Clinical Sequencing by Sanger: State of the Art in a Next-Gen World

Elaine Lyon, PhD
Associate Professor of Pathology
University of Utah
Medical Director of Molecular
Genetics and Genomics
ARUP Laboratories

Objectives

- Describe assay design considerations for complete coverage of regions to be interrogated
- Discuss validation approaches to establish performance characteristics and ensure test accuracy and robustness
- List challenges in and solutions for complex data analysis and interpretation
- Discuss workflow measures for implementing efficient Sanger sequencing assays into the clinical laboratory

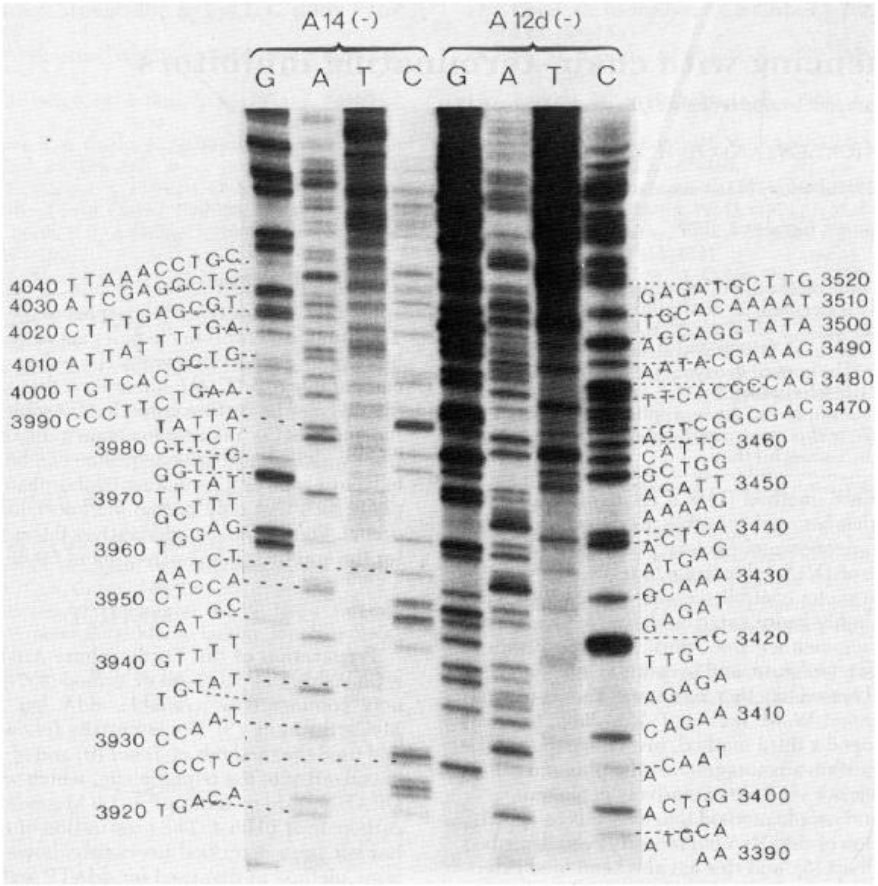
Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕX174)

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

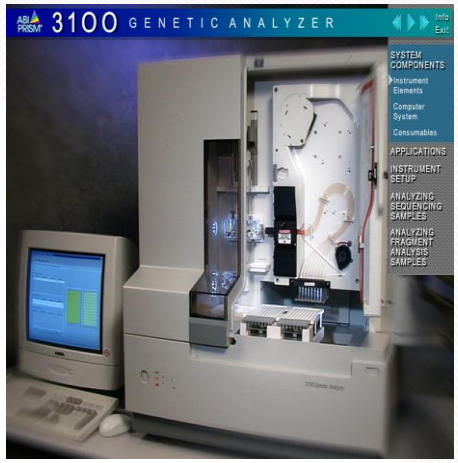
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

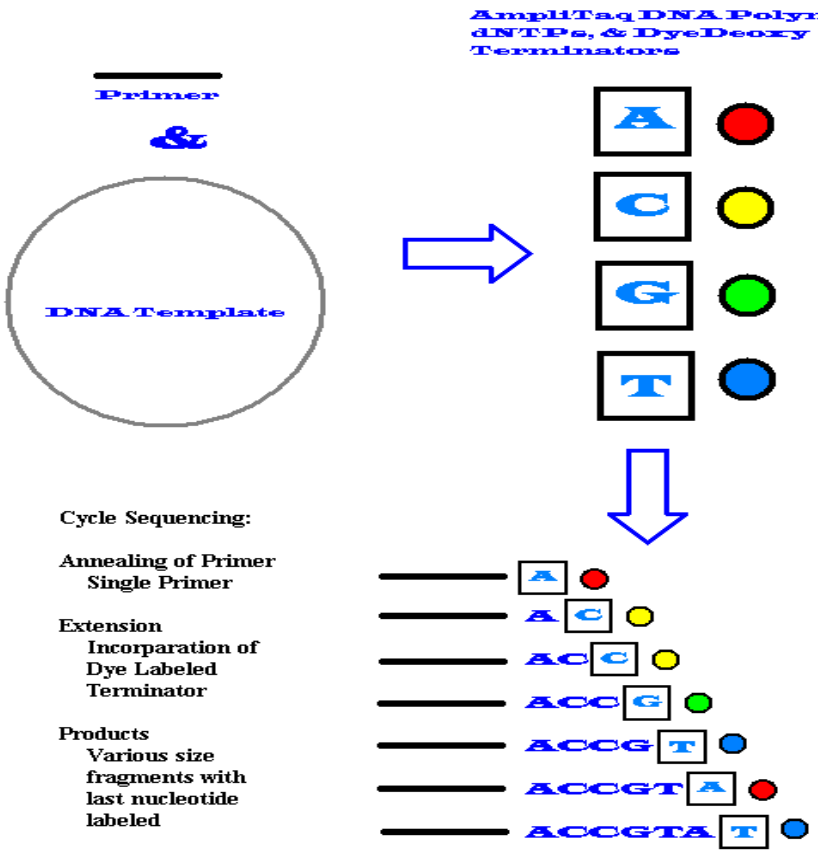


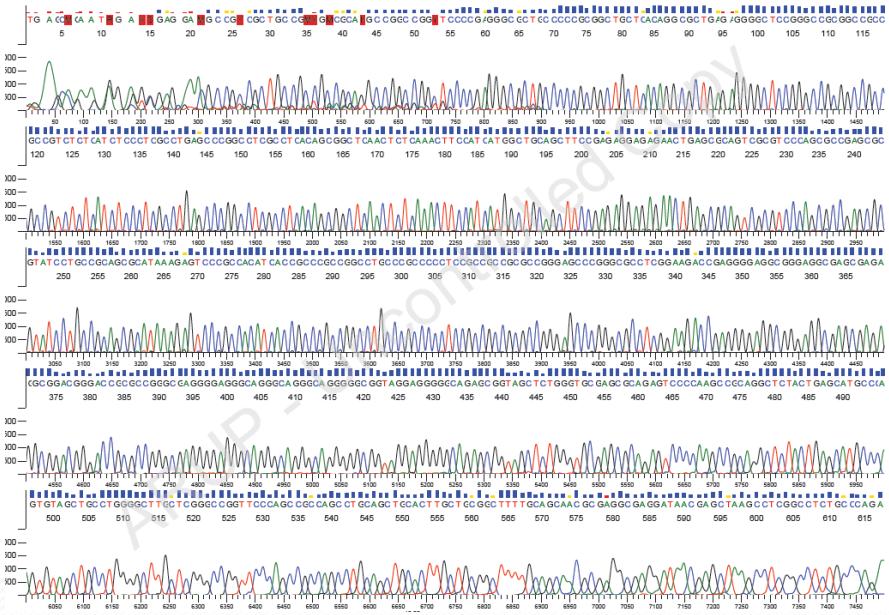
1977

Courtesy of Karl Voelkerding

1986- ABI Sequencing (Sanger with Fluorescent Terminator)







Clinical Sequencing Assays

- Analytical Validation
  - Familiarity
  - Design
  - Optimization
  - Accuracy
  - Robustness (reproducibility)
  - Interpretation
- Clinical validation
  - Clinical sensitivity
  - Clinical specificity

Familiarization and Planning

- Reference sequence

Gene	GBK file (analysis)	GBK file (reporting)	Mutation database numbering differences	MLPA exon numbering differnences	GVIE - ARUP Wiki	CDS	Inheritance
PTEN	NC_000010.10	NM_000314.4	None	NO		No	A.dominant

- Alternative transcripts
- Homology checks
  - pseudogenes
- Inheritance
- Databases
  - Locus specific
- Known benign variants

Regions Interrogated

- Targeted exons
  - Example: MEN2
- All coding exons
  - 'Full gene or full sequence analysis'
- Intron/exon boundaries
  - +20-+50
- Known deep intronic mutations
- Regulatory regions
  - 5' UTR, promoter
  - 3' UTR

Primer Design

- Often per exon
- Design around pseudogenes
- Avoid known variants
  - Interfere with PCR
- All at same PCR conditions?

