Next Generation Sequencing Analyses of Complex Dual Genome Mitochondrial Disorders: Technical Approach

ABRF Satellite Workshop
Palm Springs, CA
March 2, 2013

Lee-Jun C. Wong, Ph.D.
Molecular and Human Genetics
Baylor College of Medicine
ljwong@bcm.edu

content:

1.Application of NGS to molecular diagnosis of mitochondrial disorders: nuclear genes and mitochondrial genome

2.Validation and Quality control of Clinical tests

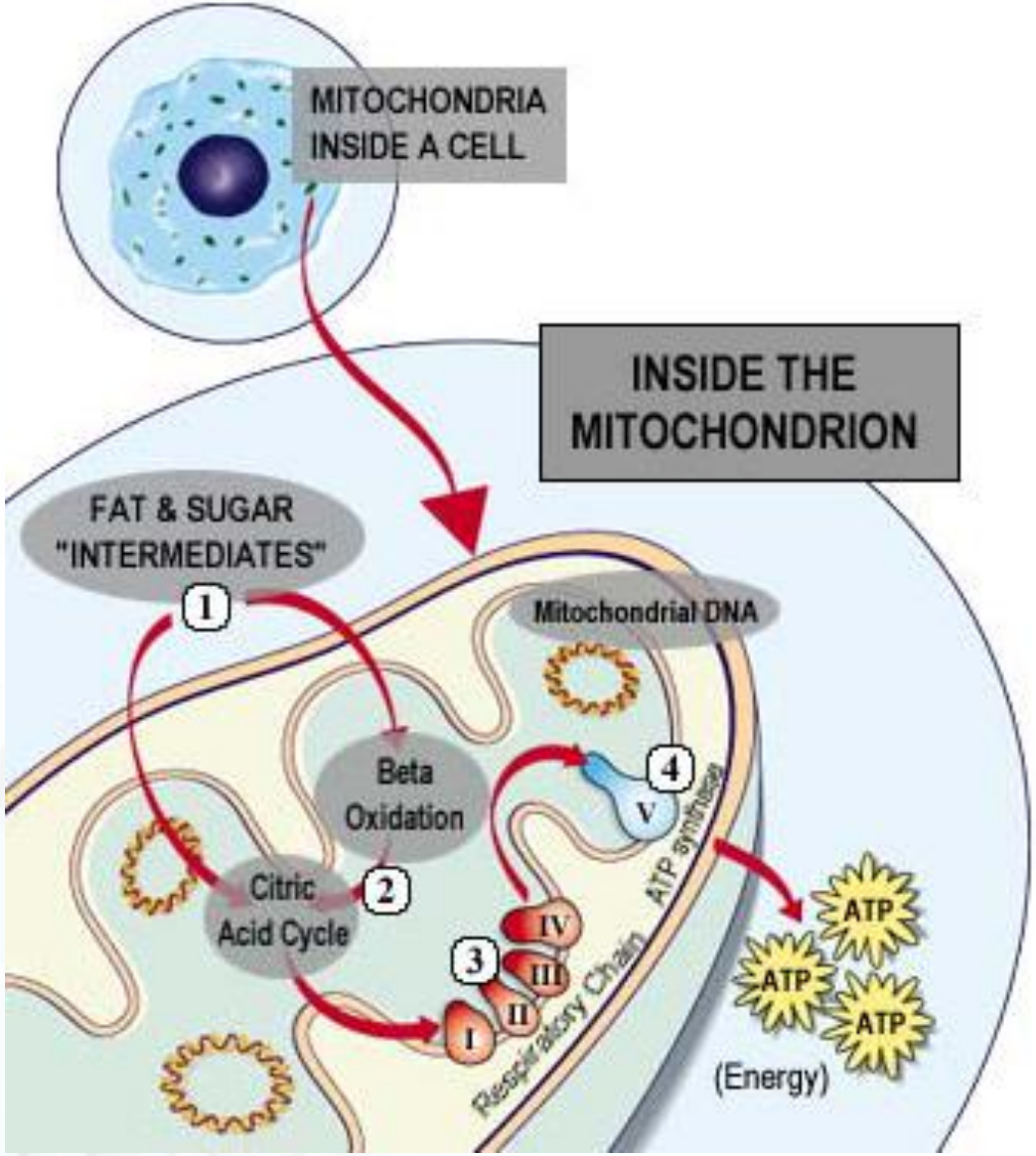
3.Types of mutations detected: point mutations, small indel, large deletions?

4.Target nuclear gene capture/sequencing

5.Mitochondrial genome: long range PCR of the whole mitochondrial genome

Mitochondrial Function: produce energy

Involves 2 genomes: mitochondrial and nuclear



Fuels:
Fatty acids
carbohydrates

↓

Combustion:
Fatty acid oxidation
TCA cycle

↓

$NADH, FADH_2$

↓

electron transport chain

↓

O_2 **Oxidative phosphorylation** H_2O

ADP **ATP**

↓

NAD, FAD

Mitochondrial Disorders:
Disease of Energy Deficiency
Respiratory chain defect
Defects in oxidative phosphorylation

Preferentially affect tissues of high energy demand

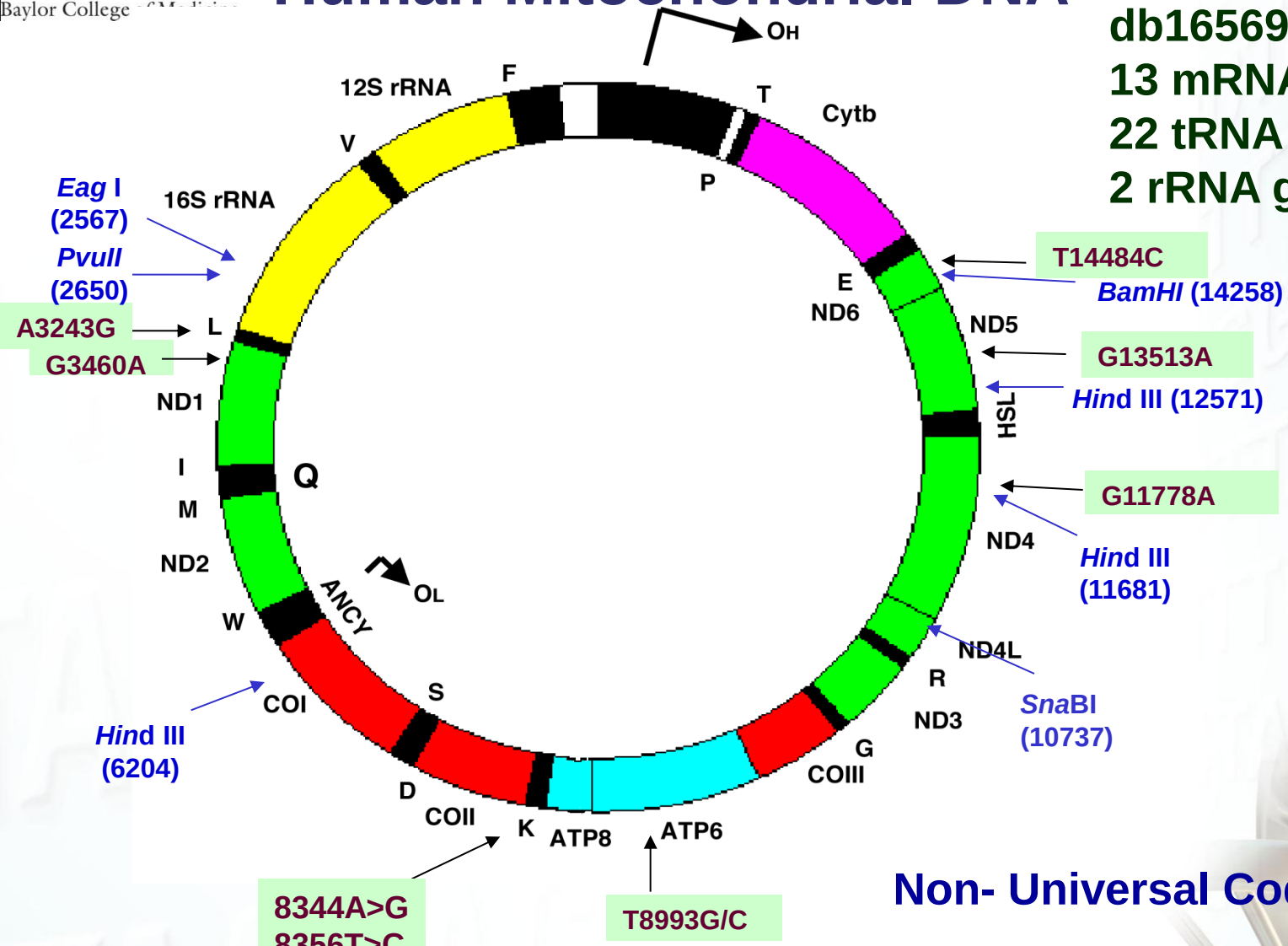
Major clinical manifestation:

neuromuscular phenotype

CNS, Brain, skeletal muscle, heart, liver, etc.

Human Mitochondrial DNA

db16569 bp
13 mRNA genes
22 tRNA genes
2 rRNA genes



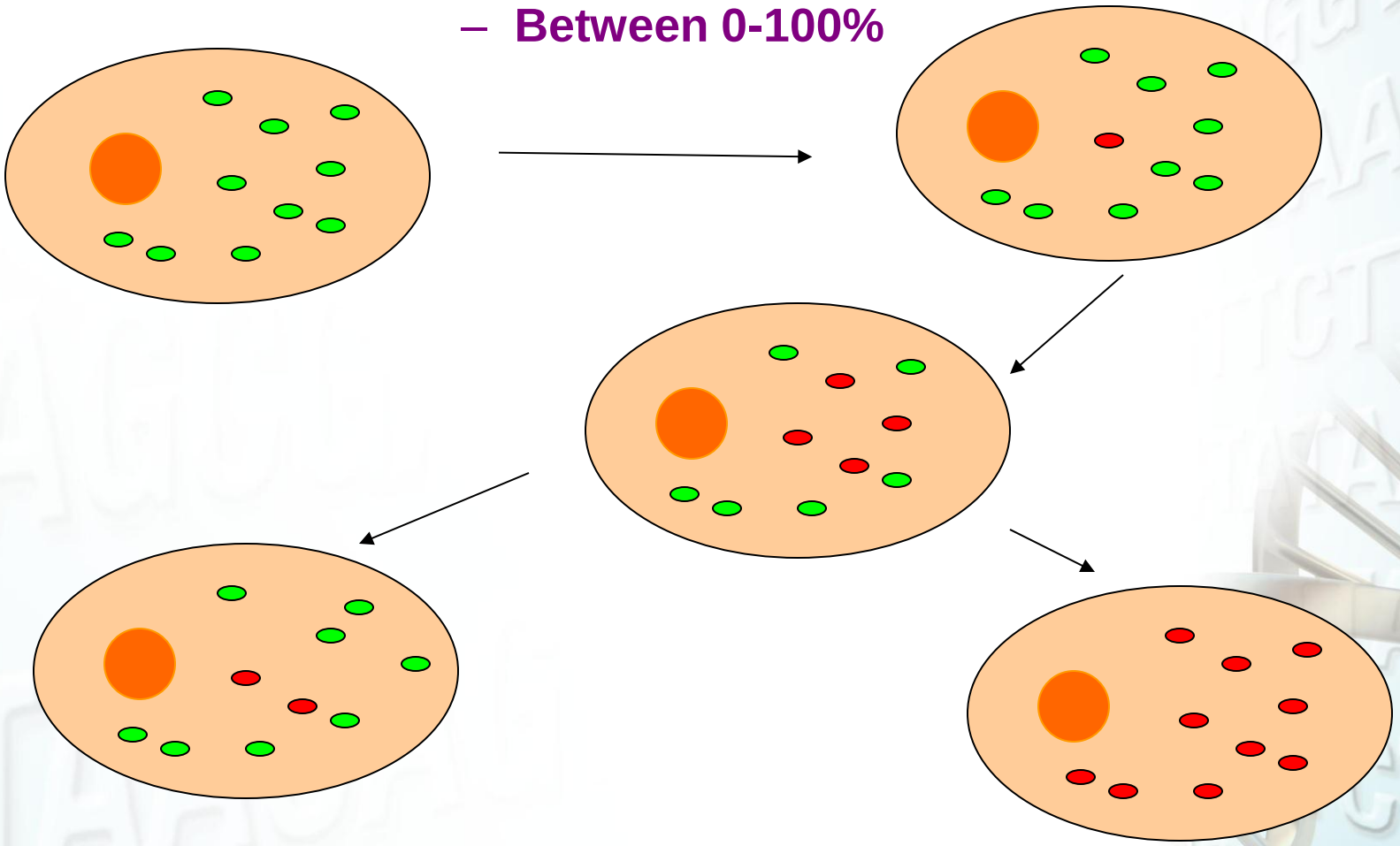
Non- Universal Codon

Light strand: 8 tRNA and 1 mRNA (ND6)
Heavy strand: 14 tRNA, 2 rRNA, and 12 mRNA
Polycistronic with posttranscriptional processing

Point mutations
Large deletions

Homoplasmy and Heteroplasmy

- **Homoplasmy**
  - 0 or 100%
- **Heteroplasmy**
  - Between 0-100%



Complex dual genome mitochondrial disorders:

- mtDNA biosynthesis+integrity maintenance
- Salvage synthesis of dNTP
- Complex assembly/Complex subunits
- Transcription/translation factors
- MRPLs/MRPSs (mito ribosomal proteins)
- Transcription and translation factors
- Mitochondrial aa-tRNA synthetases
- TIMMs and TOMMs, protein transporters
- dynamic fusion/fission proteins
- Apoptotic factors, protein kinases

