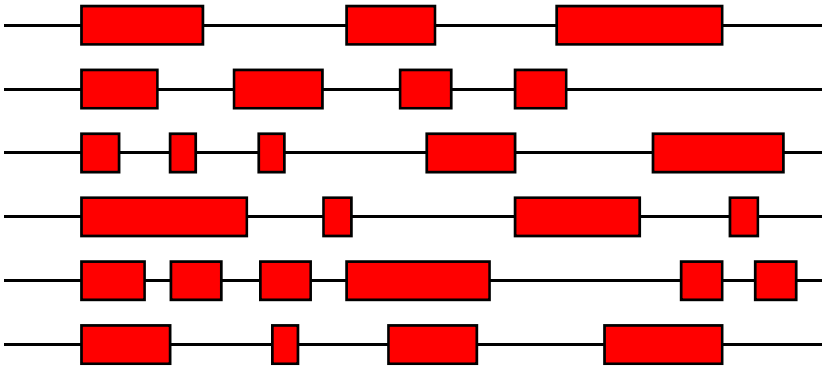
Sequencing the "Cancer Genome"

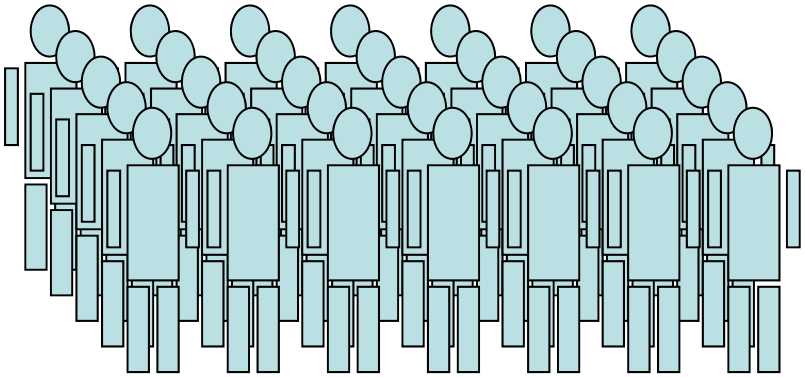
Richard K. Wilson, Ph.D.
Washington University
School of Medicine

