BARNES-JEWISH Hospital
BJC HealthCare

Washington University in St. Louis

SITEMAN CANCER CENTER
A NATIONAL CANCER INSTITUTE COMPREHENSIVE CANCER CENTER

Fundamentals of Next-Generation Sequencing: Technologies and Applications

Society for Hematopathology
European Association for Haematopathology
2017 Workshop

Eric Duncavage, MD
Washington University in St. Louis

Disclosures

- None

Outline

- **Technologies**
  - NGS data generation
  - Enrichment methods
  - Basic Analysis
- **Applications**
  - Diagnostic sequencing in myeloid malignancies
  - Sequencing for measurable residual disease (MRD)
  - Sequencing based karyotyping

Clinical NGS Overview

(in one slide)

NGS Library Generation

1 DNA

Fragmentation

2

Library Prep

3

Enrichment

4

KRAS

gene capture probes

Limited PCR

5 Enriched Library

6

Alignment

laser

Paired-end next generation sequencing

7

KRAS

exon2

exon3

Alignment

8

Tier 1 Informatics

BAM FIle

Cleaned BAM FIle

SNV

ACTAGCTAGCCAG

ACTAGCGAGCCAG

indel

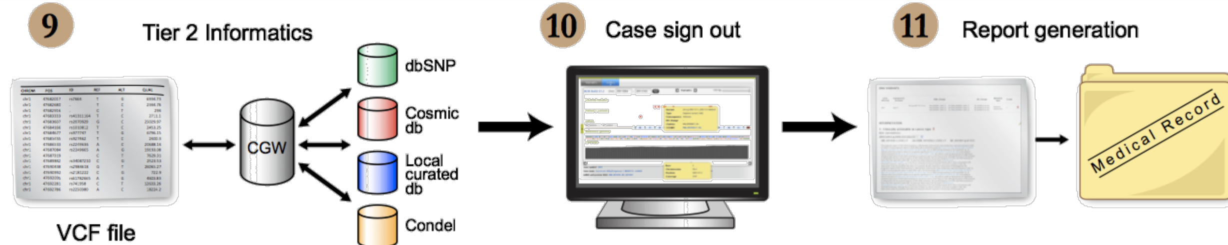
GCTAGC TAGCCG

GCTAGCTAGCTGTCTAGCCG

translocation

GCTACTCAT GCATCCAGC

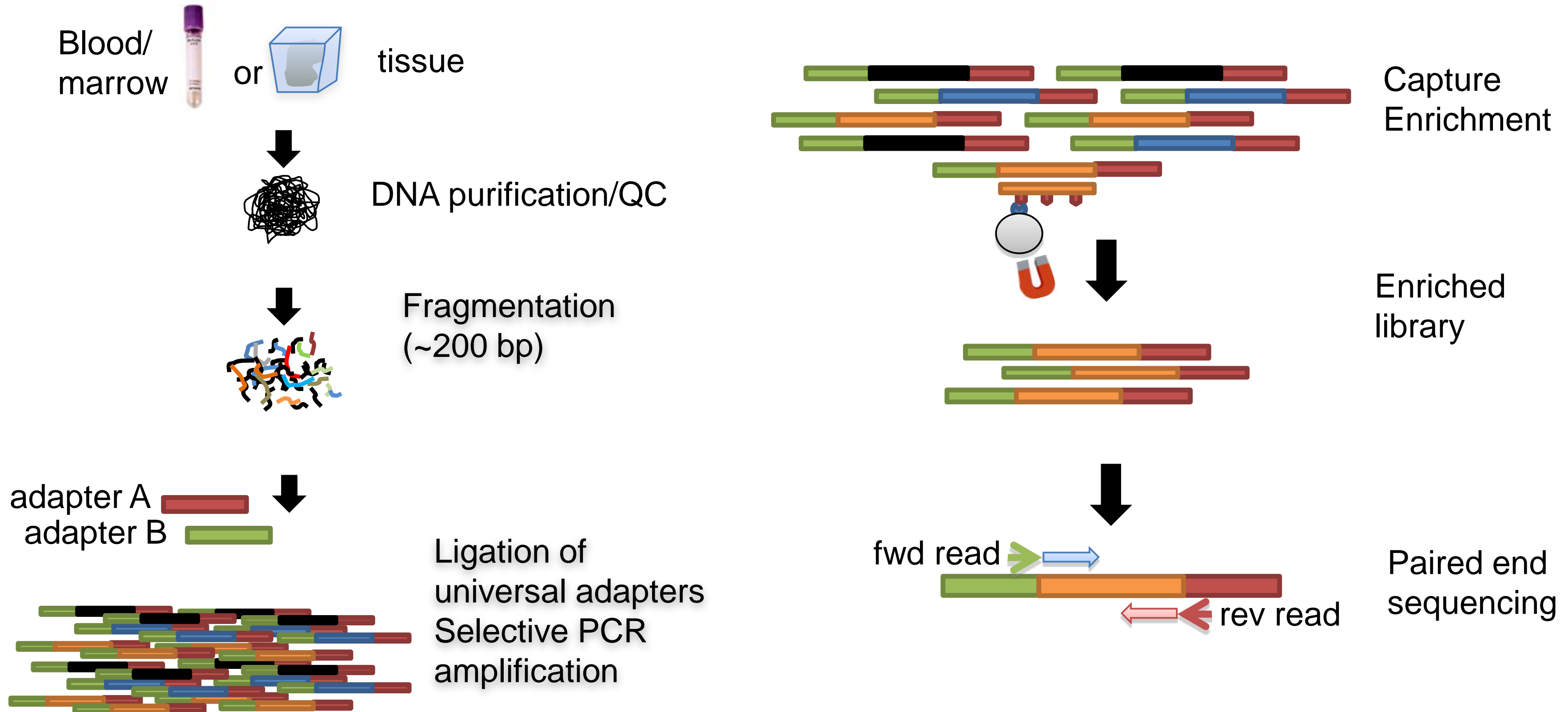
GCTACTCATGCATCCAGC



DNA Sequencing Enrichment Methods

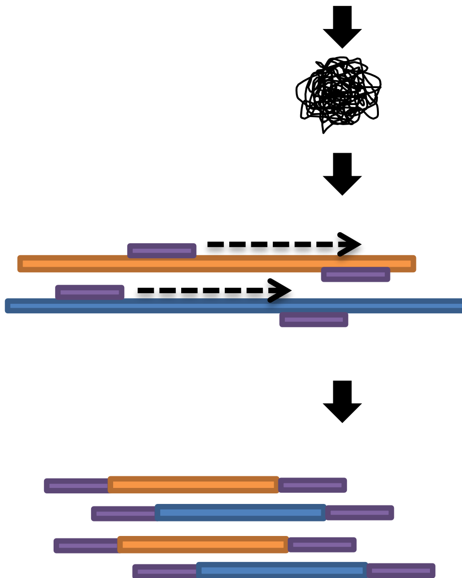
- **Multiplex PCR based enrichment**
  - Advantage: lower input DNA amounts, rapid, workflow
  - Disadvantage: cannot target large regions, or unique reads
- **Hybrid capture based enrichment**
  - Advantage: large target space, identifies unique reads
  - Disadvantage: slower, more DNA input, lower on target
- **Ligation based enrichment**
  - Advantage: fast, large target space
  - Disadvantage: some regions difficult to target

Capture-Based Enrichment



PCR-Based Enrichment

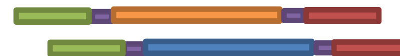
Blood/ marrow or tissue



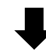
DNA purification/QC

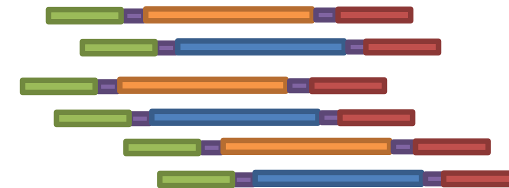
Multiplex PCR of genomic DNA with primers of interest

