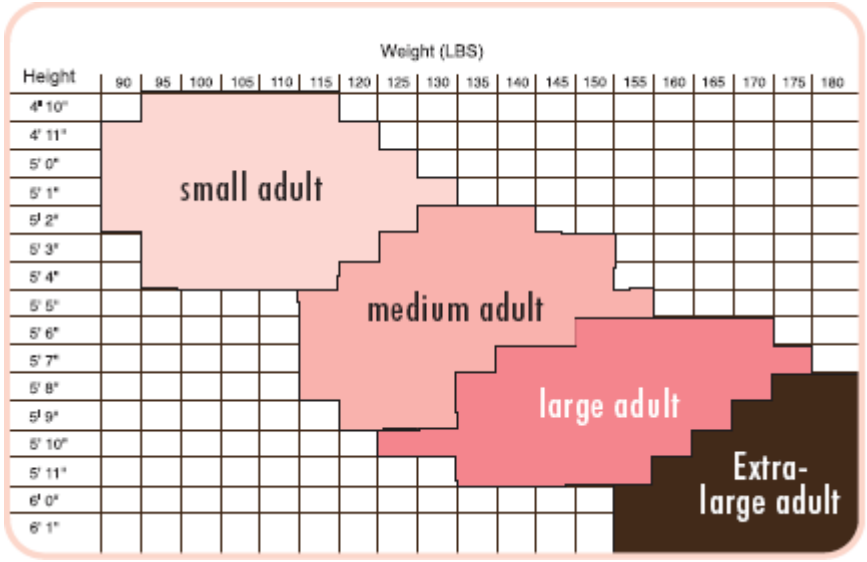
Clinical-Grade Next-Generation Sequencing: Assay Design & Validation

Richard D. Press, MD, PhD
Dept of Pathology
Knight Cancer Institute
Knight Diagnostic Labs
Oregon Health & Science University
Portland, OR

NGS Assay Design Considerations Customization is Key

- Coverage: medium, large, or X-large
- If not whole exome / genome, then:
  - Which genes to sequence?
  - Which regions?
- What cancers to cover?
  - Organ-specific?
  - Carcinoma vs sarcoma vs hematopoeitic?
  - All cancers?

What size are you?



Assay Validation Questions
(I don't have the answers)

- How to validate all known mutation possibilities?
- Tumor cellularity requirements?
- Minimal mutant allele burden detection?
- Quantitative or Qualitative reporting?
- Quality control: read depth? Other parameters?
- Unknown variants?
- Many other questions and variables

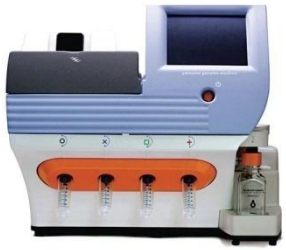
Whatôs All the Fuss About NGS?

- **Broader coverage, including tumor suppressors**
- **Better sensitivity, through deeper reads**
- **? Lower costs, through multiplexing**
- **More comprehensive cancer genome characterization for targeted therapeutic (and diagnostic) discovery**

Single gene assays | Multiplexed hotspots | Multigene panels | Whole exome | Whole genome

