National Cancer Institute

A Genomic Portrait of Tumor Progression Using Second Generation Sequencing

Peter Johansson, Ph.D.
Pediatric Oncology Branch
National Cancer Institute, NIH

Pediatric
ONCOLOGY BRANCH

National Cancer Institute

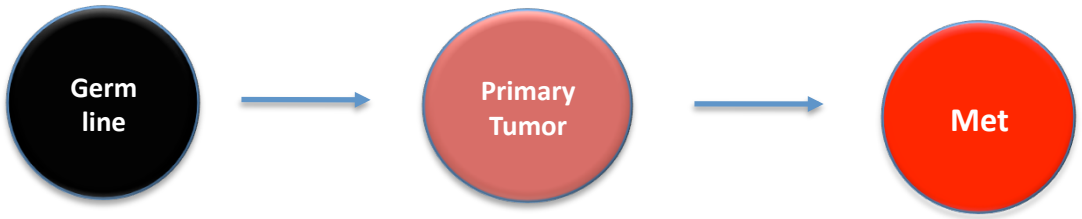
Neuroblastoma

- Most common extracranial solid tumor of childhood.
  - ~ 8 % of all childhood cancers
  - ~ 15% of deaths from childhood cancers
  - ~ 600 new cases in the U.S. annually

- Arises in Sympathetic Nervous System

National Cancer Institute

Objective

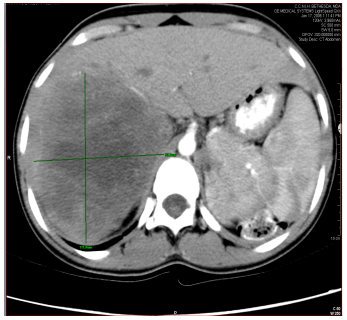


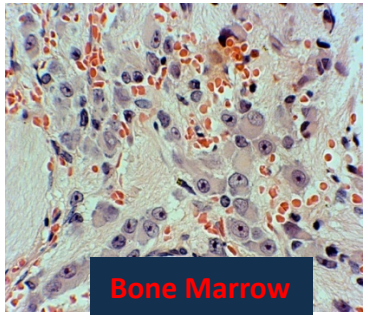
- **To portray progression of a neuroblastoma genome**
- **Single Nucleotide Variants**
- **Chromosomal Aberrations**

National Cancer Institute

Patient with Metastatic Neuroblastoma

Diagnosis

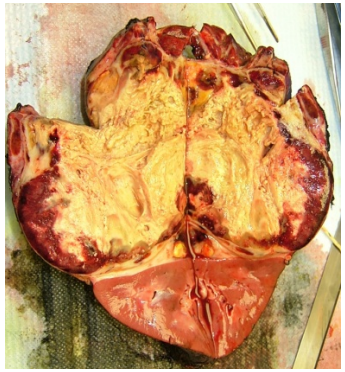


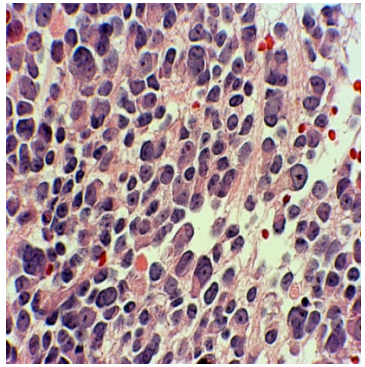


1. Met1-BM: Bone marrow biopsy at diagnosis

~4 Months

Surgery

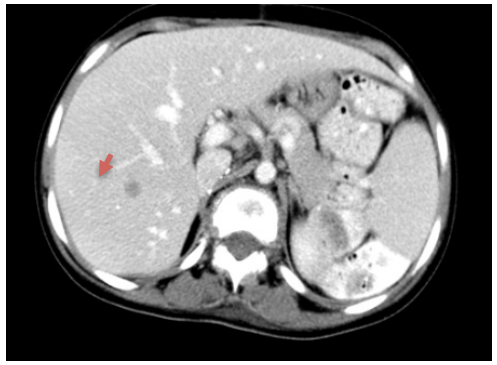




2. Primary: Primary tumor removed by surgery

3 years

Death



3. Met2-Liver: Liver metastasis at autopsy

National Cancer Institute

Outline

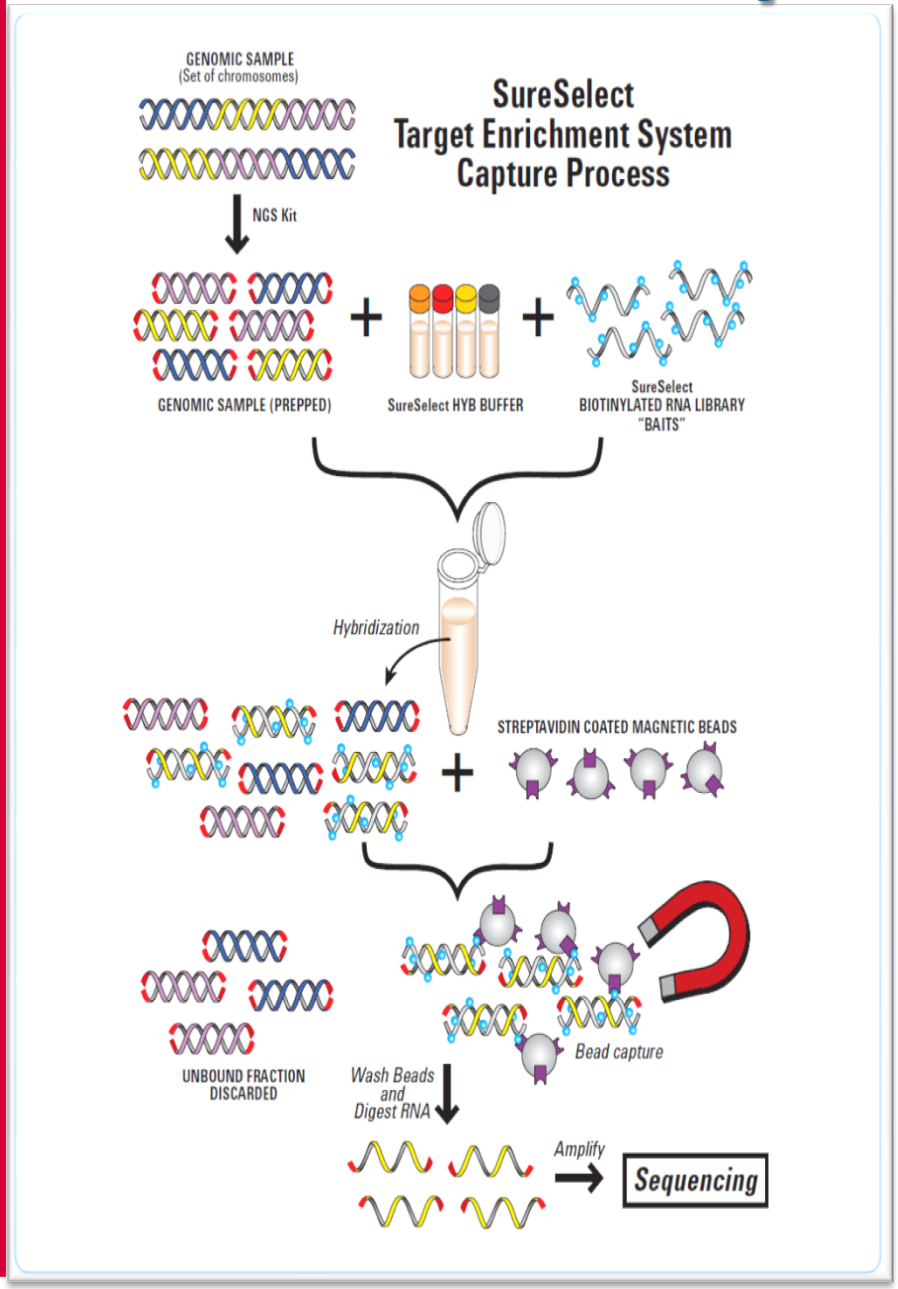
- Second Generation Sequencer
- Identify Somatic Mutations

  Two Somatic Mutations in Expressed Genes
- Chromosomal Changes

  Stable Pattern of Allelic Imbalance

National Cancer Institute

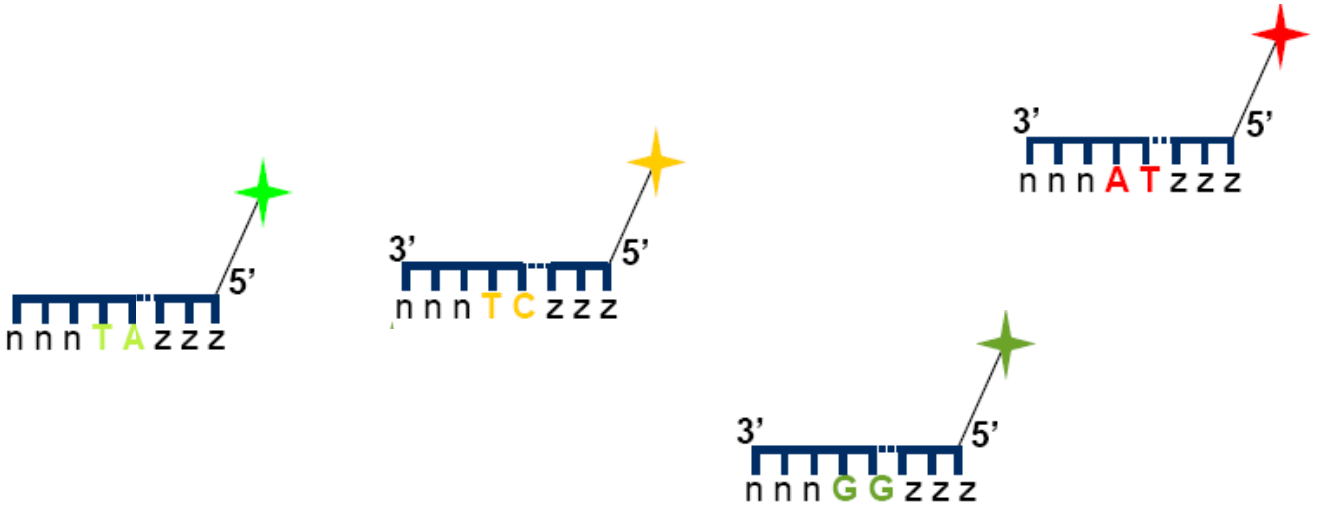
Exome Sequencing

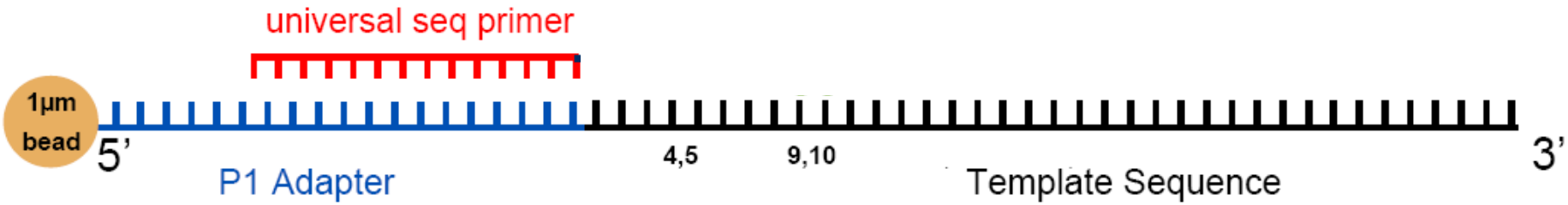


- Targeted 37.8Mb

SOLiD 4-color ligation chemistry system

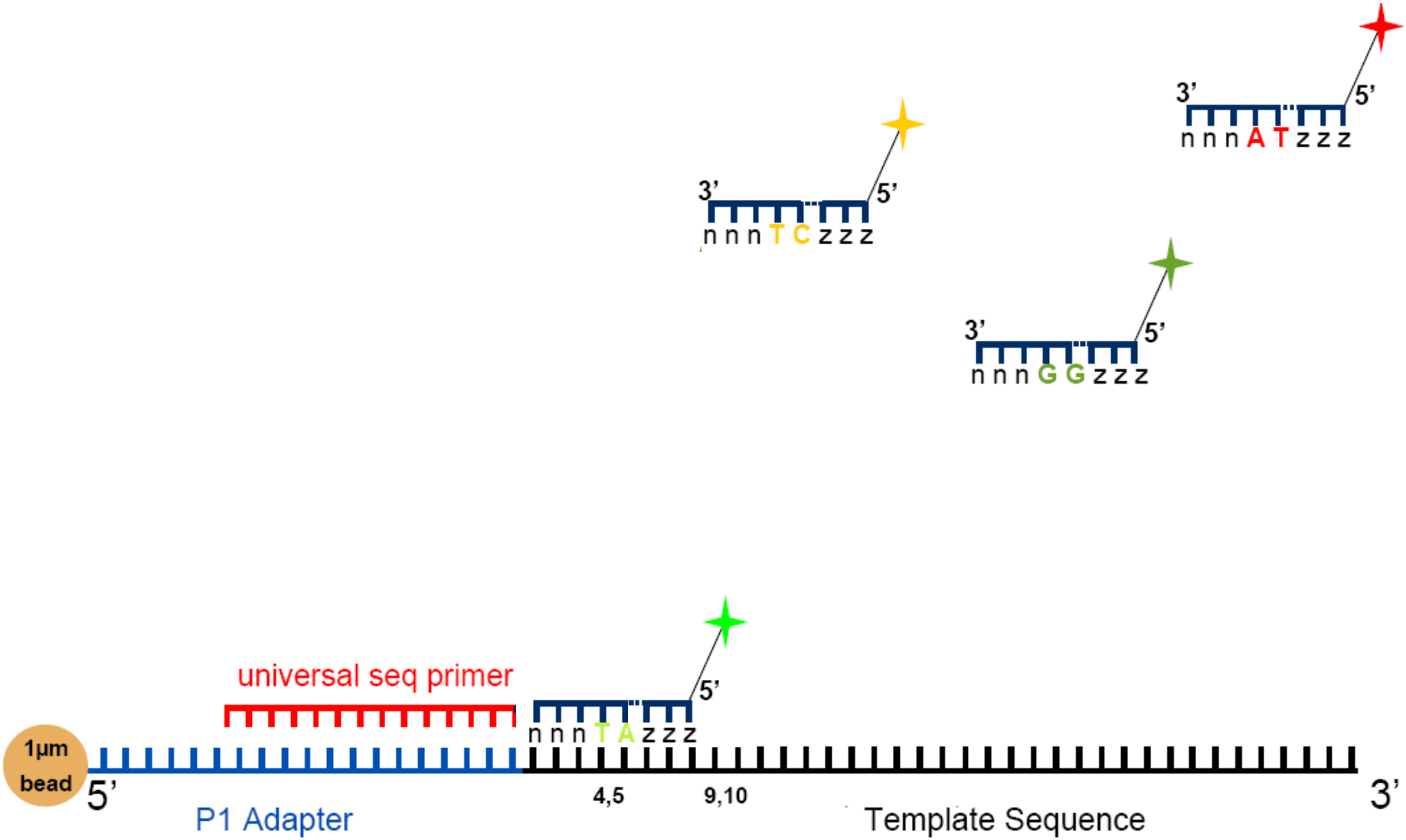
Sequencing by **O**ligonucleotide **Li**gation and **D**etection





