# Providing a More Comprehensive and Personalized Approach to Genetic Disorders through Next Generation Sequencing

**Karl V. Voelkerding, MD**
**Professor of Pathology**
**University of Utah**
**Medical Director for Genomics and Bioinformatics**
**ARUP Laboratories**

**ARUP Institute for Learning Webinar**
**June 18, 2013**

**voelkek@aruplab.com**

# Notice of Faculty Disclosure

**The individual below has no relevant financial relationships with commercial interests to disclose**

**Karl V. Voelkerding, M.D.**

# Learning Objectives

- **Describe how NGS has provided a new technological approach that has expanded the ability to improve the diagnosis of genetic disorders.**

- **Relate the essential and complex role of bioinformatics in deriving diagnostic results from NGS data.**

- **Discuss the impact of exome sequencing in the diagnostic evaluation of patients with undiagnosed disorders.**

# First Next Generation Sequencing Report - 2005

## Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies[1]*, Michael Egholm[1]*, William E. Altman[1], Said Attiya[1], Joel S. Bader[1], Lisa A. Bemben[1], Jan Berka[1], Michael S. Braverman[1], Yi-Ju Chen[1], Zhoutao Chen[1], Scott B. Dewell[1], Lei Du[1], Joseph M. Fierro[1], Xavier V. Gomes[1], Brian C. Godwin[1], Wen He[1], Scott Helgesen[1], Chun He Ho[1], Gerard P. Irzyk[1], Szilveszter C. Jando[1], Maria L. I. Alenquer[1], Thomas P. Jarvie[1], Kshama B. Jirage[1], Jong-Bum Kim[1], James R. Knight[1], Janna R. Lanza[1], John H. Leamon[1], Steven M. Lefkowitz[1], Ming Lei[1], Jing Li[1], Kenton L. Lohman[1], Hong Lu[1], Vinod B. Makhijani[1], Keith E. McDade[1], Michael P. McKenna[1], Eugene W. Myers[2], Elizabeth Nickerson[1], John R. Nobile[1], Ramona Plant[1], Bernard P. Puc[1], Michael T. Ronan[1], George T. Roth[1], Gary J. Sarkis[1], Jan Fredrik Simons[1], John W. Simpson[1], Maithreyan Srinivasan[1], Karrie R. Tartaro[1], Alexander Tomasz[3], Kari A. Vogt[1], Greg A. Volkmer[1], Shally H. Wang[1], Yong Wang[1], Michael P. Weiner[4], Pengguang Yu[1], Richard F. Begley[1] & Jonathan M. Rothberg[1]

**454 Life Sciences**

**Nature 437 (7057) 376-380**

# NGS Process Steps

Genomic DNA

↓

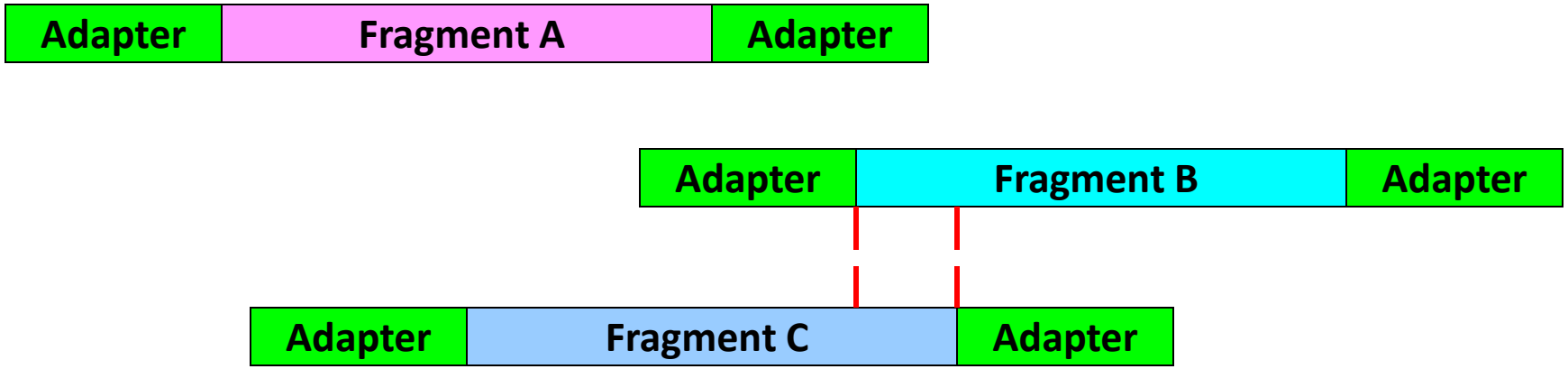
Fragmentation (150 – 400 bp)

↓

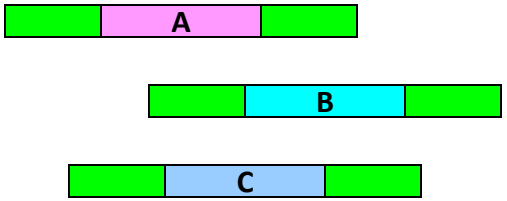
Repair Ends and Ligate Oligonucleotide Adapters

↓

"Randomly Overlapping Fragment Library"



