Introduction of Next Generation Sequencing into Clinical Diagnostics

Rong Mao, MD
Associate Professor of Pathology
University of Utah
Medical Director of Molecular Genetics and Genomics
ARUP Laboratories

August 11th, 2014

Objectives

- Describe methodology of next generation sequencing and compare different platforms
- Discuss processes, validations and use of gene panels, exome sequencing and whole genome sequencing
- Understand and appreciate the complexity of data created by next generation sequencing

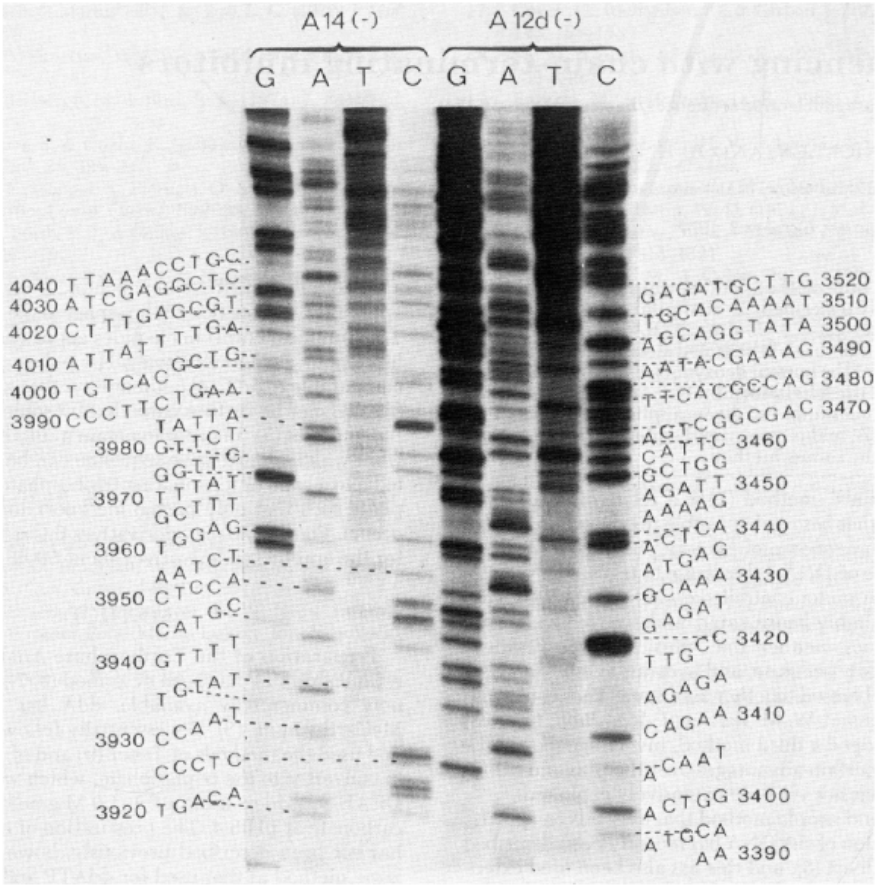
Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕX174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

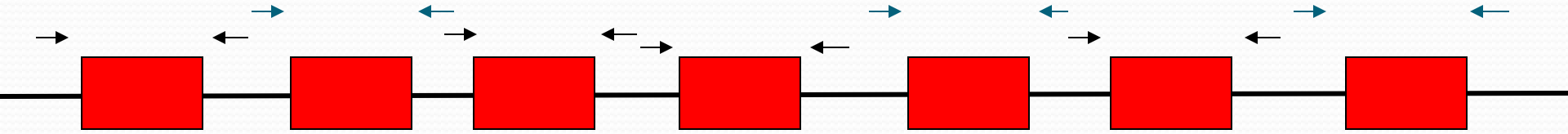
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England



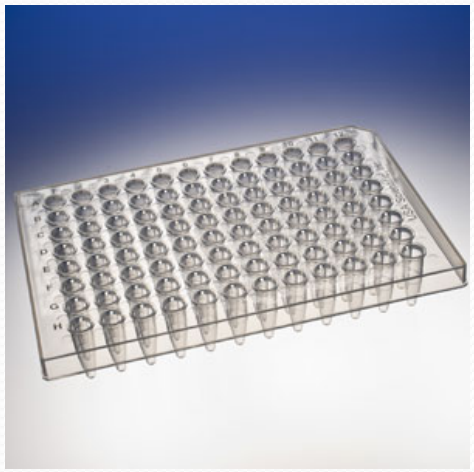
1977

Courtesy of Dr. K Voelkerding

1986 – Fluorescent Sanger Sequencing and Capillary Electrophoresis



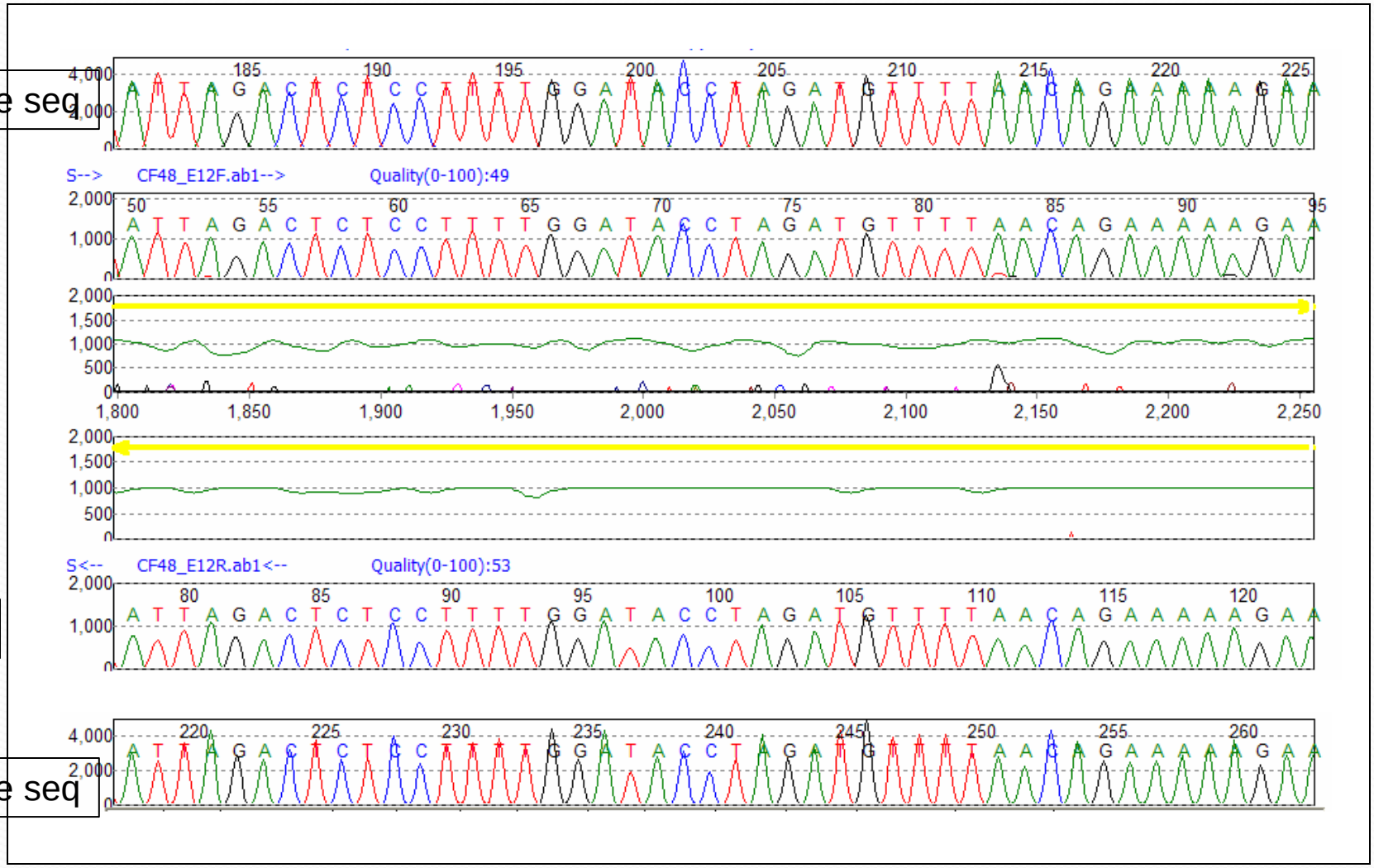
10 11 13 14 15 16



60,000-80,000bp sequences

ABI3730

Sanger Sequencing Alignment Using Mutation Surveyor Software



Sanger Sequencing

- Mature chemistry: ~ 600 - 800 Base Length Reads
- 1,000s Equal Length Termination Products per Peak
- Bi-Directional Sequencing Increases Accuracy
- Established Base Calling Algorithms

Accuracy Approaches 100% = Gold Standard

Courtesy of Dr. K Voelkerding

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]

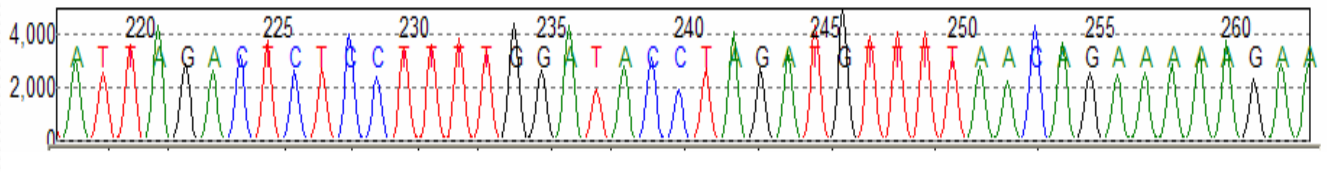
"Massively Parallel Sequencing"

Nature 437 (7057) 376-380

Paradigm Shift

Sanger Sequencing

Electrophoretic Separation of Chain Termination Products



Next Generation Sequencing

Sequence Clonally Amplified DNA Templates in a Flow Cell

Massively Parallel Configuration

Courtesy of Dr. K Voelkerding

Next Generation Sequencing Workflow

Genomic DNA

↓

Fragmentation

(Covaris)

(Sonication, Nebulization, 2-4kb fragments)

↓

Enrichment

(LR-PCR, RainDance ePCR, Array capture)

↓

Library Preparation

↓

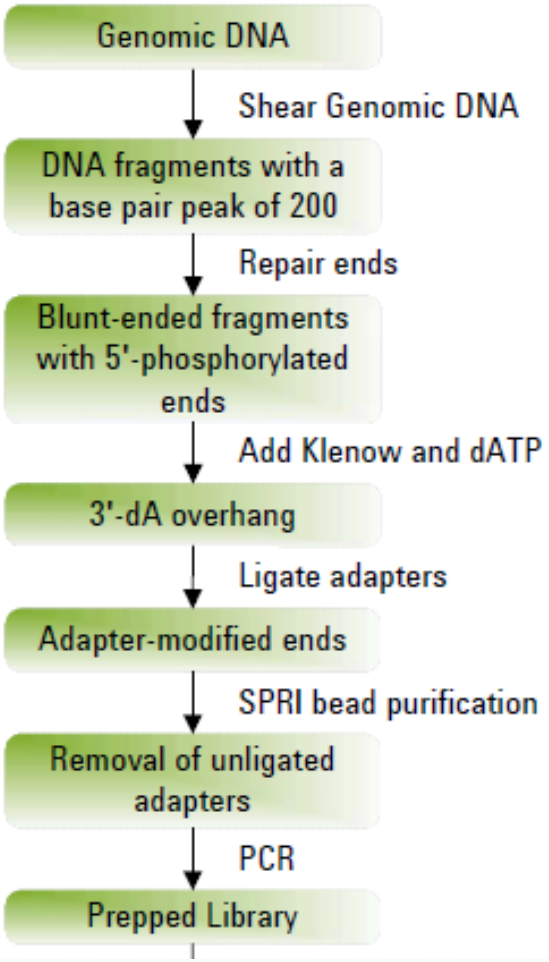
Next Gen Sequencing

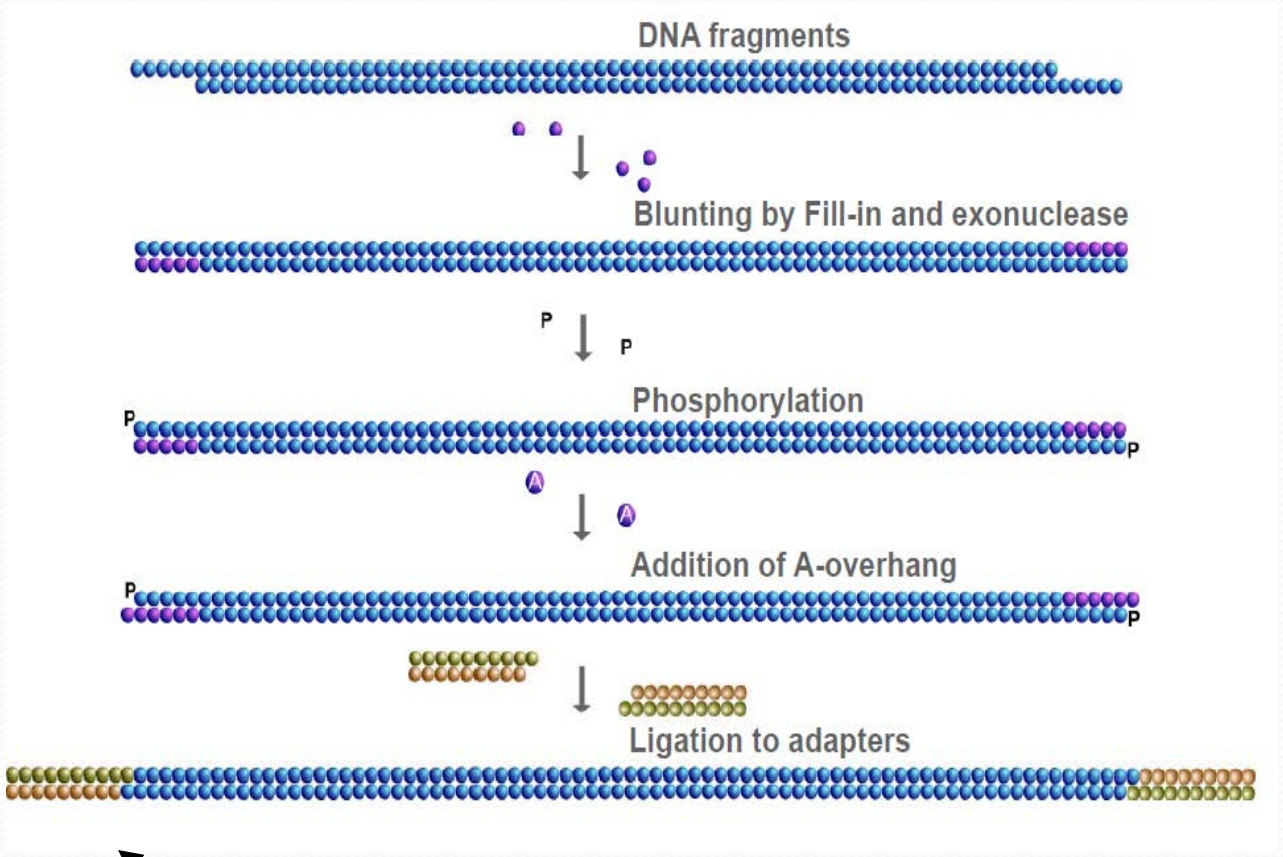
(Illumina/Hiseq, Roche/454, ABI/Solid)

↓

Bioinformatics and data analysis

Next Generation sequencing work Flow: Illumina Library Prep





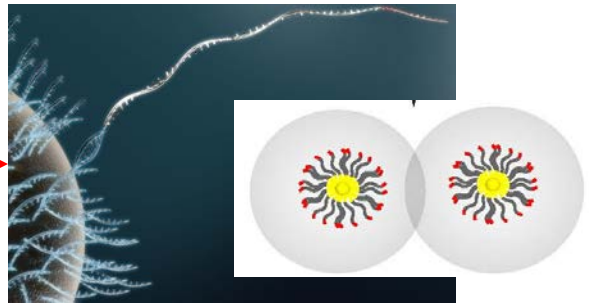
Adapters attach to flow cell during cluster formation/sequencing.