PTEN Amplicon Sizes

Amplicon	Amplicon size (bp)
Exon 1	252
Exon 2	296
~~Exon 3~~	~~220~~
Exon 3 new	400
Exon 4	234
Exon 5 short	342
Exon 5 long	396
Exon 6	383
~~Exon 7~~	~~355~~
Exon 7 new	356
Exon 8 short	299
Exon 8 long	497
Exon 9	322
Promoter	697

PTEN PCR and Sequencing Primers:

Exon	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
1	TGTAAAACGACGGCCAGTCATTTCCATCCTGCAGAAGA	CAGGAAACAGCTATGACCCAAACTACGGACATTTTCGC
2	TGTAAAACGACGGCCAGTACCTTTTATTACTCCAGCTATAGTG	CAGGAAACAGCTATGACCTGTGGCTTAGAAATCTTTTCTAAATG
~~3~~	~~TGTAAAACGACGGCCAGTTCTGTCTTTTGGTTTTTCTTGATAGTA~~	~~CAGGAAACAGCTATGACCCTCTAAGAAGCAGATAACTTTGAC~~
3 new	TGTAAAACGACGGCCAGTGGGGTATTTGTTGGATTATTTATTT	CAGGAAACAGCTATGACCCCAACAATGTTTTACCTCATCA
4	TGTAAAACGACGGCCAGTAAGATTCAGGCAATGTTTGTTAGT	CAGGAAACAGCTATGACCTCTCACTCGATAATCTGGATGAC
5s	TGTAAAACGACGGCCAGTTTATTCTGAGGTTATCTTTTTACCACA	CAGGAAACAGCTATGACCGAAACCCAAAATCTGTTTTCCA
5l	TGTAAAACGACGGCCAGTCTGTTAAGTTTGTATGCAACATTTCT	CAGGAAACAGCTATGACCTTTCCAATAAATTCTCAGATCCAG
6	TGTAAAACGACGGCCAGTAAATGGCTACGACCCAGTTAC	CAGGAAACAGCTATGACCTAATTTGTTCAAATGCTTCAGAAA
7	~~TGTAAAACGACGGCCAGTATTGGTGATATTAATCATTAAAATCGT~~	~~CAGGAAACAGCTATGACCAACAAATTATAGTTCGTTACATGTCA~~
7 new	TGTAAAACGACGGCCAGTAAAATCGTTTTTGACAGTTTGACA	CAGGAAACAGCTATGACCCACCTGCAGATCTAATAGAAAACA
8s	TGTAAAACGACGGCCAGTCAAAATGTTTCACTTTTGGGTAAA	CAGGAAACAGCTATGACCGCTGTACTCCTAGAATTAAACACACA
8l	TGTAAAACGACGGCCAGTTGCCTTATAATAGTCTTTGTGTTTACC	CAGGAAACAGCTATGACCGTCAAGCAAGTTCTTCATCAGC
9	TGTAAAACGACGGCCAGTTGTTCATCTGCAAAATGGAAT	CAGGAAACAGCTATGACCTGGTGTTTTATCCCTCTTGATAAAAA
promoter	TGTAAAACGACGGCCAGTCCATCTCAGCTTTCATCATCAG	CAGGAAACAGCTATGACCCGGTTAGAAAAGACGAAGAGGA

Notes on Primer Design:

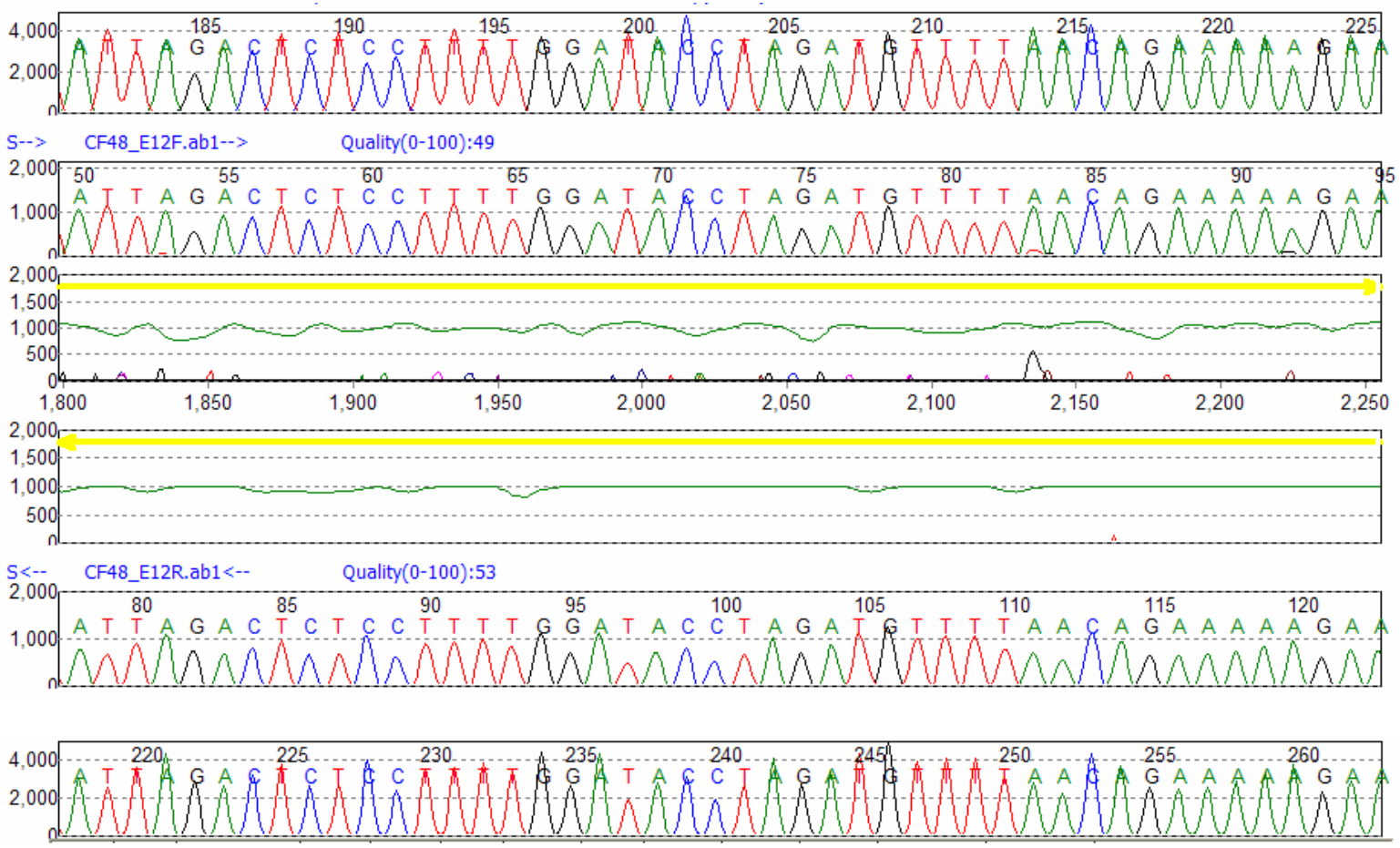
Sanger Sequencing Alignment Using Mutation Surveyor Software

Reference seq

Forward

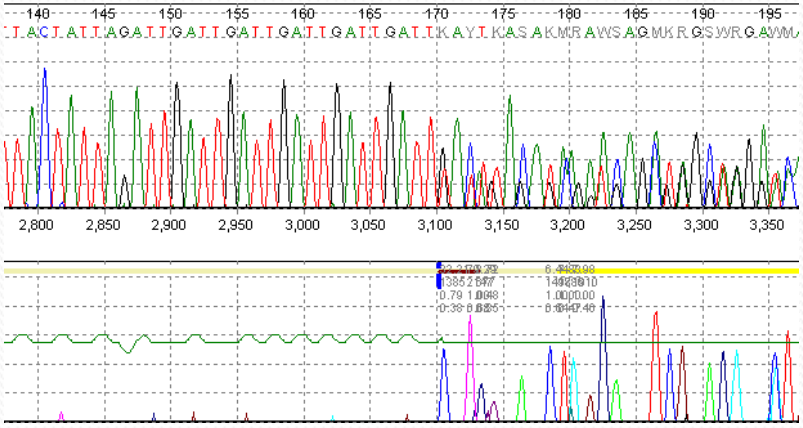
Reverse

Reference seq



Difficult Regions

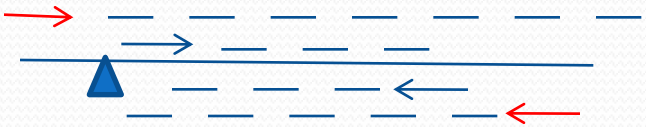
- High GC content
  - Optimization
- Secondary structure
  - Optimization or avoidance
- Benign Insertions/deletions
  - Example: CFTR GATT
- Pseudogenes
  - Example: PMS2
- Repeat motifs
  - Example: CFTR intron 8 TG/T
  - Example: Homopolymers