rwilson@watson.wustl.edu

PCR-based re-sequencing

list of candidate genes





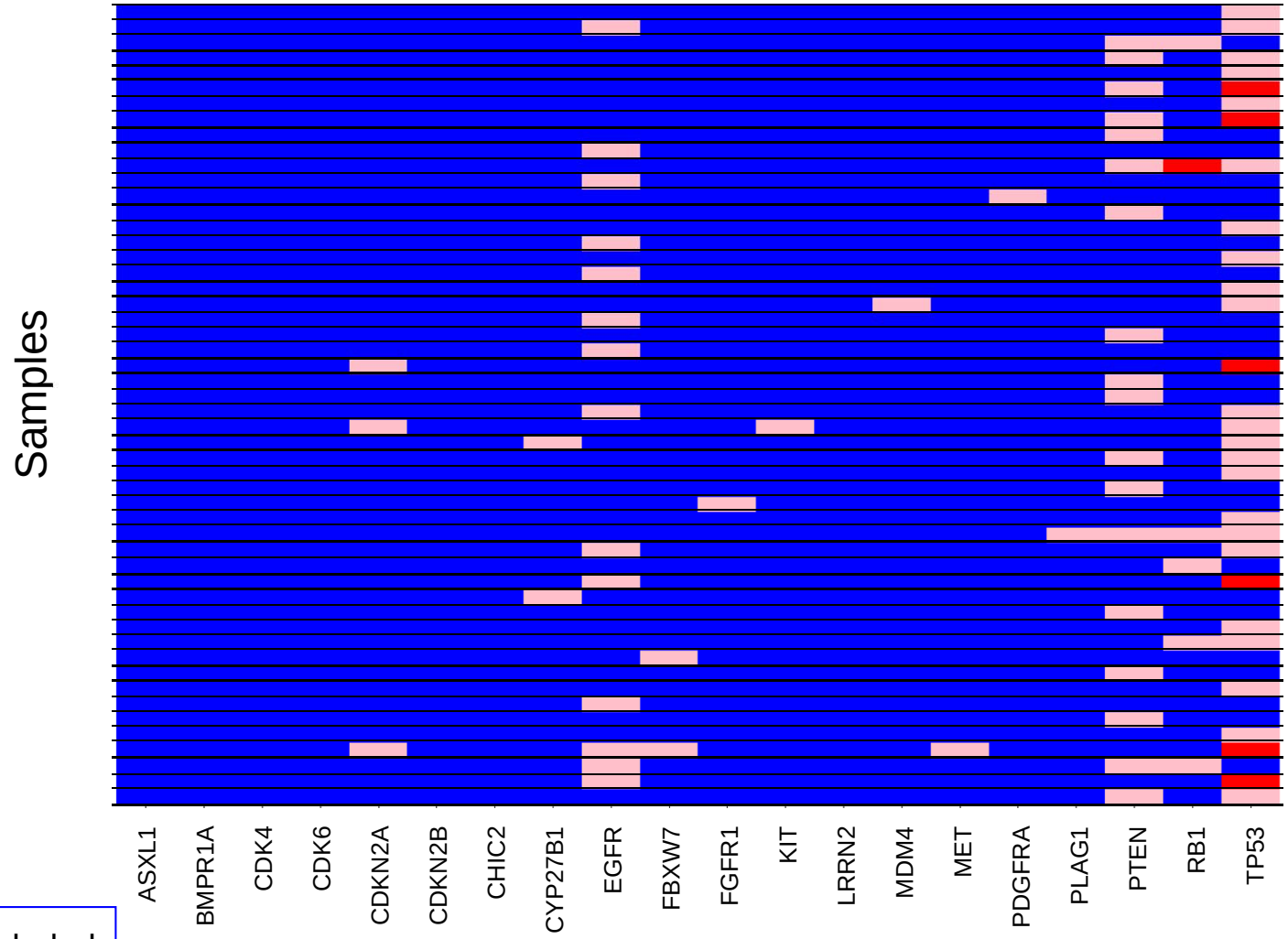
large collection of patient samples

TCGA: Targeted Re-sequencing in GBM

- Phase I (n=601)
  - Published literature and unpublished data in the GBM community
  - Emphasize genes with genetic evidence of alterations (e.g. amplification/deletion; mutation) or with clinical outcome correlation
- Phase II (n=725)
  - Initial GBM genome characterization data by CGCC
  - Genetic elements of interest (GEOI) defined by integrative analysis of copy number and expression in GBM across multiple platforms (include both coding mRNA and non-coding microRNA)
  - Conserved regions across evolution
  - Tumor-specific spliced variants

GBM: Somatic mutations

(Shared 20 genes, 84 tumors; orthogonal validation pending)