Courtesy of Dr. W Donahue

Roche/454 FLX Titanium: Workflow
Long Read Length>400bp

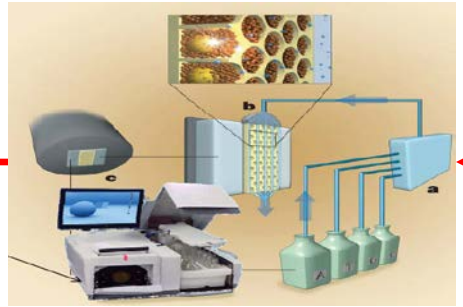
Sample Fragmentation

Adapter Ligation

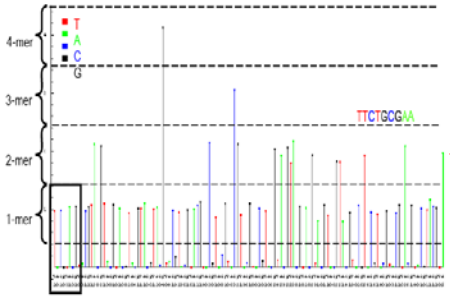


emPCR (1 fragment = 1 bead)

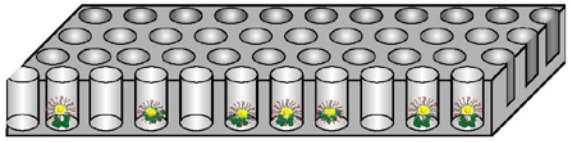
Picotiter plate (44 uM/1.6M wells)



Sequencing

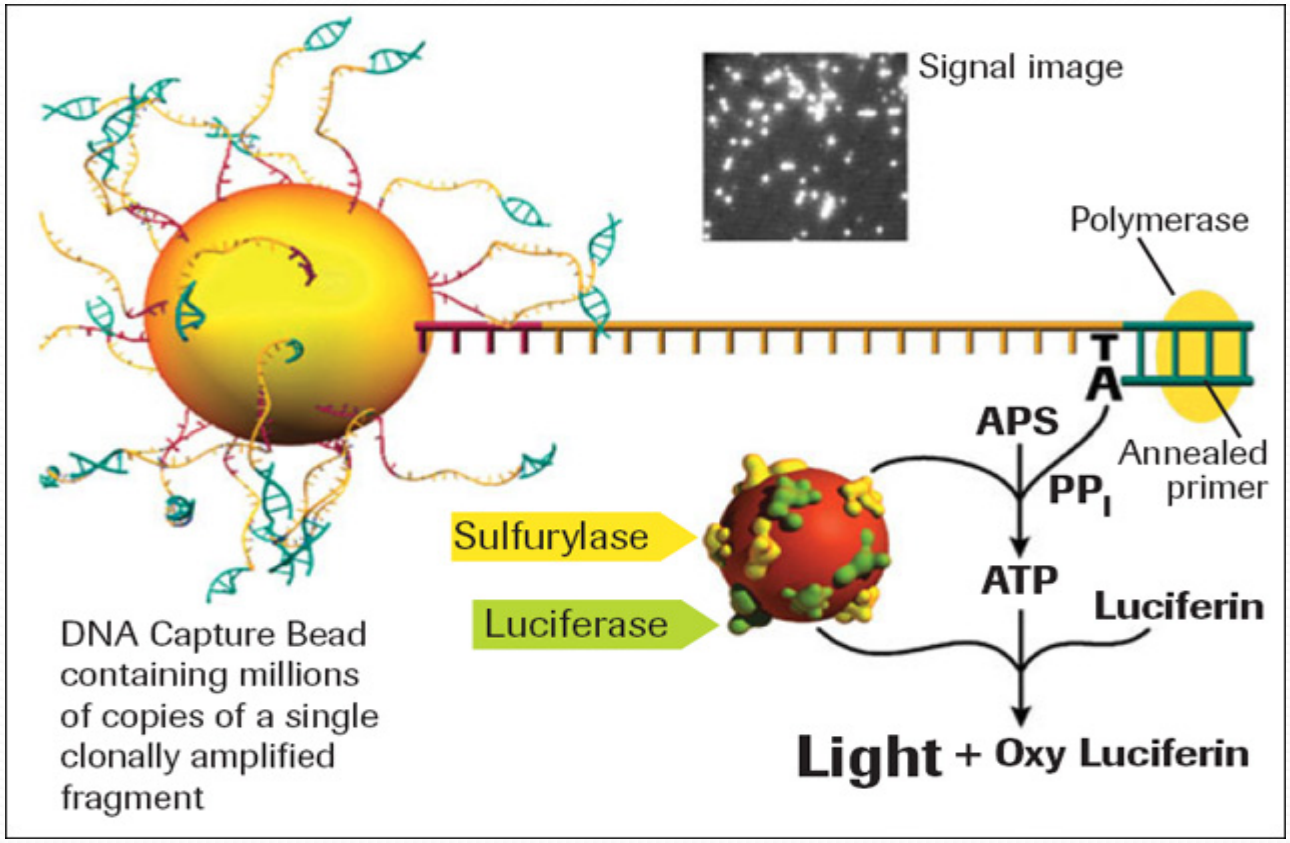


Analysis



Beads with clonally amplified template DNAs and sequencing enzymes

Roche/454 FLX Titanium-Pyrosequencing

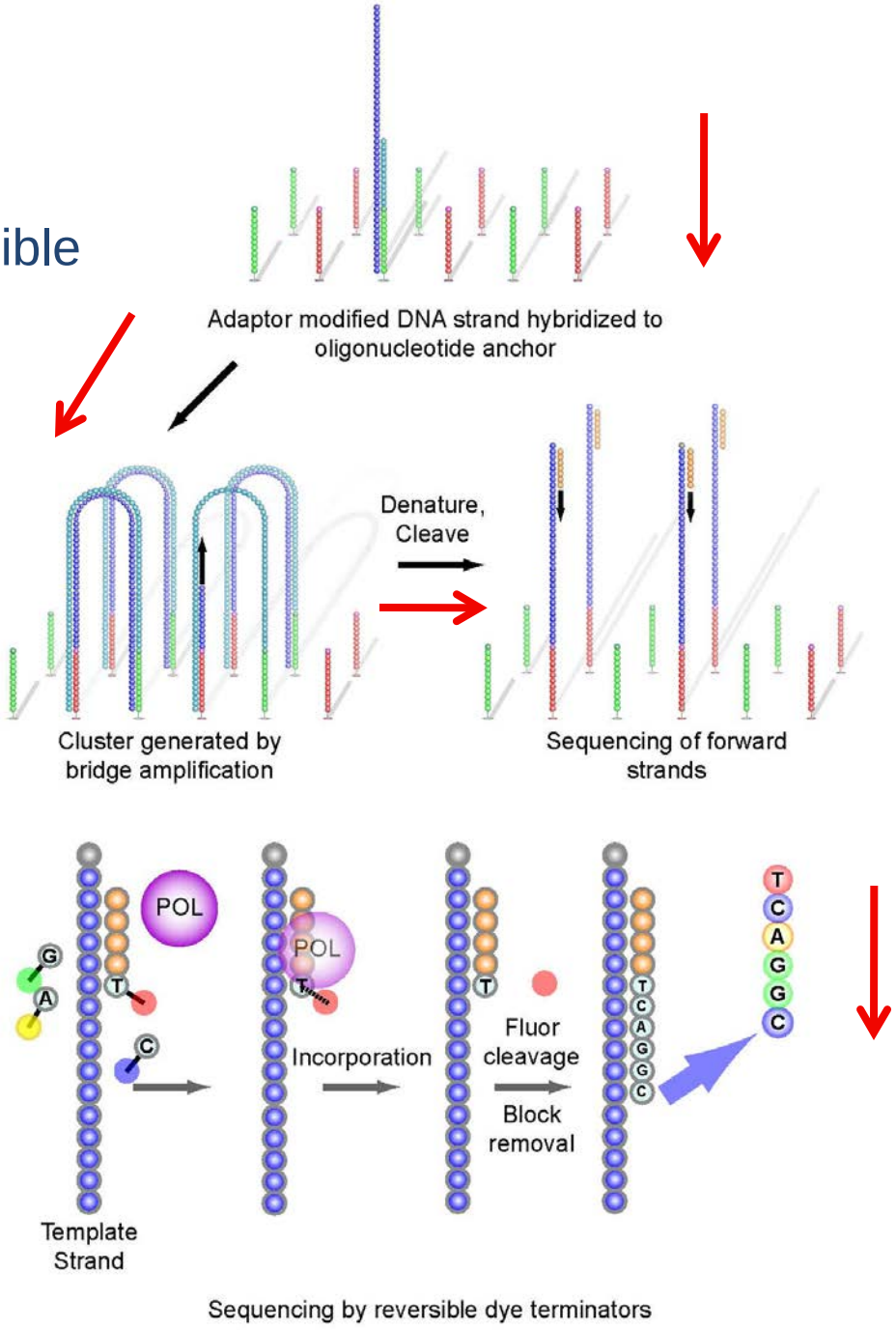


Roche/454 Life Sciences – Long Reads 400bp+

- Genome Sequencer FLX Titanium
  - > 1 million/400 bp reads=400 Mb
  - Run Time: 10 hrs
  - Reagents: $6000/run
- Roche Junior

Illumina

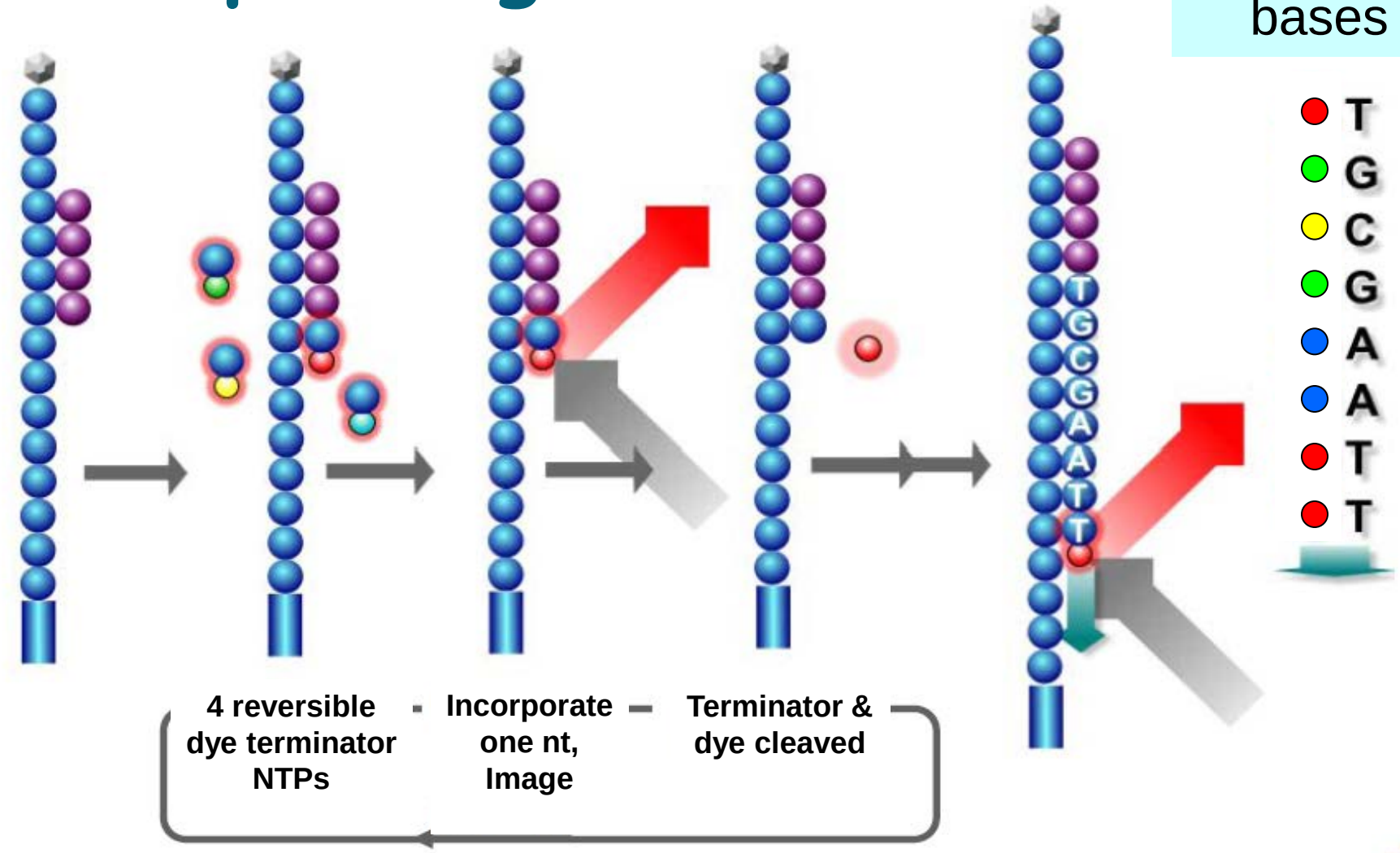
Sequencing by reversible dye terminators



Sequencing by reversible dye terminators

Sequencing

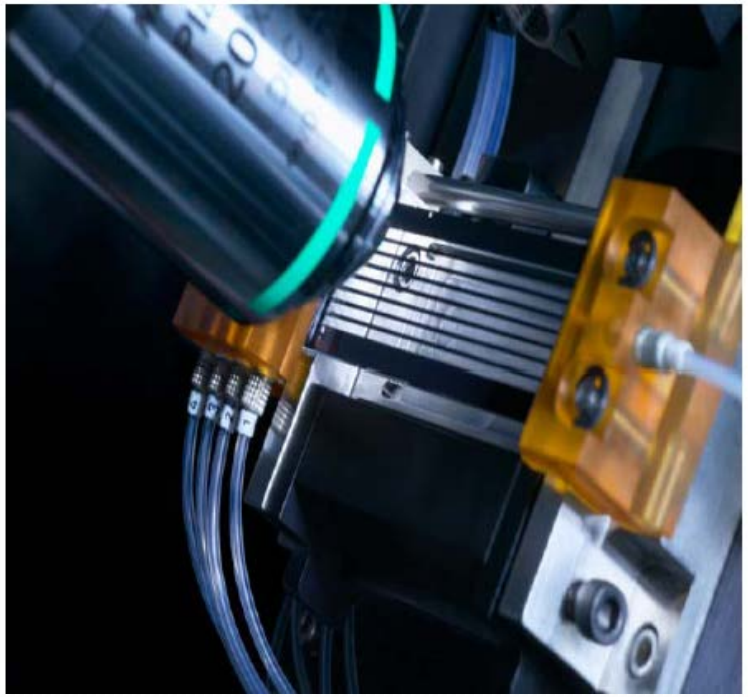
Read 100 bases

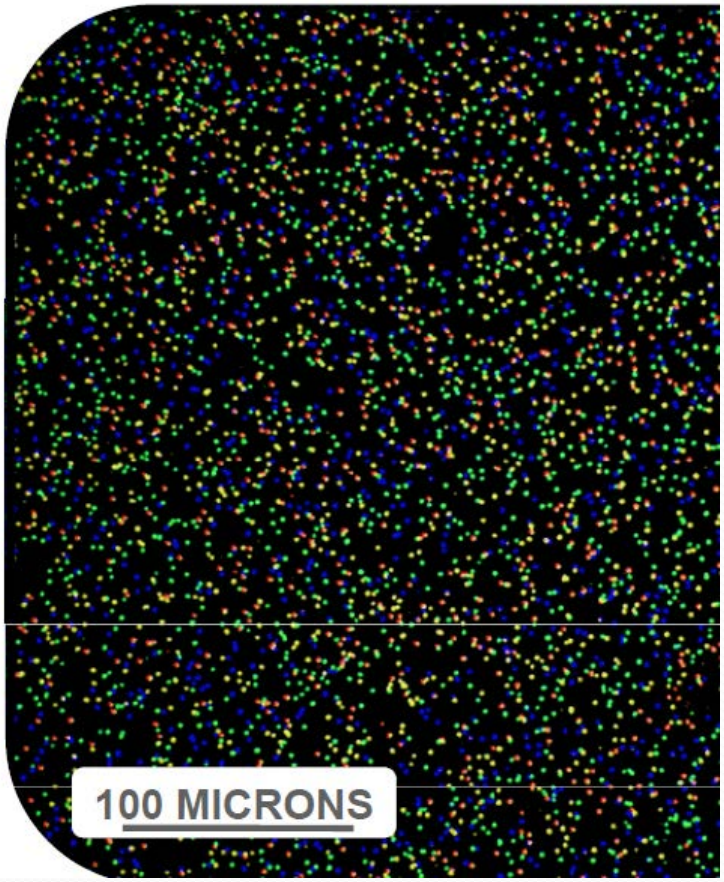


Three step cycle

illumina

Image scanning and Sequencing

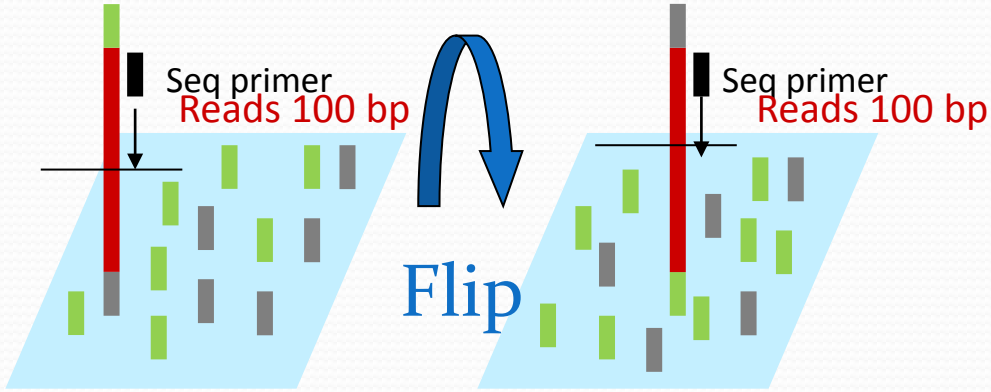




T G C A

Image of clusters during sequencing.

Paired-End Reading (2X100 bp)



Paired end reads

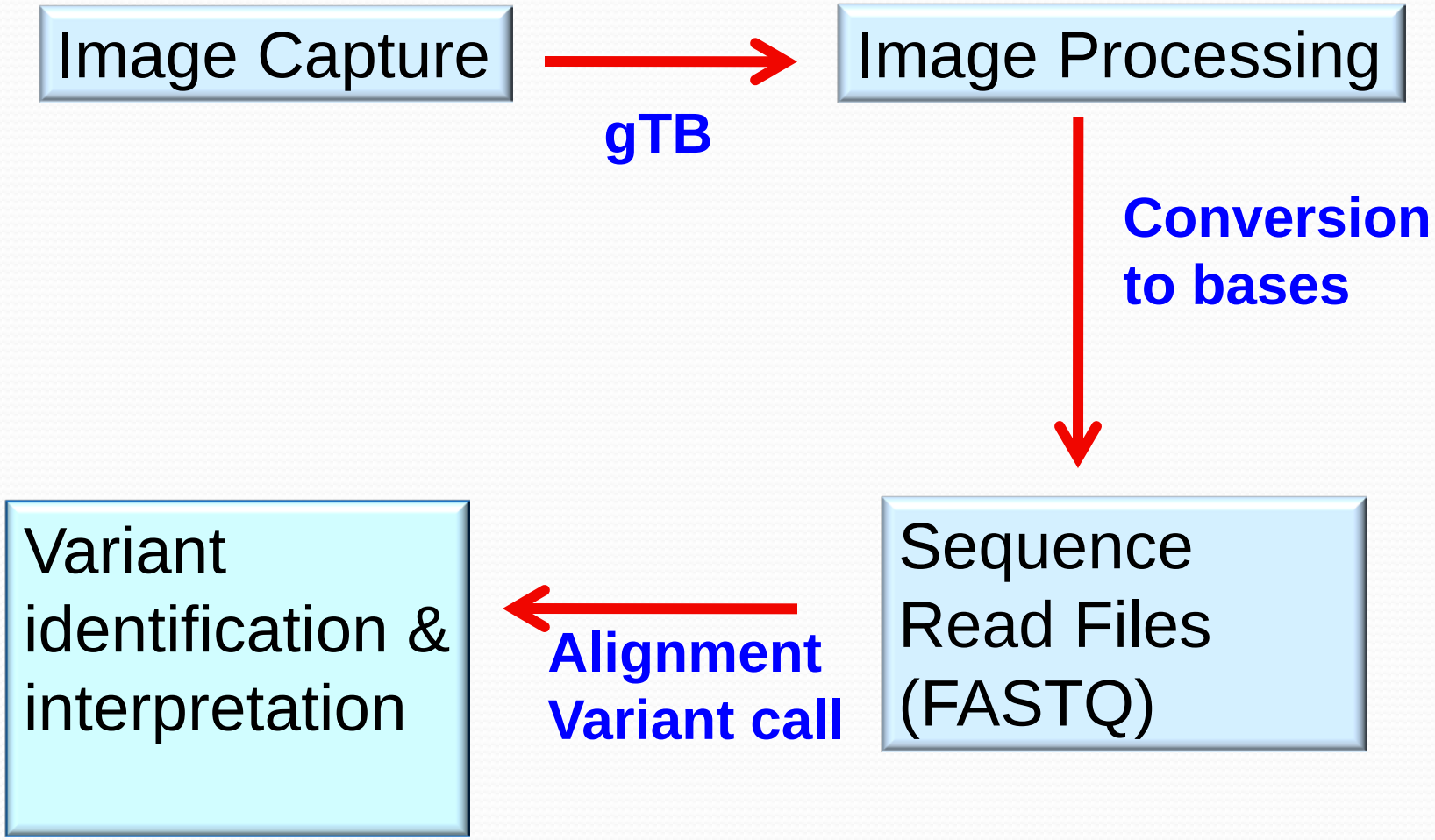
- Increase read coverage per cluster
- More accurate reading and alignment
- Detect small and large insertions, deletions, inversions, and other rearrangements