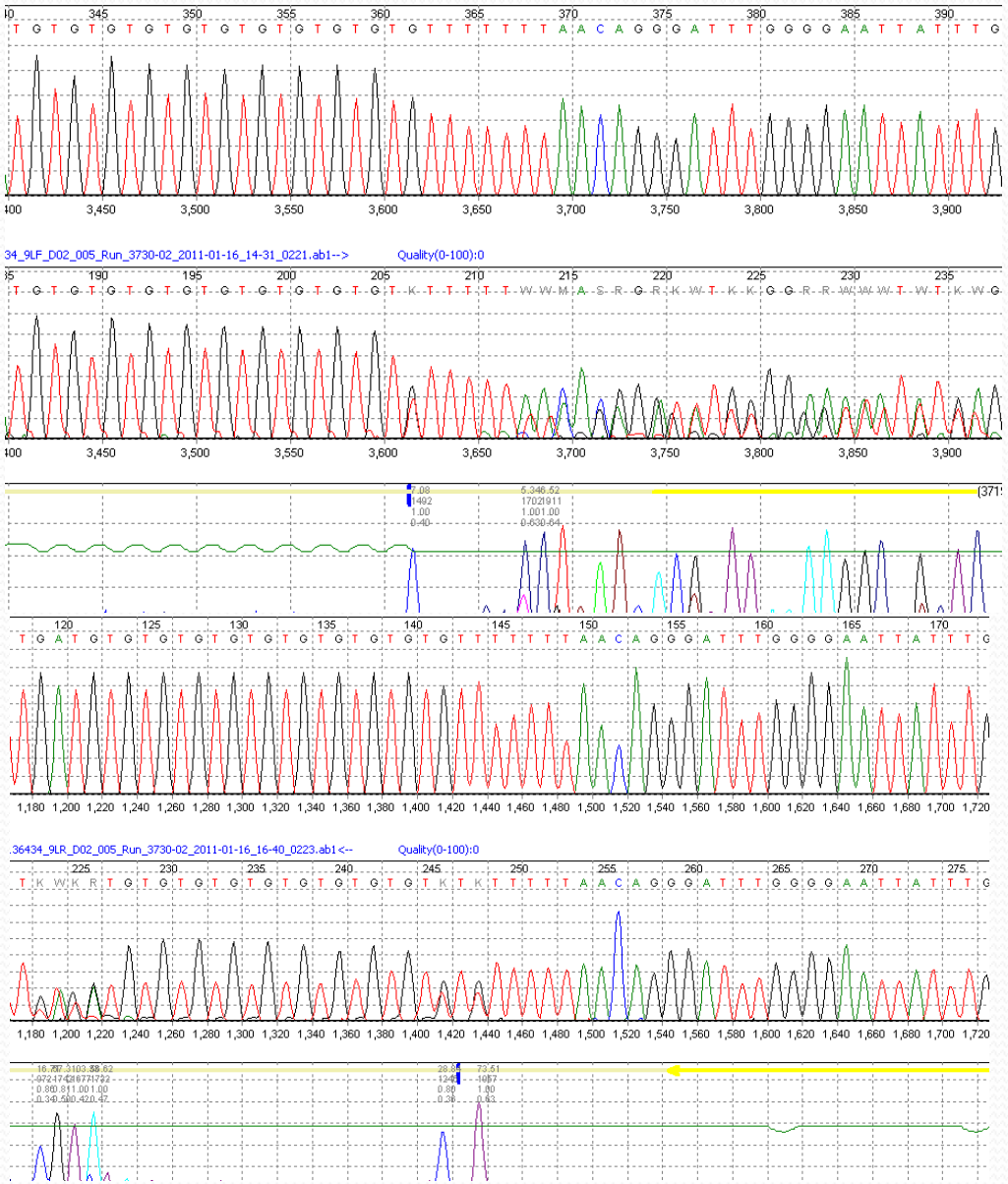
CFTR GATT

Primer Design

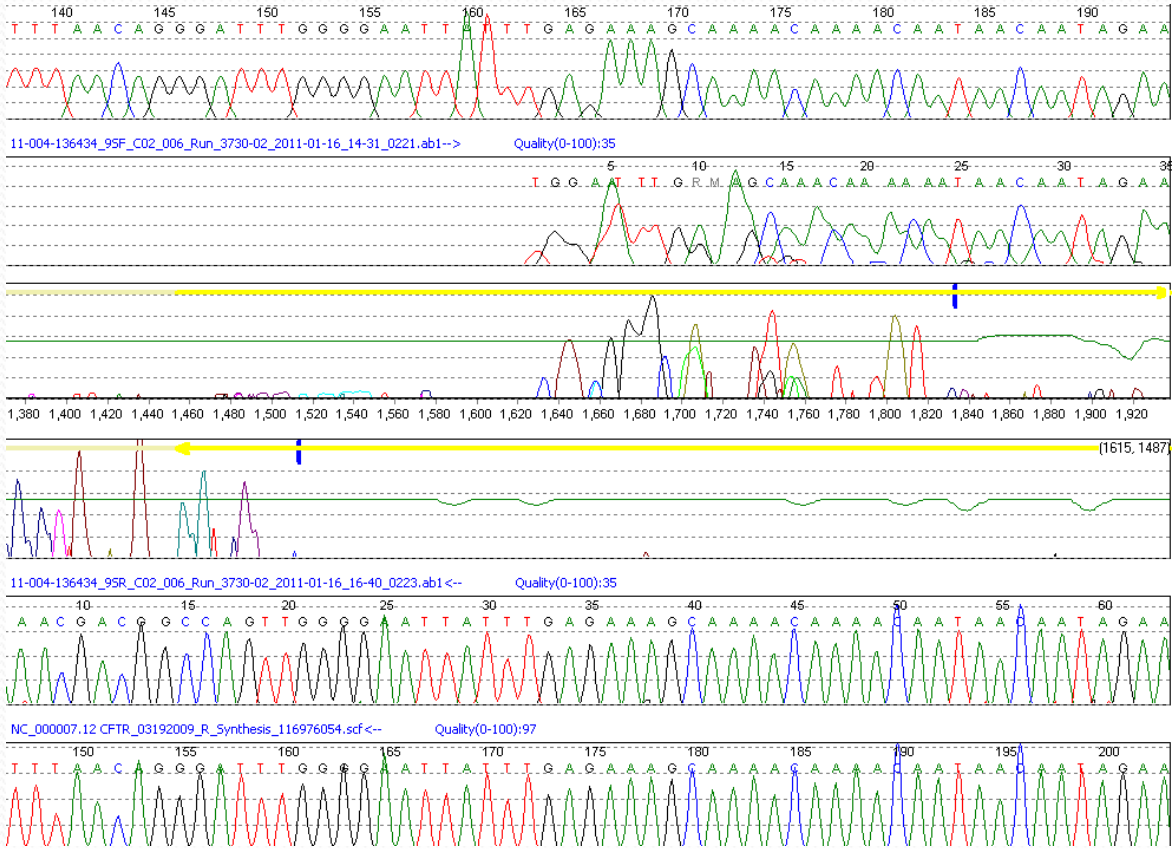
- Design Long and Short amplicons
- Cover all regions



