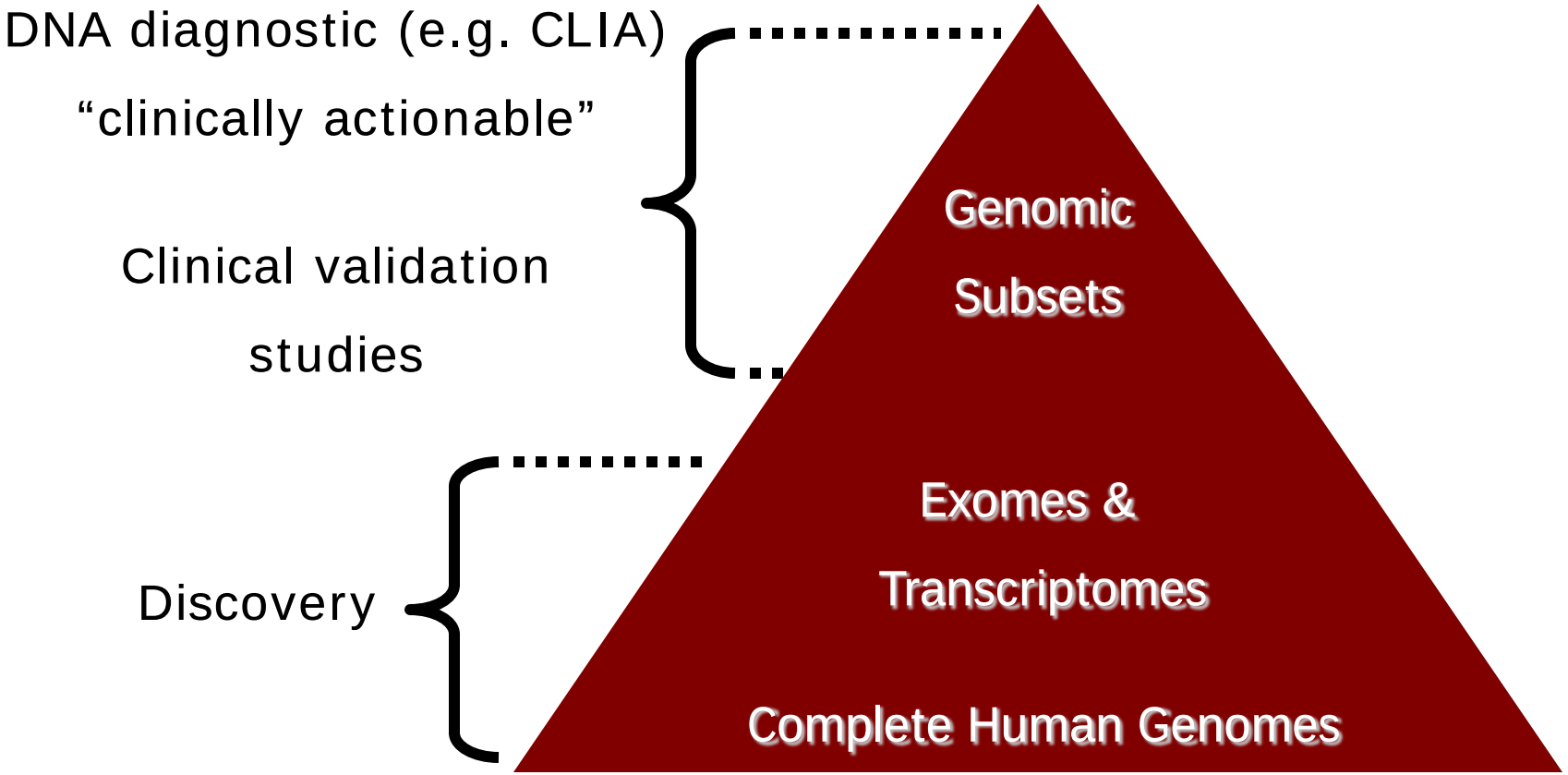
Division of Oncology
Stanford University School of Medicine
Stanford Genome Technology Center
Stanford Cancer Institute

Mass validation of variants identified by whole genome sequencing

Georges Natsoulis, PhD
Stanford University

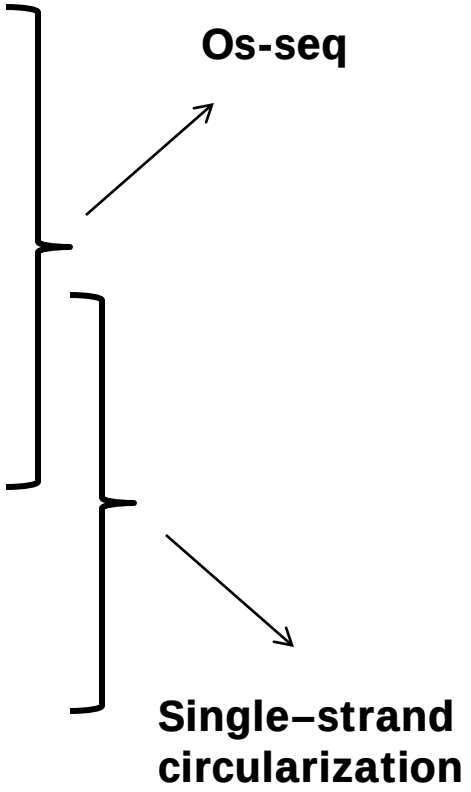
Cancer genome sequencing and personalized diagnostics



Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

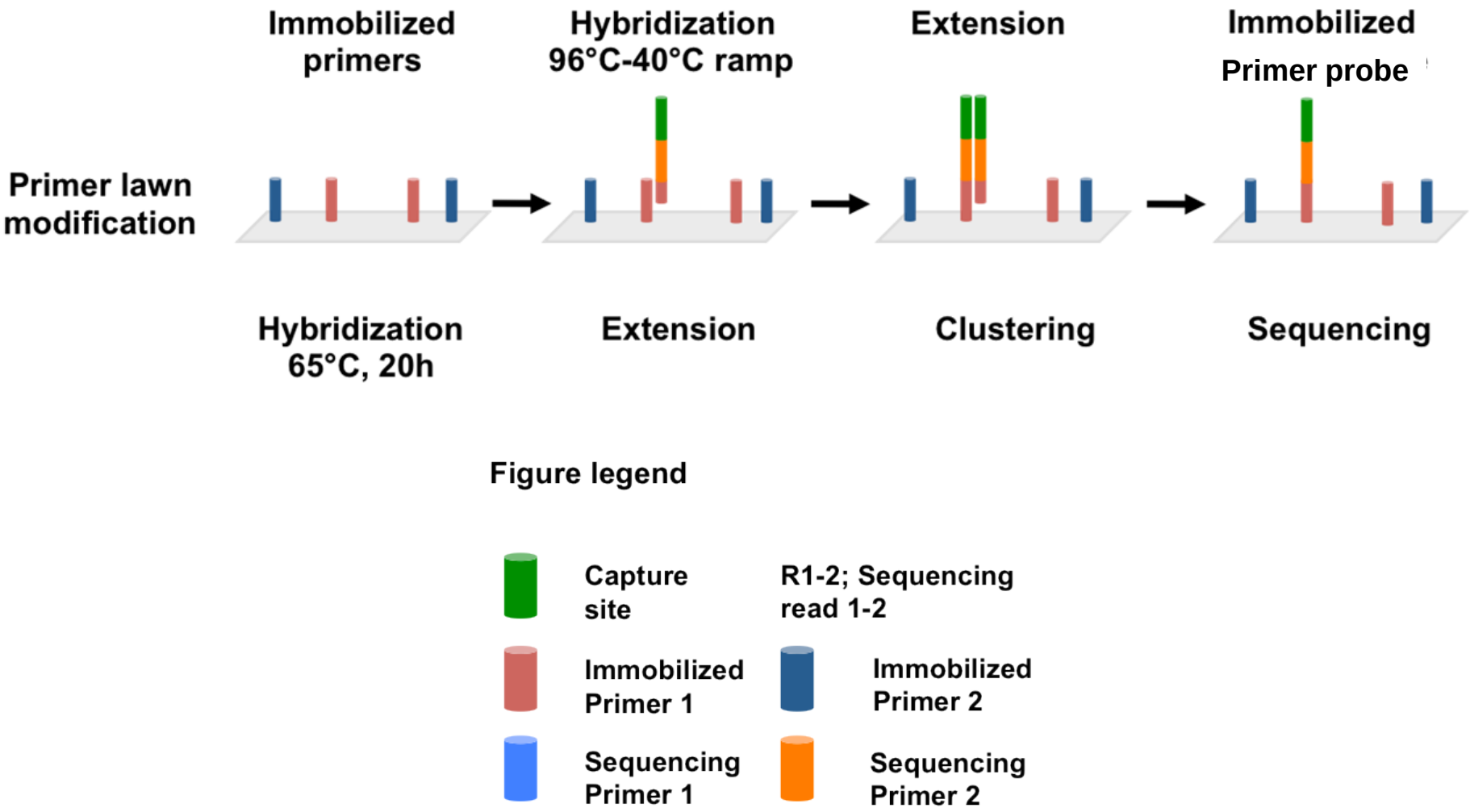
Two methods addressing multiple objectives

Objective	Advances
Whole genome sequencing (WGS) discovery	**Integration** of targeting with WGS
Validation of genome variants from cancer WGS	**Accelerating** and improving variant validation
Clinical implications from cancer populations	**Facilitating** analysis large clinical cohorts of archival cancer samples
Clinical translation as diagnostic	**Rapid, accurate analysis** for prospective clinical review



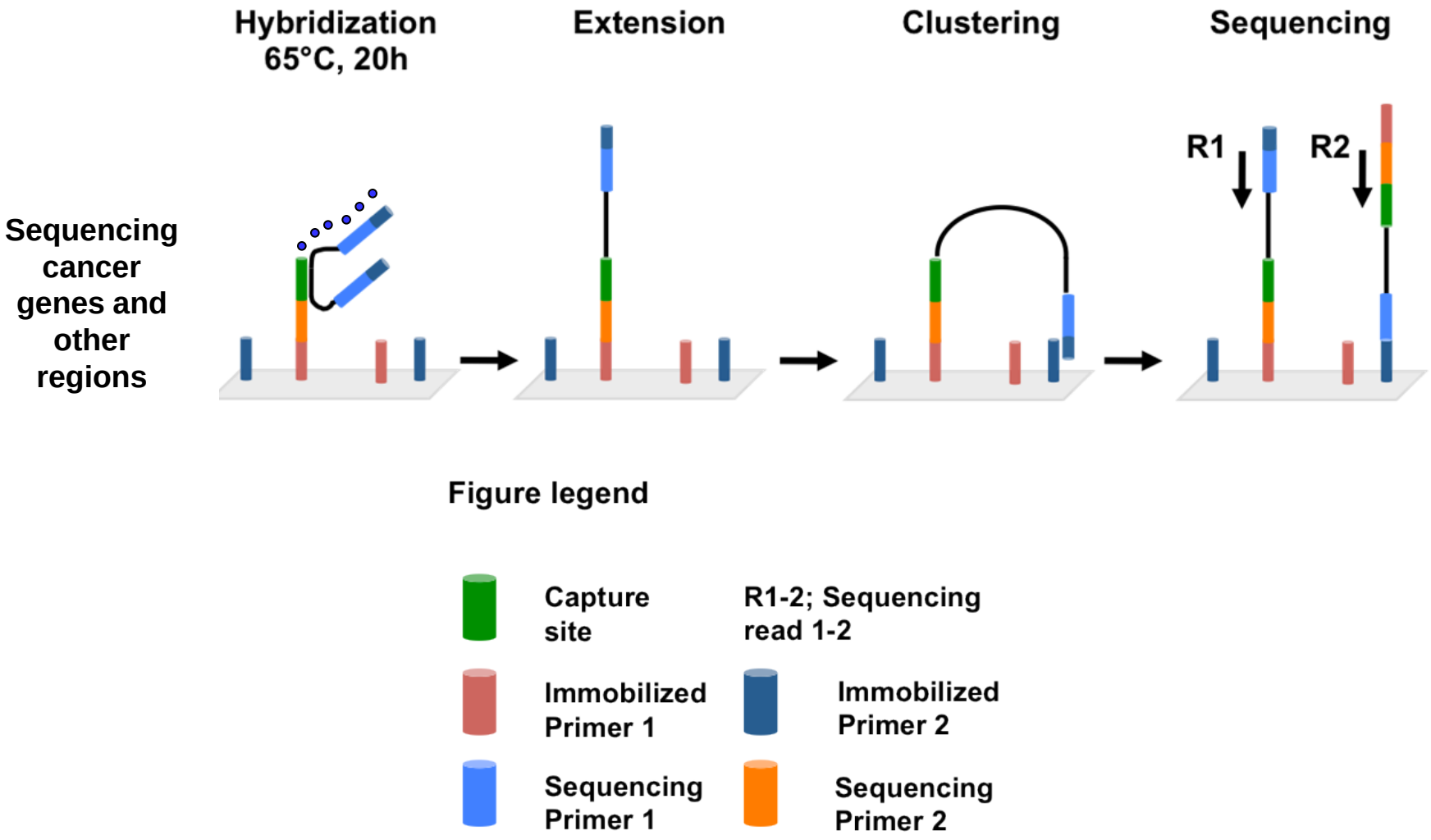
Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Method 1: OS-Seq
Step1: synthesize capture probes on flow cell lawn

