesian clustering algorithm and integration over uncertainty in number of clusters

Scott Carter
Subclonal mutations travel in packs: example sample

Before clustering

After clustering
Subclonal mutations travel in packs: extended to 2D

The same methods work for **longitudinal** analysis

Here – a CLL sampled a two timepoints
Analysis of FFPE samples
Why use FFPE?

• **Very** large numbers of samples in tissue banks and Biorepositories worldwide

• Samples often very well-characterized with histological, pathological and follow-up clinical data

→ Can fill the accrual gap in cancer genome projects
→ “We need to get to 10,000 patients per tumor type” -- Lou Staudt (Nov 2012)

• Remains part of clinical standard of care (difficult to change pathology practices for research needs alone)

→ Enable connecting to existing clinical trials and move genomic analyses into standard clinical practice
Challenges with FFPE?

- Difficulty of extracting samples
  - Deparaffinization & de-cross-linking of protein-DNA.
  - Degradation of RNA
  - Physical size of the samples can be small
  - Yield

- Poor quality of extracted material due to:
  - Warm-ischemic time in operating room
  - Type of formalin used, how fixed, & how long (un-buffered vs. buffered)
FFPE samples vary in size (TCGA samples)

FFPE Block Choices from Pilot Round #2

Group #1 (Large Tissues)

Group #2 (Medium Tissues)

Group #3 (Small Tissues)

Nationwide Children’s Hospital Biospecimen Core Resource
Samples from clinical study of drug resistance (Broad)

Small

Very small

Tiny

Where is it?

Kristin Ardlie
General protocol

- **Extraction of DNA**: FFPE DNA Extraction was done using a slightly modified version of the Qiagen QIAamp DNA FFPE protocol. Extraction from either 2mm cores or slides.

- **Library construction**: We used a slightly modified version of the protocol described in Fisher et al. *Genome biology* 12, R1 (2011). Reduced DNA input to 10ng – 100ng.

- **QC Metrics**: We use various QC metrics such as calculated in the Picard HsMetrics tool (picard.sourceforge.net): % reads on target, %bases coverage >20x, library size.
Questions regarding the use of FFPE samples

1. Can we generate high quality data from FFPE samples?
2. Can we detect the genomic events in FFPE data? Are they real?
3. What is the correlation/overlap of genomic events between FFPE and fresh frozen (FF)?
4. Can we perform cancer genome projects with FFPE samples?
5. Can we use these data to make clinical decisions?
TCGA Prostate – “trios”
- 4 FFPE Tumor samples + 4 Fresh Frozen Tumor/Normal pairs
- Sequencing Coverage:
  - FFPE samples: 200x
  - Fresh Frozen pairs: 100x

TCGA Colon – “trios”
- 3 FFPE Tumor samples + 3 Fresh Frozen Tumor/Normal pairs
- Sequencing Coverage:
  - FFPE samples: 200x
  - Fresh Frozen pairs: 100x

Lung Cancer, NSCLC Adenocarcinoma – “quartets”
- 17 FFPE Tumor/Normal sample pairs + 17 Fresh Frozen Tumor/Normal pairs
- Source = FFPE Sections (15 microns, 9 per sample), Ontario, Canada
- Age of Fixed Block = 2007 - 2010
Exome seq: (1) Data quality → similar coverage

3 Colon, 4 prostate
FFPE 200x, FF 100x

% of Target Bases Covered

Estimated Library Size

Exome-seq: (2) Can we find mutations?
Total Count of mutations is similar

<table>
<thead>
<tr>
<th>Frozen</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense_Mutation</td>
<td>83</td>
</tr>
<tr>
<td>Nonsense_Mutation</td>
<td>8</td>
</tr>
<tr>
<td>Silent</td>
<td>41</td>
</tr>
<tr>
<td>Splice_Site</td>
<td>2</td>
</tr>
<tr>
<td>Translation_Start_Site</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
</tr>
</tbody>
</table>