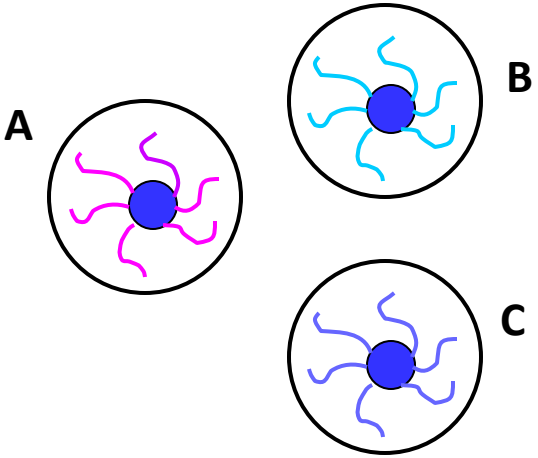
# NGS Process Steps



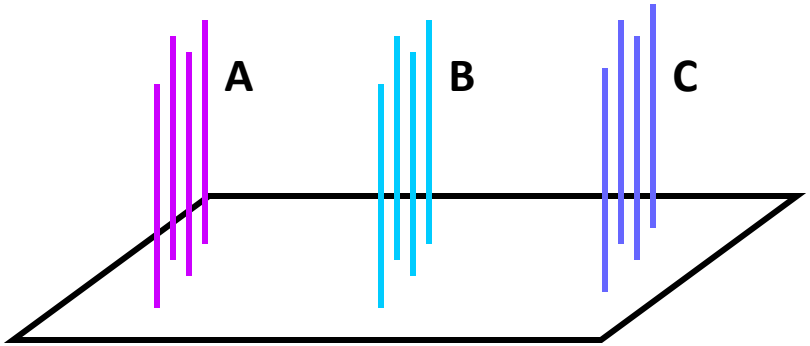

**“Fragment Library”**

**Clonal Amplification of Each Fragment**

**Emulsion PCR**




**On Flow Cell Surface**

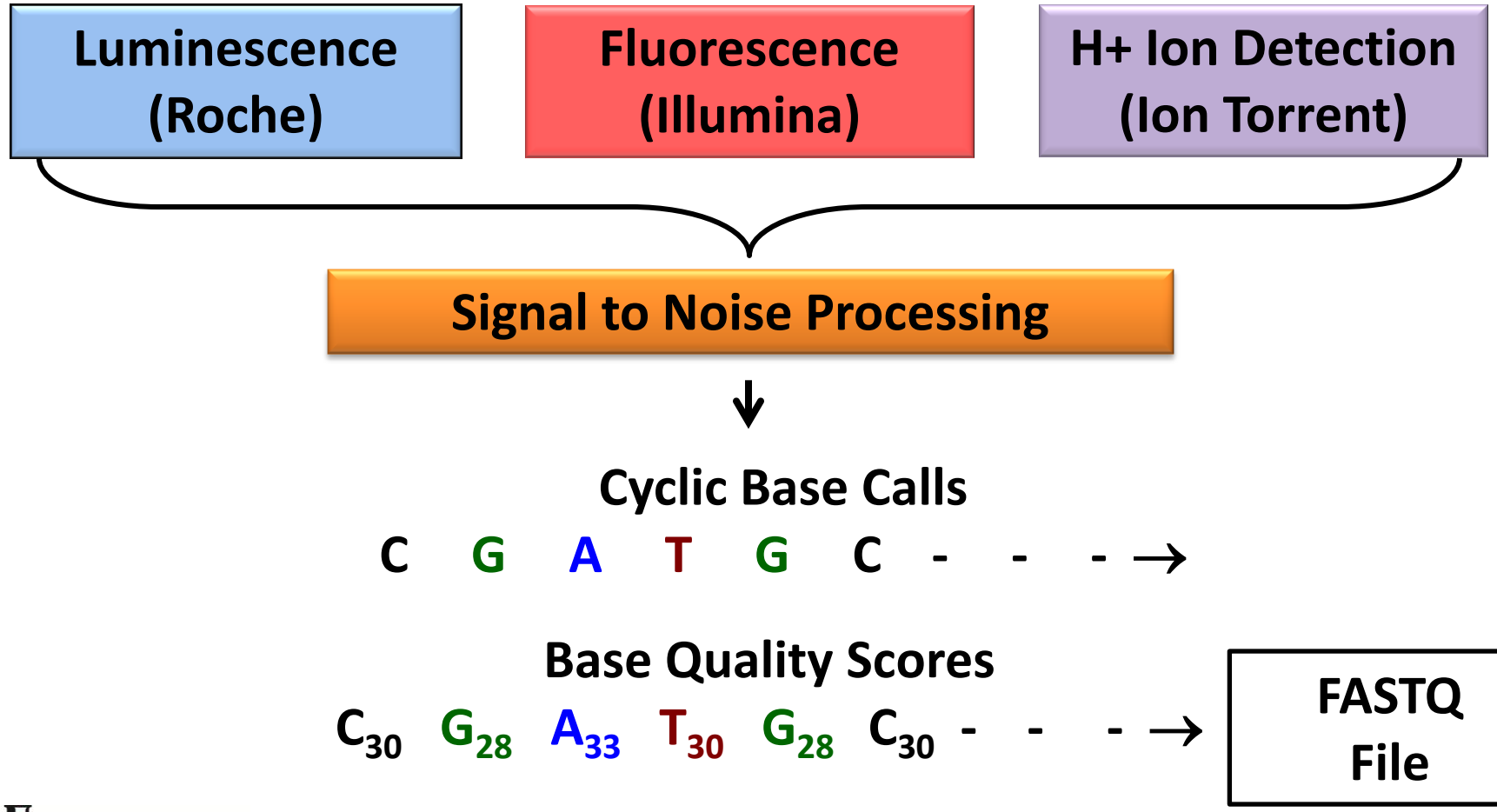



**Sequencing of Clonal Amplicons in a Flow Cell**

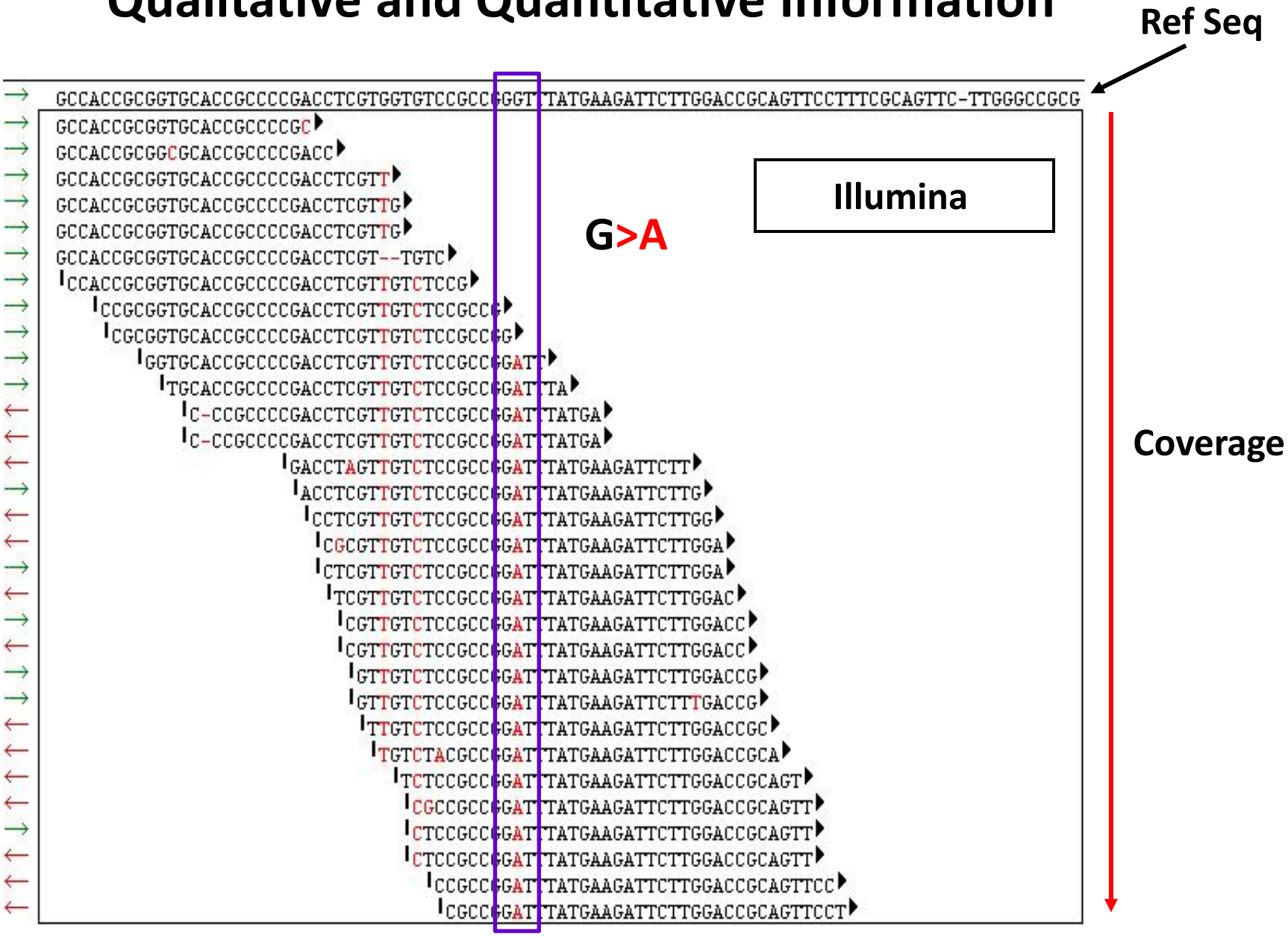
# Next Generation Sequencing

## Flow Cell – High Throughout Process

## Sequential Introduction of Nucleotides to Build Sequence



# Qualitative and Quantitative Information



# NGS Platform Summary

**Table 65-1. NGS Platforms and Specifications**

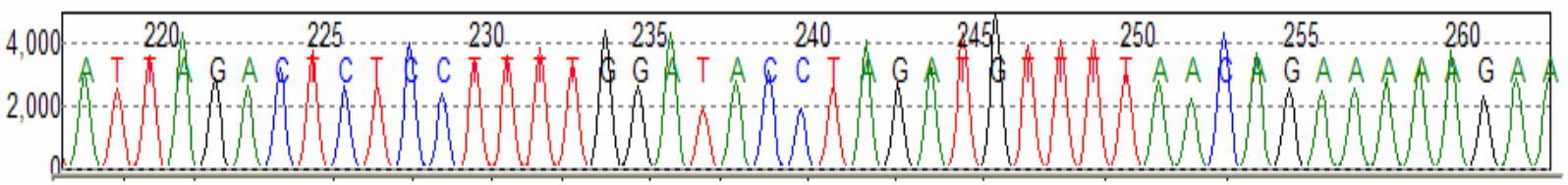
| Platform | Template Preparation | Chemistry | Read Length* | Run Time** | Throughput‡ | Primary Errors | Error Rates† |
|---|---|---|---|---|---|---|---|
| **Roche 454** | | | | | | | |
| GS Junior | ePCR | Pyrosequencing | 400 | 10 h | 35 Mb | Indel | ~1 |
| GS FLX+ | ePCR | Pyrosequencing | 700-1,000 | 23 h | 700 Mb | Indel | ~1 |
| **Illumina** | | | | | | | |
| MiSeq | Bridge Amplification | Reversible Dye Terminators | 36-250 | 4-40 h | 600 Mb-8 Gb | Substitution | ~ 0.5-1 |
| HiSeq 2000 | Bridge Amplification | Reversible Dye Terminators | 100 | 11 days | 600 Gb | Substitution | ~ 0.5-1 |
| HiSeq 2500 Rapid Run Mode | Bridge Amplification | Reversible Dye Terminators | 150 | 27 h | 120 Gb | Substitution | ~ 0.5-1 |
| **Ion Torrent** | | | | | | | |
| PGM | ePCR | Hydrogen Ion Sensing | 100-200 | 2.5-4.5 h | 500Mb-1Gb | Indel | ~ 0.5-2 |
| Proton | ePCR | Hydrogen Ion Sensing | 200 | ~ 4+ h | Up to 10 Gb | Indel | ~ 0.5-2 |

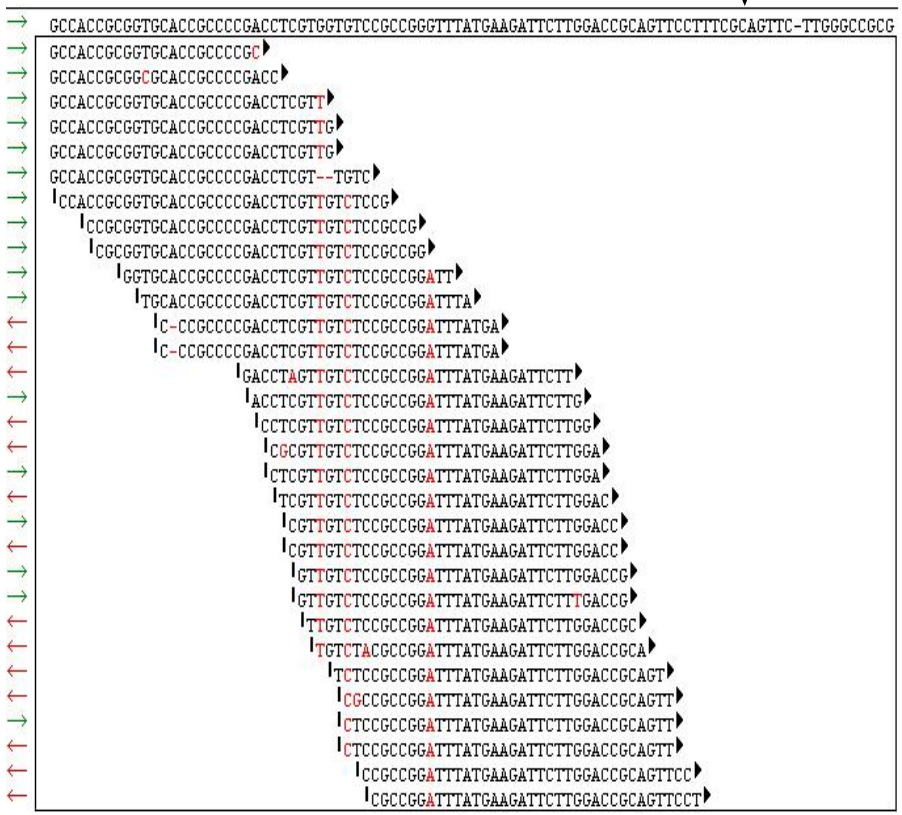
*= Read Length in Bases
**= Run Time Varies with Read Length and Single versus Pair End Sequencing
‡= Throughput Varies with Read Length and Single versus Pair End Sequencing
†= Percentage of Errors per Base within Single Reads at Maximum Read Length as Reported by Vendor and Literature

