CLINICAL APPLLICATIONS OF NEXT GENERATION SEQUENCING

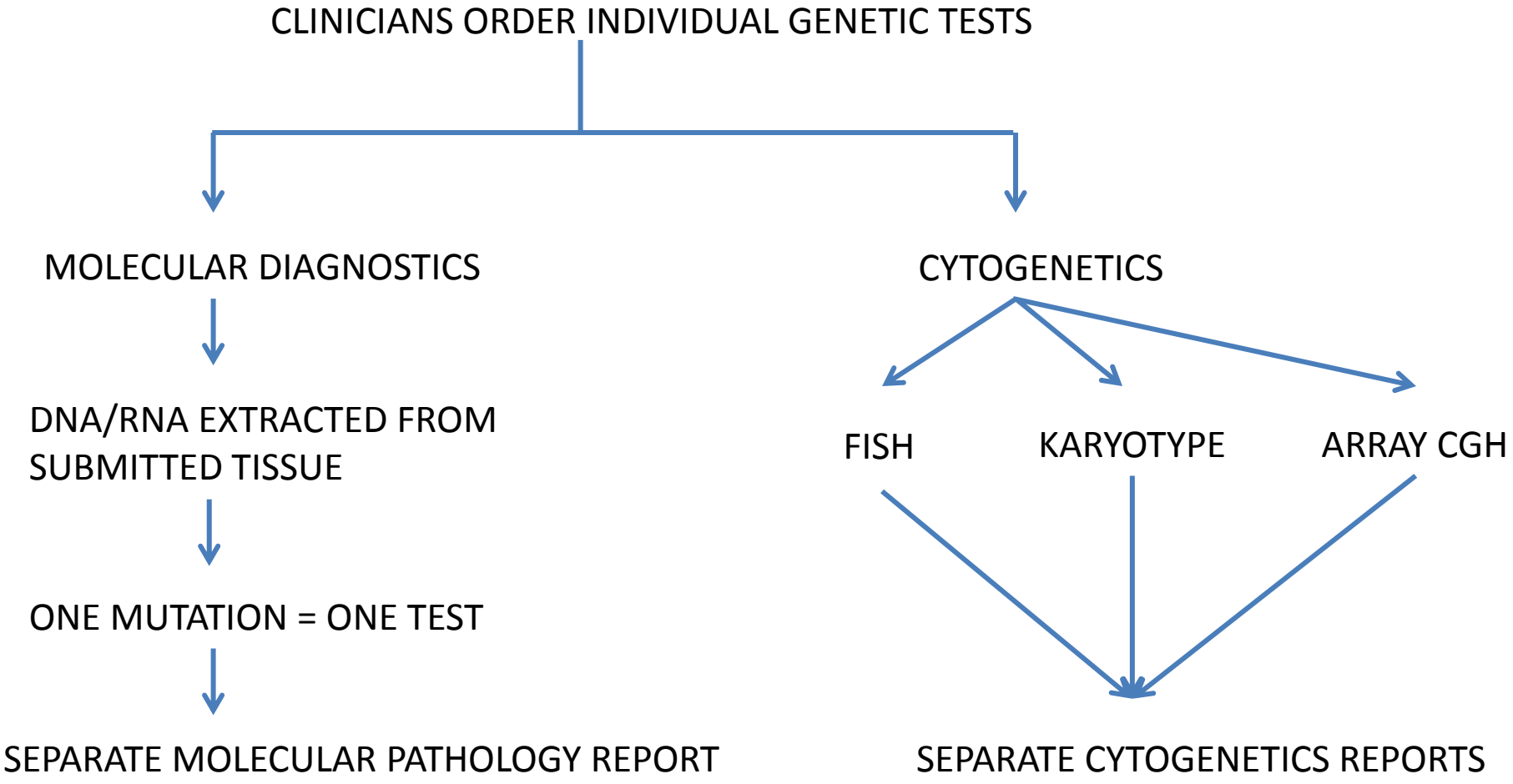
Bharat Thyagarajan
Department of Laboratory Medicine and Pathology
University of Minnesota

MOLECULAR DIAGNOTSTICS LABORATORY

- The Molecular Diagnostics Laboratory (MDL) processes around 20,000 specimens annually

- Major testing categories
  - Infectious disease testing: HPV
  - Bone marrow engraftment analyses
  - Hematological malignancies
    - Translocations, quantitative BCR-ABL, JAK2/FLT3/NPM1/CEBPA. T and B cell gene rearrangements etc.
  - Solid tumor malignancies
    - Microsatellite instability, KRAS, BRAF etc.
  - Inherited disorders
    - Factor II, V mutations, sequencing, Southern blot etc.

CURRENT MOLECULAR TESTING SCHEME FOR ONCOLOGY



LUNG CANCER

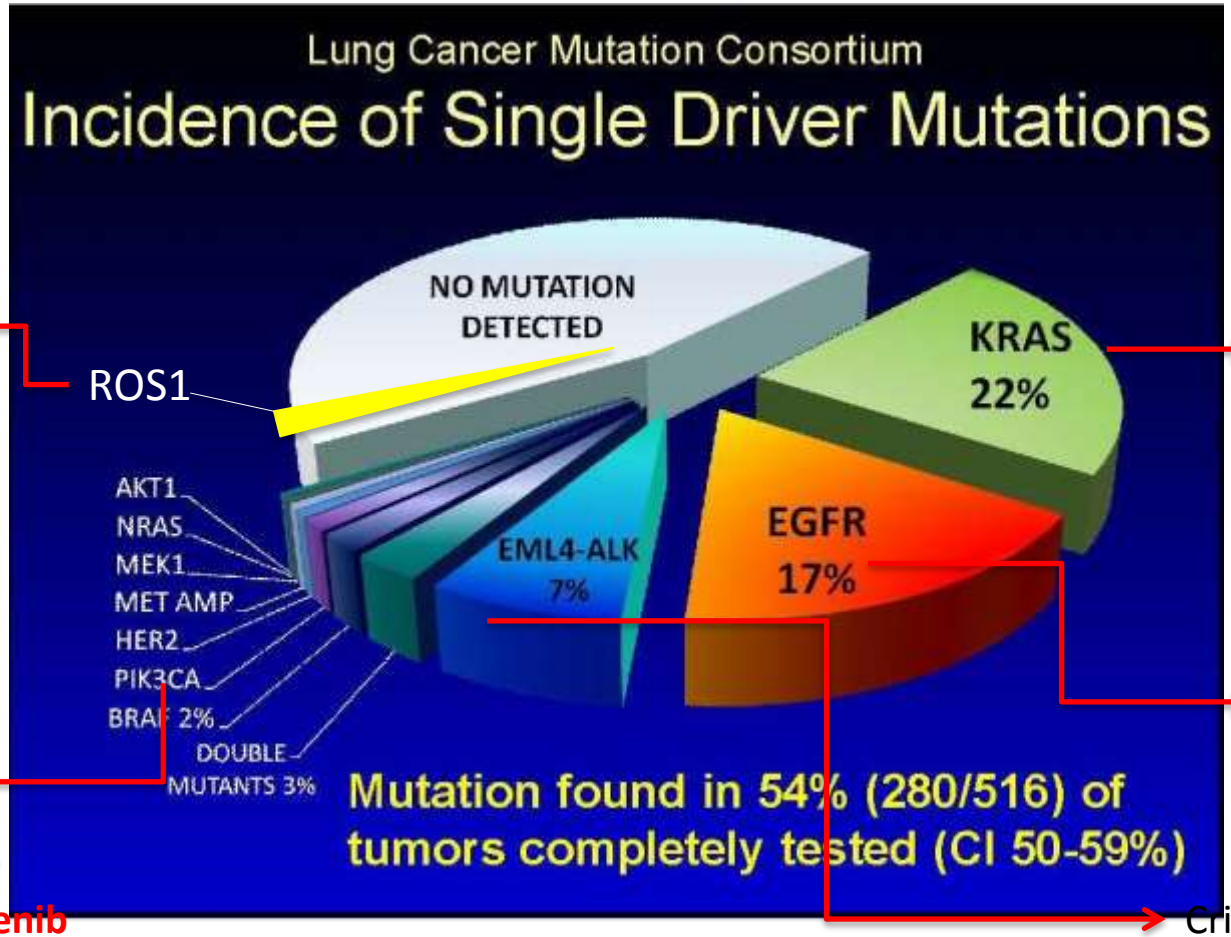
- As recently as a decade ago stage IV lung cancer had an universally poor prognosis of < 12 months irrespective of chemotherapeutic regimen
- At present, lung cancer with EGFR mutations have a mean survival of > 2 years
- Initial results of targeted therapy for other genetic alterations (e.g. ALK, ROS translocations) have shown promise
- Thus lung cancer is now considered a prototype for genetically tailored cancer therapy

CURRENT GENETIC TESTING FOR LUNG CANCER

- Molecular Diagnostics
  - *EGFR* mutation analysis
    - Wide range of genetic alterations including point mutations in various exons (18-21) and deletions in exon 19

- Cytogenetics
  - *ALK-EML4* translocations
    - Commonly detected using an *ALK* break-apart probe

MUTATIONAL PROFILE IN LUNG CANCER



Crizotinib

MEK1/2 inhibitors??