Majority of mitochondrial disorders are caused by Defects in nuclear genes

1500 nucear genes targeted to mitochondria

Currently about 200+ linked to known diseases

BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

Current Approaches: step-wise

1. Screen for mtDNA common point mutations: by PCR/ASO or other detection methods
2. mtDNA deletion: by Southern analysis
3. Quantification of heteroplasmic mtDNA point mutations: ARMS qPCR
4. Determination of mtDNA deletion and breakpoints: aCGH, PCR sequencing
5. Unknown mutations: sequence the whole mitochondrial genome by Sanger
6. mtDNA depletion: qPCR analysis for mtDNA copy number
7. Sequence relevant nuclear genes, one by one
8. aCGH to detect large deletions in nuclear genes

Gold Standard Sanger Sequencing

Pitfalls

1. Does not provide quantitative information

2. Sequence gene one by one

3. Does not detect deletions

4. Tedious and costly

5. Not comprehensive

BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

Mitochondrial Challenges

1. *The most clinically and genetically heterogeneous dual genome disorders*
2. *Primary defects in mitochondrial genome, common point mutations and large deletion*
3. *Quantification of mutation heteroplasmy*
4. *Majority (90%) of mitochondrial disorders are caused by one of ~1500 nuclear genes*
5. *Advances in technologies for diagnosis of complex disorders: array CGH and next generation sequencing approach*

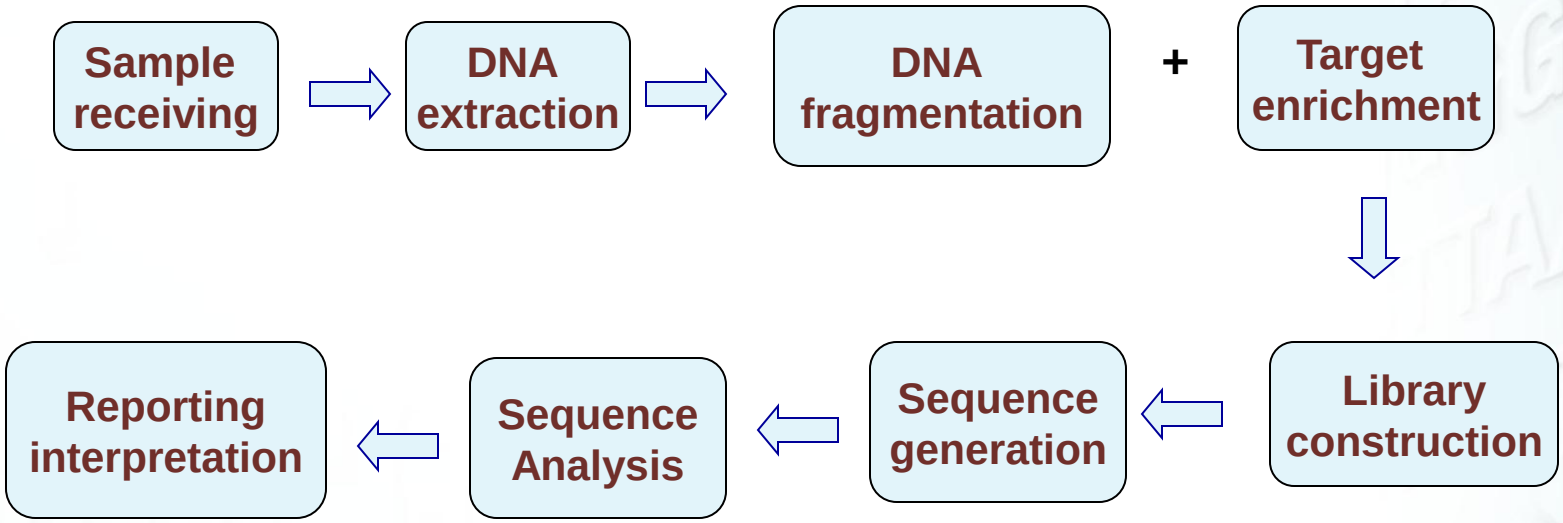
BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES
MEDICAL GENETICS LABORATORIES

Next Generation Massively Parallel Sequencing

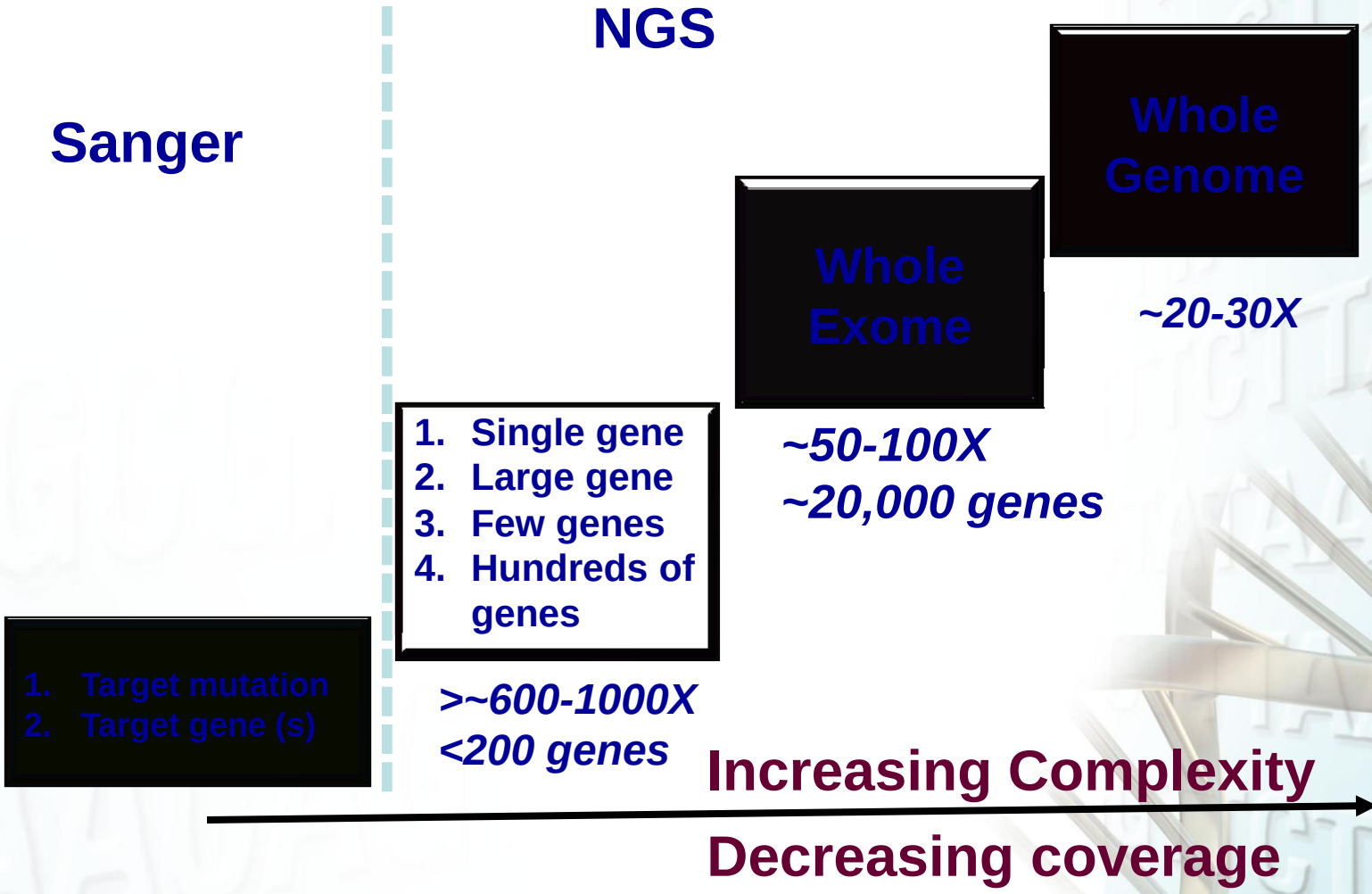
1. *Ability to sequence many genes in parallel*
2. *Identify new mutations in known genes*
3. *Discover new disease genes*
4. *Detect point mutations, small indels and large deletion/duplication (CNV)*
5. *Quantify mtDNA heteroplasmy, mosaicism*
6. *RNA sequencing, Gene expression*
7. *Quantitative DNA methylation analysis*

NGS workflow

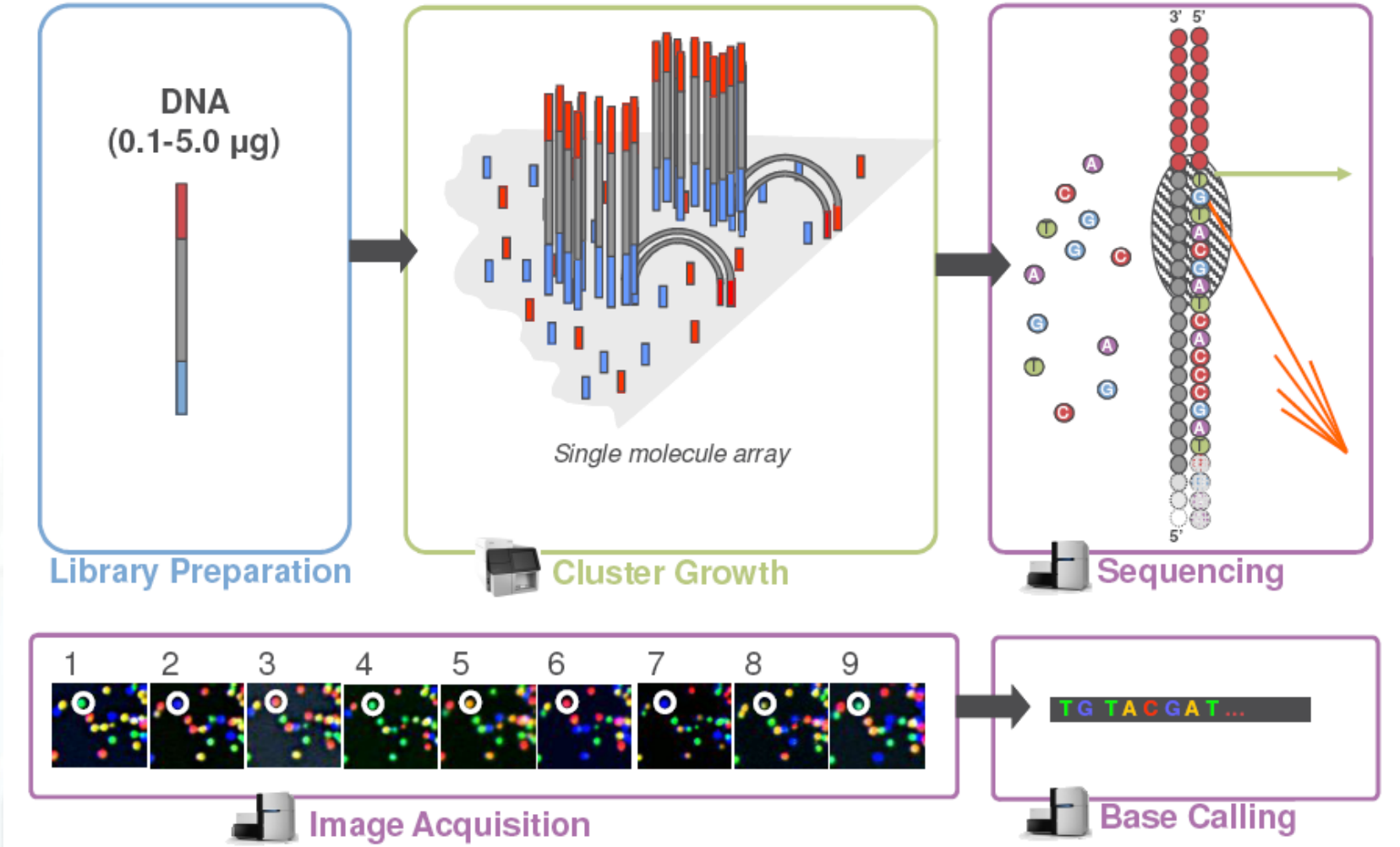


Quality control procedures are required to assure that each step works properly and results are accurate for each patient's specimen analyzed

Bring NGS to Clinical Diagnosis



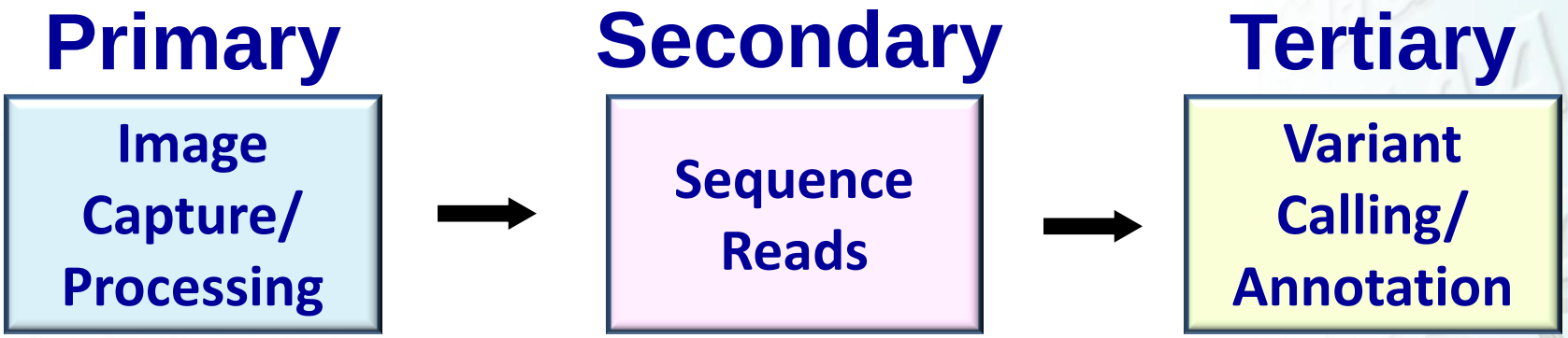
Illumina Sequencing Technology Overview



Primary analysis

(Adapted from Illumina.inc.)