CFTR intron 8
TG/T region
F and R primers for Long amplicon



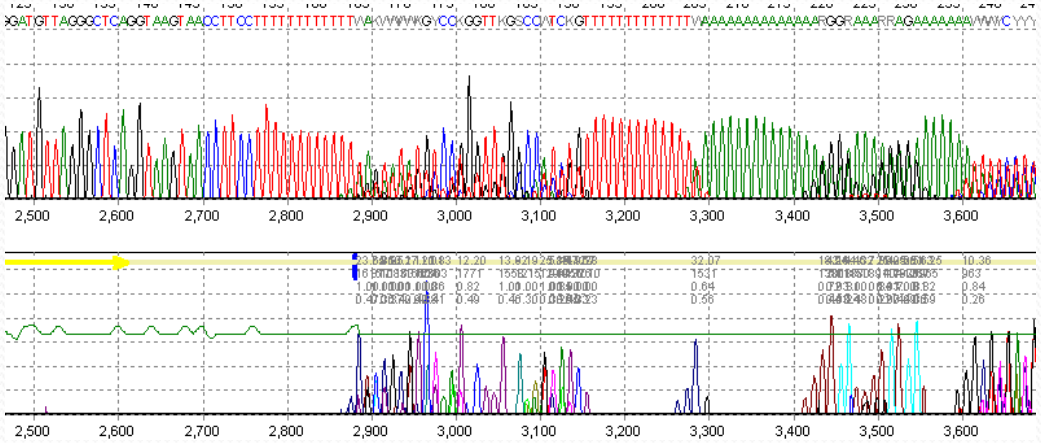
Primer Design

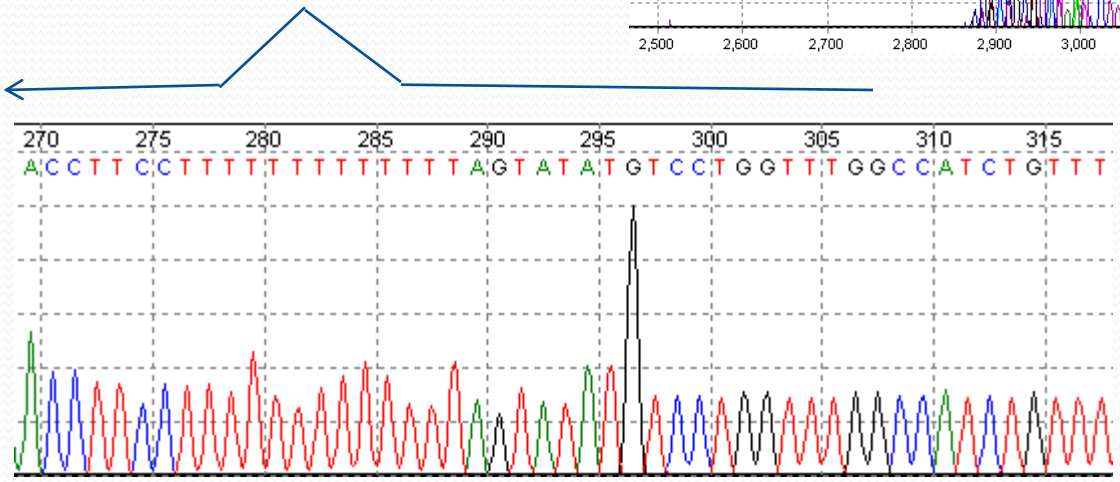


CFTR intron 8
TG/T region
F and R primers for Short amplicon

Primer Design

- Loop-out/masking





Analytical Validation

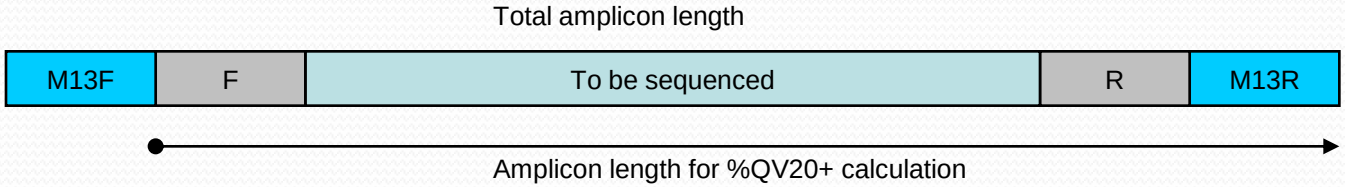
- Performance characteristics
  - *Accuracy

Results

PTEN Exon	Clinically Diagnosed Sample 1	Clinically Diagnosed Sample 2	Accuracy 1 (Cftr33)	Accuracy 2 (M3 DNA)	Accuracy 3 (CF16) 11.21.08	Accuracy 4 (CFTR41) 12.3.08	Accuracy 5 (11F)
1	wt	wt	wt	wt	wt	wt	wt
2	wt	wt	wt	wt	wt	wt	wt
3	wt	wt	wt	wt	wt	wt	wt
4	wt	Splice site mutation c.253+1G>GC	wt	wt	wt	wt	wt
5	c.1420C>CT	wt	wt	wt	wt	wt	wt
6	wt	wt	wt	wt	wt	wt	wt
7	wt	wt	wt	wt	wt	wt	wt
8	wt	wt	wt	wt	wt	wt	wt
9	wt	wt	wt	wt	wt	wt	wt
promoter	wt	c.1-1085C>CT	wt	wt	wt	wt	wt
Intron	IVS1= c.80-96A>AG IVS8= c.1026+32T>TG	IVS1= c.80-96A>AG IVS4= c.253+1G>GC IVS8= c.1026+32T>TG	IVS1= c.80-96A>AG IVS8= c.1026+32T>TG	wt	IVS1= c.80-96A>AG IVS8= c.1026+32T>TG	wt	wt

Quality checks

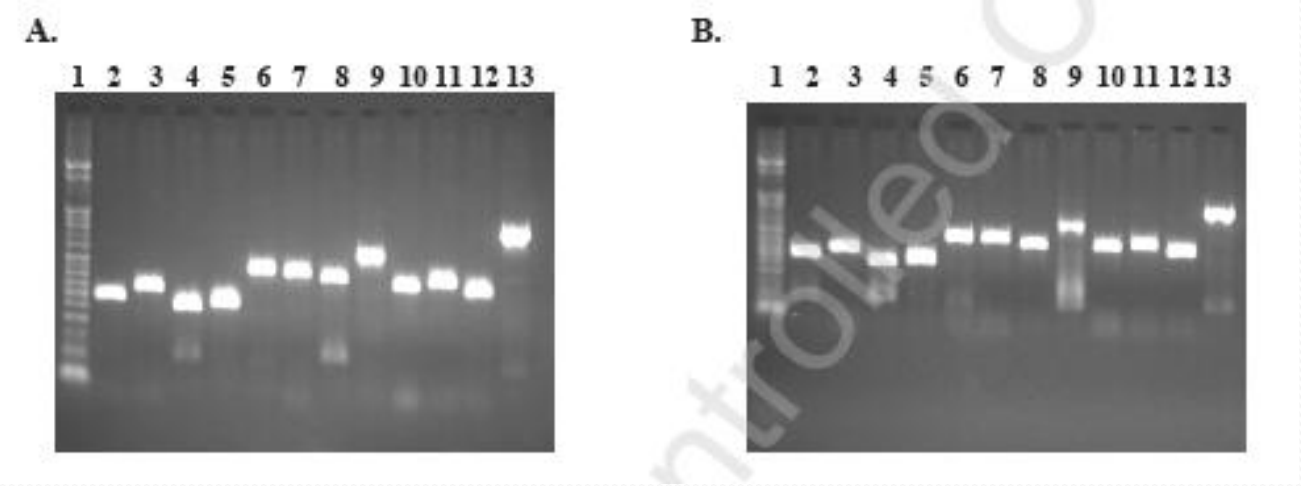
- Trace scores: average quality score
- Signal intensity
- Signal to noise ratio
- %QV20+ :percentage of bases with quality values ≥ 20



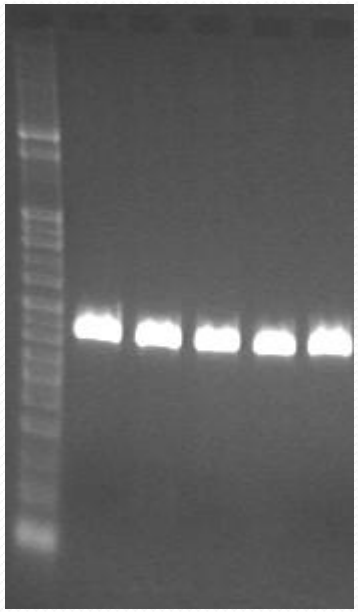
Sequencing Results for *PTEN*

EXON 1	TS	QV20+	AL (in bp)	%QV20	S/N (A)	S/N (C)	S/N (G)	S/N (T)	SI (A)	SI (C)	SI (G)	SI (T)
sample1_Acc_ex1F	43	227	270	84.074	625	308	423	275	4031	1997	2396	2284
sample1_Acc_ex1R	46	238	270	88.148	1643	801	1514	878	10706	5418	8925	7372
sample1_Within_ex1F	39	215	270	79.63	817	338	537	328	5523	2268	3047	2761
sample1_Within_ex1R	46	234	270	86.667	1641	822	1789	936	11283	5418	10631	7470
sample1_Betw_ex1F	39	221	270	81.852	347	159	233	175	3349	1330	1697	1674
sample1_Betw_ex1R	46	236	270	87.407	1304	612	1173	684	9187	4062	7181	5642
Average	43.2	228.5	270	84.63	1062.8	506.667	944.83	546	7346.5	3415.5	5646.2	4534
Standard Deviation	3.43	9.1378	0	3.3644	546.55	277.831	637.8	328.63	3477.6	1795.3	3764.8	2619