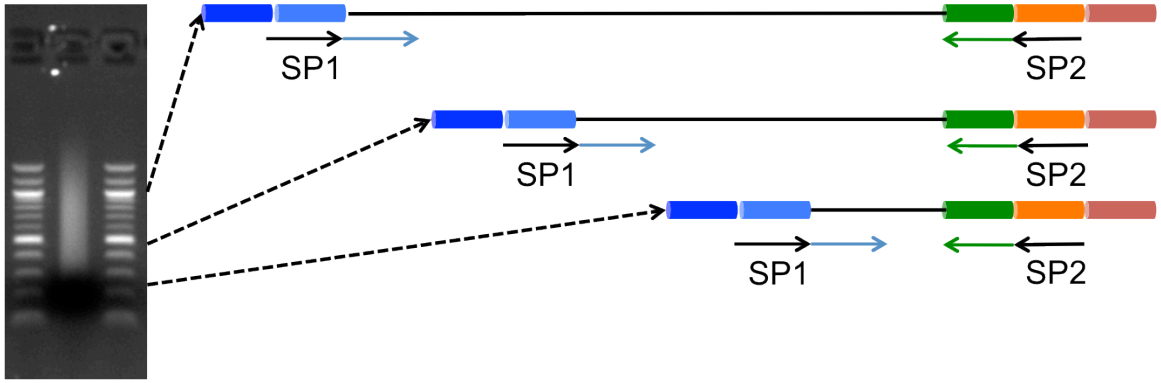


Myllykangas et al. 2011, Nat. Biotech.

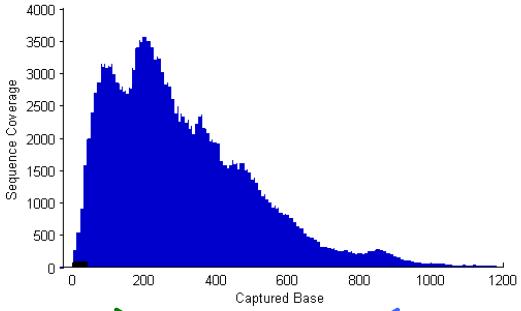
Step2: Capturing a target region from cancer genomes

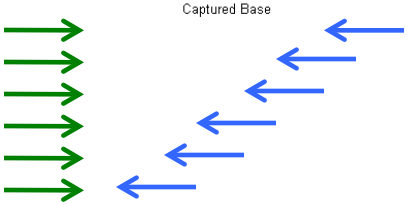


OS-Seq for targeting cancer genome regions



Strand-specific capture

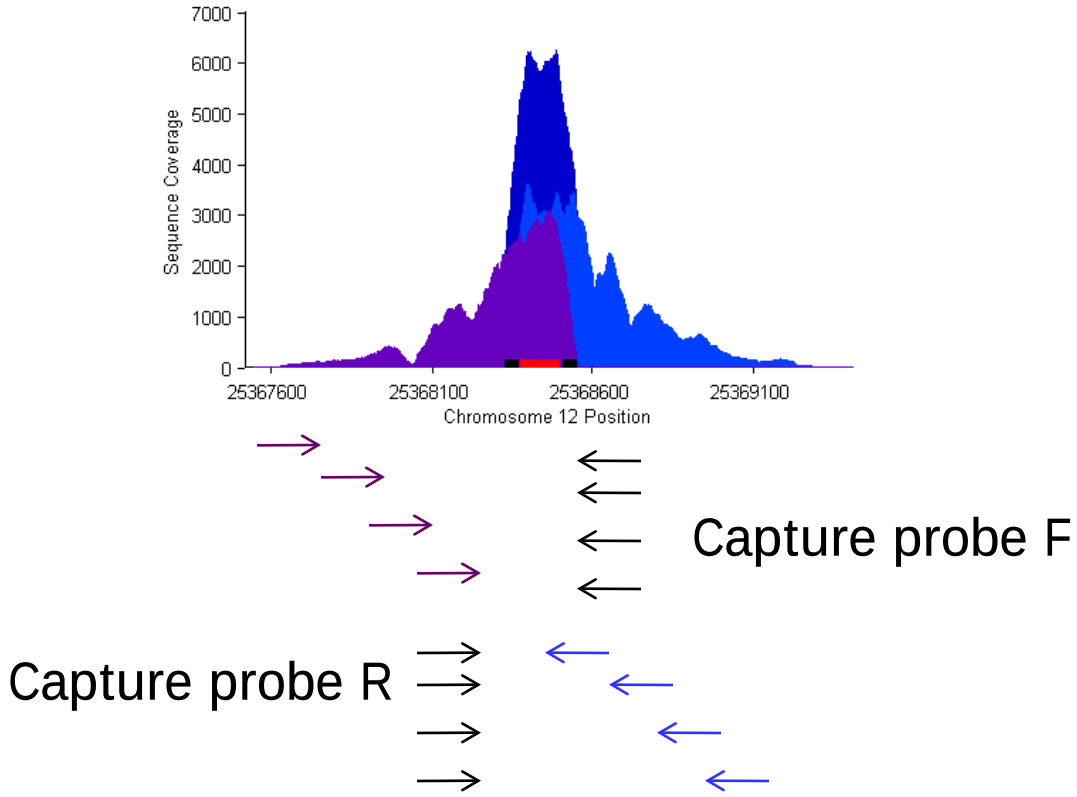




P1 reads

P2 reads

Primer probe design



- "Double strand" coverage of target with two primer probes
- Improved mutation discovery based on both strands