Total territory: 130.49 MB

<table>
<thead>
<tr>
<th>FFPE</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense_Mutation</td>
<td>83</td>
</tr>
<tr>
<td>Nonsense_Mutation</td>
<td>6</td>
</tr>
<tr>
<td>Nonstop_Mutation</td>
<td>1</td>
</tr>
<tr>
<td>Silent</td>
<td>40</td>
</tr>
<tr>
<td>Splice_Site</td>
<td>6</td>
</tr>
<tr>
<td>Translation_Start_Site</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>137</td>
</tr>
</tbody>
</table>

Total territory: 130.66 MB

<table>
<thead>
<tr>
<th>Frozen</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>De_novo_Start_OutOfFrame</td>
<td>2</td>
</tr>
<tr>
<td>Frame_Shift_Del</td>
<td>25</td>
</tr>
<tr>
<td>Frame_Shift_Ins</td>
<td>9</td>
</tr>
<tr>
<td>In_Frame_Del</td>
<td>11</td>
</tr>
<tr>
<td>In_Frame_Ins</td>
<td>2</td>
</tr>
<tr>
<td>Missense_Mutation</td>
<td>3551</td>
</tr>
<tr>
<td>Nonsense_Mutation</td>
<td>236</td>
</tr>
<tr>
<td>Nonstop_Mutation</td>
<td>7</td>
</tr>
<tr>
<td>Silent</td>
<td>1225</td>
</tr>
<tr>
<td>Splice_Site_DNP</td>
<td>4</td>
</tr>
<tr>
<td>Splice_Site_SNP</td>
<td>109</td>
</tr>
<tr>
<td>Start_Codon_Del</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>5332</td>
</tr>
</tbody>
</table>

Total territory: 499.2 Mb

<table>
<thead>
<tr>
<th>FFPE</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>De_novo_Start_OutOfFrame</td>
<td>1</td>
</tr>
<tr>
<td>Frame_Shift_Del</td>
<td>17</td>
</tr>
<tr>
<td>Frame_Shift_Ins</td>
<td>4</td>
</tr>
<tr>
<td>In_Frame_Del</td>
<td>8</td>
</tr>
<tr>
<td>In_Frame_Ins</td>
<td>3</td>
</tr>
<tr>
<td>Missense_Mutation</td>
<td>3428</td>
</tr>
<tr>
<td>Nonsense_Mutation</td>
<td>270</td>
</tr>
<tr>
<td>Nonstop_Mutation</td>
<td>3</td>
</tr>
<tr>
<td>Silent</td>
<td>1152</td>
</tr>
<tr>
<td>Splice_Site_DNP</td>
<td>4</td>
</tr>
<tr>
<td>Splice_Site_SNP</td>
<td>122</td>
</tr>
<tr>
<td>Splice_Site_TNP</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>5013</td>
</tr>
</tbody>
</table>

Total territory: 512.2 Mb

4 prostate

17 lung
Exome-seq: Are the FFPE mutations artifacts? No – similar mutation spectra

4 prostate

17 lung

Broad GSC
Exome-seq: (2) Can we detect copy number changes? (3) Are they the same?  Example #1

Using CapSeg, Aaron McKenna, Scott Carter  Broad GSC
Exome-seq: (2) Can we detect copy number changes? (3) Are they the same? Example #2
Exome-seq: (3) Are we finding the same mutations in FFPE and frozen?