ABI

Sequenom MassArray

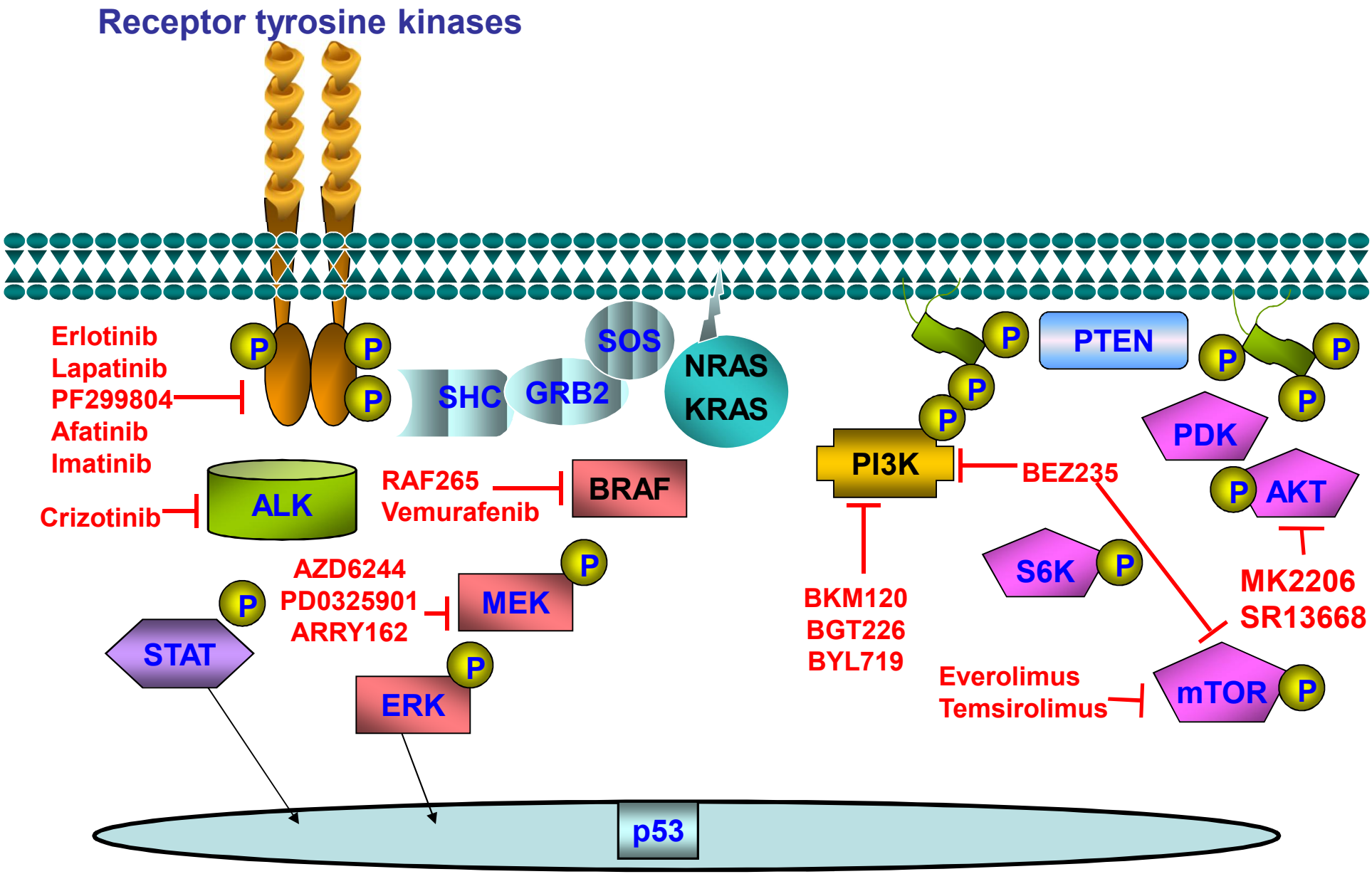


Ion Torrent PGM

454

Illumina HiSeq

Targeted Therapeutics



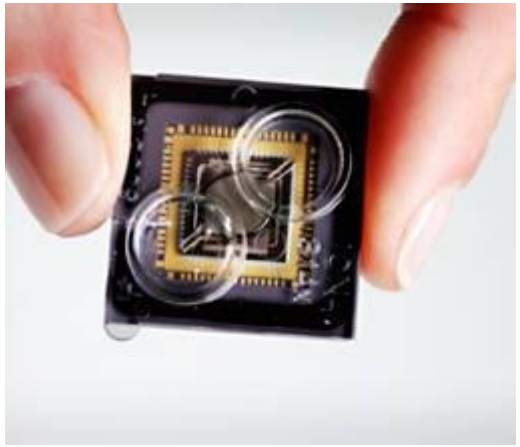
Coverage:
How High Do You Want to Go?

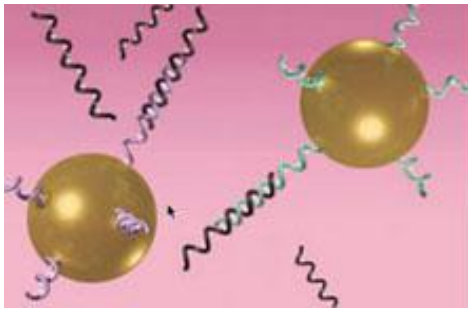
- Mutation Hotspots
- Disease-Specific Gene Panels
- Generic "Cancer" Gene Panels
- Whole Exome
- Whole Genome

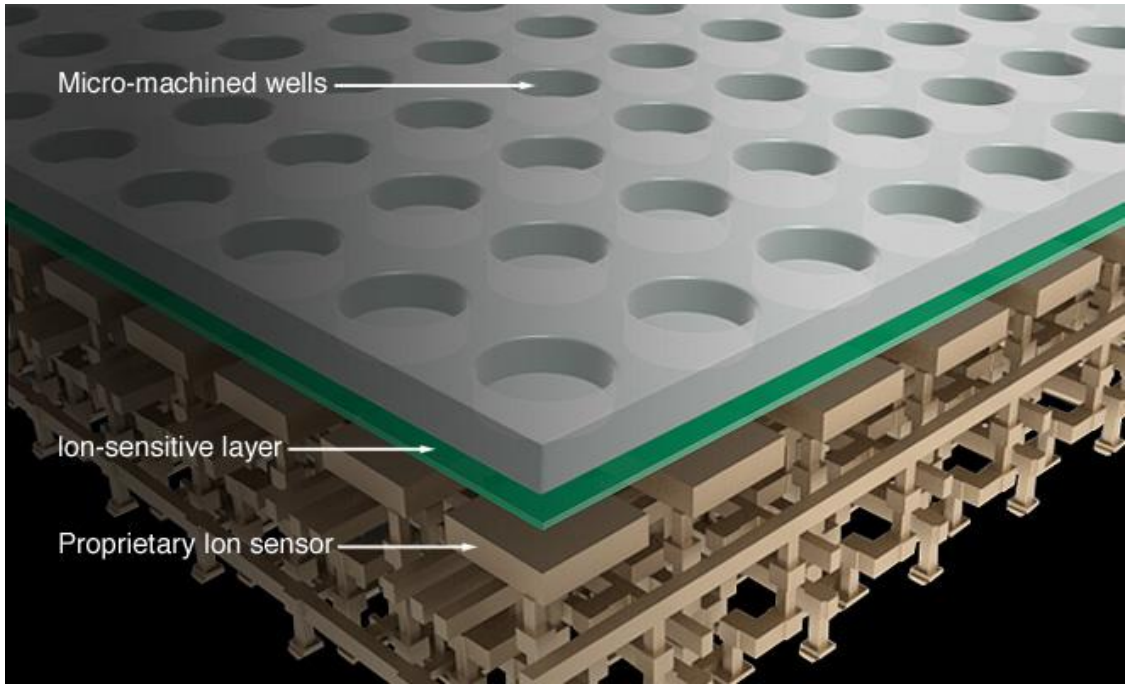
Higher Coverage means:

- More complexity & cost
- More unknown variants
- ? Overkill for clinical care: who cares if its not drugable?