Amplified genes of interest

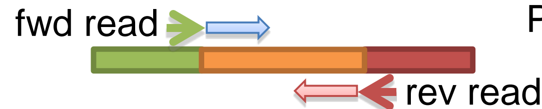


Digest primers and add sequence adapters





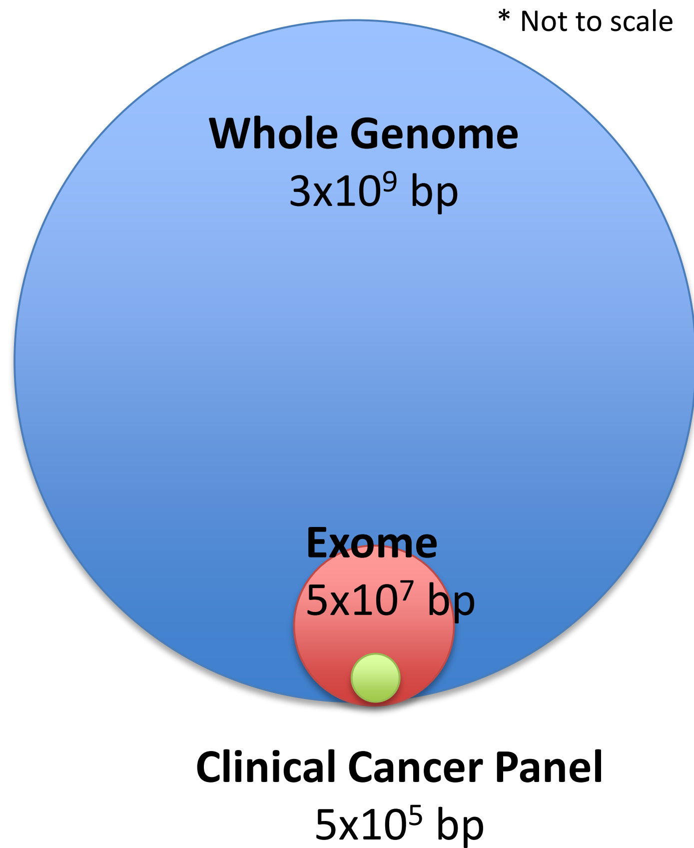
Limited cycle PCR



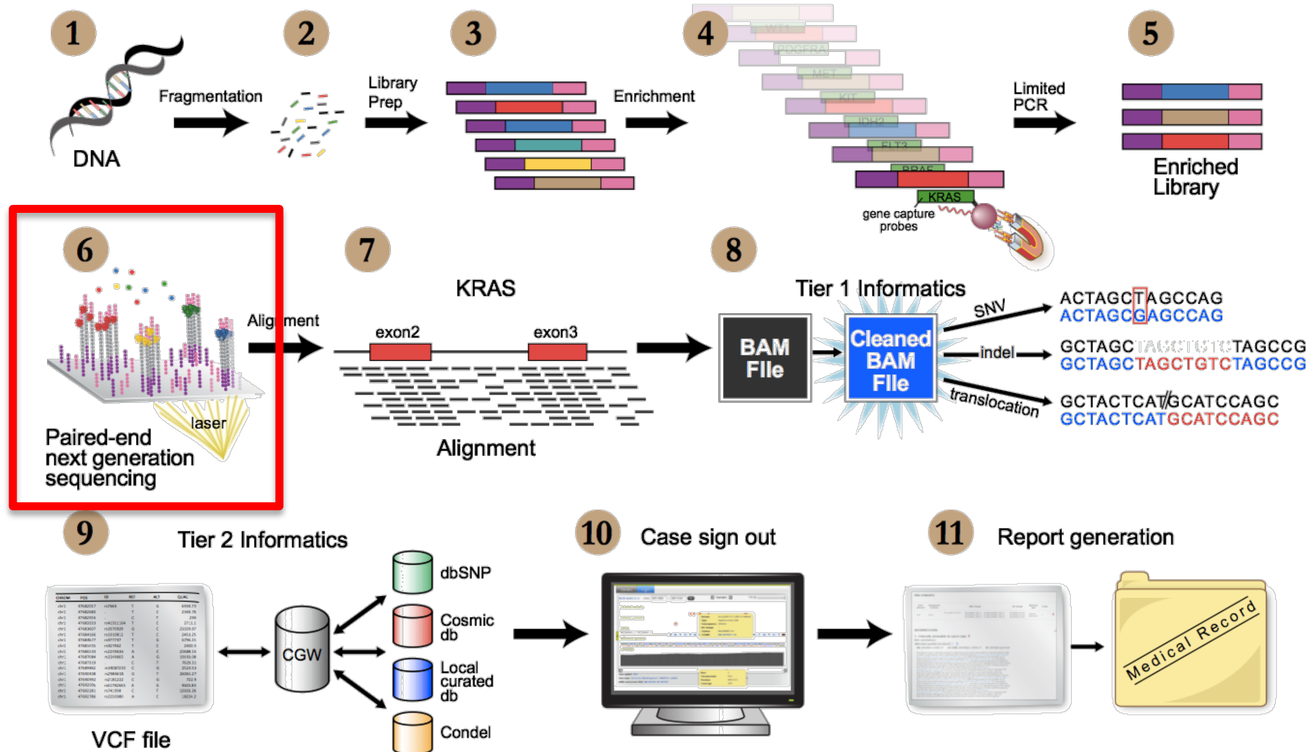
Paired end sequencing

Gene Targeting Strategies

- **Whole genome sequencing**
  - Advantages: Unbiased, detects fusions, CNVs, etc
  - Disadvantages: Expensive, limited sensitivity
- **Exome Sequencing**
  - Advantages: Detects any coding mutation, CNVs
  - Disadvantages: Expensive, limited sensitivity, no fusions
- **Hematologic Malignancy Panels**
  - Range from <50 genes to >400 genes
  - May use amplion, capture, or ligation based enrichment
  - Advantages: inexpensive, rapid, highly sensitive
  - Disadvantages: biased to targeted genes, limited structural variant detection



Sequencing



Next Generation Sequencing

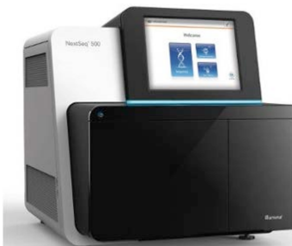
- NGS or massively parallel sequencing refers to a group of sequencing technologies that:
  - Perform numerous sequencing reactions simultaneously through nano-scale engineering
  - Use of a ‘sequencing by synthesis’ approach
  - Generate short reads
- Two major platforms in clinical labs
  - ThermoFisher Ion Torrent
  - Illumina Instruments

Instrumentation

- Higher throughput
- Higher cost per run
- Lower cost per base

ThermoFisher

Illumina



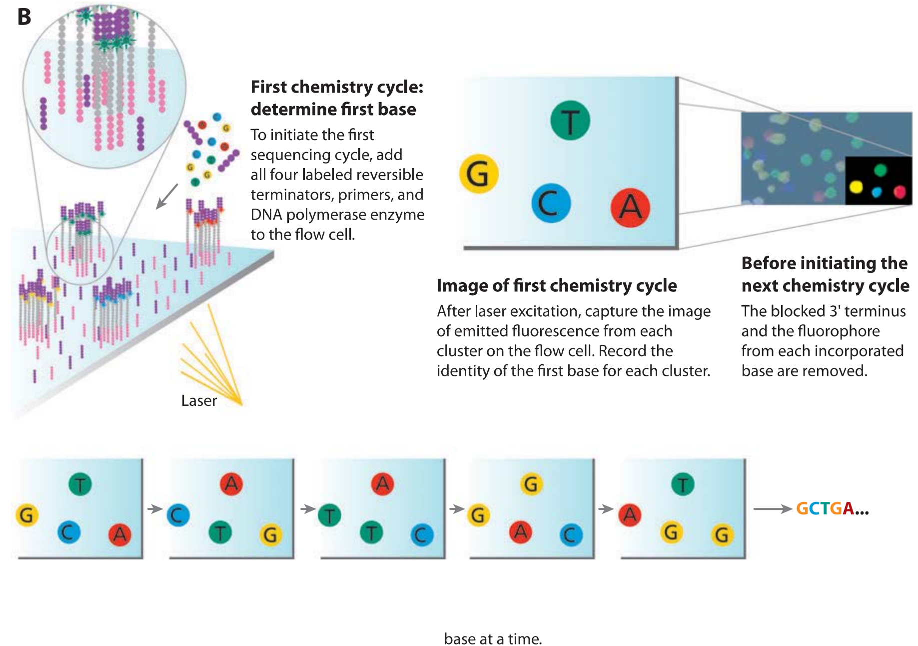
NGS Data Generation (Illumina)