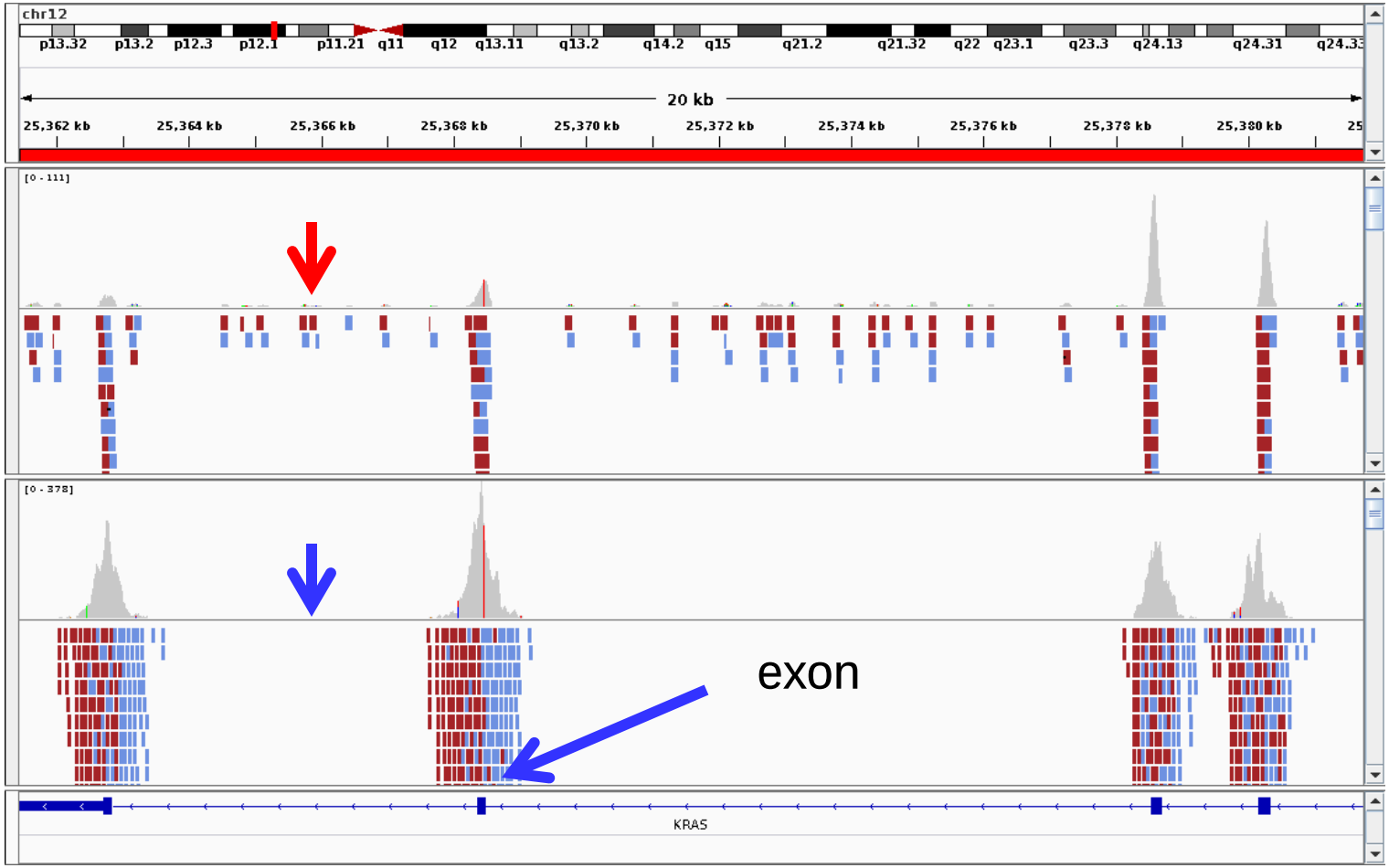
Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

OS-Seq: accurate targeting compared to other methods

- *KRAS* oncogene

Nimblegen exomes

OS-Seq



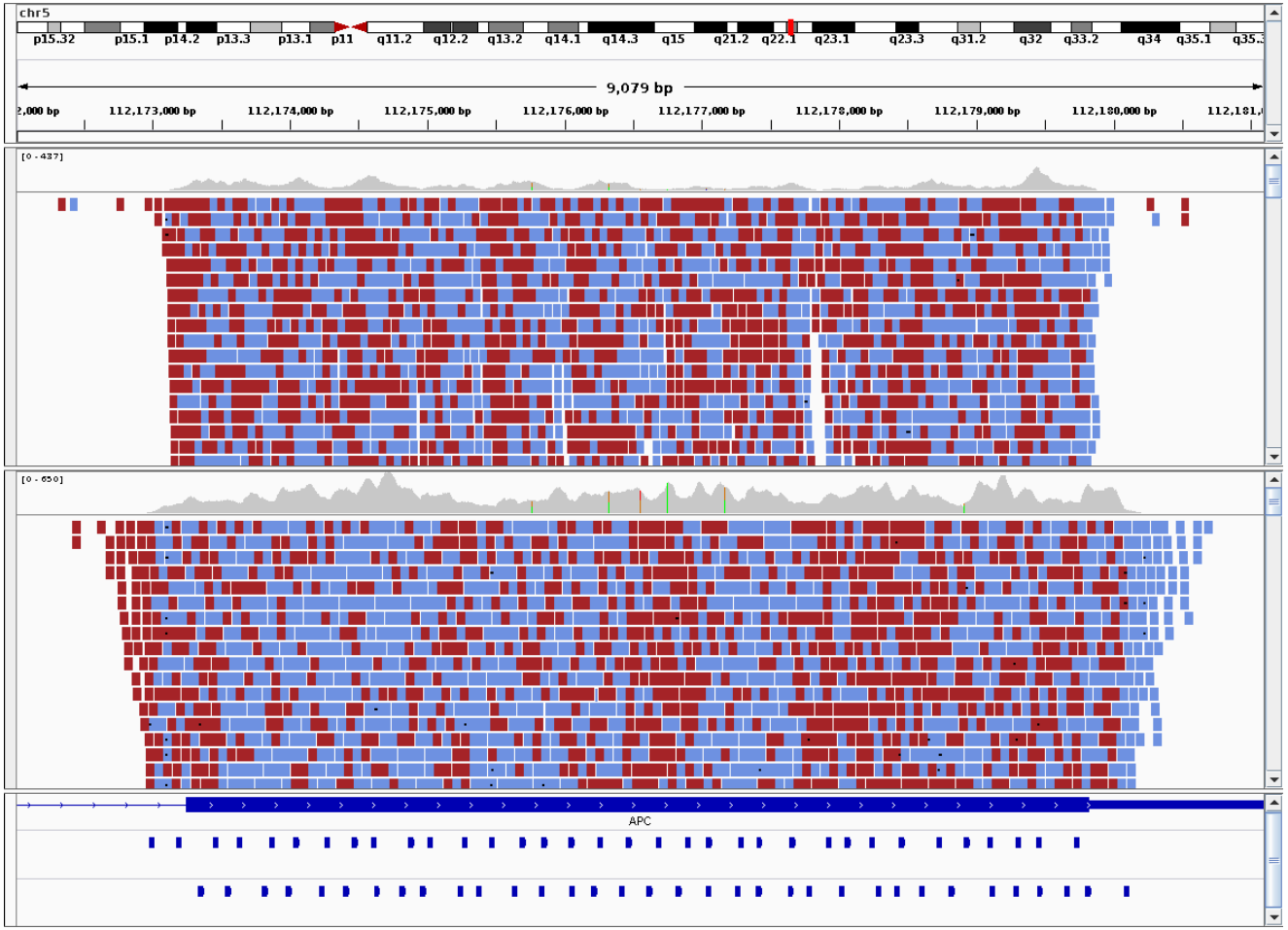
OS-Seq: Targeting loci like extended exons

- APC exon 15 (6.5 Kb)