sequence analysis: three steps
With built-in QA/QC samples



Convert image to base calls
Base quality scores assigned

Filtering of reads
Based on quality
Alignment / Assembly

Results interpretation

BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

To bring NGS to clinical settings requires

1. Validation:

Cover all bases in all CDS + 20 bp in flanking introns all mutations/VUS confirmed by a second method

Phase I: specificity, sensitivity, reproducibility, accuracy, compared to Sanger

Phase II: detection of different mutation types using positive control samples

Phase III: Blinded Samples without molecular diagnosis to obtain diagnostic yield

2. Define experimental error, limit of detection, alignment and analytical steps

3. Variants interpretation and reporting

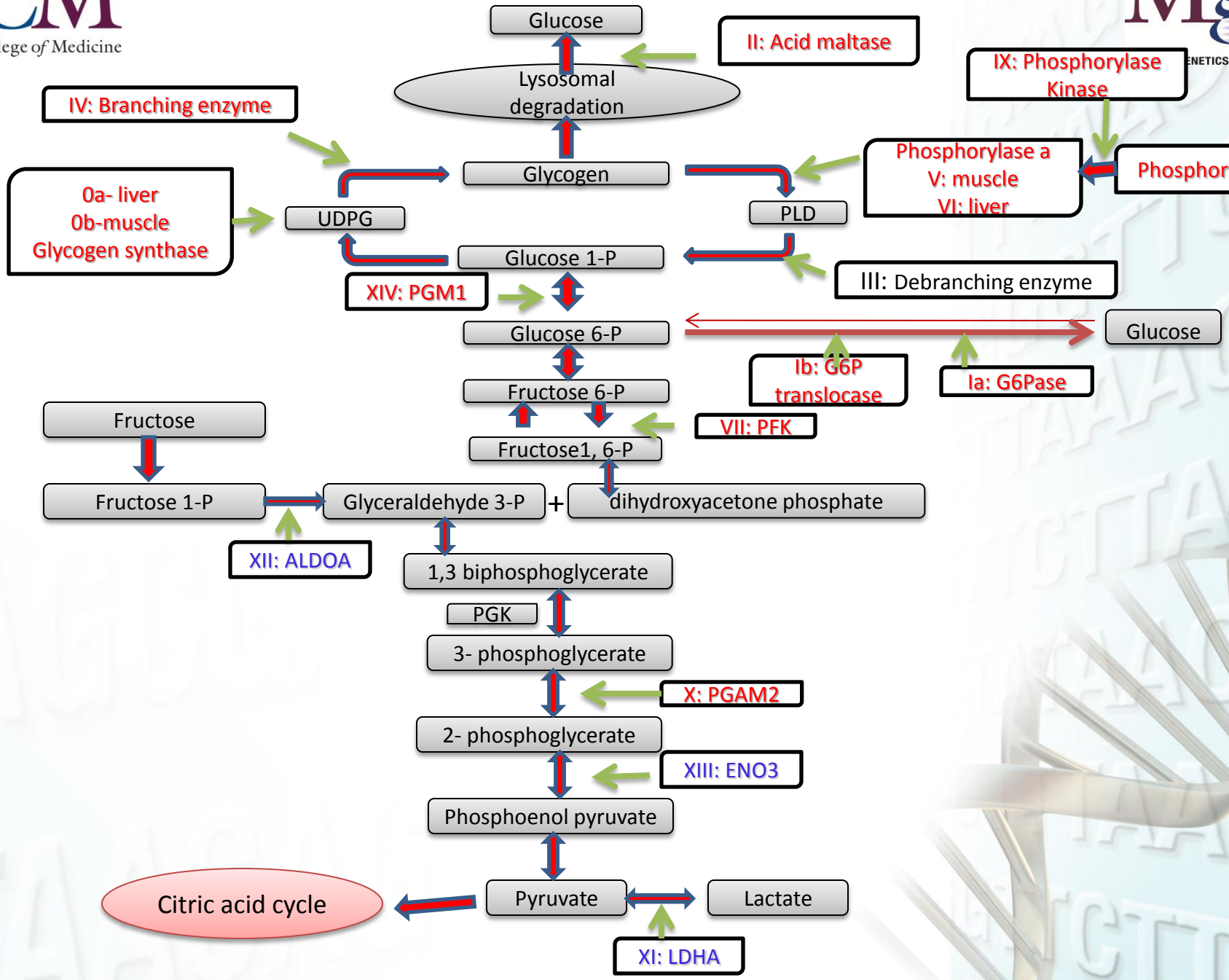
Nuclear Genes Capture Sequencing

1. Genes responsible for mtDNA Depletion and maintenance of integrity
2. GSD (liver and muscle forms)
3. Complex I-V panel, CoQ panel
4. Usher panel
5. PDH panels
6. Metabolic myopathy
7. RP (retinitis pigmentosa) panel 66 genes
8. Mitome200
9. Mitome500
10. Mitome1500
11. Exome
12. Whole Genome

MEDICAL GENETICS LABORATORIES

Example of Nuclear Gene Capture Sequencing

1. Panel testing:
   NGS analysis of a group of genes involved in Glycogen Metabolism: synthesis and breakdown
   Glycogen Storage Disorders (GSD)

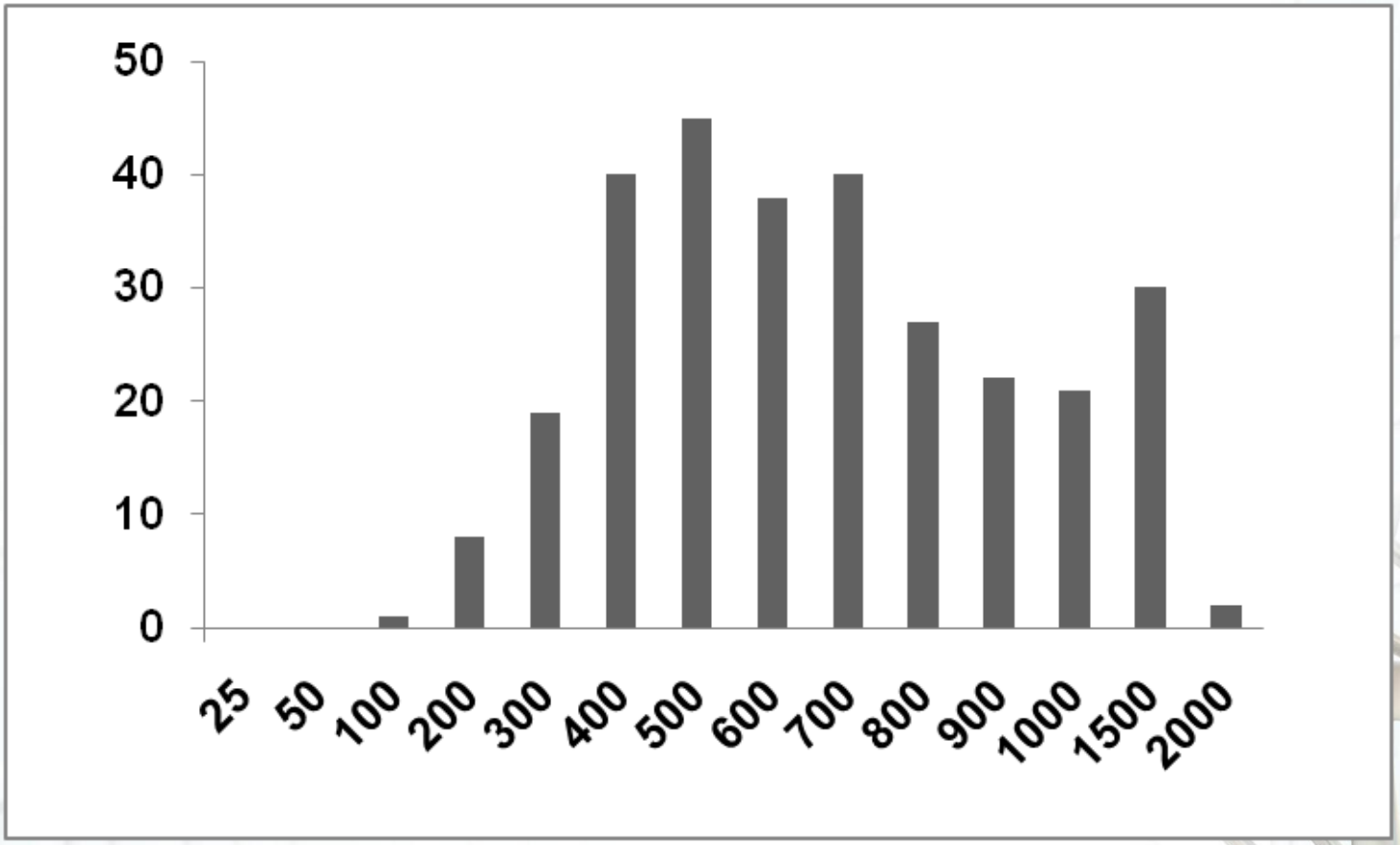

Glucose
II: Acid maltase
Lysosomal degradation
IX: Phosphorylase Kinase
IV: Branching enzyme
Glycogen
Phosphorylase a
V: muscle
VI: liver
Phosphorylase b
0a- liver
0b-muscle
Glycogen synthase
UDPG
PLD
Glucose 1-P
III: Debranching enzyme
XIV: PGM1
Glucose 6-P
Glucose
Ib: G6P translocase
Ia: G6Pase
Fructose 6-P
Fructose
VII: PFK
Fructose1, 6-P
Fructose 1-P
Glyceraldehyde 3-P
+
dihydroxyacetone phosphate
XII: ALDOA
1,3 biphosphoglycerate
PGK
3- phosphoglycerate
X: PGAM2
2- phosphoglycerate
XIII: ENO3
Phosphoenol pyruvate
Citric acid cycle
Pyruvate
Lactate
XI: LDHA

Genes involved in Glycogen Metabolism

GSD Types	Genes	Liver Panel	Muscle panel	NM#
GSD 0A	*GYS2*	√		NM_021957.3
GSD 0B	*GYS1*		√	NM_002103.4
GSD IA	*G6PC*	√		NM_000151.2
GSD IB	*SLC37A4*	√		NM_001467.5
GSD II	*GAA*	√	√	NM_000152.3
GSD III	*AGL*	√	√	NM_000642.2
GSD IV	*GBE1*	√		NM_000158.3
GSD V	*PYGM*		√	NM_005609.2
GSD VI	*PYGL*	√		NM_002863.4
GSD VII	*PFKM*		√	NM_000289.5
GSD IX A	*PHKA2*	√		NM_000292.2
GSD IX B	*PHKB*	√	√, mild	NM_000293.2
GSD IX C	*PHKG2*	√		NM_000294.2
GSD IX D	*PHKA1*		√	NM_002637.3
GSD X	*PGAM2*		√	NM_000290.3
GSDXIV	*PGM1*		√	NM_002633.2

NGS Panel name	GSD-16gene-panel
Genes included	*AGL, G6PC, GAA, GBE1, GYS1, GYS2, PFKM, PGAM2, PGM1, PHKA1, PHKA2, PHKB, PHKG2, PYGL, PYGM, SLC37A4* (16 genes)
Number of CDS	294
Target size	50,062 bp (CDS ± 20 bp)
Enrichment	In solution capture library
Sequencing info	Illumina HiSeq 2000, 75 cycle, single-end