A

Nucleotides

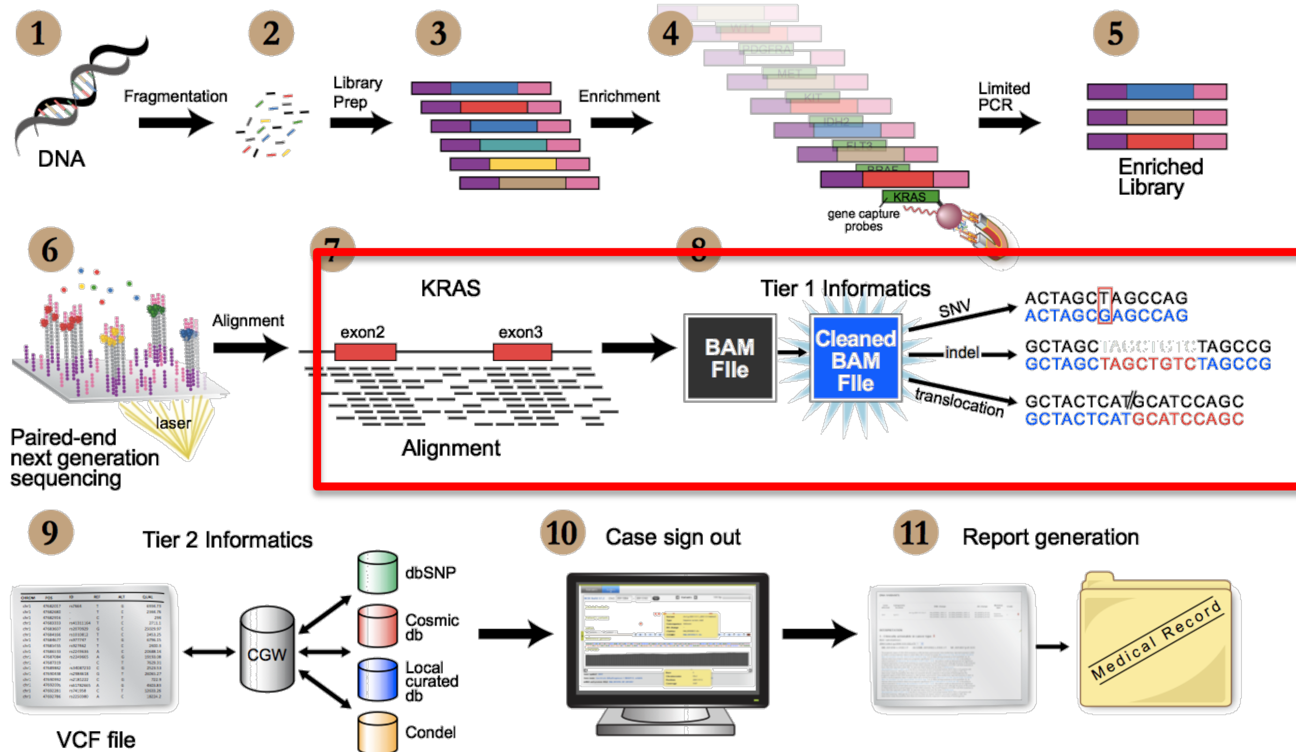
and enzyme to initiate solid-
phase bridge amplification.

**Denature the double
stranded molecules**



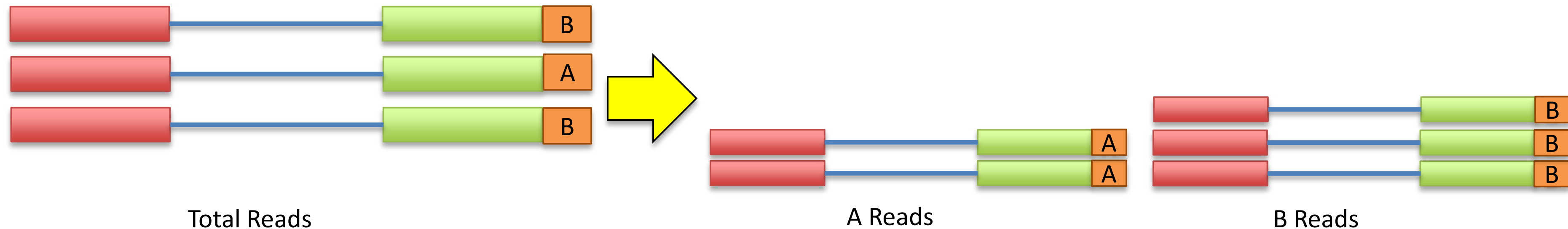
Mardis ER. Annu Rev Genomics Hum Genet, 2018

Data Analysis



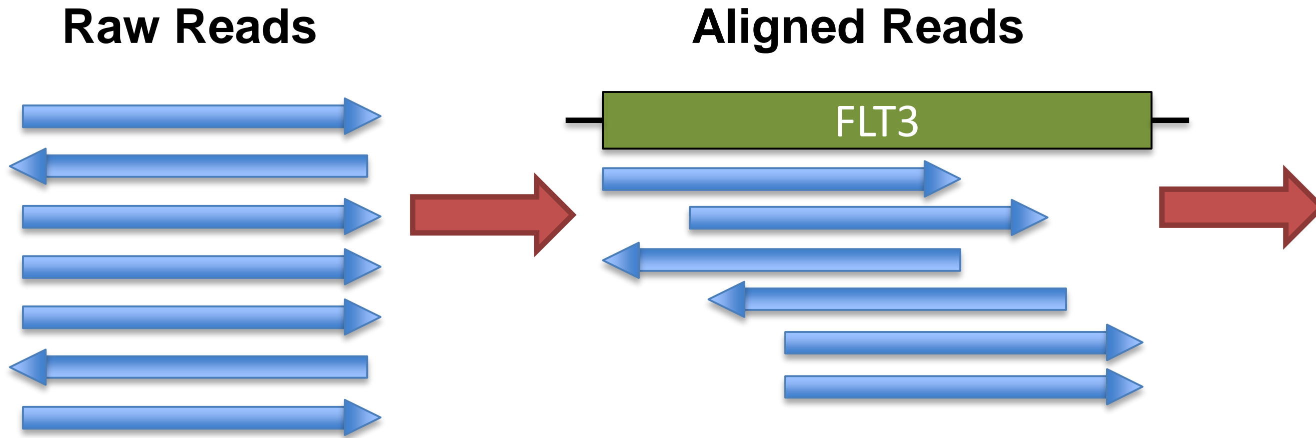
Sample De-multiplexing

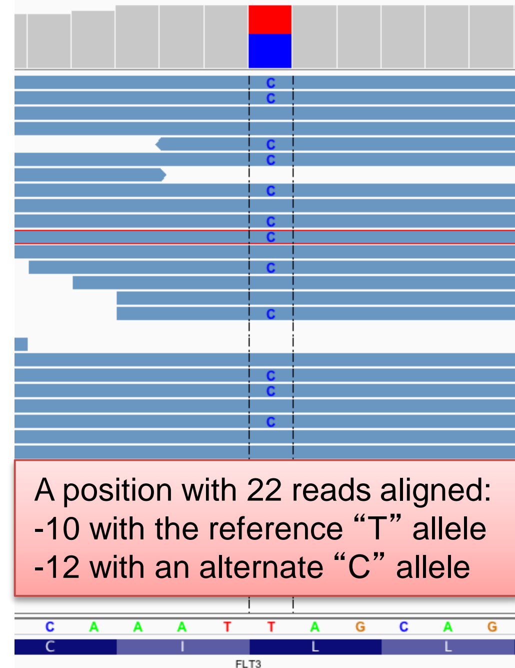
- For most clinical applications, multiple cases are sequenced in the same NGS lane.
- FASTQs must first be **demultiplexed** based on sample-specific barcodes and split into separate files.



Read Mapping and Variant Calling

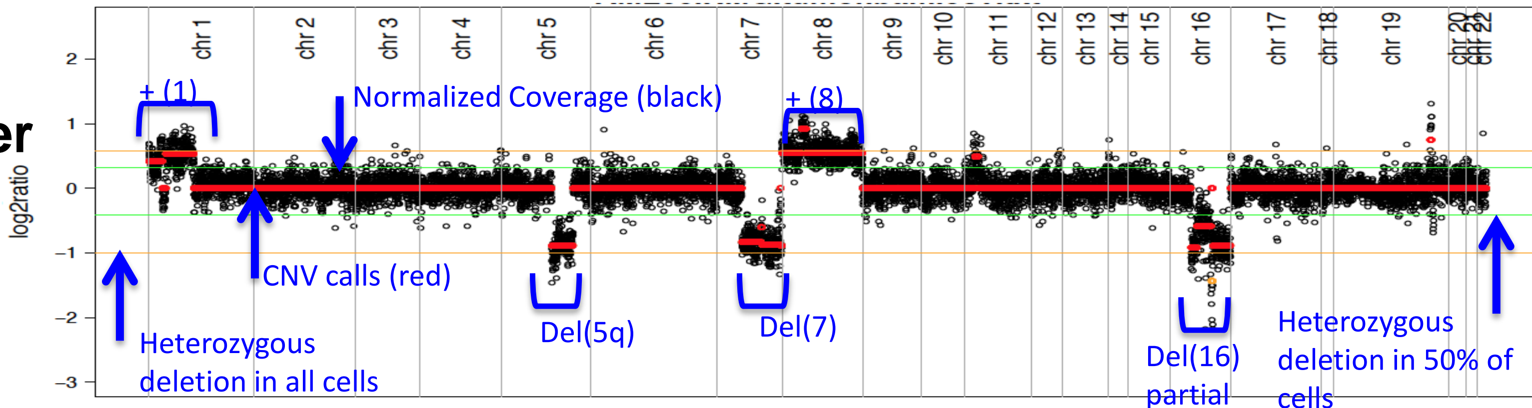
- To make sense of short reads they have to first be aligned to the human genome
- Most computationally intensive step of the process