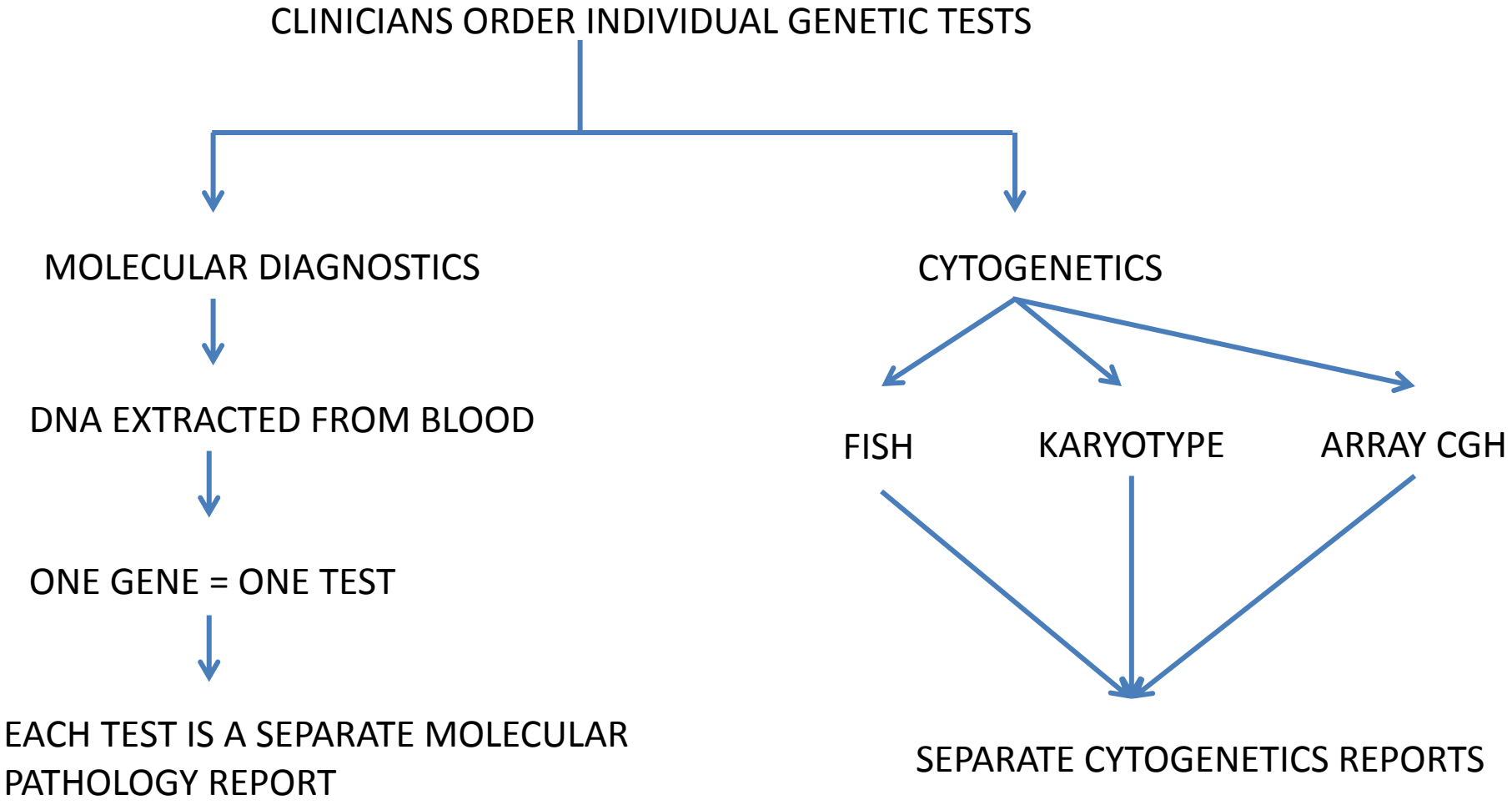
Erlotinib/Gefitinib

Dabrafenib

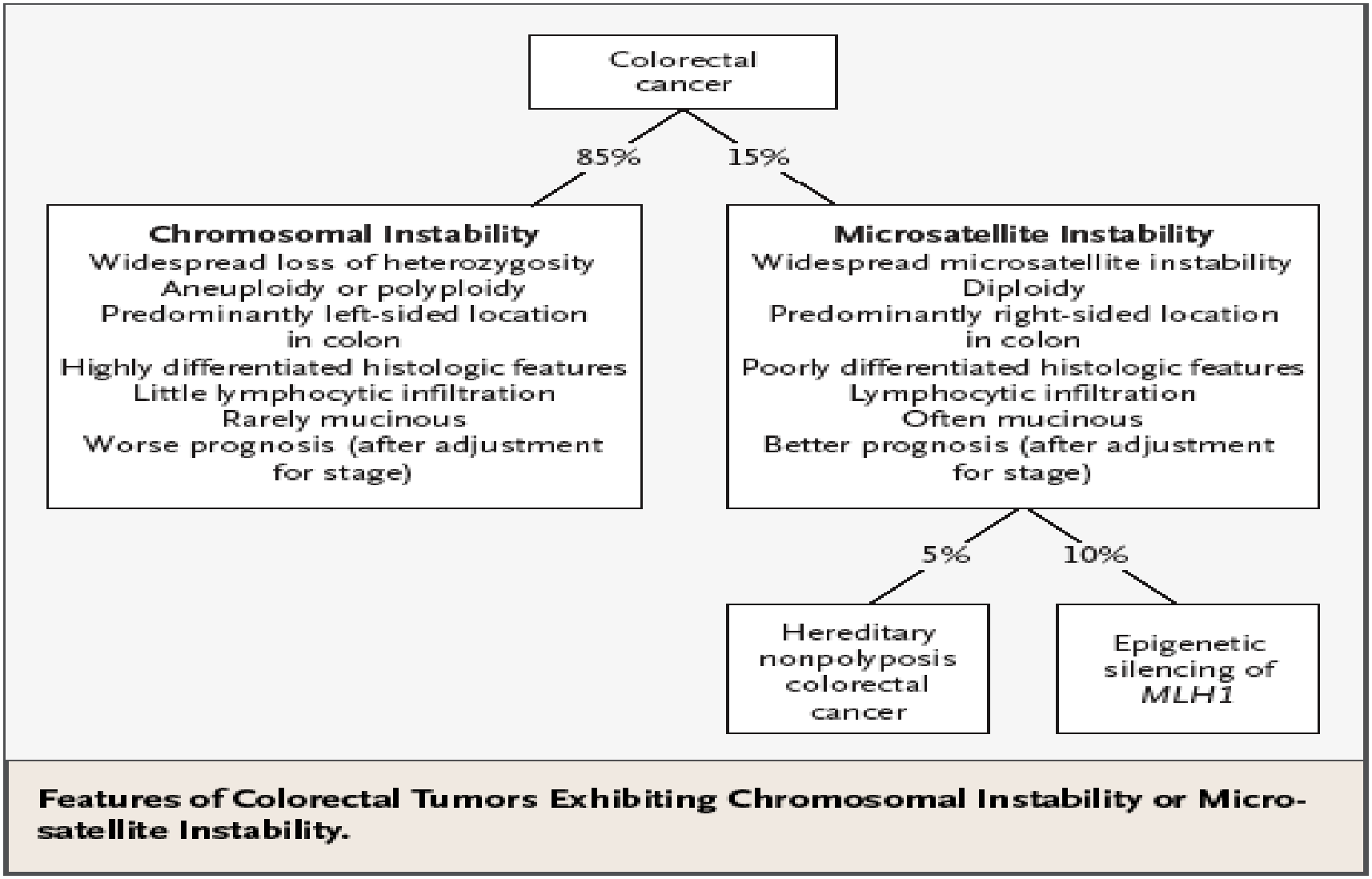
Crizotinib

Kris, et al ASCO 2011

CURRENT MOLECULAR TESTING SCHEME FOR INHERITED DISEASE

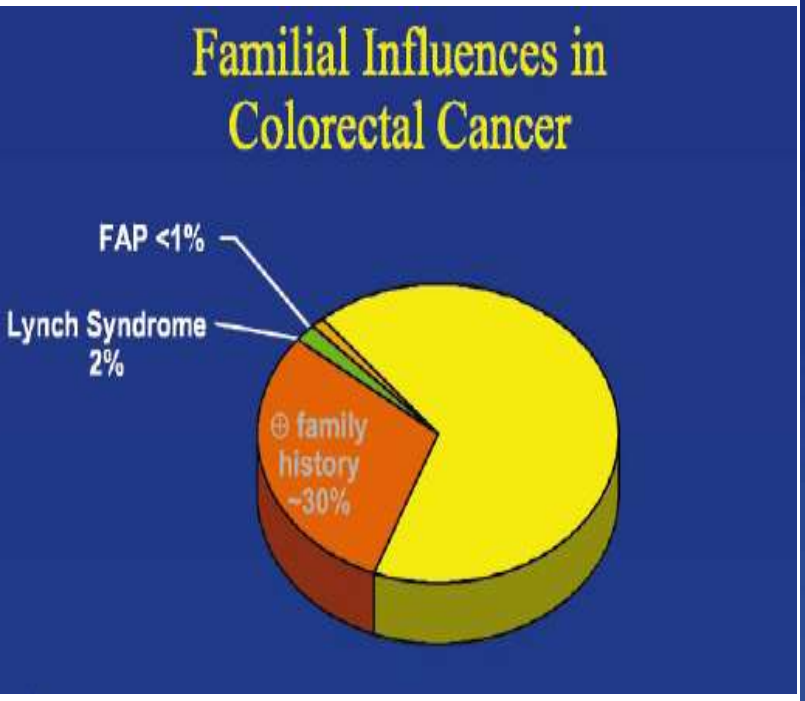


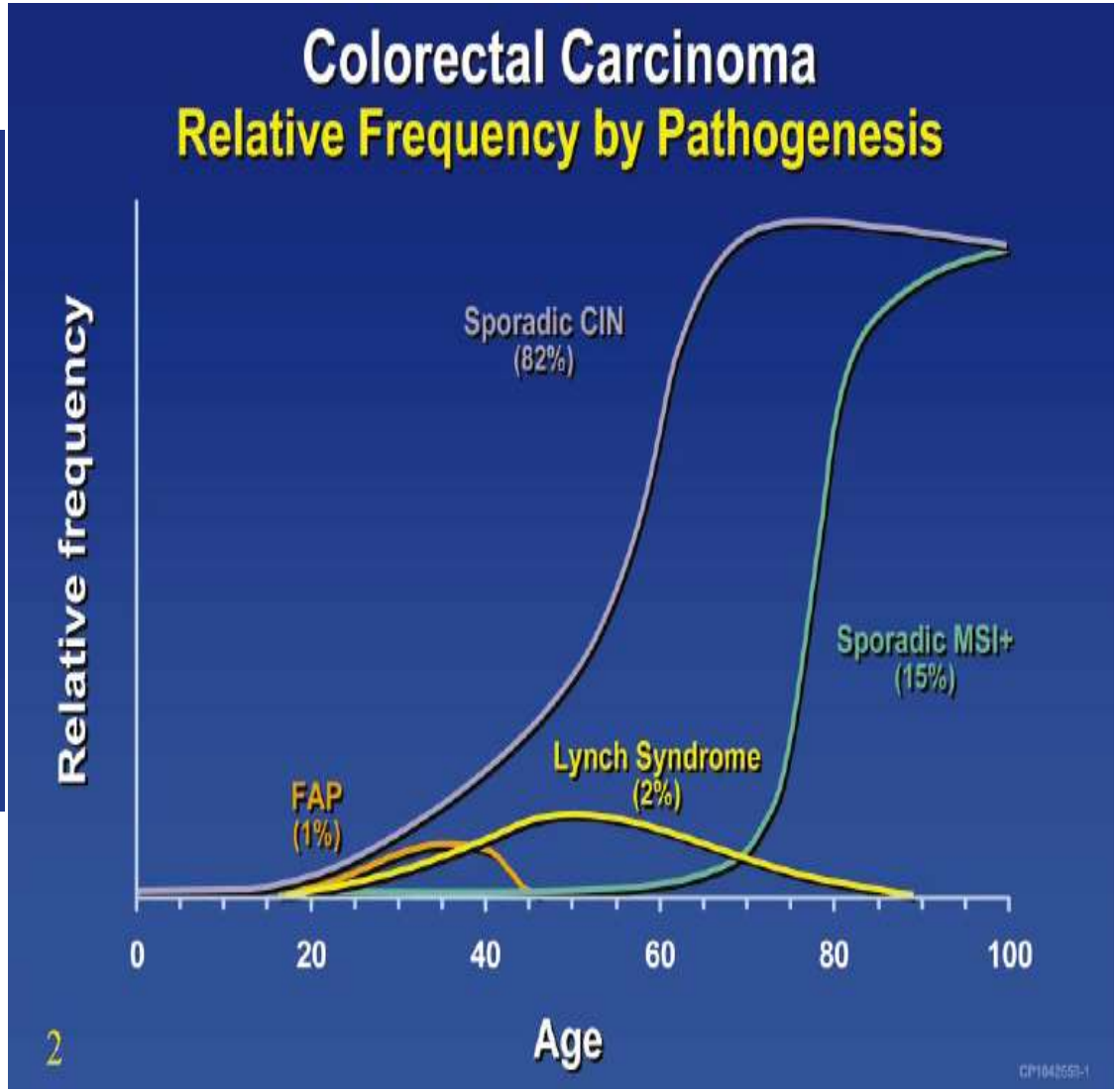
DISTRIBUTION OF MSI vs. MSS COLON CANCERS



Features of Colorectal Tumors Exhibiting Chromosomal Instability or Microsatellite Instability.

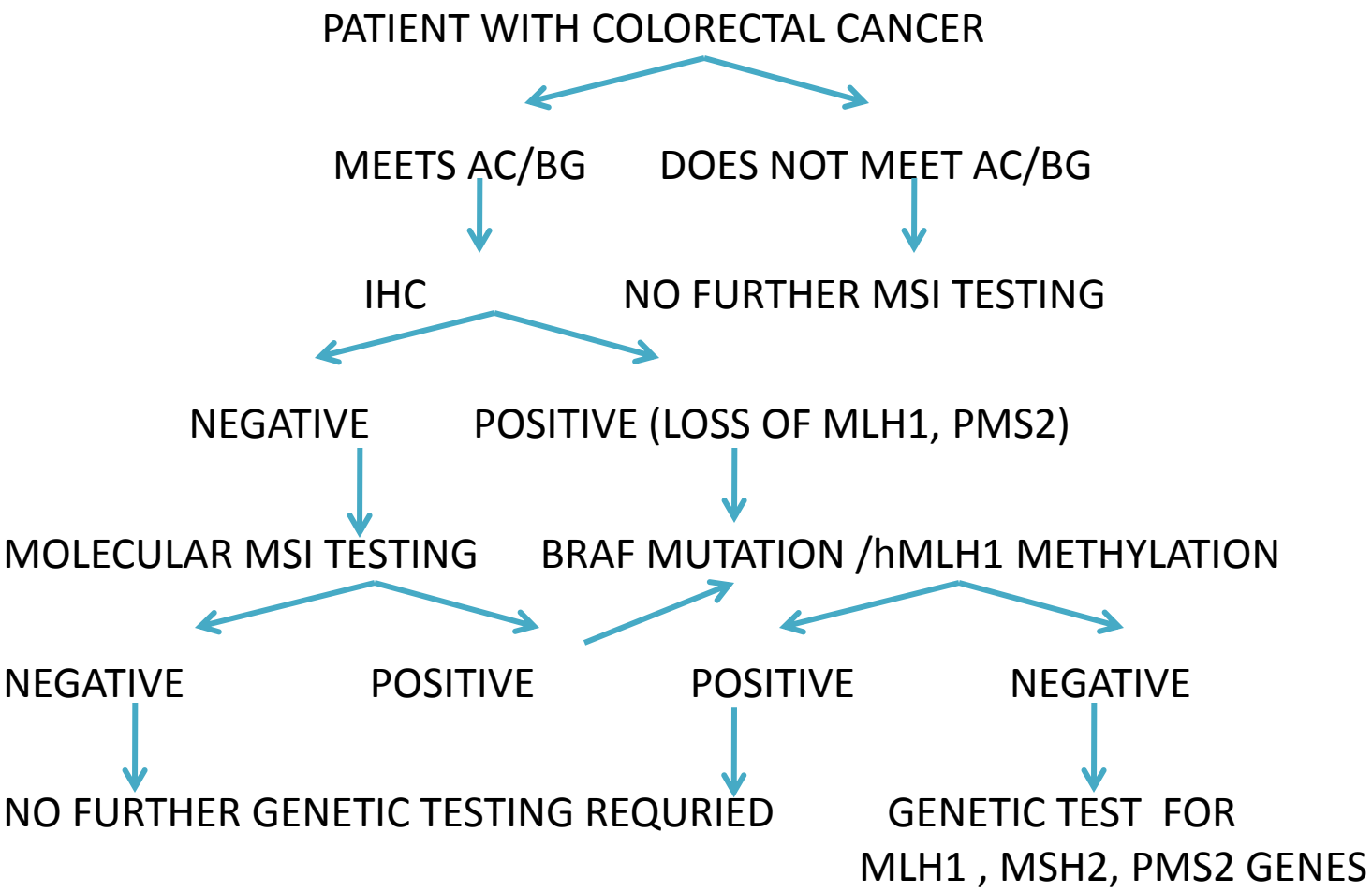
TESTING FOR LYNCH SYNDROME





(*Arch Pathol Lab Med.* 2005;129:1385–1389

CURRENT TESTING ALGORITHM



LIMITATIONS OF CURRENT TESTING PARADIGM

- INHERITED DISORDERS
  - Comprehensive genetic testing for several syndromes frequently involve simultaneous testing for several genes
  - Increasing demand for detection of point mutations and structural genetic alterations within tested genes