Illumina-Short reads 35bp-150bp

- HiSeq 2500
  - 1000Gb reads
  - Run Time: 14 days/125bp
  - Reagents: $14,000
  - 8 separate flow cells /run

- MiSeq: 15Gb
- NextSeq500: 120Gb

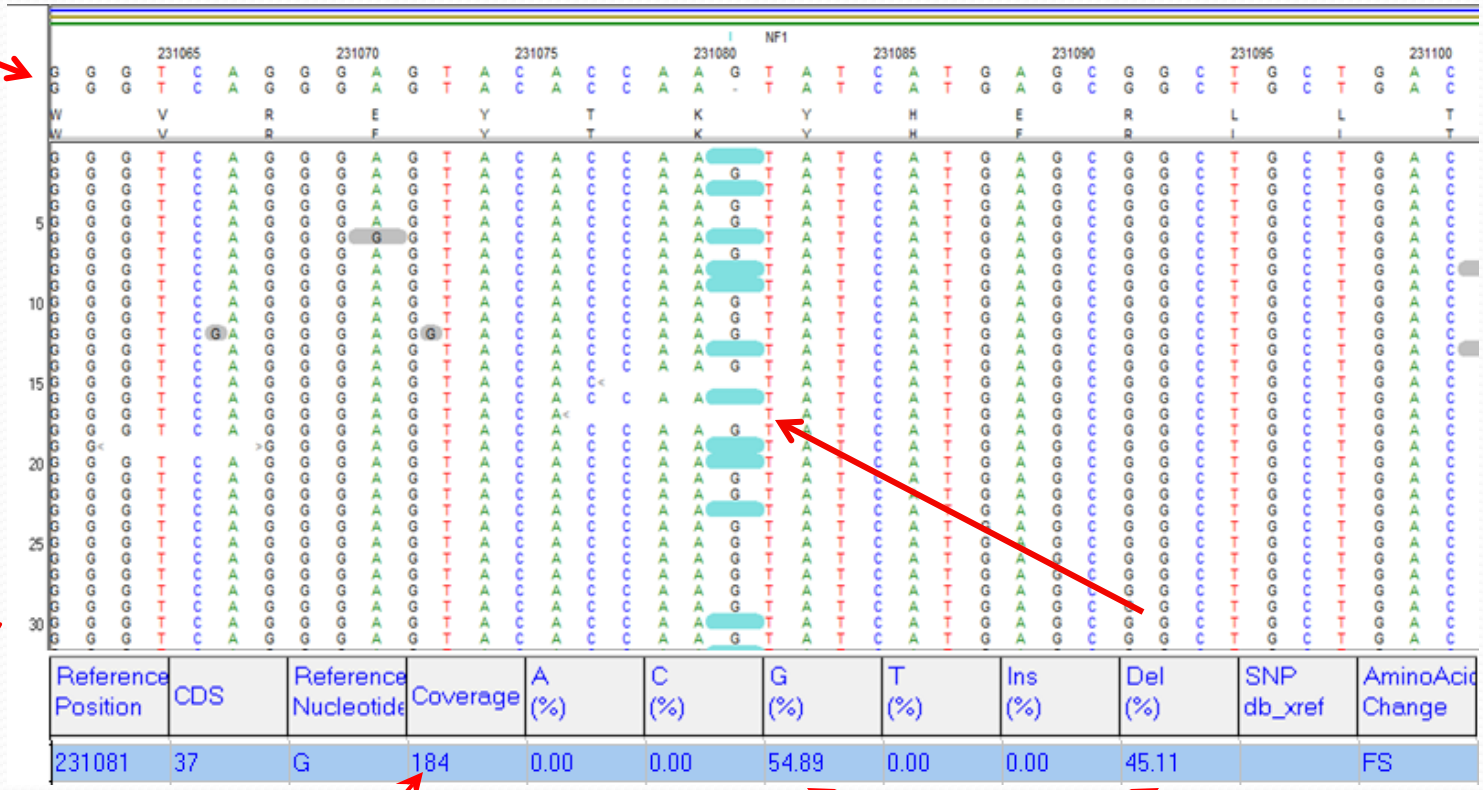
Bioinformatics and Data Analysis



Important Parameters for NGS Data Analysis

Reference

Coverage



Reference Position	CDS	Reference Nucleotide	Coverage	A (%)	C (%)	G (%)	T (%)	Ins (%)	Del (%)	SNP db_xref	AminoAcid Change
231081	37	G	184	0.00	0.00	54.89	0.00	0.00	45.11		FS

Sequencing coverage 15X

Smith et al. Genome Research 18(10):1638 (2008)

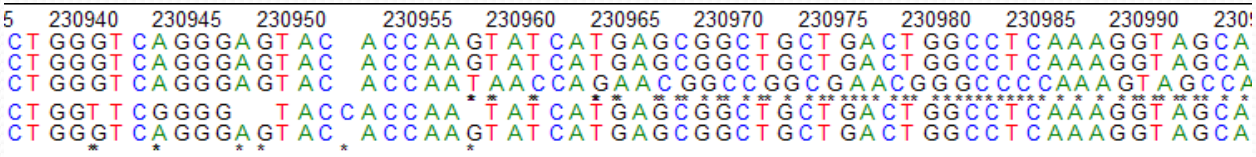
Base change percentage:

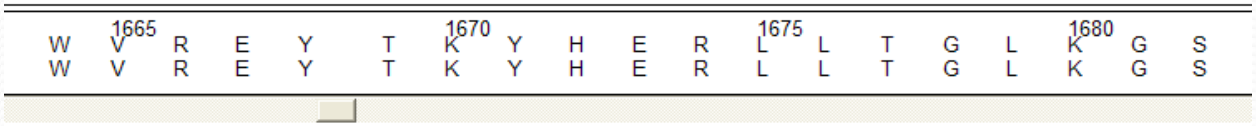
Percentage to make a heterozygous call 30% /70%

Sanger Sequencing Confirmation

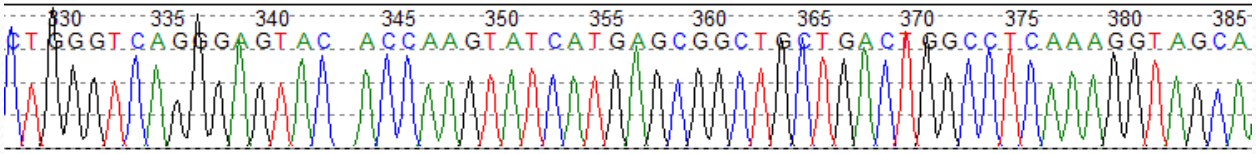
Exon37

Genome Sequence

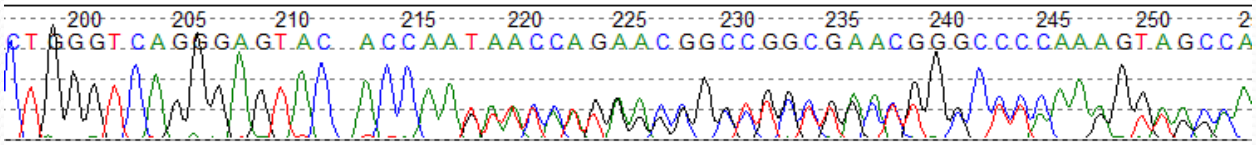




Reference Trace File

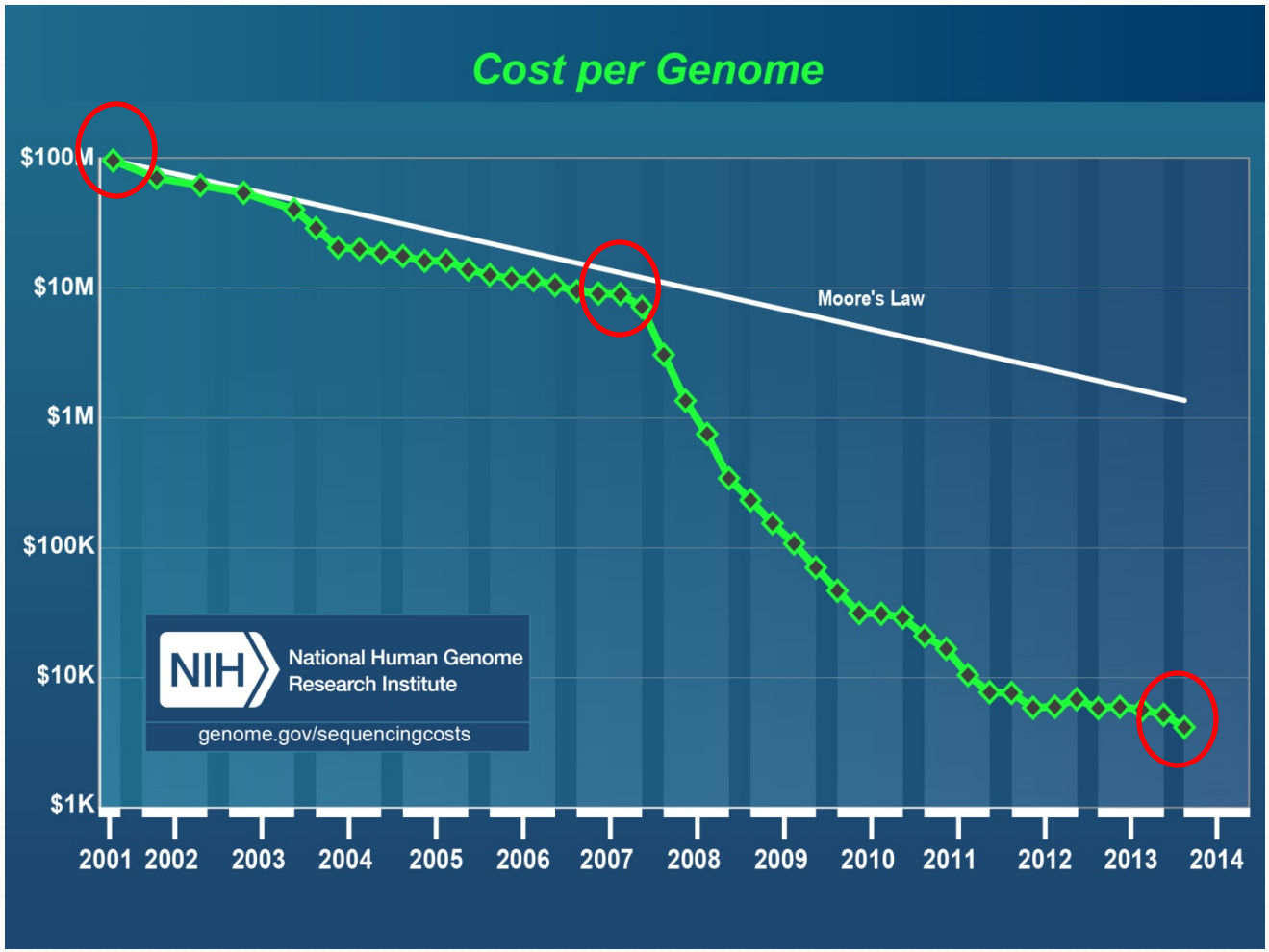


Sample 1 Trace File



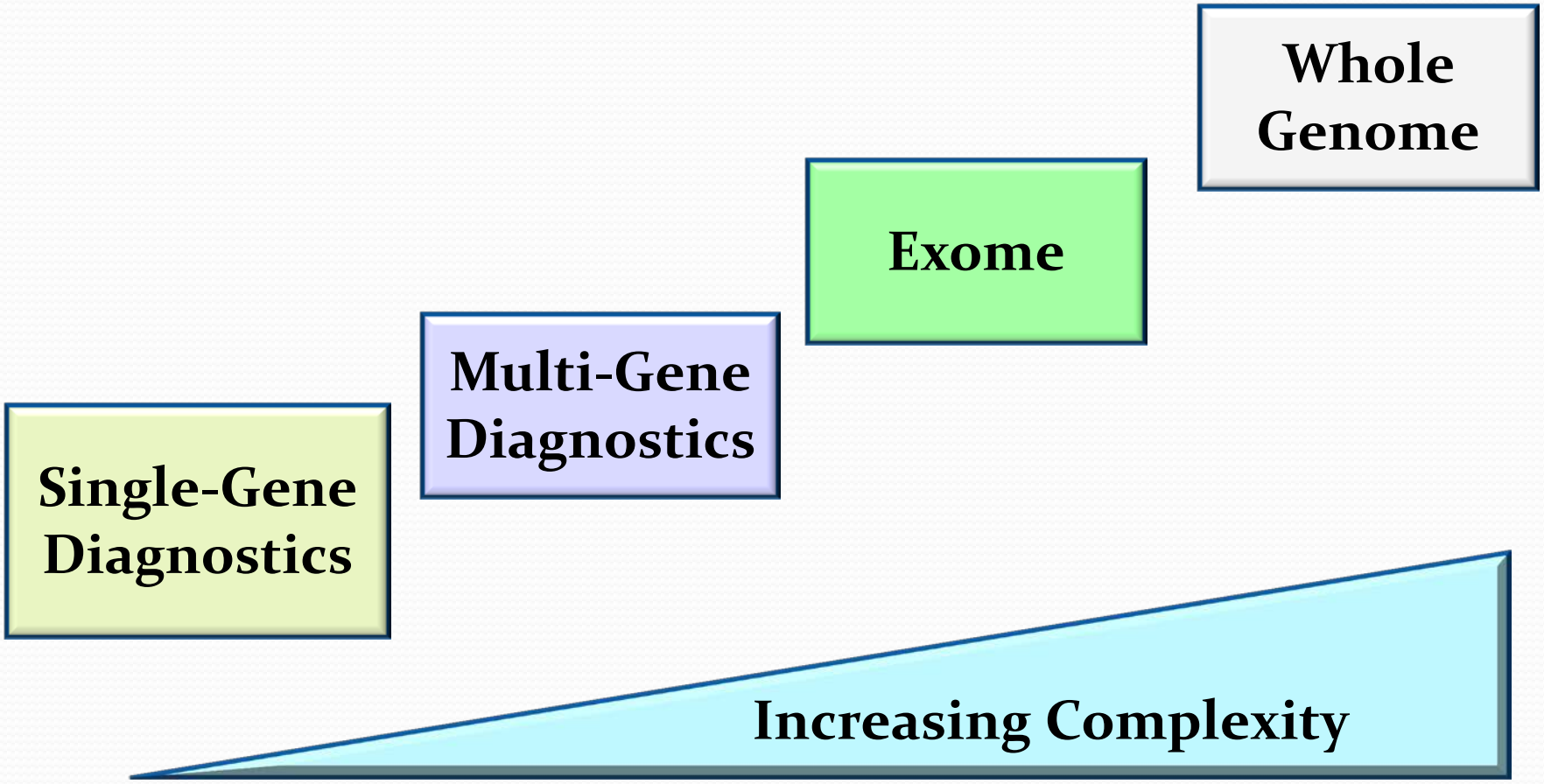
230959het_del G

Next Generation Sequencing Cost Dropping



Jan -2014 Cost per genome=$4,008

http://www.genome.gov/sequencingcosts/

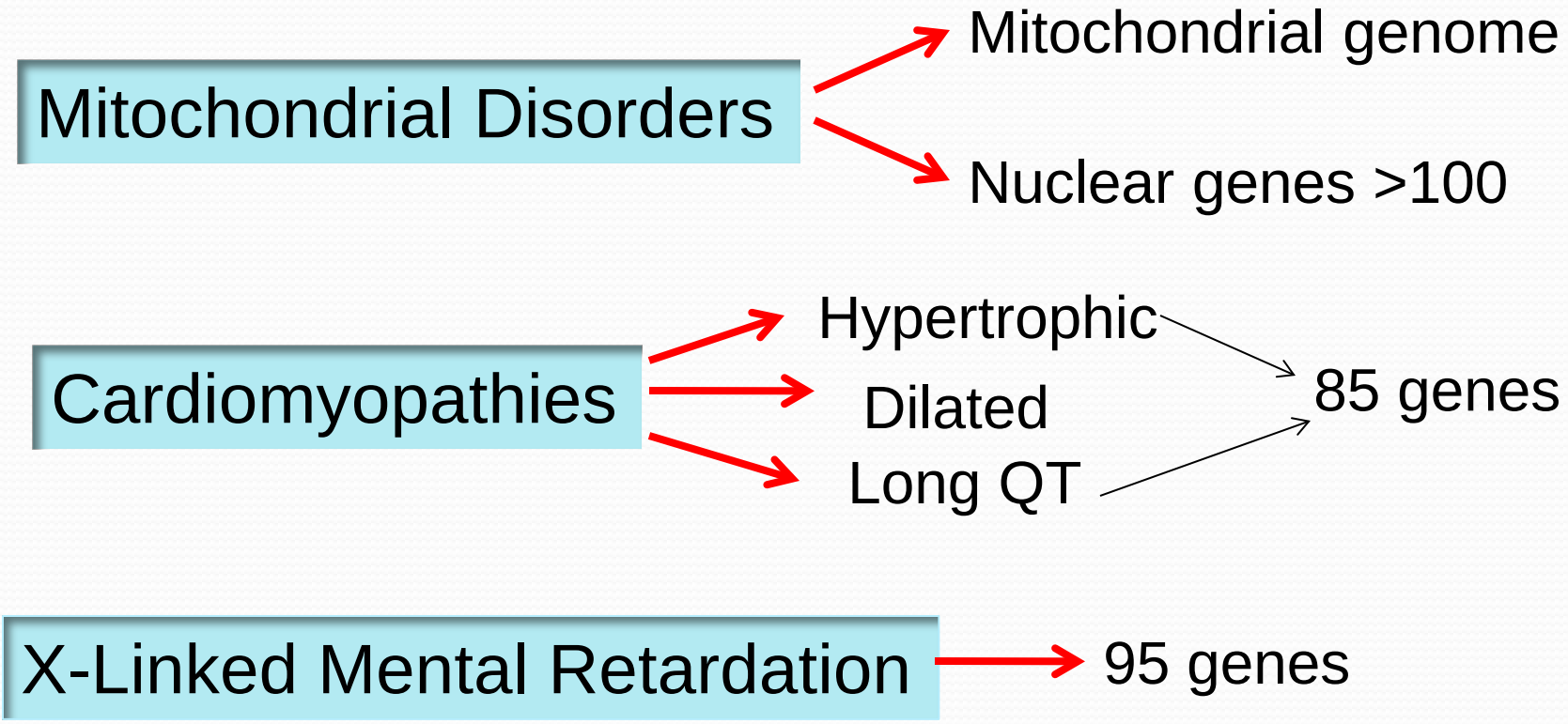

Whole Genome
Exome
Multi-Gene Diagnostics
Single-Gene Diagnostics
Increasing Complexity