**Lung**
- FFPE: 1909
- Frozen: 3230
- Intersect: 2225
- Percentage: 44%

**Prostate**
- FFPE: 77
- Frozen: 56
- Intersect: 26.5%

![Bar chart showing mutation rates](chart.png)
A fundamental observation: When comparing frozen to FFPE we are changing TWO variables at once

(1) Frozen vs FFPE
(2) Two different pieces of the tumor
    -- Different in terms of tumor purity
    -- Different with respect to sub-clonal composition

THIS AFFECTS ALL COMPARISONS BETWEEN FFPE AND FROZEN SAMPLES (DNA, RNA, PROTEINS)
Sensitivity to detect (and even observe) a mutation – depends on coverage and allelic fraction

Purity = 67%
Absolute copy number in tumor = 4
Mutation multiplicity = 1
⇒ Allelic fraction = 2/10 = 0.2


ABSOLUTE: SNP arrays / exome-seq → purity, ploidy & abs. copy-number profile
Allelic fraction in frozen and FFPE are different due to differences in purities

17 lung
Instead of independent calling → validate

- The missing mutations actually do exist
- Instead of calling → validate

**Validation**

- Convert allelic-fraction from sample that it was detected to the other
- Consider only mutations that we had statistical power to detect in the other sample

Andrey Sivachenko, Mara Rosenberg
But not all mutations are clonal ... 
We can distinguish between clonal & sub-clonal mutations

A single ovarian cancer

ABSOLUTE can identify clonal and sub-clonal mutations

Clonal mutations, validate = 2+ reads, corrected allelic fraction, power ≥ 95%

~90% validation rate, even better at high allele fractions

Petar Stojanov, Eli Van Allen
Interpretation

How can we distinguishing “driver” from “passenger” alterations?

- Model background mutational processes
- Identify genes/regions/pathways with more mutations than predicted by the background model

candidate ‘driver’ events

or

inaccurate background mode
systematic artifacts in mutation calling
Tools for detecting significantly altered cancer genes / regions / pathways

GISTIC 1.0
Beroukhim et al. PNAS (2007)
GISTIC 2.0
Mermel et al. Genome Biol. (2011)

Uses: Frequency and amplitude of events
Separates broad and focal gains and losses

MutSig
Lawrence et al. in development
Uses: Number and types of mutations;
corrects BMR for expression, gene footprint size etc.
Works on genes, genesets and conserved regions (intervals on the genome)

NetSig (in development)
Zou et al., in development
Uses: all types of alterations to identify clusters of mutated genes in protein-protein networks

Craig Mermel, Rameen Beroukhim, Steve Schumacher, Mike Lawrence, Lihua Zou, Alex Ramos, Gregory Kryukov, Petar Stojanov
MutSig: Approach -- Score genes/genesets according to number and types of mutations

Gene-specific models, expression levels using intron mutations (as in InVEx), and more
(4) Long tailed distribution of cancer genes

**GBM (n=84)**

**Lung Adenocarcinoma (n=188)**

**Head & Neck (n=74)**

**Ovarian (n=316)**

+ BRAF, PIK3CA, KRAS, NRAS
(5) Can we perform cancer genome projects using FFPE samples?

Yes! Very similar MutSig lists

Orange background if within top 30 of other list

Old MutSig version

<table>
<thead>
<tr>
<th>Frozen</th>
<th>FFPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>rank</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

Broad GSC
(6) Can we sequence clinical FFPE samples for clinical decision making? Yes!

<table>
<thead>
<tr>
<th>Gene</th>
<th># in Frozen</th>
<th># in FFPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>KRAS</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>EGFR</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>STK11</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>KEAP1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ATM</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>NF1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

17 lung samples
The importance of subclonal mutations in clinical sequencing: Example from Chronic Lymphocytic Leukemia

Landau, Carter, Stojanov et al.  
Cell (Feb 14, 2013)

Collaboration with Catherine Wu’s lab from DFCI
Study design

149 CLL samples, 18 sampled at 2 timepoints

WES / SNP array

Allelic fractions / copy ratios

ABSOLUTE (Carter et al., 2012)

Cancer Cell Fractions (CCF)

Clonal mutations: Affecting *all* cells;
- Founder and earlier events

Subclonal mutations: Affecting a *subpopulation*;
- Later events
25 recurrent drivers in CLL

* 9 novel putative CLL drivers identified

16 previously reported drivers in:
Wang et al., NEJM, 2011
Quesada et al., Nat Gen, 2011
Fabbri et al., JEM, 2011
Brown et al., Clin Can Res, 2011
Edelmann et al., Blood 2012
Patterns of clonal evolution in longitudinal samples (n=18)