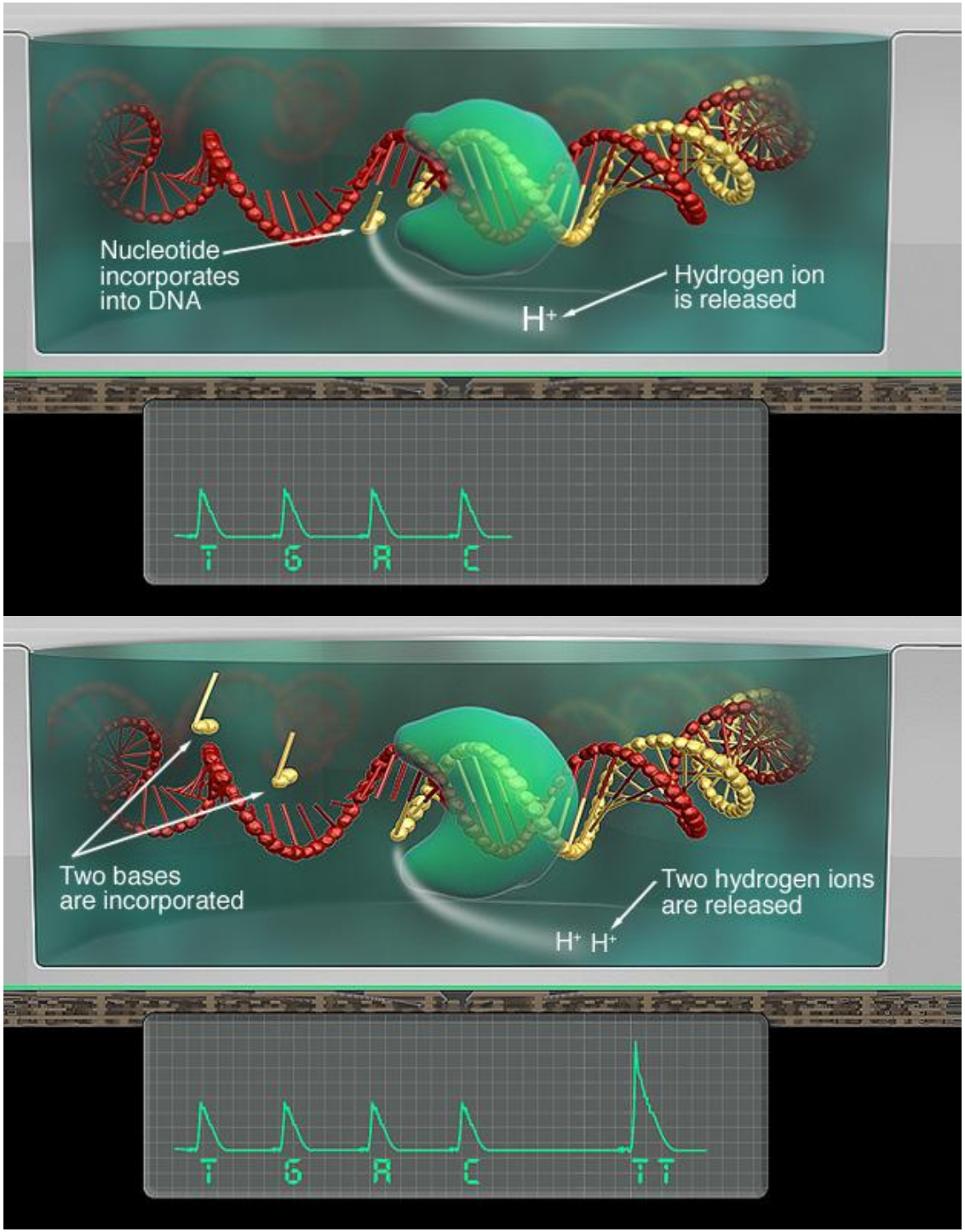
Ion Torrent PGM





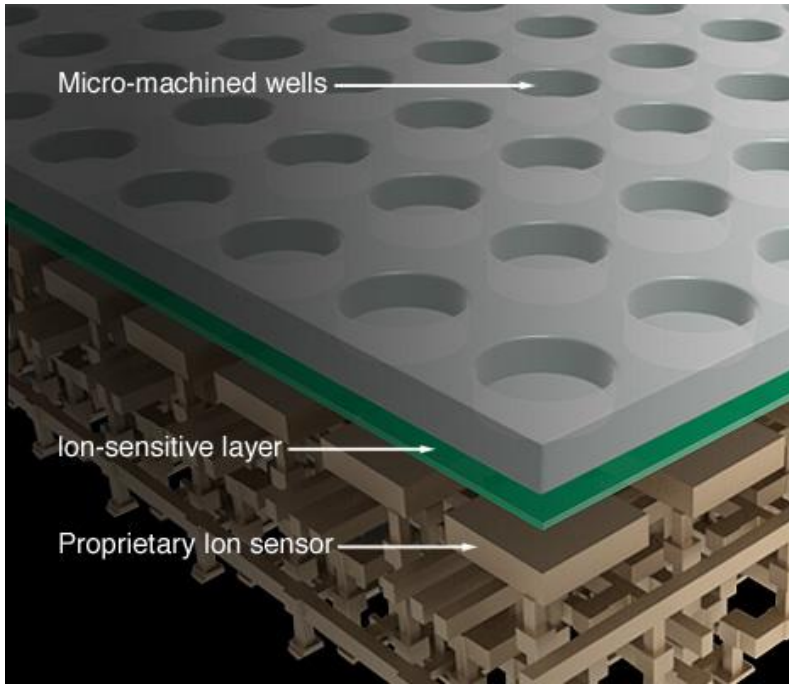


- Moderate throughput
- Massively parallel 3rd generation sequencing
- Performed on a semi-conductor chip



Ion Torrent PGM

Each well is a tiny solid-state pH meter



Library Preparation: Hybridization-Capture Approaches

Nimblegen
Agilent
Ion Torrent

50 to 1,000 ng Genomic DNA

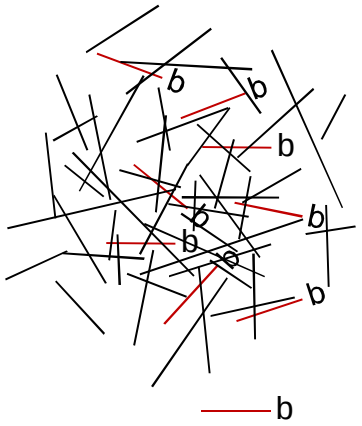
↓

Shear, End Repair, Ligate adapters

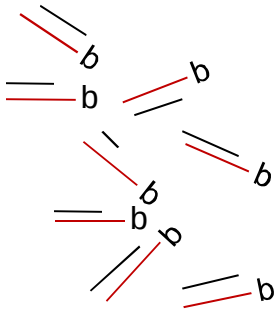
↓

Gel purify at ~180 bp

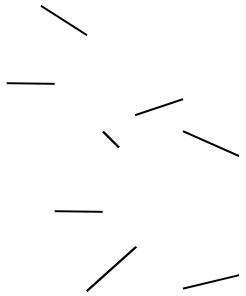
Add biotinylated probes to genes/exons of interest



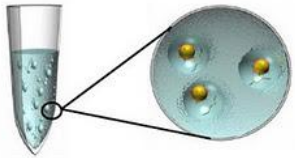
Purify hybridized RNA probes with magnetic beads