Multi-Gen Panel Diagnostics

- Large gene with no mutation hotspot, e.g. *NF1* 58 coding exons; *DMD* 79 exons
- Multiple genes responsible for the phenotype (genetics heterogeneity)
- Phenotypic overlapping

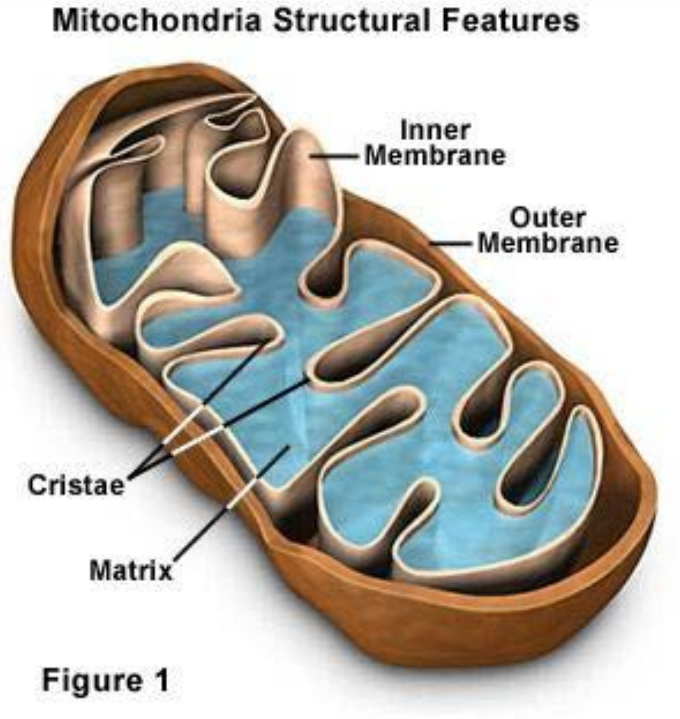
Difficult for Sanger sequencing

Multi-Gen Panel Diagnostics



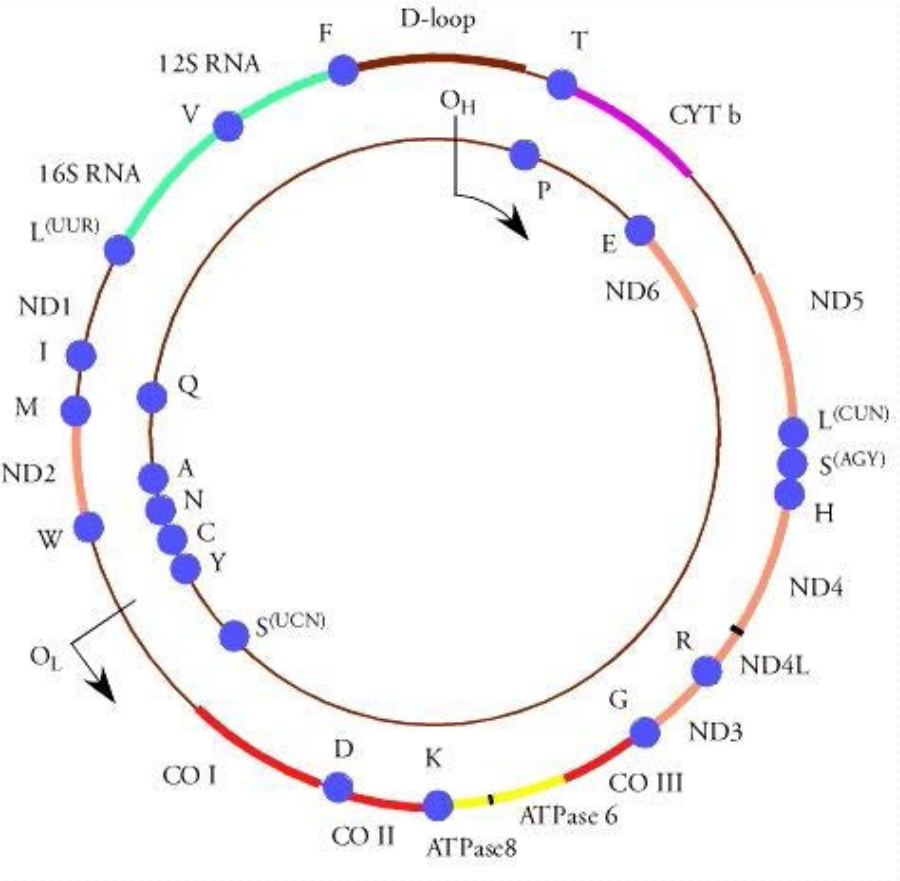
Feasible for next generation sequencing

Mitochondrial Disorders-Model for Multi-Gene Panel



- >1500 genes
  - Nuclear DNA
  - mtDNA
- ATP generation
  - ATP production via oxidative phosphorylation
- Energy resource:
  - – supplies 90% of energy for the body

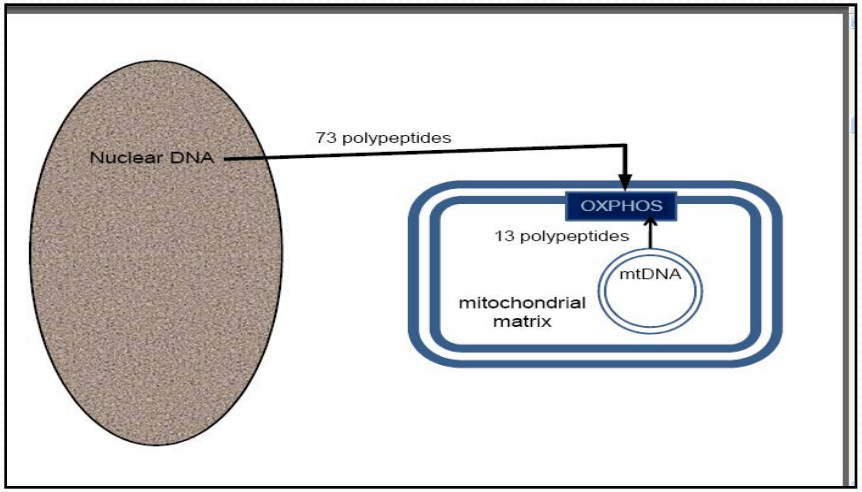
Mitochondrial genome



- Double stranded, circular
- No intron, 80 - 93% coding gene
- Lack histone and DNA repair mechanism damage, mutations (free radicals)
- 37 gene: 22 tRNA, 2 rRNA &13 protein
- Heteroplasmy
- Maternal inheritance

Mitochondrial Nuclear Genes

- >1500 genes encode proteins in the mitochondria



- Mendelian Inheritance
  - Dominant
  - Recessive
  - X-linked

108 Disease Causing Mito Nuclear Genes

Gene	NM no.	Chr
ABCB7	NM_004299	chrX
ACAD9	NM_014049	chr3
ACADL	NM_001608	chr2
ACADM	NM_000016	chr1
ACADS	NM_000017	chr12
ACADVL	NM_000018	chr17
ACAT1	NM_000019	chr11
APTX	NM_175073	chr9
ASS1	NM_000050	chr9
ATPAF2	NM_145691	chr17
ATXN7	NM_000333	chr3
BCKDHA	NM_000709	chr19
BCKDHB	NM_183050	chr6
BCS1L	NM_004328	chr2
c10orf2	NM_021830	chr10
CABC1	NM_020247	chr1
COQ9	NM_020312	chr16
COX10	NM_001303	chr17
COX15	NM_078470	chr10
COX6B1	NM_001863	chr19
CPT1A	NM_001876	chr11
CPT2	NM_000098	chr1
DARS2	NM_018122	chr1
DBT	NM_001918	chr1
DGUOK	NM_080916	chr2
DLAT	NM_001931	chr11
DLD	NM_000108	chr7

Gene	NM no.	Chrr
DNAJC19	NM_145261	3
DNM1L	NM_012062	12
ETFA	NM_000126	15
ETFB	NM_001014763	19
ETFDH	NM_004453	4
ETHE1	NM_014297	19
FH	NM_000143	1
FXN	NM_000144	9
GFM1	NM_024996	3
HADH	NM_001184705	4
HADHA	NM_000182	2
HADHB	NM_000183	2
HMGCL	NM_000191	1
HMGCS2	NM_005518	1
HSPD1	NM_002156	2
LRPPRC	NM_133259	2
MCCC2	NM_022132	5
MFN2	NM_014874	1
MPV17	NM_002437	2
MRPS16	NM_016065	10
MRPS22	NM_020191	3
NDUFA1	NM_004541	X
NDUFA11	NM_175614	19
NDUFAF1	NM_016013	15
NDUFAF2	NM_174889	5
NDUFAF3	NM_199069	3
NDUFAF4	NM_014165	6

Gene	NM no.	Chr
NDUFS1	NM_005006	2
NDUFS2	NM_004550	1
NDUFS3	NM_004551	11
NDUFS4	NM_002495	5
NDUFS5	NM_004552	1
NDUFS6	NM_004553	5
NDUFS7	NM_024407	19
NDUFS8	NM_002496	11
NDUFV1	NM_007103	11
NDUFV2	NM_021074	18
OPA1	NM_130837	3
OXCT1	NM_000436	5
PC	NM_001040716	11
PCK2	NM_004563	14
PDHB	NM_000925	3
PDHX	NM_003477	11
PDP1	NM_001161778	8
PDSS1	NM_014317	10
PDSS2	NM_020381	6
PINK1	NM_032409	1
POLG	NM_002693	15
POLG2	NM_007215	17
PPM1B	NM_002706	2
PREPL	NM_006036	2
PUS1	NM_025215	12
RRM2B	NM_015713	8
SCO1	NM_004589	17

Gene	NM no.	Chr
SCO2	NM_005138	22
SDHA	NM_004168	5
SDHB	NM_003000	1
SDHC	NM_003001	1
SDHD	NM_003002	11
SLC22A5	NM_003060	5
SLC25A13	NM_001160210	7
SLC25A15	NM_014252	13
SLC25A19	NM_001126121	17
SLC25A20	NM_000387	3
SLC25A22	NM_024698	11
SLC25A3	NM_002635	12
SLC25A4	NM_001151	4
SLC3A1	NM_000341	2
SPG7	NM_003119	16
SUCLA2	NM_003850	13
SUCLG1	NM_003849	2
SURF1	NM_003172	9
TAZ	NM_000116	X
TIMM8A	NM_004085	X
TK2	NM_004614	16
TRMU	NM_018006	22
TSFM	NM_001172696	12
TUFM	NM_003321	16
UQCRB	NM_006294	8
UQCRQ	NM_014402	5
WFS1	NM_006005	4

Mitochondrial Disease Prevalence

- Incidence of 1:5000 live births (Smeitink 2006)
- 20% are due to mtDNA mutations (200 pathogenic mutations), 80% to nuclear DNA mutations

Mitochondrial Disorders-Model for Multi-Gene Panel

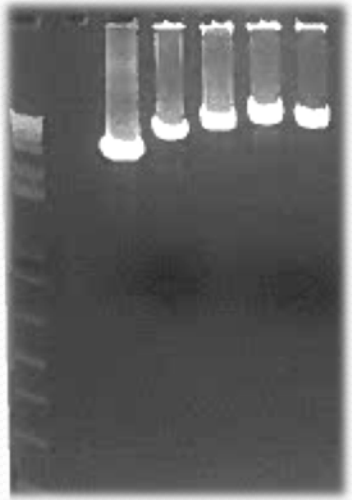
- Next Generation Sequencing (NGS)
  - Mitochondrial genome sequencing
  - 108 Mito Nuclear genes sequencing

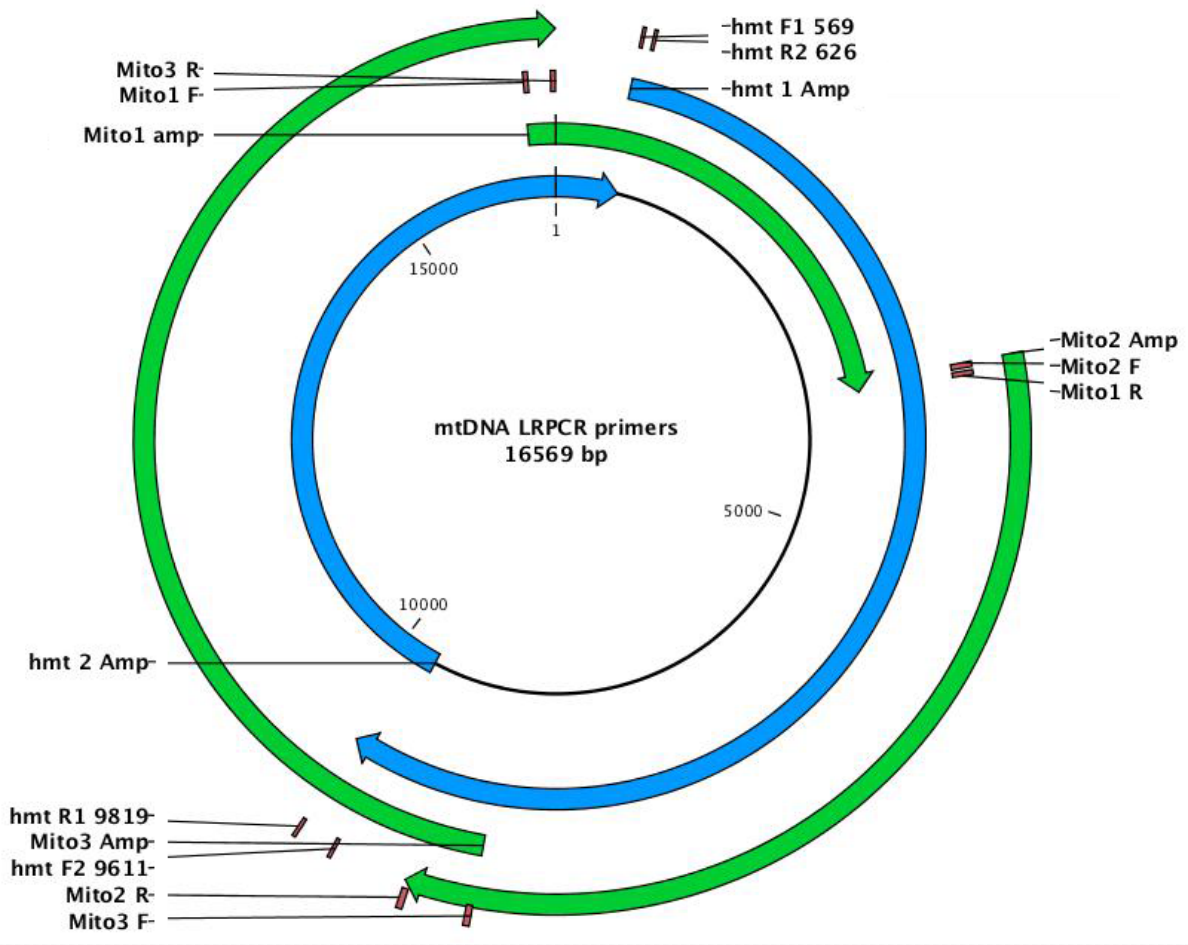
Point mutations and small ins/del Low heteroplasmy

- Large deletions and duplications in mitochondrial genome and >100 nuclear gens by high density exonic CGH Microarray