No evolution

Evolution

Red contours denote a CCF increase of > 0.2
Stable relative subpopulation sizes in CLLs \textit{without} intervening therapy (n=6)
Frequent clonal evolution in CLLs with intervening therapy (n=12)

- Expanding mutations are highly enriched in drivers
- Patients receiving therapy have frequent clonal evolution

Are subclonal drivers associated with clinical outcome?
CLLs *with* and *without* subclonal drivers in single timepoint samples

Sequenced CLL samples (n=149)

Driver mutations

Subclonal driver present (n=68)

Subclonal driver absent (n=81)
Subclonal drivers impact clinical outcome in single time-point analysis.
Emergence of fitter subclones following treatment
Summary 1/2

- Exome Sequencing of FFPE samples is robust – we can extract DNA, capture and sequence
- We need to carefully calculate overlap between FFPE and frozen samples controlling for relative coverage and adjust for different allelic fractions
- Mutation rates and categories are very similar
- Sub-clonal mutations contribute to the differences

➡️ We can perform cancer genome project based on FFPE material
➡️ We can use clinical FFPE samples for exome sequencing

A fundamental problem when comparing frozen to FFPE is that we are analyzing two different parts of the tumor.
In TCGA we are now taking multiple samples from the same tumor, frozen and FFPE, to compare within and between frozen and FFPE.
Summary 2 / 2

Tasks of cancer genome projects: **characterization** and **interpretation**

**Characterization** is also needed for **clinical use**.

**Characterization**

1. Systematic benchmarking of analysis methods is needed: ROC curves
2. MuTect can detect low-allele fraction mutations.
3. Need estimates of purity and local absolute copy-numbers (using ABSOLUTE) to estimate cancer cell fractions – can distinguish clonal and sub-clonal mutations.

**Interpretation**

4. Significance analysis yields a long-tailed distribution of significantly mutated genes
5. Existence of sub-clonal drivers may affect the outcome of CLLs and potentially other tumor types.
Acknowledgements

Analysis Team
Mike Lawrence
Kristian Cibulskis
Petar Stojanov
Scott Carter
Andrey Sivachenko
Mara Rosenberg
Eli Van Allen
Nickhil Wagle
Craig Mermel
Yotam Drier
Gordon Saksena
Marcin Imielenski
Lihua Zou
Alex Ramos
Mike Berger
Trevor Pugh
David DeLuca
Mike Chapman
Aaron McKenna
Gregory Kryukov
Alex Kostic
Nicolas Stranksy
Derek Chiang
Roel Verhaak
Shantanu Benerji
Chip Stewart

Firehose
Doug Voet
Pei Lin
Lee Lichtenstein
Raktim Singh
Peter Carr

Cathy Wu
Jennifer Brown
Lili Wang
Youzhong Wan
Dan-Avi Landau

GDAC
Mike Noble
Robert Zupko
Dan DiCarra
Haili Zhang
Juok Cho
Spring Liu

Rameen Beroukhim
Steve Schumacher
Barbara Tabak

Broad Institute
Genomics Platform
Tim Fennell
Sara Chauvin
Lauren Ambrogio
Sheila Fisher
Joshua Levin
Xian Adiconis
Andreas Gnirke
David Jaffe
Toby Bloom
Chad Nusbaum
Wendy Winckler
Robb Onofrio
Brendan Blumenstiel
Huy Nguyen
Mellisa Parkin

Broad Institute Biological Samples Platform
Kristin Ardlie

Project Management
Carrie Sougnez
Erica Shefler
Elizabeth Nickerson
Daniel Auclair
Marisa Cortes
Kristin Thompson

Jim Robinson
Helga Thorvaldsdottir
Marc-Danie Nazaire
Jill Mesirov

Mark DePristo
Eric Banks
Kiran Garimella