Nimblegen exomes

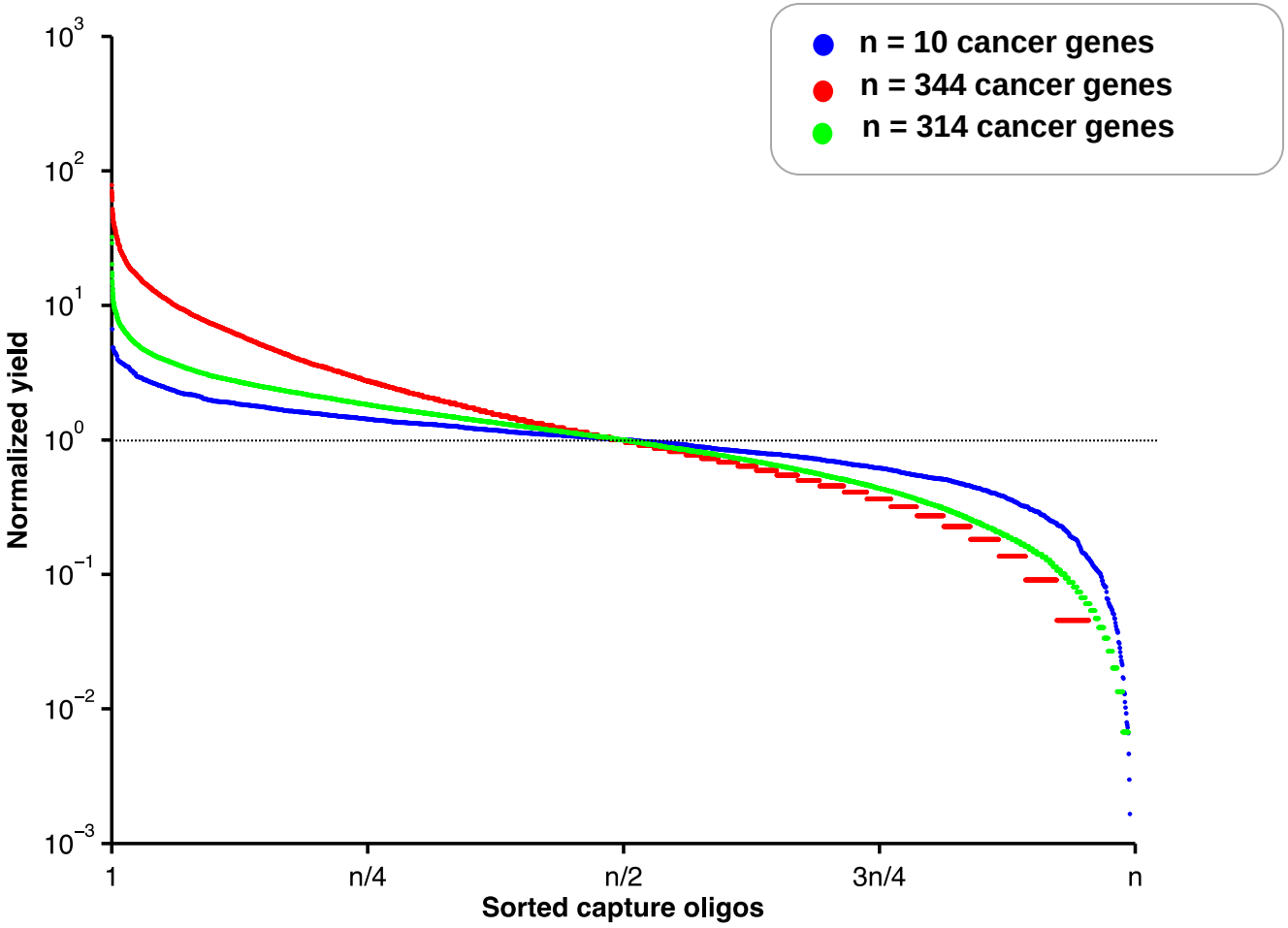
OS-Seq

Primer probe placement



OS-Seq: Even coverage of genomic region targets

- Primer probe yield



OS-Seq advances and advantages

- **Higher sensitivity and specificity mutation** detection with “deep” targeted resequencing
- **Higher accuracy** targeting of any nonrepetitive human genome region
- **Accurate variant discovery** - overlapping primer probe design improves variant detection
- **Identification** of rearrangement breakpoint sequences
- **Efficient** workflow of 1 day reduces experimental errors
- **Low** sample requirements (<1 ug DNA)

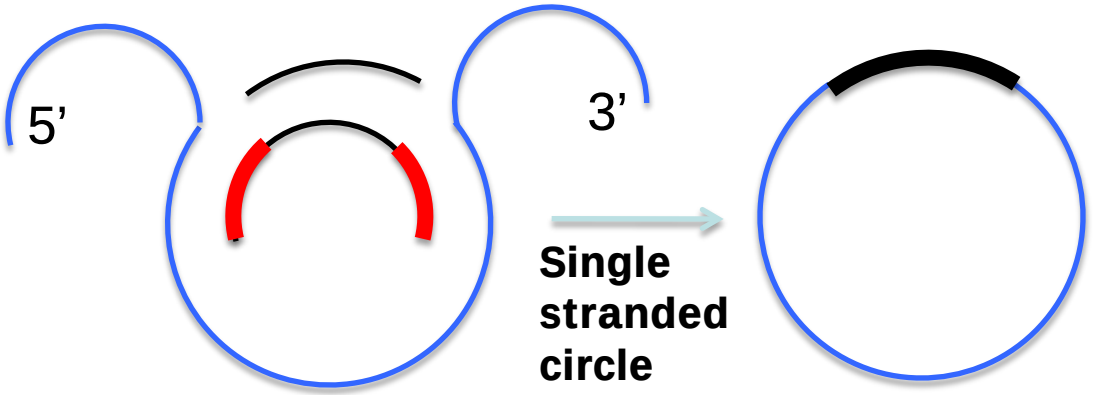
Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Method 2: single strand genomic circularization

Capture probes

Single-stranded fragmented DNA

Mix in presence of Ampligase, TaqPol, exoI



Single stranded circle

Amplify with common primers

Population of double stranded linears

Illumina Nextgen Seqencing

Key features:
-Single-stranded substrate compatible with FFPE material.
-Capture probes can be placed anywhere.