Reproducibility – PCR product

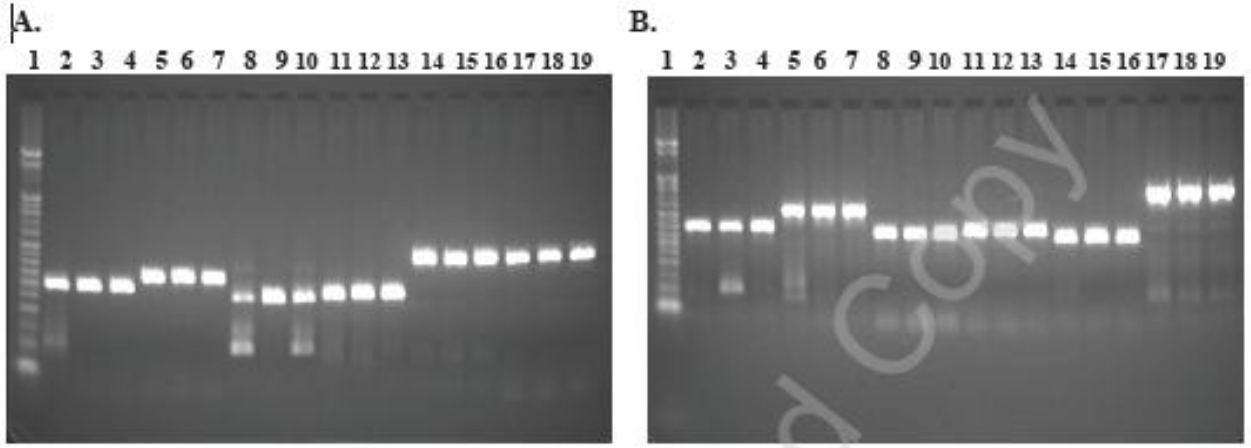


*Intra-run variability



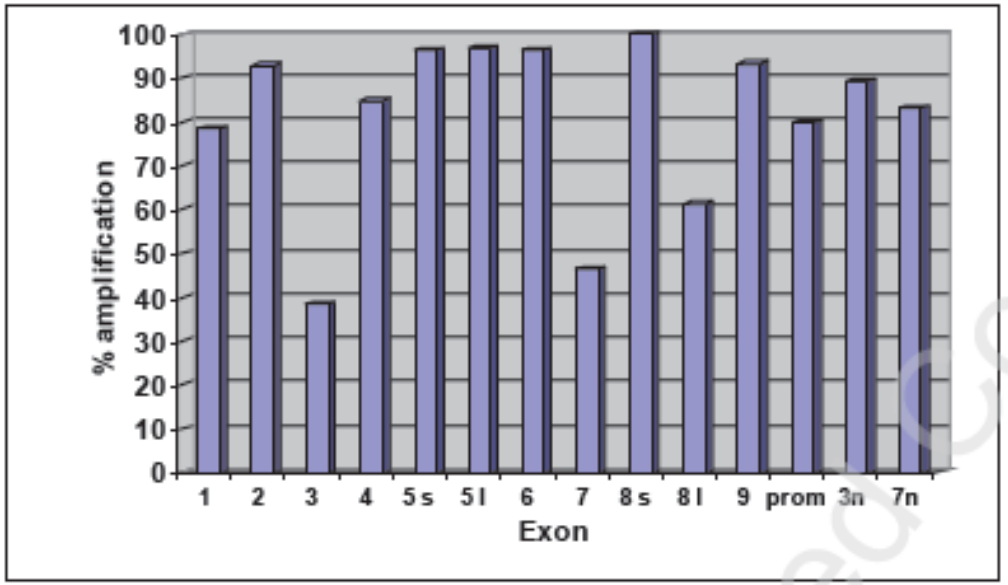
Re-design of exon 3

Reproducibility – PCR products

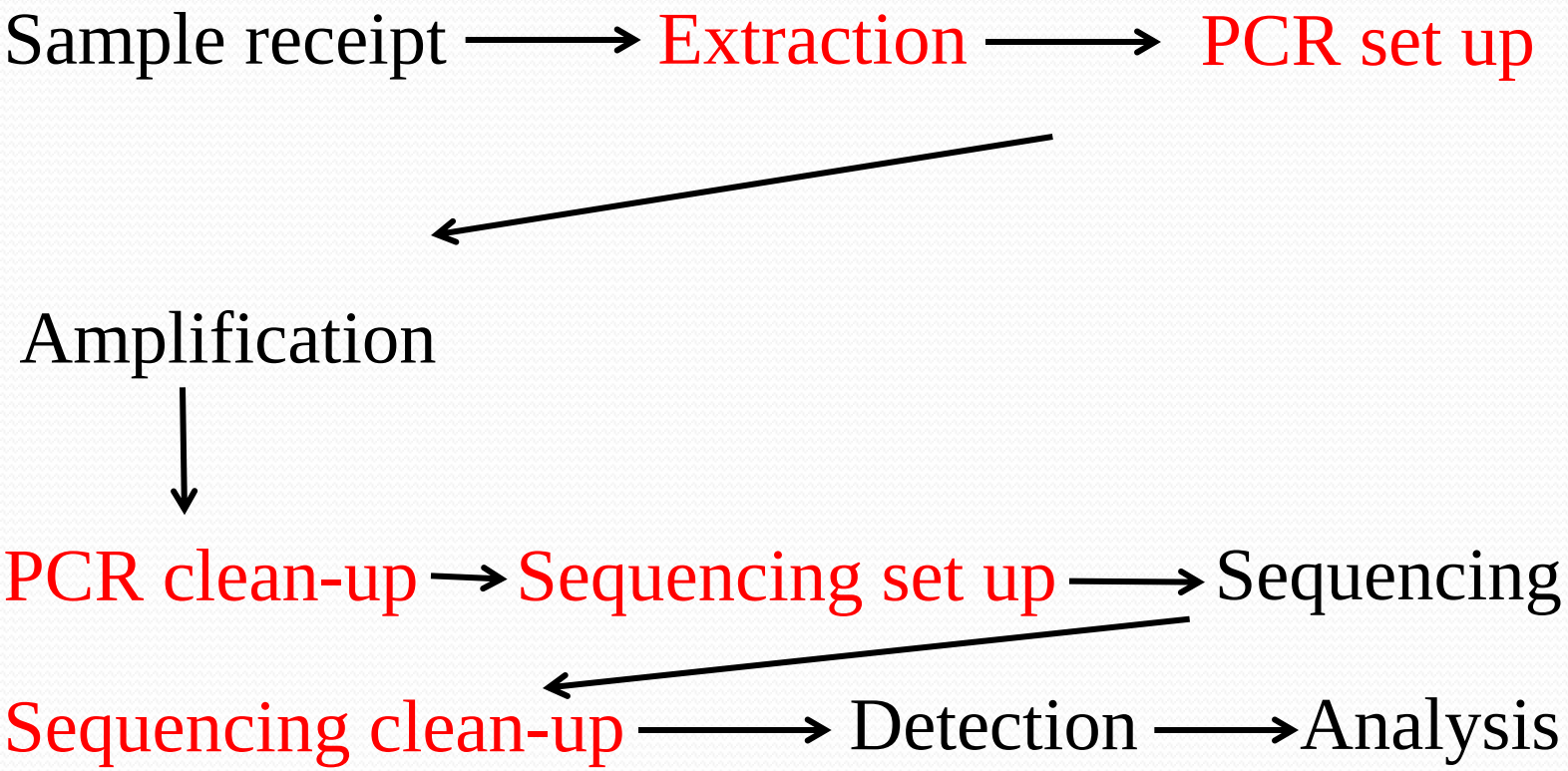


*Inter-run

All reactions



Workflow


Sample receipt
Extraction
PCR set up
Amplification
PCR clean-up
Sequencing set up
Sequencing
Sequencing clean-up
Detection
Analysis

Workflow

- M13 tagged primers
- Workflow
  - Low throughput – per sample
  - High throughput – per exon
- Primer plate

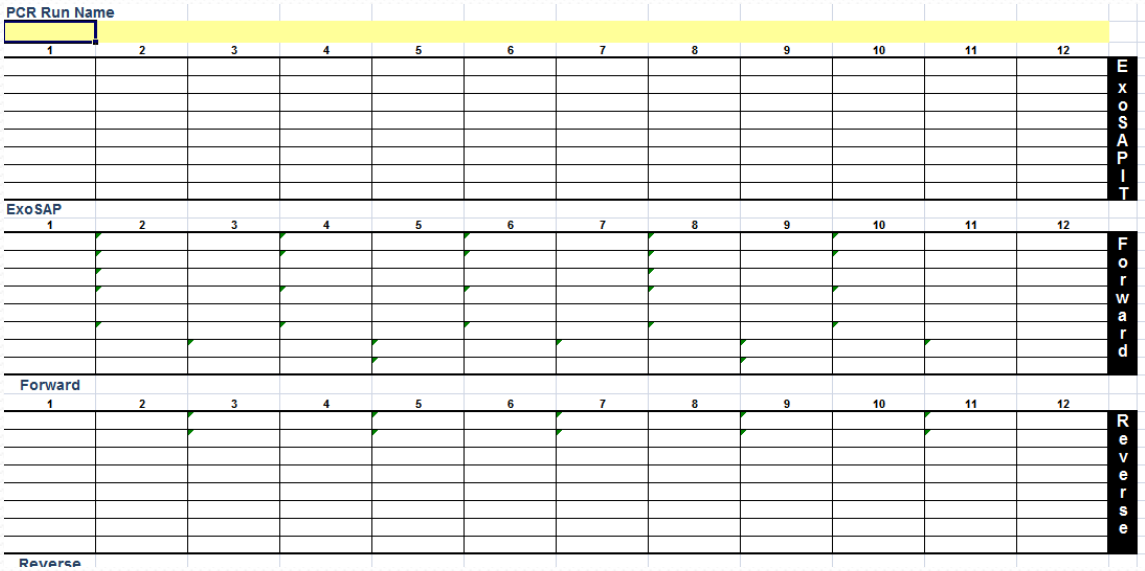
PCR Tray Map

	Sample #1	Sample #2	Sample #3	Sample #4	Sample #5	Sample #6	Sample #7	Sample #8	Sample #9	Sample #10	Sample #11	Sample #12
	1	2	3	4	5	6	7	8	9	10	11	12
A	10	10	10	10	10	10	10	10	10	10	10	10
B	11	11	11	11	11	11	11	11	11	11	11	11
C	13	13	13	13	13	13	13	13	13	13	13	13
D	14	14	14	14	14	14	14	14	14	14	14	14
E	15	15	15	15	15	15	15	15	15	15	15	15
F	16	16	16	16	16	16	16	16	16	16	16	16
G												
H												

Thermocycler method: pcr- men

Sequencing Throughput

- High throughput
  - 96 samples, one exon (amplicon)/plate
- Medium throughput
  - 1 plate – 1-8 samples, 3-48 reactions/sample
- Low throughput
  - Manual is faster



Clinical Parameters

- Clinical sensitivity
  - Percent affected individuals in which mutations can be found in the gene
  - Mutation detection rate
- Clinical specificity
  - Percent of unaffected individuals in which mutations are found in a gene
  - Penetrance
- Reference or reportable range
  - Description of gene regions interrogated
  - Mutations tested
  - Zygosity