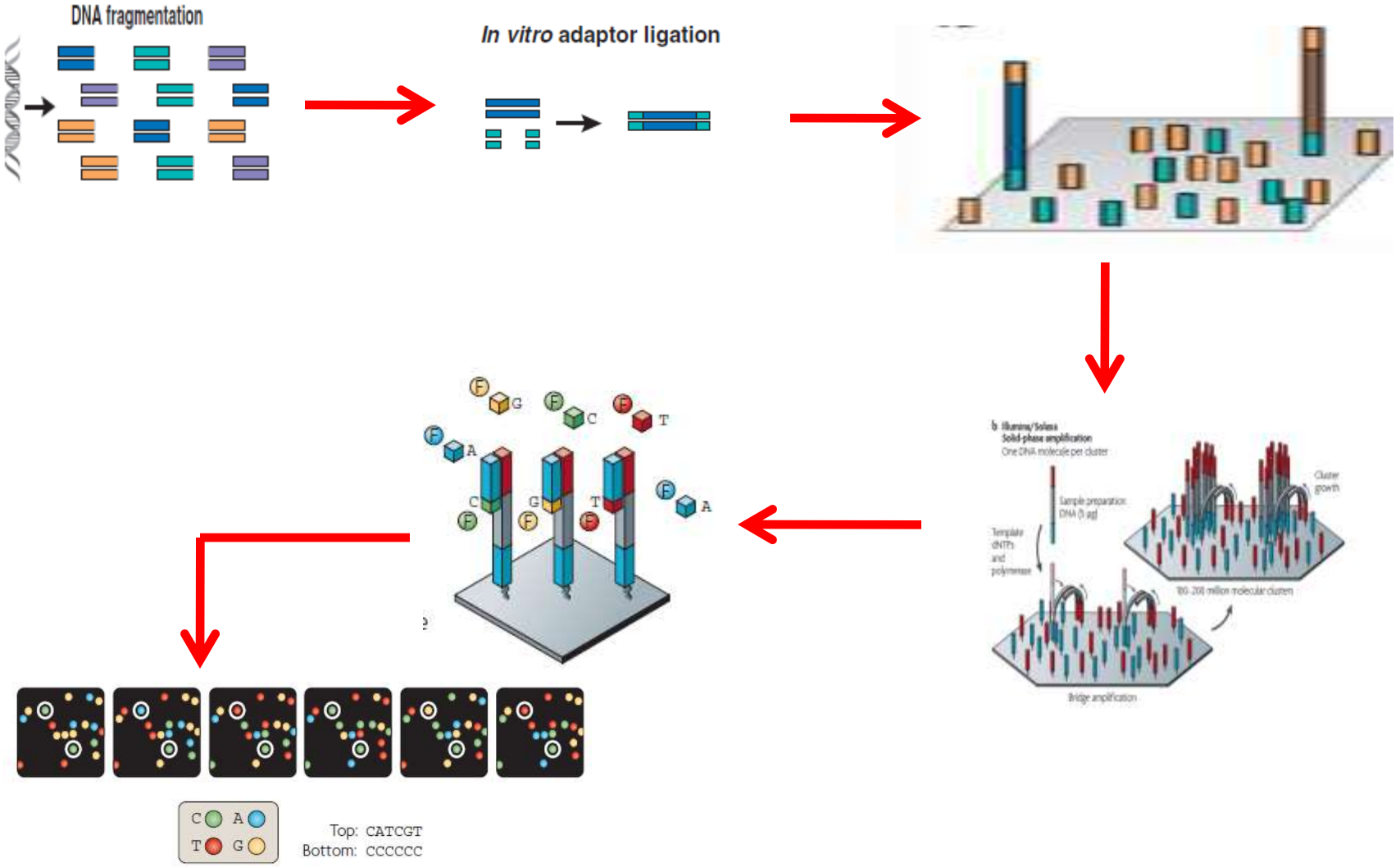
- CANCER DIAGNOSTICS
  - Comprehensive prognostic and predictive testing in near future will involve testing at least a few dozen genes
  - Various types of genetic alterations (point mutations, translocations etc.) will need to be evaluated simultaneously
  - Limited amount of sample available will be available for testing

PROPOSED SOLUTION

- NEXT GENERATION SEQUENCING technology was specifically designed to simultaneously evaluate variation in several genes
- This technology can also be used to detect different types of genetic alterations

- TYPES OF SEQUENCERS
  - HiSeq 2000/2500
  - Desktop sequencers: MiSeq/IonTorrent

TECHNOLOGY: NEXT GENERATION SEQUENCING



WHOLE GENOME VS. TARGETED CAPTURE

Genomic DNA library

Whole-genome sequencing

Sequence Reads

For example, 3 Mb region of interest in Human genome

Sequence entire 3 Gb genome to get it…

One flow cell

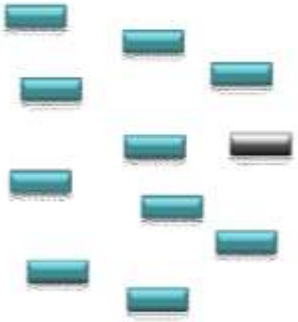
1 sample

- 8 days instrument time
- $15k sequencing cost
- 30x coverage

$5,000/sample

Genomic DNA library

Sequence Enrichment



x-fold Enrichment of Reads

Enrich and sequence 3 Mb only…

One lane of flow cell

100 samples

- 8 days instrument time
- $21k sequencing cost
- 30x coverage

$210 / sample

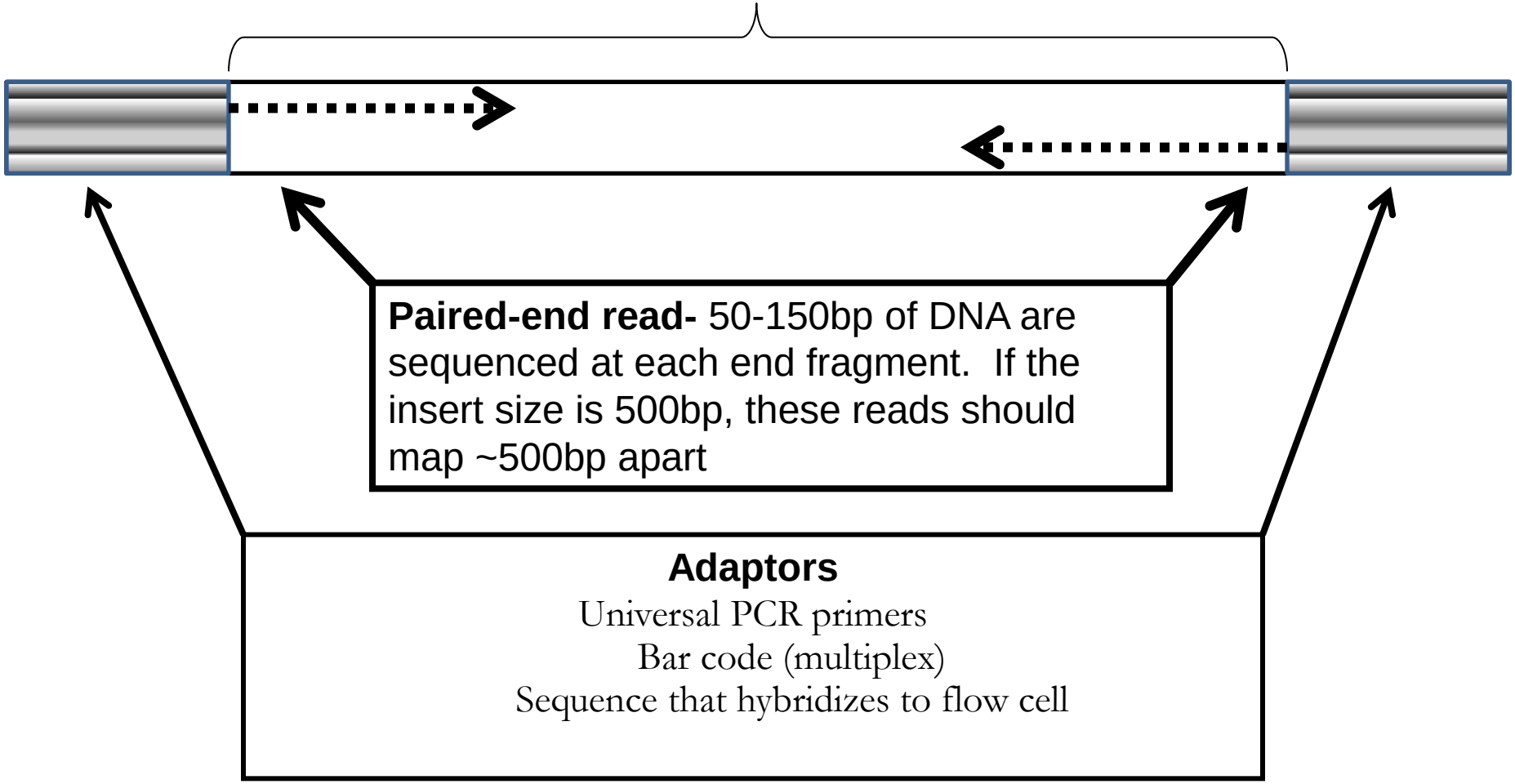
MAJOR STEPS OF NGS

- DNA library preparation
- Target enrichment
- Cluster generation & sequencing (Illumina HiSeq 2000)
- Bioinformatics analysis of sequence data
- Data interpretation

LIBRARY PREPARATION

Library = fragments of DNA that have been prepared for amplification and sequencing

Genomic DNA fragment 150-600bp in length
The size of the fragment is called the "**insert size**"

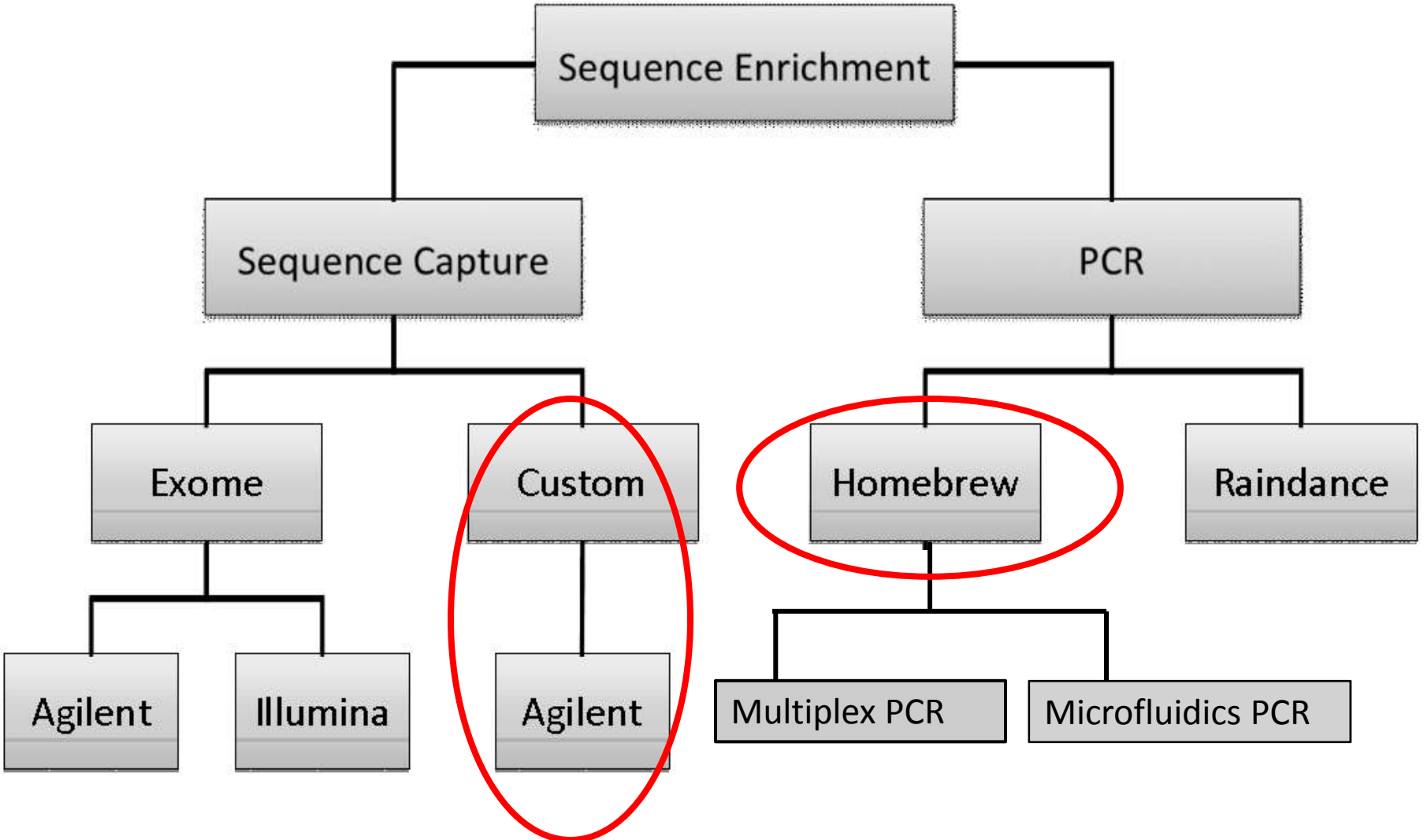


Paired-end read- 50-150bp of DNA are sequenced at each end fragment. If the insert size is 500bp, these reads should map ~500bp apart