Pilot demonstration of targeting and accuracy

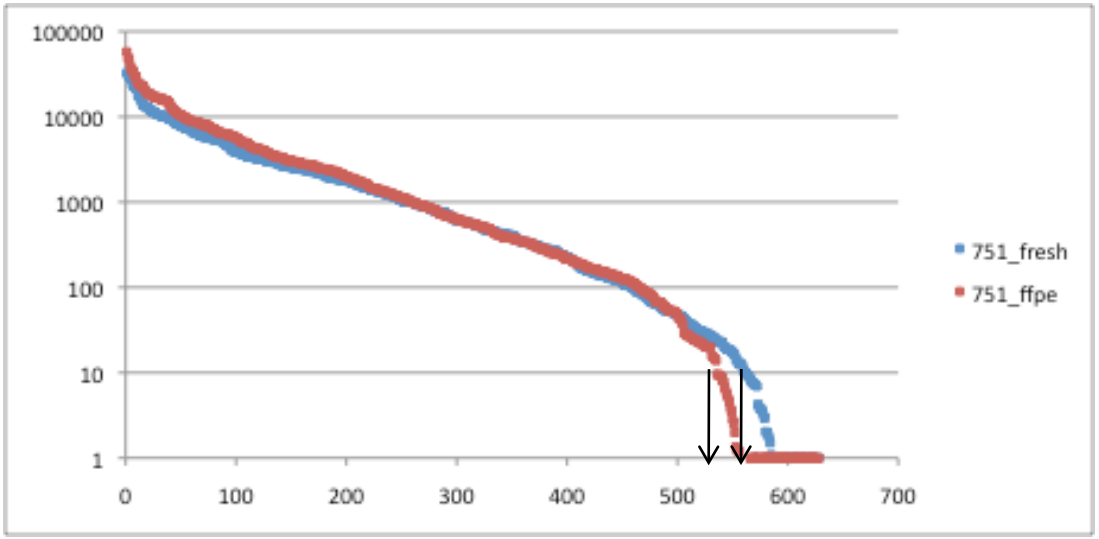
- 628 genomic regions targeted (~200bp average size)
- 123 Kb of total size of genomic targets
- Samples
  - Matched tissues from the same organ and individual
    - High quality genomic DNA from flash frozen tissue
    - Low quality DNA from matching FFPE tissue.
- Sequencing performed in triplicate

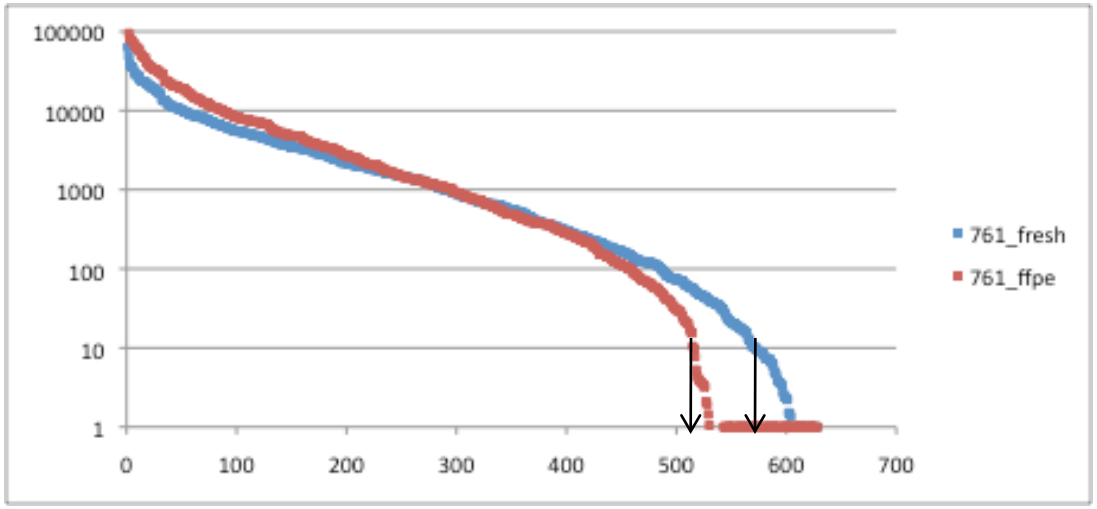
Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Mutation discovery from clinical archival samples

1. Compare capture yields from high quality versus from FFPE genomic DNA
2. Determine sensitivity of detection of heterozygote variants in high quality genomic DNA compared to matched archival genomic DNA (FFPE)
3. Evaluate FFPE-related DNA damage in variants in FFPE genomic DNA but not in high quality genomic DNA

Capture uniformity of high quality versus FFPE DNA





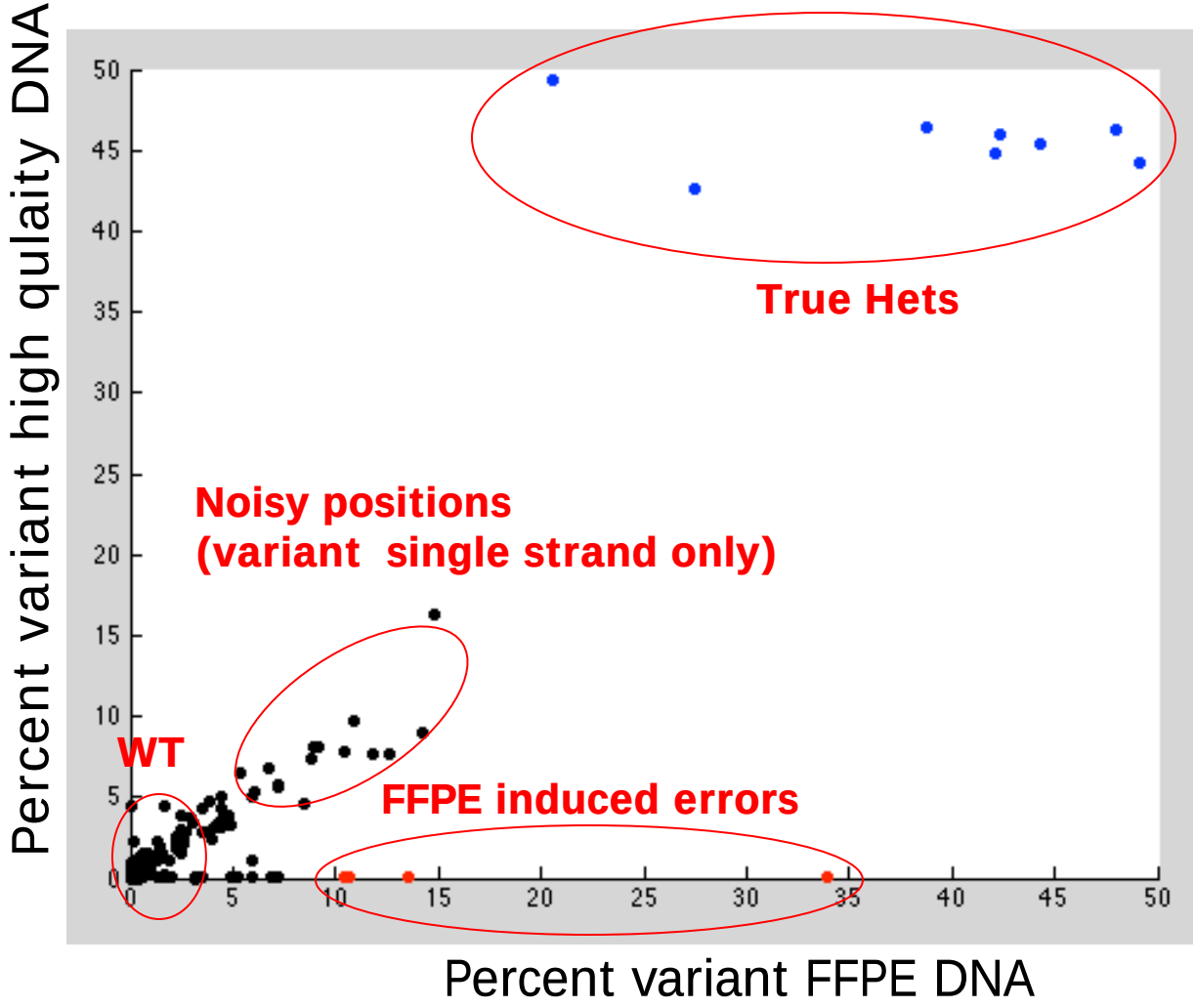
5% of the regions captured with coverage < 10X