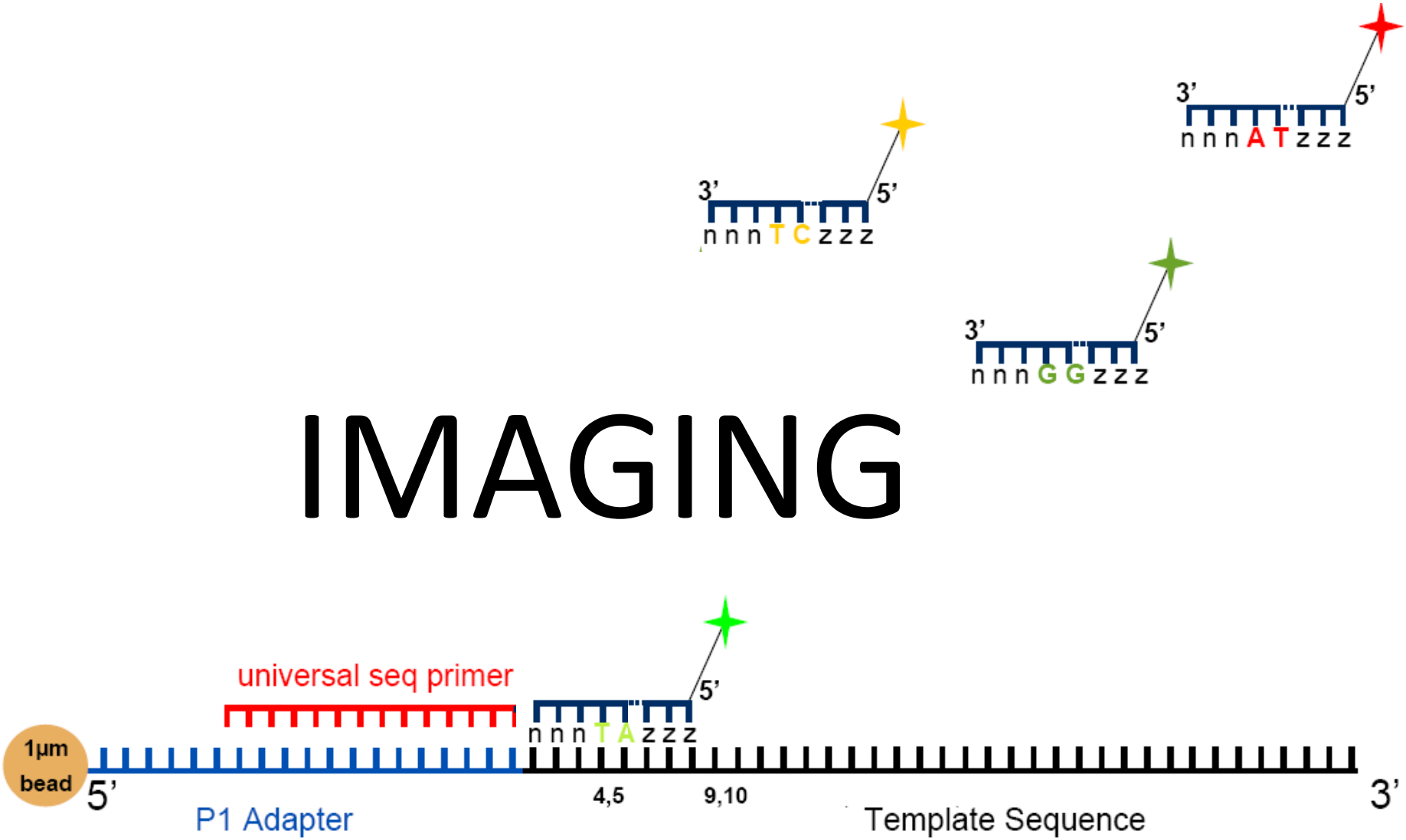
SOLiD 4-color ligation chemistry system

Sequencing by **O**ligonucleotide **Li**gation and **D**etection



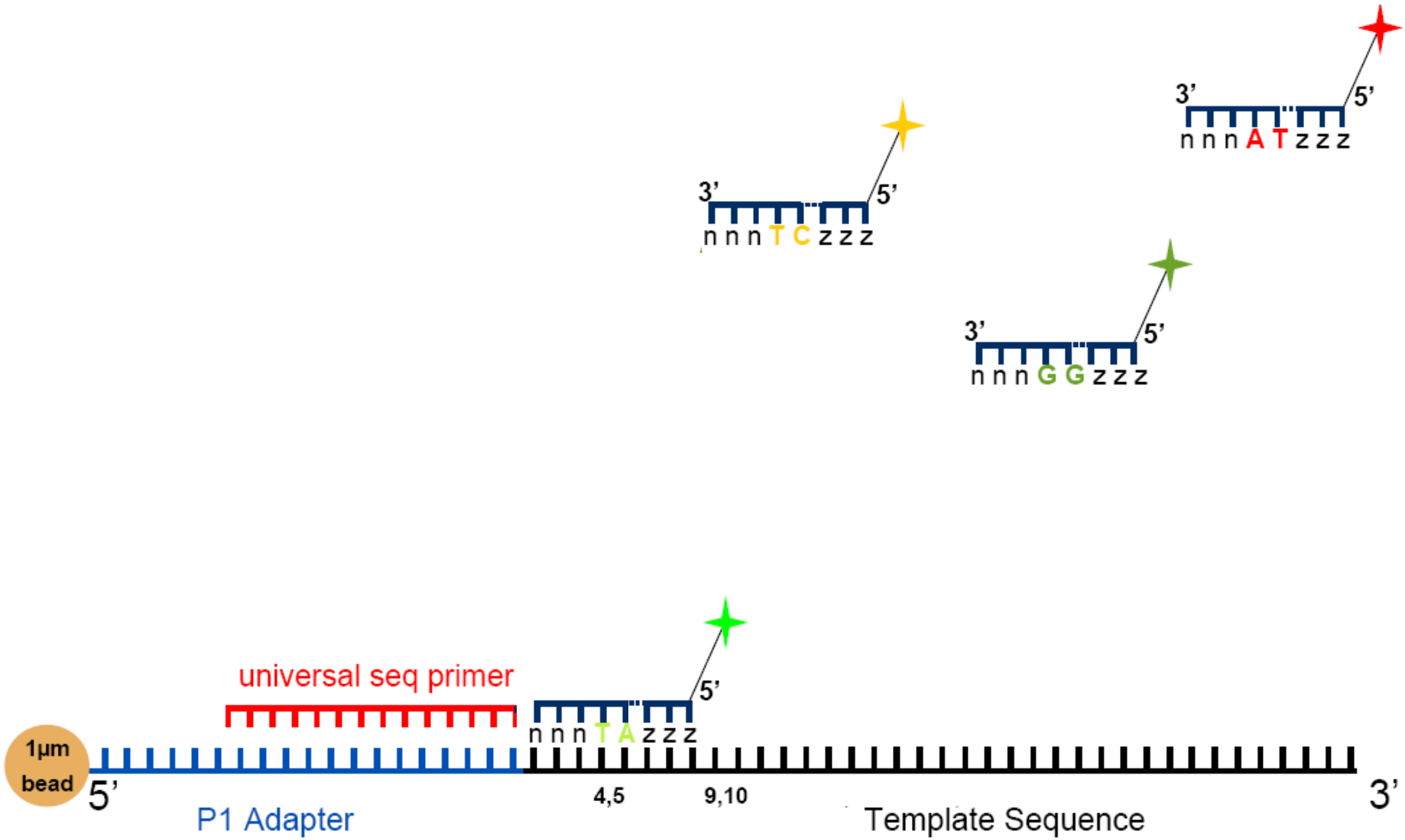
SOLiD 4-color ligation chemistry system

Sequencing by **O**ligonucleotide **Li**gation and **D**etection



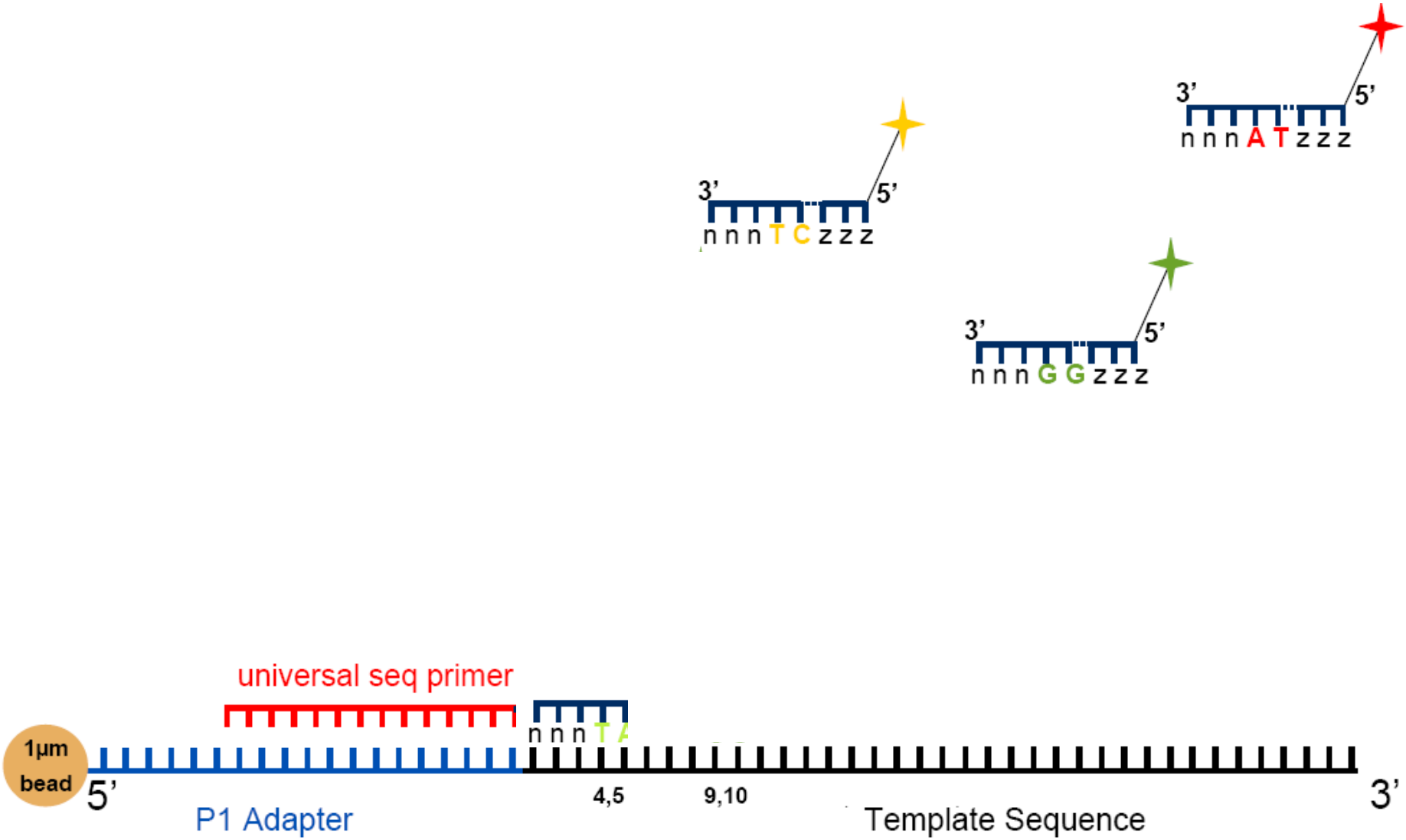
SOLiD 4-color ligation chemistry system

Sequencing by **O**ligonucleotide **Li**gation and **D**etection



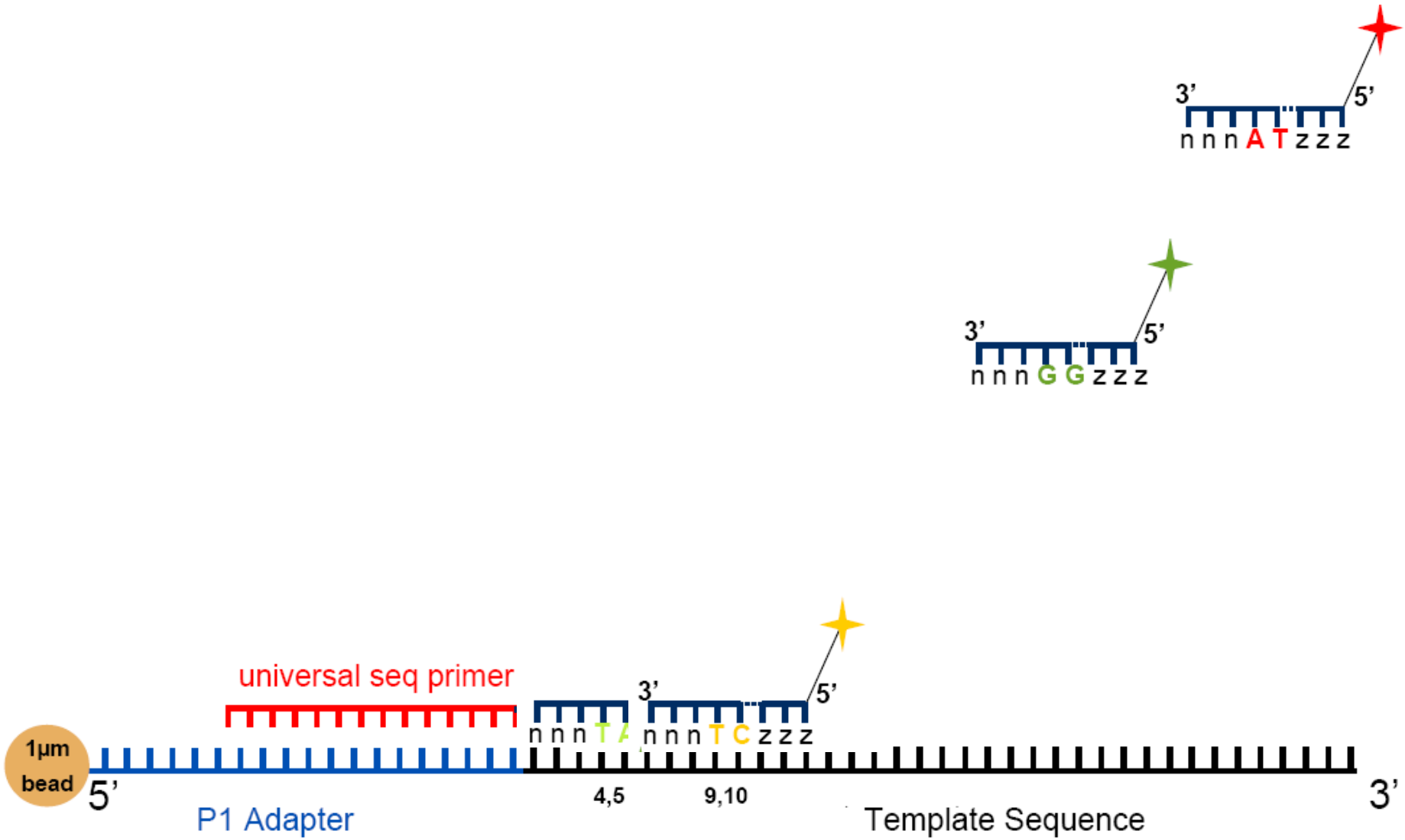
SOLiD 4-color ligation chemistry system

Sequencing by **O**ligonucleotide **Li**gation and **D**etection



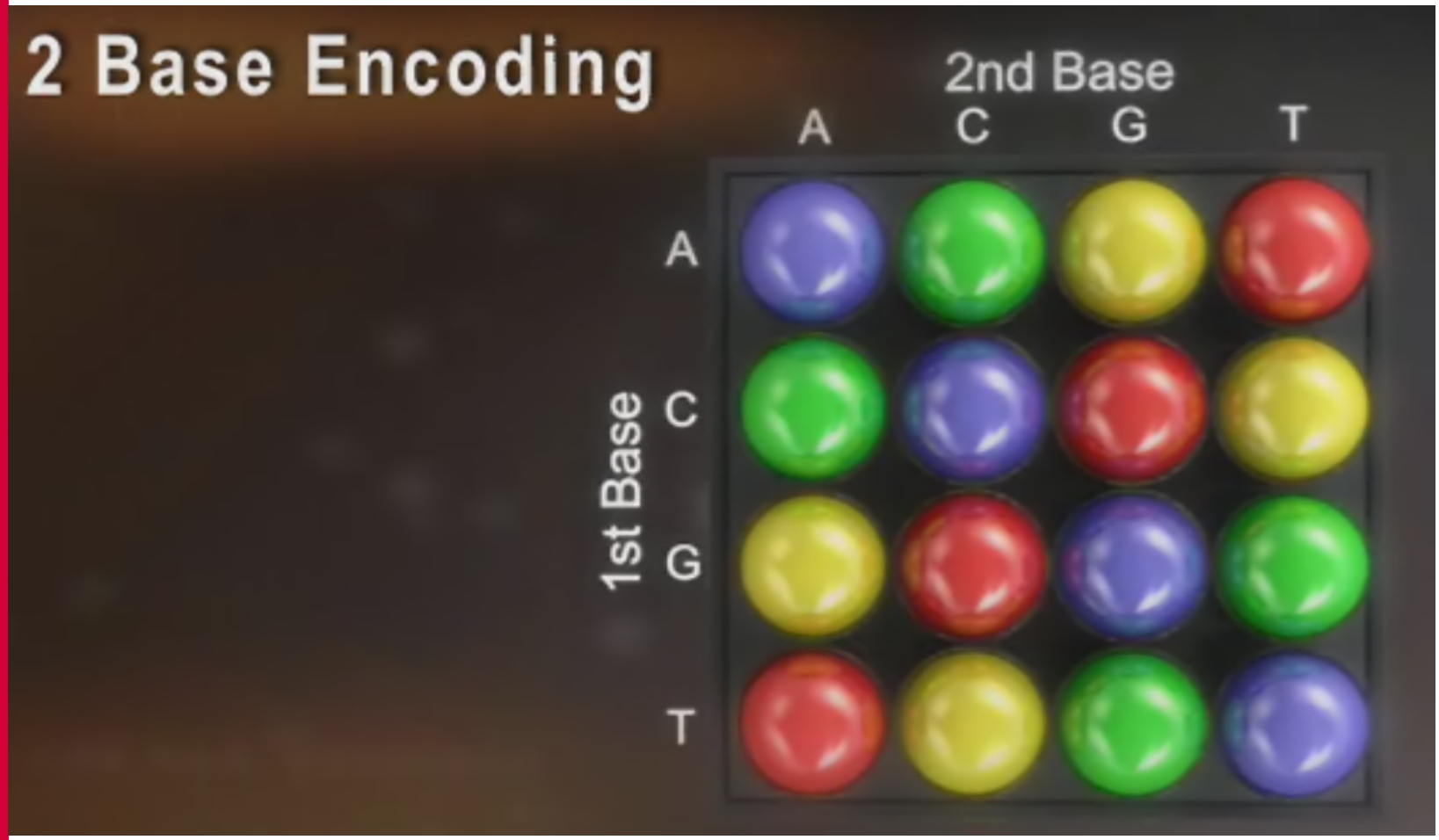
SOLiD 4-color ligation chemistry system

Sequencing by **O**ligonucleotide **Li**gation and **D**etection



National Cancer Institute

2 Base Encoding Improves Accuracy



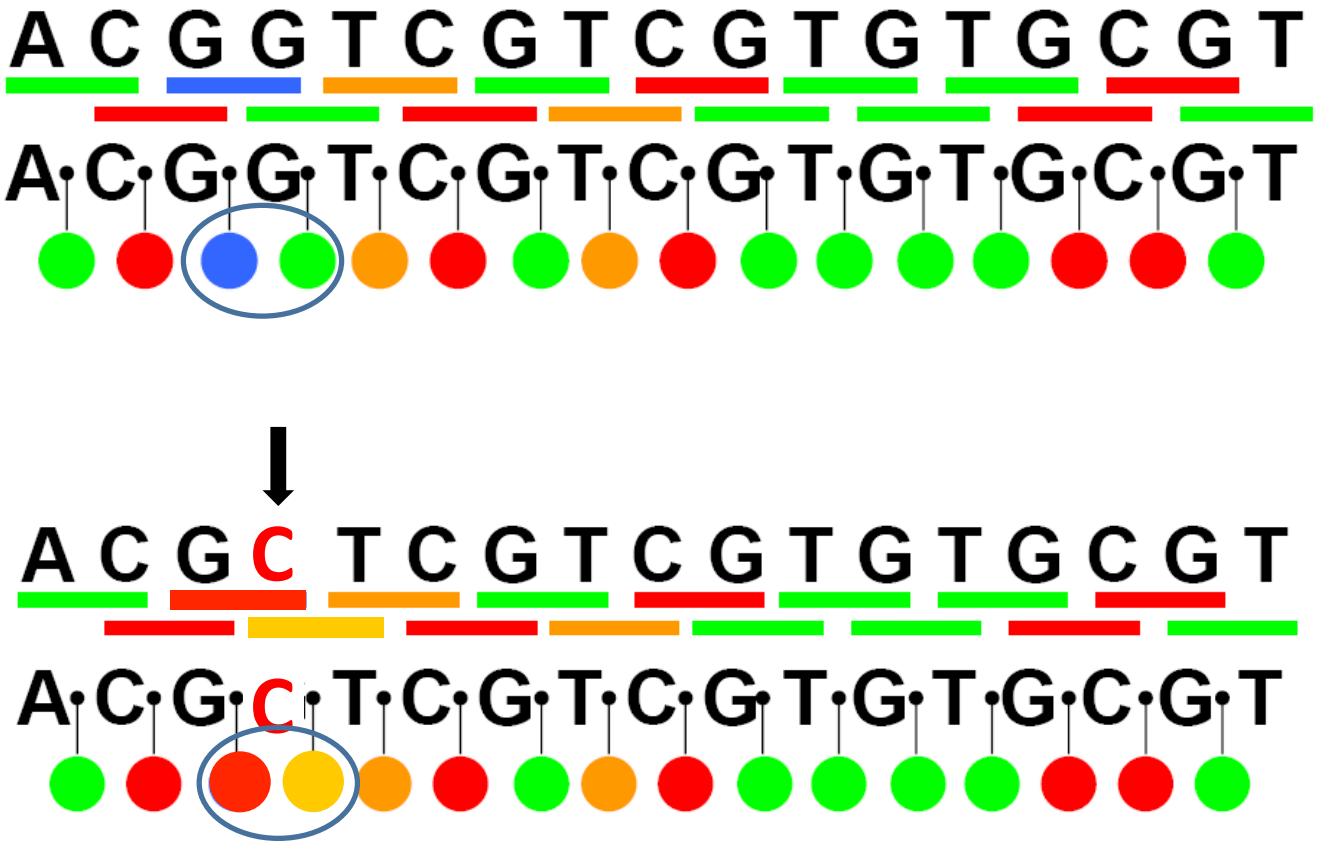
Advantages of 2 base encoding

- Double base interrogation eases the discrimination between system ***errors*** and ***true*** polymorphism

True polymorphisms produce 2 color changes while system errors produce a single color change

Advantages of 2 base encoding

- Double base interrogation eases the discrimination between system ***errors*** and ***true*** polymorphism

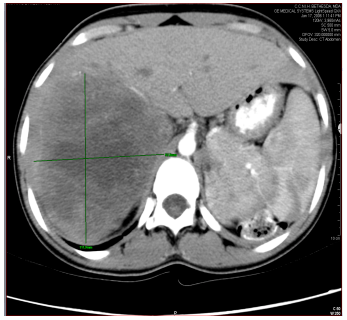


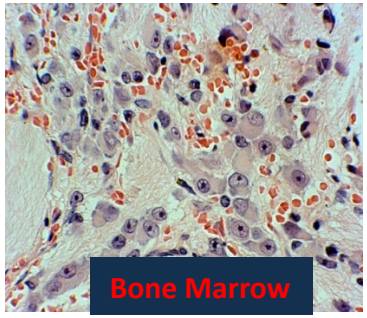
True polymorphisms produce 2 color changes while system errors produce a single color change