20% of del in mito DNA and 5-10% large del/dup in nuclear gens

Mitochondrial Genome Enrichment- Long Range PCR



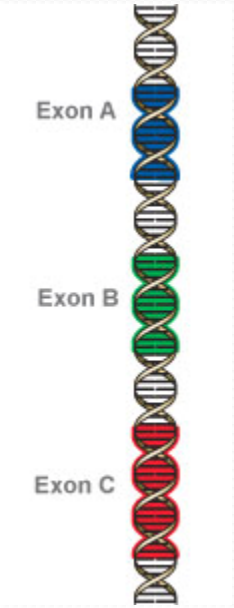


mt genome enriched by long range PCR

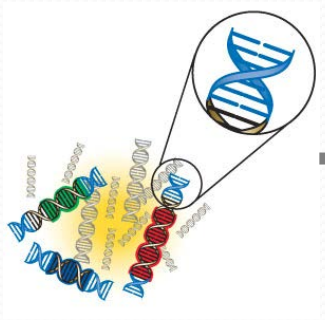
Courtesy of Shale Dames

Mitochondrial 108 Nuclear Genes: Roche/NimbleGen SeqCap for Targeted Enrichment

Genomic DNA Targeted Region

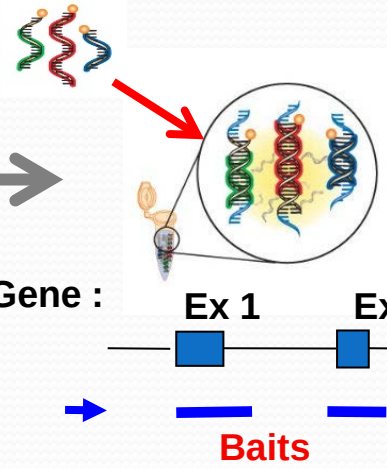


Prepare with Next-Gen Sequencing Adaptors



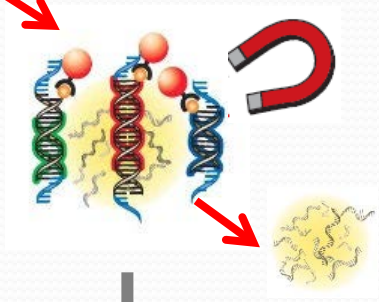
Hybridization

Biotinylated Library Baits



Capture and Washing

Streptavidin Coated Magnetics Bead



Unbound DNA

Amplify DNA and Enrichment QC

Sequencing

Mitochondrial Genome NGS

SPRI-TE

Illumina HiSeq 2000

Long Range PCR

Amplicons equimolar pooled

SPRI-TE and index

Illumina HiSeq

Sequence Alignment

Variant calls

Day 1

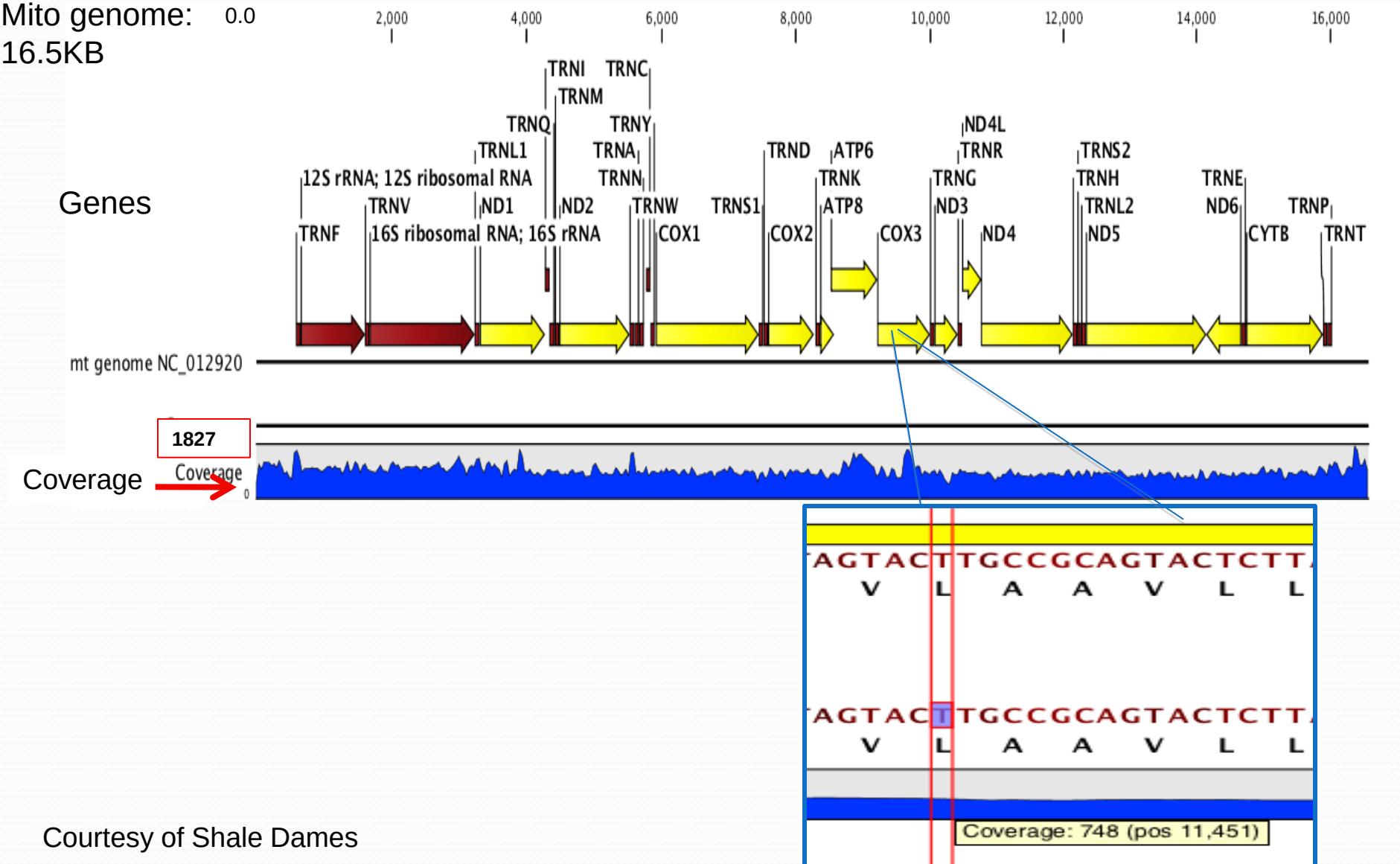
Days 2-3

Days 4-9

Days 10+

Mitochondrial Genome NGS

- CLCbio Genomics Workbench



Courtesy of Shale Dames

CLCBio Output

m.3460G>A

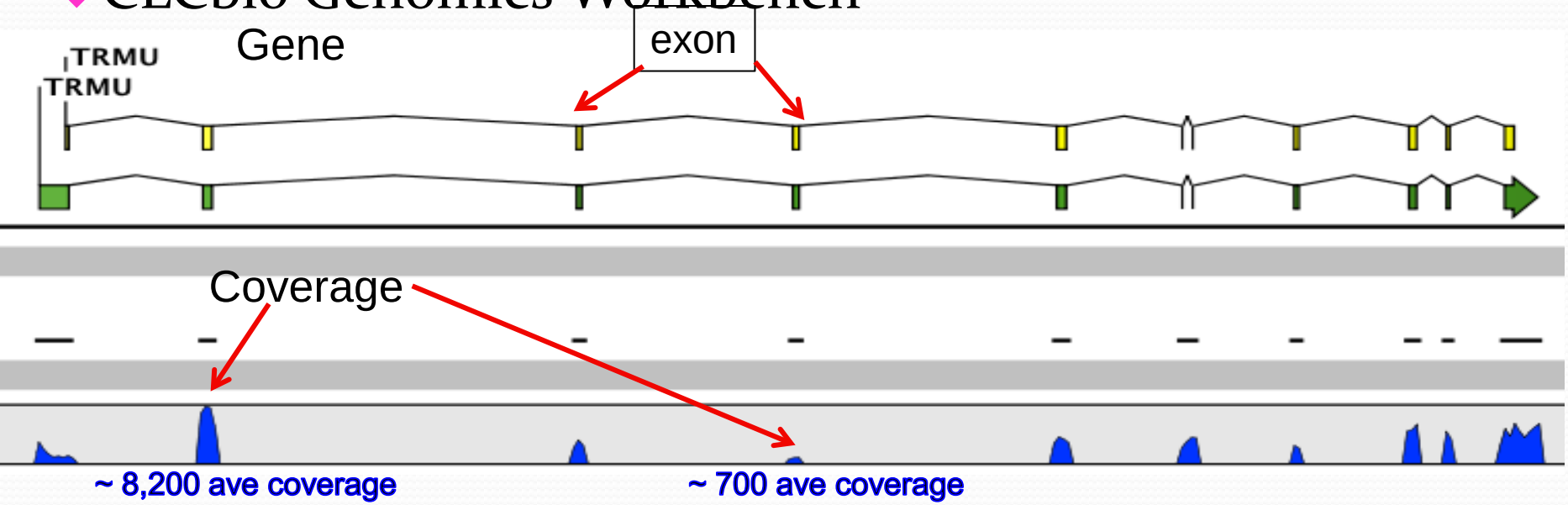
Sample ID: NA11605		Flowcell ID: 81C03ABXX	
Fastq file: NA11605_7_1		Cluster kit ID: 0745788 L/N 5836181	
Start date of run: 022111		Index sample, Single read	
Date of analysis: 03082011		Technician: S. Dames	

Reference Position	Amino Acid Change	Frequencies	Coverage	Clinical Significance
3460	Ala52Thr	99.7	6517	Significant: Peripapillary microangiopathy; Gene ND1

Courtesy of Shale Dames

mt 128 Nuclear Gene Panel

- CLCbio Genomics Workbench



- Alignment/variant call parameters:
  - Aligned to dbSNP132 annotated and masked reference sequence
  - Minimum coverage: 50-fold
  - Heterozygous allele frequency range: 30-70%
  - Report all CDS SNP/DIP variants
  - Filter out common polymorphisms

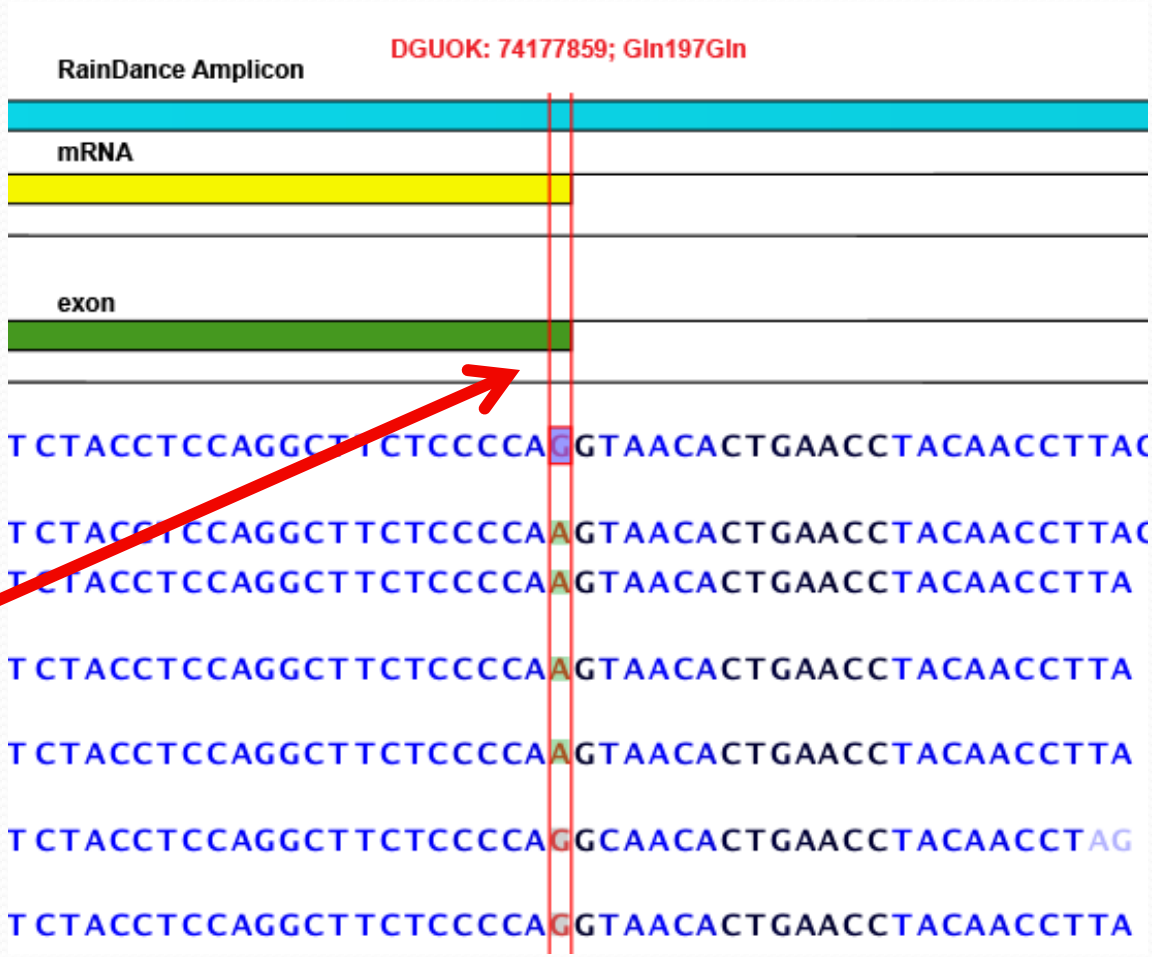
Example 1:

- **Clinical history:** Newborn with abnormal phenylalanine on NBS. Follow-up plasma AA showed elevated tyrosine and methionin.
- He had significant failure to thrive, feeding difficulties and fat malabsorption.
- Liver failure and transplant at 7 weeks of age.
- Subsequent development of hypotonia and psychomotor regression. Died at 23 months from a cardiac arrest.

DGUOK Mut1:

Reference Seq

c.591G>A
p.Gln197Gln
Splice site



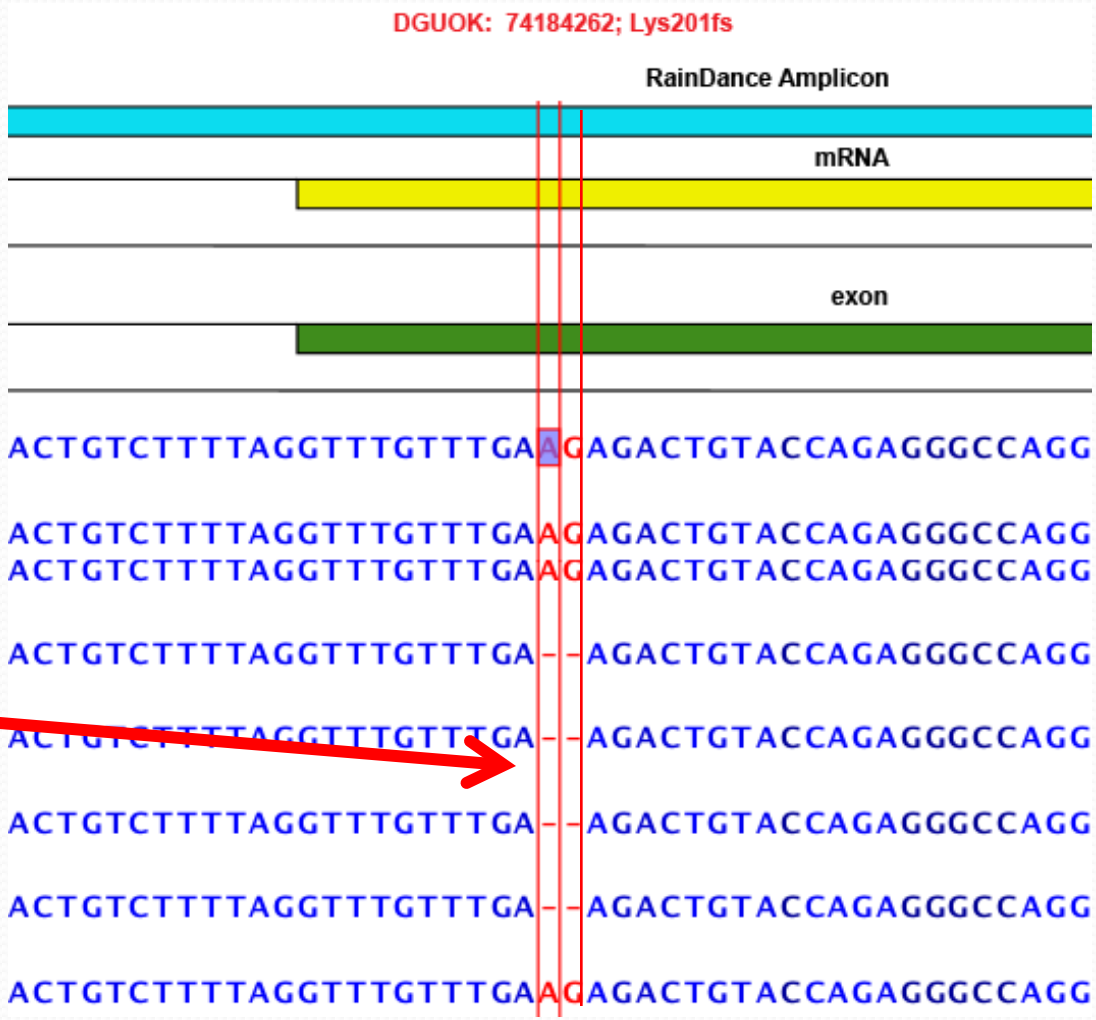
Mapping	Reference	Variation	Reference	Allele	Frequencies	Counts	Coverage	Amino	rs	Mutation
DGUOK	74177859	SNP	G	A/G	53.7/46.3	2695/2327	5023	Gln197Gln	not reported	**MDS compound het with AKA R202TfsX13.**

Courtesy of Shale Dames

DGUOK Mut 2

Reference Seq

c.605-606delAG
p.Lys201fs



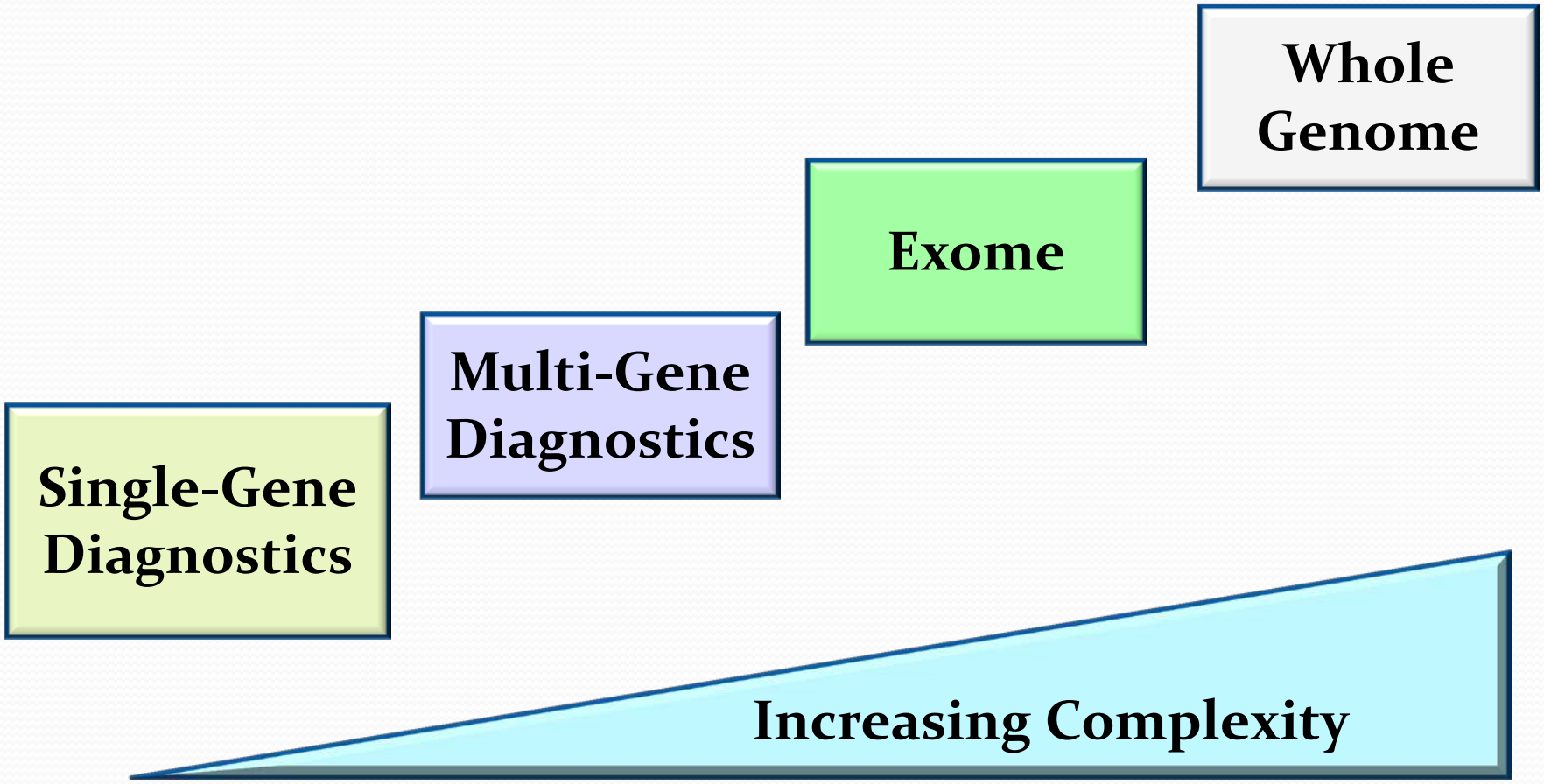
Mapping	Reference	Variation	Reference	Allele	Frequencies	Counts	Coverage	Amino	rs	Clinical
DGUOK	74184262	DIP	AG	AG/--	59.9/39.9	2977/1983	4967	Lys201fs	not reported	MDS. AKA R202TfsX13. Introduces stop codon at aa position Glu214Ter (alt trans VCLKTVPEGQGGGERN*)