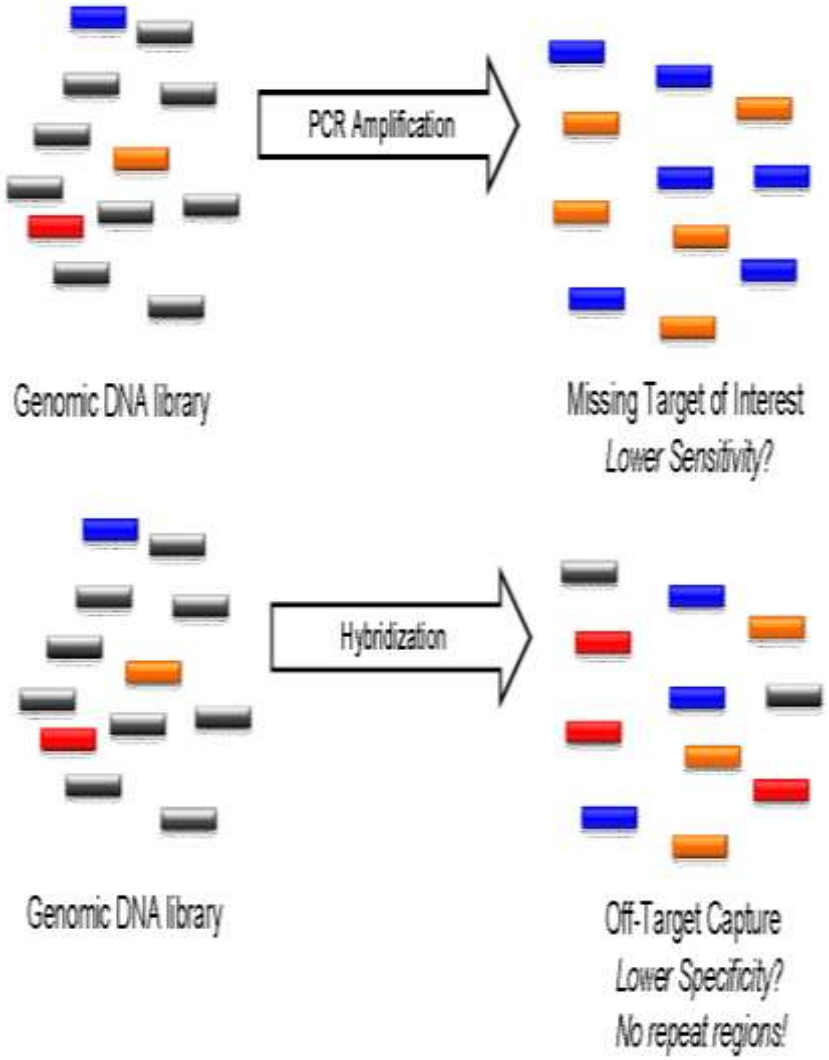
Adaptors

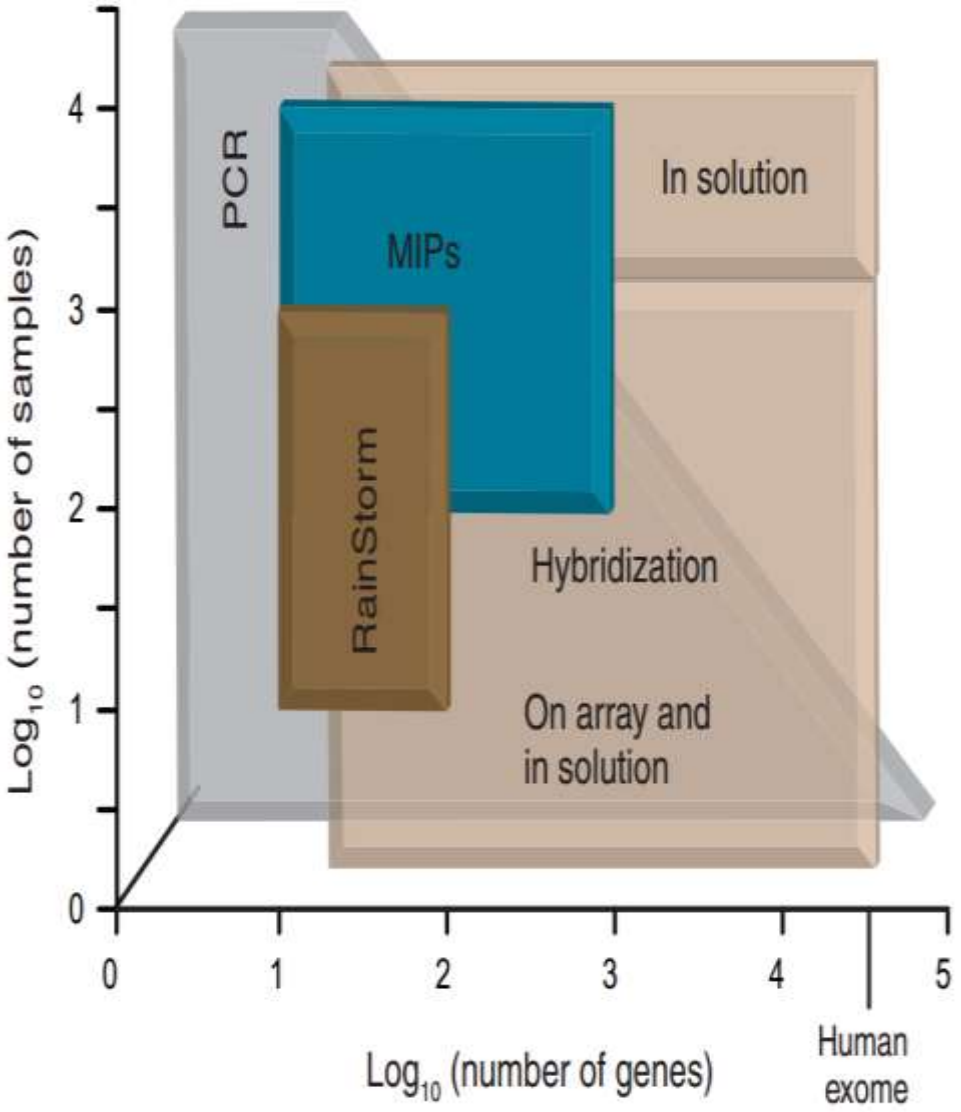
Universal PCR primers
Bar code (multiplex)
Sequence that hybridizes to flow cell

SEQUENCE ENRICHMENT OPTIONS

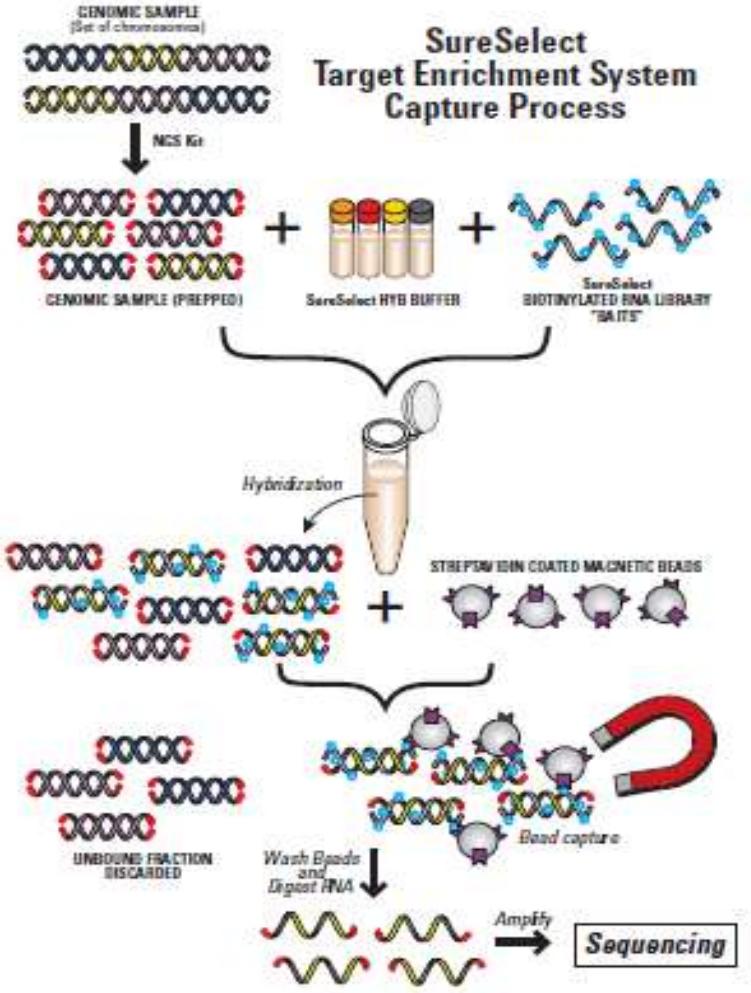


PCR VS. SEQUENCE CAPTURE





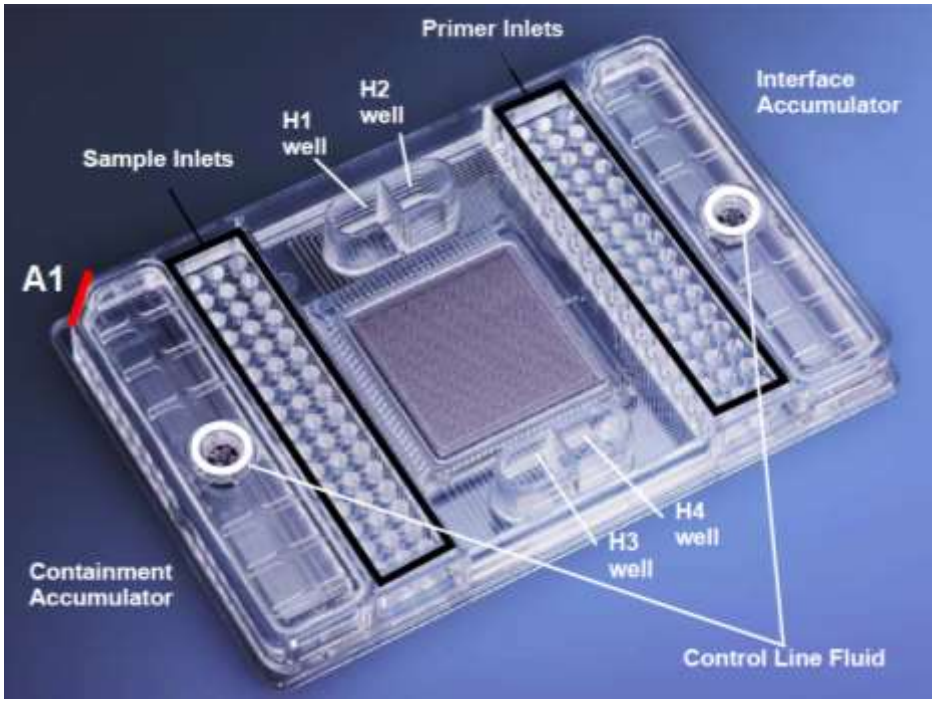
SEQUENCE CAPTURE



- Up to 80% enrichment for the targeted DNA
- 120 bp “baits” bind to DNA and magnetic beads
- Unbound DNA is discarded
- Baits are digested

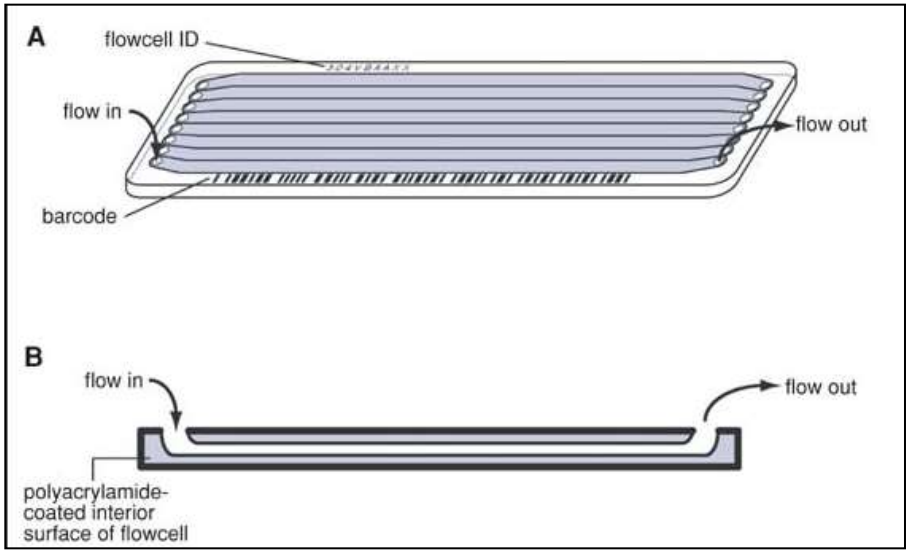
METHODOLOGIES FOR PERFORMING PCR BASED ENRICHMENT

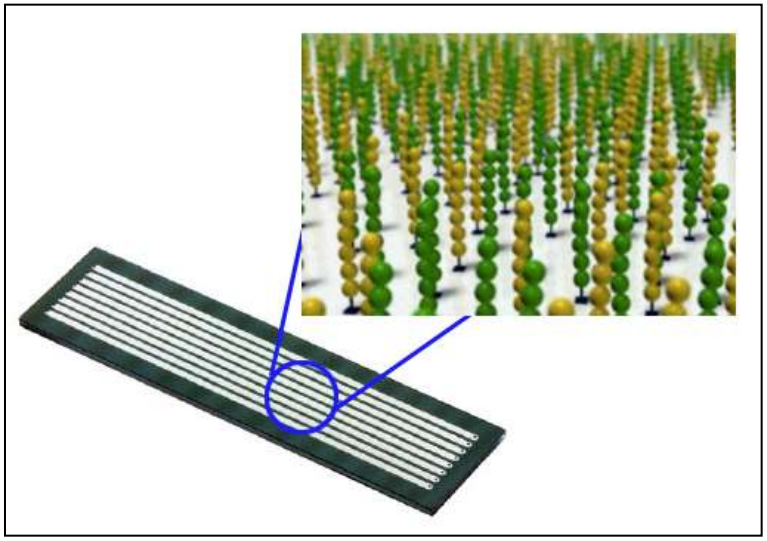
- Uniplex PCR: 1 reaction = 1 amplicon
- Multiplex PCR: 1 reaction = 10-50 amplicons
- Droplet PCR: 1 reaction = 4,000 amplicons
- Microfluidics PCR
  - Multiplex 10 PCR/well
  - Can simultaneously amplify 480 amplicons