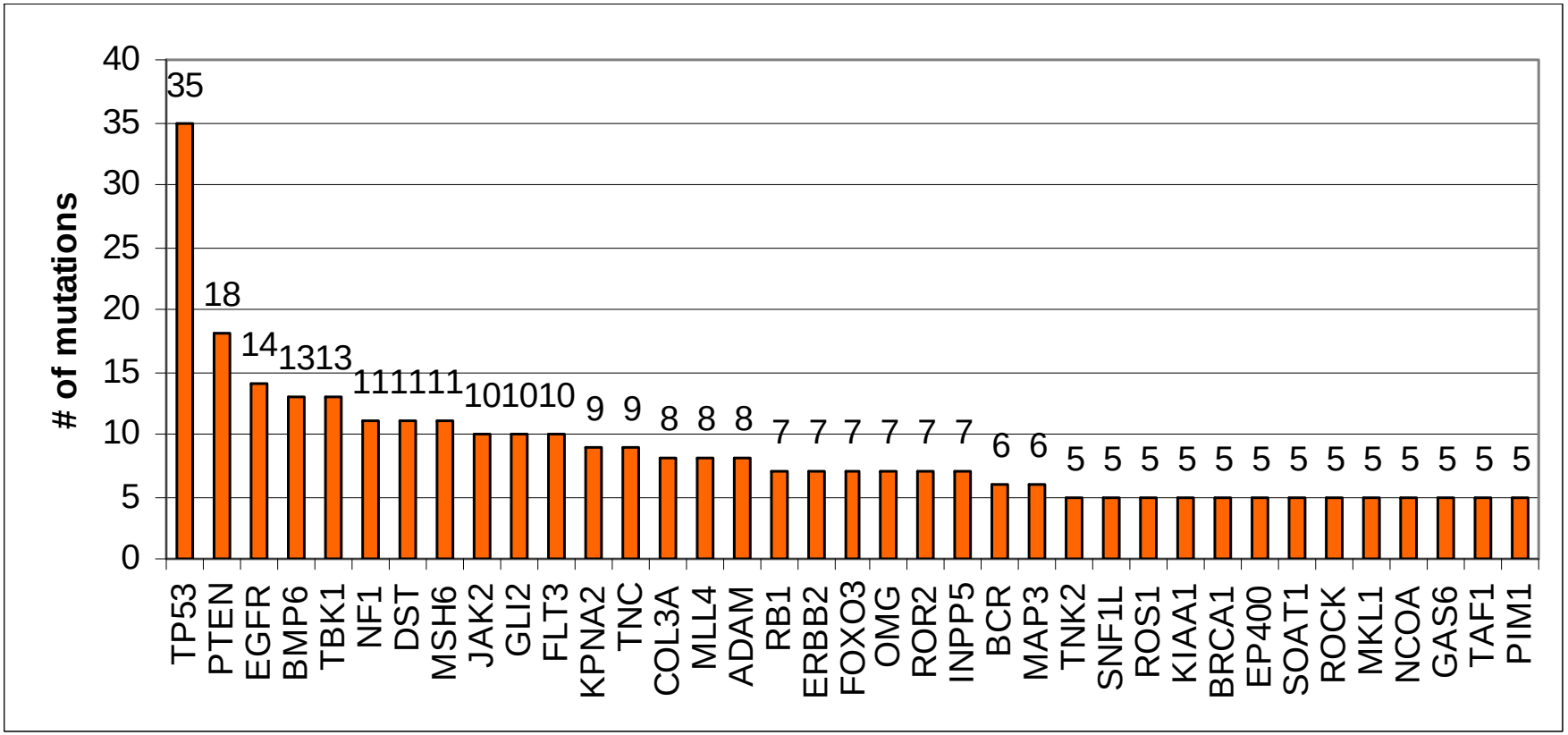


*Indels included

Pink: 1 mutation Red: 2 mutations

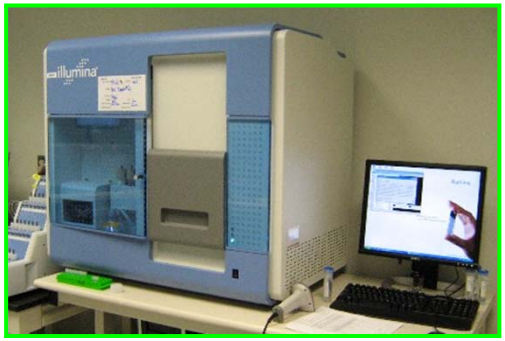
GBM: Somatic mutations

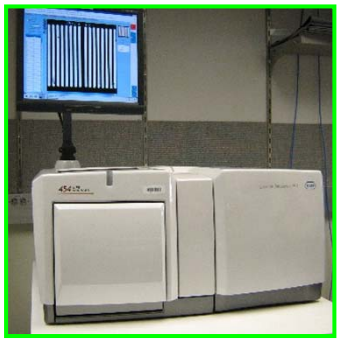
(601 Phase I genes, 84 tumors; orthogonal validation pending)



❖TCGA WU/Broad/Baylor/Shared genes used for this analysis

Next Generation DNA Sequencing Technologies




Genome size:		3000 Mb	
Req'd coverage	6	12	25

	3730	454 FLX	Solexa
bp/read	600	250	32
Reads/run	96	400,000	40,000,000
bp/run	57,600	100,000,000	1,280,000,000
#/runs req'd	312,500	360	59
Cost per run	$ 48	$ 6,800	$ 9,300
Total cost	**$15,000,000**	**$ 2,448,000**	**$ 544,922**