Courtesy of Shale Dames

Mito Multiple Gene Panel-Clinical Utility

- Differential diagnosis (NICU), confirm genetics etiology
- Specific disease causing mutation identified
- Family risk consultation and testing for other family members

- Management and treatment


Whole Genome
Exome
Multi-Gene Diagnostics
Single-Gene Diagnostics
Increasing Complexity

ARUP: Exome Sequencing

- Exome Sequencing with Symptom-Guided Analysis Test Code: 2006332
  - Proband plus up to five family members (same price)
- Exome Sequencing with Symptom-Guided Analysis, proband Only
  - Test Code: 2006332
  - Prefer to have parental samples for Sanger sequencing controls
- Turnaround Time: 12-16 weeks
- Specimen Type: Blood, other sample types are acceptable

Human Exome

- Exome: the portions of a gene or genome that code information for protein synthesis

- Est, 21, 000 genes
- 180,000 exons
- 1.5% of whole genome

Online Mendelian Inheritance in Man (OMIM)

- Mendelian basis of inheritance
- Total: 22,340 Entries
- Gene Description: 14,569
- Known disease associated genes: 4,105 (19%)
- Autosomal: 1,739; X-linked: 1,210, Y-linked: 59; Mitochondrial: 65

http://omim.org/# Updated May 2 2014%

Why Exome Sequencing?

Focuses on the part of the genome we understand best, the exons of the genes

Exons comprise 1% of the genome

~85% of all known disease causing mutations are located on exons

Exome sequencing costs 1/6 of the cost of whole genome sequencing

What is Exome sequencing ?

The sequence of all exons of the genome

What is missing?

Some protein coding genes
Some exons of some genes
Non-genic control elements
Copy number changes
Structural changes
mtDNA
Some microRNA genes

Exome Sequencing

- A powerful tool for gene discovery
- Over 200 genes have been discovered in a couple of years
- Now a powerful diagnostic tool !

Diagnostic Yield

Based on the NIH Undiagnosed Diseases Program clinical sensitivity of exome sequencing is around 25%

Possibly selection of “best” cases

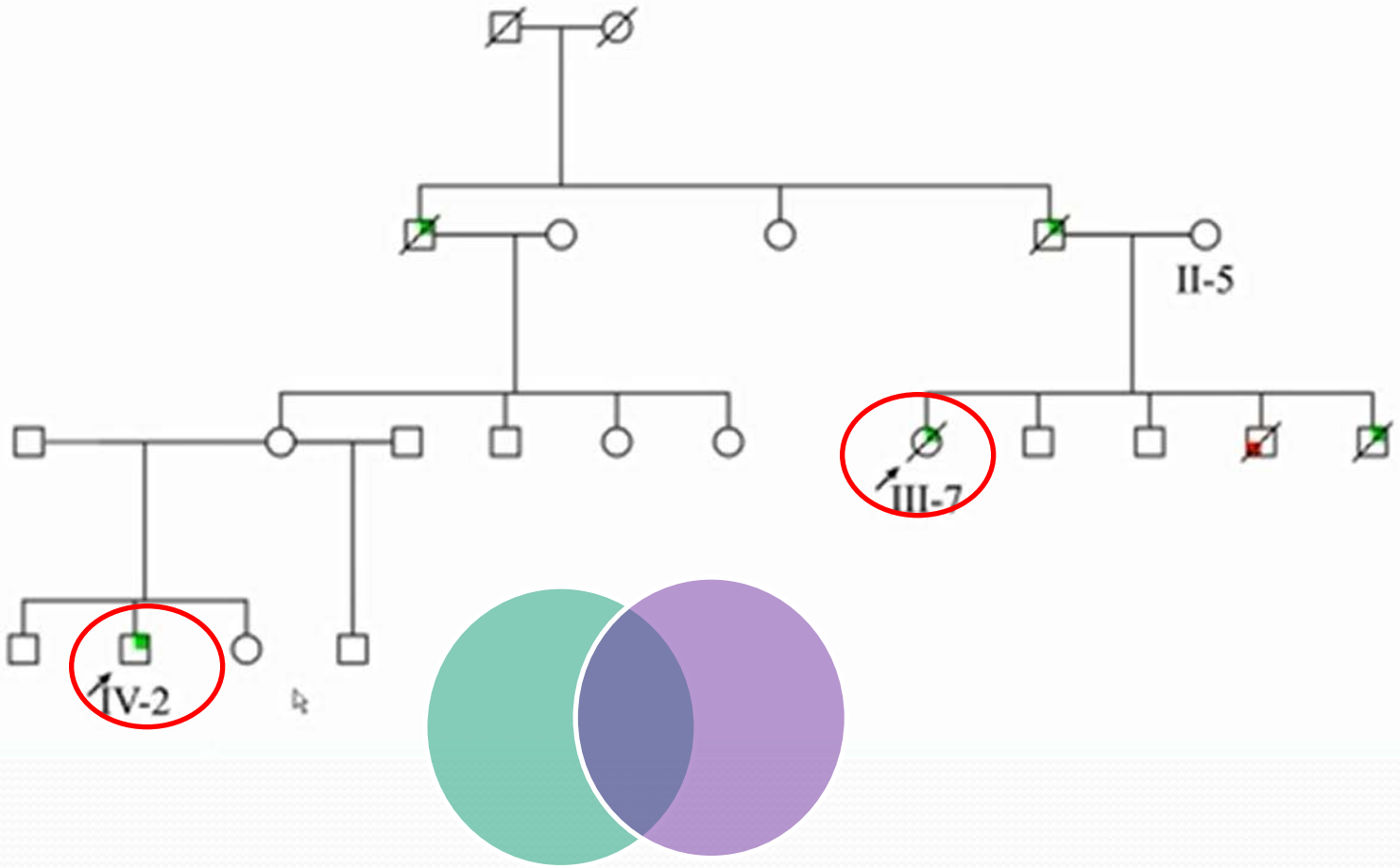
Consistent with ARUP clinical exome sensitivity

Gahl et al., Vol14 (1) Jan 2012 | **Genetics in medicine**

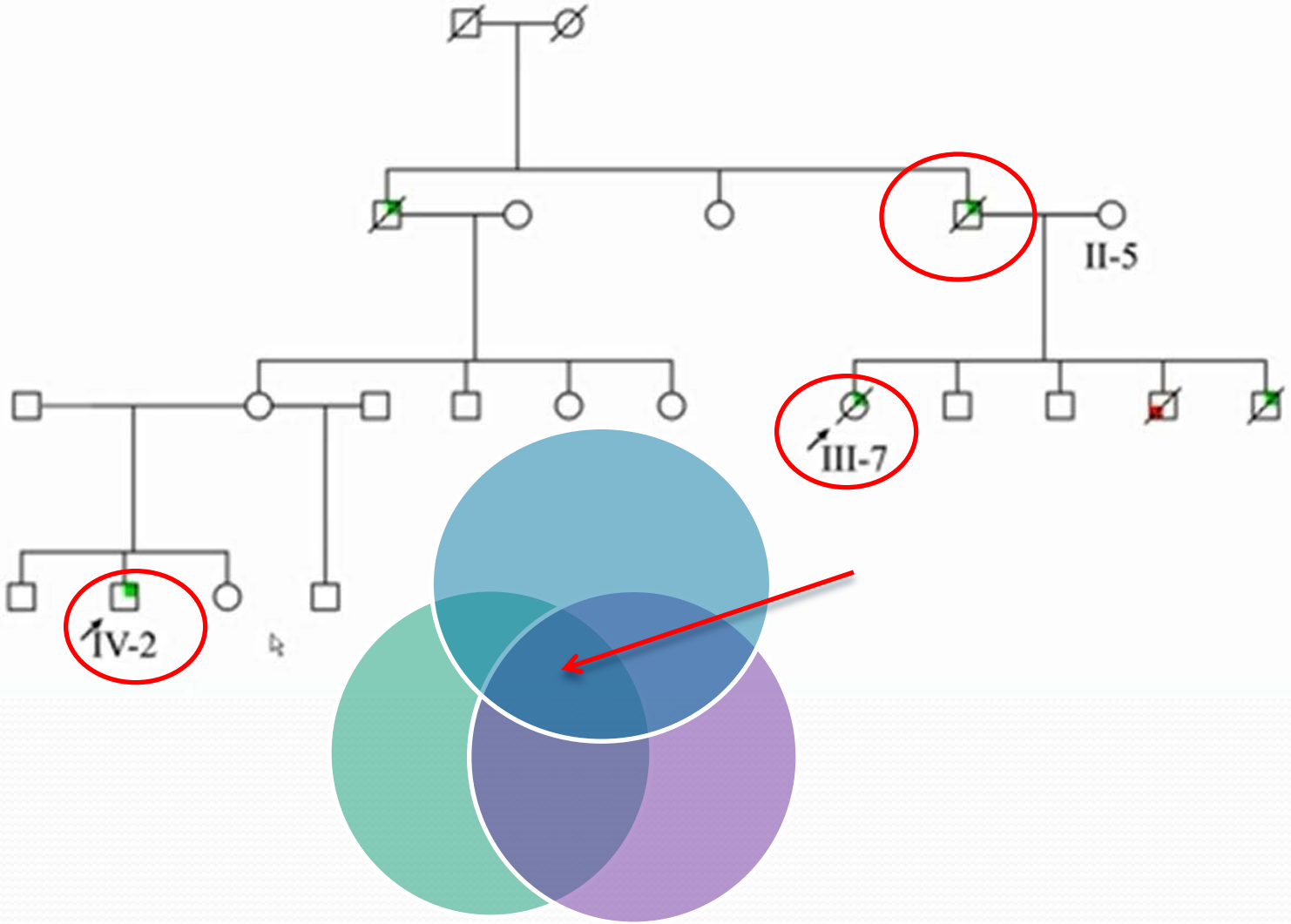
Diagnostic Odyssey

- Multiple congenital abnormalities
- Intellectual disability
- Unexplained developmental delay or declining

Sequencing Strategy



Sequencing Strategy



Clinical Exome Sequencing

- Indexing of samples (barcoding)
- Agilent and Nimblegen liquid capturing
- Illumina HiSeq 2500
- Alignment / Variant calling / Phenotype scoring
- Candidate mutation list
- Interpretation

CLINICAL EXOME SEQUENCING

Work flow :

Time Frame:

DNA (Sheared DNA)

↓

Library prep

↓

Enrichment (RNA or DNA beads in solution)

→ 2 days with automation

↓

Barcoding

Cluster generation → 1 day

↓

Sequencing → 14 days for paired-end

↓

Data Analysis → 5-10 days

CLINICAL EXOME SEQUENCING

Work flow :

DNA (Sheared DNA)

Library prep

Gene : Ex 1 Ex 2

Coverage

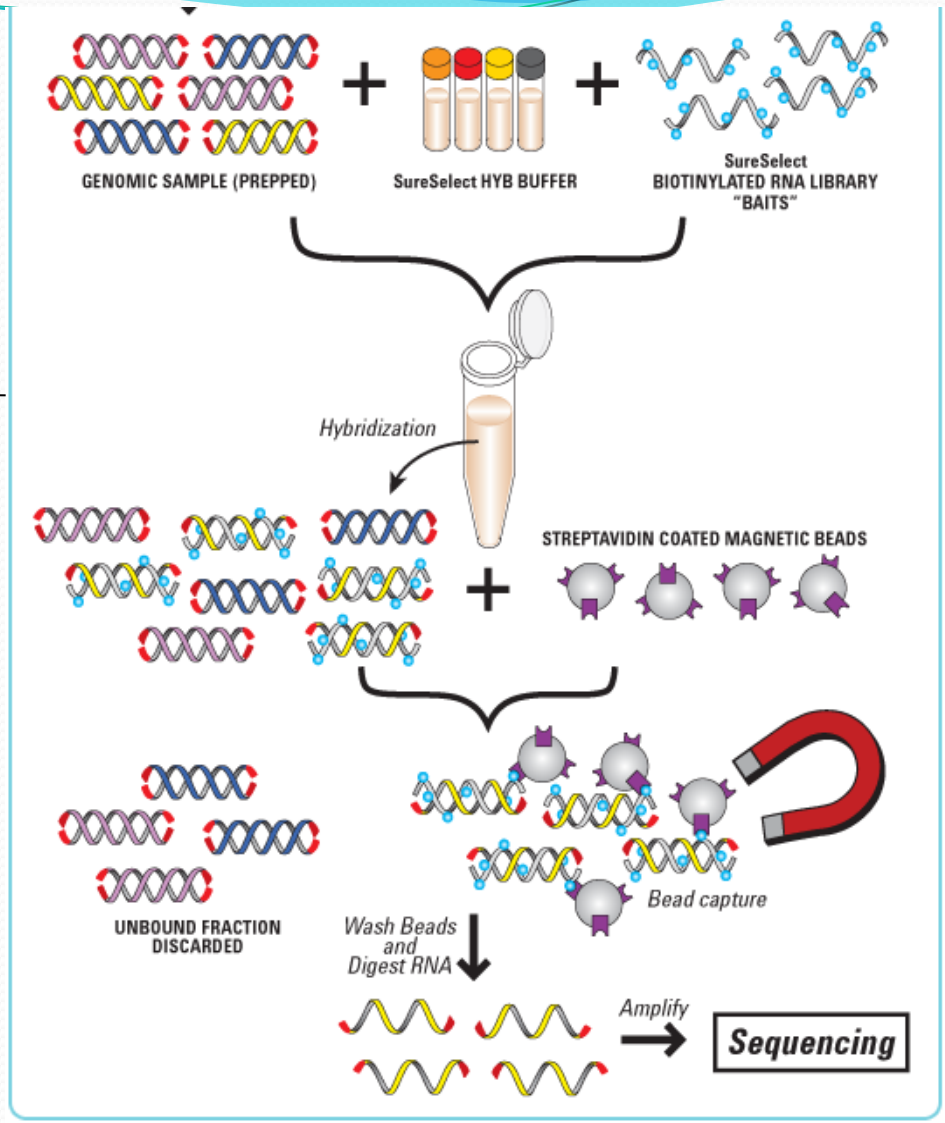
Enrichment

Barcoding

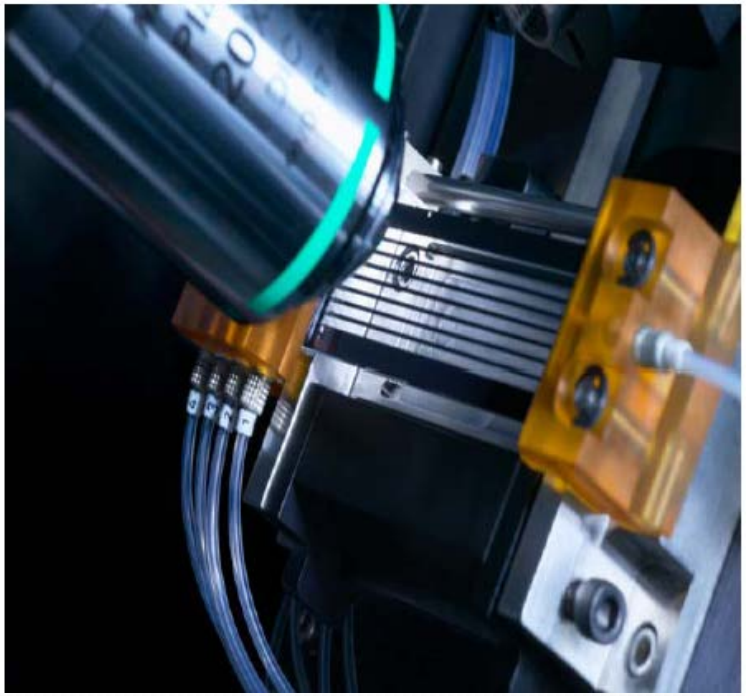
Cluster generation

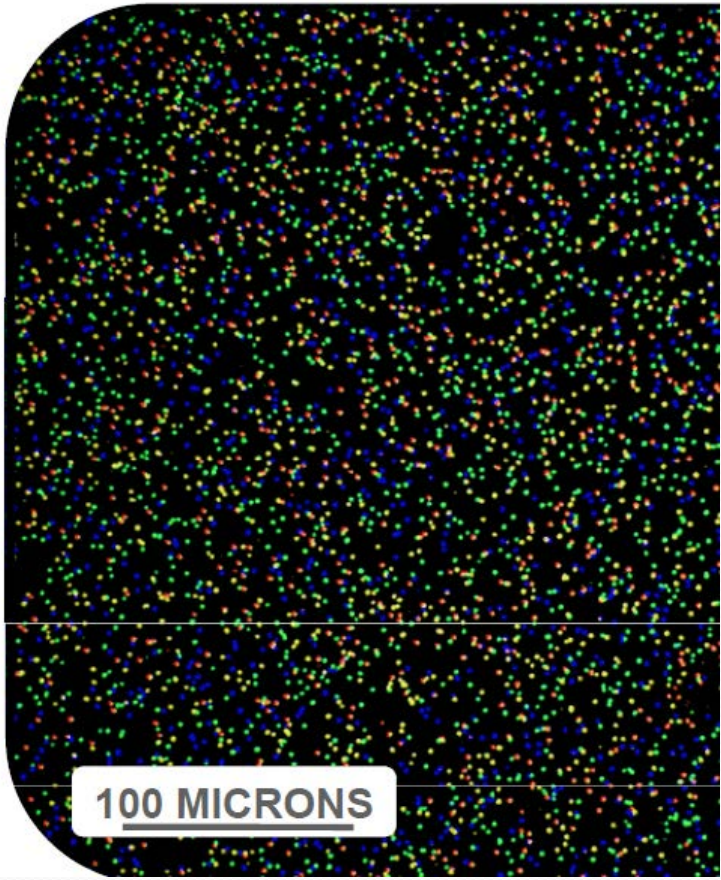
Sequencing

Data Analysis



Biotinylated RNA library baits covers all exons annotated in the consensus CDS database as well as flanking sequence for each targeted region and small non-coding RNAs