CLUSTER GENERATION AND SEQUENCING

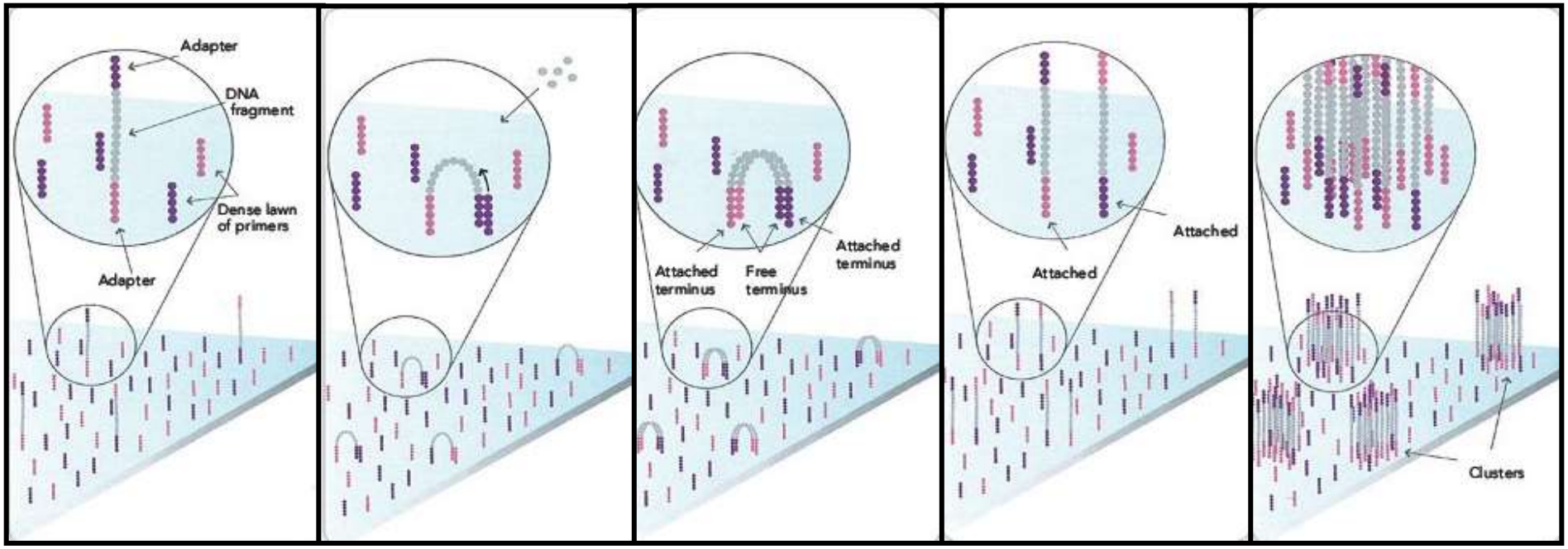
- Oligonucleotides attached to flow cell hybridize to the adaptors
- Individual DNA library fragments are immobilized





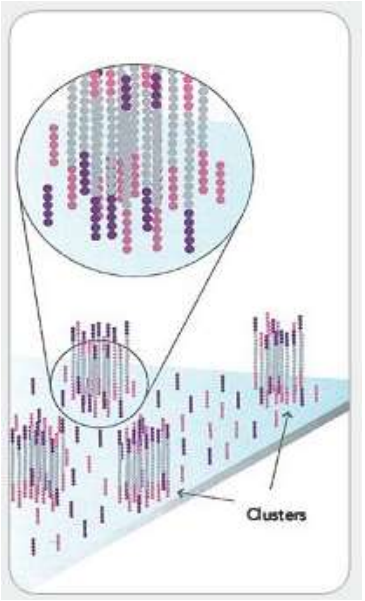
CLUSTER GENERATION AND SEQUENCING

- Starting DNA template concentration is crucial to avoid overcrowding of clusters
- Each unique DNA molecule undergoes “bridge amplification”



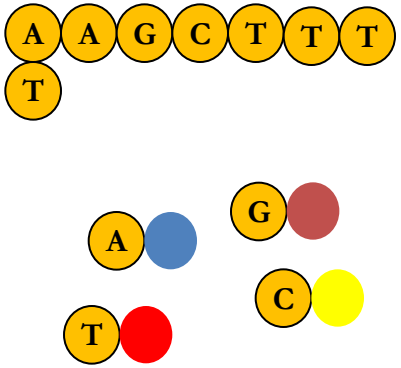
CLUSTER GENERATION AND SEQUENCING

- Simultaneous generation of millions of clusters (“polonies”)
- One cluster:
  - Derives from a single parent DNA molecule
  - Made up of ~1000 identical copies
  - Unique
  - Physically isolated from other clusters

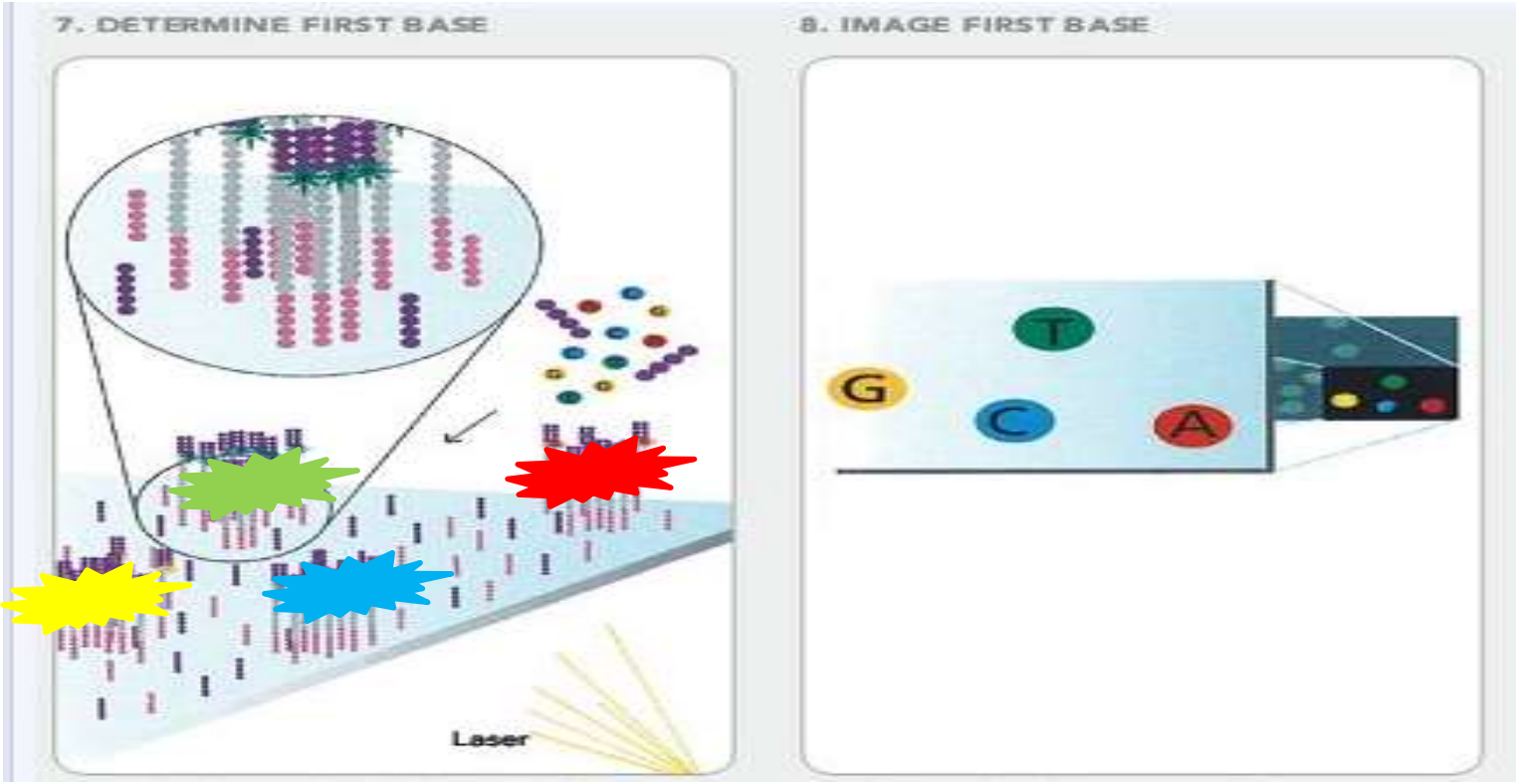


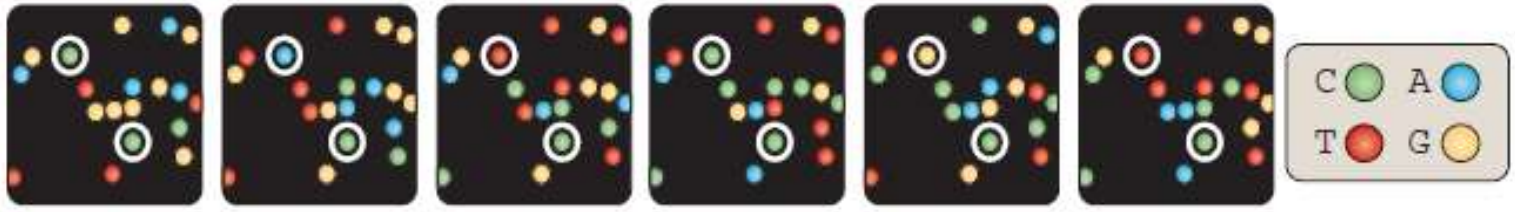
SEQUENCING BY SYNTHESIS

- All clusters are sequenced in parallel, one base at a time
- Fluorescently tagged nucleotides compete for next space
- Fluorescent tag blocks addition of more than 1 nucleotide per round
- Each round
  - Addition of one base
  - Laser excitation -> fluorescence
  - One “base” read from each cluster
  - Removal of fluorescent tag