National Cancer Institute

Patient with Metastatic Neuroblastoma

Diagnosis

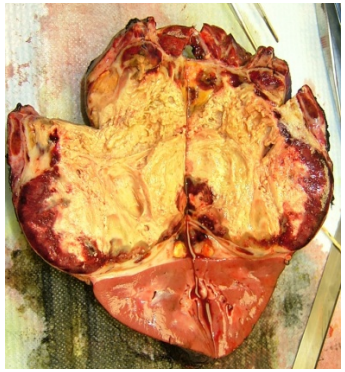


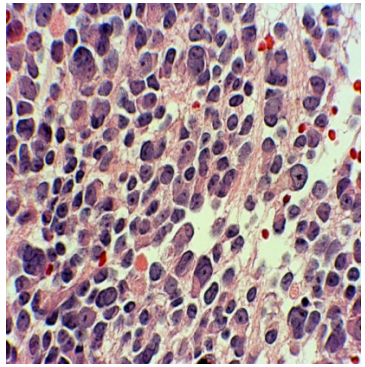


1. Met1-BM: Bone marrow biopsy at diagnosis

~4 Months

Surgery

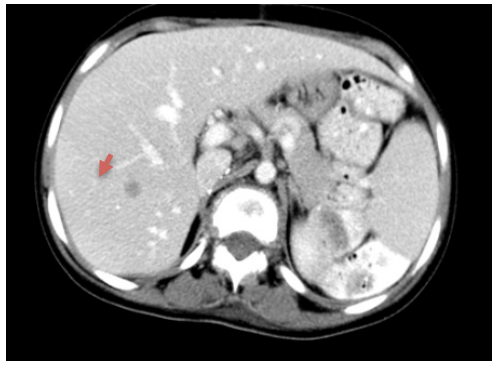




2. Primary: Primary tumor removed by surgery

3 years

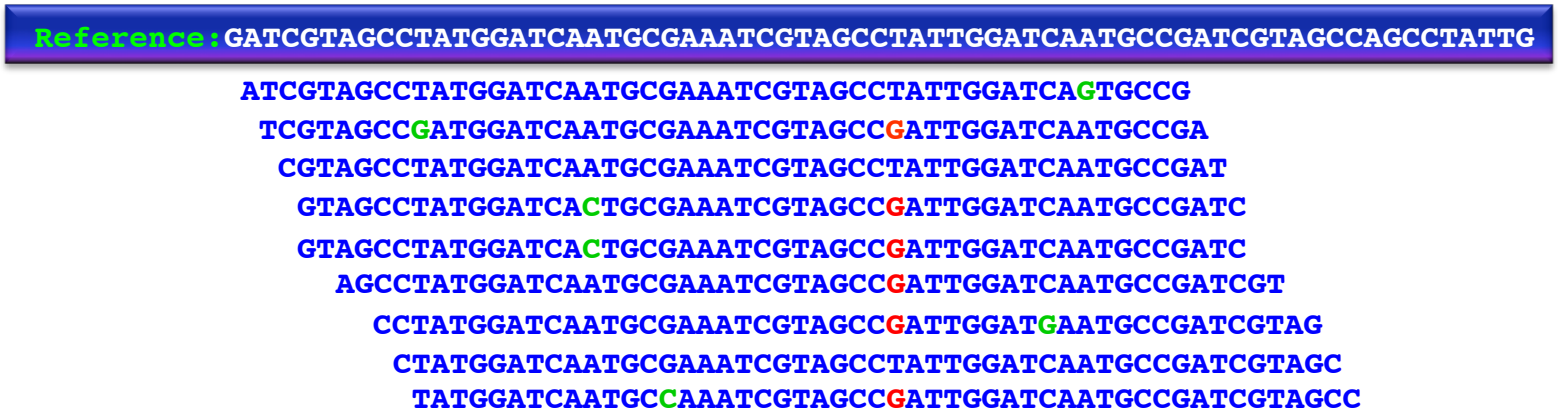
Death



3. Met2-Liver: Liver metastasis at autopsy

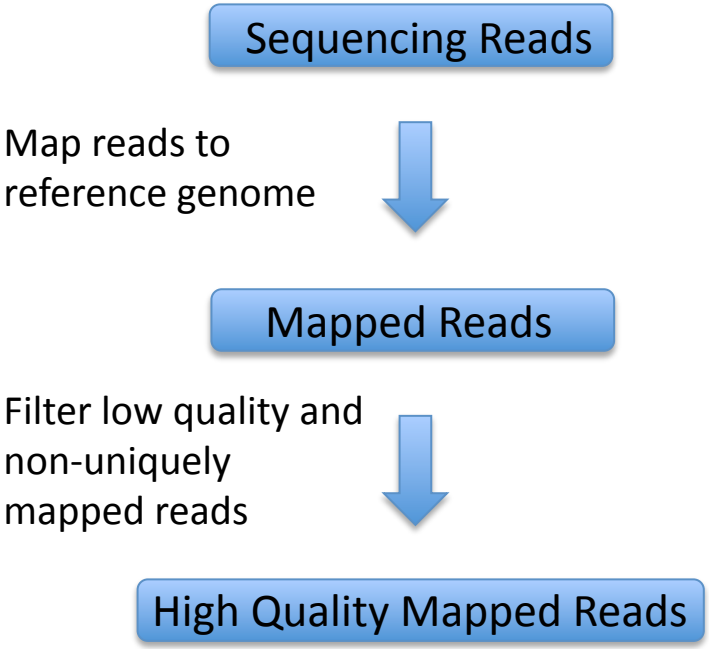
National Cancer Institute

Align To Reference Genome



National Cancer Institute

Align and Filter Reads



National Cancer Institute

High Coverage

Sample	Primary Tumor	Met1	Met2	Liver	Skin
Number of Mapped Reads	279M	216M	217M	259M	181M
High Quality Reads	231M	186M	180M	223M	153M
On Target Reads	165M	146M	143M	175M	134M
Coverage	187X	182X	184X	193X	205X

National Cancer Institute

Second Generation Sequencing

DNA

Mutations
- Insertions/Deletions
- Copy Number

Allelic Imbalance
- Translocations
- Inversions

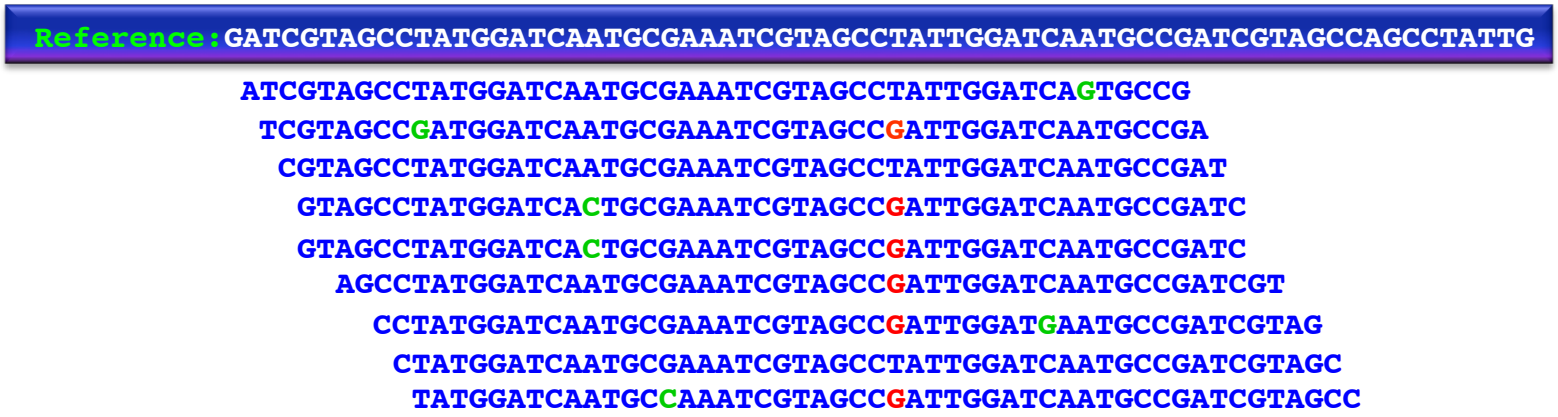
RNA

Mutations

Gene Expression
- Splice Variants
- Fusion Genes
- Novel Transcripts

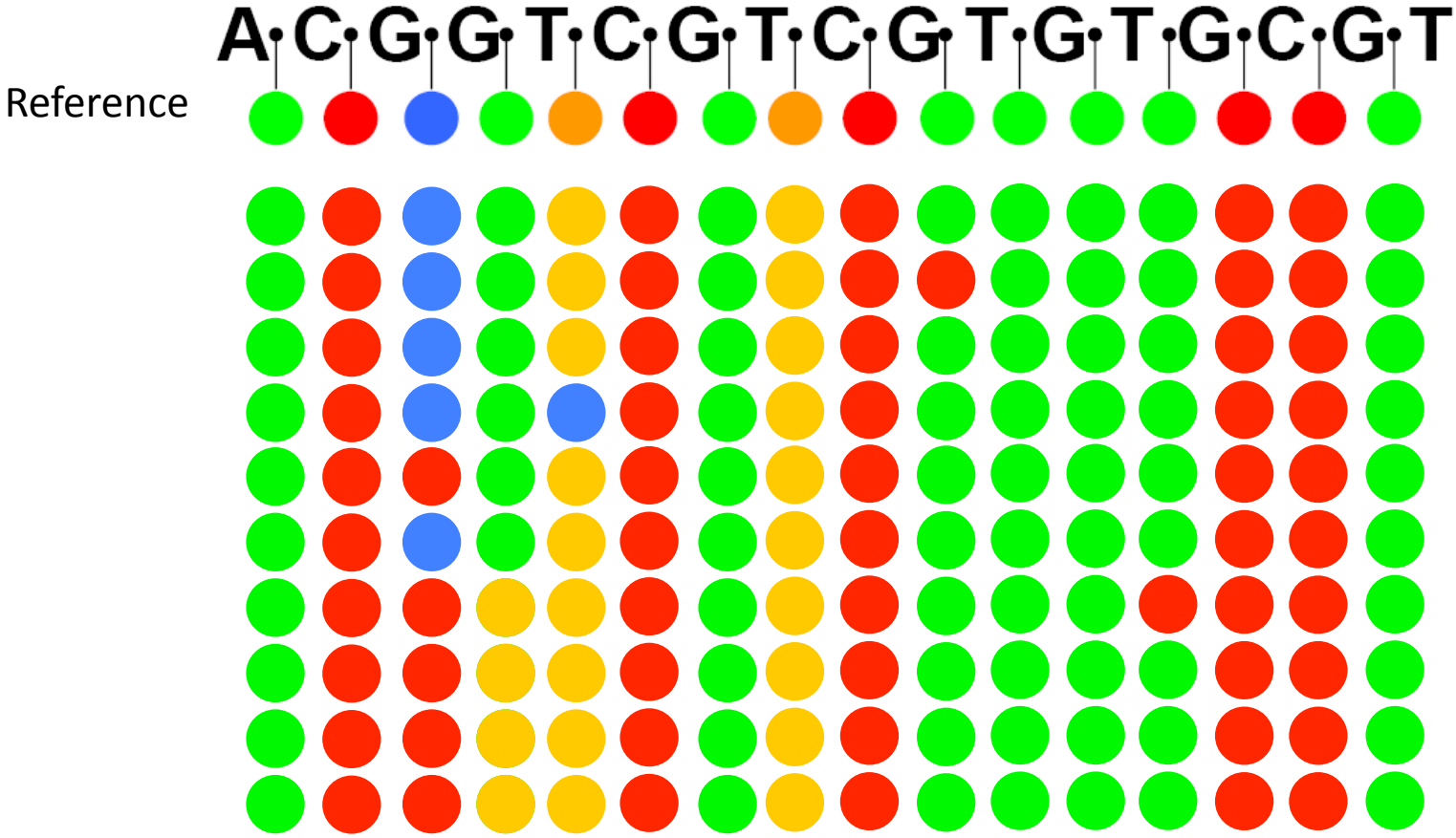
National Cancer Institute

Identify a Mutation



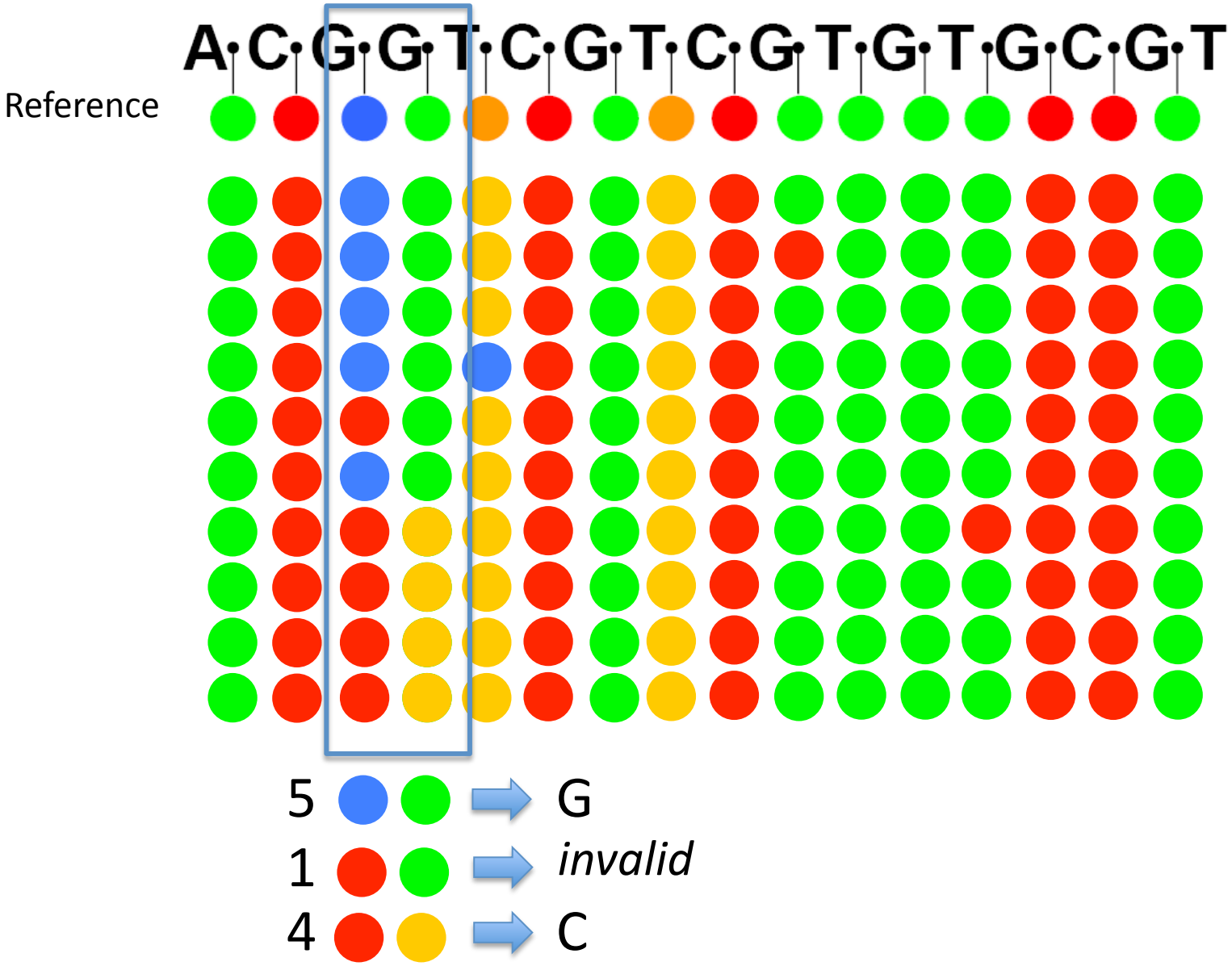
National Cancer Institute

Identify a Mutation in Color Space



National Cancer Institute

Identify a Mutation in Color Space



National Cancer Institute

Identify a Mutation - diBayes

Bayesian Algorithm:

$$P(\text{Genotype}|\text{Data}) = \frac{P(\text{Data}|\text{Genotype})\ P(\text{Genotype})}{P(\text{Data})}$$

National Cancer Institute

Identify a Mutation - diBayes

Bayesian Algorithm:

$$P(\text{Genotype}|\text{Data}) = \frac{P(\text{Data}|\text{Genotype})\ P(\text{Genotype})}{P(\text{Data})}$$