**FDA Submission Cystic Fibrosis**

Voelkerding et al

# Genetic Testing Paradigm Shift

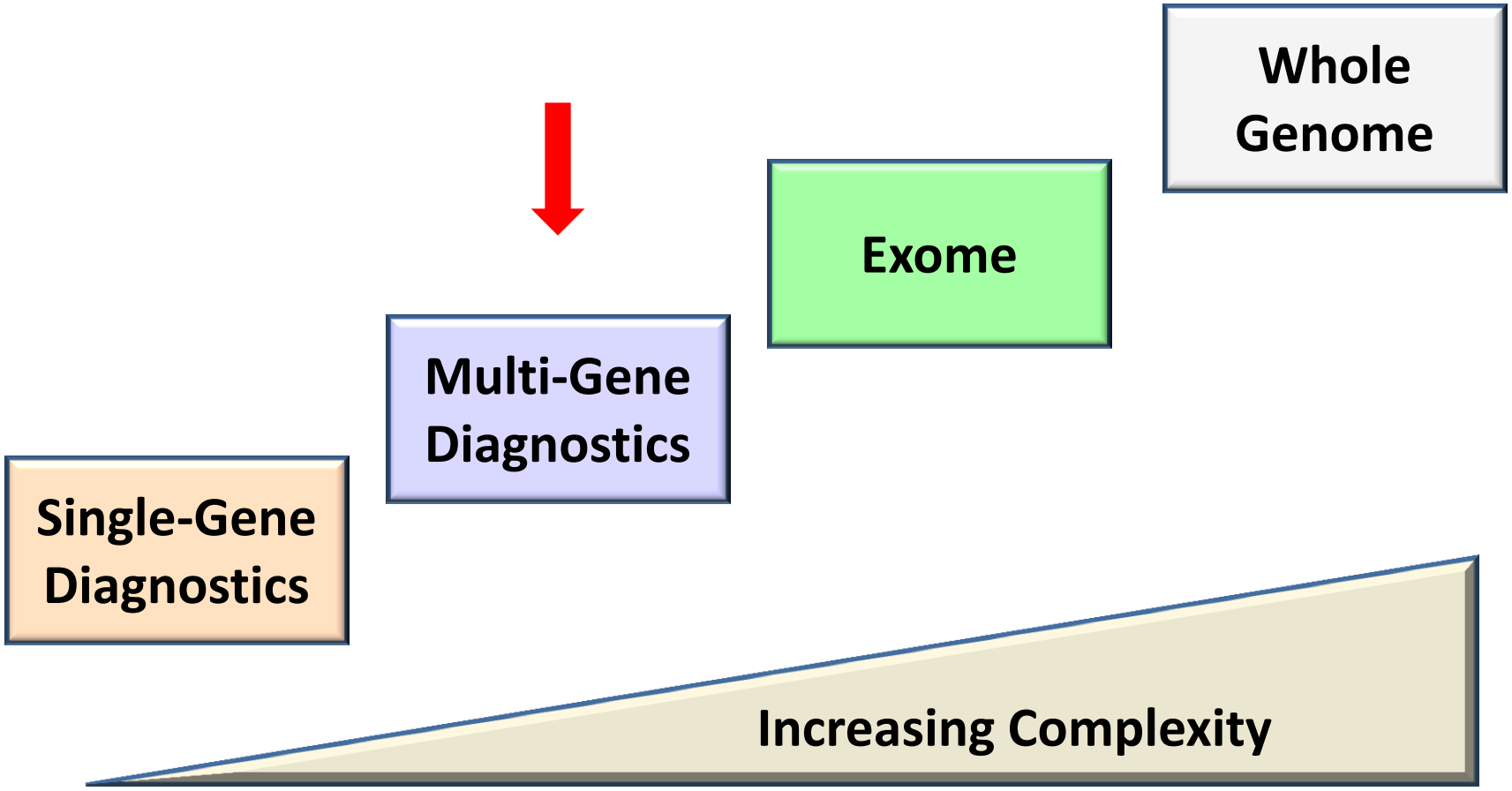


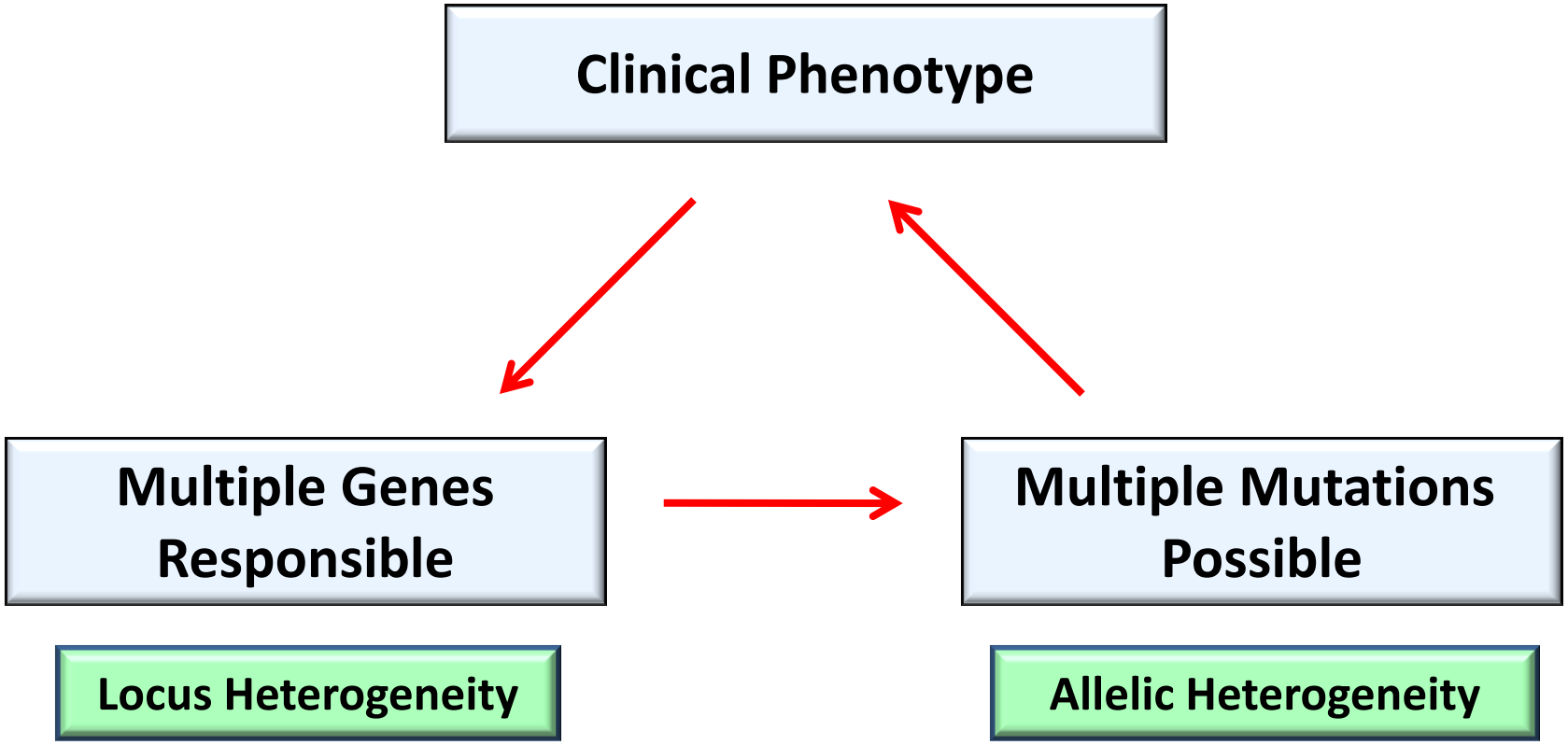
**Sanger Sequencing**
**Qualitative**

**Next Generation Sequencing**
**Qualitative and Quantitative**
**High Throughput**

# New Landscape of Genetic Testing

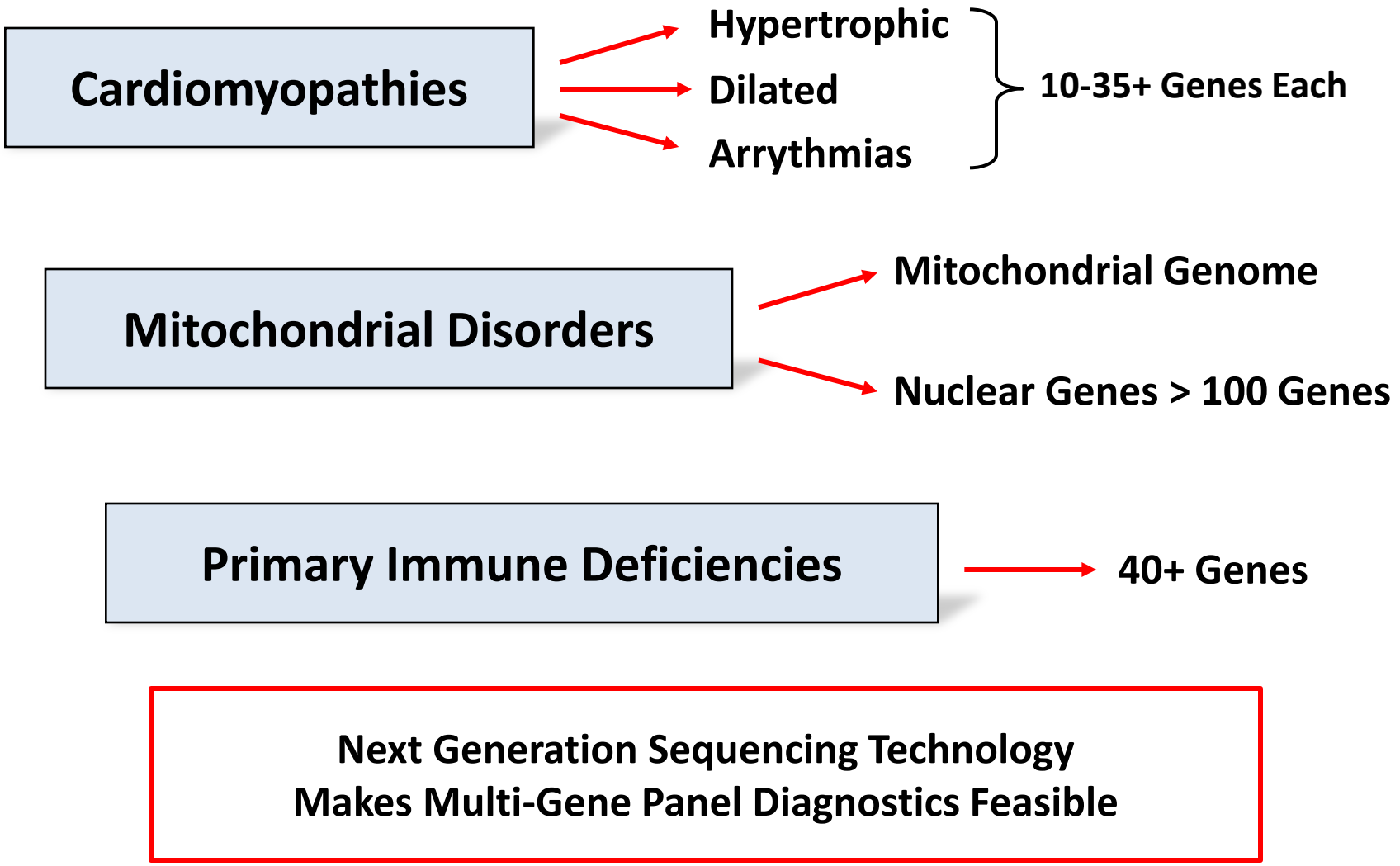


# Multi-Gene Panel Diagnostics



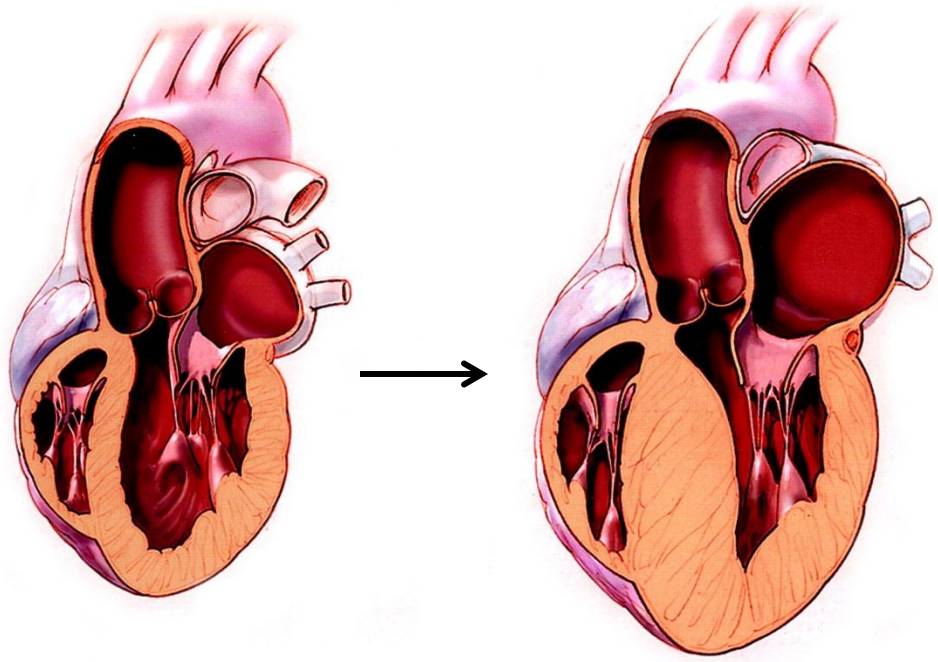

**Technically Difficult to Test For by Sanger Sequencing**

# Multi-Gene Panel Diagnostics



# Hypertrophic Cardiomyopathy – Model for Multi-Gene Diagnostics

**Prevalence = ~ 1 in 500 – 1,000**



**Normal** **Hypertrophic**

**Teenage to Adult Onset**

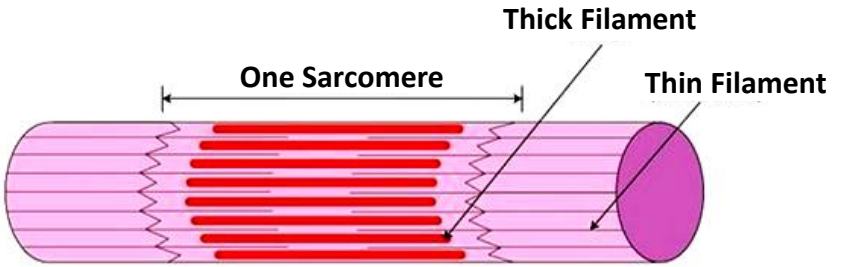
**Autosomal Dominant**

**Arrythmias/Angina**
**Sudden Death**

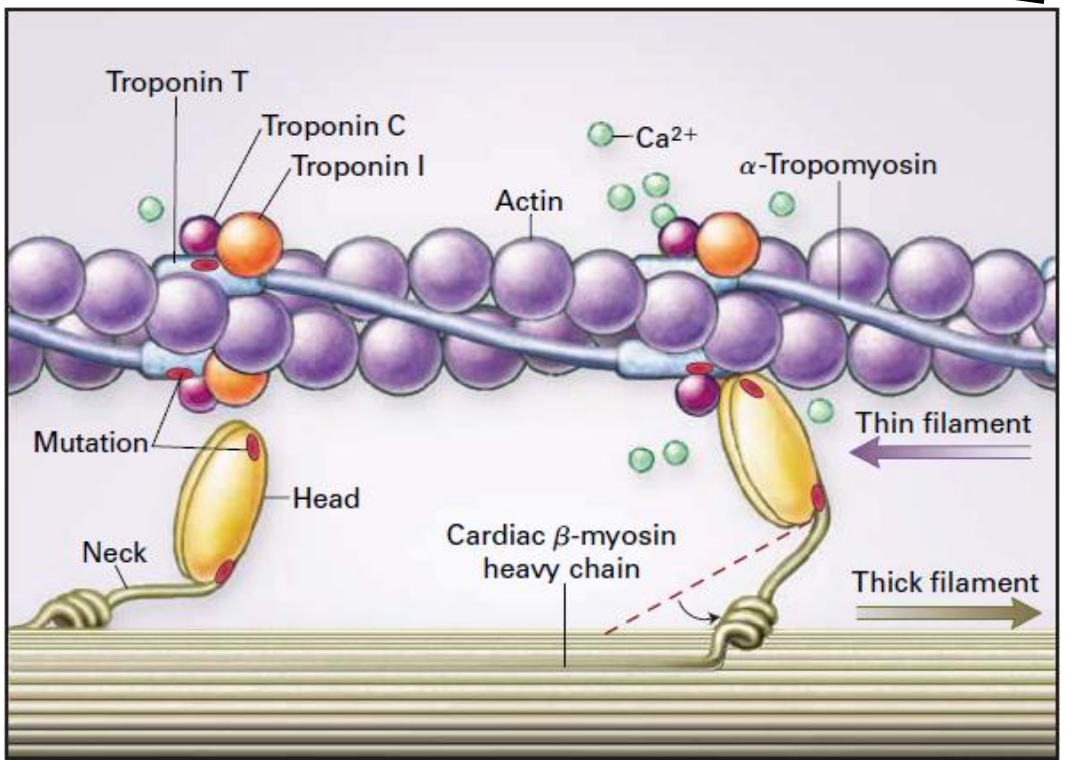
Nishimura RA, et al. Circulation 11;108(19)

# HCM – Genetic Disorder of Cardiac Sarcomere

**Myofibril**



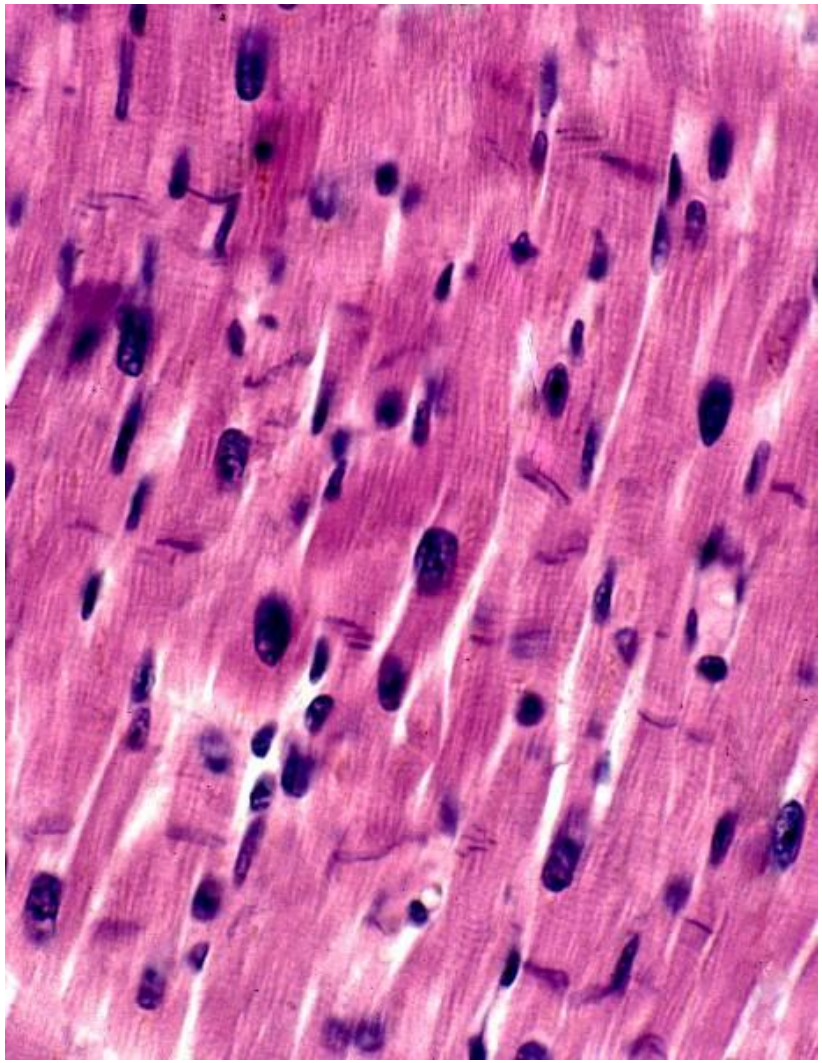

**Sarcomere**





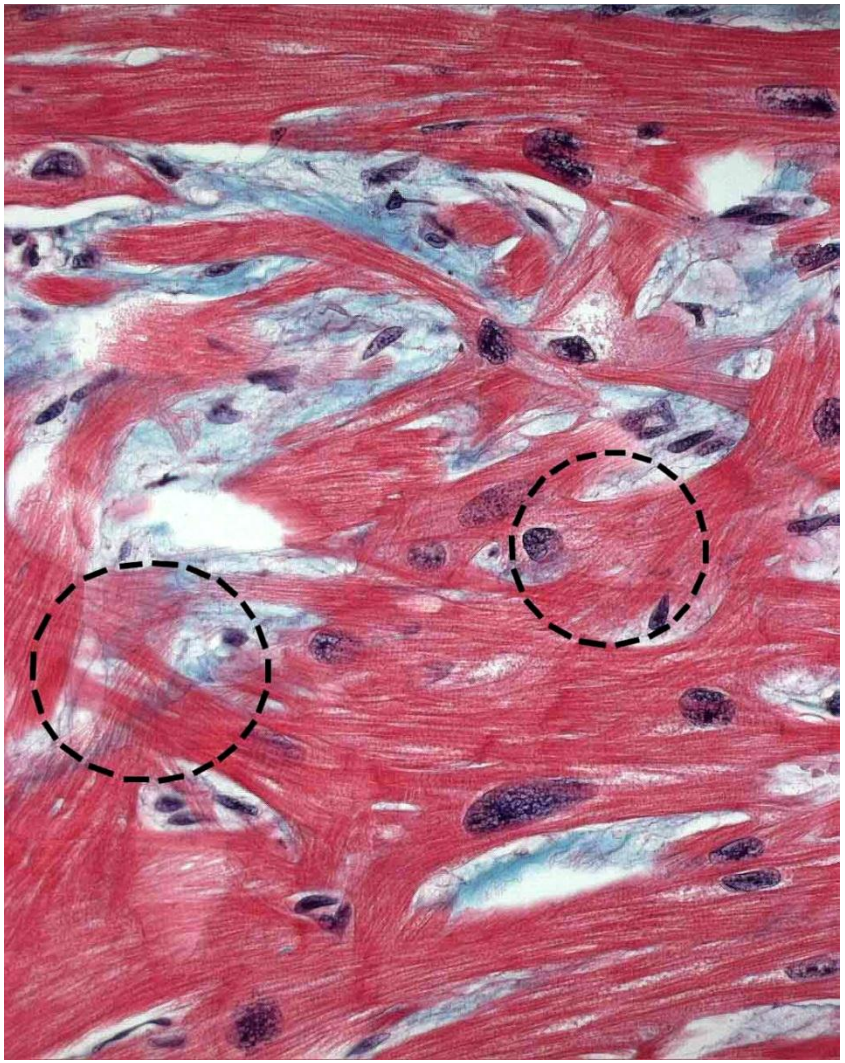

Kamisago et al. NEJM 343(23):1688

# Normal Myocytes



http://www.umanitoba.ca

# HCM – Myocyte Disarray



Soor et al, J Clin Pathol 2008

# Hypertrophic Cardiomyopathy Genes

| Protein | Gene | Mutations | | Gene Size bp |
|---|---|---|---|---|
| Myosin, heavy chain 7 | *(MYH7)* | 193 | Sanger | 32,628 |
| Myosin binding protein C | *(MYBPC3)* | 138 | | 28,280 |
| Troponin T type 2 | *(TNNT2)* | 33 | | 25,673 |
| Troponin I type 3 | *(TNNI3)* | 32 | | 12,963 |
| Cysteine and glycine-rich protein 3 | *(CSRP3)* | 12 | | 27,024 |
| Tropomyosin 1, α | *(TPM1)* | 11 | | 36,274 |
| Myosin, light chain 2 | *(MYL2)* | 10 | | 16,758 |
| Actin | *(ACTC)* | 7 | | 14,631 |
| Myosin, light chain 3 | *(MYL3)* | 5 | | 12,617 |
| Protein kinase, AMP-activated, γ2 | *(PRKAG2)* | 4 | | 328,114 |
| Phospholamban | *(PLN)* | 2 | | 19,112 |
| Troponin C type 1 | *(TNNC1)* | 1 | | 9,041 |
| Titin | *(TTN)* | 2 | | 281,434 |
| Myosin, heavy chain 6 | *(MYH6)* | 2 | | 32,628 |
| Titin-cap | *(TCAP)* | 2 | | 9,361 |
| Caveolin 3 | *(CAV3)* | 1 | | 20,199 |
| | | 455 | | 906,737 |