Treat with NaOH to remove RNA

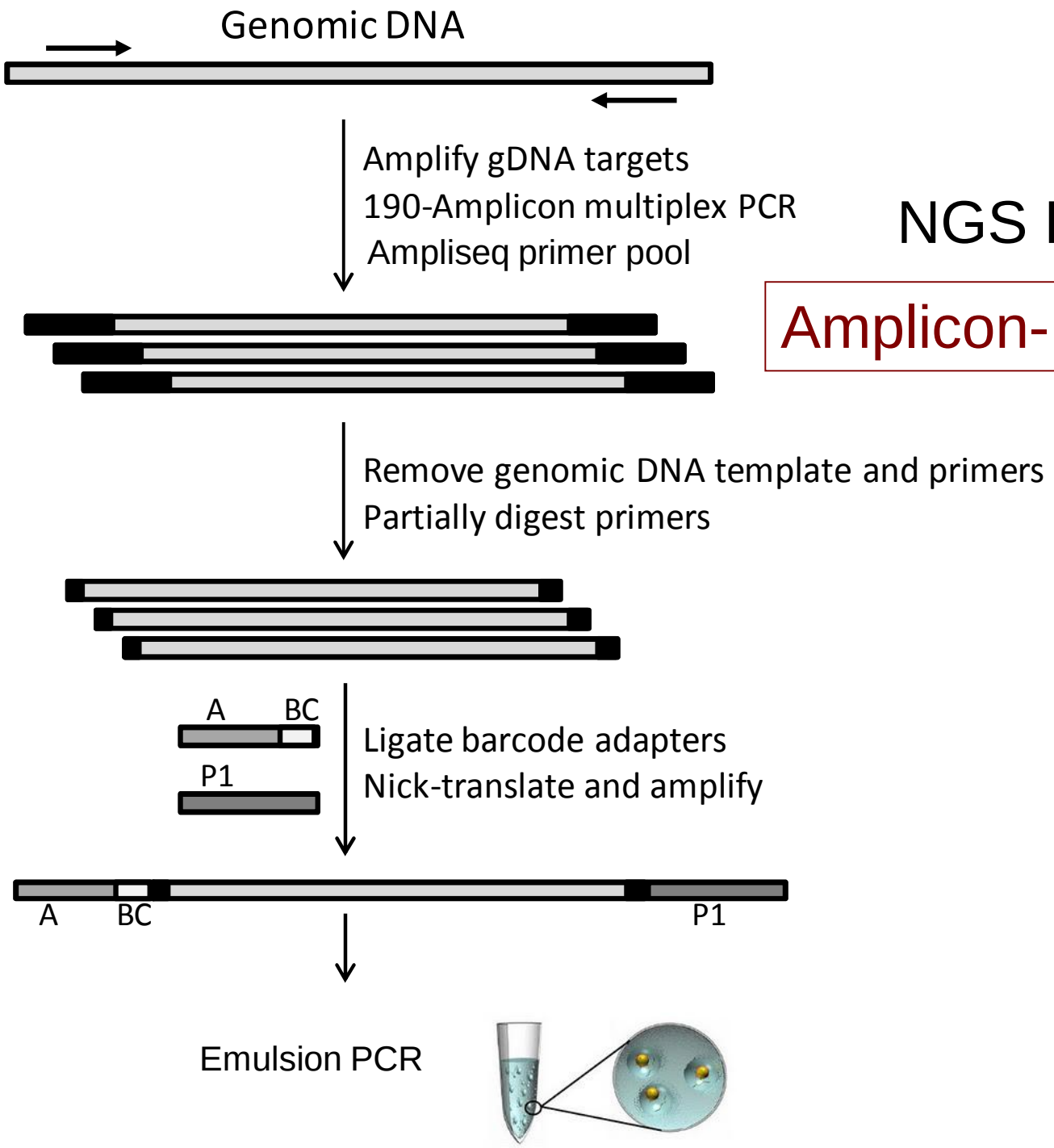


Emulsion PCR



NGS Library Prep

Amplicon-Based Approach



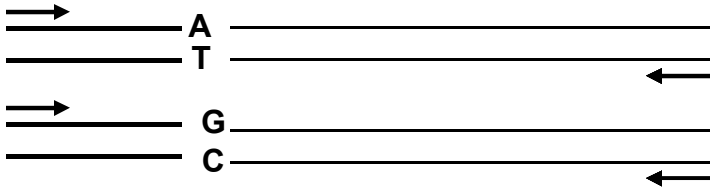
AmpliSeq Cancer Panel

- **739 hotspots covered by 190 amplicons**
- **Single tube amplification**
- **Average amplicon length: 119 bp (100-169 bp)**
- **Input DNA: 10 ng (Fresh or FFPE)**
- **Turn-around time: 48 hours**
- **46 genes:**

  ABL1, AKT1, ALK, APC, ATM, BRAF, CDH1, CDKN2A, CSF1R, CTNNB1, EGFR, ERBB2, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, FLT3, GNAS, HNF1A, HRAS, IDH1, JAK2, JAK3, KDR, KIT, KRAS, MET, MLH1, MPL, NOTCH1, NPM1, NRAS, PDGFRA, PIK3CA, PTEN, PTPN11, RB1, RET, SMAD4, SMARCB1, SMO, SRC, STK11, TP53, VHL

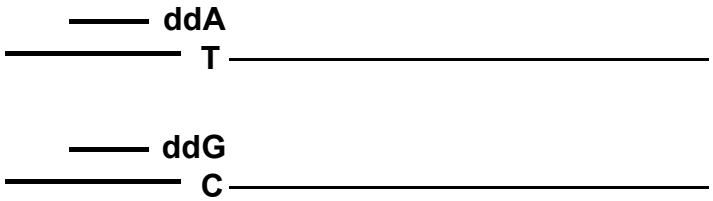
Mass Spectrometry-Based Detection of Genomic Mutations