Genotypes: GG, GC, GT, GA, CC, CT, CA, TT, TA, AA

Prior: P(Genotype) = 0.999 if same as reference (e.g. GG)
(1-0.999)/9 otherwise

National Cancer Institute

Identify a Mutation - diBayes

Bayesian Algorithm:

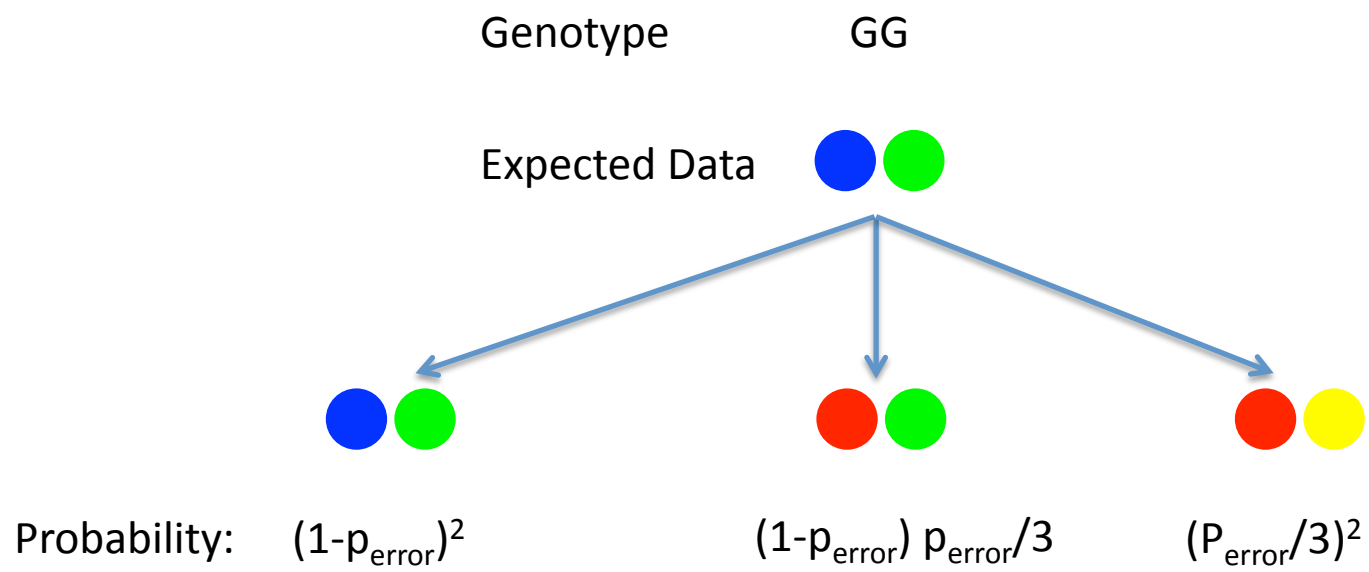
$$P(\text{Genotype}|\text{Data}) = \frac{P(\text{Data}|\text{Genotype})\ P(\text{Genotype})}{P(\text{Data})}$$

$$P(\text{Data}|\text{Genotype}) = P(\text{Read}_1|\text{Genotype})\ P(\text{Read}_2|\text{Genotype})\ldots$$

National Cancer Institute

Identify a Mutation - diBayes

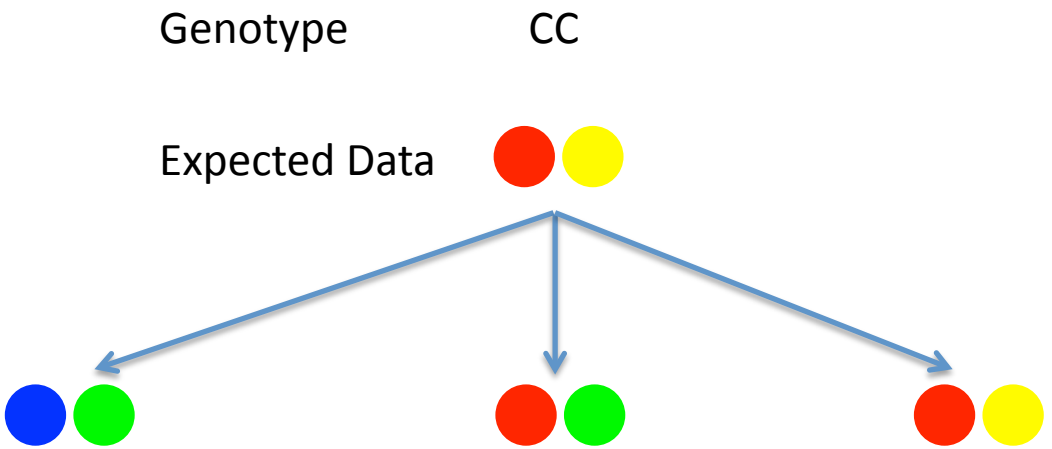
$P(Read_1 | Genotype)$



National Cancer Institute

Identify a Mutation - diBayes

$P(Read_1 | Genotype)$

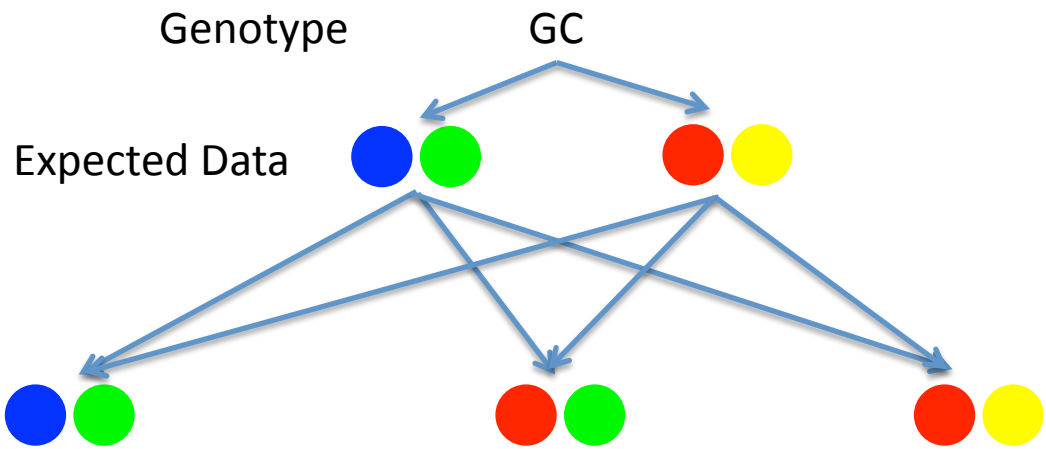


Probability: $(P_{error}/3)^2$ $(1-p_{error})\ p_{error}/3$ $(1-P_{error})^2$

National Cancer Institute

Identify a Mutation - diBayes

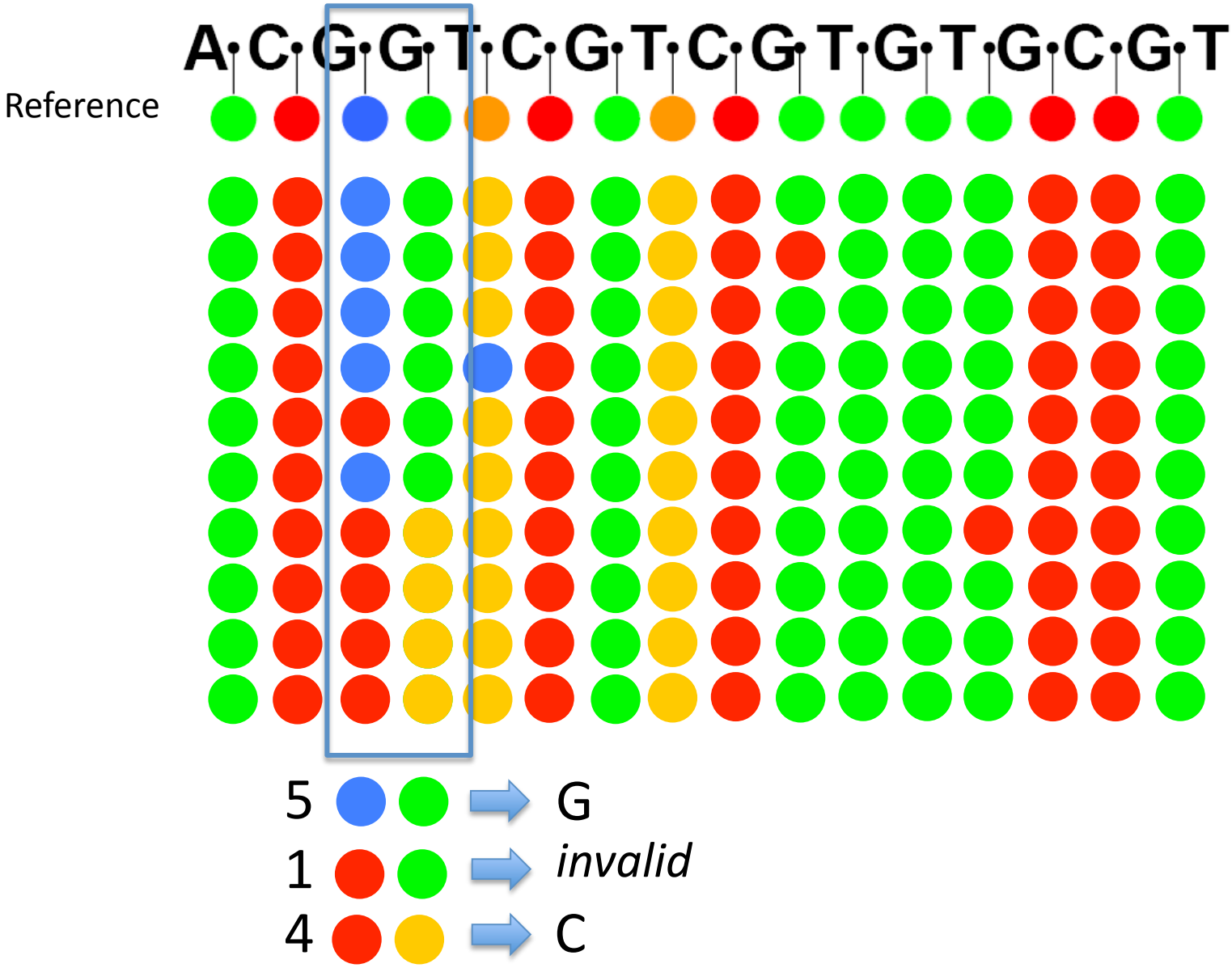
$P(Read_1 | Genotype)$



Probability: $0.5(1-P_{error})^2+0.5(P_{error}/3)^2$ $(1-p_{error})\ p_{error}/3$ $0.5(P_{error}/3)^2+0.5(1-P_{error})^2$