Minimal coverage per base of Exons

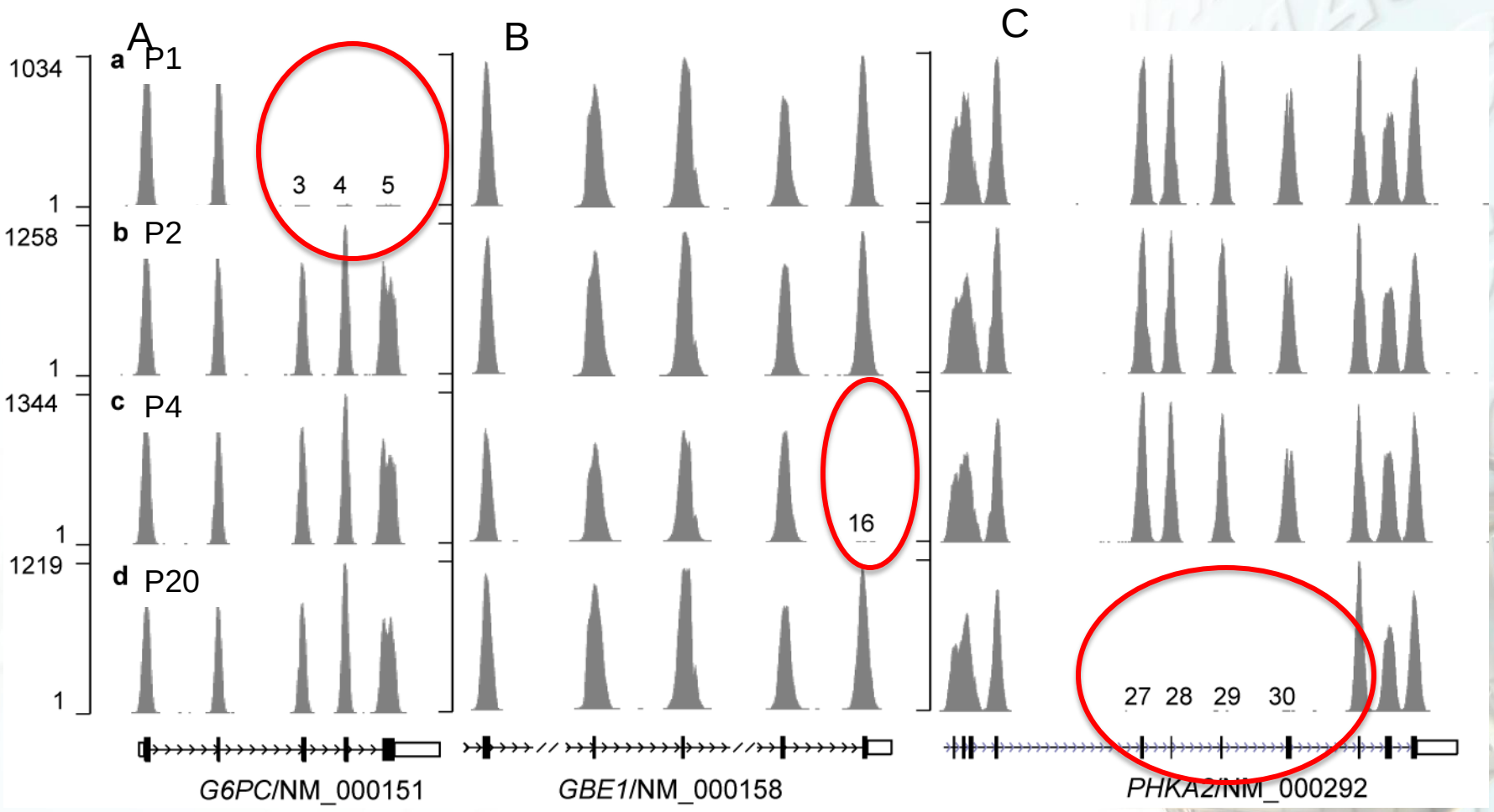


BCM
Baylor College of Medicine

Phase I Validation: known samples

MGL
MEDICAL GENETICS LABORATORIES

Sample ID	Mean coverage	Total reads Per 100 bp	**Min coverage**	**# of CDS < 20X**	**Multiplexing factor**
547	722±233	988±317	68	0	8
755	837±195	1146±266	77	0	8
700	783±187	1072±255	93	0	8
833	803±189	1099±258	74	0	8
264	686±225	939±306	47	0	8
041	841±220	1151±300	51	0	8
203	747±238	1021±324	92	0	8
941	674±220	921±299	1/92	3	8
545	731±230	1004±313	77	0	8
206	727±243	998±330	1/77	1	8
067	706±228	971±311	93	0	8
504	623±177	856±244	59	0	8
531	850±345	1169±418	69	0	8
255	878±267	1028±355	93	0	8


A
B
C
a P1
b P2
c P4
d P20
1034
1
1258
1
1344
1
1219
1
3 4 5
16
27 28 29 30
G6PC/NM_000151
GBE1/NM_000158
PHKA2/NM_000292

BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

GSD1A negative

Patients	Age	Gender	Clinical Indication
24547 P8	4m	M	hypoglycemia, hepatomegaly
28755 P9	13yr	F	fat,encephalopathy, abnormal liver function
30700 P10	1.5yr	F	hyperlipidemia, hyperlactatemia,ftt, hepatomegaly
31833 P11	10m	F	hypoglycemia, hepatomegaly
34264 P12	3yr	M	hypoglycemia, hyperuricemia, reccurent infections, bone fractures
36041 P13	3m	F	pulmonary hypertension, large liver, elevated lipids/uric acid/lactate
37203 P14	2yr	M	Hepatomegaly

NGS results summary

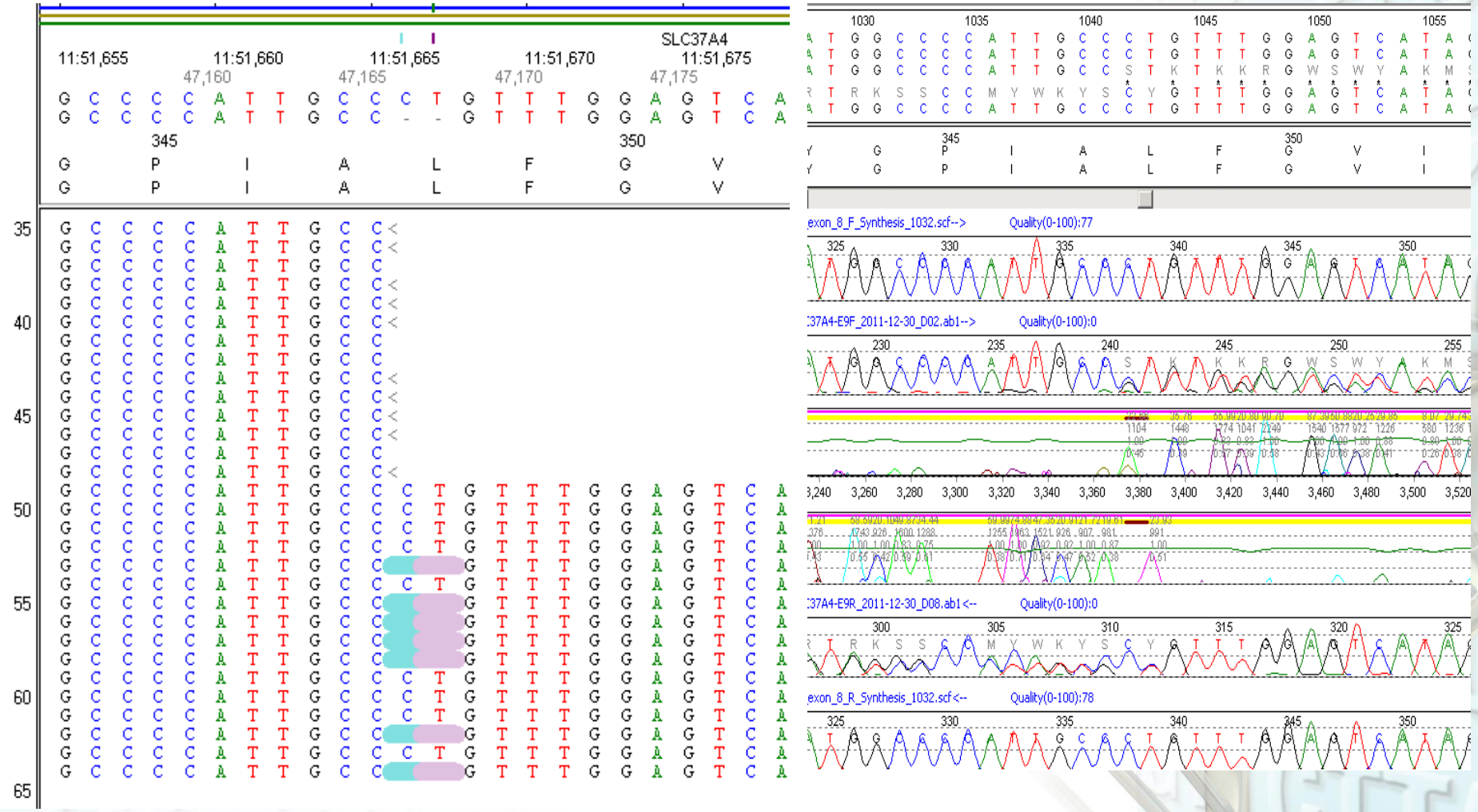
Patient	Gene	CDS	exons	mutations
30700 P10	SLC37A4	5	6	c.817G>A (p.G273S)
		7	8	c.1042_1043delCT (p.L348Vfs*53)
31833 P11	SLC37A4	5	6	c.785-3_786del5
		5	6	c.785-3_786del5
37203 P14	AGL	2	4	c.256C>T (p.Q86X)
		20	22	c.2723T>G (p.L908R)

c.817G>A (p.G273S): conserved from C. elegans to human, predicted to be deleterious.
c.2723T>G (p.L908R): conserved from yeast to human, predicted to be deleterious.

Baylor College of Medicine

MEDICAL GENETICS LABORATORIES

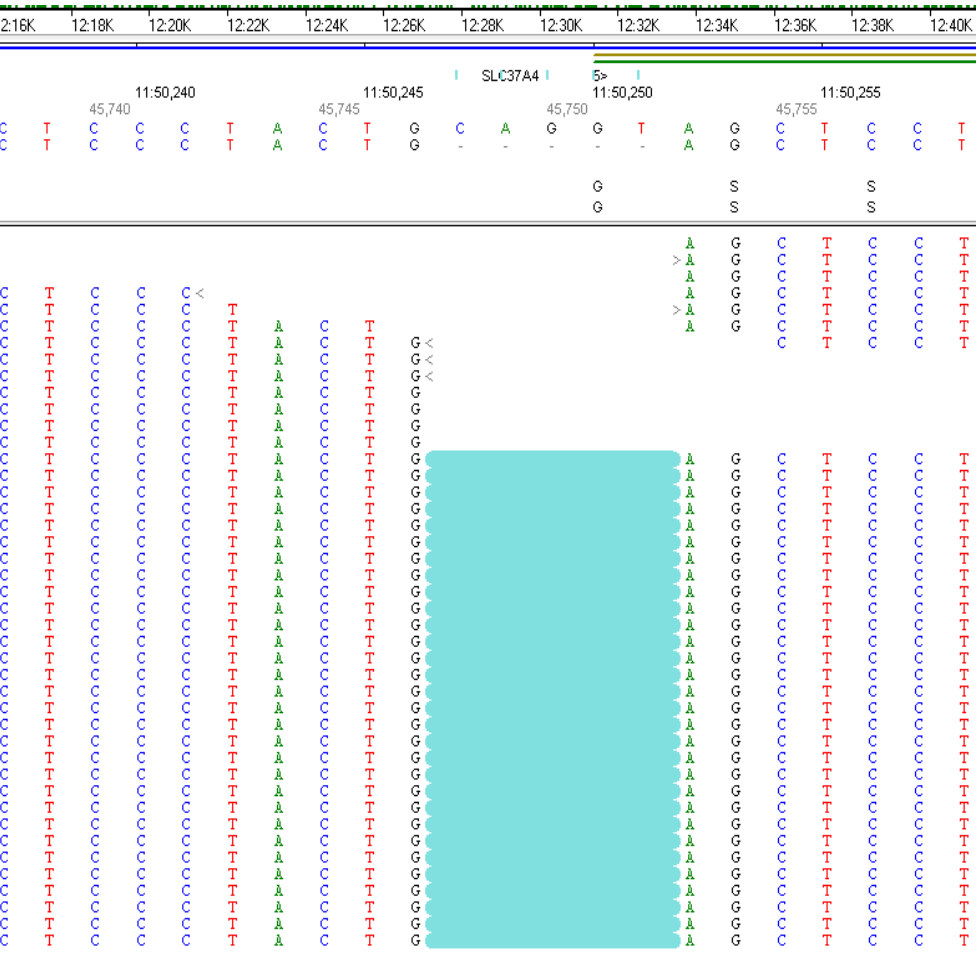
P10 SLC37A4
c.1042_1043delCT het

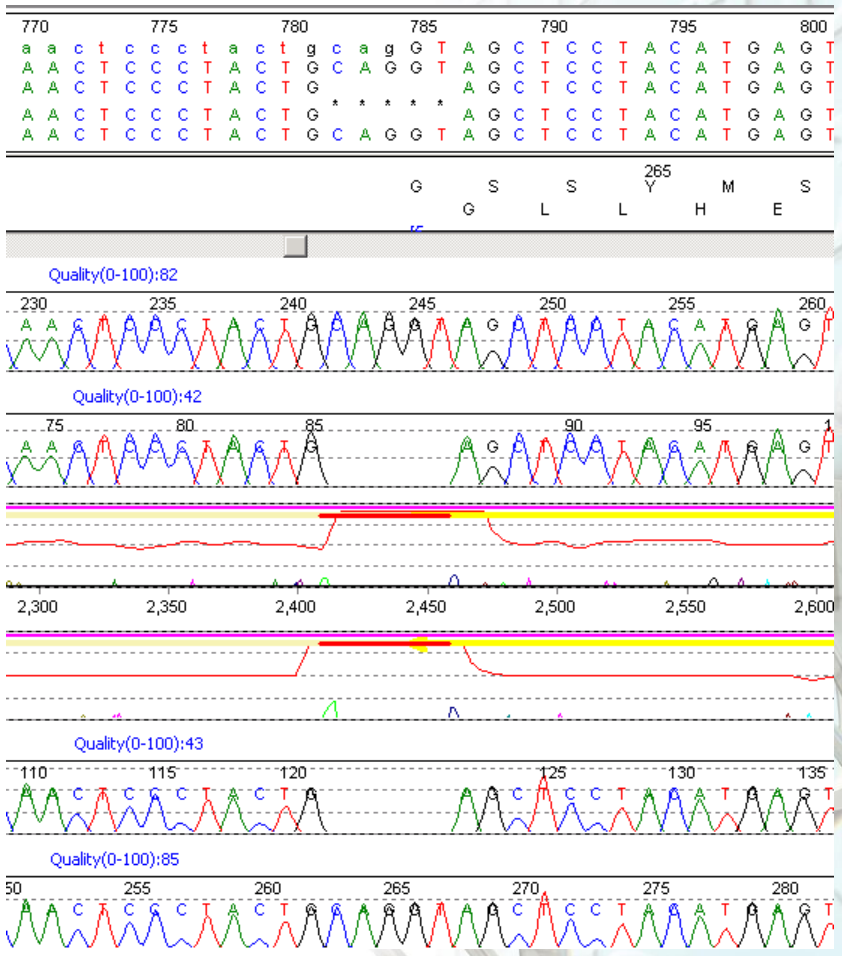


Baylor College of Medicine

MEDICAL GENETICS LABORATORIES

P11: SLC37A4
c.785-3_786del5 homozygous





GSD by panel NGS diagnostic yield: >65%

Usher syndrome

Hearing loss and retinitis pigmentosa
USH1, USH2, USH3
9 huge genes
Clinical overlap