# Hypertrophic Cardiomyopathy Genes

| Protein | Gene | Mutations | | Gene Size bp |
|---|---|---|---|---|
| Myosin, heavy chain 7 | *(MYH7)* | 193 | | 32,628 |
| Myosin binding protein C | *(MYBPC3)* | 138 | | 28,280 |
| Troponin T type 2 | *(TNNT2)* | 33 | | 25,673 |
| Troponin I type 3 | *(TNNI3)* | 32 | | 12,963 |
| Cysteine and glycine-rich protein 3 | *(CSRP3)* | 12 | | 27,024 |
| Tropomyosin 1, α | *(TPM1)* | 11 | | 36,274 |
| Myosin, light chain 2 | *(MYL2)* | 10 | NGS | 16,758 |
| Actin | *(ACTC)* | 7 | | 14,631 |
| Myosin, light chain 3 | *(MYL3)* | 5 | | 12,617 |
| Protein kinase, AMP-activated, γ2 | *(PRKAG2)* | 4 | | 328,114 |
| Phospholamban | *(PLN)* | 2 | | 19,112 |
| Troponin C type 1 | *(TNNC1)* | 1 | | 9,041 |
| Titin | *(TTN)* | 2 | | 281,434 |
| Myosin, heavy chain 6 | *(MYH6)* | 2 | | 32,628 |
| Titin-cap | *(TCAP)* | 2 | | 9,361 |
| Caveolin 3 | *(CAV3)* | 1 | | 20,199 |
| | | 455 | | 906,737 |

# Hypertrophic Cardiomyopathy – Model for Multi-Gene Diagnostics

## Value of Genetic Testing

- Confirm Genetic Etiology
- Specific Mutation Identification
- Family Risk Counseling/Testing

## Medical Management

Beta and Calcium Channel Blockers

Antiarrythmics – Cardioversion – Implantable Defibrillators

Transplantation

# Multi-Gene Panel Diagnostics

**More Comprehensive**
**Compared to Single Gene Sanger Sequencing**

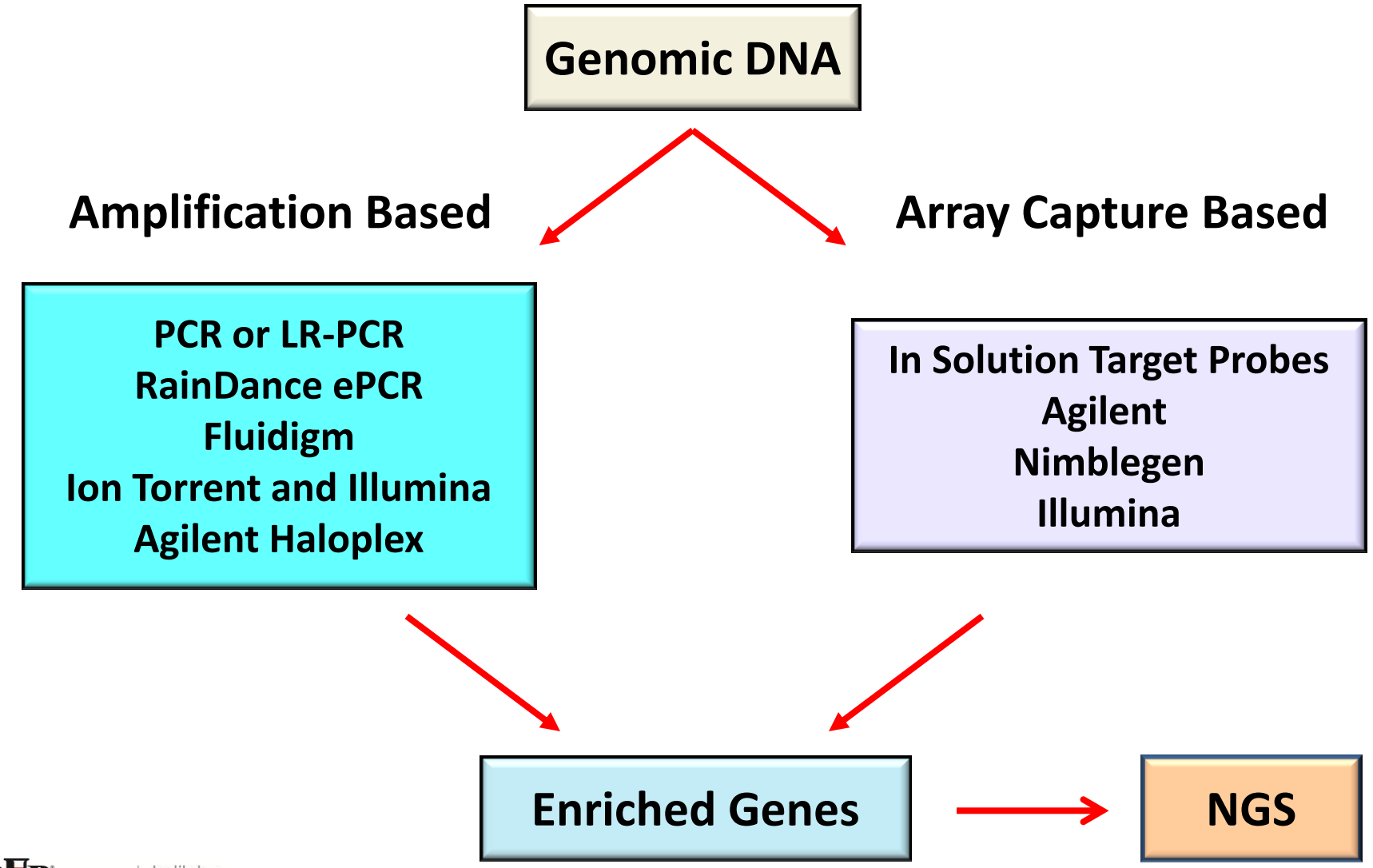
**Gene Content = Based on Current Knowledge**

Illumina MiSeq

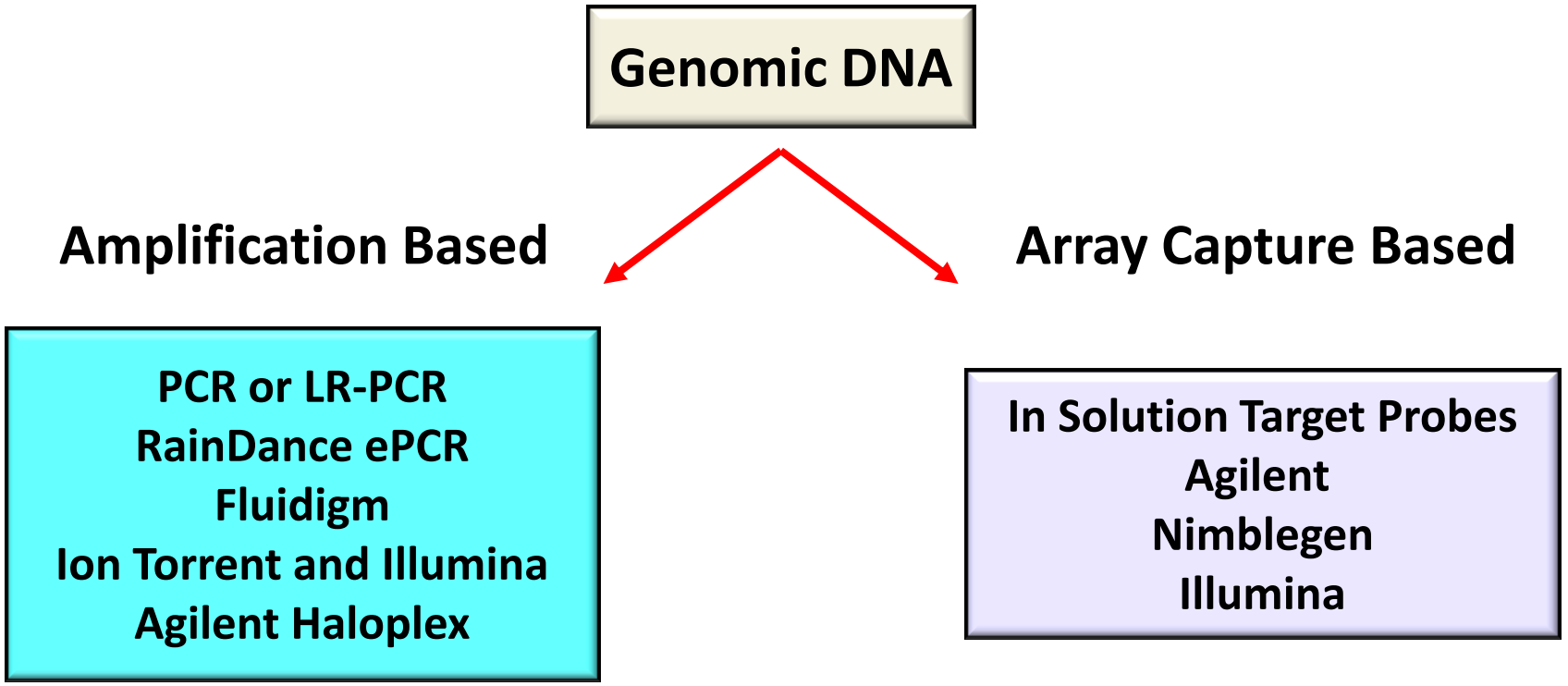
**Facilitated by New Platforms**
**Lower Capital Costs**
**Faster Sequencing Process**

Ion Torrent PGM

# Multi-Gene Diagnostics Require Gene Enrichment



# Multi-Gene Diagnostics Require Gene Enrichment




**Enrichment Method - Difficult Choice - Substantial Cost Investment**

# Considerations in Designing Multi-Gene Panels

## Suitability of Enrichment Method for Laboratory

- Is the Technical Workflow (Manual) Adoptable in Your Setting?
  - Is it Possible to Automate the Workflow?
- Is the Enrichment Method Compatible with Your Sequencing Platform?
  - How Many Samples can be Barcoded and Pooled for Sequencing?
- What Data Analysis Pipeline will be Required?
  - Vendor Supplied or In House Custom Developed

# Considerations in Designing Multi-Gene Panels

## Perform *In Silico* Designs with Enrichment Methods

- **Free Designs Using Vendor Software**
  - **Valuable to Compare Design Results between Method Options**
- **What Percentage of Gene Targets will be Enriched?**
  - **Are there *In Silico* Predicted Problem Areas?**

# Considerations in Designing Multi-Gene Panels

## Expect *In Silico* versus Empiric Results Differences

- **Characterize Problem Areas**
  - **Inadequate Sequence Coverage of Some Target Regions**
  - **Regions where Data Analysis indicates Homologous Sequence Interference**

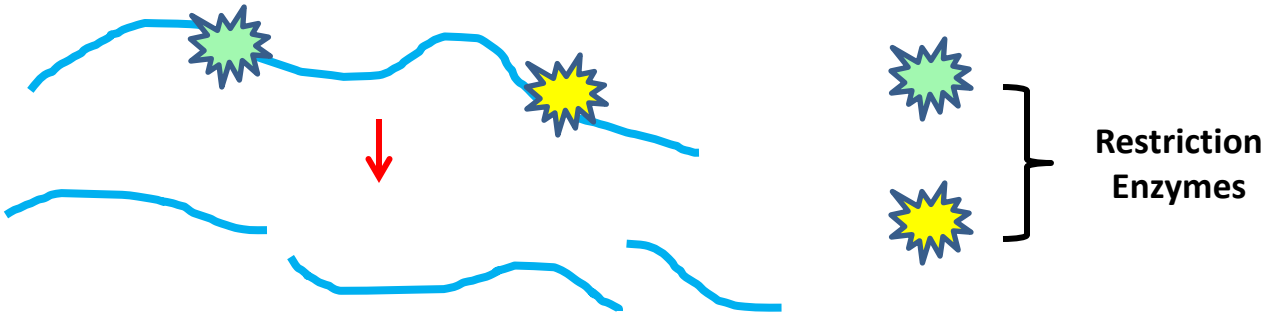
# Case Example Multi-Gene Panel Design

## Project Goal

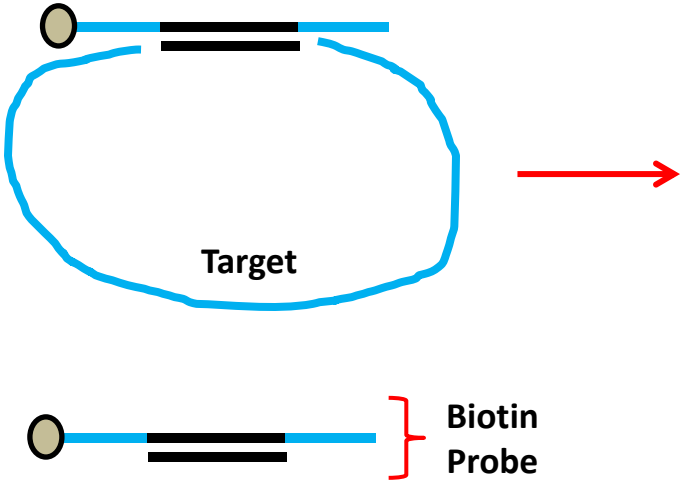
- Multi-Gene Panel for Primary Immune Deficiencies
- Sequencing Platform – Illumina MiSeq
- In Silico Designs Performed and Agilent Haloplex Chosen
- In House Custom Data Analysis