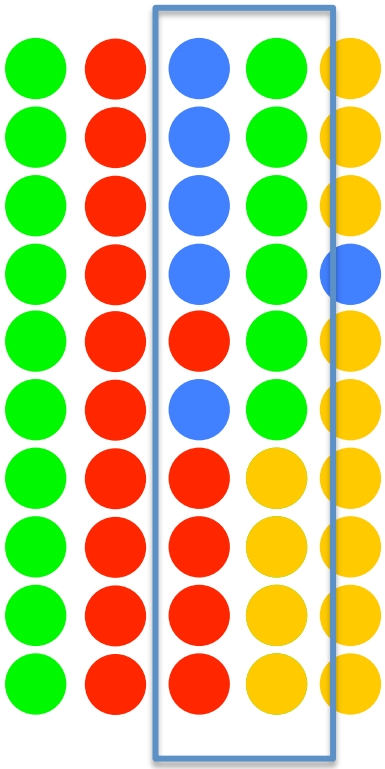
National Cancer Institute

Identify a Mutation in Color Space



National Cancer Institute

Identify a Mutation in Color Space



P(GG|Data) = P(GG) P(●●|GG)^{5}P(●●|GG)^{1}P(●●|GG)^4

P(GC|Data) = P(GC) P(●●|GC)^{5}P(●●|GC)^{1}P(●●|GC)^4

P(GG|Data) = P(CC) P(●●|CC)^{5}P(●●|CC)^{1}P(●●|CC)^4

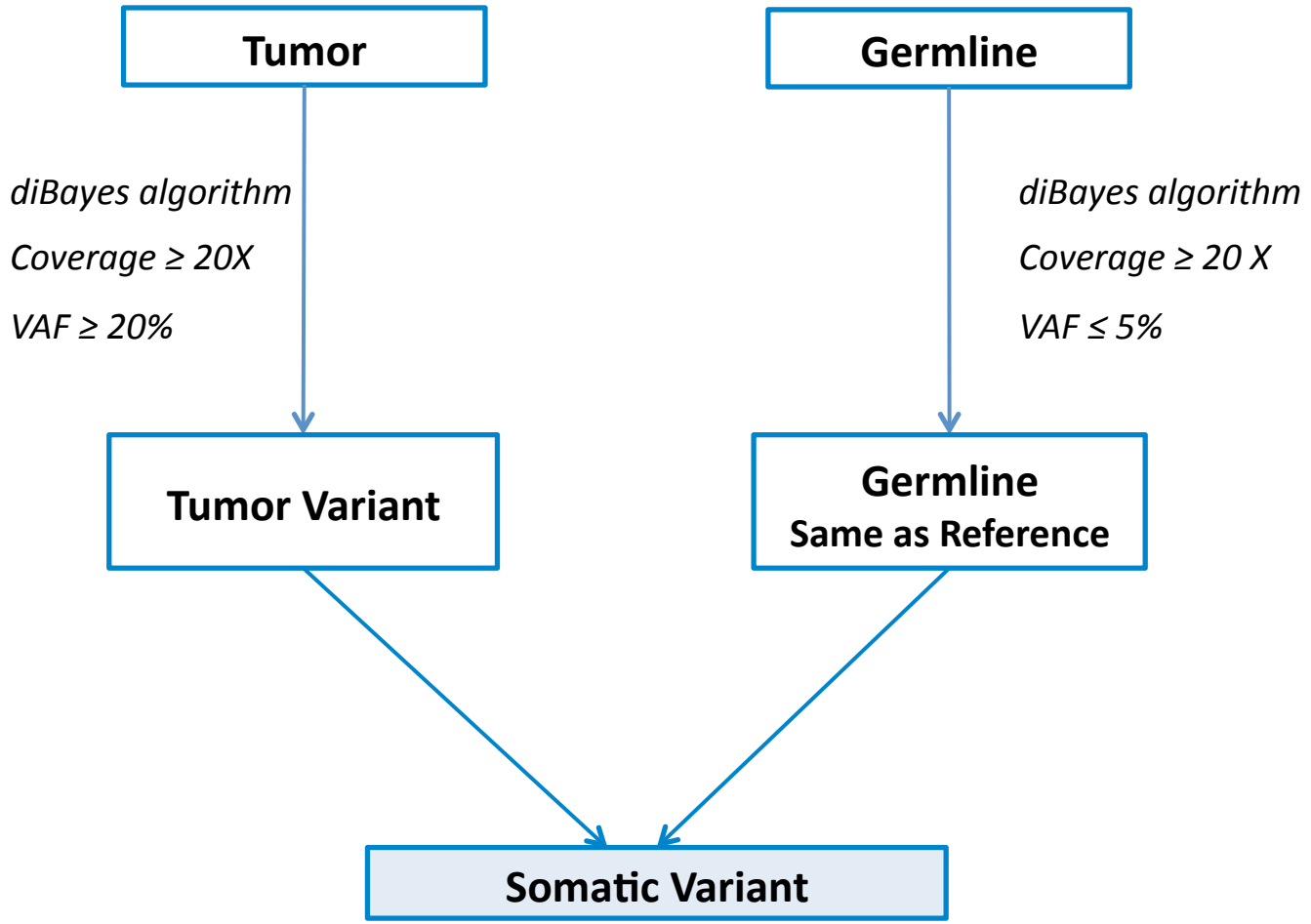
5 ●● → G

1 ●● → *invalid*

4 ●● → C

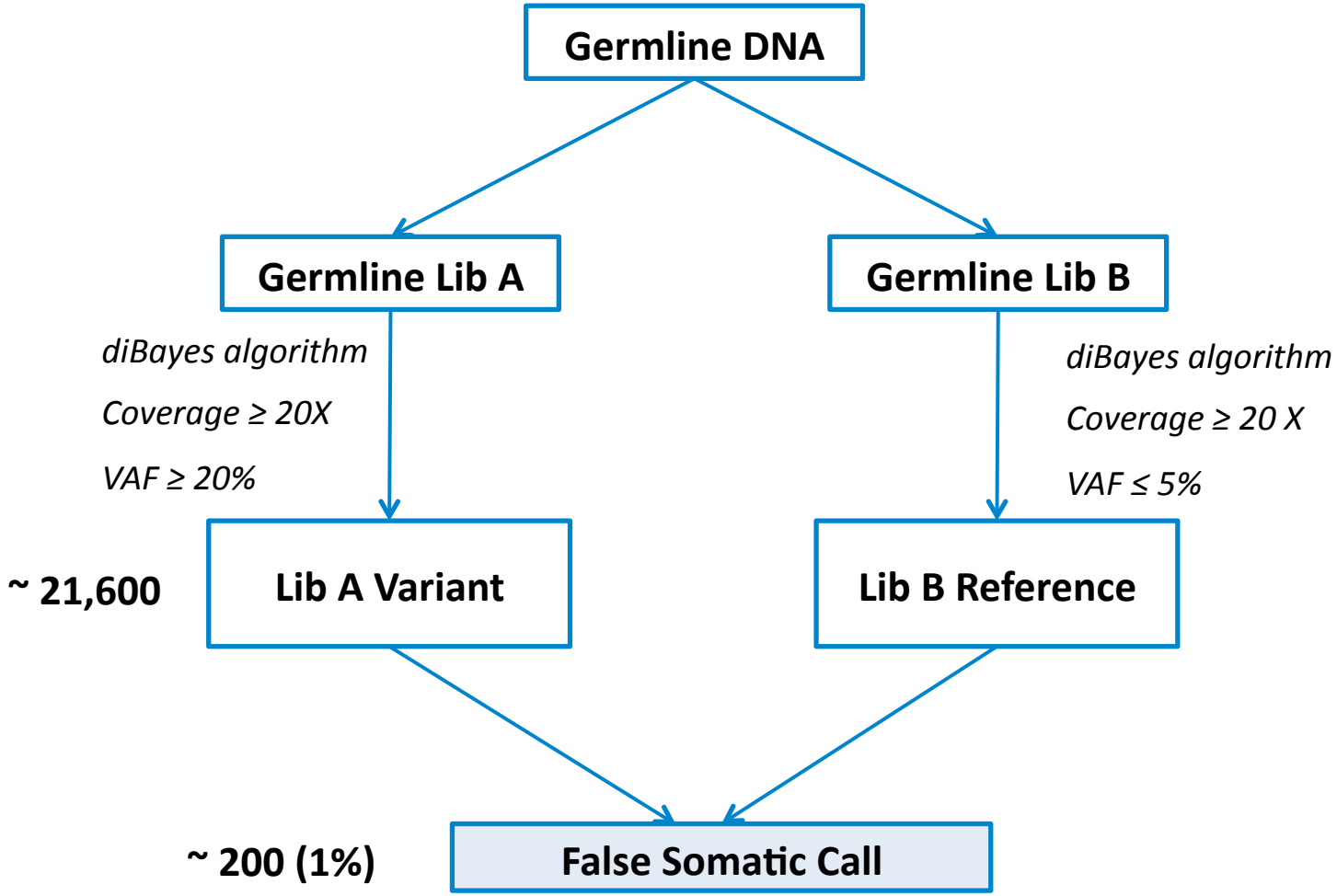
National Cancer Institute

Somatic Mutation Detection



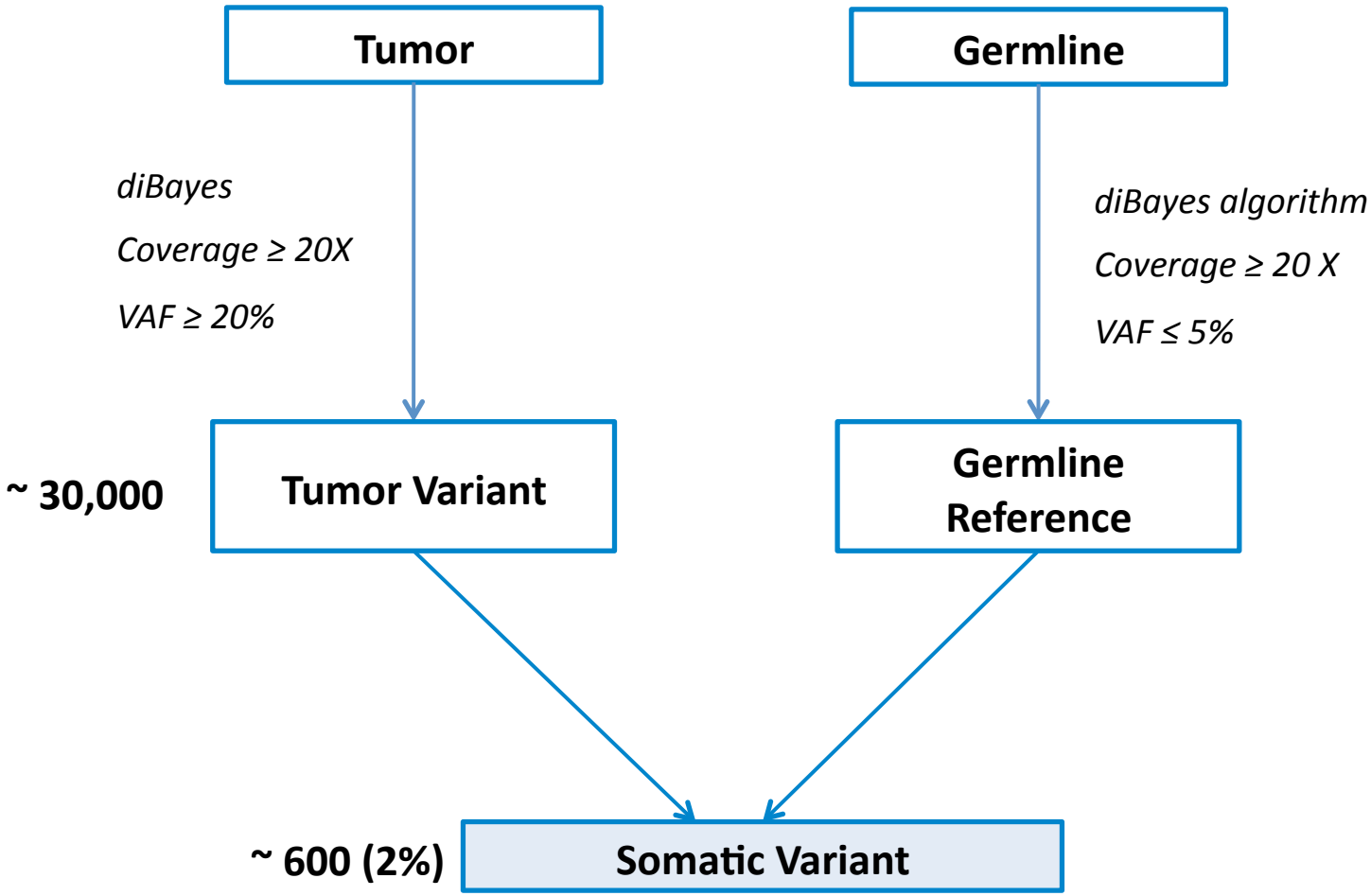
National Cancer Institute

Estimation of False Call Rate



National Cancer Institute

Somatic Mutation Detection



False Calls: ~300 (1% of 30,000)

National Cancer Institute

Somatic Mutations

	Primary Tumor	Met1 Bone Marrow	Met2 Liver
Variants	29,593	30,006	29,131
Somatic Variants	627	573	634
Protein Coding Region	239	243	287
Non-synonymous Variant	163	170	195

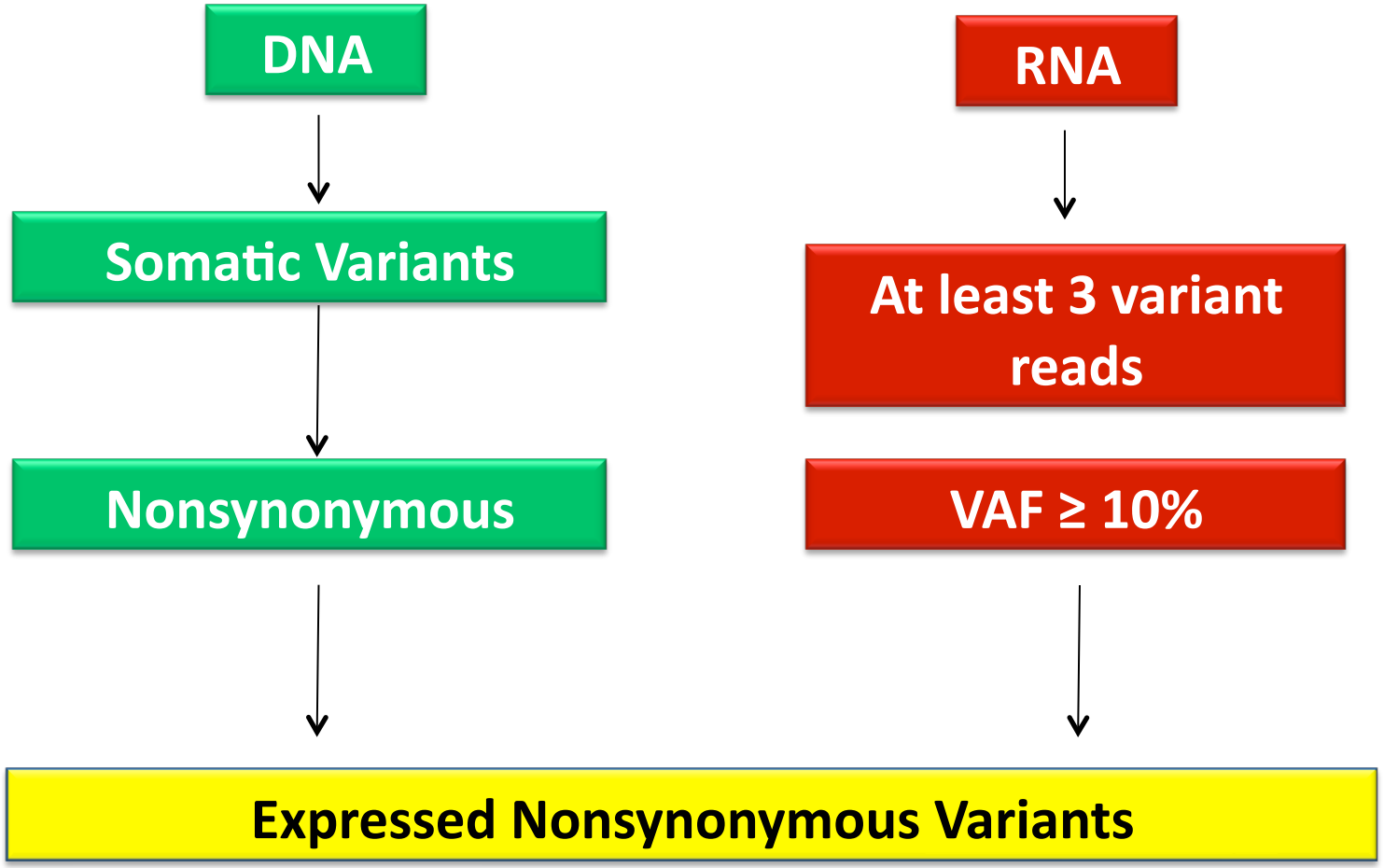
National Cancer Institute

Overlap with RNA data

- Reduce false positives by focusing on variants detected by both methods
- Identify therapeutic target – expressed by nature