PCR targets of interest
~ 100 bp amplicons



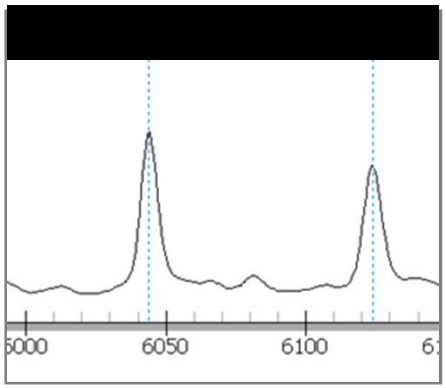
Clean-up steps

Primer extension reaction

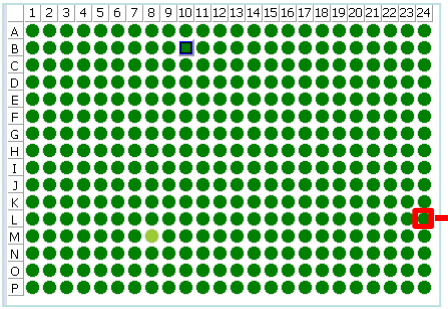


De-salting step

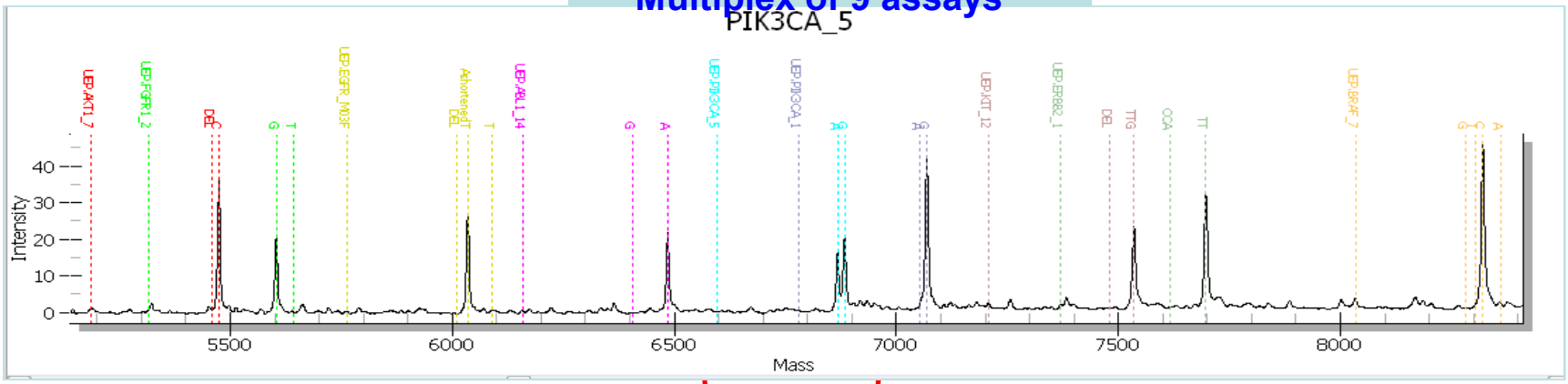
MALDI-TOF Mass spectrometry



Mass spec profile for a 9-plex reaction



Multiplex of 9 assays

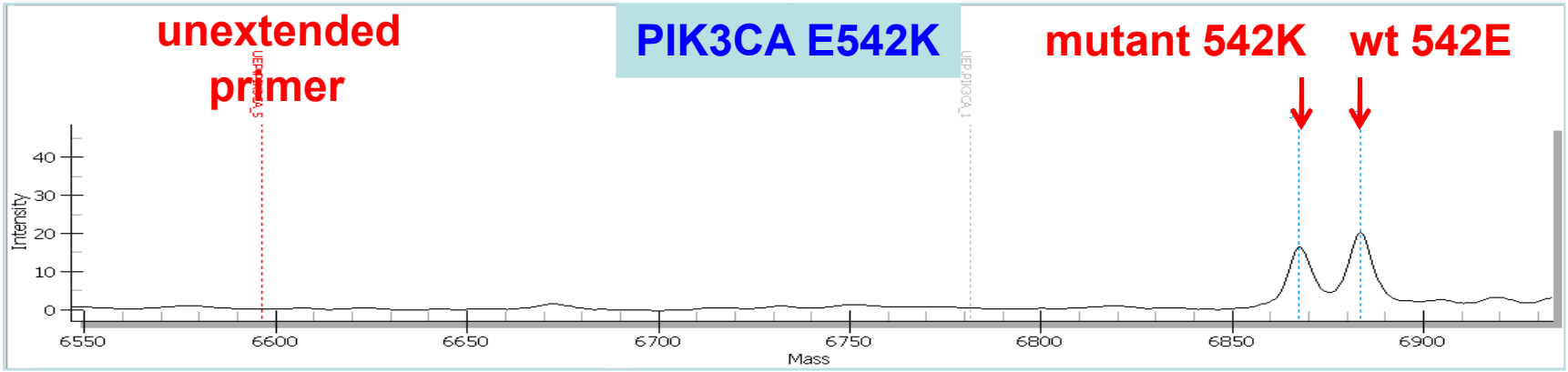


unextended primer

PIK3CA E542K

mutant 542K

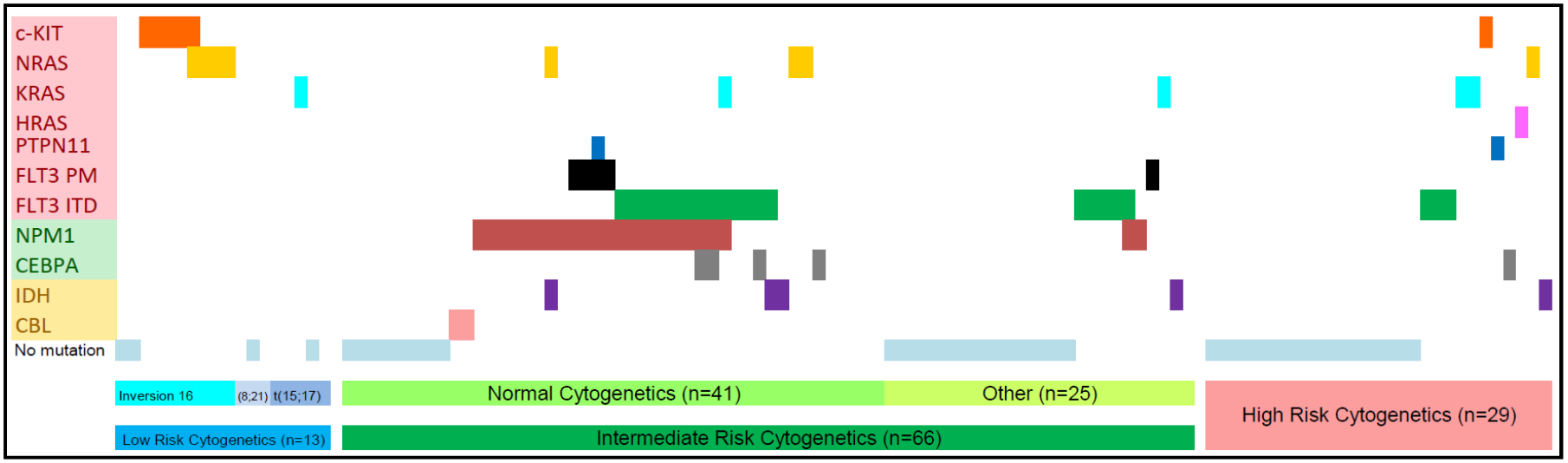
wt 542E



Mass Spec Leukemia Panel: 370 mutations / 31 genes

ABL
AKT1
AKT2
AKT3
BRAF
CBL
CBLB
FBXW7
FES
FGFR4
FLT3
FMS
GATA1
HRAS
IDH1
IDH2
JAK1
JAK2
JAK3
KIT
KRAS
MET
MPL
NOTCH1
NPM1
NRAS
NTRK1
PAX5
PDGFRB
PTPN11
SOS1