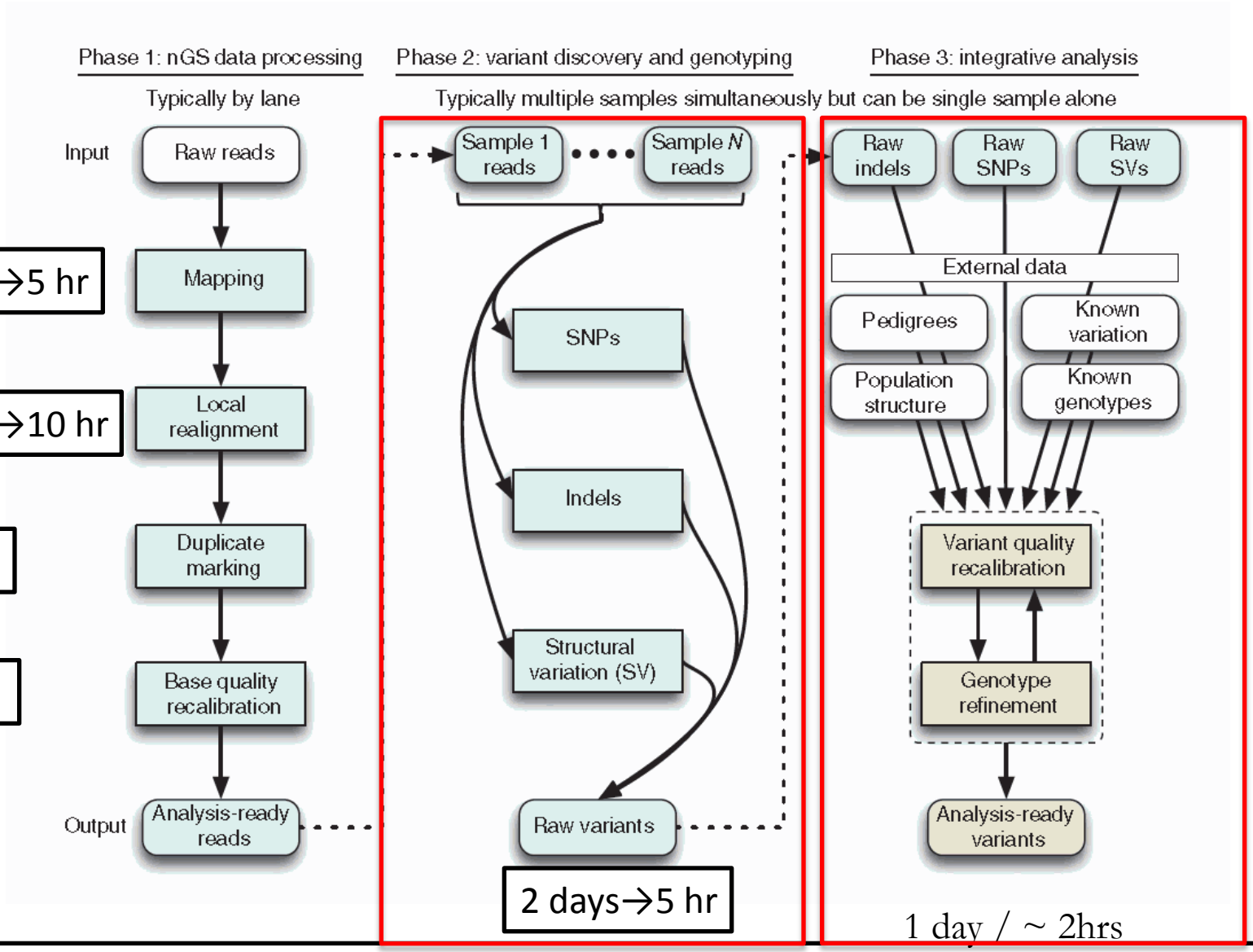
SEQUENCING BY SYNTHESIS





Top: CATCGT
Bottom: CCCCCC

BIOINFORMATICS ANALYSIS



INHERITED DISEASES VS. ONCOLOGY

	INHERITED DISEASES	ONCOLOGY
NUMBER OF GENES	Range: 1-150	Range: 1-50
MUTATION DISTRIBUTION	Across entire gene	Hotspot mutations/deletions /translocations
MINIMUM COVERAGE	20X	~250X – 1000X
STRUCTURAL VARIATION	Existing methods work well	Sensitivity depends on tumor percentage
TISSUE TYPES	Blood	Blood, Fresh frozen, FFPE, cytology
INPUT DNA	3 µg	5 ng – 1µg
TURNAROUND TIMES	4-6 weeks	5-7 days
COST	Range: $1000 - $10,000	Range: $400 - $1000

TESTING SCHEME FOR INHERITED DISEASES

SAMPLE PREPARATION

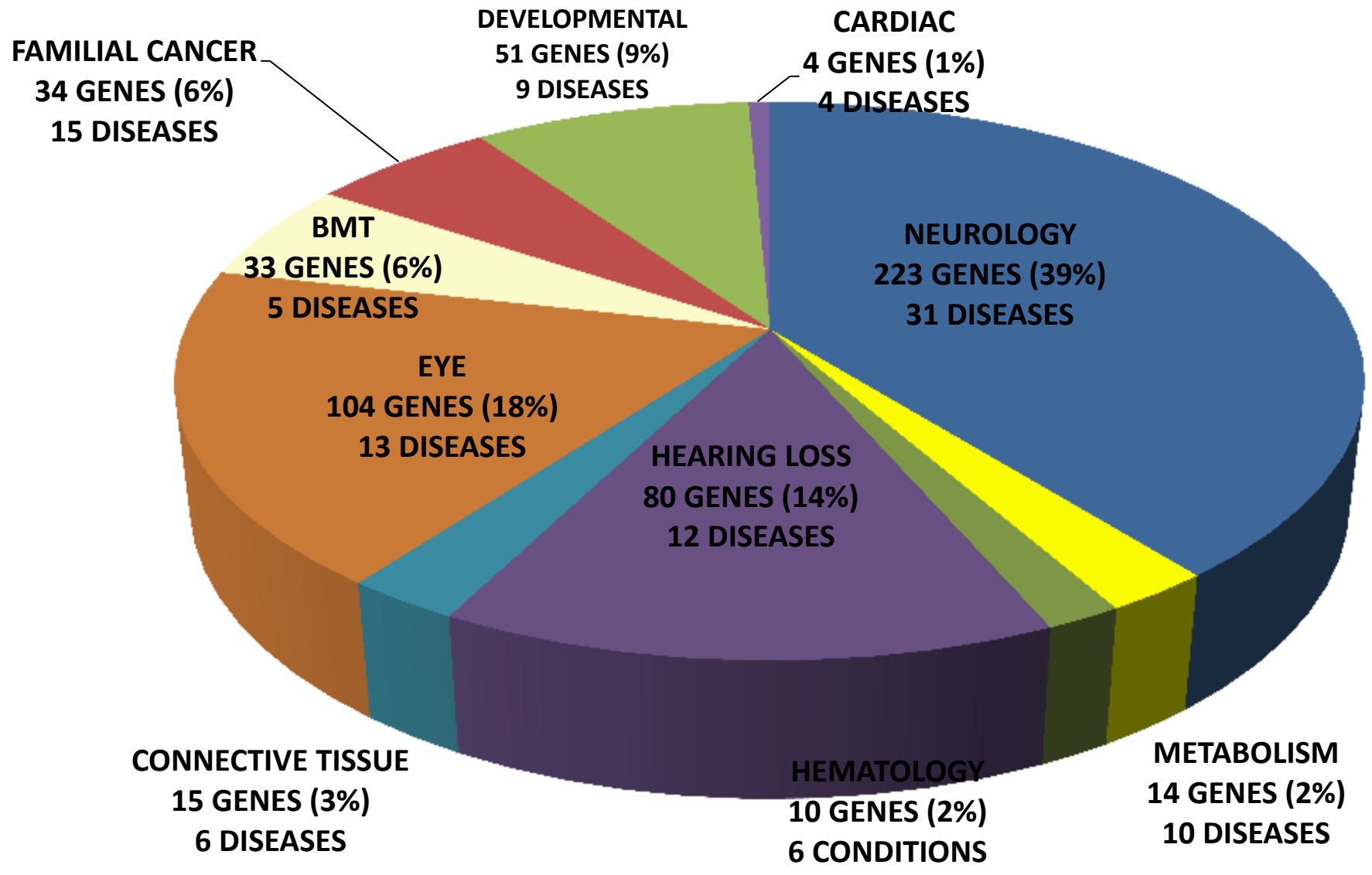
TARGET CAPTURE

SEQUENCING

SMALL PANELS
RAPID TURNAROUND TIMES
(5-7 DAYS)

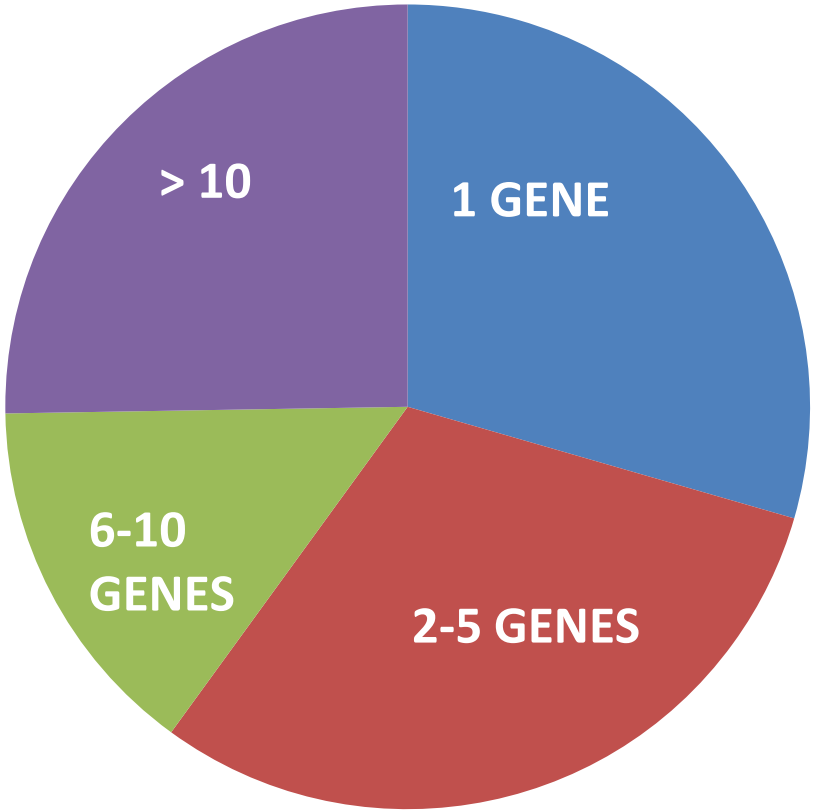
LARGE PANELS
EXOME SEQUENCING
SLOW TURNAROUND TIMES
(8-10 WEEKS)