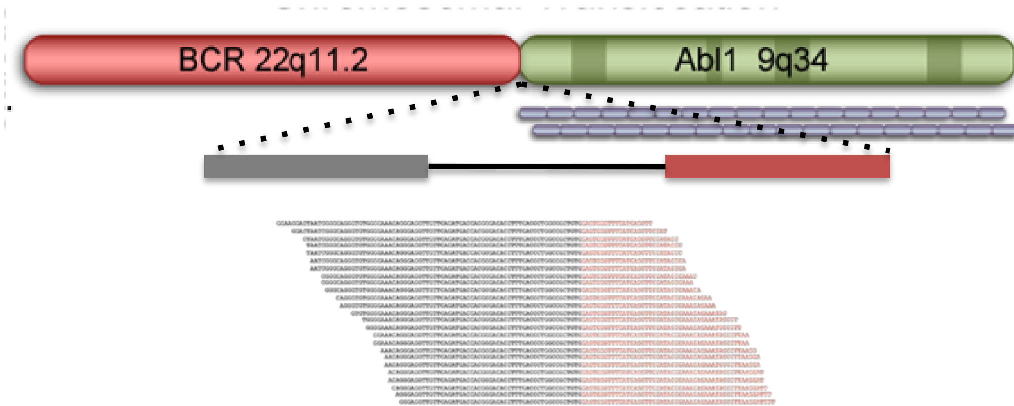


NGS can Identify all Mutations Types

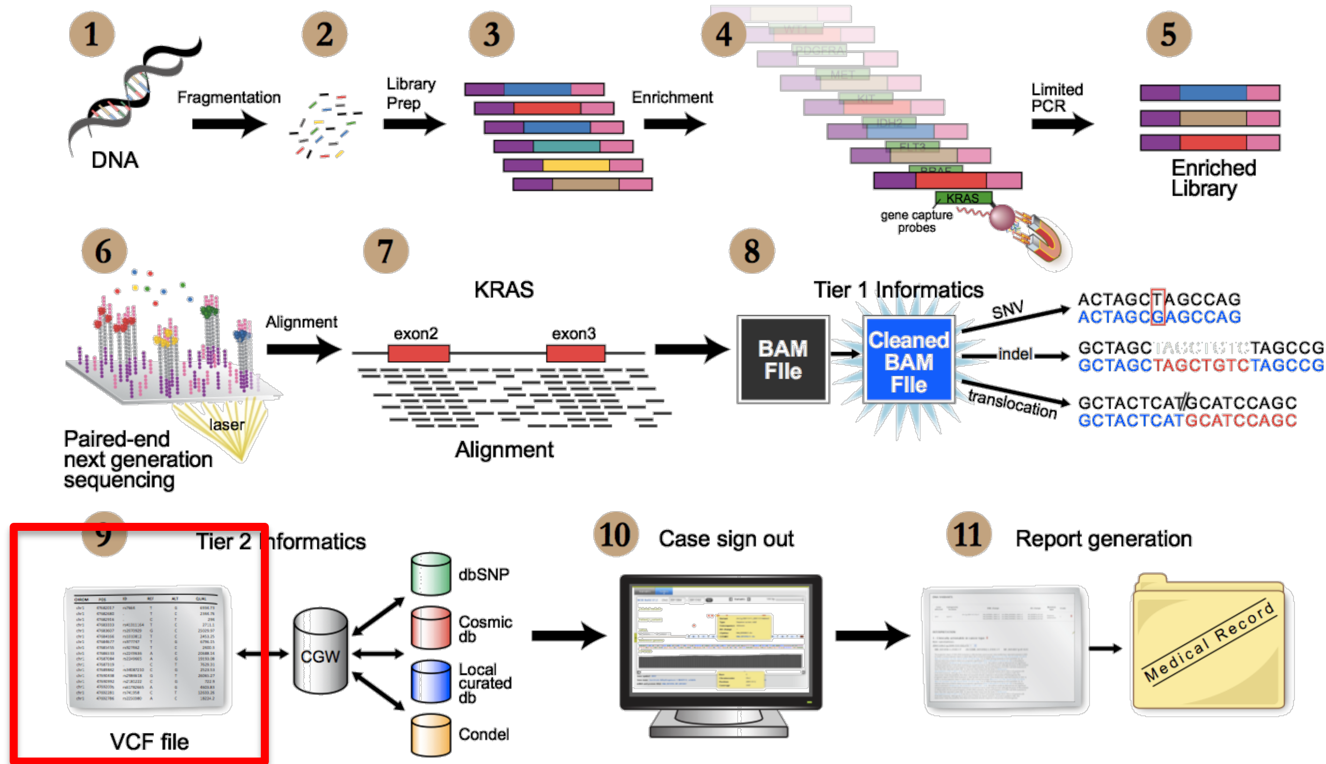
Copy Number Alterations



Translocations



Variant Calls



Key Concept

NGS coverage

- What's the advantage of high coverage?
  - Gives confidence in the variants identified
  - Makes it possible to detect tumor mutations despite admixture of nontumor tissue
  - Makes it possible to detect variants in subclones of tumor cells

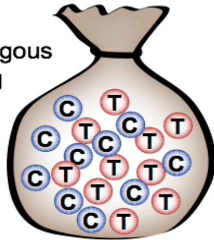
Coverage and Sensitivity

- To a large extent coverage determines the sensitivity of an NGS assay

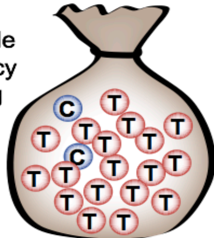
Sampling of Alleles

True sequence: ACTAGA

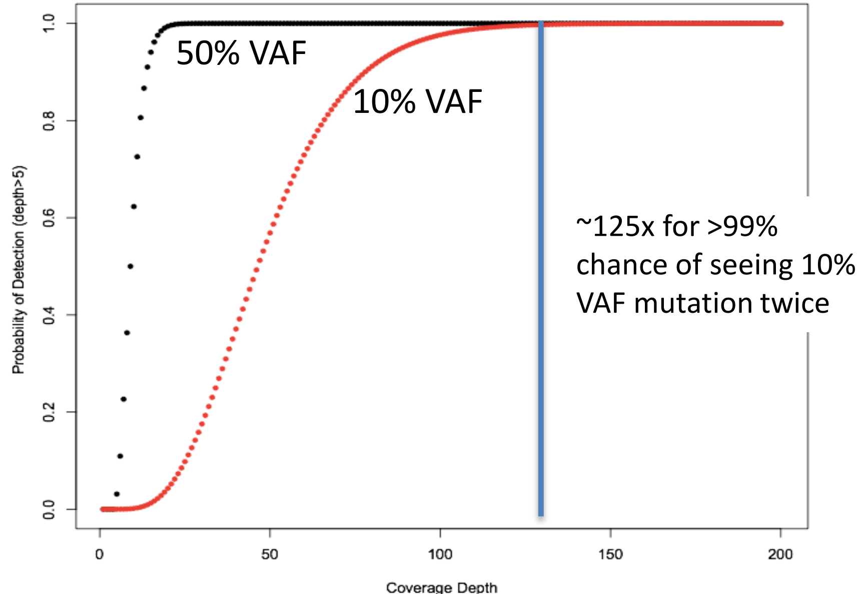
Heterozygous Sampling

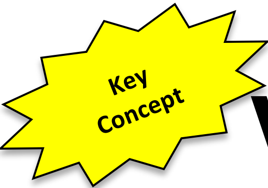


10% Allele Frequency Sampling



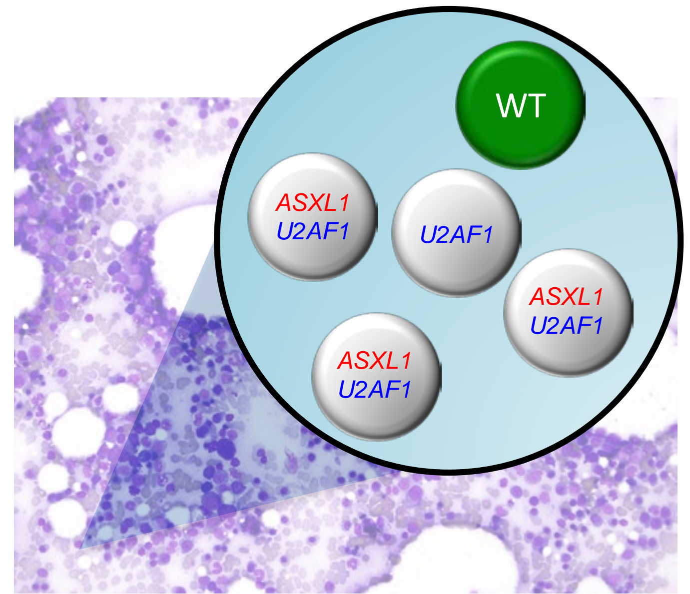
Minimum Predicted Coverage Required to Call Variants