# Haloplex Enrichment Theory and Practice

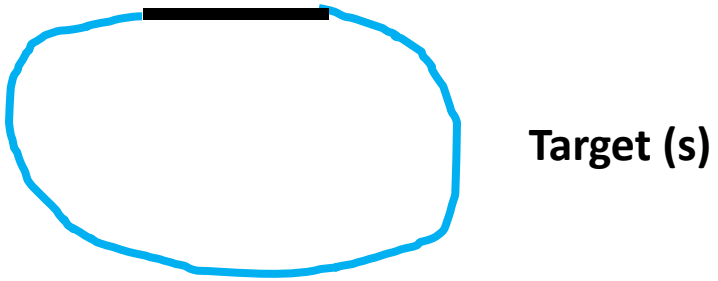
## 1. Digest and Denature Genomic DNA




## 2. Hybridize Biotin Target Probe Library to Form "Tri-Molecular" Circular Complexes

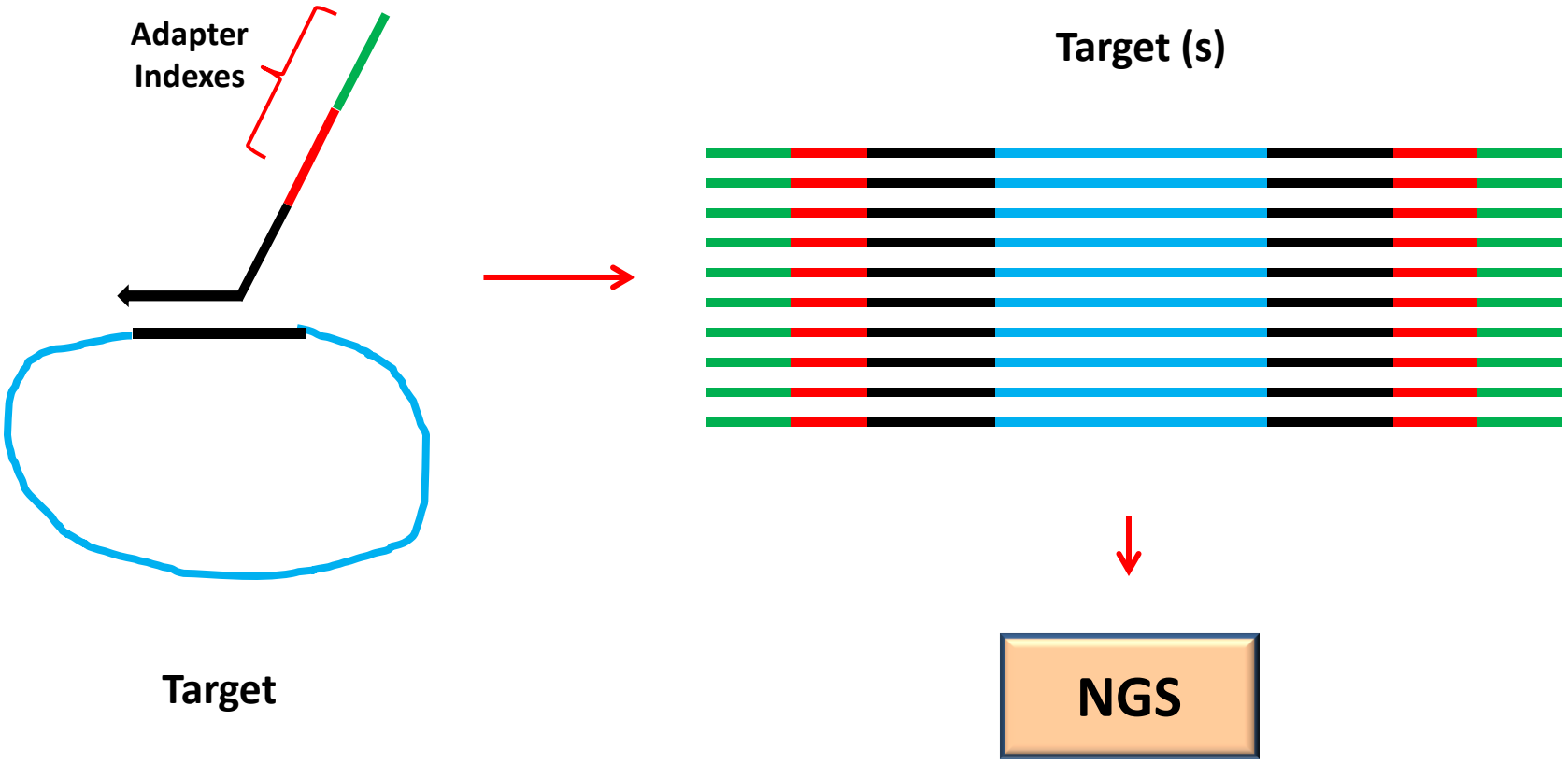


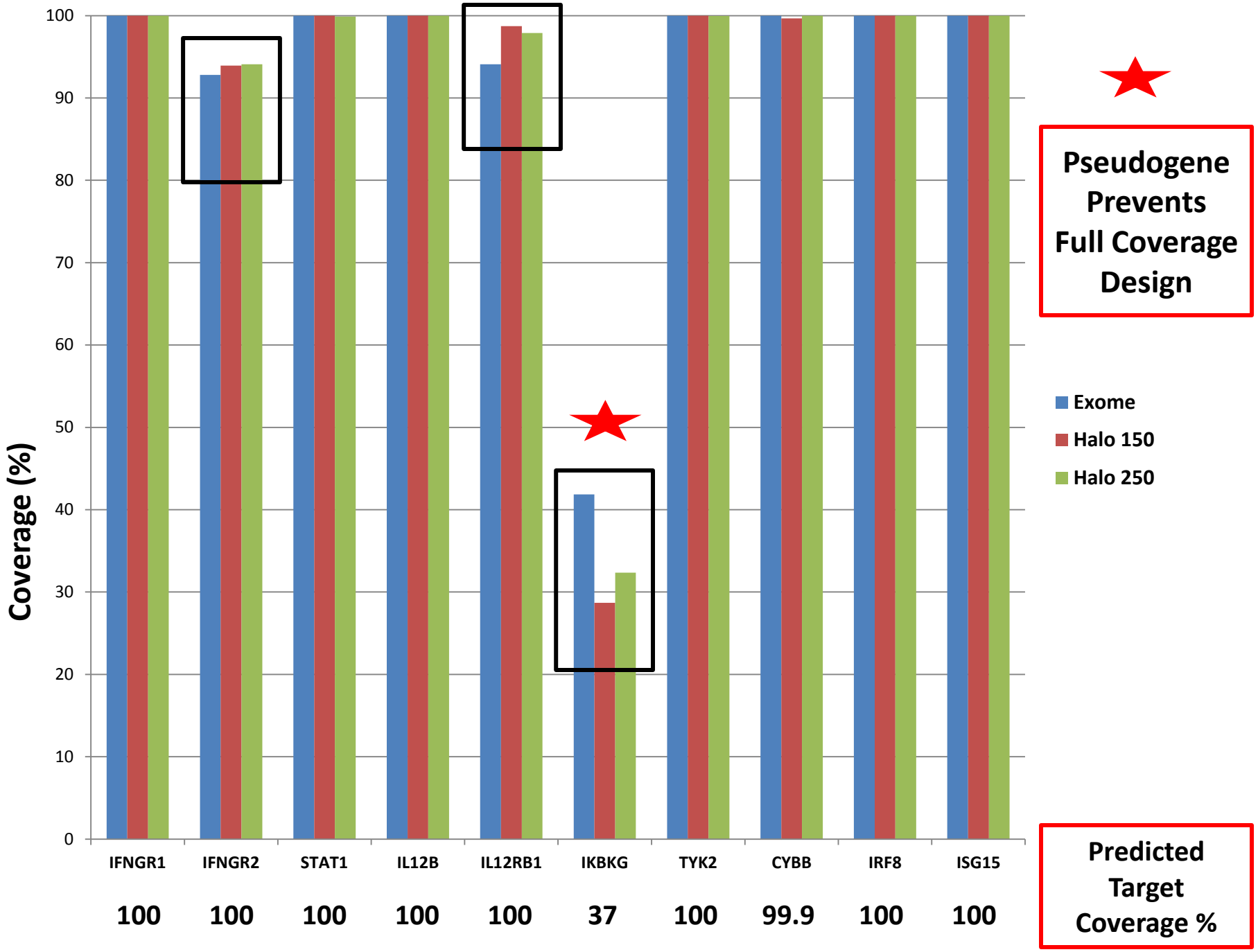

## 3. Capture and Ligate to Form Closed Target Circles



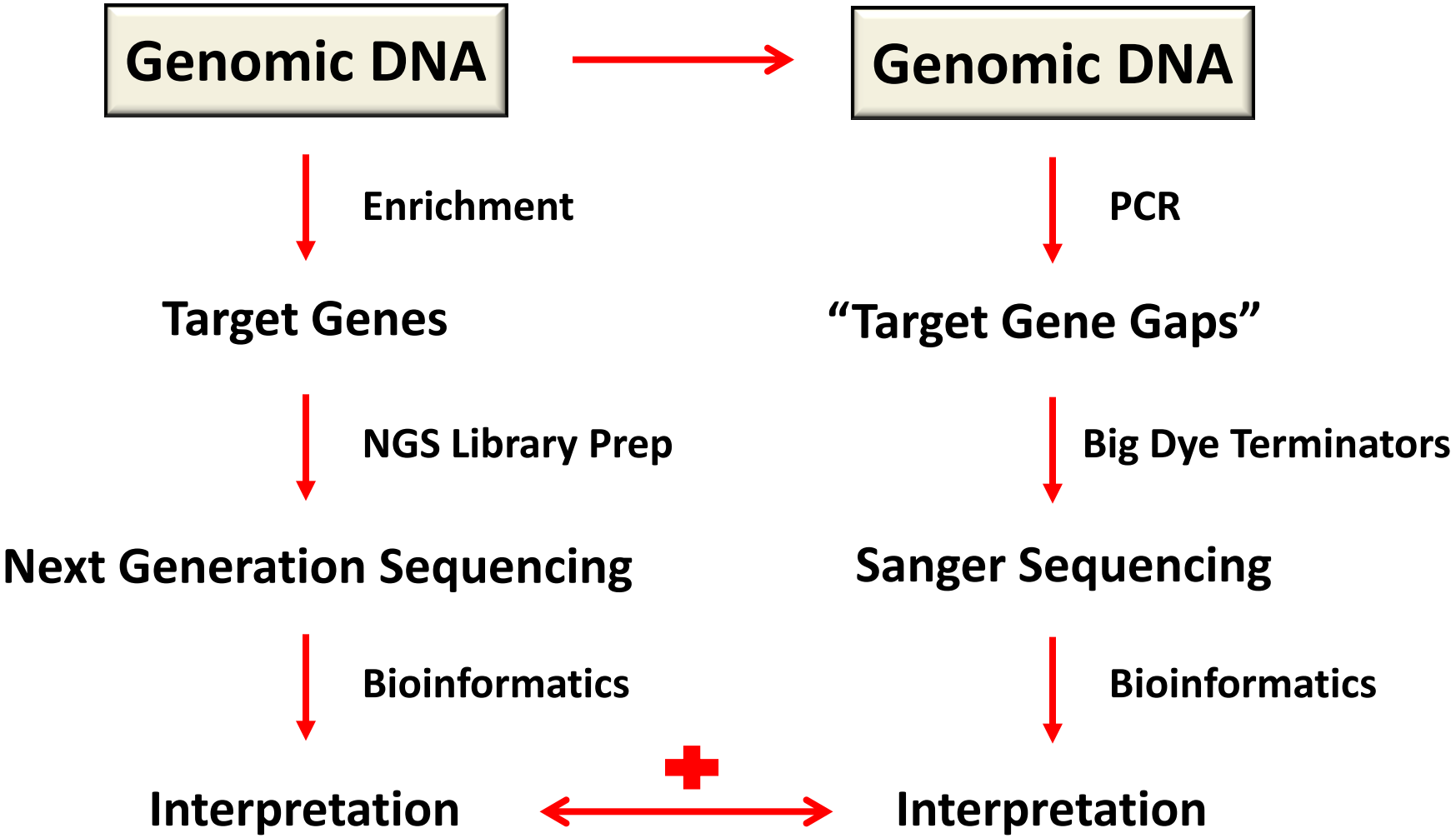
# Haloplex Enrichment Theory and Practice

## 4. PCR Amplify Targets and Incorporate Sequencing Adapters and Indexes




Coverage (%)
100
90
80
70
60
50
40
30
20
10
0
IFNGR1
IFNGR2
STAT1
IL12B
IL12RB1
IKBKG
TYK2
CYBB
IRF8
ISG15
100
100
100
100
100
37
100
99.9
100
100
Exome
Halo 150
Halo 250
Pseudogene Prevents Full Coverage Design
Predicted Target Coverage %

# Addressing "Gaps" in Multi-Gene Panels



# Multi-Gene Panel Diagnostics - Summary

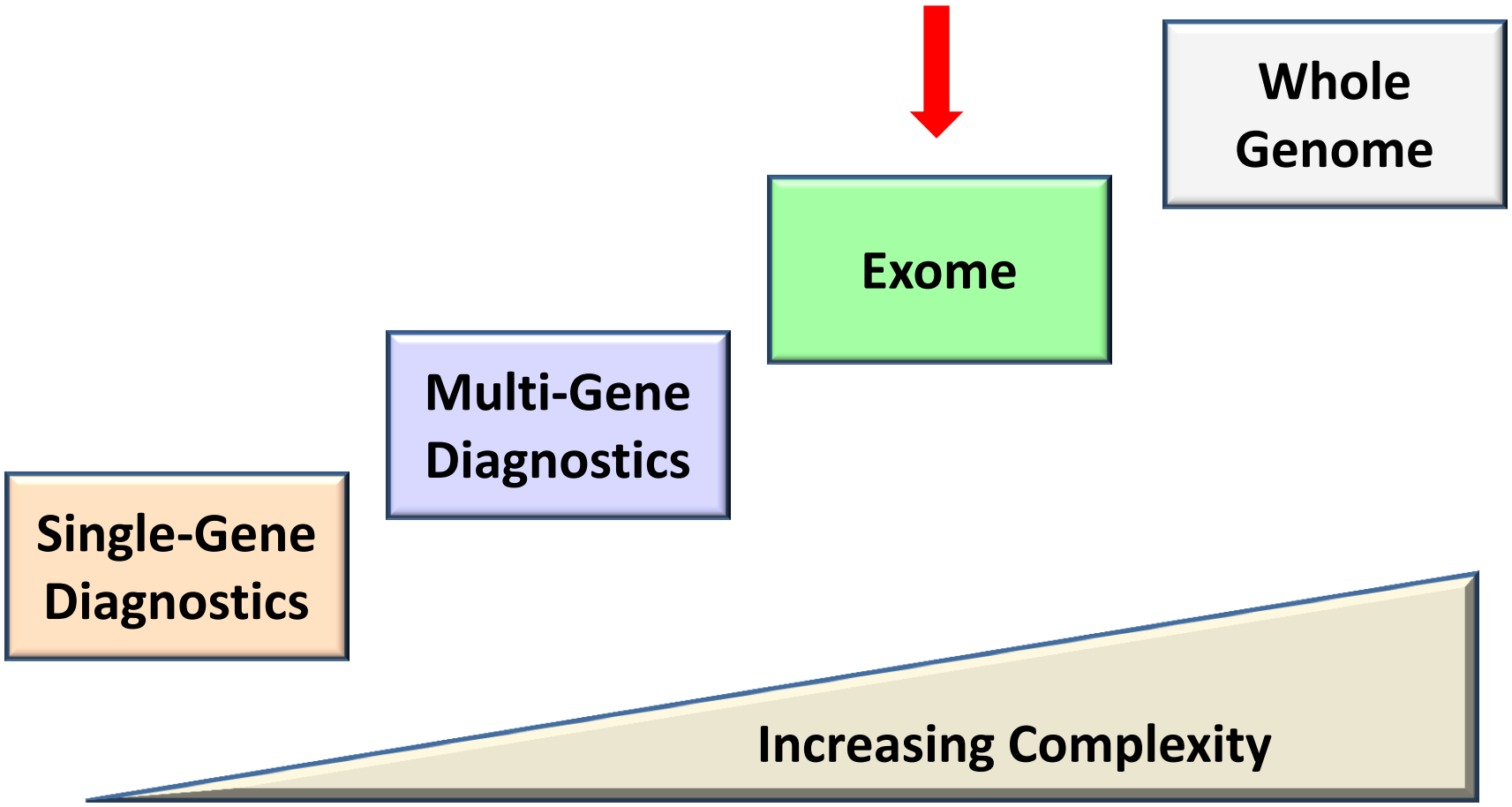
## Becoming a "New First Tier" Approach

- Application to a Growing Number of Inherited Disorders

## Implementation Challenges for Laboratories

- Choosing a Technical Approach
- Assay Optimization and Data Analysis
- Scaling Gene Numbers Increases Interpretive Review Time