454-based Mutation Detection

DNA from X tumor samples

↓ Pooled with equal concentration

PCR amplification with Y primer pairs

↓ Pool PCR products

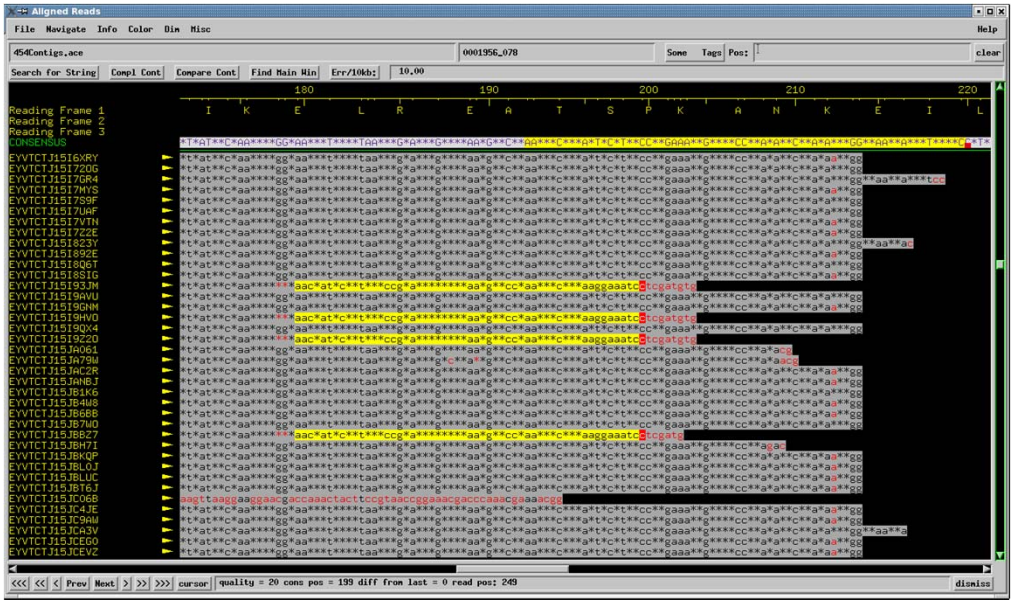
454 sequencing

↓

SNP/Indel Detection Using ssahaSNP and BreakPointRead

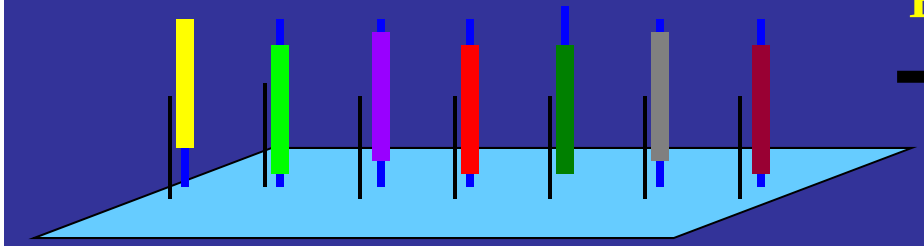
Reads with G12 mutation in KRAS

Reads with 15 bp deletion in EGFR

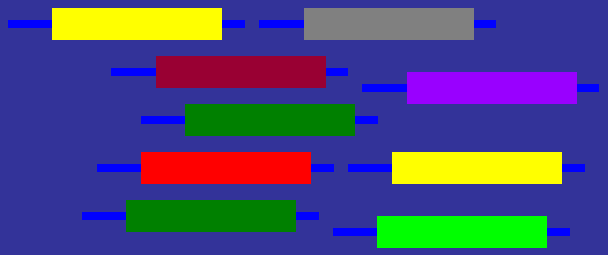


BCM/Nimblegen Approach to Exon Sequencing

Fragment and hybridize to Nimblegen capture array

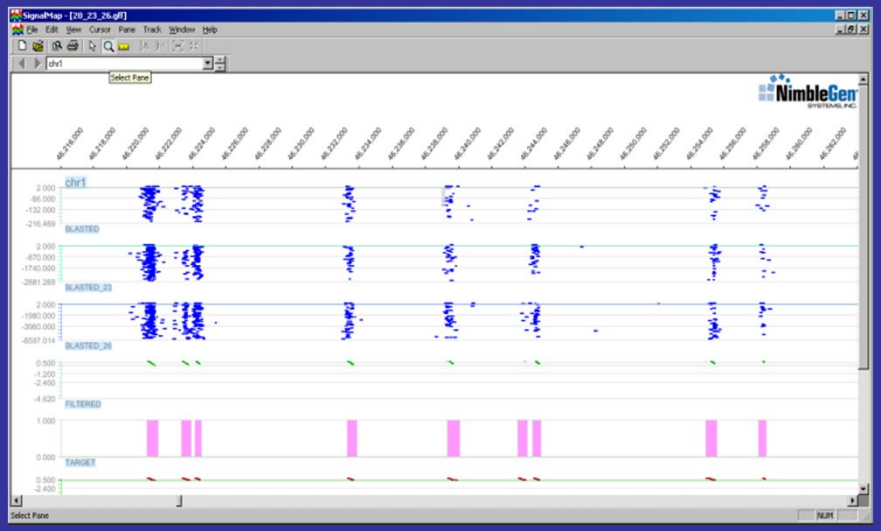


Elute



454 Sequencing

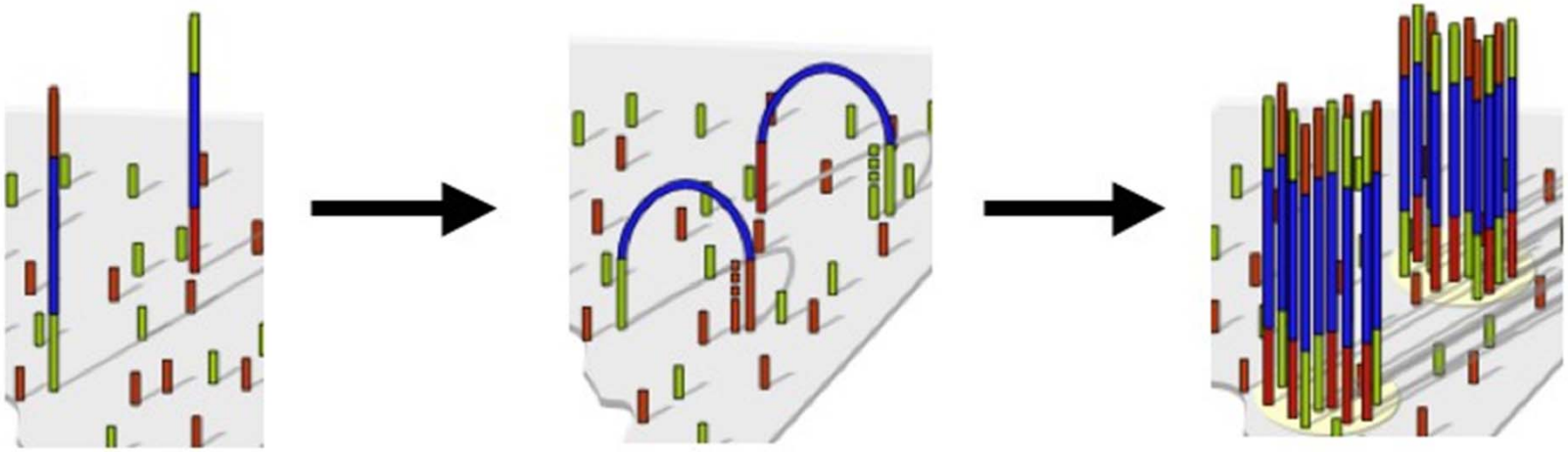
Analyze



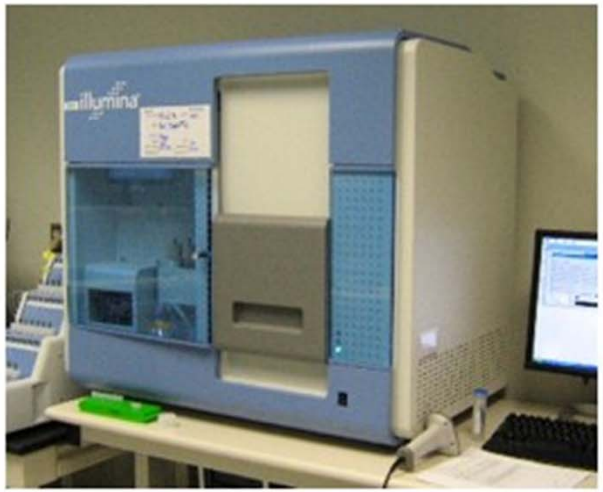