Variant Allele Fraction (VAF)

U2AF1 Reference Sequence	*ASXL1* Reference Sequence
TGTGCAACCGAGAGCACCTG	CTGCCTACTACAGAGGGCTA
TGTGCAACCGAG	CTGCCTACTAC
GTGCAACCGAGA	TGCCTACTACA
TGCAACCGAGAG	GCCTACTACAG
CAACCGAGAGC	CCTACTACAGA
AACCGAGAGCA	CTACTACAGAG
ACCGAGAGCAC	TACTACAGAGG
Mutant: CCGAAAGCACC	ACTACAGAGGG
Mutant: CGAAAGCACCT	Mutant: CTATAGAGGGC
Mutant: GAAAGCACCTG	Mutant: TATAGAGGGCT
Mutant: AAAGCACCTGT	Mutant: ATAGAGGGCTA
VAF=4/(4+6)=40%	VAF=3/(3+7)=30%
Heterozygous mutation in 80% of cells	**60% of cells**



VAF=**V**ariant **A**llele **F**raction=variant reads/total reads

Jacoby et al, Expert Reviews, 2016

Myeloid Sequencing Panels

Sequencing for Measurable Residual Disease

Sequencing Based Karyotyping

APPLICATIONS

Application #1

Sequencing Panels for Myeloid Malignancy

- The first AML genome was reported in 2009 for an approximate cost of $2M
- Identified 12 coding region mutations including *NRAS*, *NPM1*, and *IDH1*
- In less than 10 years clinical sequencing is commonly performed on MDS and AML
- **Small sequencing panels work great for myeloid malignancies due to the small number of recurrently mutated genes**

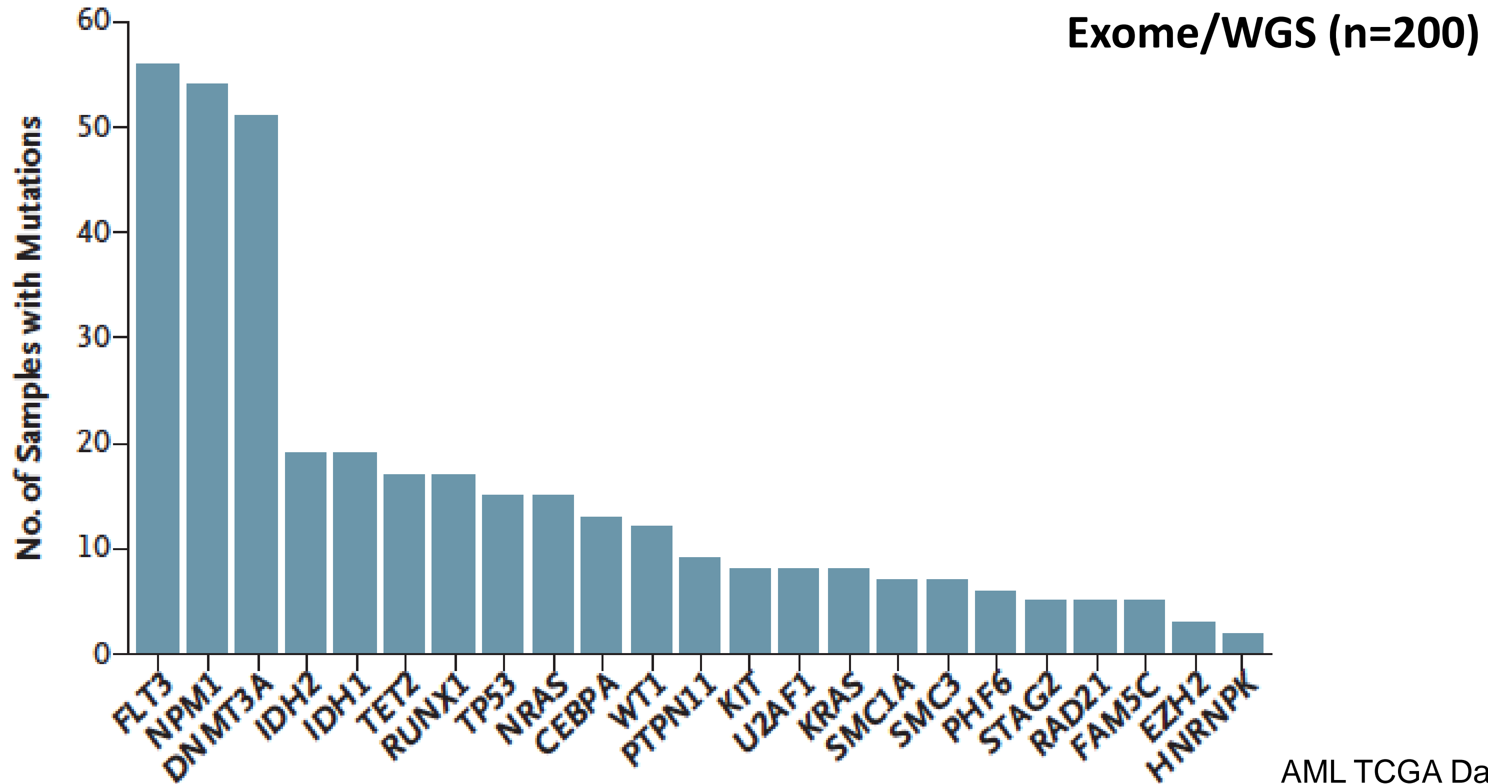
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome

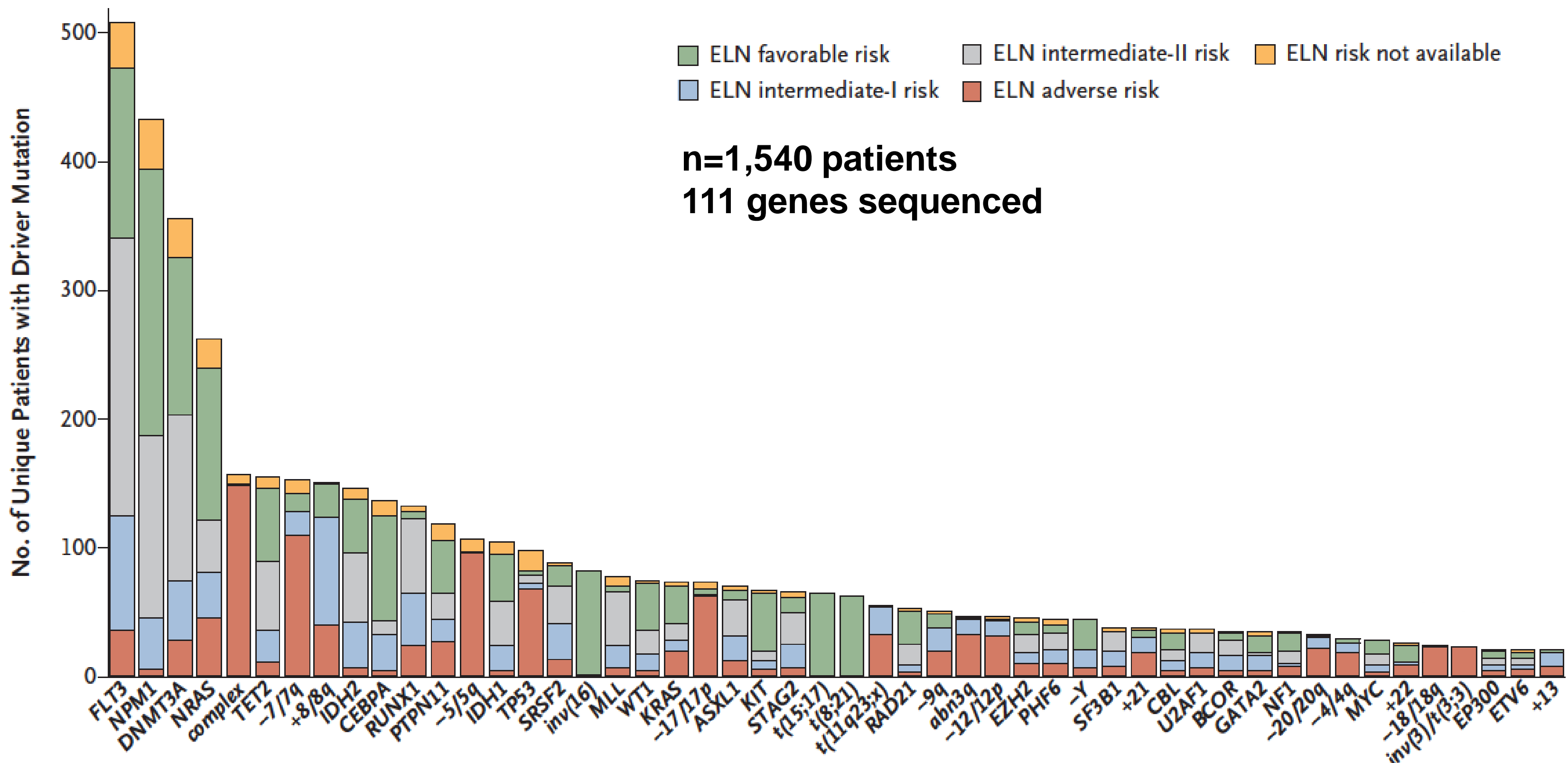
Elaine R. Mardis, Ph.D., Li Ding, Ph.D., David J. Dooling, Ph.D., David E. Larson, Ph.D., Michael D. McLellan, B.S., Ken Chen, Ph.D., Daniel C. Koboldt, M.S., Robert S. Fulton, M.S., Kim D. Delehaunty, B.A., Sean D. McGrath, M.S., Lucinda A. Fulton, M.S., Devin P. Locke, Ph.D., Vincent J. Magrini, Ph.D., Rachel M. Abbott, B.S., Tammi L. Vickery, B.S., Jerry S. Reed, M.S., Jody S. Robinson, M.S., Todd Wylie, B.S., Scott M. Smith, Lynn Carmichael, B.S., James M. Eldred, Christopher C. Harris, B.S., Jason Walker, B.A., B.S., Joshua B. Peck, M.B.A., Feiyu Du, M.S., Adam F. Dukes, B.A., Gabriel E. Sanderson, B.S., Anthony M. Brummett, Eric Clark, Joshua F. McMichael, B.S., Rick J. Meyer, M.S., Jonathan K. Schindler, B.S., B.A., Craig S. Pohl, M.S., John W. Wallis, Ph.D., Xiaoqi Shi, M.S., Ling Lin, M.S., Heather Schmidt, B.S., Yuzhu Tang, M.D., Carrie Haipek, M.S., Madeline E. Wiechert, M.S., Jolynda V. Ivy, M.B.A., Joelle Kalicki, B.S., Glendoria Elliott, Rhonda E. Ries, M.A., Jacqueline E. Payton, M.D., Ph.D., Peter Westervelt, M.D., Ph.D., Michael H. Tomasson, M.D., Mark A. Watson, M.D., Ph.D., Jack Baty, B.A., Sharon Heath, William D. Shannon, Ph.D., Rakesh Nagarajan, M.D., Ph.D., Daniel C. Link, M.D., Matthew J. Walter, M.D., Timothy A. Graubert, M.D., John F. DiPersio, M.D., Ph.D., Richard K. Wilson, Ph.D., and Timothy J. Ley, M.D.