By NGS: diagnostic yield is >83%
10/12 found 2 deleterious mutant alleles
2/12: one heterozygous allele

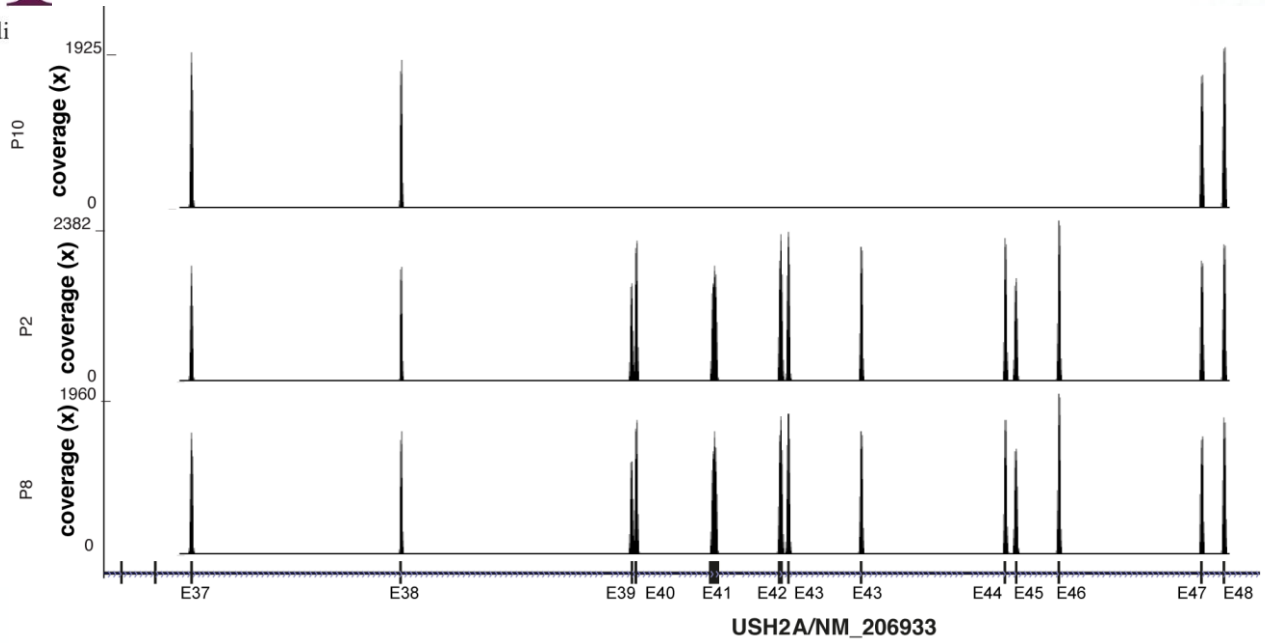
Usher syndrome

NGS Panel name	***Usher panel-2195***							
Genes included	**9 nuclear genes: CDH23, CLRN1, DFNB31, GPR98, MYO7A, PCDH15, USH1C, USH1G, USH2A**							
Number of CDS	**362CDS**							
Target size	**81,170 bp (CDS ± 20 bp)**							
Enrichment	**In solution capture library**							
Sequencing info	**Illumina HiSeq 2000, 75 cycle, single-end**							
Sample ID#	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mean coverage / base	749± 235	1564± 499	902± 297	1636± 539	1623 ± 505	1433± 515	1345± 425	1627± 536
Number of CDS < 20X	4	3	4	4	3	4	3	3

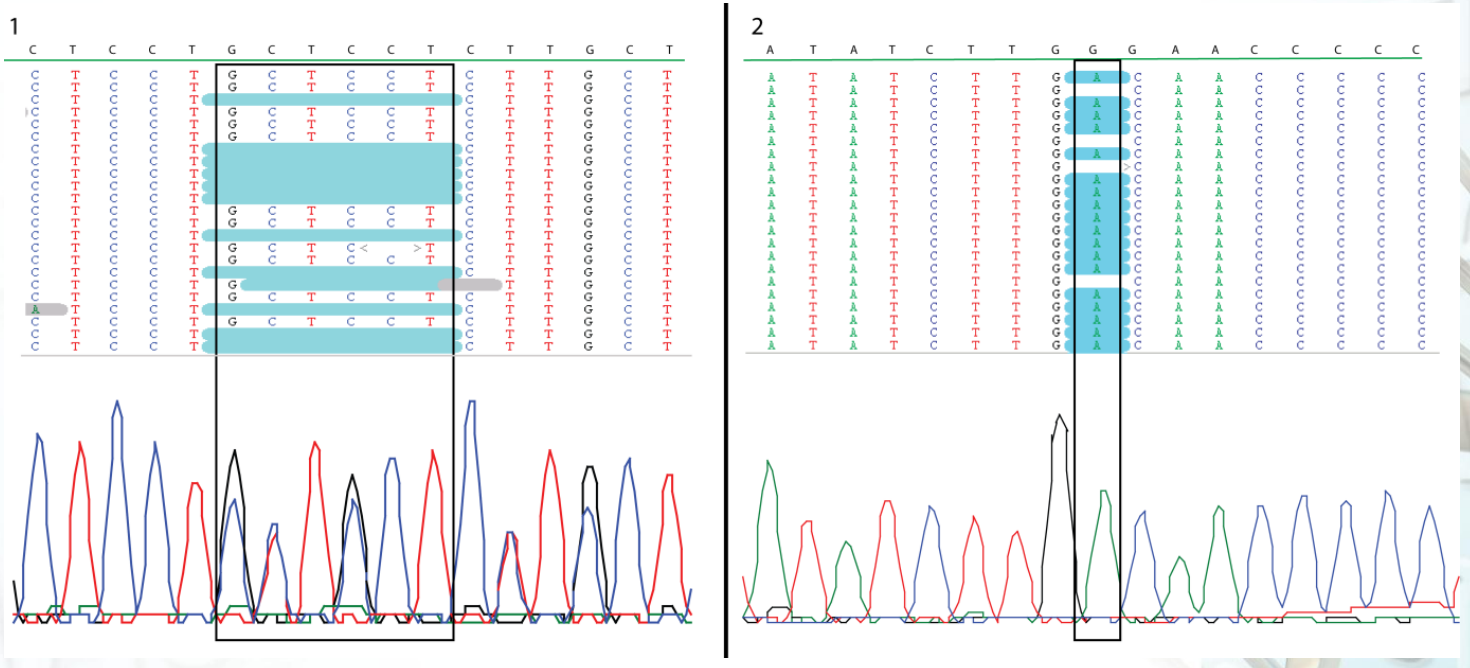
BCM
Baylor College of Medi

MGL
MEDICAL GENETICS LABORATORIES

a



b

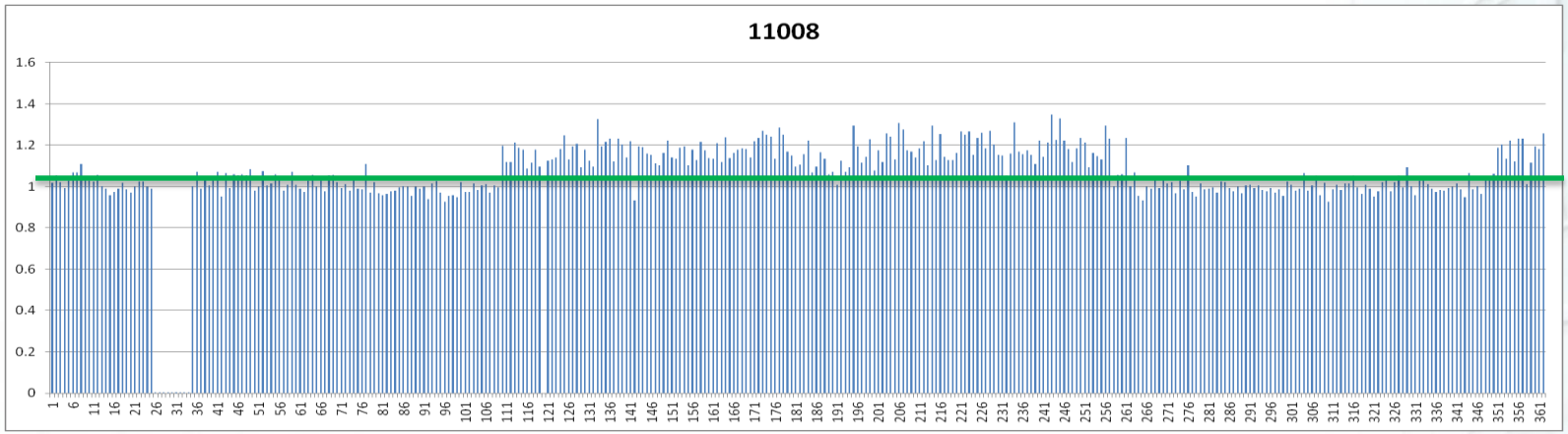


Baylor College of Medicine

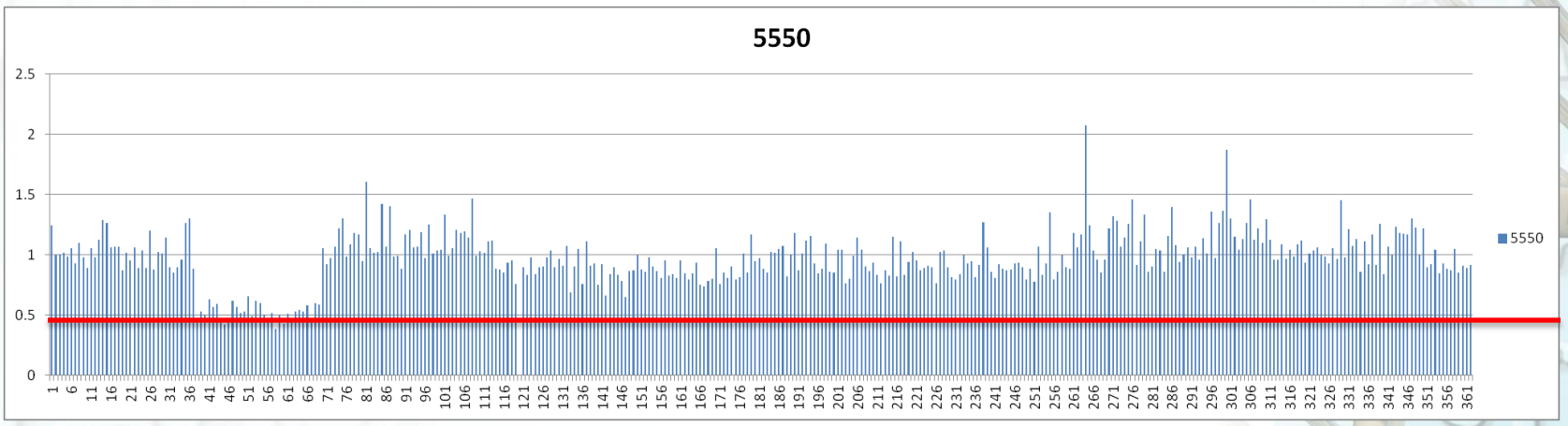
MEDICAL GENETICS LABORATORIES

NGS detects large deletions (CNV) Usher panel (previously not identified)

USH2A: CDS38-46 homozygous del



USH2A: CDS3-34 heterozygous deletion.



BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

Abnormal Bone Mass related disease

High Bone Mass Panel (14 genes)

ANKH, CA2, CLCN7, CTSK, FAM123B, FAM20C, LEMD3, OSTM1, SOST, TCIRG1, TGFB1, TNFRSF11A, TNFSF11, TYROBP

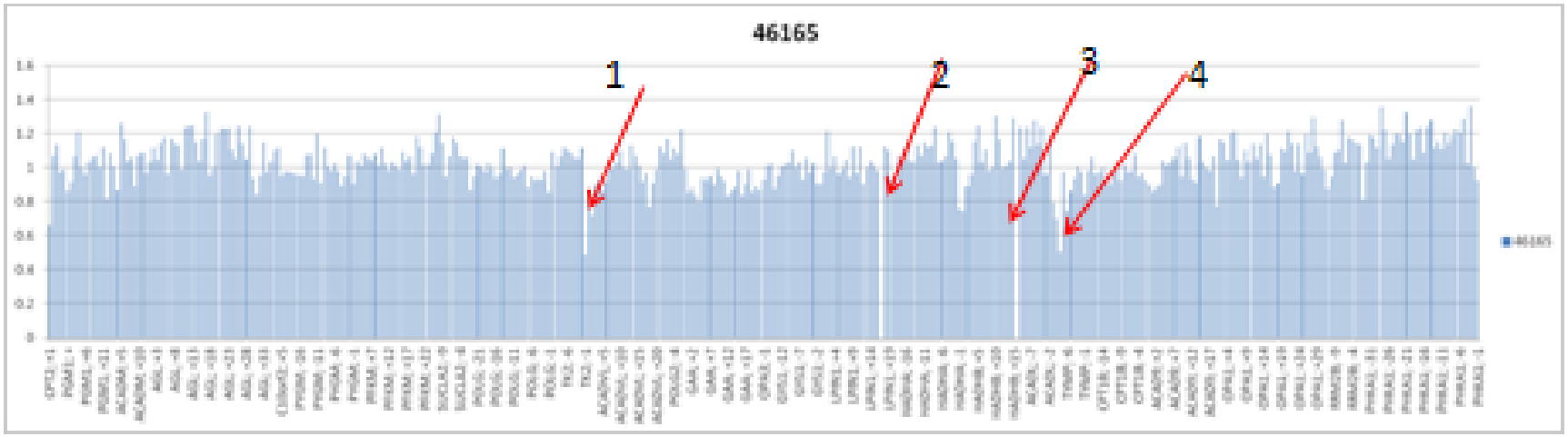
Low Bone Mass Panel (21 genes)

ALPL, B4GALT7, COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, CRTAP, FBN1, FKBP10, LEPRE1, PLOD2, PLOD3, PPIB, SERPINF1, SLC34A1, SLC39A13, SLC9A3R1, SP7, TNFRSF11A, TNFRSF11B

Abnormal Bone Mass related disease

NGS Panel name	***Skeletal panel***
Genes included	**34nuclear genes including highbonemass-panel +lowbonemass-panel**
Number of CDS	**602 CDS**
Target size	**98962bp (CDS ± 20 bp)**
Enrichment	**In solution capture library**
Sequencing info	**Illumina HiSeq 2000, 75 cycle, single-end**

	Mean coverage per base	# of Exons with one base covered <20X (Exon)	
Panel	**663x**	**20**	**3.3%**
Reduced coverage	**165x**	**134**	**22.1%**



1:TK2, E1: GC-rich. Capture not consistent between samples

2: LPIN1: E18 homozygous deletion

3: ACADL:E1: many probes but consistently among samples never been captured and sequenced to sufficient depth

4: TYMP, CDS7-9 (E8-E10): low coverage

NGS Panel testing)

Tests	number of genes	# cds	target size (bp)	cds needs PCR/Seq
GSD	16	294	50,062	0
Usher Synd	9	363	81,171	4
Bone-High Mass	14	129	27,318	13
Bone-Low Mass	20	432	67,419	6
Myopathy/ rhabdomyolysis	26	401	70,178	4
RP	66	939	202,800	16
mtDNA Depletion	14	145	26,537	4
Mitome200	162	1,788	307,144	31

BCM
Baylor College of Medicine

MGL
MEDICAL GENETICS LABORATORIES

Mitochondrial Disease: a Complex Dual Genome Disorders

Genetically and Clinically Heterogeneous

Mitochondrial Genome: 16.6 kb

- Point mutations: common, novel
- % mutant loads: heteroplasmy
- large mtDNA deletions
- copy number changes: mtDNA depletion

Nuclear Genes: ~ 1,300 genes

- Most commonly autosomal recessive
- Severe, present at early age of life
- Point mutations and large deletions
- Autosomal dominant, X-linked

Mitochondrial Disease: a Complex Dual Genome Disorders

Genetically and Clinically Heterogeneous

Nuclear Genes: ~ 1,500 genes

- specific panels:
  - depletion
  - complex subunits and assembly genes
  - aa tRNA synthetases
- Mitome200
- Mitome 500
- Mitome1500
- Exome
- Whole genome

MEDICAL GENETICS LABORATORIES

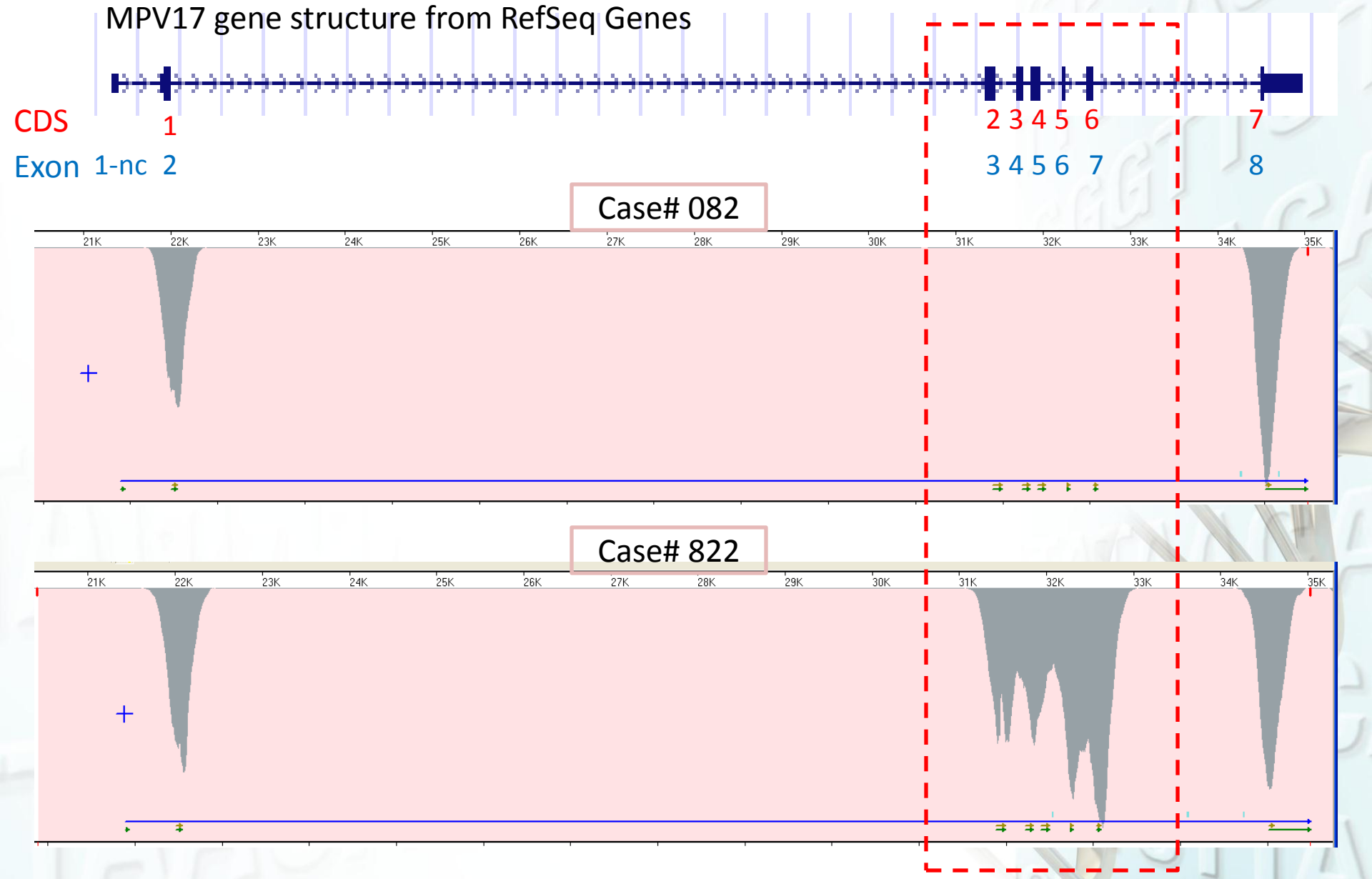
Example

- 11 month old girl
- Presented with hepatomegaly and hypoglycemia
- Previous tests revealed:
  - mtDNA depletion in liver: 9% of control
  - Whole mitochondrial genome sequencing is unremarkable

Baylor College of Medicine

NGS detects homozygous deletions: MPV17 (Exons 3-7)

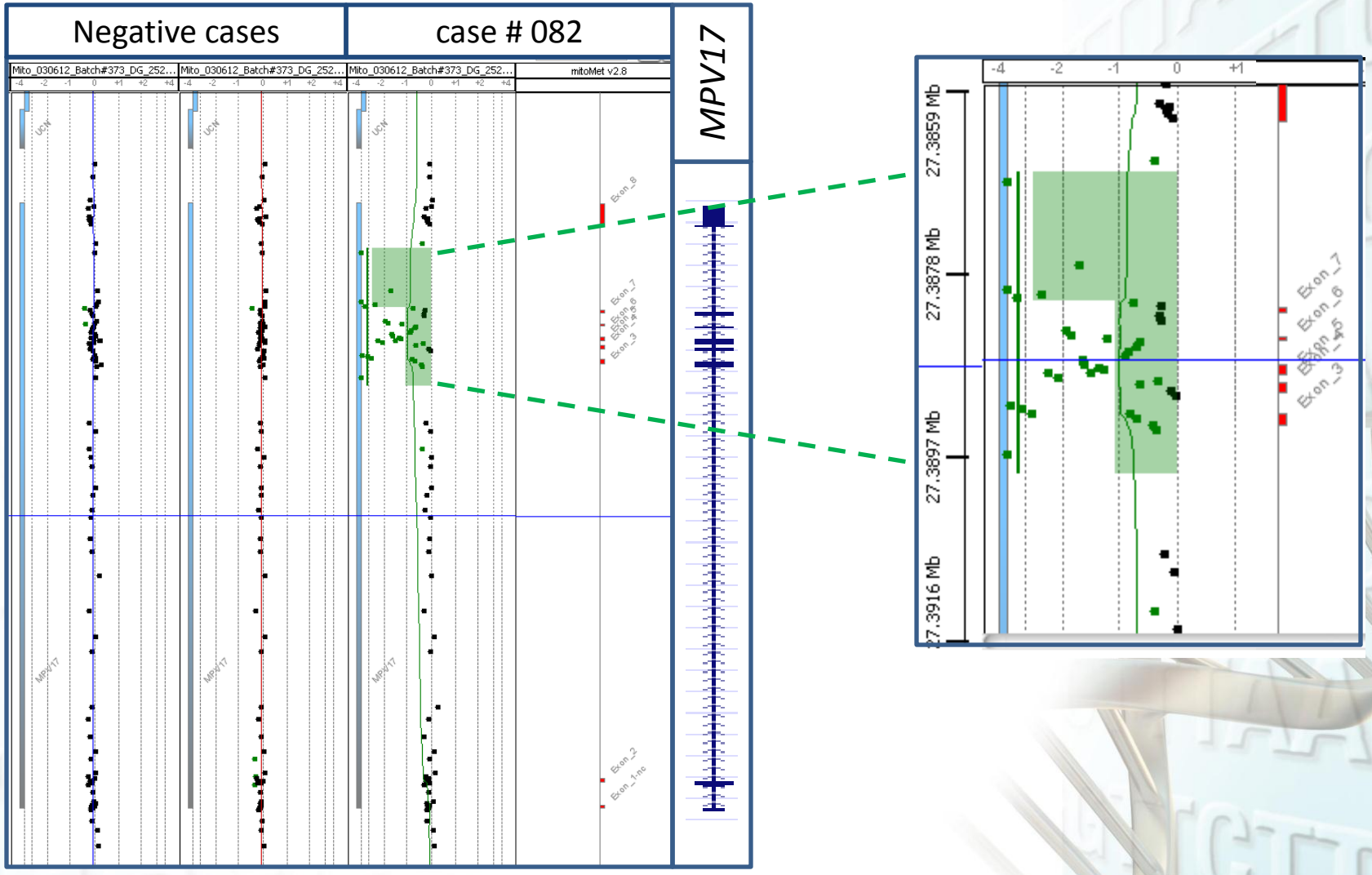
MEDICAL GENETICS LABORATORIES



Baylor College of Medicine

MEDICAL GENETICS LABORATORIES

Deletions in *MPV17* are confirmed by arrayCGH

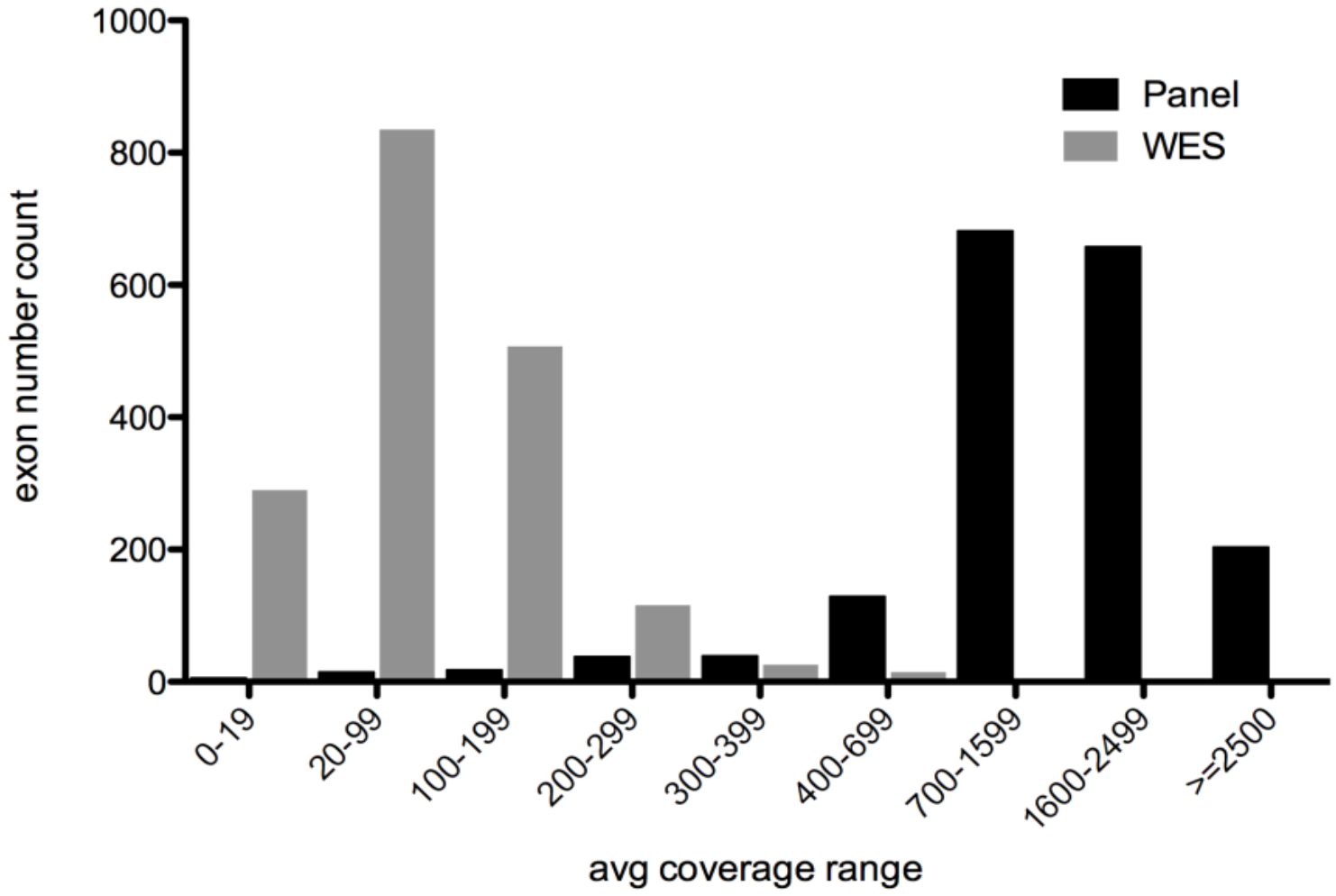


Mitome200

NGS Panel name	Mitome200
Genes included	*162 nuclear genes related to mito diseases*
Number of CDS	1,789
Target size	308,281 bp (CDS ± 20 bp)
Enrichment	In solution capture library
Sequencing info	Illumina HiSeq 2000, 75 cycle, single-end