Mutation Spectrum in AML (108 OHSU cases)



Normal cytogenetics: 78% mutation frequency
Abnormal cytogenetics: 43% mutation frequency

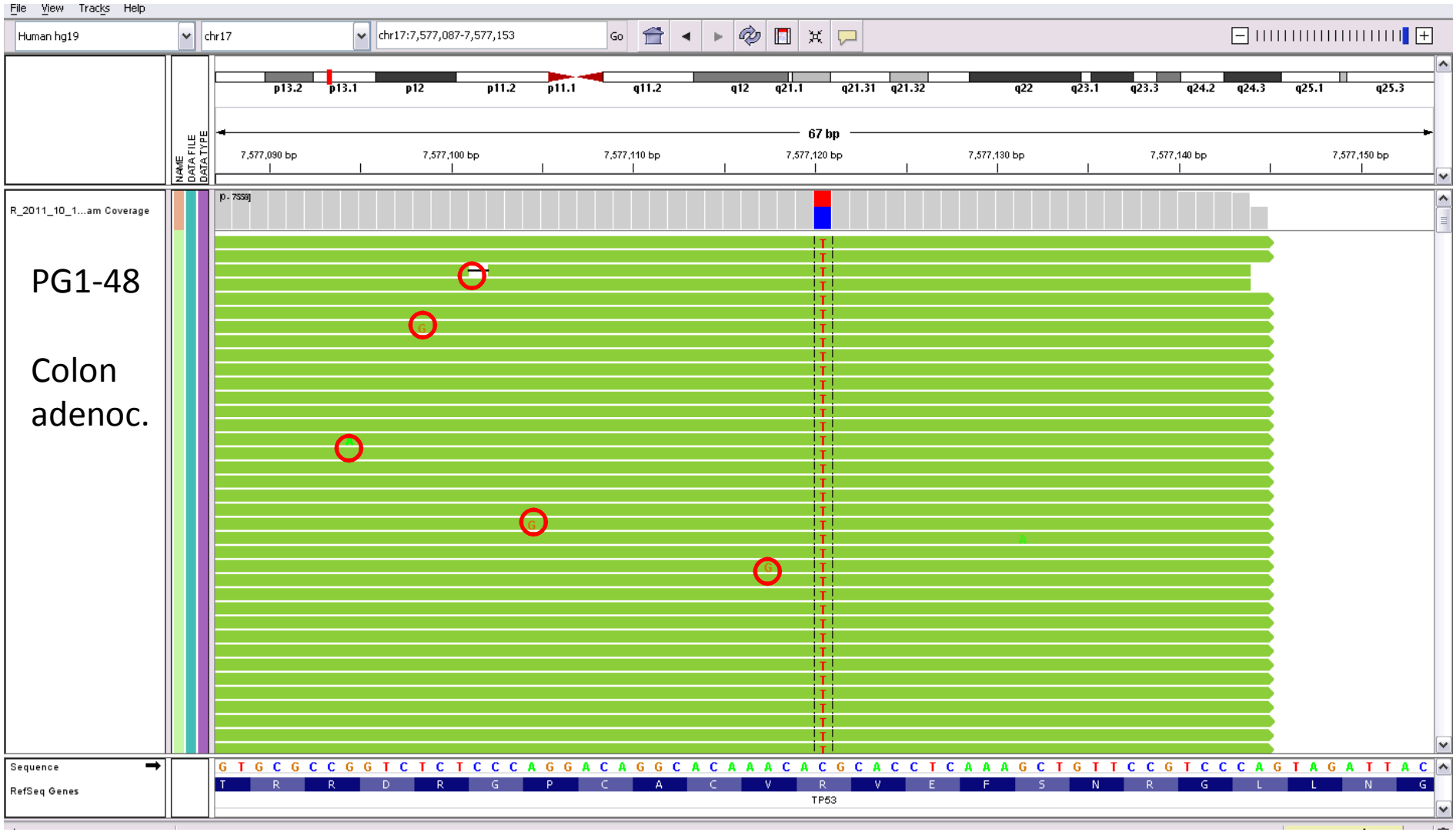
Mutation discovery with 31-gene mass spec panel plus single gene in-del assays (FLT3, CEBPA, KIT)

J Dunlap, in press

Ampliseq Validation Study

- 45 FFPE tumor DNA samples with known mutations previously quantitated on Sequenom MassArray (mass spectrometry)
  - 53 point mutations
  - 19 in/dels (range 4 - 63 bp)
- 7 unmatched FFPE normal tissue DNA
- 100 bp single-end sequencing runs
  - 22 samples run singly on 314 chips
  - All samples run as 4-plexes on 316 chips