Exon sequences

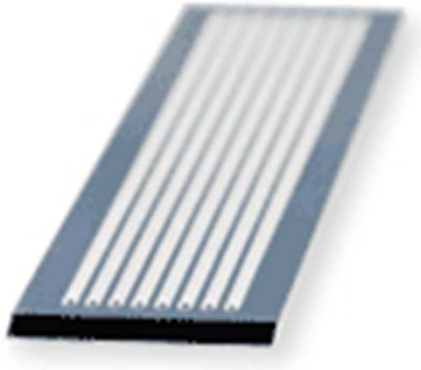
Courtesy of R.A. Gibbs

Solexa-based Whole Genome Sequencing



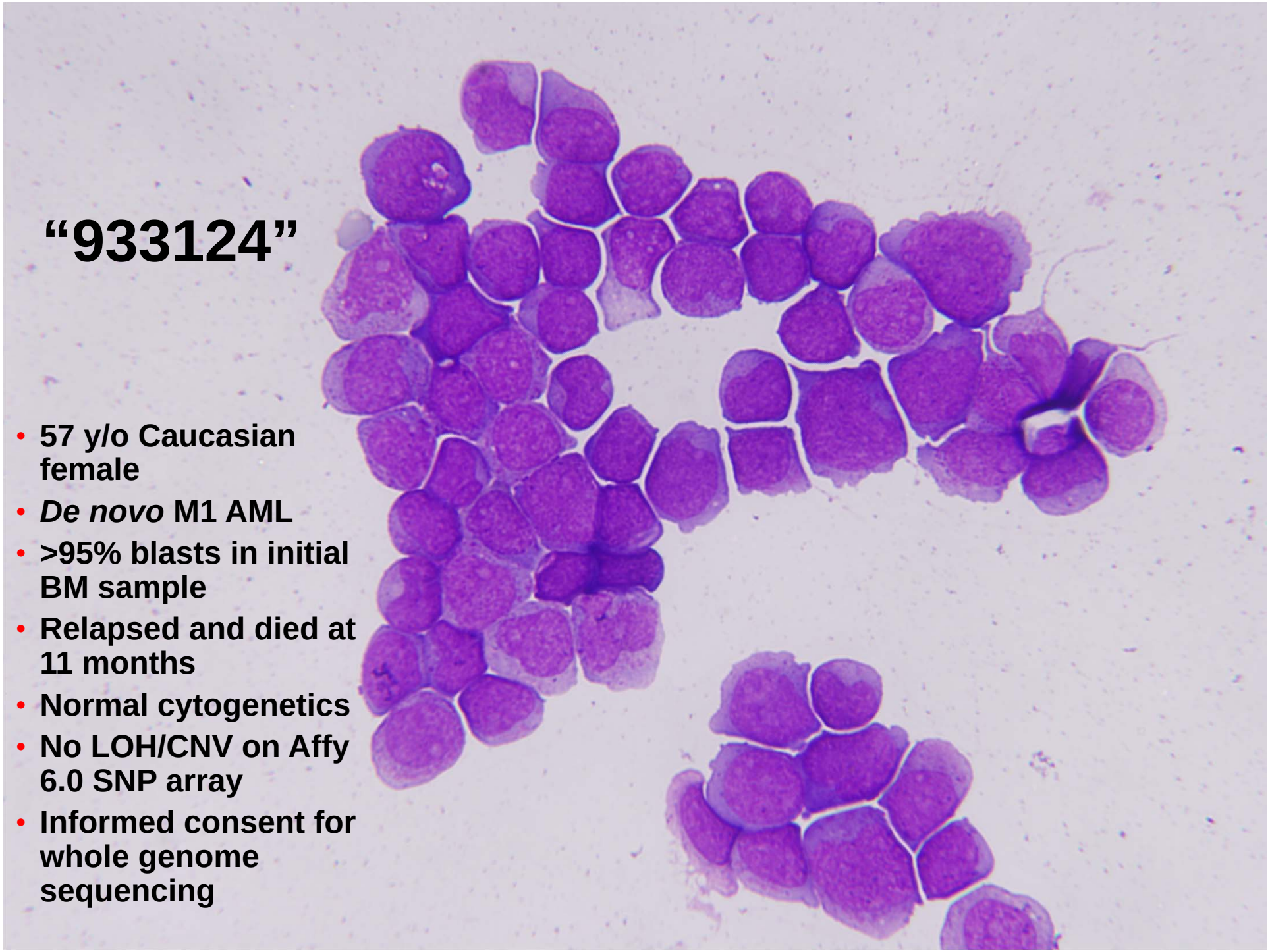
Solexa flow cell



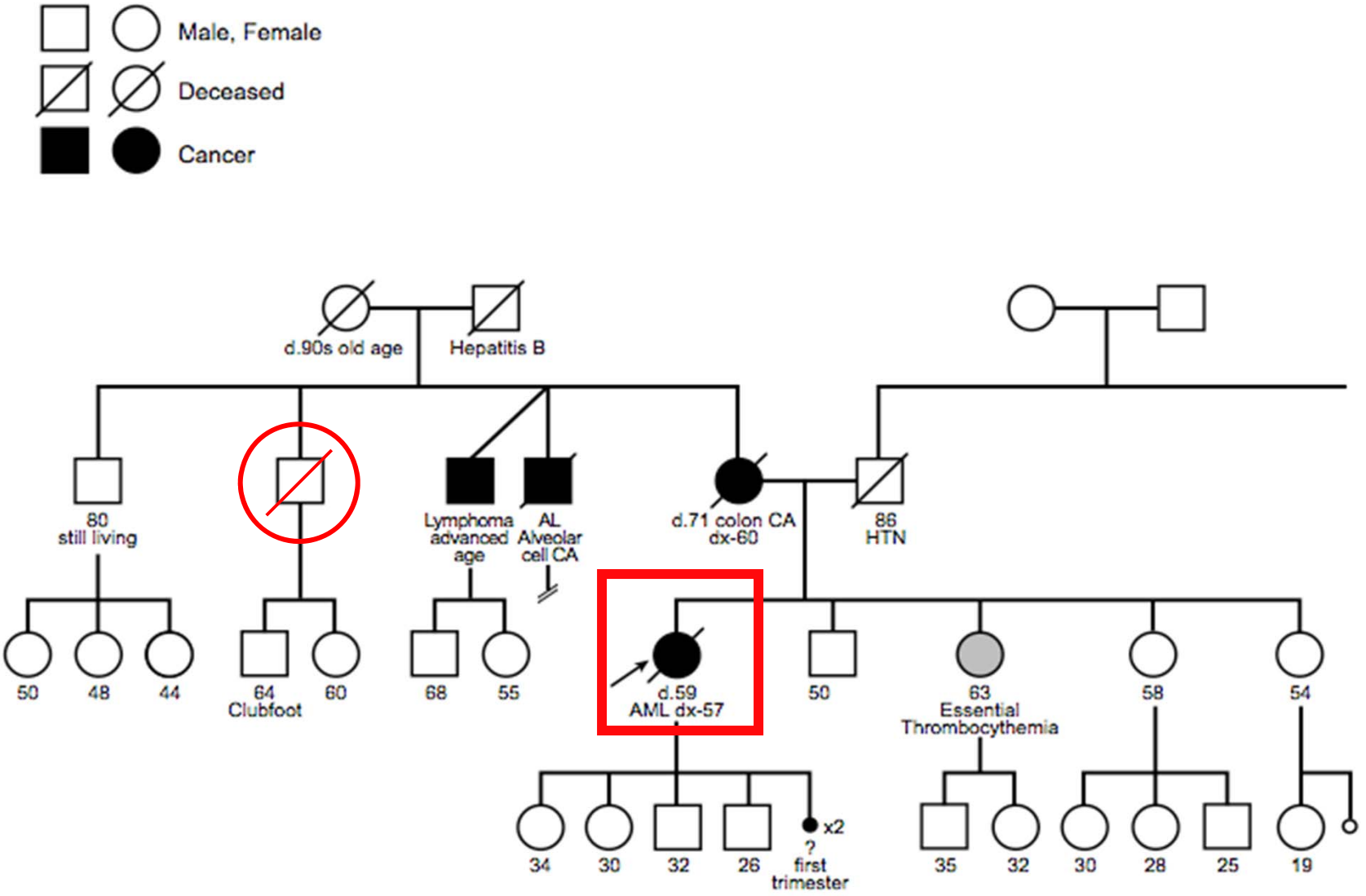


~50M clusters are sequenced per flow cell.

“933124”



- 57 y/o Caucasian female
- *De novo* M1 AML
- >95% blasts in initial BM sample
- Relapsed and died at 11 months
- Normal cytogenetics
- No LOH/CNV on Affy 6.0 SNP array
- Informed consent for whole genome sequencing



AML Tumor Genome Sequence

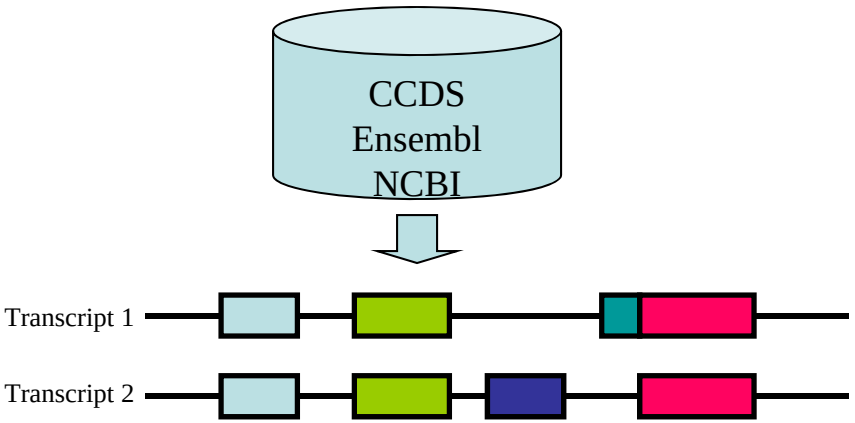
- **77% diploid coverage** was obtained for an AML M1 tumor genome with 22x sequence coverage (**97.4% haploid coverage**) using Solexa (Illumina) sequencers.