INHERITED DISEASES:NGS TESTING

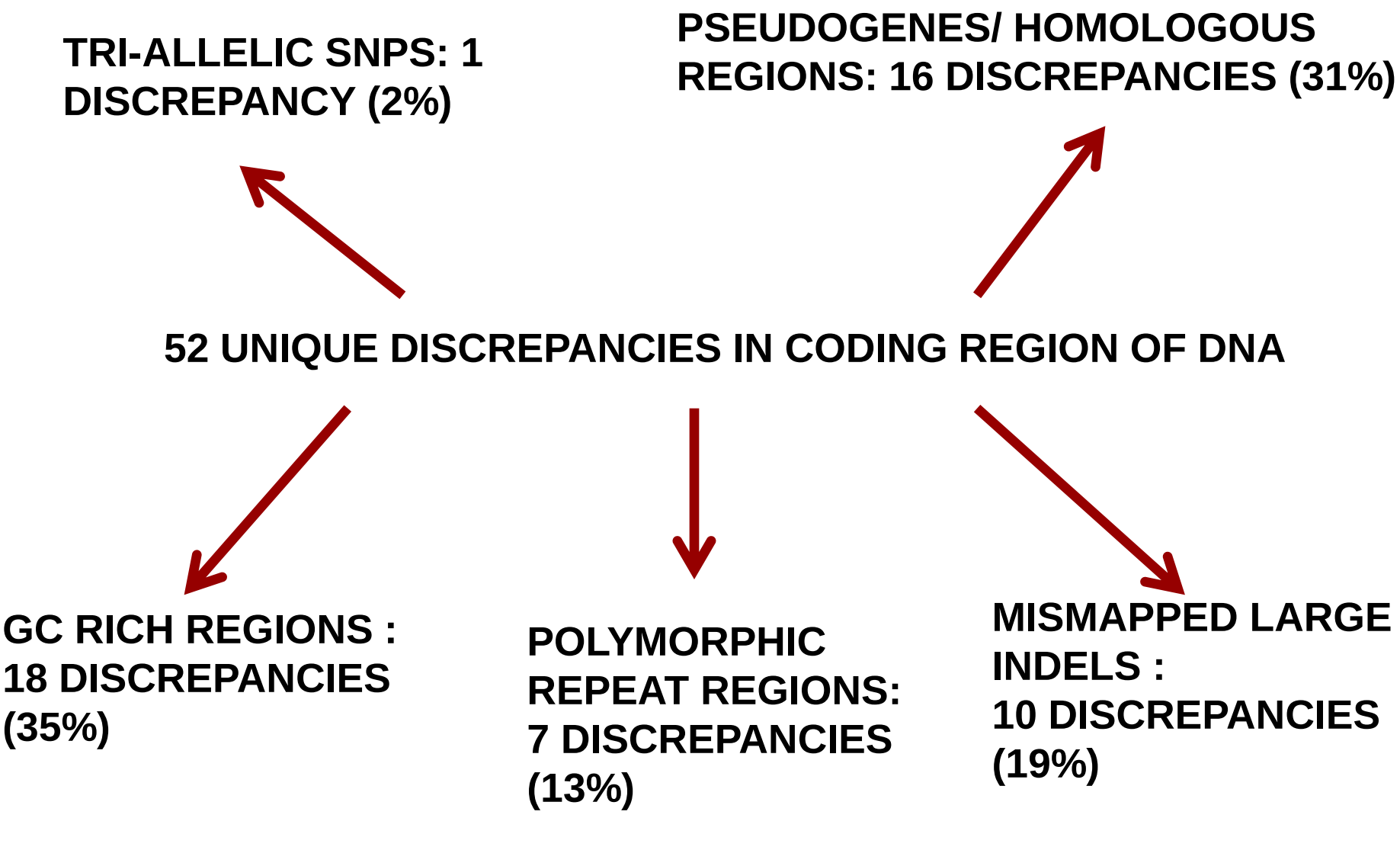


NGS EXPERIENCE AT MDL

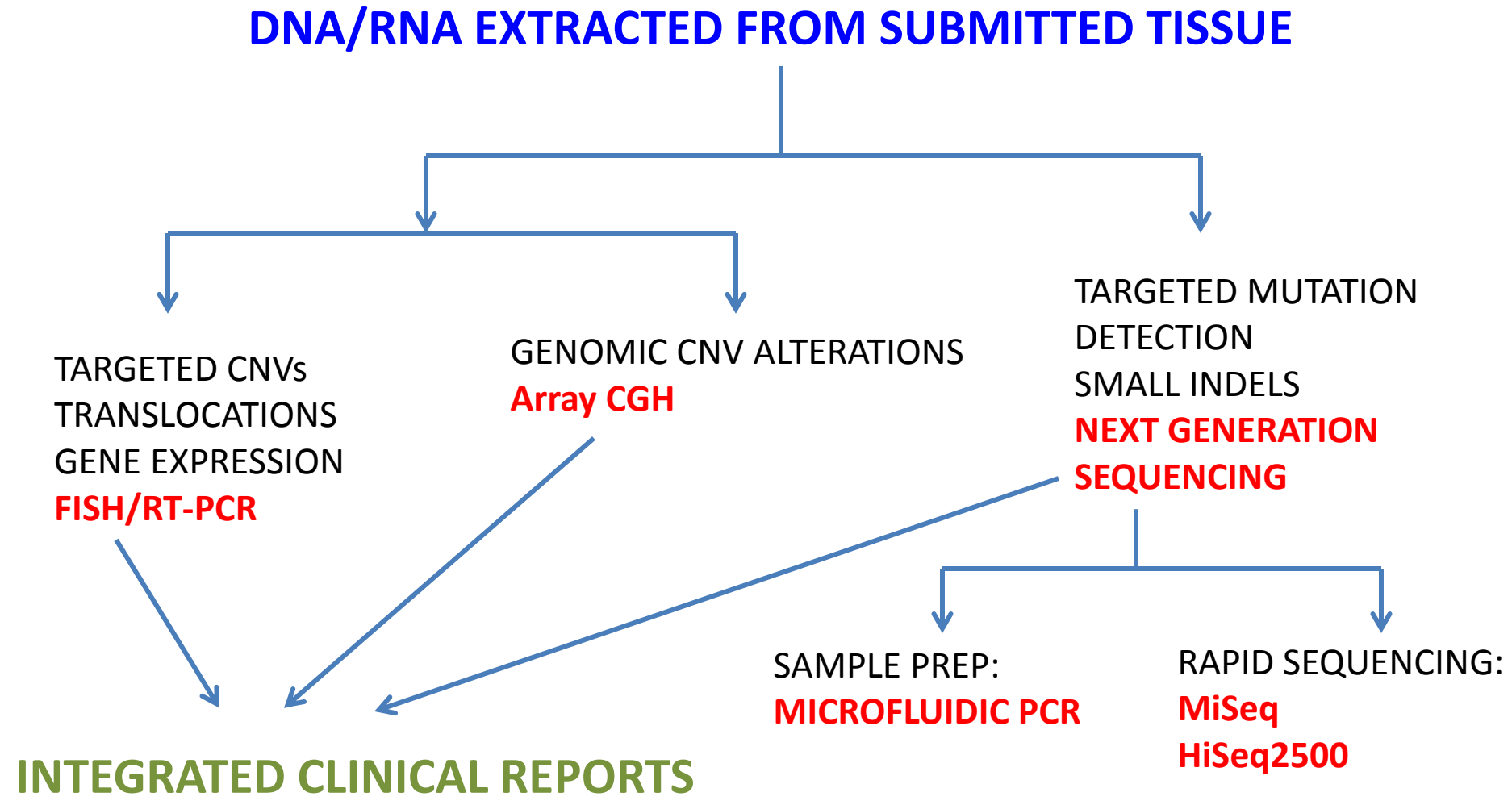
- Have offered NGS testing for over 130 Mendelian disorders since August 2012
  - We have tested 300 samples
  - We have detected mutations in approximately 30% of all samples tested
  - Mutation detection rate is dependent on clinical diagnosis
    - Mutation identified in 80% of inherited thrombophilias
    - Mutation identified in 25% of ataxias
    - No mutations identified in disorders of sexual development



GENOMIC REGIONS THAT ARE PROBLEMATIC FOR NGS



TESTING SCHEME FOR ONCOLOGY



ONCOLOGY:NGS TESTING

Lung Cancer Panel

- **Somatic mutation testing**
  - *KRAS (NRAS/HRAS)*
  - *EGFR*
  - *BRAF*
  - *PIK3CA*
  - *ERBB2*
  - *MET*
  - *TP53*
  - *AKT1*
  - *MAP2K1*
  - *EGFRvIII* (RT-PCR assay)
- **Translocation**
  - *ALK* (EML4-ALK, but other partners up to 20)
  - *ROS* (up to 7 partners)
  - *KIF5B/RET*
  - *CCDC6/RET* (aka RET/PTC1)
- **Amplification**
  - *EGFR*
    - *MET*
    - *MAPK1* (p42/ERK2)
    - *FGFR1*
    - *FGFR2*

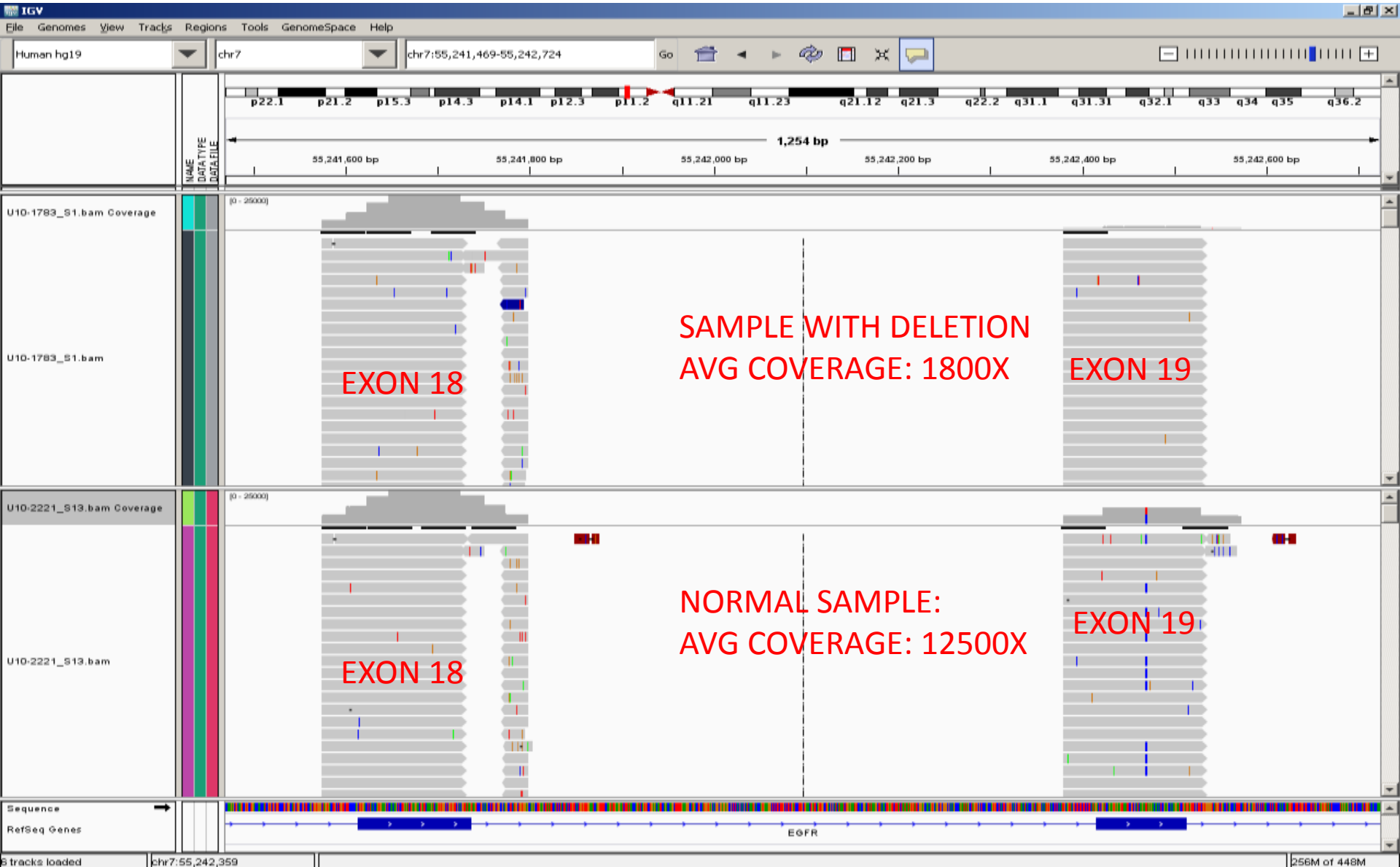
Gastrointestinal Cancer Panel

- **Somatic mutation testing**
  - *EGFR*
  - *KRAS (HRAS/NRAS)*
  - *BRAF*
  - *PIK3CA*
  - *TP53*
  - *ERBB2*
  - *MET*
  - *KIT*
  - *PDGFRA*
  - *AKT1*
  - *PTEN*
  - *APC*
- **Amplification/Deletion**
  - *ERBB2*
  - *IGF2* (11p15.5)
  - *PTEN*
  - *MDM2*
  - *EGFR* (rare)

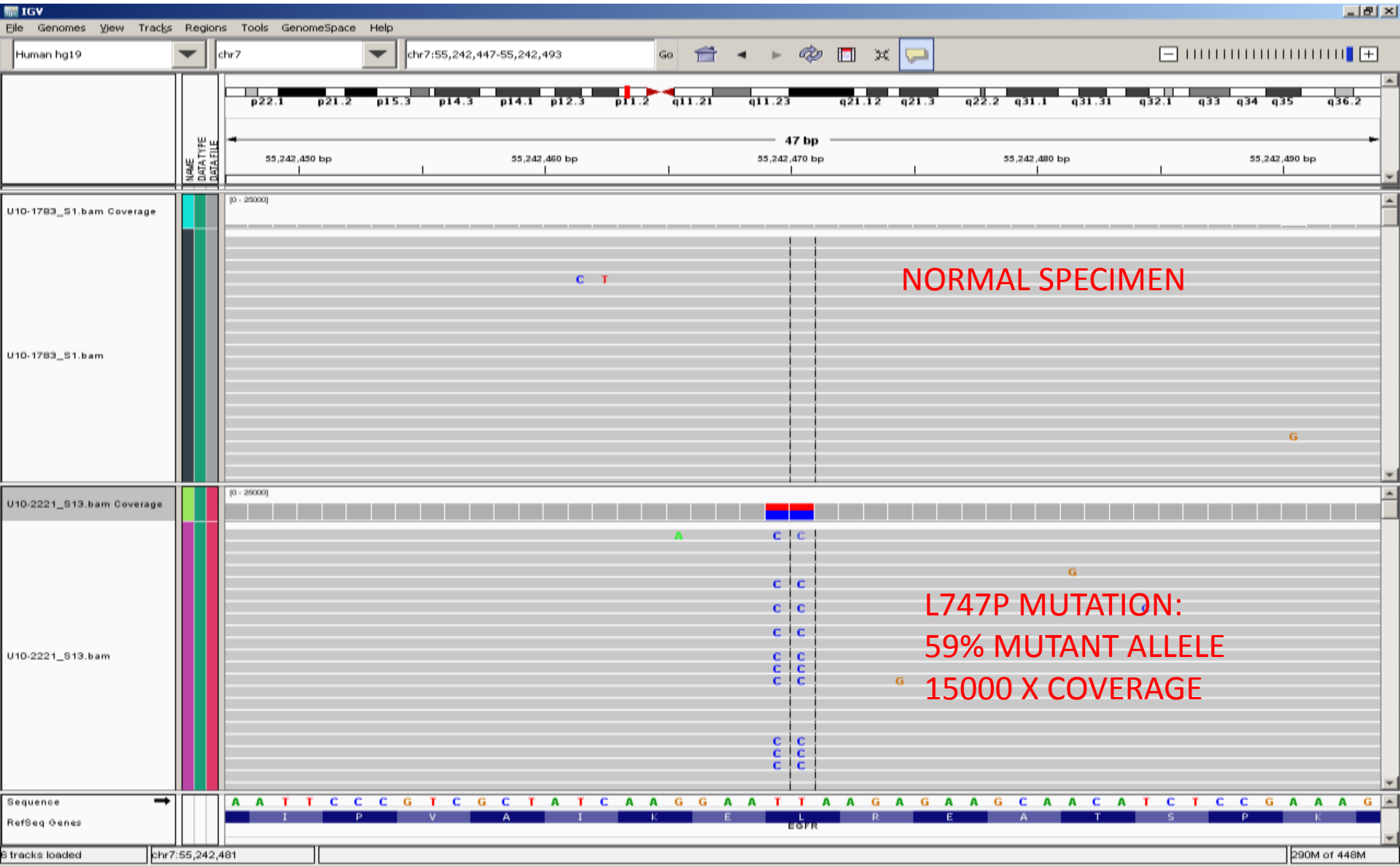
DETECTION OF EGFR MUTATIONS

- Targeted sequencing of exons 18-21 (visualizing 18-19)
- Input DNA: 50 ng of DNA for two lung cancer specimens
- Specimen 1 (exon 19 deletion)
  - 80% tumor
- Specimen 2 (L747P mutation due to sequential T>C mutations)
  - 70% tumor