# New Landscape of Genetic Testing

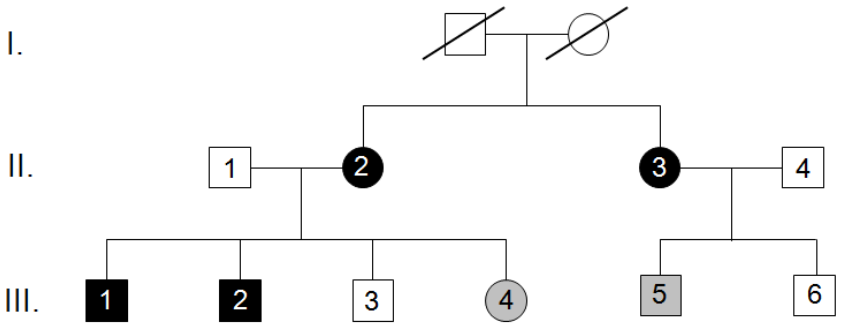


# Human Exome

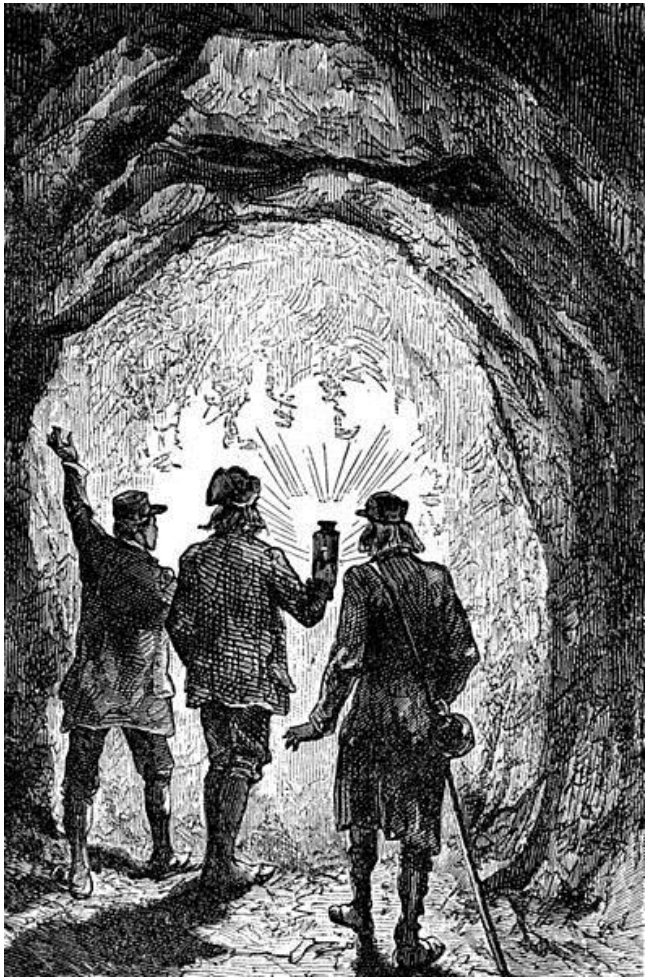
~ 1.5% of the genome

~ 20,500 genes

"Repository" of Mendelian Mutations




# "Center of the Genome"



"Journey to the Center of the Earth"
Jules Verne 1864

# History of Exome Sequencing

"Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing"

Choi et al PNAS 2009 – Congenital Chloride Diarrhea Gene

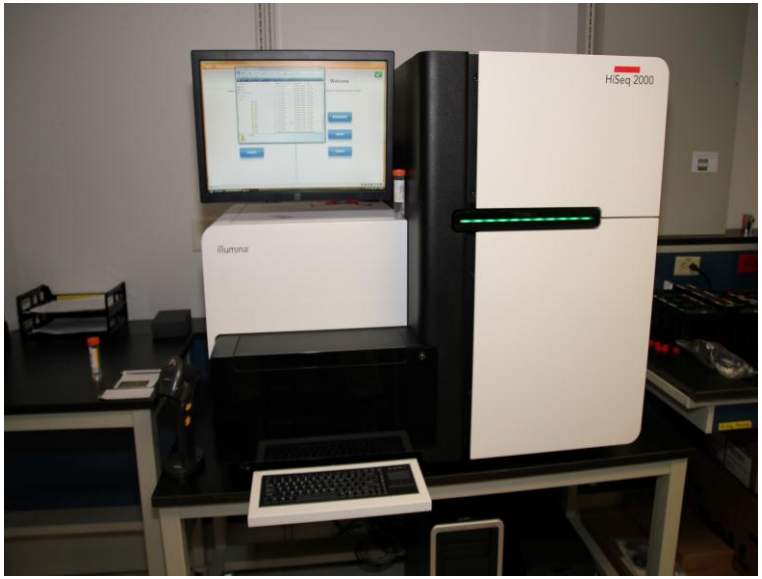
>200 Gene Discoveries
Recessive-Dominant-De Novo
June 2013

OMIM Database - June 2013
7430 Disorders with Known or Suspected Mendelian Inheritance

3,805 Disorders with Molecular Basis Known
Potential for Further Molecular Diagnoses is Substantial