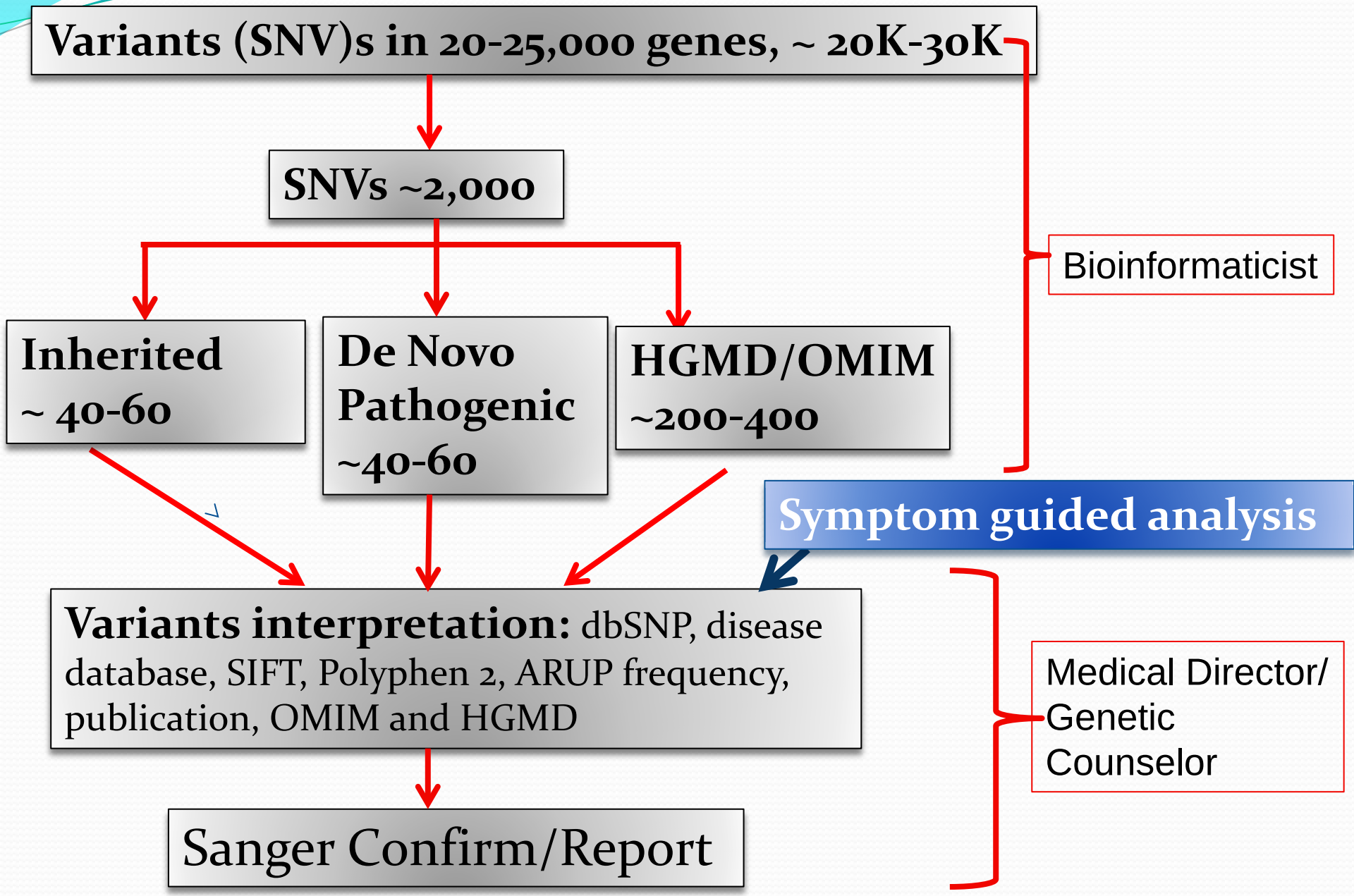


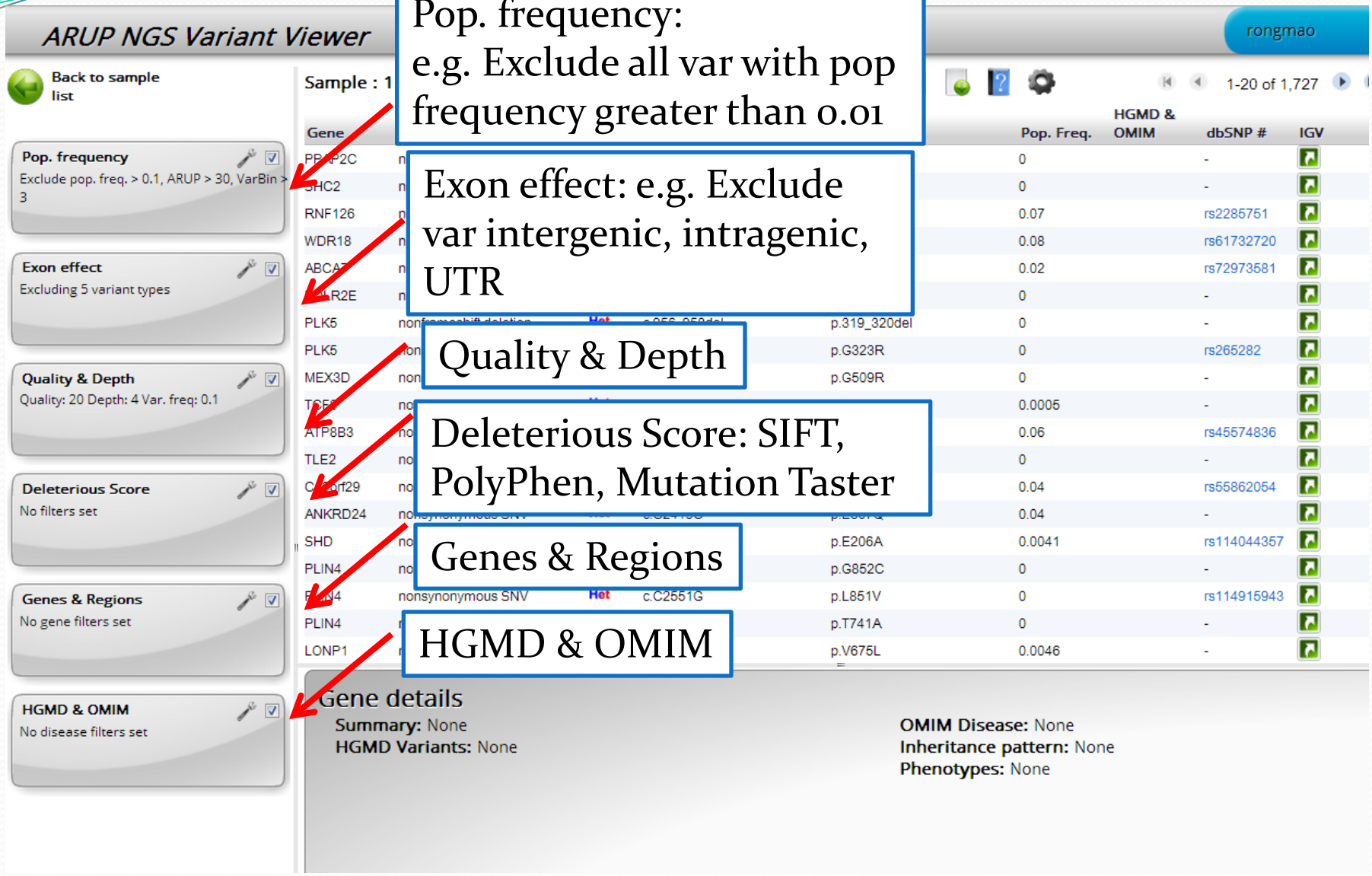


T G C A

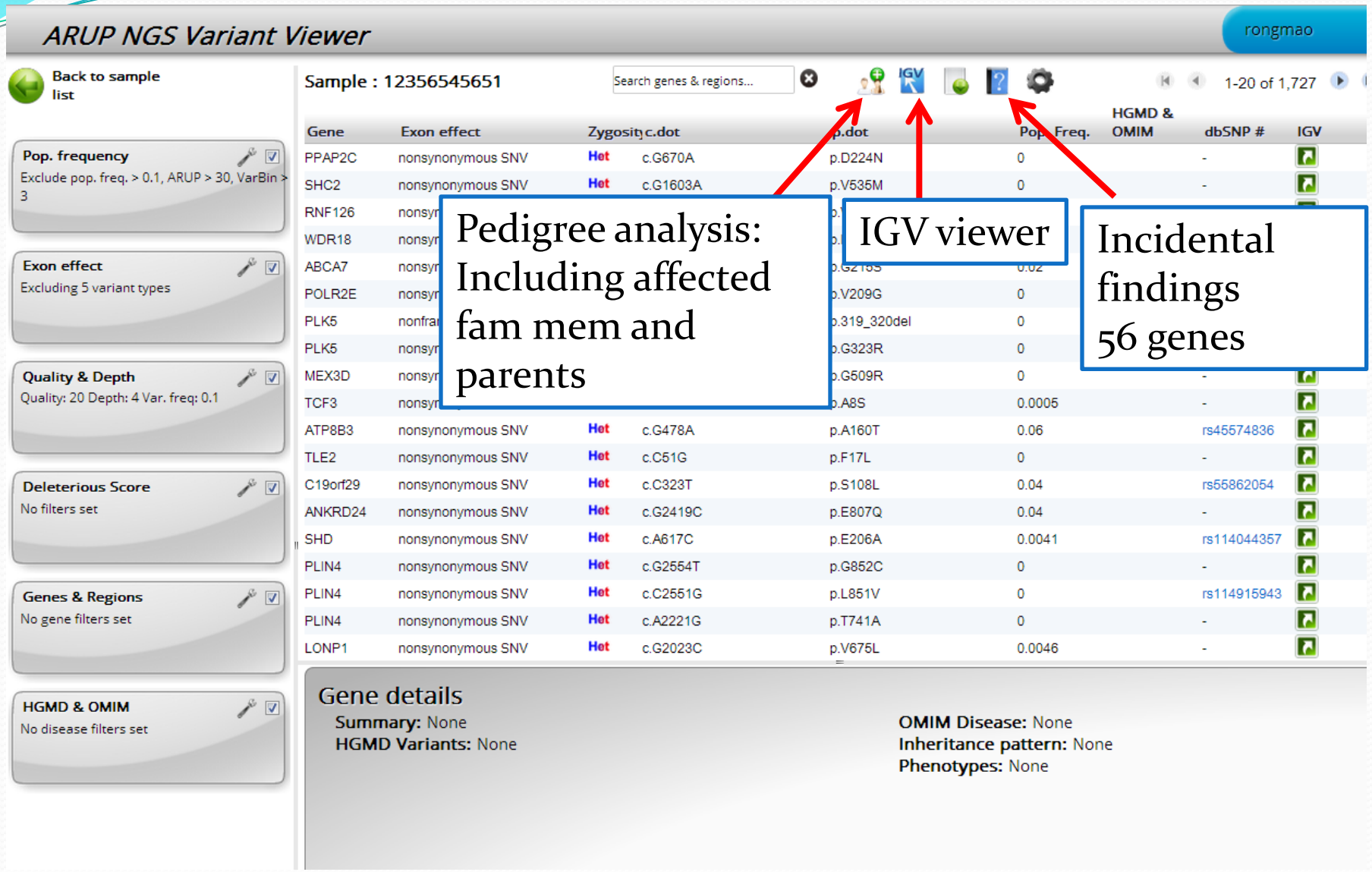
Image of clusters during sequencing.

Exome Interp Algorithm



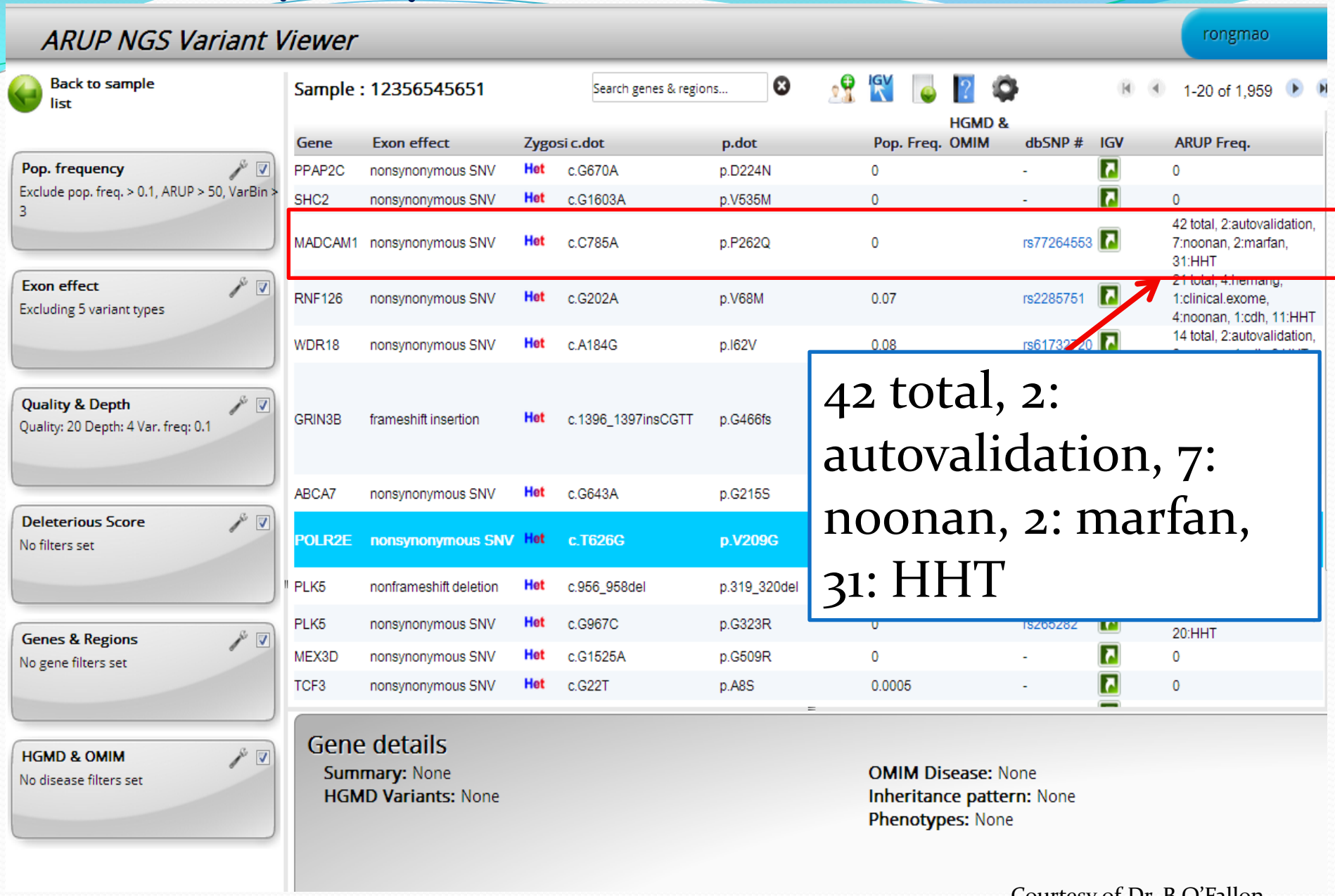


Courtesy of Dr. B O'Fallon



Courtesy of Dr. B O'Fallon

ARUP frequency:



42 total, 2: autovalidation, 7: noonan, 2: marfan, 31: HHT

Courtesy of Dr. B O'Fallon

What are incidental findings?

Variants found by exome/genome sequencing , which are unrelated to the disease of interest

- majority of them are benign
- a small number of them (between 1-5) might be well-described, disease-associated mutations

Courtesy of Dr. P. Bayrak-Toydemir

American College of Medical Genetics and Genomics

ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing

Robert C. Green, MD, MPH[1,2], Jonathan S. Berg, MD, PhD[3], Wayne W. Grody, MD, PhD[4-6], Sarah S. Kalia, ScM, CGC[1], Bruce R. Korf, MD, PhD[7], Christa L. Martin, PhD, FACMG[8], Amy McGuire, JD, PhD[9], Robert L. Nussbaum, MD[10], Julianne M. O'Daniel, MS, CGC[11], Kelly E. Ormond, MS, CGC[12], Heidi L. Rehm, PhD, FACMG[2,13], Michael S. Watson, MS, PhD, FACMG[14], Marc S. Williams, MD, FACMG[15], Leslie G. Biesecker, MD[16]

Direct laboratories to return with each genomic sequencing order results from 56 genes in which mutations greatly increase risk of 24 serious, but treatable diseases, even if clinicians do not suspect patients have them.

Incidental Findings

The ACMG Working Group recommended that the laboratory **actively search** for the specified types of mutations in the specified genes listed in these recommendations.

Mandatory reporting of known mutations for the disorders:
- Hereditary cancers,
- Marfan syndrome,
- Long QT syndrome,
- Brugada syndrome,
- Certain cardiomyopathies

Courtesy of Dr. P. Bayrak-Toydemir

Returning incidental findings in children

Recommendations for seeking and reporting incidental findings are not to be limited by the age of the person being sequenced.

The ethical concerns about providing children with genetic risk information about adult-onset diseases were outweighed by the potential benefit to the future health of the child and the child's parent of discovering an incidental finding where intervention might be possible.

Courtesy of Dr. P. Bayrak-Toydemir

Patient Consent and Opt-in/out option

- Proband and family members needs to consent for exome sequencing and incidental finding

- the ACMG Working Group revised document offering the patient a preference as to whether or not to receive the minimum list of incidental findings described in these recommendations.