Blue: high quality DNA

Red: FFPE only

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Artifacts introduced by FFPE processing



Blue: high quality DNA and FFPE DNA

Red: FFPE only

Specificity and sensitivity of detection

- **Sensitivity:** 85% heterozygote detection over 120 Kb target region
  - Related to capture coverage
- **Specificity:** 1 False positive heterozygote per 10-15kb (1 error per 5 genes)
- **Specific classes** of artifacts observed
  - transitions: G→ A: 7 times and C→ T: 8
  - transversions: C→ A: 4 times and G→T: 5

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Single lane mass-validation of whole genome sequencing

Whole Genome sequencing and exome sequencing of matched Normal blood/Primary gastric tumor/Ovarian metastasis

↓

386 coding variants including SNVs, Indels and SVs

↓

Validate all positions in parallel in a single lane of sequencing

↙ ↘

From flash frozen tissue

↓

OS-seq capture

↓

GAIIx or HiSeq

From FFPE

↓

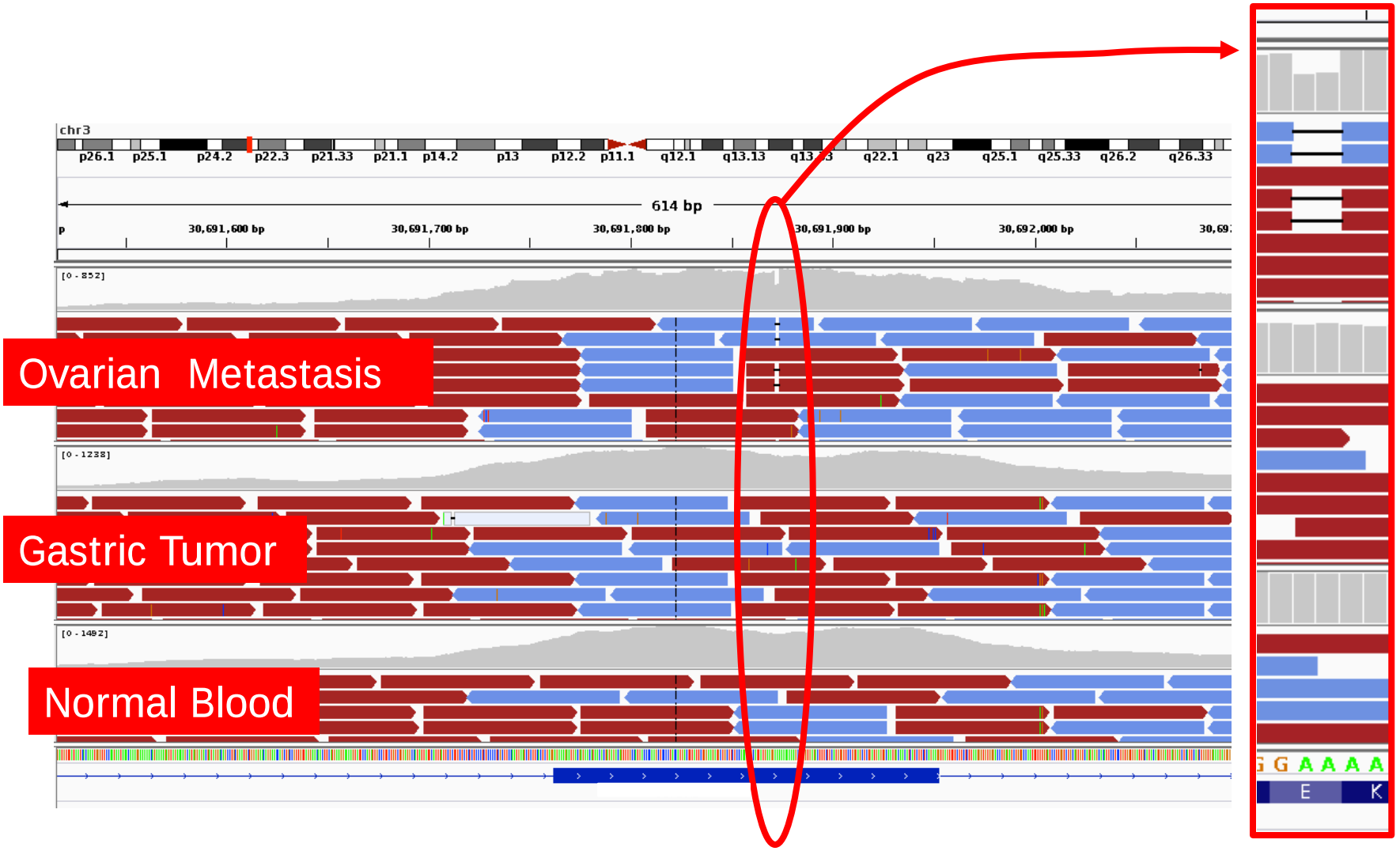
Single Strand Circularization

↓

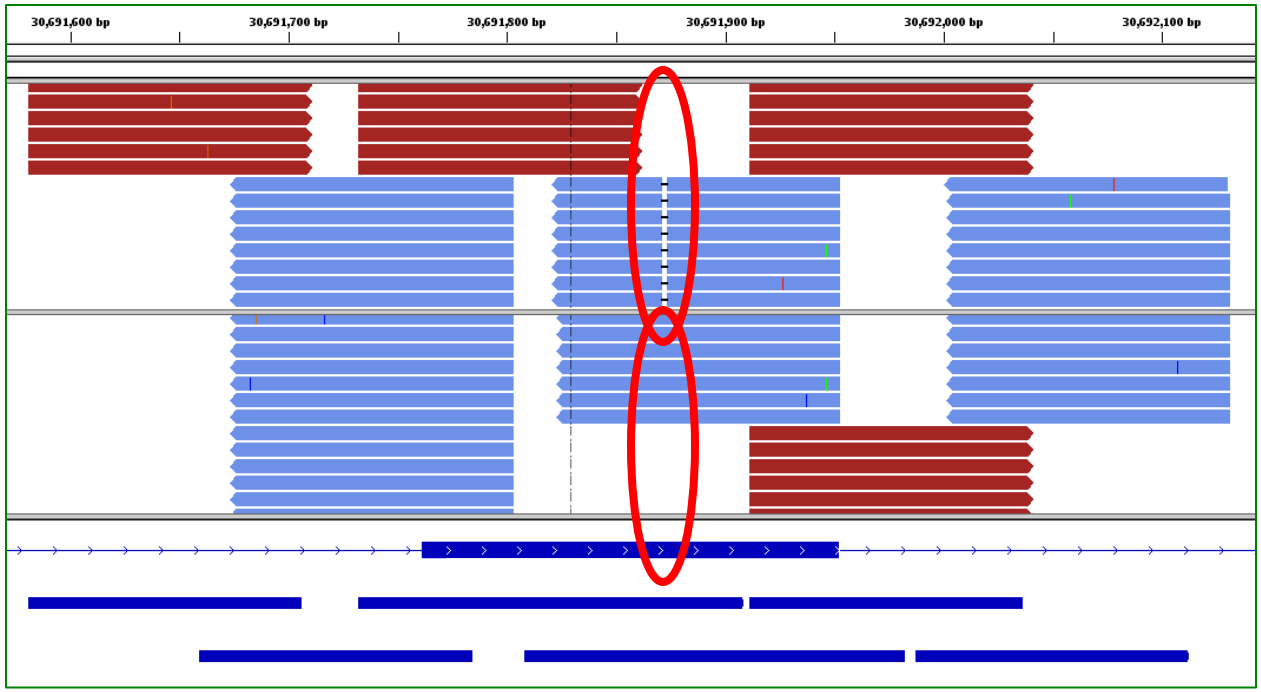
MiSeq

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Os-seq



Single strand genomic circularization



Ovarian Metastasis
FFPE

Normal
Flash frozen

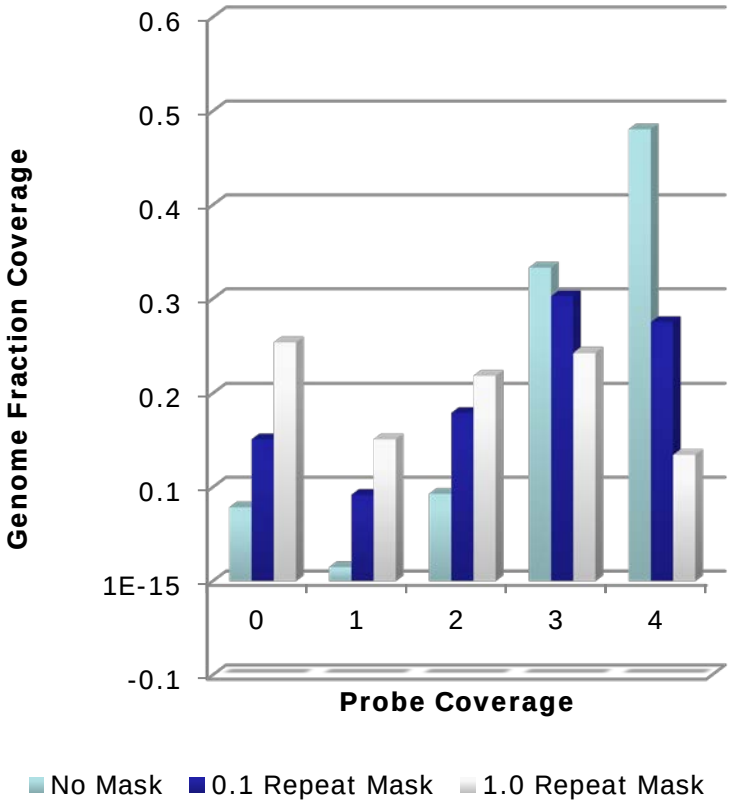
Note : targeted amplicons are end-sequenced (150 by 150 bp) on MiSeq