National Cancer Institute

Expressed Nonsynonymous Variants



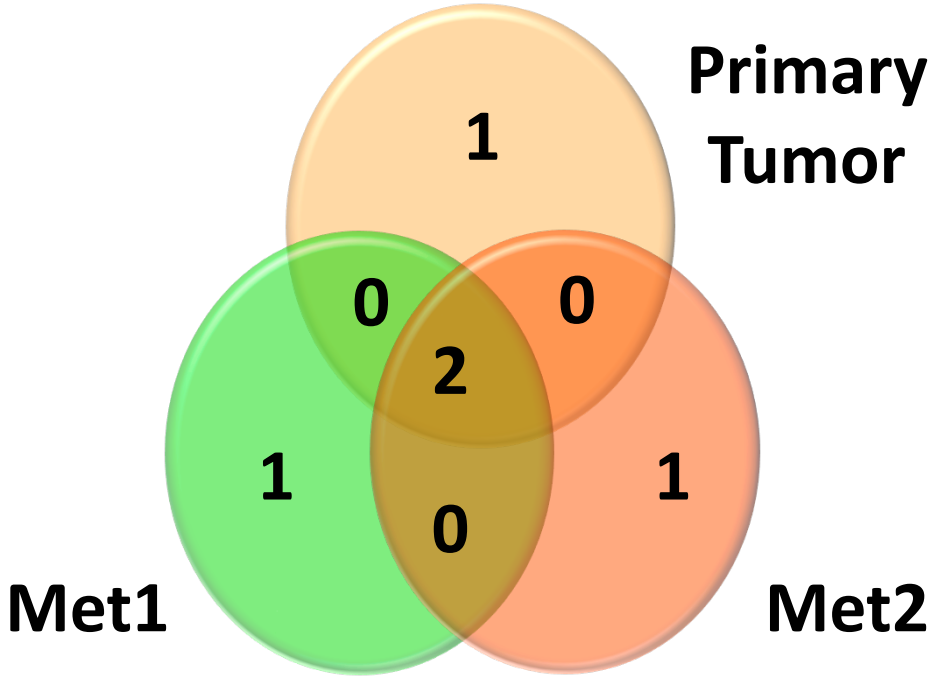
National Cancer Institute

Expressed Nonsynonymous Variants

	Primary Tumor	Met1 Bone Marrow	Met2 Liver
Variants	29,593	30,006	29,131
Somatic Variants	627	573	634
Translated Region	239	243	287
Non-synonymous	163	170	195
Expressed	3	3	3

National Cancer Institute

Expressed Nonsynonymous Variants



National Cancer Institute

Outline

- Second Generation Sequencer
- Identify Somatic Mutations
  - Two Somatic Mutations in Expressed Genes
- Chromosomal Changes
  - Stable Pattern of Allelic Imbalance