# Platform Options for Exome Sequencing

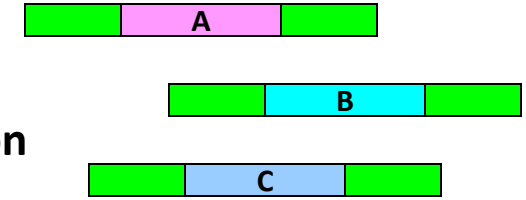
## Illumina HiSeq 2000 or 2500




## Ion Torrent Proton

# Exome Sequencing Laboratory Workflow

**Genomic DNA**




↓ Library Preparation

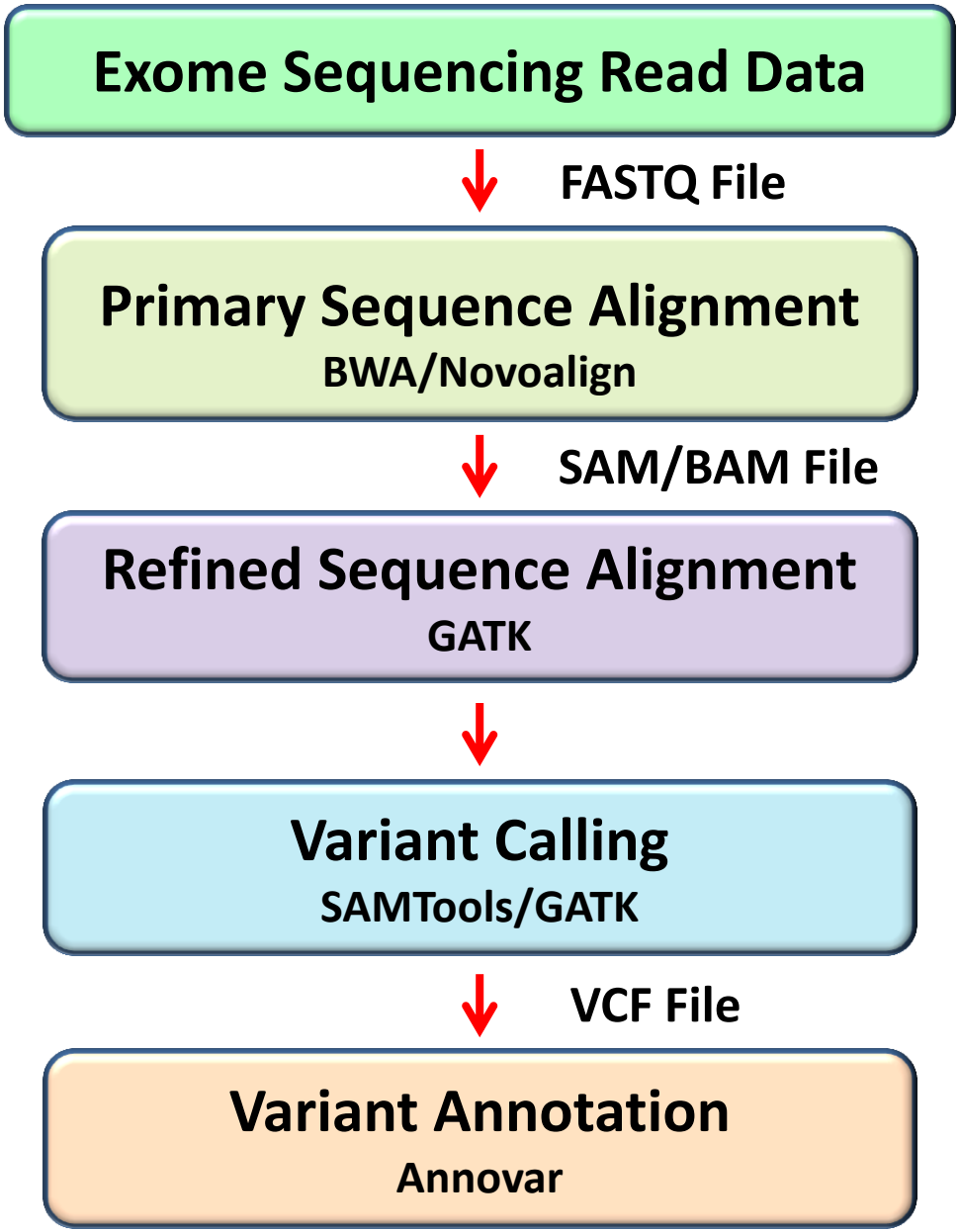
**Next Generation Sequencing Library**

↓ Hybridize to Exome Capture Probes

**Exome Enriched Library**

↓ Next Generation Sequencing

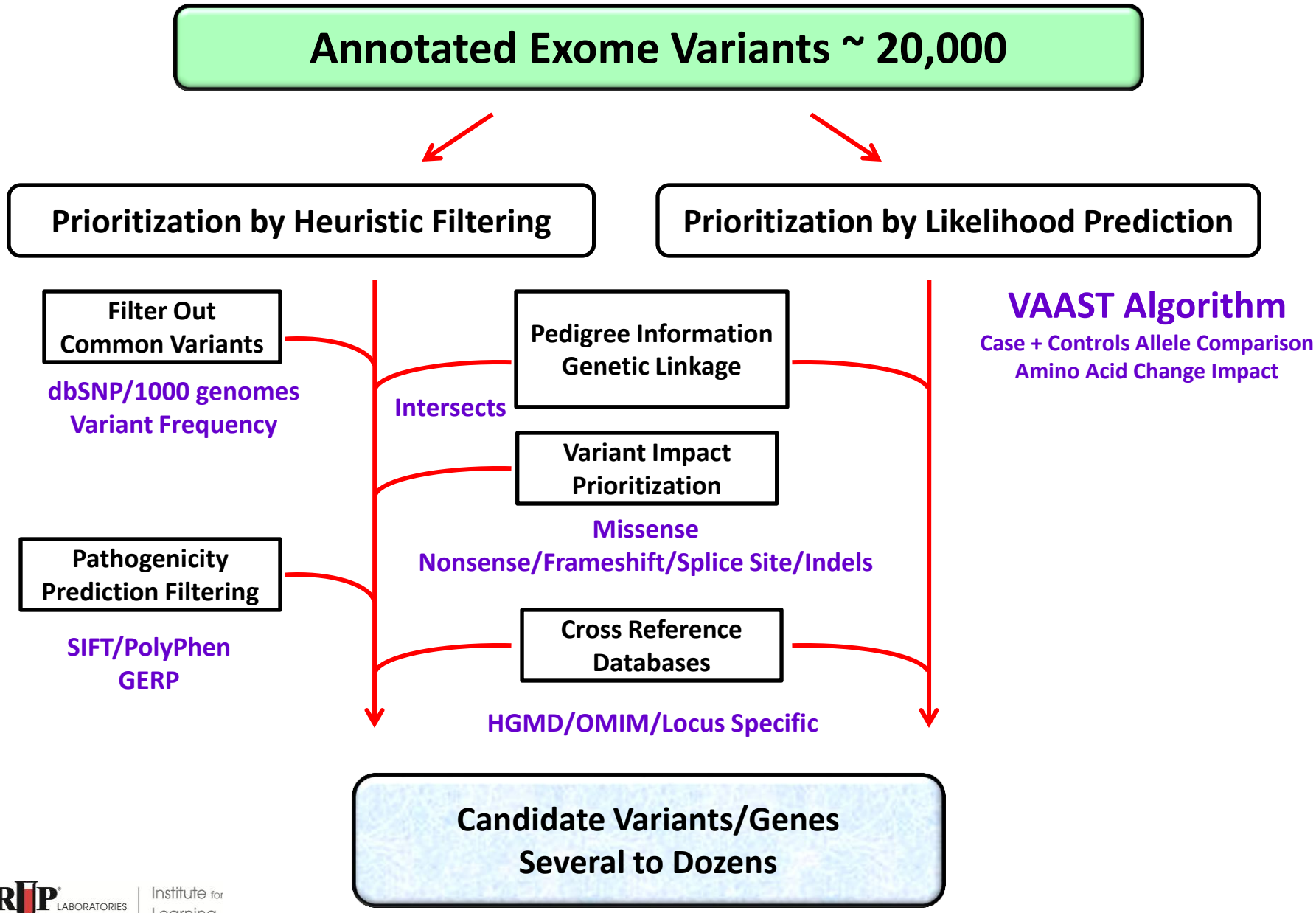
**Bioinformatics Analysis**

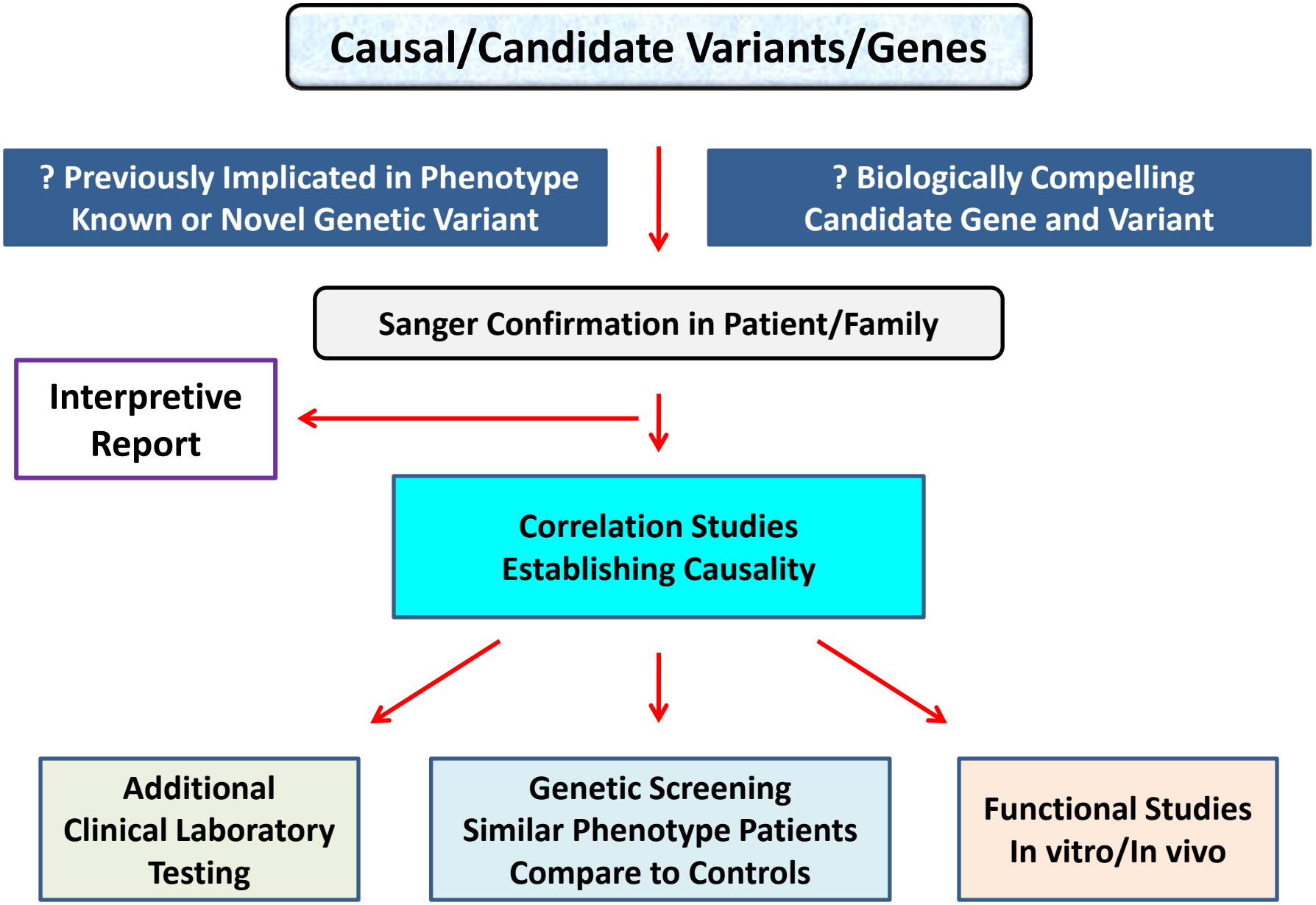



## FastQ File Format

```
@HW-ST573_75:1:1:1353:4122/11
CAATCGAATGGAATTATCGAATGCAATCGA
ATAGAATCATCGAATGGACTCGAATGGAAT
CATCGAA
+
ggfgggggggggggfgggggggfgeggggg
fdfeefeggggggggegbgegegggdeYed
ggggggeg
@HW-ST573_75:1:1:1347:4151/11
ATCTGTTCTTGTCTTTAACTCTCAAGGCAC
CACCTTCCATGGTCAATAATGAACAACGCC
AGCATGC
+
effffgggggggggggfgggggggggggg
gdggggfgggfgdggaffffgfggffgdgg
ggggdfg
@HW-ST573_75:1:1:1485:4153/11
GAGGAGAGATATTTTGACTTCCTCTCTTCA
TATTTGGATGCTTTTTACTTATCTCTCTTG
ACTAATT
+
dZdddbXc`_ccccbeeedbeaedeeeee^
aeeedcaZca_`^c[eeeeed]eeecd[dd
^eeba[d
```

# Workflow for Causal/Candidate Gene Identification




Causal/Candidate Variants/Genes
? Previously Implicated in Phenotype
Known or Novel Genetic Variant
? Biologically Compelling
Candidate Gene and Variant
Sanger Confirmation in Patient/Family
Interpretive Report
Correlation Studies
Establishing Causality
Additional
Clinical Laboratory
Testing
Genetic Screening
Similar Phenotype Patients
Compare to Controls
Functional Studies
In vitro/In vivo

# Criteria for Choosing Patients for Exome Sequencing

**Genetic Etiology Strongly Suspected**

**Standard Testing Negative or Impractical**

**Diagnosis Likely to Impact Treatment and/or Management Decisions**

**Diagnostic Yield is Greater in Family Studies**

- **Families with Multiple Affected Members**

## Exome Sequencing – "Diagnostic Yield"

**Difficult to Determine [Yet]**

**Currently: Largely Single Case Reports**
**Anecdotal Series ~20-30% Diagnosis**

**NIH Undiagnosed Disease Program – 2011 Report**
**5 Molecular Diagnoses in 30 Patients/Families (17%)**
**Several Compelling Candidate Genes**

# Exome Sequencing – "Diagnostic Yield"

## Diagnostic Yield Expected to Increase - By How Much ?

**Driving Forces**

**Increasingly Sophisticated Bioinformatics Will Improve Variant Detection**

**Growth in Knowledge Base of Disease Causing Genes and Variants**

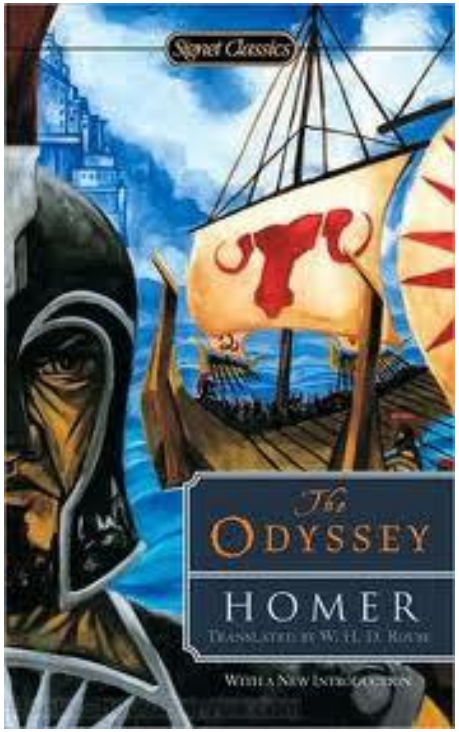
+

**Conversion to Whole Genome Sequencing**

- **Filling in the Gaps**

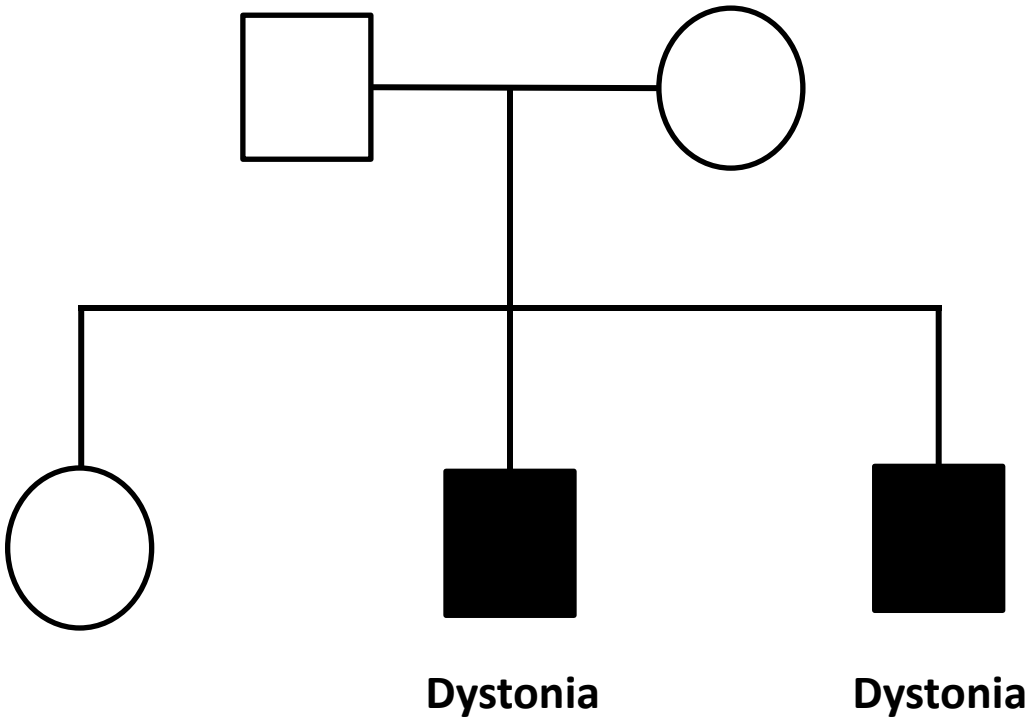
# Exome Sequencing – Case Vignette

## “Diagnostic Odyssey”




**8th Century BC**

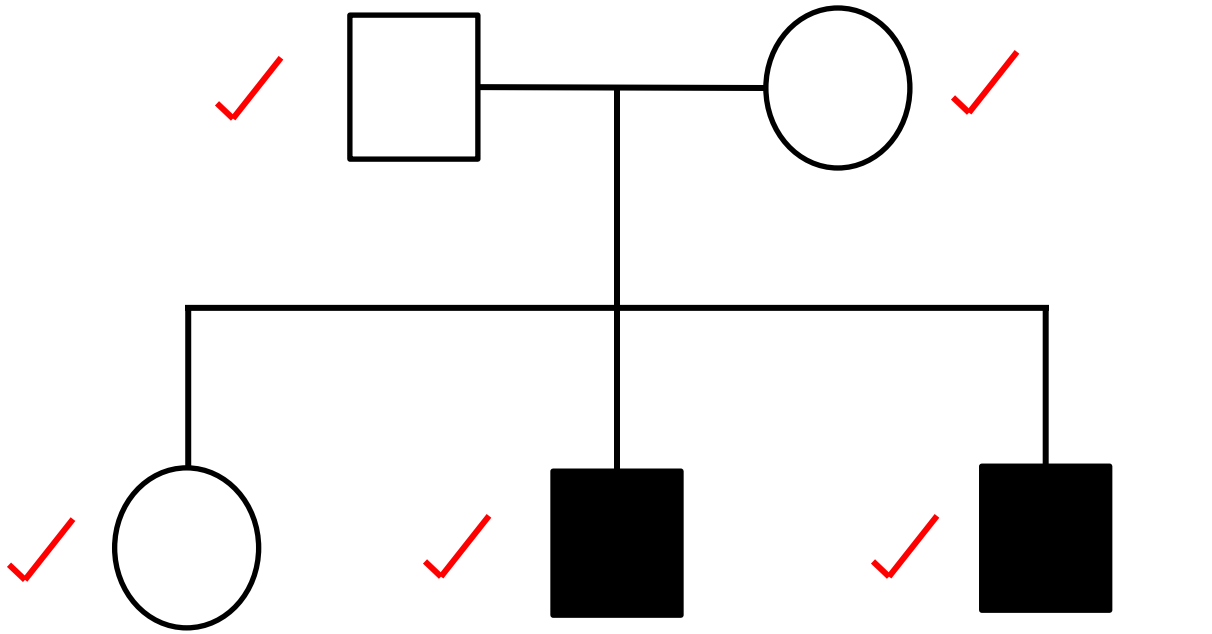
# Exomes for “Diagnostic Odyssey”