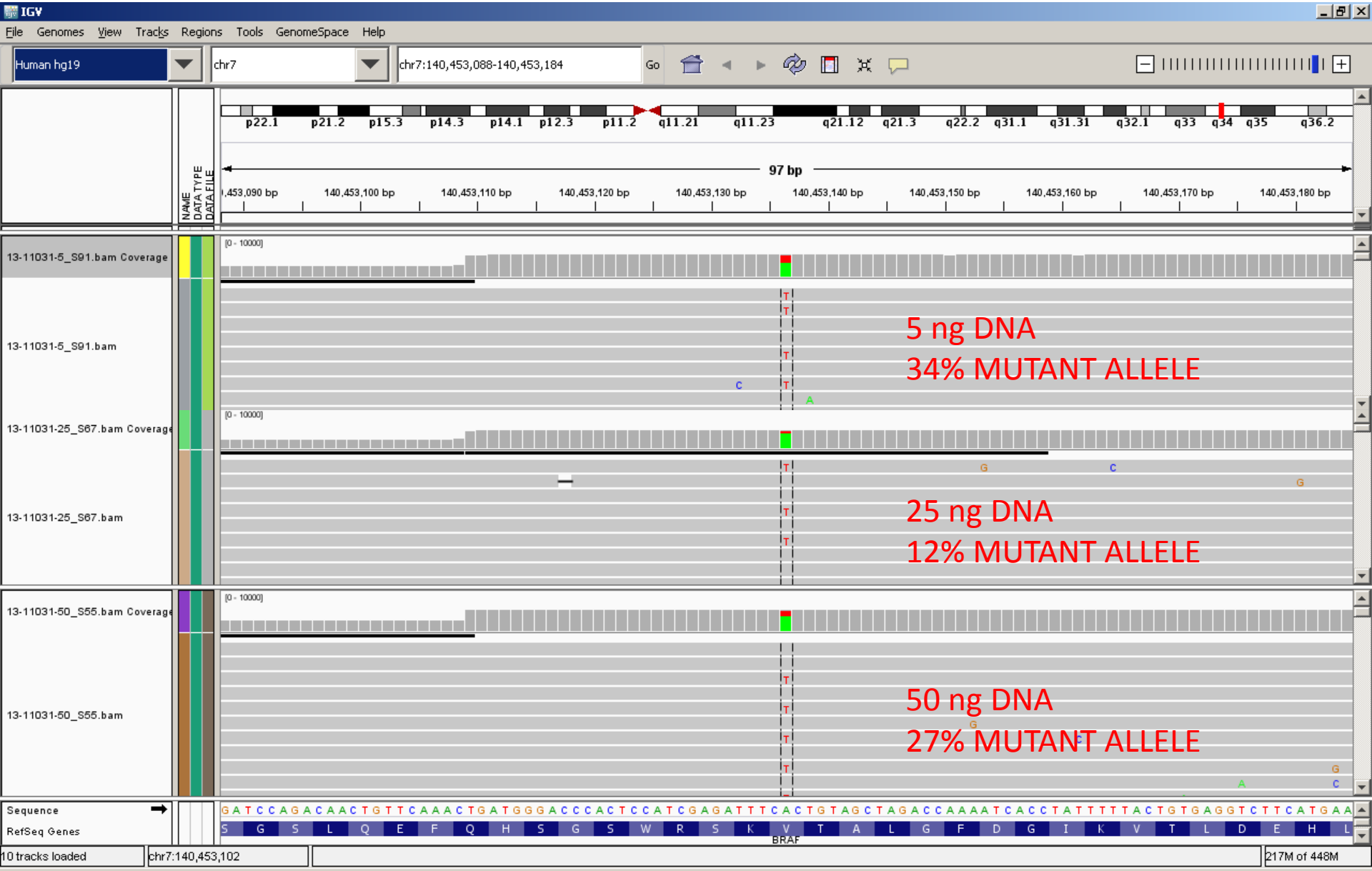
EGFR EXON 19 DELETION



EGFR L747P MUTATION



BRAF V600E MUTATION: MINIMUM DNA INPUT



TECHNICAL ISSUES WITH NGS IMPLEMENTATION

- Bioinformatics methods for sequence alignment keep undergoing rapid improvements
  - Need to update bioinformatics pipeline at frequent intervals
- Structural genetic variation:
  - Optimal algorithms for detection of copy number variation remain unclear
- Several regions with inadequate coverage
  - Backup Sanger sequencing/alternative methodology necessary for several exons in the context of inherited disorders
  - Sensitivity to detect somatic mutations will not be the same in all the analyzed regions

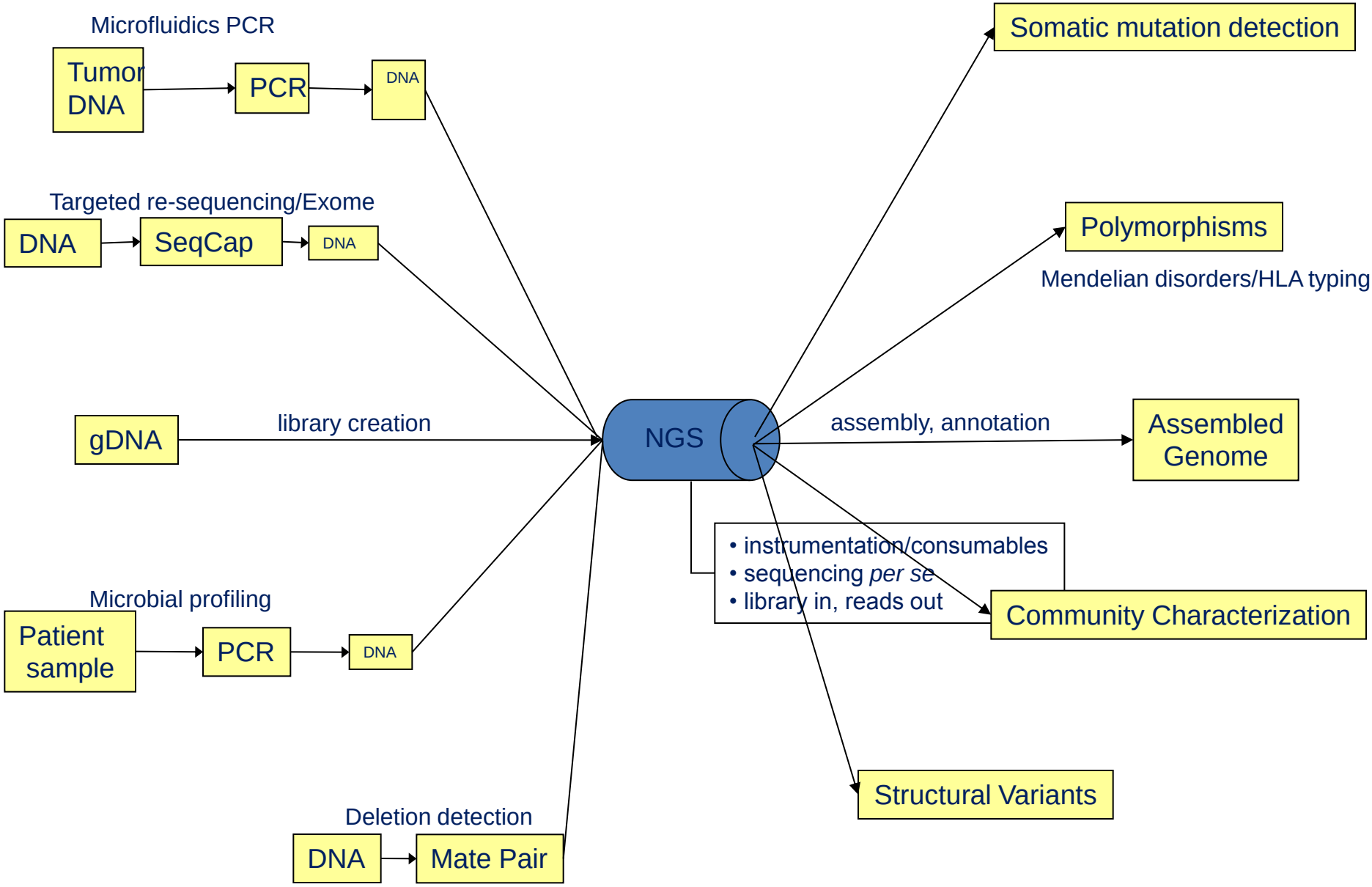
CLINICAL ISSUES WITH NGS IMPLEMENTATION

- Interpretation of clinical significance of many variants is unclear
  - Communication of these results to the clinician is problematic
  - Often results in additional testing of family members to determine clinical significance of a particular variant
- Incidental genetic findings need to reported and appropriate clinical follow up procedures need to be in place

OTHER ISSUES WITH NGS IMPLEMENTATION

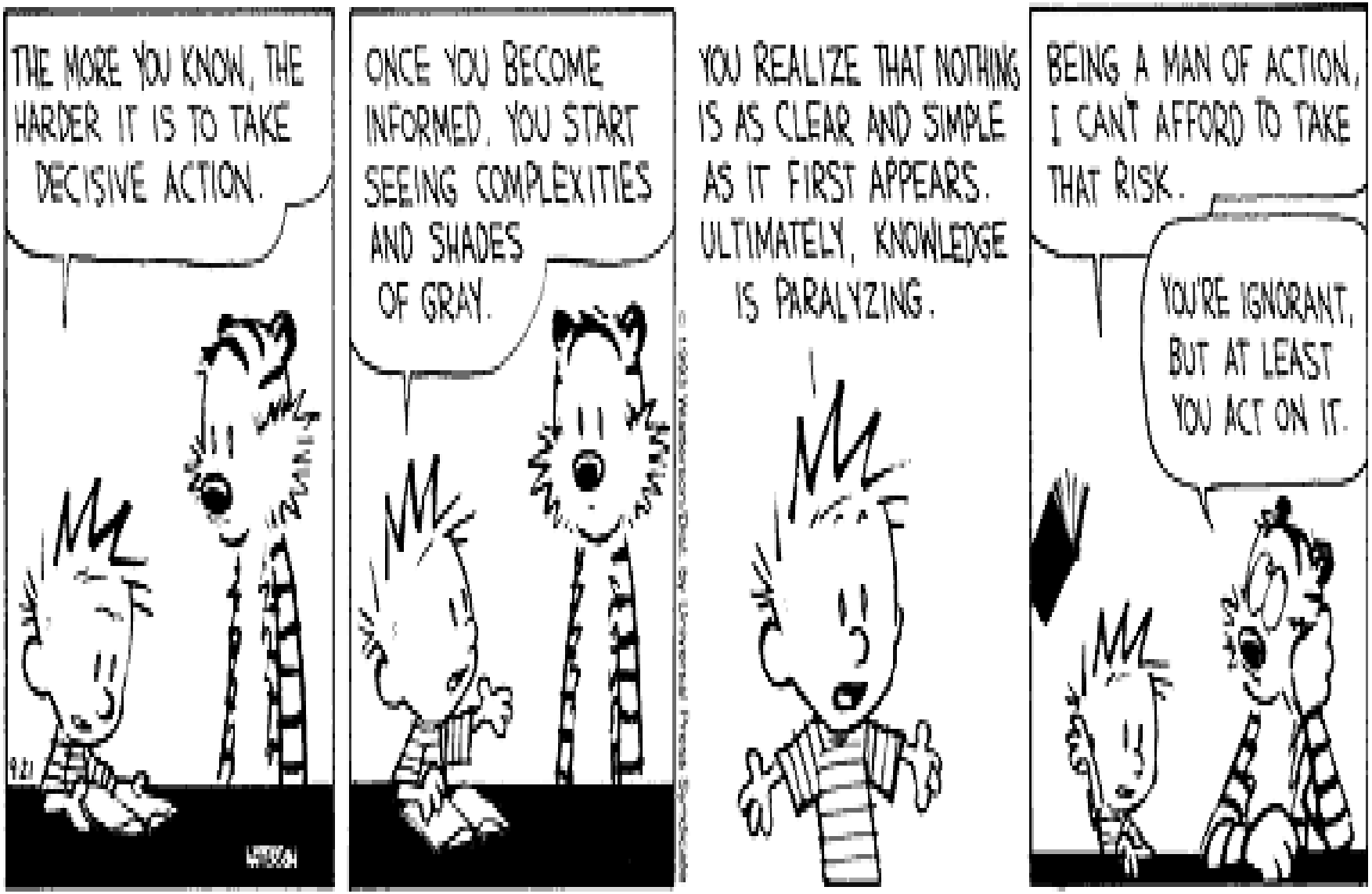
- High upfront costs for test validation
  - Substantial reagent costs
- High sequencing run costs
  - Need to batch samples to reduce assay costs
    - Need to offer a large test menu to increase sample volume
  - Limited ability to repeat samples
    - Robustness of assays need to be adequately validated

FUTURE APPLICATIONS OF NGS



ACKNOWLEDGEMENTS

- FUNDING SOURCES
  - Institute for Translational Neuroscience
  - Biomedical Genomics Center
- BIOMEDICAL GENOMICS CENTER
  - Kenneth Beckman, Karina Bunjer, Adam Hauge, Archana Deshpande
- BIOINFORMATICS CORE FACILITIY
  - Kevin Silverstein, Getiria Onsongo, Jesse Erdmann
- MOLECULAR DIAGNOSTICS LABORATORY
  - Matt Bower, Teresa Kemmer, Matt Schomaker, Sophia Yohe, Jon Wilson, Michael Spears, Andrew Nelson
- FAIRVIEW
  - Klint Kjeldahl, Karin Libby


THE MORE YOU KNOW, THE HARDER IT IS TO TAKE DECISIVE ACTION.
ONCE YOU BECOME INFORMED, YOU START SEEING COMPLEXITIES AND SHADES OF GRAY.
YOU REALIZE THAT NOTHING IS AS CLEAR AND SIMPLE AS IT FIRST APPEARS. ULTIMATELY, KNOWLEDGE IS PARALYZING.
BEING A MAN OF ACTION, I CAN'T AFFORD TO TAKE THAT RISK.
YOU'RE IGNORANT, BUT AT LEAST YOU ACT ON IT.