National Cancer Institute

Allelic Imbalances

Maternal Copy

ACTGACTGGTACTGATTT

ACTCACTGGTACTGATTT

Paternal Copy

National Cancer Institute

Allelic Imbalances

Maternal Copy

ACTGACTGGTACTGATTT

ACTCACTGGTACTGATTT

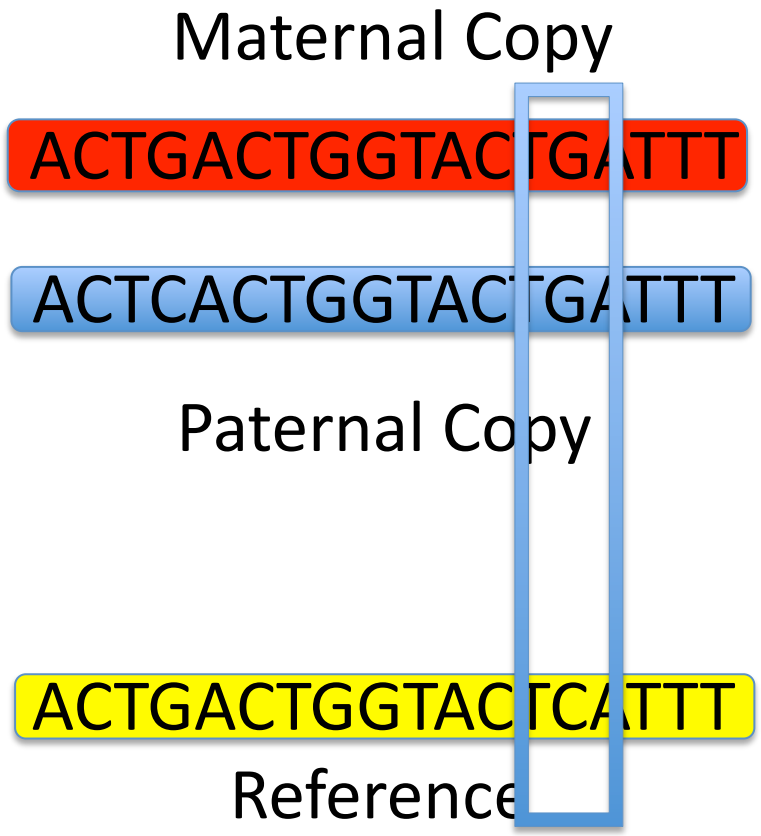
Paternal Copy

ACTGACTGGTACTCATTT

Reference

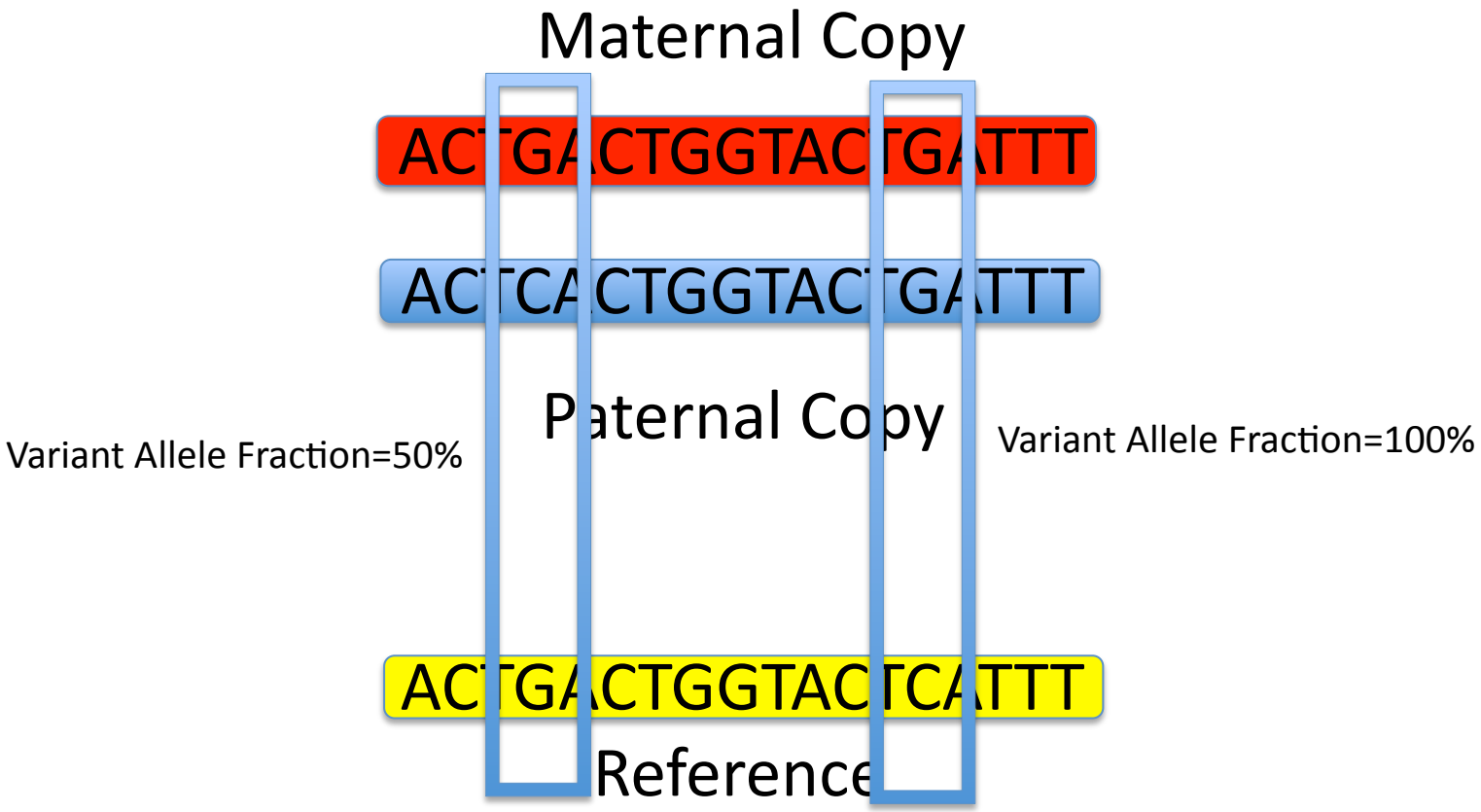
National Cancer Institute

Allelic Imbalances



National Cancer Institute

Allelic Imbalances

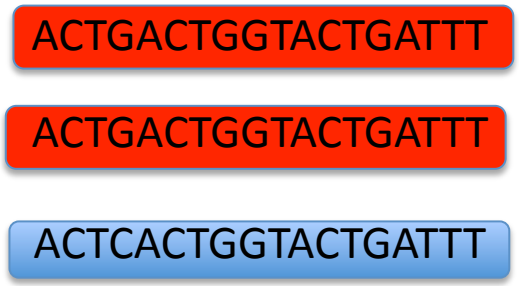


National Cancer Institute

Allelic Imbalances

NORMAL
VAF: 0%, 50%, 100%

Loss of Heterozygosity (LOH)
VAF: 0%, 100%



Gain a Copy
VAF: 0%, 33%, 67%,100%

National Cancer Institute

Allelic Imbalances

ACTGACTGGTACTGATTT

ACTCACTGGTACTGATTT

NORMAL
VAF: 0%, 50%, 100%

ACTGACTGGTACTGATTT

Loss of Heterozygosity (LOH)
VAF: 0%, 100%

ACTGACTGGTACTGATTT