- 10x sequence coverage of normal genome (epidermal).
- **2.1M sequence variants** were detected (consistent with the known human polymorphism rate).
- **~495,000 are novel variants**: SNPs vs. somatic mutations
- **3,731 non-synonymous coding variants** were detected; most are likely rare SNPs (compare to germline & cDNA).
- **Two somatic mutations** have been detected and confirmed; both are coding sequence insertions.
- Currently validating a small number (<30) of additional candidate somatic mutations...

AML Tumor Genome Sequence

- FLT3: FMS-like tyrosine kinase
- NPM1: nucleophosmin
  - Both are small insertion mutations.
  - Both have been previously implicated in myeloproliferative disease.

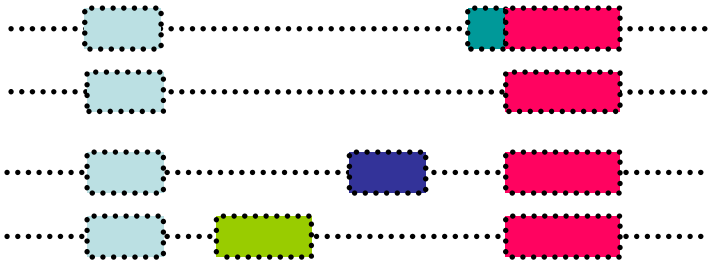
HWI-EAS68__10151_5_84_616_593	1	8	NPM1	854	861
HWI-EAS68__10151_5_84_616_593	2	9	NPM1	855	862
HWI-EAS68__10151_5_84_616_593	3	10	NPM1	856	863<-4bp ins?
HWI-EAS68__10151_5_84_616_593	15	22	NPM1	864	871<-4bp ins?
HWI-EAS68__10151_5_84_616_593	17	24	NPM1	866	873
HWI-EAS68__10151_5_84_616_593	22	29	NPM1	871	878
HWI-EAS68__10151_5_84_616_593	23	30	NPM1	872	879
HWI-EAS68__10151_5_84_616_593	24	31	NPM1	873	880
HWI-EAS68__10151_5_84_616_593	25	32	NPM1	874	881
HWI-EAS68__10151_5_84_616_593	20	27	NPM1	365	372