Limited Number of Recurrent Mutations in AML



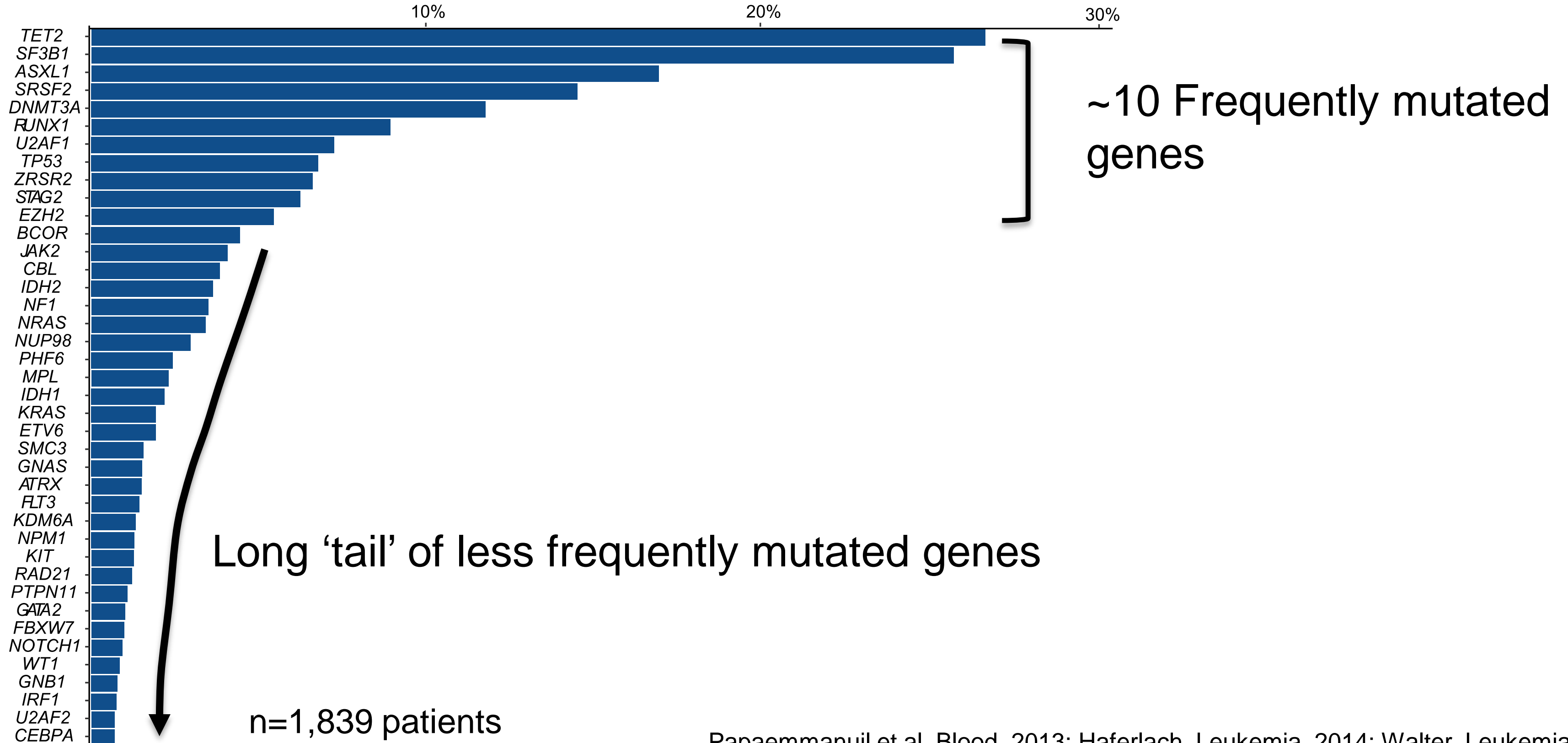
AML TCGA Data, NEJM 2013

Limited Number of Recurrent Mutations in AML



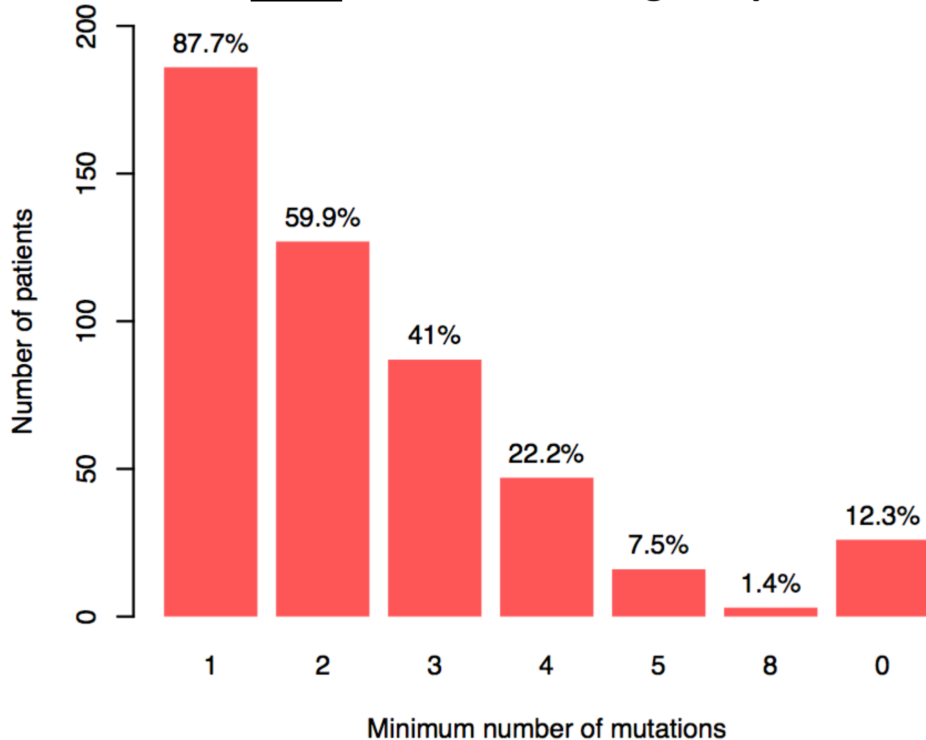
Papaemmanuil et al, NEJM 2016

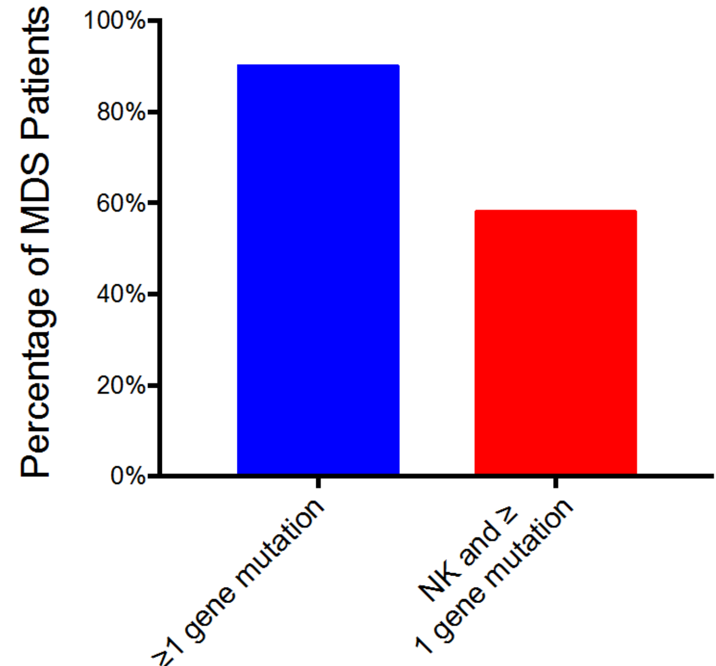
Limited Number of Recurrent Mutations in MDS



Papaemmanuil et al, Blood, 2013; Haferlach, Leukemia, 2014; Walter, Leukemia, 2013

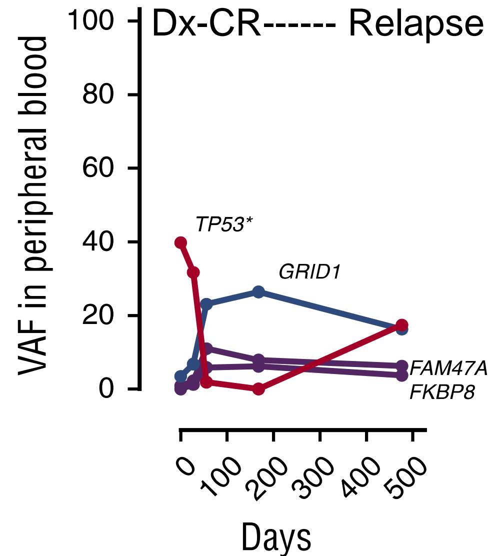
Small Diagnostic Sequencing Panels are Ideally Suited to AML and MDS Evaluation

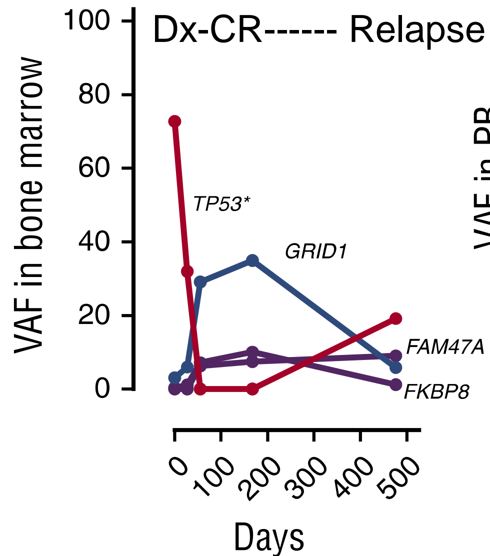


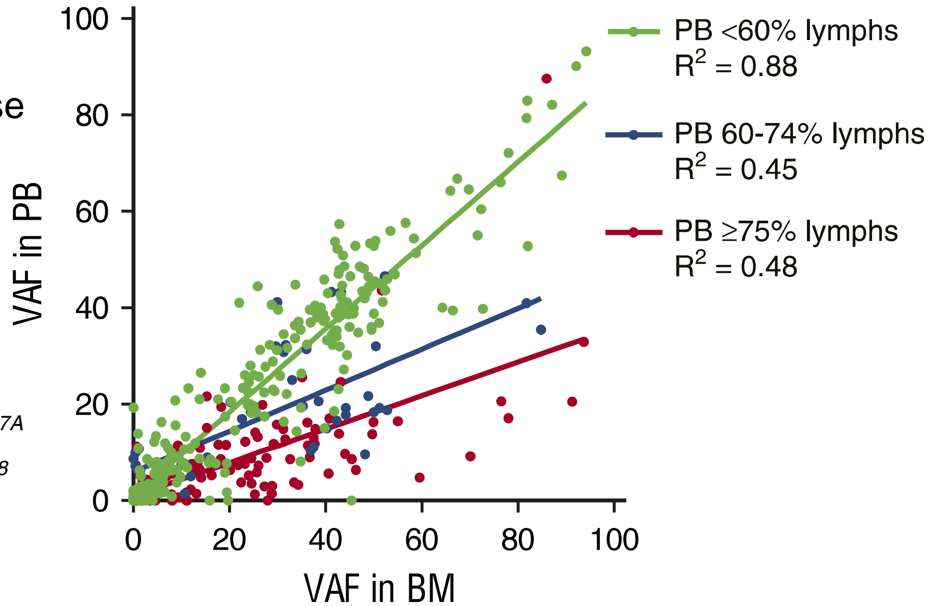


Adapted from Haferlach, Leukemia, 2014

Sequencing Blood is Equivalent to Marrow







Duncavage et al, Blood, 2017

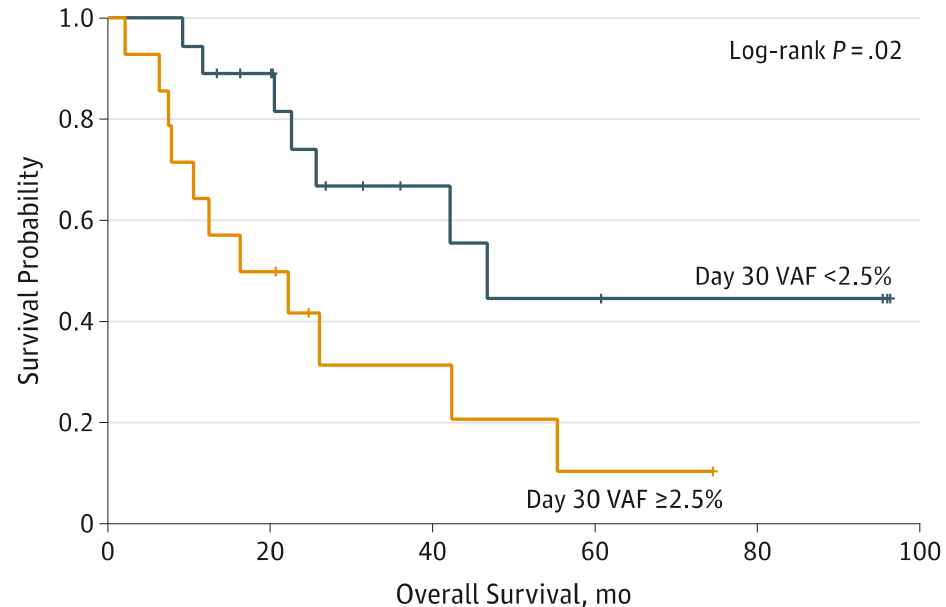
Clinical Sequencing Panels

Metric	Range	Consideration
Sequencing Coverage	500x to > 2,000x	Higher coverage result in high sensitivities
VAF Sensitivity	3-10% VAF	Sensitivity depends on coverage and analysis approach
Number of genes	7-400+ genes	More genes is not necessarily better
Spectrum of reported mutations	SNV and indels to translocation and CNAs	Some labs my report rearrangements and CNAs [e.g. del(5), del(7)] or translocations from NGS panel data
Cost	$1,000-$3,000	Generally reimbursed by private insurance and Medicare
Turn around Time	~2 weeks	May be as long as 4 weeks

Application #2

Measurable Residual Disease

- Molecular risk stratification systems in AML rely on the prognostic significance of recurrent mutations identified at initial diagnosis
- In ‘tumor clearance’ methods of risk stratification AML is sequencing at diagnosis and again 30 days after 7+3 induction to determine if mutations have been cleared
- Patients in which mutations (VAF >2.5%) could be detected 30 days after treatment almost all died within 5 years
- **Doesn't matter what the mutations are—if they are detectable after treatment it is a bad prognostic indicator**