Preanalytic Considerations

- Genetics counselors discuss the case with physician
  - Patient specific
    - well defined findings
    - good evidence for a genetic basis

    Obtain clinical information, lab results, MRI, etc.

  - Family specific:
    - affected family members
    - inheritance pattern

- Consenting patient and family members

Analytic Considerations

Limitations of exome testing

- capturing efficiency

Bioinformatic aspects

- variant calling
- filtering
- analyzing genes only in Human Genome Mutation Database or OMIM
- analyzing genes on mandatory reporting

Postanalytic Considerations

- Genetics Counselor and Medical Directors:
  - Reporting
    - negative, positive, uncertain for primary patient findings
    - incidental findings
    - limitations and quality of exome sequencing, coverage
  - Ethical and counseling issues
  - Patient consent
  - Education of consumers (patients, clinicians, payers)

Example 2: Clinical Information

- 7 yrs. male, Caucasian
- Suspected MPS, Cornelia De Lange
- Neuro: delayed speech, fine motor delays, cognitive delays
- Dysmorphic: hirsutism, coarse
- Growth: short (less than 32 centile), overweight
- Skeletal: diffuse osetopenia, significantly delayed bone age
- Craniofacial: cleft palate, macrodontia, cleft earlobe
- Dermatologic: eczema

Example 2 : Lab Results

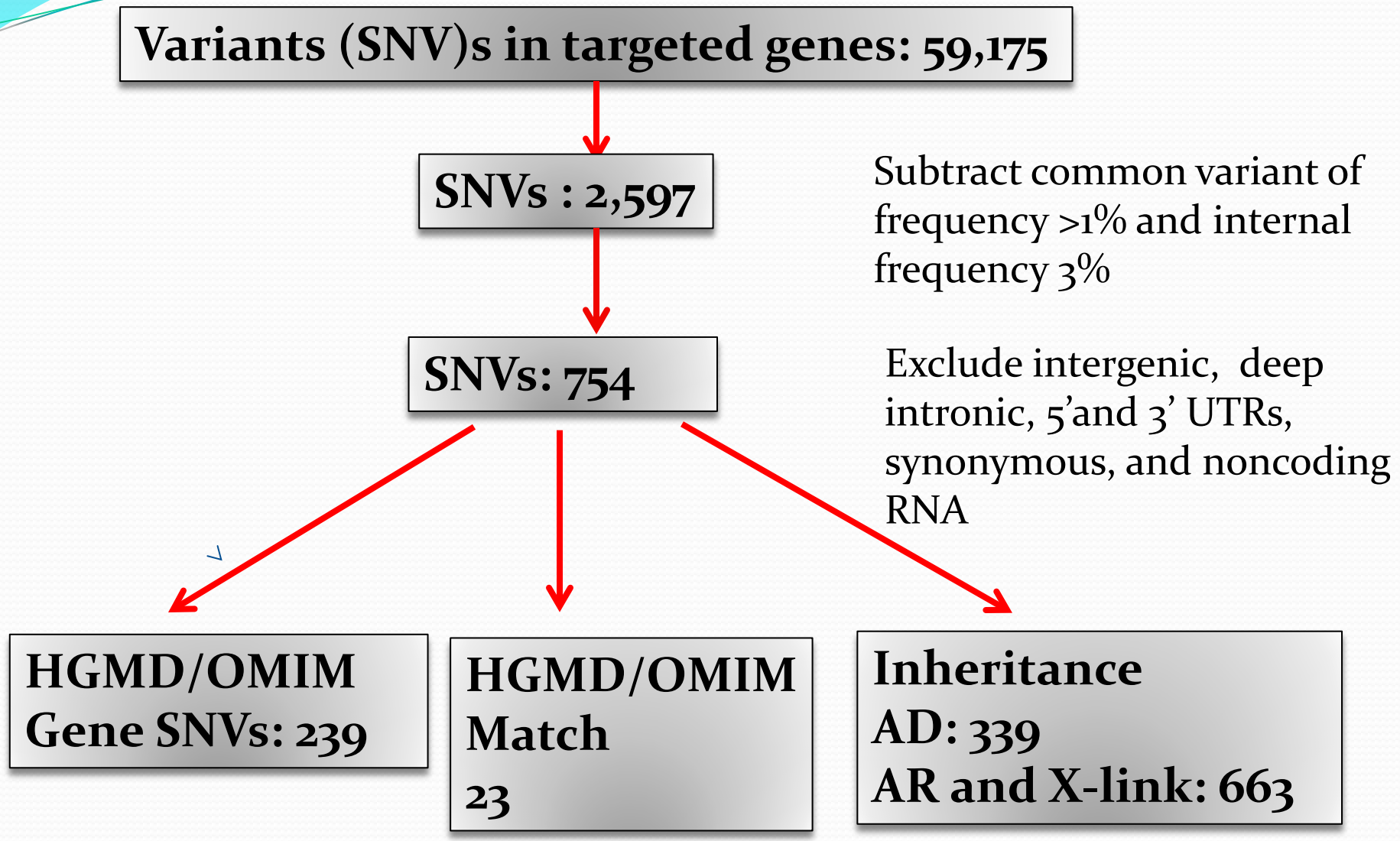
- Metabolic: Normal for MPS screening, UOA nonspecific elevations, phenylacetic acid (PAA) normal.
- Genetics: Normal Karyotype, microarray, Fragile X, and 22q deletion
- Brain: EEG suggests nonspecific diffuse cerebral dysfunction, MRI demonstrated mild cerebral atrophy and changes compatible with a Dx of cerebellitis.

- Family history: NO

Example 2 : Exome sequencing

- Exome sequencing performed on proband specimen
- Parental samples were available for Sanger confirmation

Example 2: Exome Data:



Subtract common variant of frequency >1% and internal frequency 3%

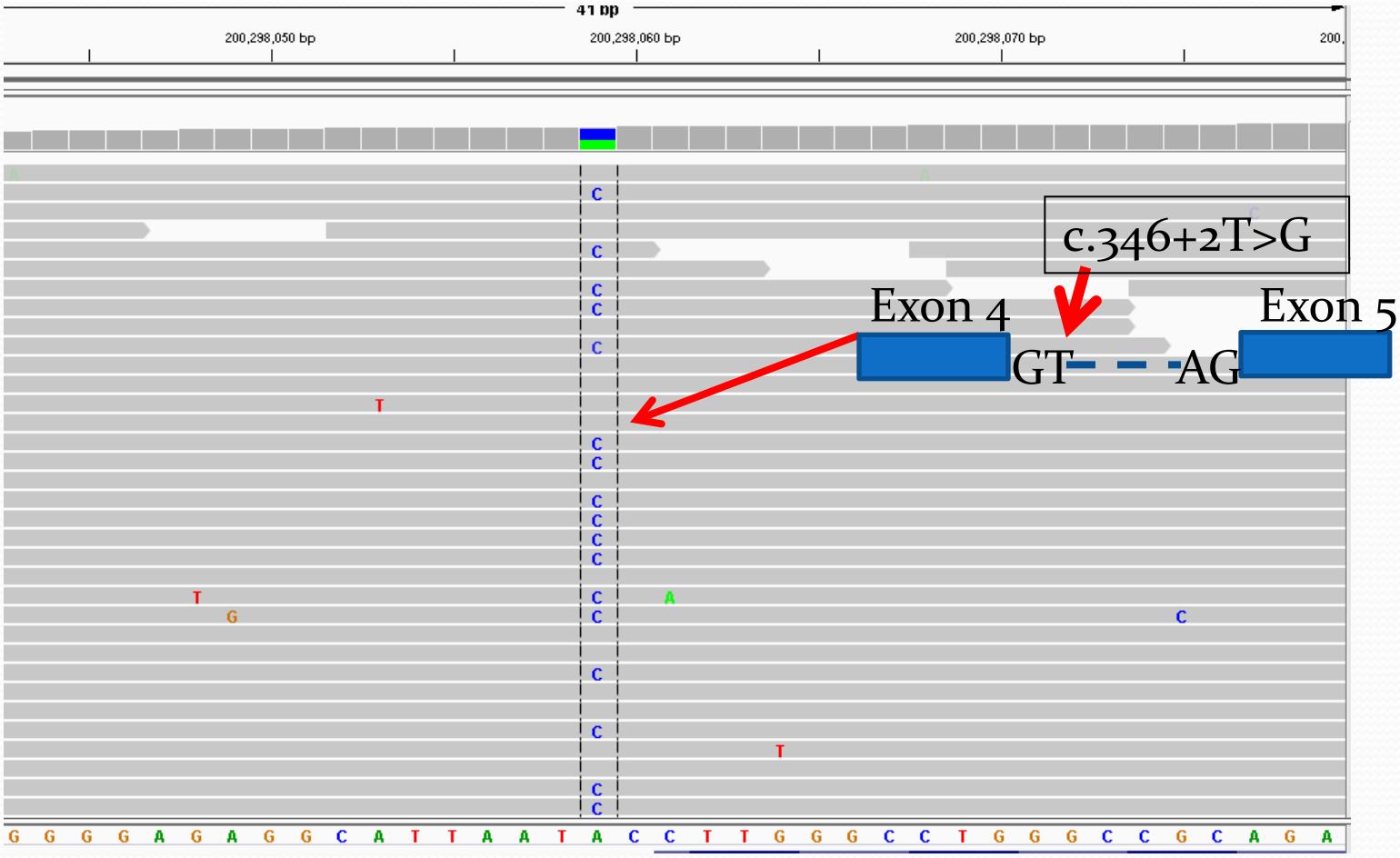
Exclude intergenic, deep intronic, 5'and 3' UTRs, synonymous, and noncoding RNA

Example 2: Candidate Genes/Variants

- No mutation in Cornelia de Lange genes: NIPBL, SMC1A, SMC3 and HDAC8
- No mutations in MPS genes: IDUA, IDS, SGSH, NAGLU, HGSNAT, GNS, GALNS, GLB1, ARSB, HYAL1, GUSB

Example 2: Candidate Gene/Variants

- Gene: SATB2
- Variant: c.346+2T>G (one copy)



Example 2: SATB2

- Sanger confirmation:
  - Confirmed on proband.
  - Testing both parents, none of them carried this mutation: *De Novo*

Example 2: *SATB2*

- *SATB2*: AT-rich sequence-binding protein 2 (*SATB2*) gene
- Encodes a protein binds nuclear DNA matrix attachment regions.
- Function: involving in transcription regulation, chromatin remodeling, play an important role in craniofacial patterning and brain development.

Example 2: *SATB2* mutations

- FitzPatrick, 2003. Reported 2 **de novo** chromosomal translocations involving 2q32-q33 in unrelated individuals with isolated cleft palate. One breakpoint was localized to intron 2 of SATB2, and the other breakpoint was located 130 kb 3-prime to the *SATB2* polyadenylation signal region.

Example 2: *SATB2* mutations

- Leoyklang, 2007, identified a het de novo mutation, R239X in the patient had isolated cleft palate, generalized osteoporosis, and profound mental retardation, consistent with Glass syndrome (OMIM 612313). The findings suggested a role for the SATB2 gene in malformation syndromes involving craniofacial patterning and brain development.

Example 2: *SATB2* mutations

- Docker, 2013, a de novo het R239X detected by whole exome sequencing in a 3 y.o. girl with cleft palate, severely delayed speech, hypotonia, and mental retardation.
- Dysmorphic facial features included hypotonic face with hypersalivation, hypertelorism, down slanting palpebral fissures, long eyelashes, upturned nose with broad tip, microretrognathia, long philtrum, low-set and posteriorly rotated ears, and crowded teeth. She also had severe sleeping disturbances, restlessness/hyperactivity, and recurrent temper tantrums

Example 2: report

- Positive
- A pathogenic mutation in SATB2 gene caused patient's phenotype
- No incidental finding

GUIDELINES/REGULATIONS
CLIA/CAP/ACMG

Next Generation Sequencing

Next Generation Sequencing (NGS) incorporates two processes: (1) the analytical wet bench process of sample and library preparation and sequence generation and (2) the bioinformatics process or pipeline of sequence alignment, annotation and variant calling. These two processes are inextricably linked as the output from each process supports the optimization of the other. The large volumes of data produced by NGS platforms put substantial demands on laboratories in terms of the requirements for documentation, validation, quality control and assurance, monitoring, data storage, as well as assessment and implementation of new technology and software releases.

Inspector Instructions:

- Sampling of next generation sequencing policies and procedures
- Records of wet bench processing and bioinformatics process validation
- QM program records with corrective action for component failure
- Sampling of exception log records

Guide validation of samples, analysis and reporting

Clinical Laboratory Standards for Next Generation Sequencing

Heidi L. Rehm, PhD[1,2], Sherri J Bale, PhD[3], Pinar Bayrak-Toydemir, PhD[4], Jonathan S Berg, MD[5], Kerry K Brown, PhD[6], Joshua L Deignan, PhD[7], Michael J Friez, PhD[8], Birgit H Funke, PhD[1,2], Madhuri R Hegde, PhD[9], Elaine Lyon, PhD[5]; A Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee

Clinical Laboratory Improvement Amendments

Assuring the Quality of Next-Generation Sequencing in Clinical Laboratory Practice

Next-generation Sequencing: Standardization of Clinical Testing (Nex-StoCT) Workgroup Principles and Guidelines

Supplementary Guidelines

CAP Proficiency Test

- CAP proficiency test available and the first shipment is in 2015
- Methodology PT

NGS Validation-Accuracy

- Concordance of results using 2 different assays
  - Known samples from another lab or Coriell repository (HapMap sample-NA12878)
  - ARUP samples analyzed by 2 different methods (Exome-NGS, Sanger)
  - Known positive samples with characterized mutations
  - Confirmed polymorphisms (not disease causing) can be used for method accuracy (HapMap sample NA12878)
- Bioinformatics pipeline parameters and output are captured
- Data from HapMap sample used to determine analytical sensitivity and specificity

Courtesy of Dr. G. Pont-Kingdon

Analytical Sensitivity

- Performance of the assay **to detect the known variants** in the HapMap Sample
- TP/TP+FN
- Will often be >99% because all targeted variants are detected
- Some type of variants (large del-dups) will not be detected
- FN can result from unknown variants in the sequences used for capture

Courtesy of Dr. G. Pont-Kingdon

Analytical Specificity

- Performance of the assay **to detect normal sequences** in the HapMap sample
- TN/TN+FP
- Will often be >99%
- Pseudogenes and difficult regions are examined carefully to limit detection of FP
- Sanger sequencing always performed before reporting a positive

Example from a small gene panel:

Table 4. HapMap Sample comparison*

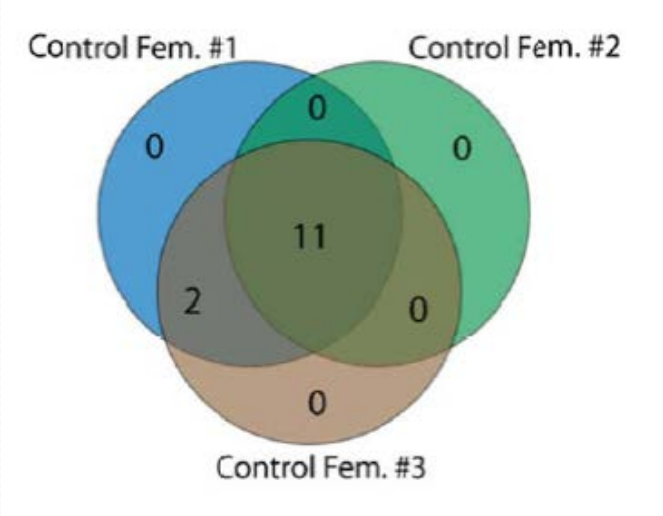
Total sites examined	15,981
Total true SNPs	13
Total SNPs identified	14
True positives	13 (100%)
False positives	1 (7.1%)
False negatives	0 (0%)

Courtesy of Dr. G. Pont-Kingdon

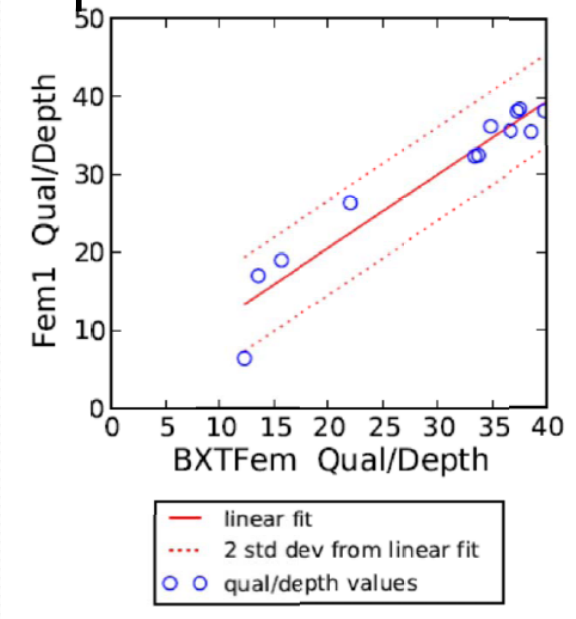
NGS Validation-Precision

- Within run:
  - Concordance of variant calls in triplicate from one sample
- Between run
  - Concordance of variant calls and quality between two independent experiments from one sample

Example of “within run”



Example of “Between run”



Courtesy of Dr. G. Pont-Kingdon

NGS Validation-Reportable Range & Reference Interval

- Reportable range:

  List of genes and regions analyzed

- Reference range:

  Common Polymorphisms found in normal samples

Courtesy of Dr. G. Pont-Kingdon

Conclusions

- Next generation sequencing technology provides opportunities for large scale genomic sequencing
- The complexity increases from gene panel to exome to whole genome sequencing
- Next generation sequencing requires advanced informatics for data analysis and annotation tools are rapidly advancing
- Variants detected need confirmation, and causality needs evidence
- Clinical and family information is critical in assessing significance

Acknowledgement

ARUP R&D and Bioinformatics

- Whitney Donahue
- Shale Dames
- Brendan O'Fallon
- Genevieve Pont-Kingdon
- Karl Voelkerding

University of Utah

- Nicola Longo
- Kathryn Swoboda

ARUP Institute for Clinical & Experimental Pathology

ARUP Genomics Lab

- Pinar Bayrak-Toydemir
- Jennifer Stocks
- Tatiana Tvrdik
- Tyler Wayman
- Lisa Robles
- Josh Raney
- Scott Pew
- Chad VanSant Webb

ARUP Genetics Counselors

- Chris Miller
- Patti Krautscheid

ARUP Laboratories