**First Year of Life: Seizures/Dystonia**

**Third Year of Life: MRI with Leukodystrophy**

# Exomes for “Diagnostic Odyssey”

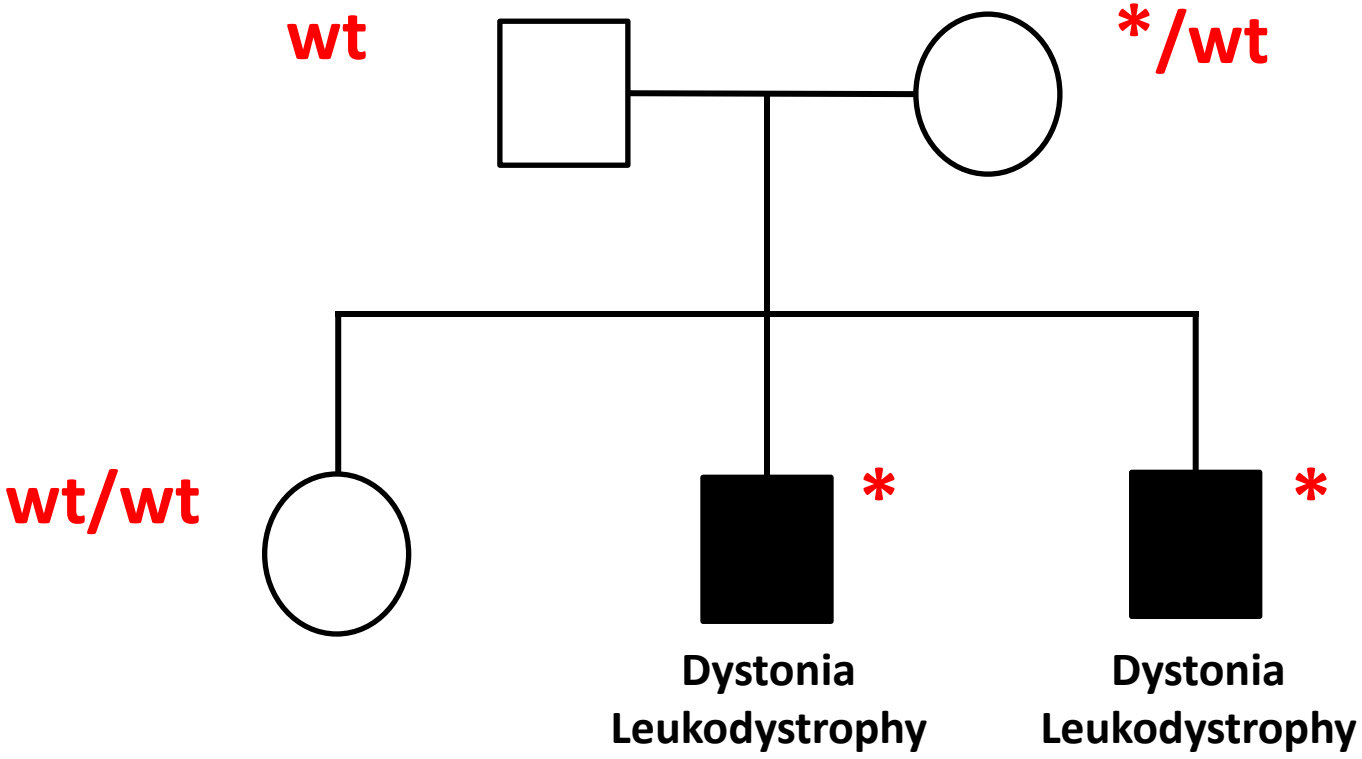



## Heuristic Filtering + VAAST + Interpretive Review

**Top Three Candidate Genes**

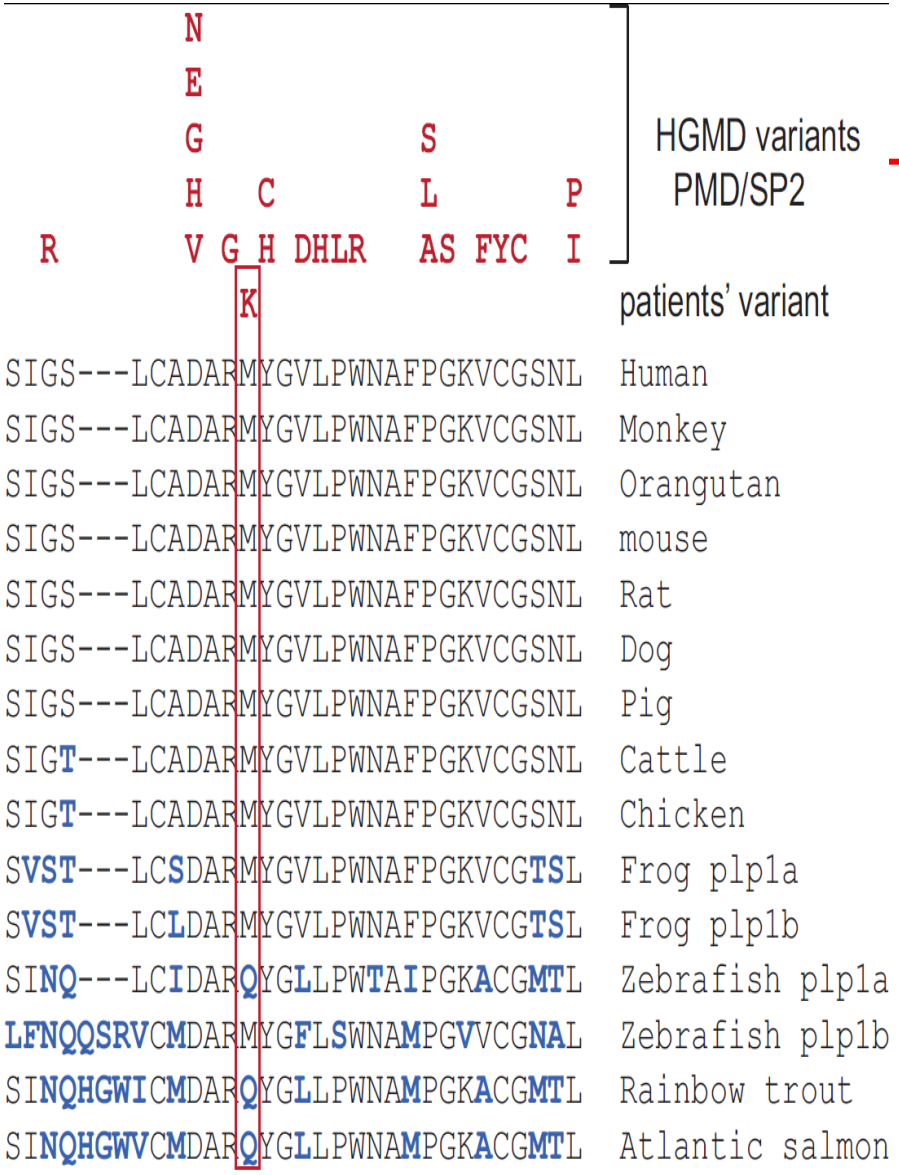
- 1 Recessive
- 2 X-Linked

# Exomes for “Diagnostic Odyssey”




**X-Chromosome PLP1 (Proteolipid Protein 1) Gene Mutation**

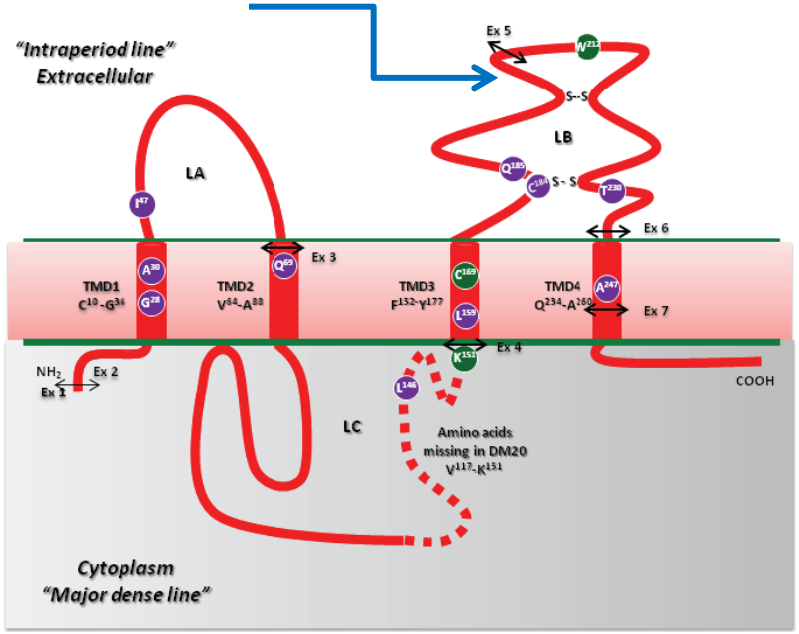
**c.617T>A, p.M206K – Novel Mutation***




**SIFT Score 0.01**

**PMD = Pelizaeus-Merzbacher Disorder**
**Dysmyelination/Leukodystrophy**
**PLP1 Mutations**

**p.M206K**




**PLP1 = Major Myelin Protein**

# Exome Sequencing – Summary

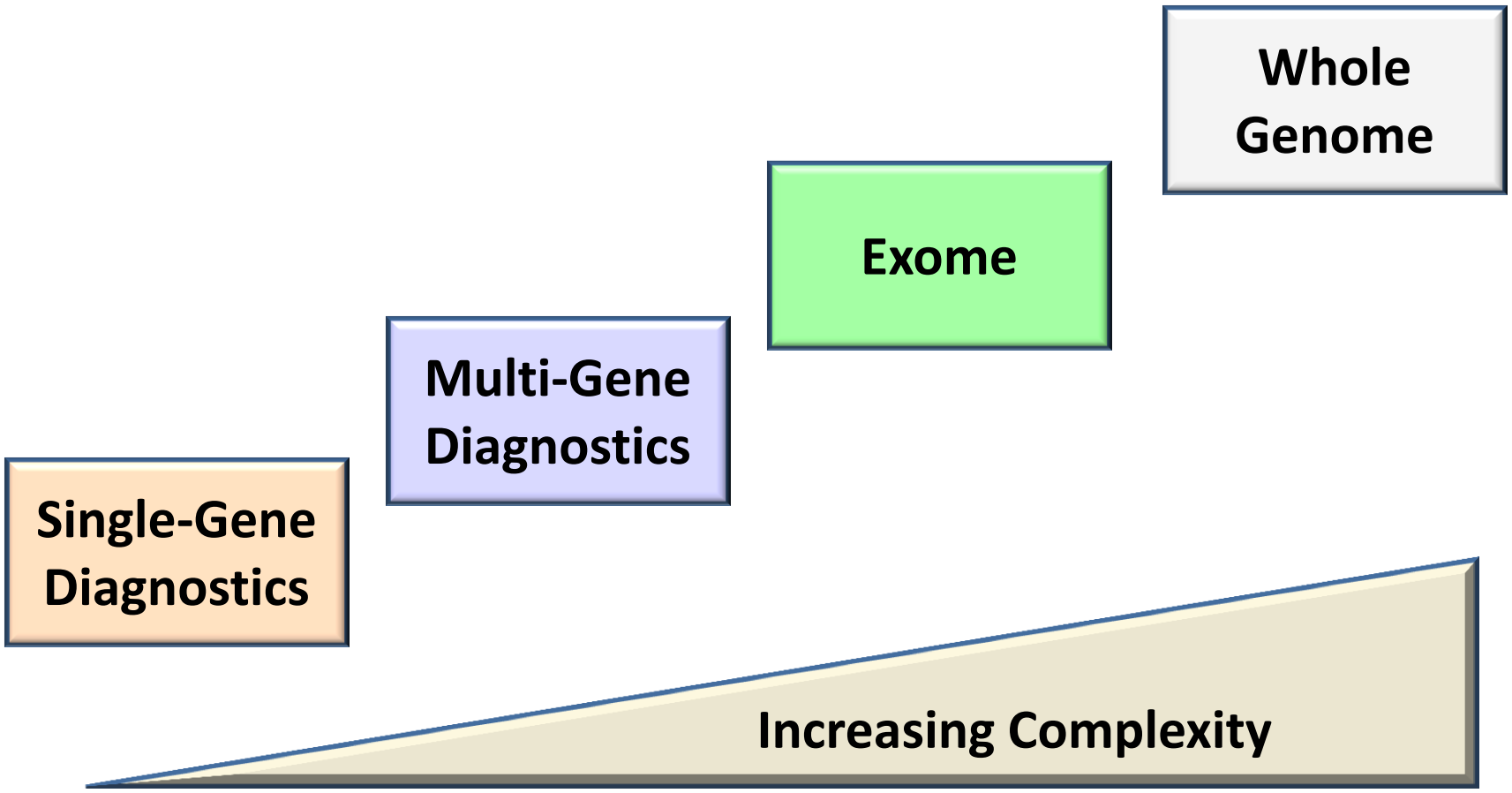
## Powerful New Approach to Inherited Disorders

- Now Available as a Diagnostic in Several Reference Laboratories

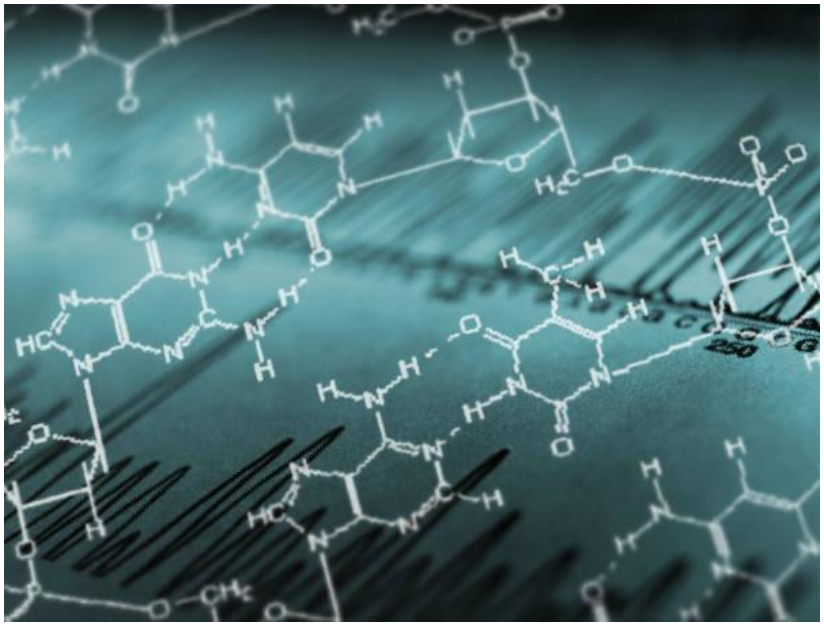
## Implementation Challenges for Laboratories

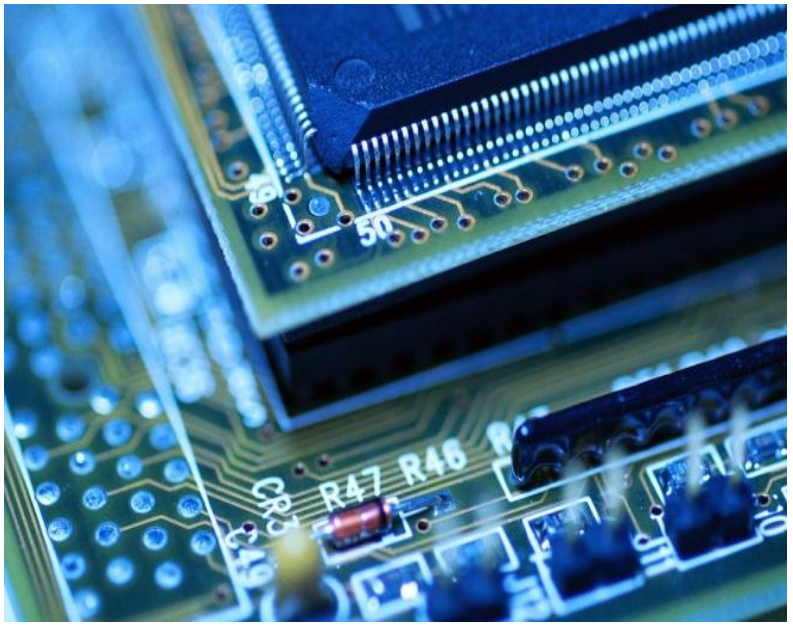
- Technically Demanding and Capital Equipment Intensive
- Complex and Evolving Data Analysis Requirements
- Diagnostic Yield Needs Management of Expectations

# New Landscape of Genetic Testing



# Thank You






**voelkek@aruplab.com**