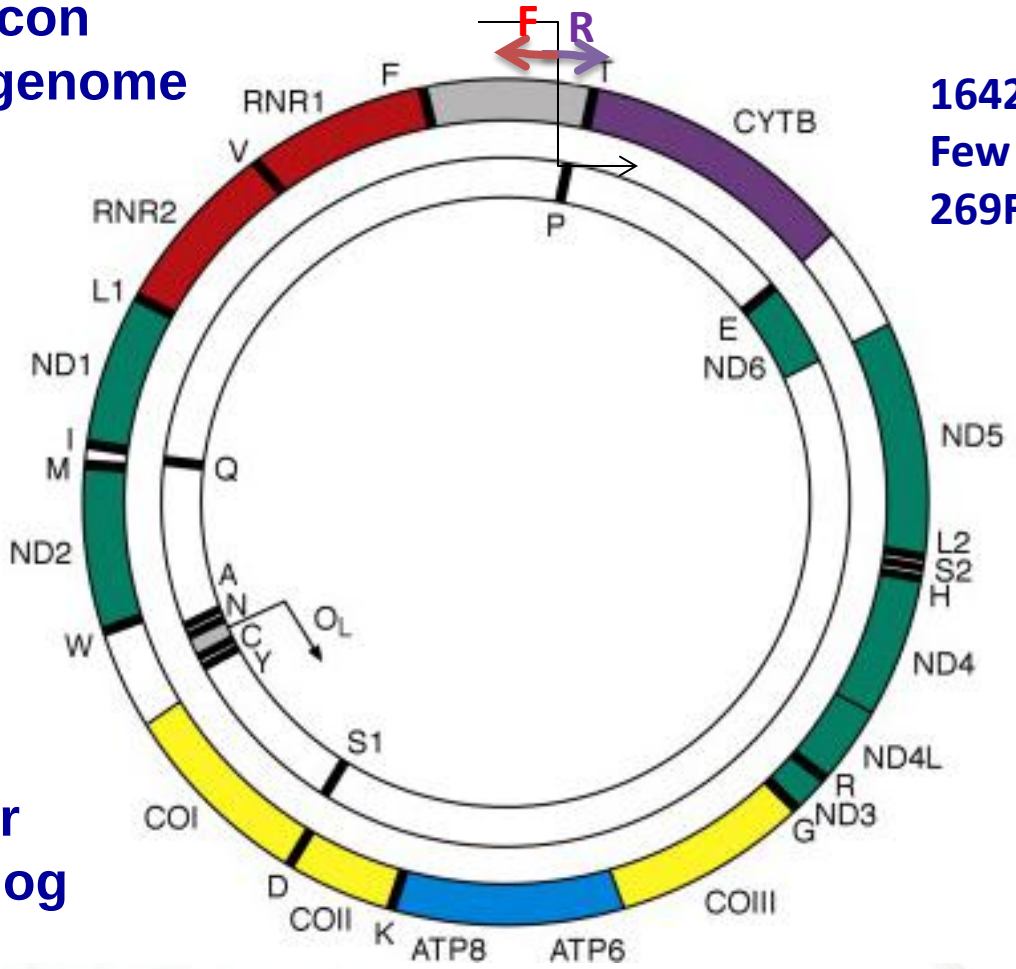
	Mean coverage per base	# of Exons with one base covered <20X (Exon)		# of bases covered <20X (Base)	
Panel	1569x	14	0.78%	1,107	0.36%
Low coverage	92x	499	27.89%	61,626	19.99%

Mitome200 vs low coverage exomes



Next Generation deep seq
1 pair of primers to avoid NUMT

Long Range PCR
Single amplicon
Whole mito genome



16426F/16425R
Few more backup sets
269F/268R

NUMT: nuclear mtDNA homolog

Sensitivity and specificity

	NGS				Sanger	NGS	
ID#	TP	FN	TN	FP	Positives	Sensitivity (%)	Specificity (%)
309	15 (1het)	0	16,554	0	14	100	100
286	15	0	16,554	0	15	100	100
964	41	0	16,528	0	41	100	100
614	45	0	16,524	0	45	100	100
798	37	0	16,531	0	37	100	100
914	16 (1het)	0	16,553	0	16	100	100
085	38	0	16,531	0	38	100	100
799	32	0	16,537	0	32	100	100
926	37	0	16,532	0	37	100	100
563	46	0	16,523	0	46	100	100
889	23 (1het)	0	16,546	0	22	100	100
820	40	0	16,529	0	40	100	100
Sum	**385**	**0**	**198,442**	**1**	**383**	**100%**	**100%**

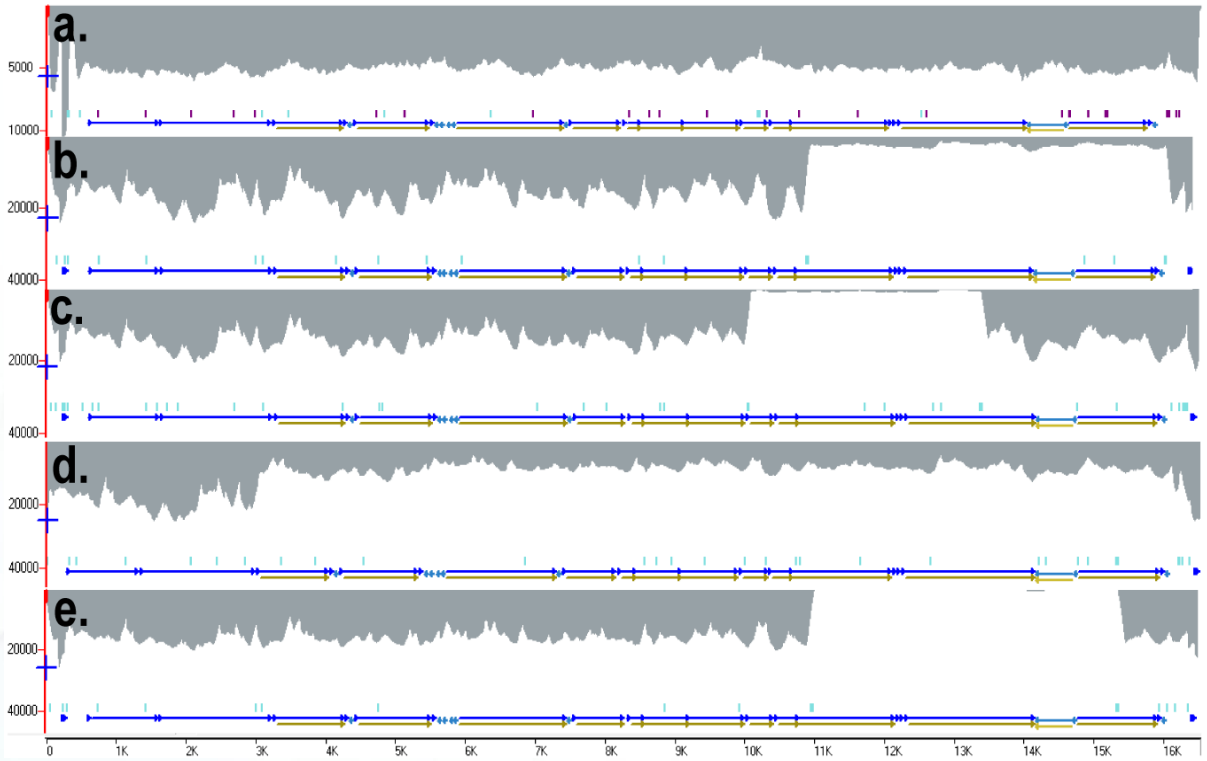
>500 samples analyzed by MPS so far

Baylor College of Medicine

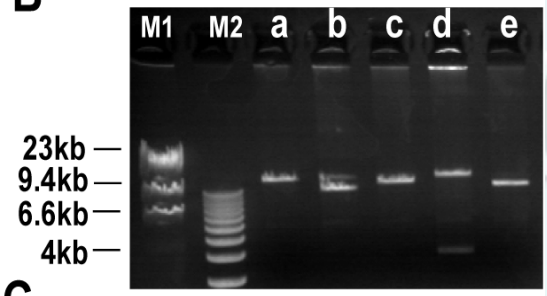
MEDICAL GENETICS LABORATORIES

Detection of mtDNA deletions by whole mtDNA amplification followed by NGS

A



B



C

Sample ID	Size	Level
a	Normal	0%
b	5,182 bp	82%
c	3,361 bp	90%
d	13,031 bp	58%
e	4,420 bp	96%

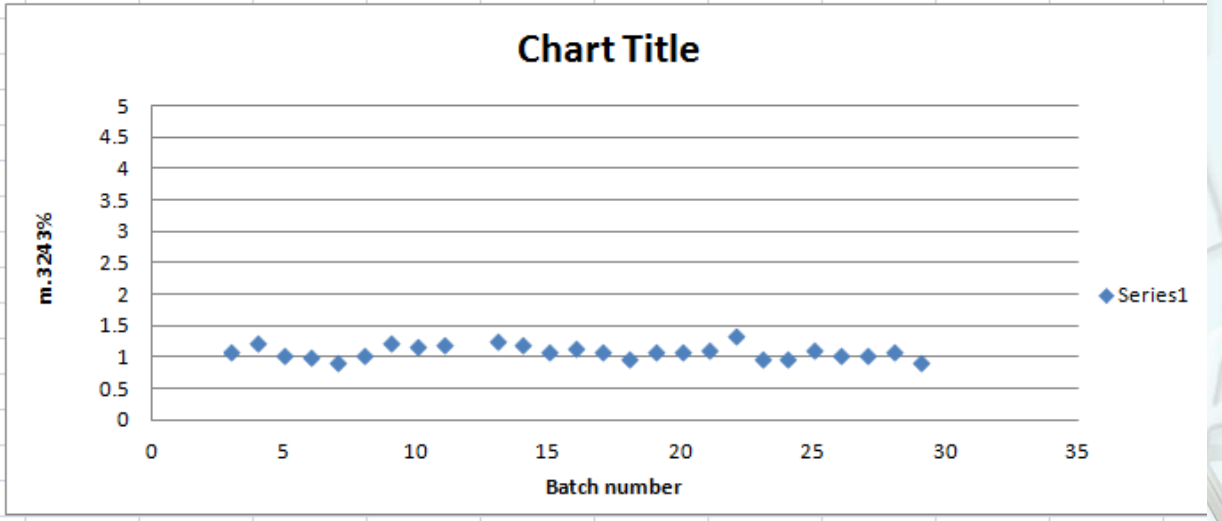
Deletions are confirmed by MitoMet array CGH and PCR/sequencing

Quantification of heteroplasmy

ID#	Position	Base change	Heteroplasmy (%)	
			NGS	qPCR
263	normal		NA	ND
062	m.3243	A>G	1.1	3
367	m.3243	A>G	2.3	8
030	m.3243	A>G	6.8	16
085	m.3243	A>G	11	32
362	m.3243	A>G	27	50
761	m.3243	A>G	36	48
074	m.3243	A>G	68	95
626	m.8344	A>G	84	73
611	m.8344	A>G	86	82
926	m.8993	T>C	88	87
799	m.10191	T>C	28	ND
994	m.11778	G>A	90	83
027	m.11778	G>A	91	91
285	m.13513	G>A	37	84
487	m.13513	G>A	54	95
563	m.14484	T>C	45	20