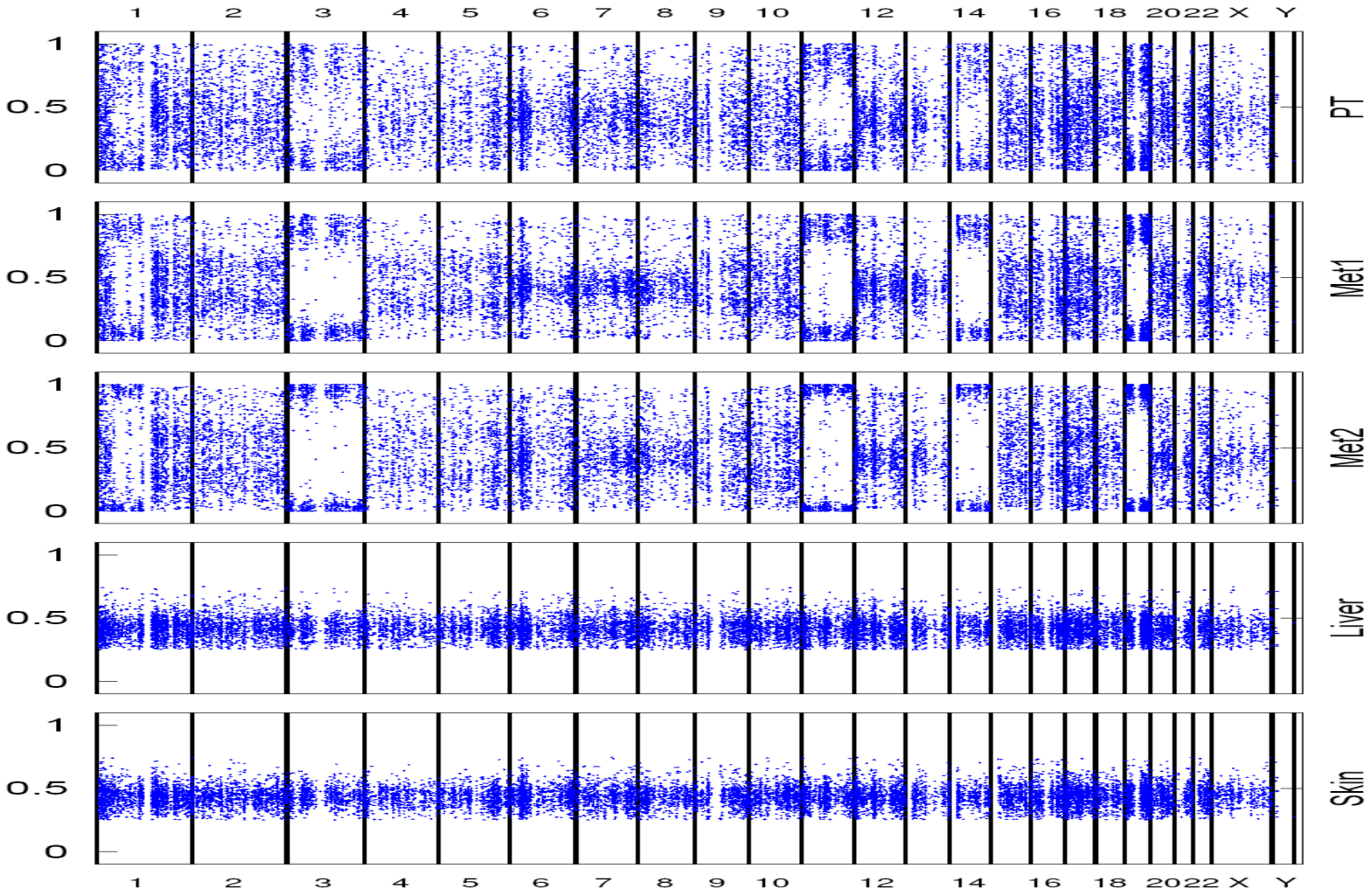
ACTGACTGGTACTGATTT

ACTCACTGGTACTGATTT

Gain a Copy
VAF: 0%, 33%, 67%,100%

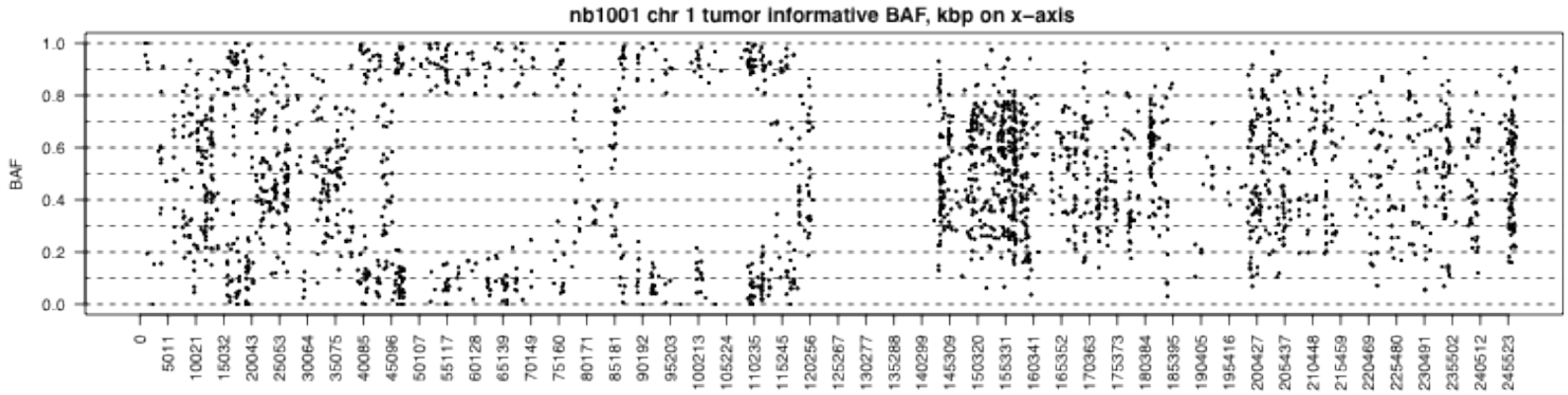
National Cancer Institute

Allelic Imbalances



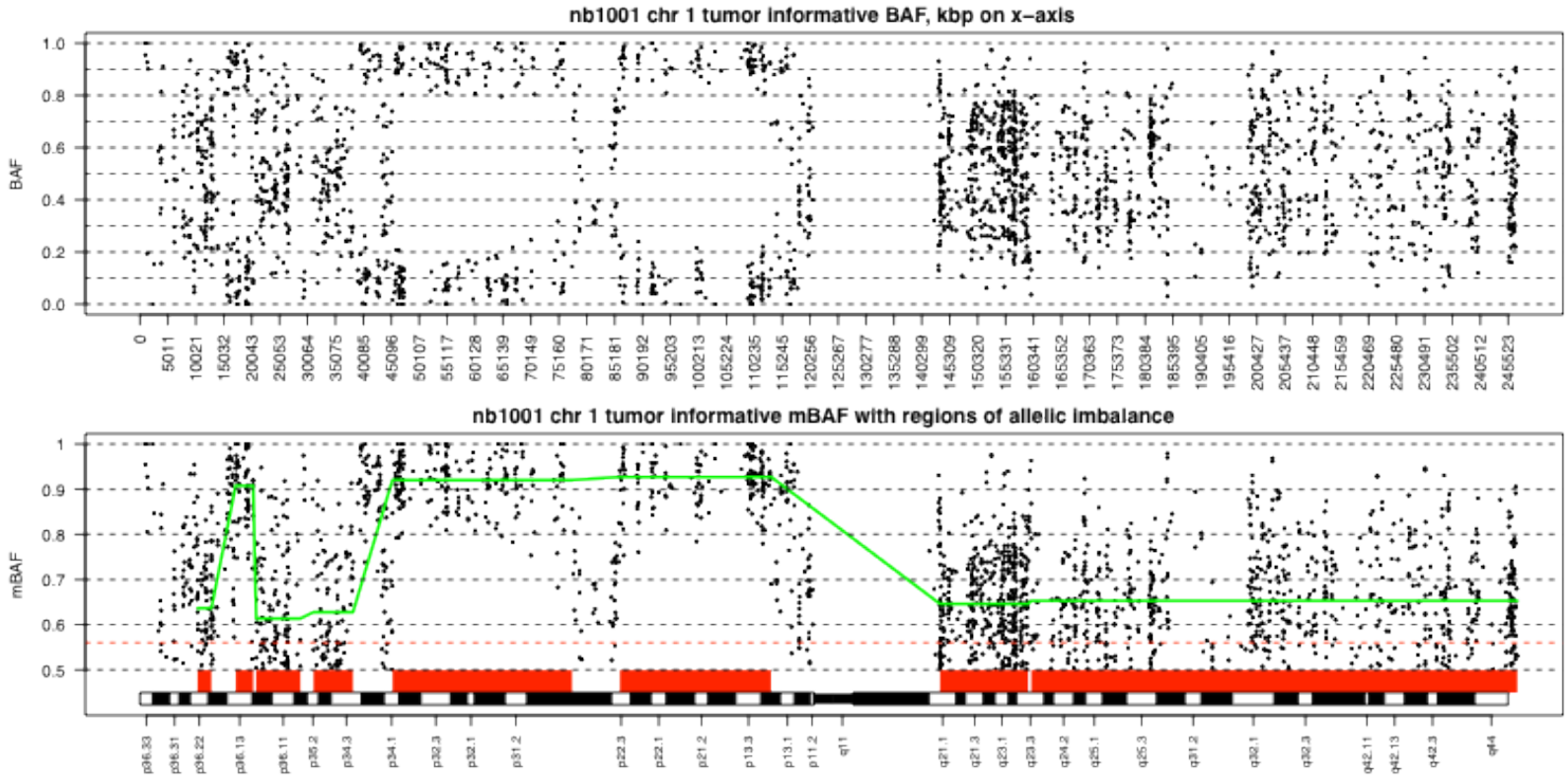
National Cancer Institute

VAF Segmentation



National Cancer Institute

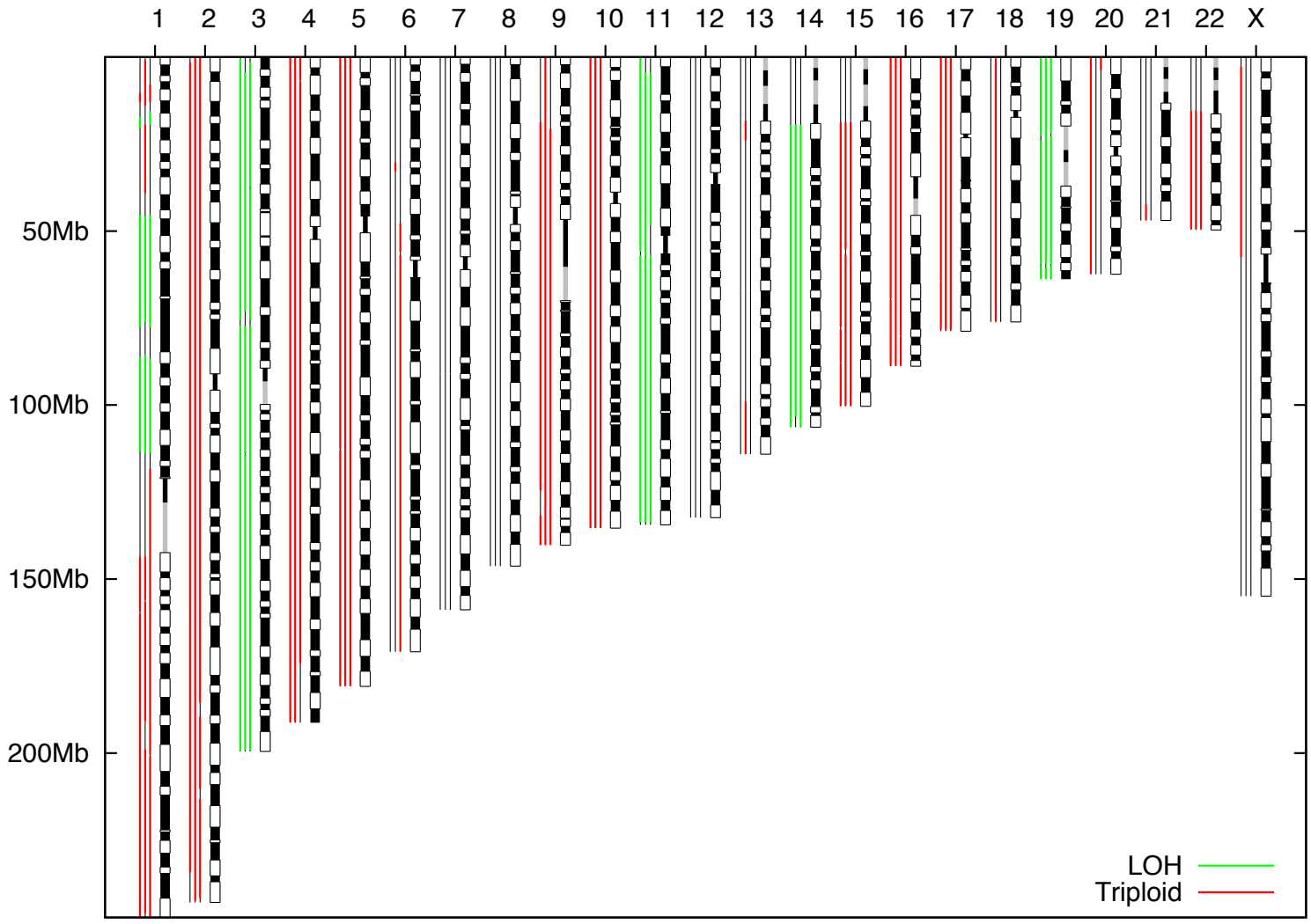
VAF Segmentation



Staaf *et al.*, *Genome Biology* **9**, R136 (2008)

National Cancer Institute

Similar Patterns of Allelic Imbalance



National Cancer Institute

Summary

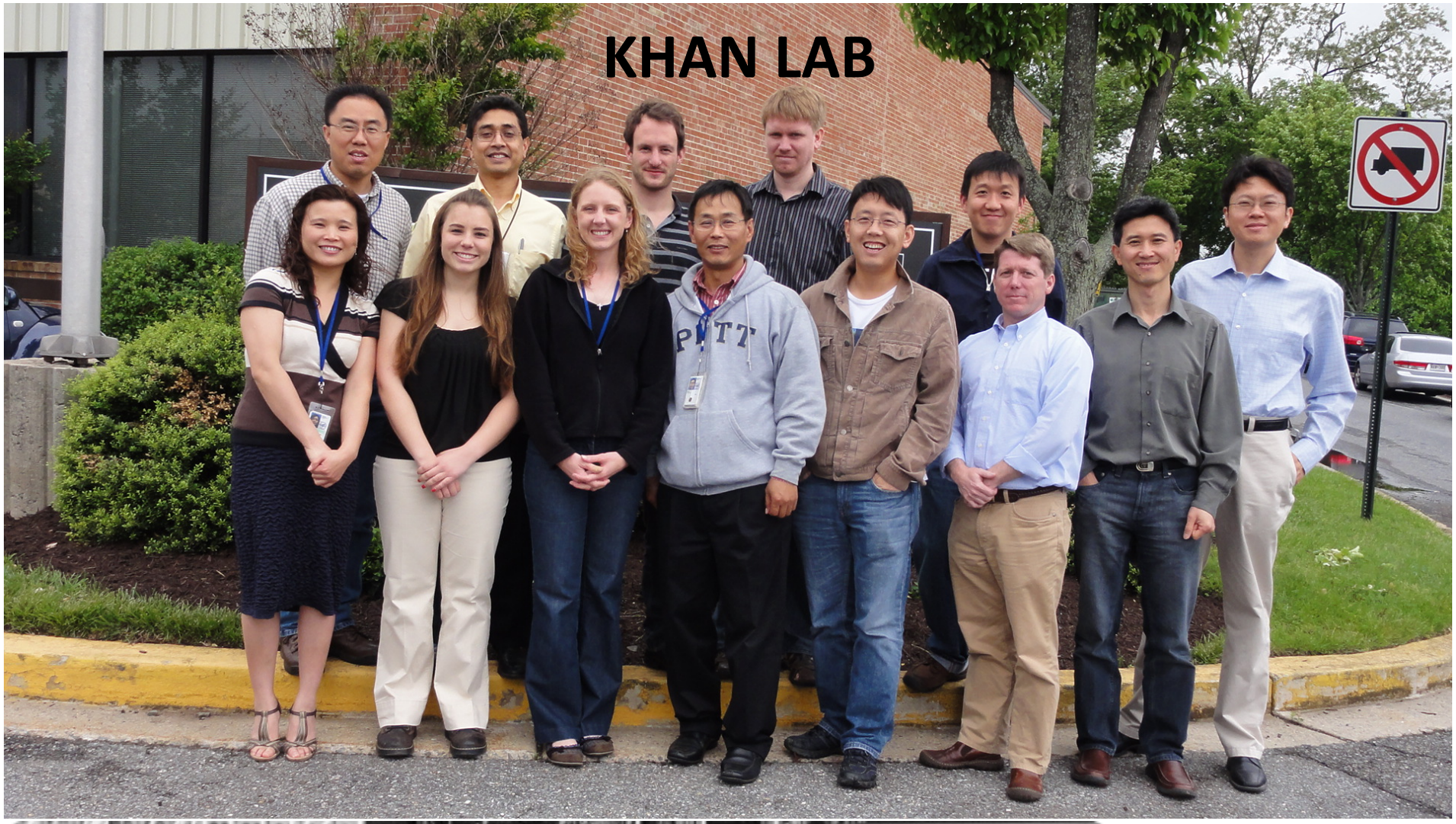
Somatic Mutations

- 2% of found variants in tumors are somatic
- We have identified two expressed somatic mutations present in all three samples

Allelic Imbalance

- Tumor showed a stable pattern of extensive allelic imbalances indicating a common origin

Acknowledgements



Biowulf Cluster