Klco et al, JAMA, 2015

Error Corrected UMI-based Methods

Genomic DNA

PCR & sequencing errors

T

T

True mutation

33% VAF

Barcoded DNA Library

Barcode Families

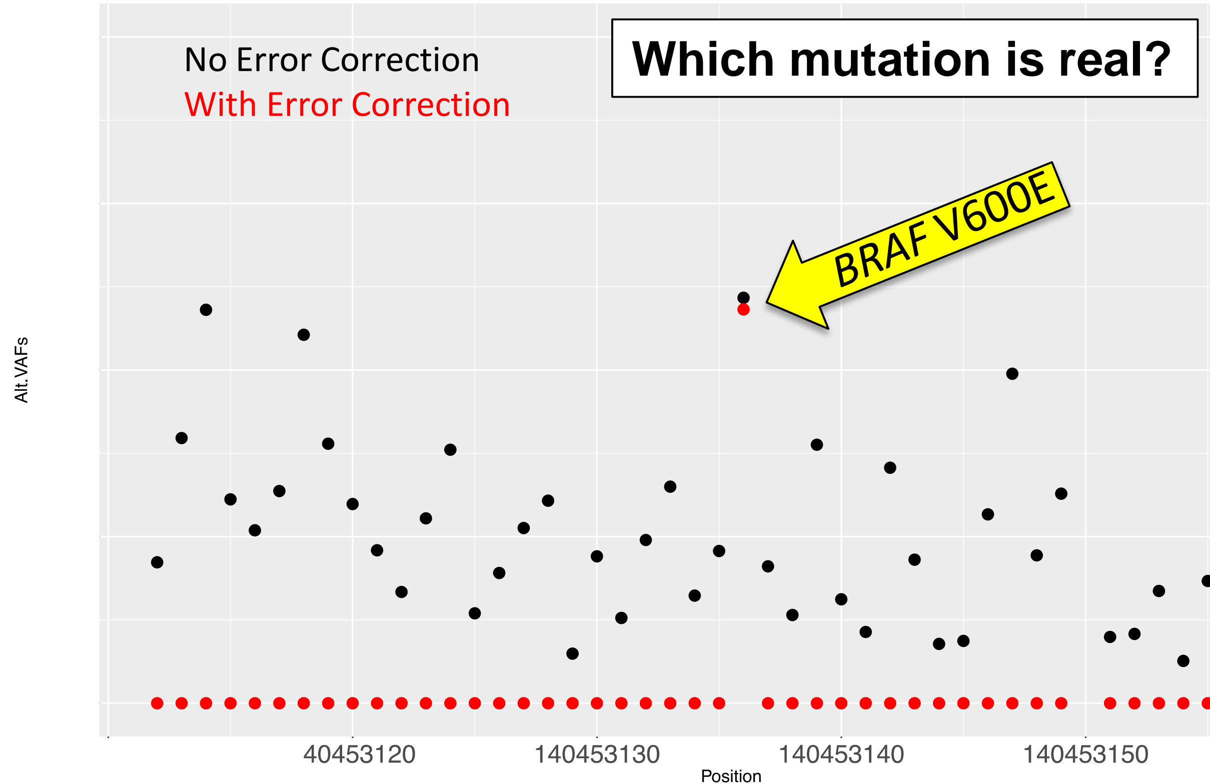
Collapse Families

Variant Calling

VAF, Variant Allele Fraction

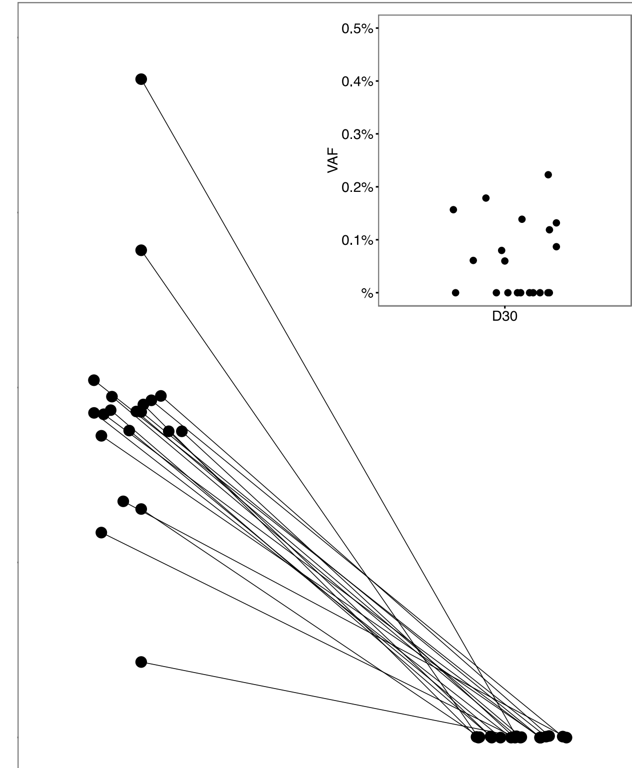
UMI Based Error Correction for Ultra Sensitive Sequencing

- *BRAF* V600E mutated cell line diluted to a VAF of 1.3%
- Sequenced to 10,000x coverage using amplicon-based enrichment
- High background noise below 2% VAF



Sequencing Based MRD

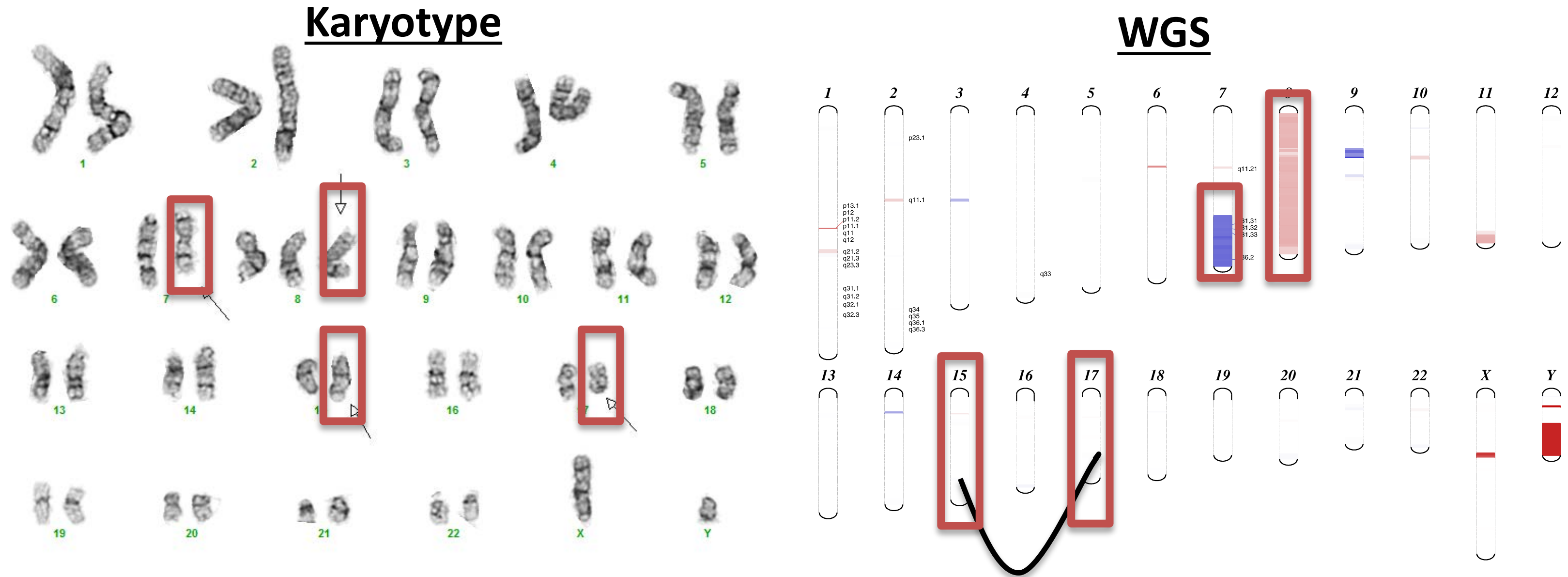
- We can now use error corrected sequencing methods (UMIs) to find low level MRD representing one mutated cell in 1,000 normal cells.



Application #3

Sequencing Based Karyotyping

- WGS can be used to recapitulate cytogenetic findings including gains/losses and rearrangements
- Cost of WGS continues to fall—currently $1000



Acknowledgements

Patients

McDonnell Genome Institute

Richard Wilson
Elaine Mardis
Li Ding
Robert Fulton
Dave Larson
Chris Miller
Haley Abel
Gue Su Chang
Vince Magrini
Heather Schmidt
Joelle Kalicki-Veizer
Michelle O'Laughlin

WU Oncology Division

Matt Walter
Tim Ley
John DiPersio
Dan Link
John Welch
Peter Westervelt
Meagan Jacoby
Geoff Uy
Jin Shao, Josh Robinson
David Spencer

Funding

NCI R33 CA217700-01
Edward P. Evans Foundation
SPORE in Myeloid Malignancies Career Enhancement Award

The Elizabeth H.
and James S. McDonnell III

Washington University

Washington University in St. Louis