Implementation

- Validation summary
  - With refseq, known SNPs, known double mutations, database information
- Standard operating procedure
- Training
- Costs
- Test information
- Reporting
- Internal databases
- Proficiency testing

Reporting

- Result
  - Standard *vs* Traditional nomenclature
    - Example: Beta globin amino acids are commonly known from the mature protein (-1 amino acid)
  - Nucleic Acid
    - Example: c.2183delAA
  - Amino Acid
    - Example p.G542X
- Reference sequence (version) and numbering scheme
- Interpretation
- Recommendations

ACMG Recommendations

- Report clinical significance
  - "… the laboratory must provide the interpretive information and a best estimate of clinical significance for the variants….."
    - ACMG recommendations for standards and interpretation and reporting of sequence variations. Richards et al. Genet Med 2008 10:294-300

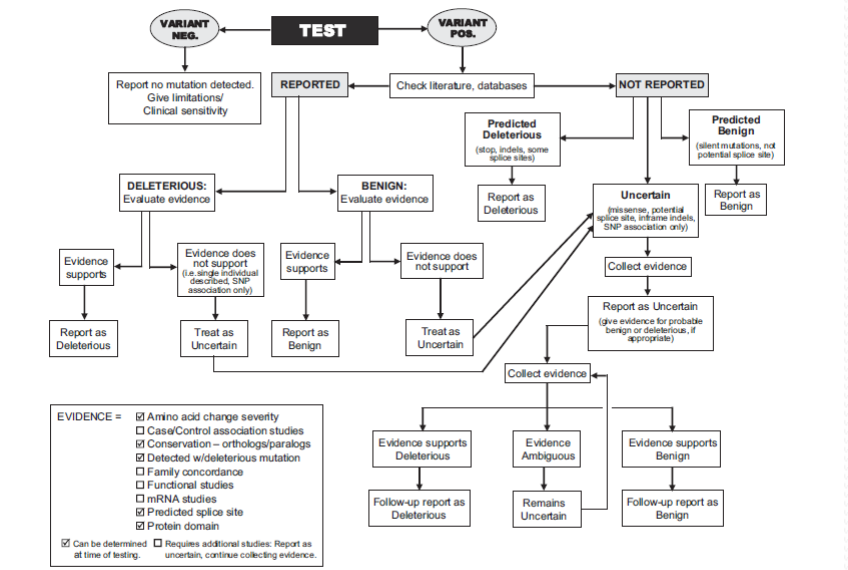


Fig. 1. A decision tree for interpretation of sequence variants and clinical reporting. Evidence that can be used to support sequence variant interpretation is shown in the box at the bottom.

Mutation Categories

- Previously reported
  - Pathogenic
  - Benign
  - But check original reports
- Previously unreported
  - Expected pathogenic
  - Suspected pathogenic
  - Uncertain
  - Suspected benign
- Further classification
  - Severe, moderate, mild, very mild

Interpretation

- Exonic
  - Frameshift (presumed pathogenic)
  - Nonsense (presumed pathogenic, except 3’ end?)
  - In-frame deletion/duplication (may or may not be pathogenic)
  - Missense (may or may not be pathogenic)

Missense Mutation

- Evidences:
  - Reported before?
    - Seen in affected or control individuals?
  - Conserved amino acid?
    - Over gene families or species?
  - Active site in the protein?
  - Affect mRNA levels?
  - Occur in the general population?
  - Co-occurrence with causative mutations
  - Track with disease in the family?
  - Functional studies available?
    - IHC, structural analysis, RNA, biochemical studies

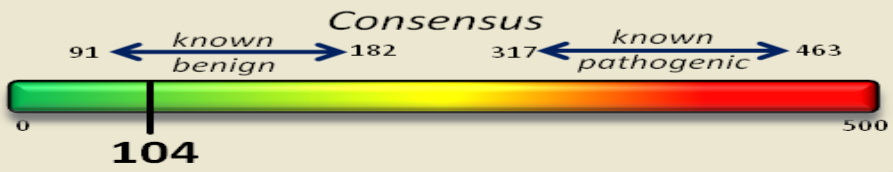
Amino Acid Prediction

- Existing predictions programs
  - PolyPhen 2, SIFT, Pmut, PhD-SNP, nsSNPAnalyzer, AlignGVGD
- Predictions using machine learning classification tools.
  - Gene-specific algorithms outperform generalized tools
  - Developed a standardized metric for evaluation of uncertain gene variants.
  - Visualization models for clinical implementation
- Emerging “authoritative” (clinically curated) gene variant/disease archives

ACADM UNCERTAIN VARIANTS

ACADME - A170S

Predictor	Call	Score
SIFT	tolerated	21
PolyPhen	benign	27
PMut	neutral	7
MutPred	benign	45
PSAAP	benign	4

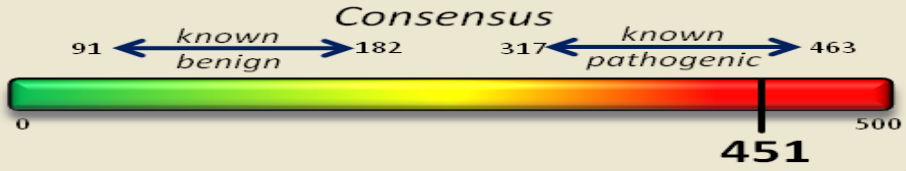


PREDICTED BENIGN

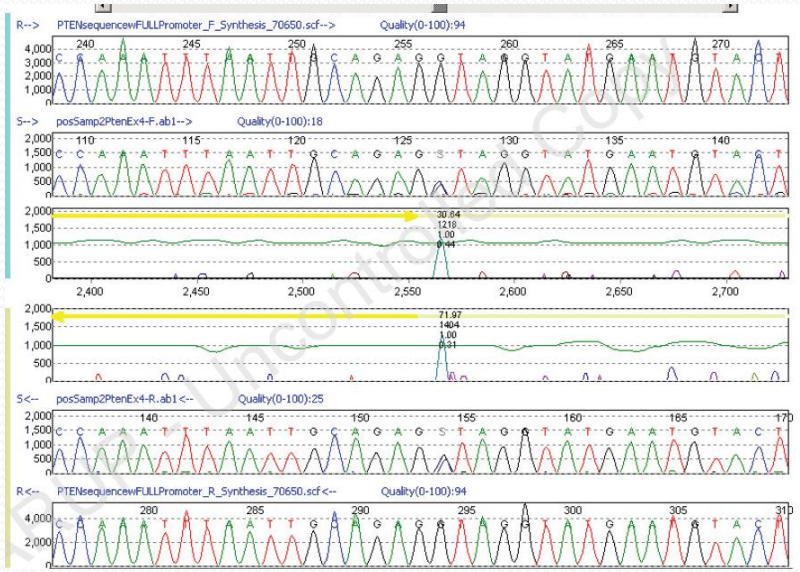
PREDICTED PATHOGENIC

ACADME - A372D

Predictor	Call	Score
SIFT	affects function	98
PolyPhen	probably damaging	99
PMut	pathological	82
MutPred	pathogenic	88
PSAAP	pathogenic	84



Intronic Mutation



- Intronic
  - has it been reported before?
  - approximately 20-50 bases
  - potential splice site http://www.fruitfly.org/seq_tools/splice.html
    - consensus sequence GT …..AG
      - Donor GT(start of intron)
      - Acceptor GA (end of intron)
      - Branch site U(18-40 upstream of 3' splice site)

Finding Rare Variants

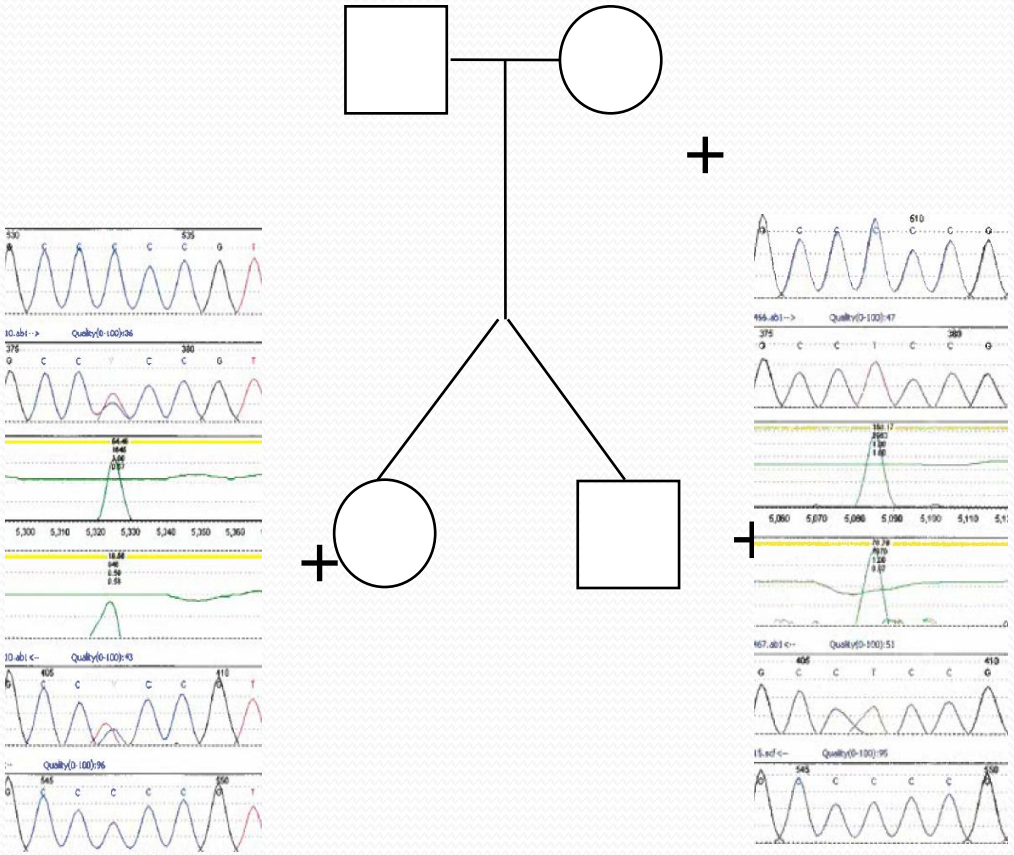
- CFTR Example
  - Child with F508del/I1028T
  - Mother also with F508del/I1028T
  - *In cis*
  - Does not explain symptoms in child
- Alpha globin Example
  - Apparent homozygous for p.X143Glu (Hb Seal Rock)
  - Subsequent deletion analysis showed -3.7Kb deletion
  - Compound heterozygous
  - Mild Hb H disease