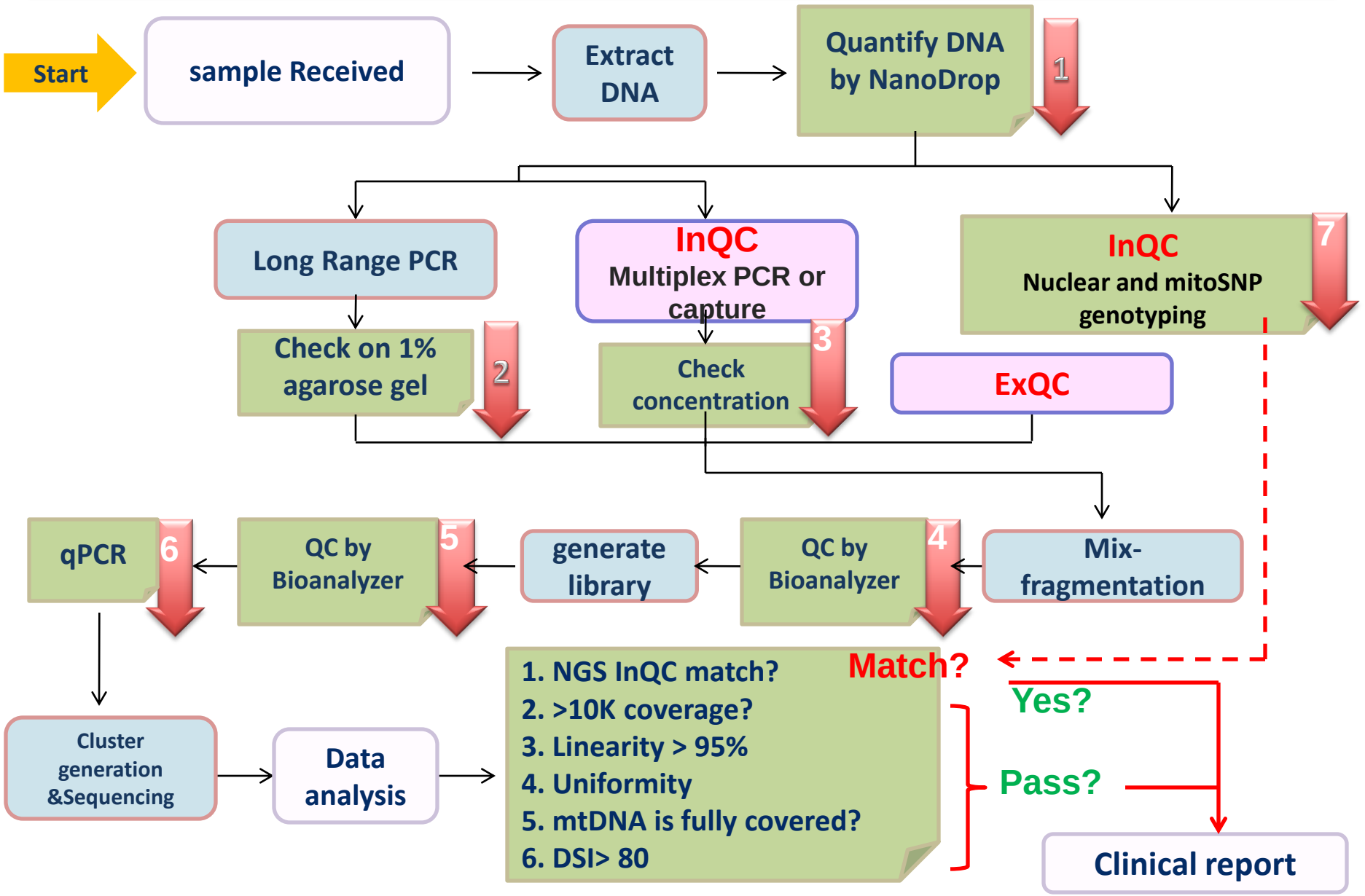
Spike-in 1.1% positive control

HiSeq run batch#	m.3243A>G%
3	1.10
4	1.25
5	1.05
6	1.03
7	0.94
8	1.05
9	1.26
10	1.20
11	1.21
13	1.27
14	1.21
15	1.10
16	1.18
17	1.10
18	1.01
19	1.12
20	1.10
21	1.14
22	1.35
23	1.01
24	1.00
25	1.15
26	1.05
27	1.04
28	1.11
29	0.95
average	1.11
STD	0.10



Average: 1.11+/- 0.1% heteroplasmy

Mito genome and Mitome NGS QA/QC workflow

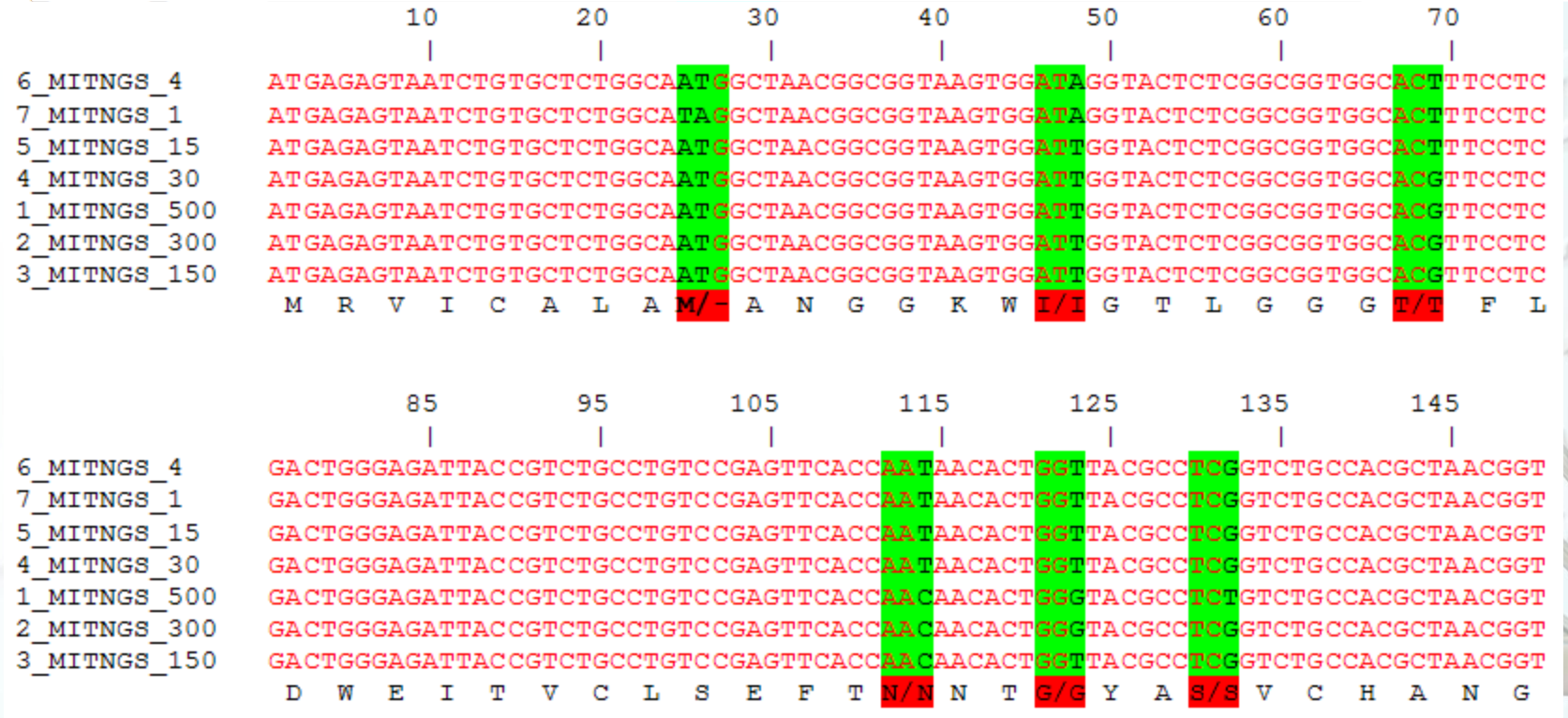


Baylor College of Medicine

MEDICAL GENETICS LABORATORIES

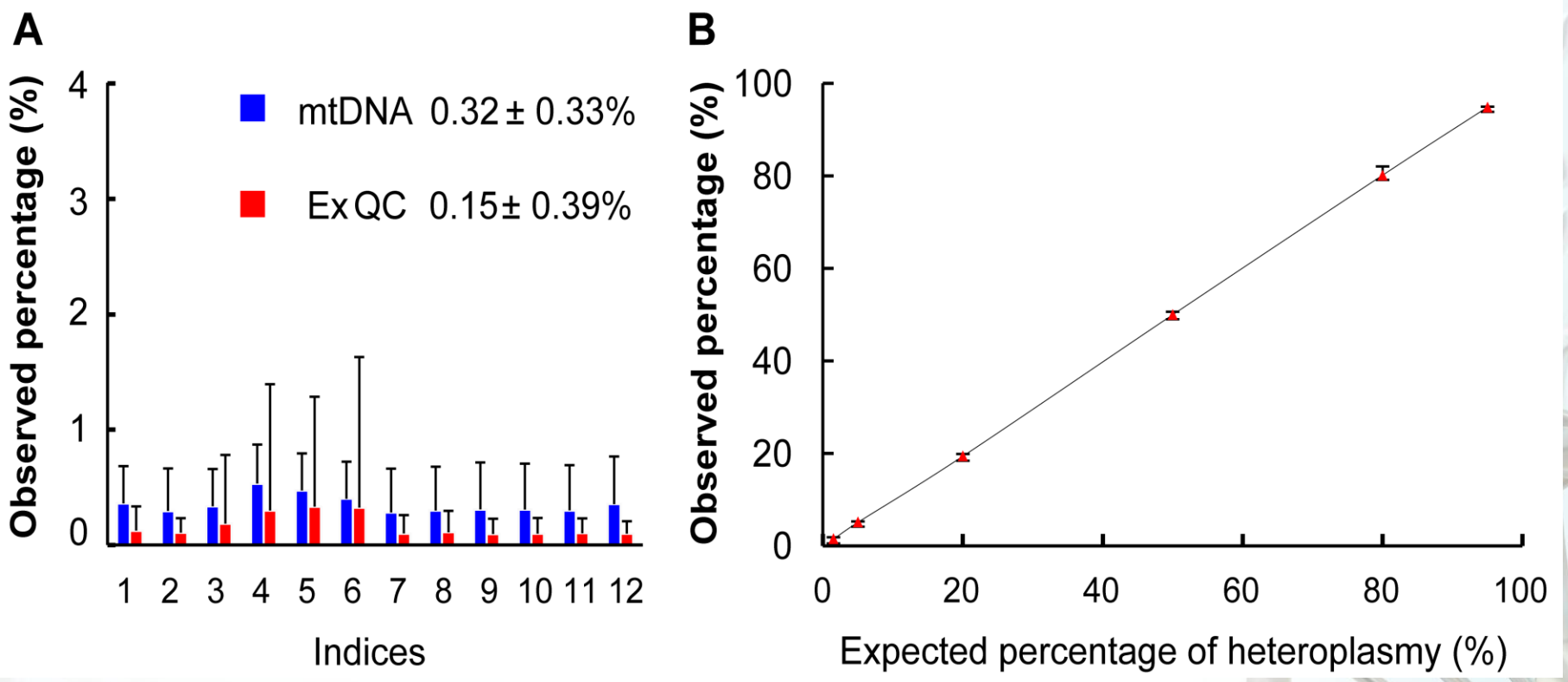
"External quality control for each indexed sample"

Spike in: 7 fragments with different codons at 6 different sites on phage DNA sequence



Spike in Quantitative control standards: for every sample

						0.10%	0.50%	2.00%	5.00%	20.00%	50.00%
X	I	T	W	G	S	0.10%	0.10%	0.10%	0.10%	0.10%	0.10%
M/X	I	T	W	G	S	0.40%	0.40%	0.40%	0.40%	0.40%	0.40%
M/X	I/I	T	W	G	S	1.50%	1.50%	1.50%	1.50%	1.50%	1.50%
M/X	I/I	T/T	W	G	S	3%	3%	3%	3%	3%	3%
M/X	I/I	T/T	W/W	G	S	15%	15%	15%	15%	15%	15%
M/X	I/I	T/T	W/W	G/G	S	30%	30%	30%	30%	30%	30%
M/X	I/I	T/T	W/W	G/G	S/S	50%	50%	50%	50%	50%	50%
						99.90%	99.50%	98.00%	95.00%	80.00%	50.00%
						M/X	I/I	T/T	W/W	G/G	S/S



Limit of detection~ 1.5%

Conclusion

1. Bringing NGS to clinical dx lab is practical. Proper QA/QC procedures should be instituted according to CLIA/CAP guidelines
2. Target gene capture/NGS: all procedures should be validated and positives confirmed
3. WES in research: discovery of new disease genes and/or new clinical phenotype
4. WES in clinical settings: currently report confirmed mutations in genes known to cause diseases (based on OMIM, HGMD db, and PubMed).
5. Novel gene/variants require functional confirmation.

Acknowledgement

Grant support: MDA

Assistant Directors

Jing Wang: variants interpretation
Victor Wei Zhang: design and analysis
Megan Louise Landsverk: interpretation
Fangyuan Li: multiple deletion mapping

Genetic counselors

Eric Schmitt
Sandra Peacock
Andrea Ybarra

Laboratory Staff

David Chen: bioinformatics
Meagan Palculict
Megan E. Cornwell
Zui Hung Ng
Avian N. Nedd
Zuzie Tien
J Michael Luchak
Michelle C Halberg
Rakhade Mrudula
Chang Jocelyn
Linh Phuong Trieu
Ramirez, Elisa M
Nguyen Christy
Gonzalez Dimas

Postdoctoral Fellows

Sha Tang
Hui Yu
Xia Tian
Hao Wang
Zhi-Yu Niu

Medical Director

William Craigen