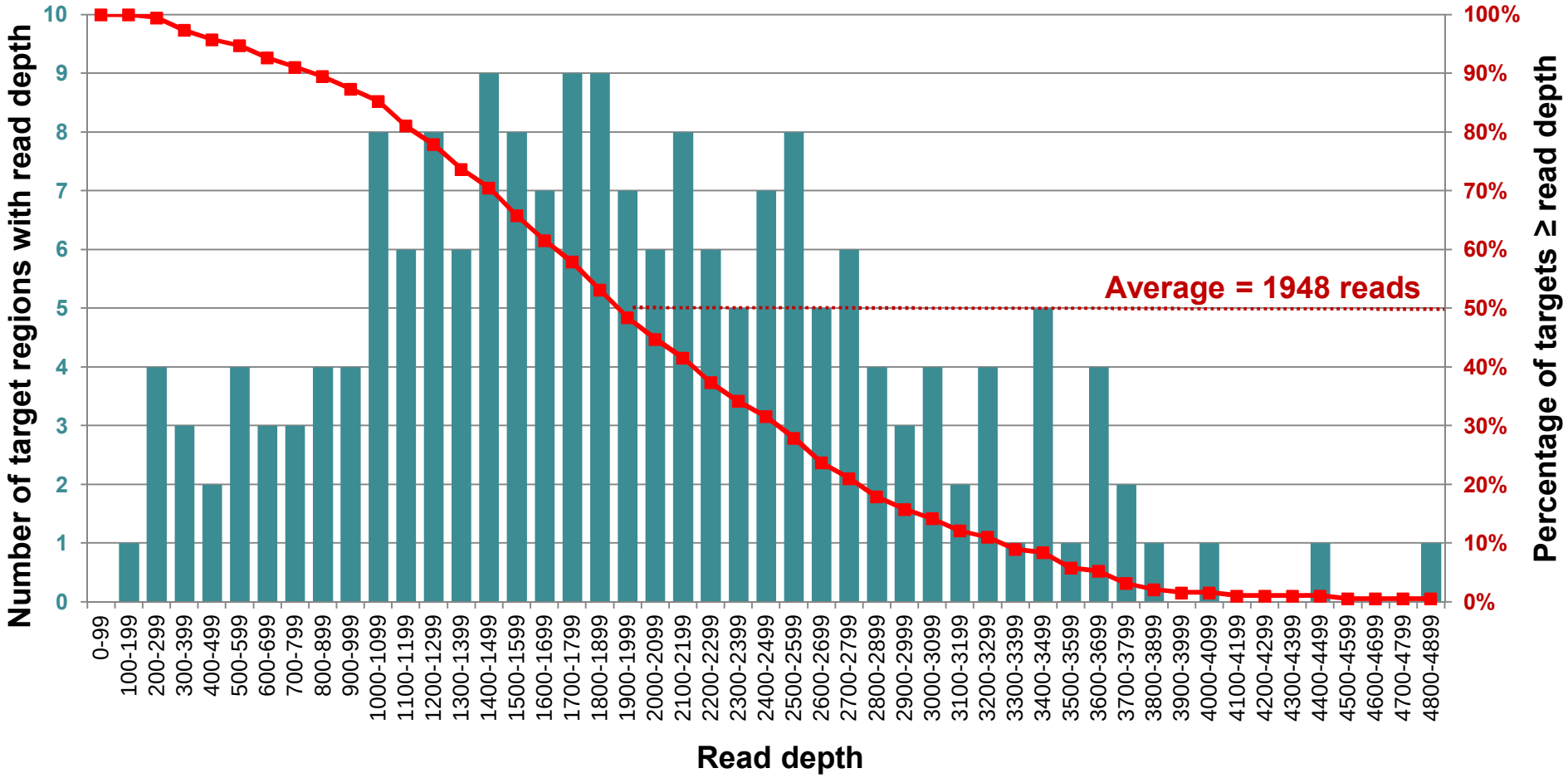
TP53 R273H



ÅGeneral background is low

ÅSome homopolymers (e.g. a run of Cs) lead to false positive calls

Read Coverage Distribution:
190 Amplicons in each of 45 Tumor Samples
(Normalized to 400,000 reads)



Å95% of amplicons average >400 reads

Å91% of amplicons >650 reads (estimated 5% sensitivity)

Å78% of amplicons with >1200 reads (estimated 1% sensitivity)

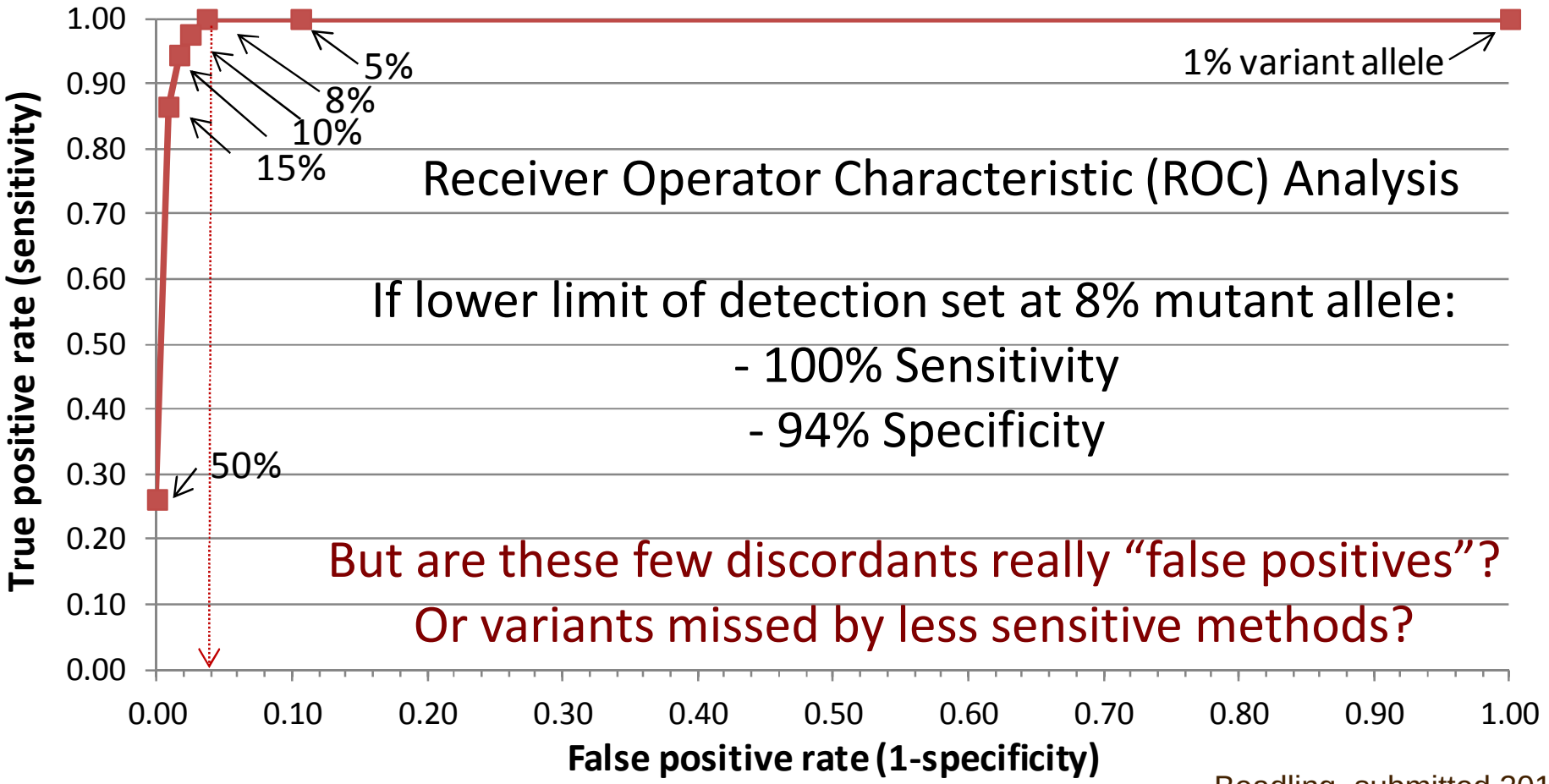
Beadling, submitted 2012

Sequencing Performance

- 4-plex samples run on 316 chips (6.2 million wells)
  - Avg 4 million beads loaded
  - Avg 3.7 million beads had library templates (92%)
  - Avg 1.7 million beads yielded quality sequence
- For individual samples
  - Avg 428,000 reads/sample (range 178K - 710K)
  - Mean read length 76 bp
  - On-target reads: >95%
- Across all samples (normalized to 400K reads)
  - Avg 1,941 reads per amplicon
  - 95% of amplicons with $\geq$ 400 reads

What is the Optimal Mutant Allele Burden Cutoff?