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

OligoGenome Resource – open access for capture assays

http://oligogenome.stanford.edu

Repeat Masking	% Genome Coverage	Probes
No Mask	~90%	26M
Low Repeat	~75%	20M
No Repeats	~50%	15M

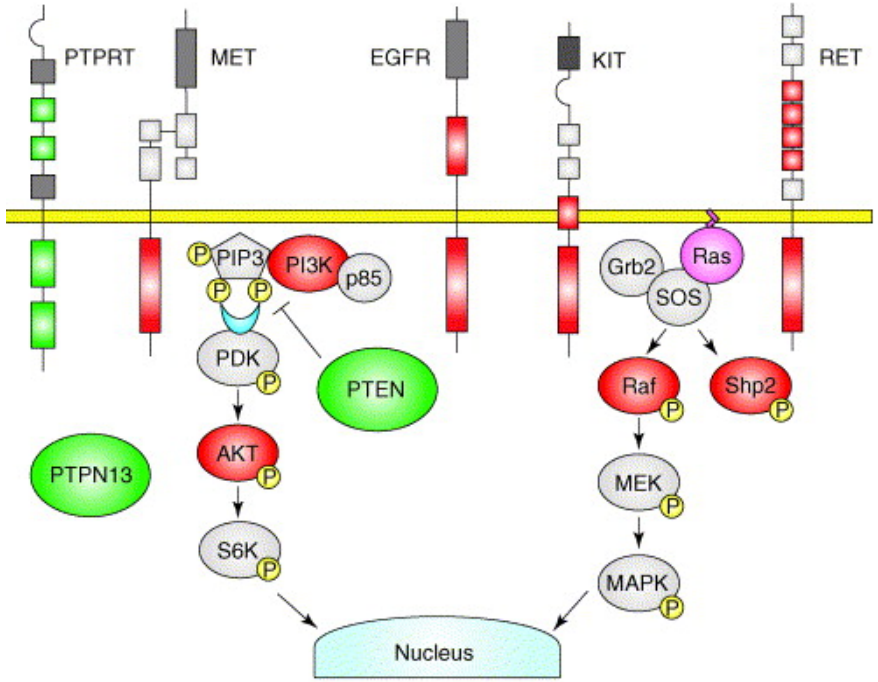


Newberger et al., *Nucleic Acids Research*, 2011

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Application of both methods to analysis of cancer genomes

- OS-seq:
  - Validation of mutations and rearrangements from cancer genomes
  - ”Onconome” and exome applications
- Single-strand circularization:
  - Follow-up clinical applications using archival samples (FFPE)



Both methods are scalable → single lane validation of cancer genomic projects

Stanford University Division of Oncology
Stanford Genome Technology Center
Stanford Cancer Institute

Acknowledgements

- *Ji Research Group*
  - Georges Natsoulis
  - Samuel Myllykangas
  - Jason Buenrostro
  - Erik Hopmans
  - Daniel Newburger
  - Laura Miotke
  - Hua Xu
  - Chris Xu
  - Sue Grimes
- *Division of Oncology*
  - Lincoln Nadauld
- *Stanford Genome Technology Center*
  - Michael Jensen

- *Funding:*
  - *NIH*
    - IMAT - National Cancer Institute (NCI)
    - National Human Genome Research Institute (NHGRI)
  - *Doris Duke Foundation*
  - *Howard Hughes Medical Foundation*