Detection of alternative splicing

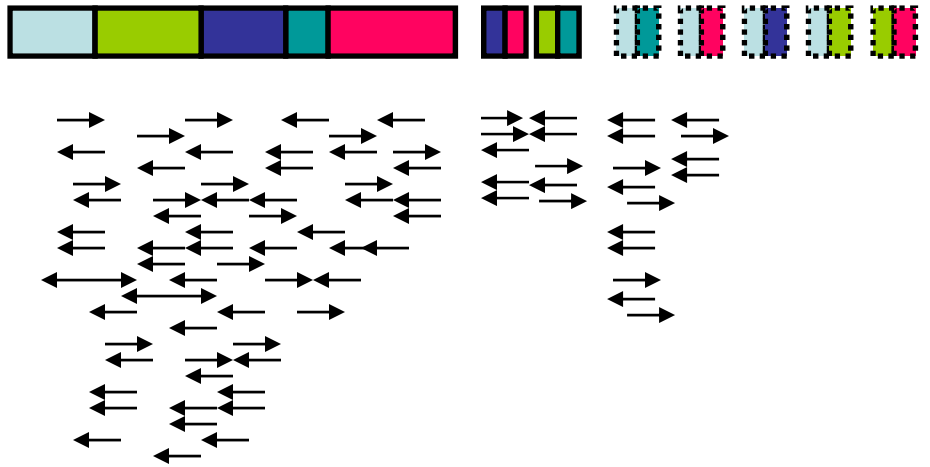


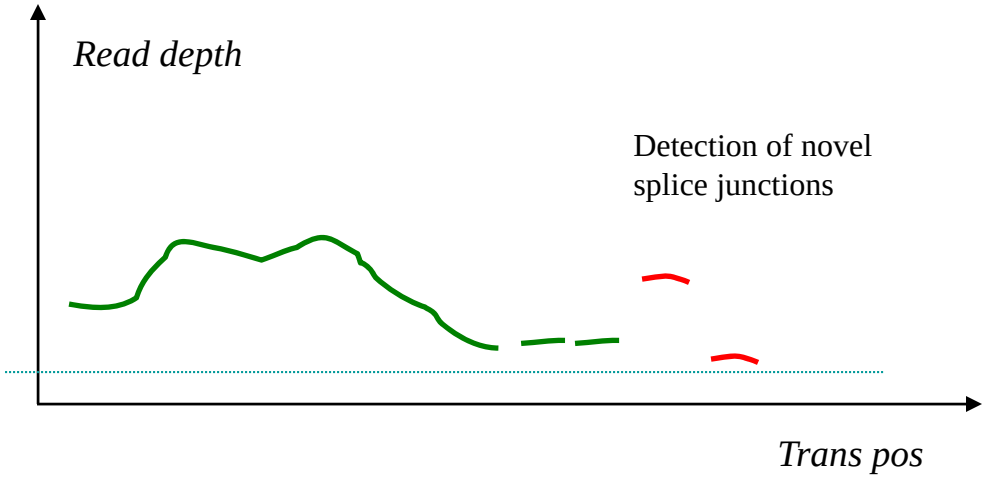
Merged non-redundant transcript:

Putative transcripts:



Putative alternative splice junctions:





- 454/Solexa AML cDNA sequencing: 350 novel transcripts in ~300 genes.

Acknowledgments

- **WU Genome Sequencing Center**

  Elaine Mardis, Li Ding, Mike McLellan, Brian Dunford-Shore, Ken Chen, David Dooling, Sean McGrath, Matt Hickenbotham, Lisa Cook, Tracie Miner, Lucinda Fulton, *et al*.

- **WU Siteman Cancer Center**

  Tim Ley, Mark Watson, Matt Walter, Rhonda Ries, Jackie Payton, John DiPersio, Dan Link, Michael Thomasson, Tim Graubert, Sharon Heath

- **TSP/TCGA Colleagues**

  Baylor HGSC, Broad Institute, many others…

- **Funding sources**

  NHGRI (Wilson), NCI (Ley), Alvin J. Siteman (AML WGS)

genome.wustl.edu