53 known point mutations in 45 tumor samples
All 53 “detected” by both NGS & Mass Spectrometry



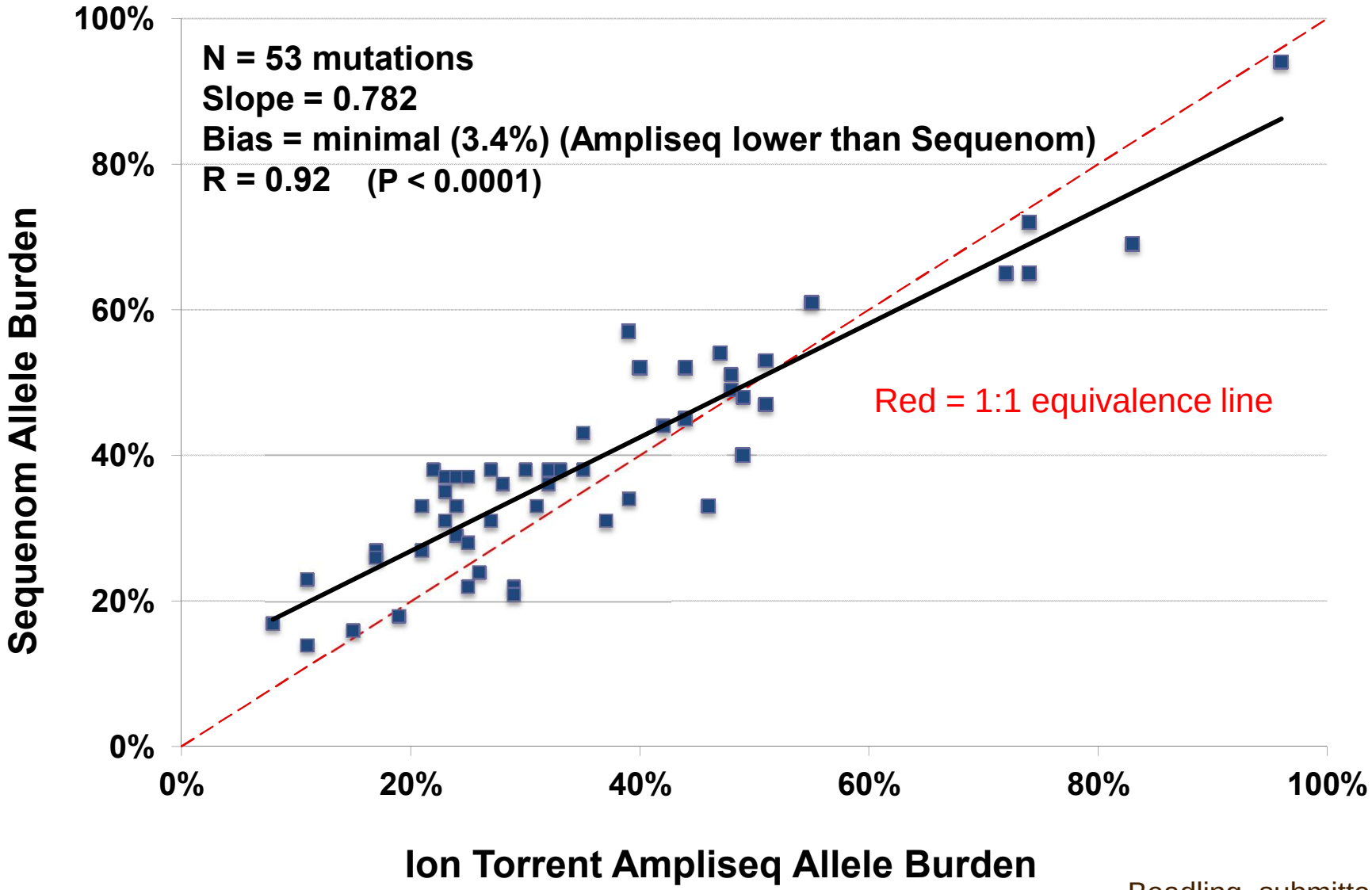
Beadling, submitted 2012

Summary of Variants

- All 53 known point mutations were identified by the variant caller software
- 26 new mutations were also identified
  - APC in colon ca; PTEN in endometrial ca; STK11 in lung ca
- 19 in/dels were included in the analysis; range: 4-63 bp
  - 2 called exactly
  - 5 flagged as point mutations
  - 12 visible on manual inspection but not flagged
- 54 óvariantsôturned up reproducibly in both tumor and normal DNA, likely reflecting sequencing aberrations

Is NGS Data Quantitative?

Allele Ratios: MassArray *vs* AmpliSeq



Beadling, submitted 2012

NGS Clinical Validation: Status Report (OHSU; May ‘12)

Ampliseq 46 gene panel (multiplex PCR library prep)

- Analytical validation essentially complete
- 100% sensitive compared to “gold standards”
- Variants of unknown function?
- Low-level variants?
  - We have validated an 8% mutant allele threshold
- Indels remain a challenge for variant caller
  - New software any better?
- Aiming for summer ‘12 launch in our clinical lab

The Future: Disease-Specific Gene Panels

Cancer Site	Target genes	# Exons	Kilobases	Ampli-cons	New Genes (not in Ampliseq)
Lung	23	224	34.3	502	9
Colon	16	157	31.8	405	6
Melanoma	21	113	13.9	231	9
AML / ALL / MDS	42	342	62.6	863	28

- Ion Torrent custom primer design software used to design primers for library prep
- Proprietary primer modifications allow massive multiplexing: 2-4 PCR reactions per library prep

Many Unanswered Questions

- How to re-validate assays given continuing rapid pace of improvements to chemistries, hardware, and software?
- Sanger sequencing confirmation?
- Unknown variants?
- Matched normal tissue required?
- Quality control?
- Gene patent implications?
- FDA? Can you image the approval process?
- Will anyone pay us for this service?

Acknowledgements

OHSU Knight Diagnostic Labs

Carol Beadling, PhD
Tanaya Neff
Andrea Warrick
Fei Yang
Jennifer Dunlap MD
Chris Corless MD, PhD

Oregon Health & Science
University