Genetic Evidence
Family Concordance Studies

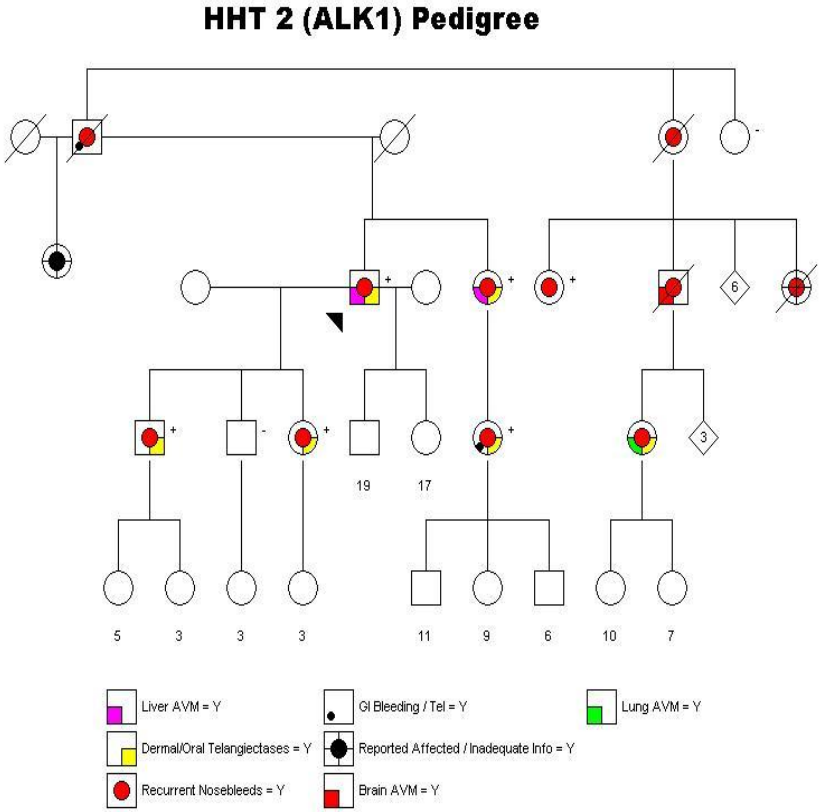
- Autosomal dominant/ X-linked/*de novo* mutations
- Single (affected) individual from a family tested
  - Results: sequence variant of unknown significance
- Test additional family members
  - Affected/Unaffected
  - Greater statistical power with affected distant relatives
- Evaluate pedigree data for evidence of causality
- Test hypothesis: Variant confers specified risk against the hypothesis of complete neutrality
- Determine likelihood ratio for causation

MECP2 Missense Mutation

- In silico prediction
  - Polyphen2: unknown
  - SIFT: Tolerated
- Present in 'normal' mother
  - Variable phenotype due to X inactivation?
- Present in unaffected brother



Extended Pedigree from Clinical Case



(R479Q)

Bayesian Factor = 461:1 in favor of causality

Likelihood Ratios: In Favor of Causality

Pedigree. Gene/Mutation	**Bayesian Factor**
1. *ACVRL1* p.R479Q	461.58
2. *ACVRL1* p.G402S	19.31
3. *ACVRL1* p.C344R	139.15
4. *ACVRL1* p.E407G	63.63
5. *ENG* p.W196R	121.35
6. *ENG* p.L300P	31.82
7. ENG p.R529H	7.98

Bayrak-Toydemir P. et al Experimental and Molecular Pathology 85 (2008) 45–49

Variant Annotation Summary

- Current manual method:
  - Check internal database for variant
  - Locus-specific databases
  - dbSNP, frequency (not all benign)
  - Prediction algorithms (Polyphen-2, Sift, others)
  - Literature search
  - Google

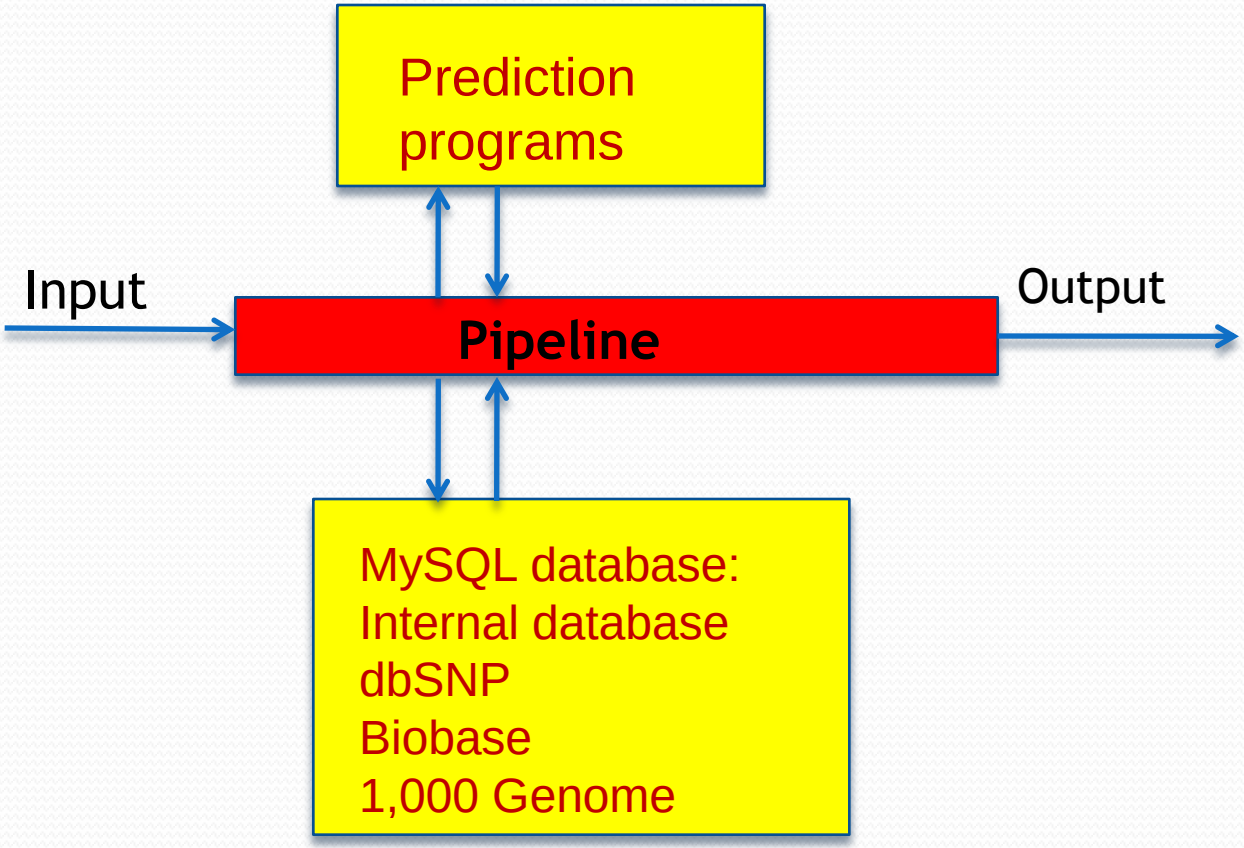
GALT Database

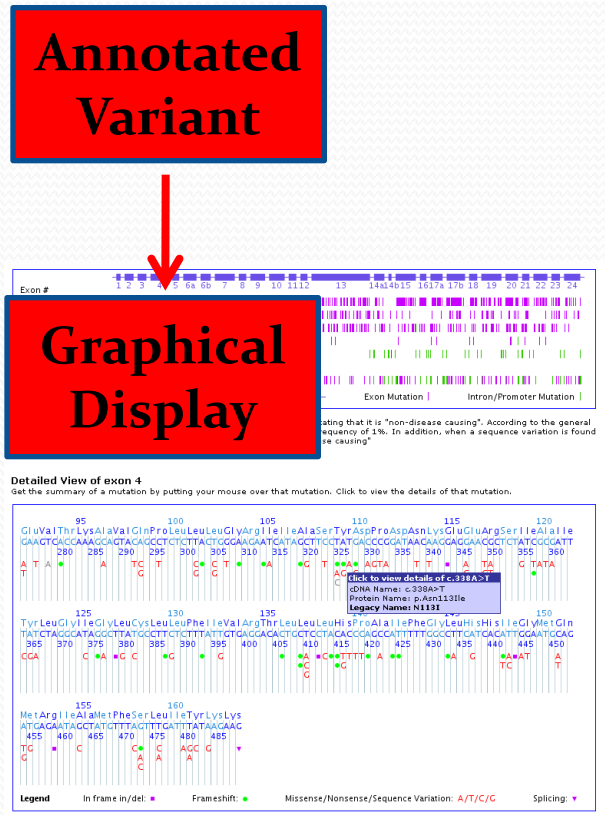
95 records found.

Location	Mutation Type	Nucleotide Change	Protein Change	Transport Activity	Expression	References	Comments
5'UTR	Polymorphism	c.-207G>C				Peltekova et al. 2004	
5'UTR	Uncertain	c.-185A>C		33		Calderon et al., unpublished	
5'UTR	Uncertain	c.-149G>A		33		Calderon et al., unpublished	
5'UTR	Deletion	c.-91_22del				Nezu et al. 1999	
5'UTR	Polymorphism	c.-78C>T		33		Koizumi et al., 1999	
5'UTR	Polymorphism	c.-77G>A		33		Koizumi et al., 1999	
5'UTR	Polymorphism	c.-38A>C				Calderon et al., unpublished	
Exon 1	Missense	c.3G>T	p.M1I	<5		Dobrowolski et al. 2005	
Exon 1	Insertion	c.4_5insC	p.R2PfsX136			Nezu et al. 1999	
Exon 1	Nonsense	c.12C>G	p.Y4X	<1		Wang et al. 2001	
Exon 1-8	Deletion	c.33_1427del	p.G12_L477del	2			
Exon 1	Missense	c.43G>T	p.G15W	3			
Exon 1	Missense	c.51C>G	p.F17L	14			
Exon 1	Missense	c.56G>C	p.R19P	<5			
Exon 1	Missense	c.59T>A	p.L20H			Calderon et al., unpublished	
Exon 1	Deletion	c.67_69delTTC	p.F23del	2		Lamhonwah et al. 2002	
Exon 1	Missense	c.83G>T	p.S28I	<12		Rahbeeni et al. 2002	

Large deletion found in two patients. Patient 1 of Italian descent was heterozygous for this mutation and p.G218VfsX68. Patient 2 of Mexican descent was heterozygous for this mutation and p.T219SfsX20.

Automated Pipeline





Courtesy of P. Ridge

Revolutionary Approach

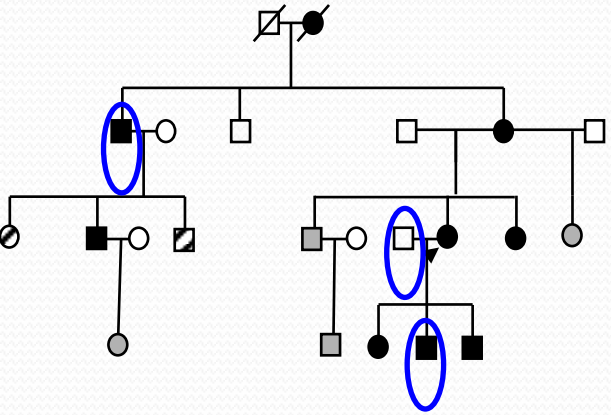
- Next-generation sequencing (NGS) 2005
  - Massive Parallel Sequencing in a flow cell (400 Mb to 30Gb)
  - Large scale sequencing/re-sequencing of the chromosome possible
    - Clonally amplified templates
    - Single molecule templates

Next Gen Sequencing

- Gene panels
  - Genes known to cause disease
  - Variant discovery
- Whole exome
  - Gene discovery
- Whole genome
  - Gene discovery

Data Analysis: Variant Filtering

Pinpoint which gene causes HHT4!



Kept affected SNVs (2 shared) ~36%

↓

Remove unaffected SNVs ~25%

↓

Remove 8 HapMap SNVs ~27%

↓

Remove SNVs in dbSNP, pseudogenes, repeat regions ~7%

↓

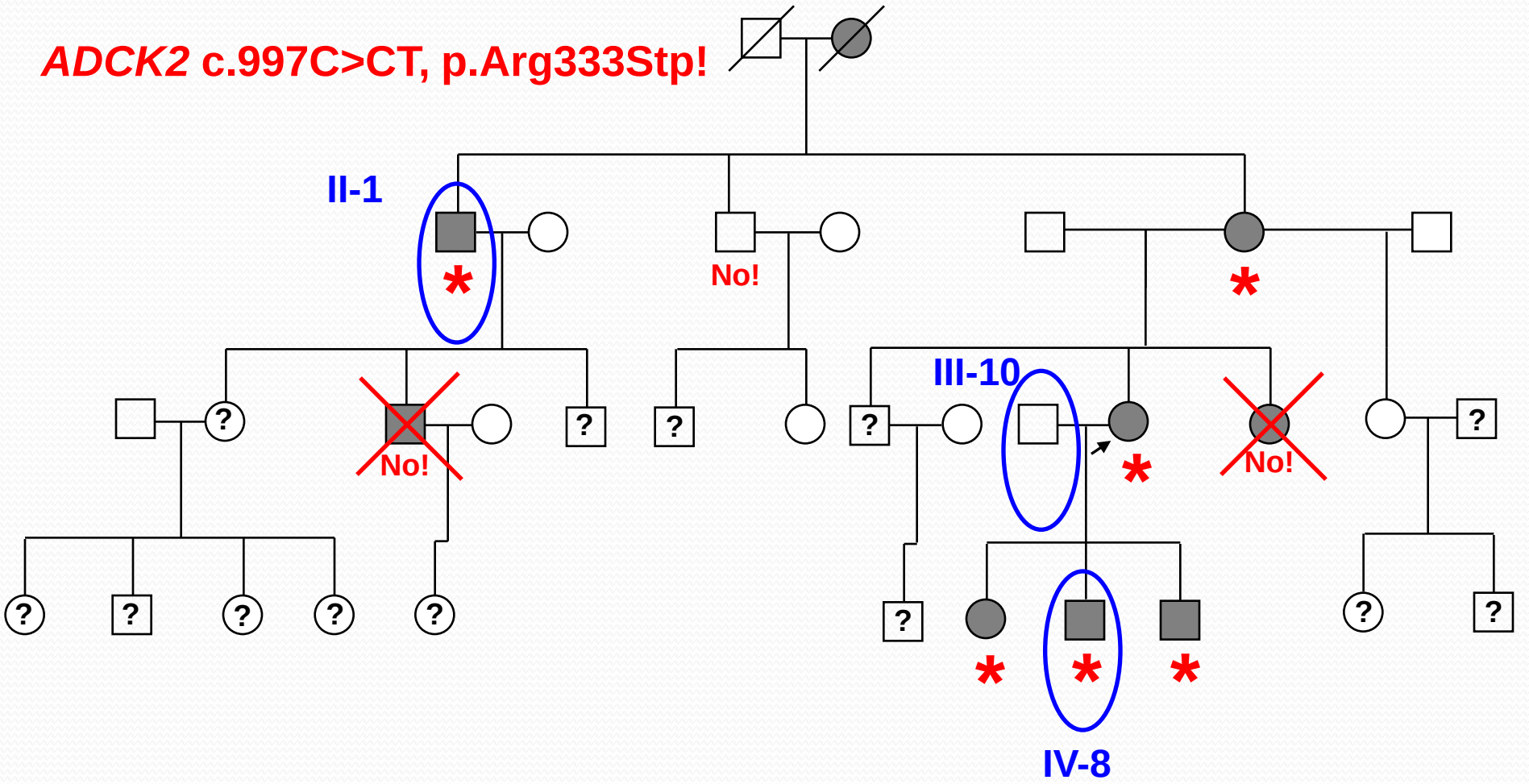
~5% remaining!

Focus on NS-SNVs in protein coding regions, UTR, splice sites

Lead candidate gene: *ADCK2* c.997C>CT, p.Arg333Stp

Courtesy of Drs. P. Bayrak-Toydemir, W Donahue

ADCK2 Variant Segregation



Did not track according to affected status

Courtesy of Drs. P. Bayrak-Toydemir, W Donahue

Sanger Sequencing Continued Role

- Complex regions difficult to align with NextGen software
- Confirm that variants are “real”
- Confirm that variants are “significant”
  - Family concordance studies
- Familial testing

Conclusions

- Sanger sequencing has allowed clinical testing for numerous diseases
- Proper design and validation of sequencing tests can prevent analytical errors
- Sequence complexity can be addressed by primer design
- Interpretation complexity still a challenge
- Mutation databases with evidences for classification are needed
- Sanger sequencing will remain important as companion to next generation sequencing

Thanks to:

University of Utah
ARUP Laboratories

- Rong Mao, MD
- Pinar Bayrak-Toydemir, MD PhD
- Genevieve Pont-Kingdon, PhD
- Perry Ridge, MS
- Karl Voelkerding, MD
- Whitney Donahue, PhD
- Friederike Gedge
- David Crockett, PhD