Confirmation Sensitive Capillary Electrophoresis

Chris Mattocks

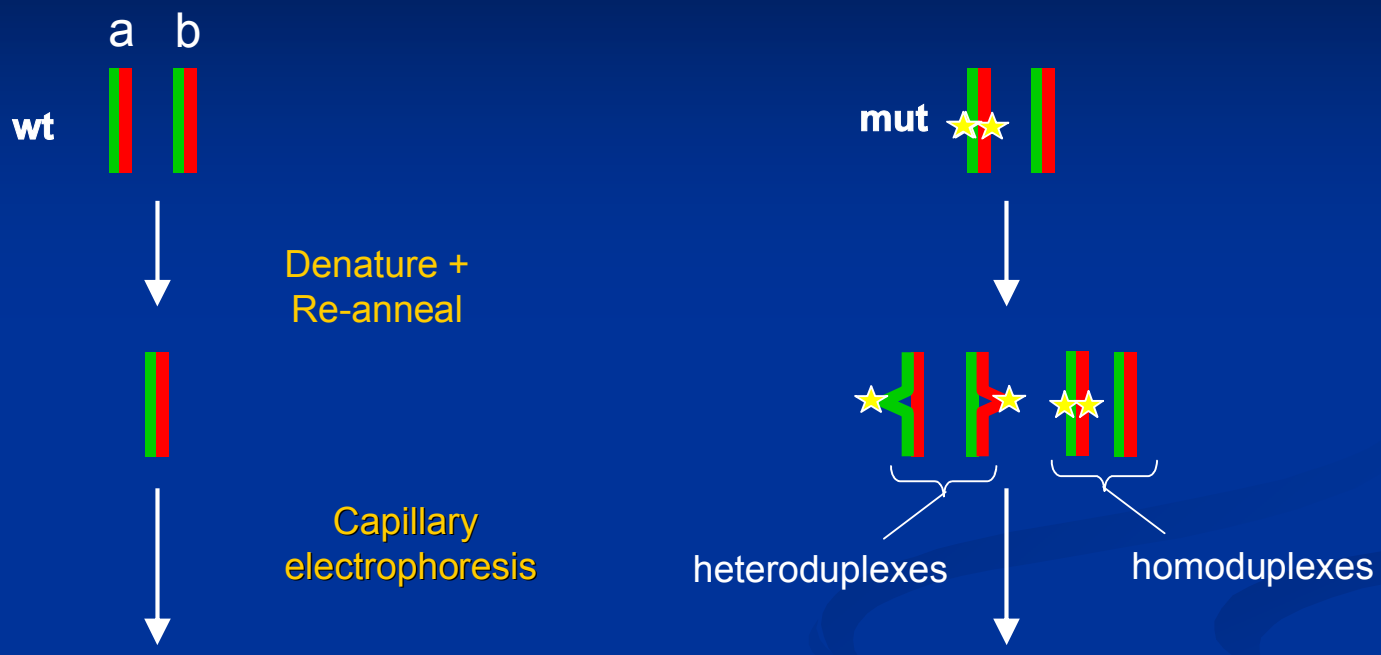
National Genetics Reference Laboratory (Wessex-Salisbury)

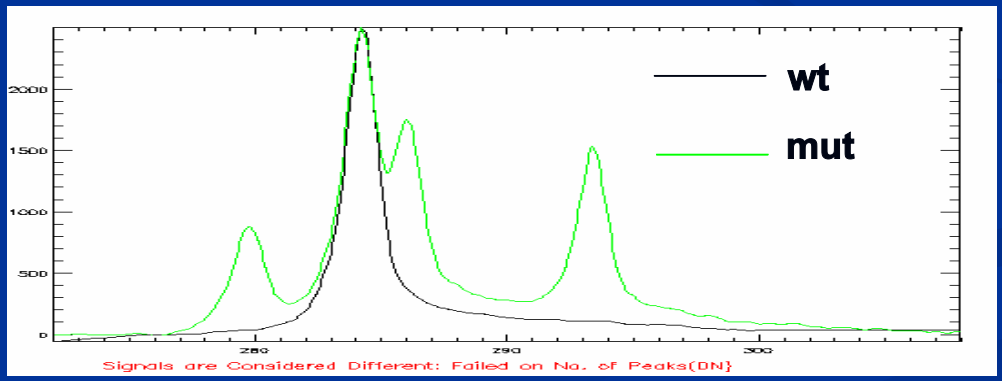
HTSF

- Setup in response to genetics White paper 2003
- Aims provide system and infrastructure for rapid efficient and timely processing of mutation scanning for large genes.
- Flexibility w.r.t. referral rates
- Key target = TAT

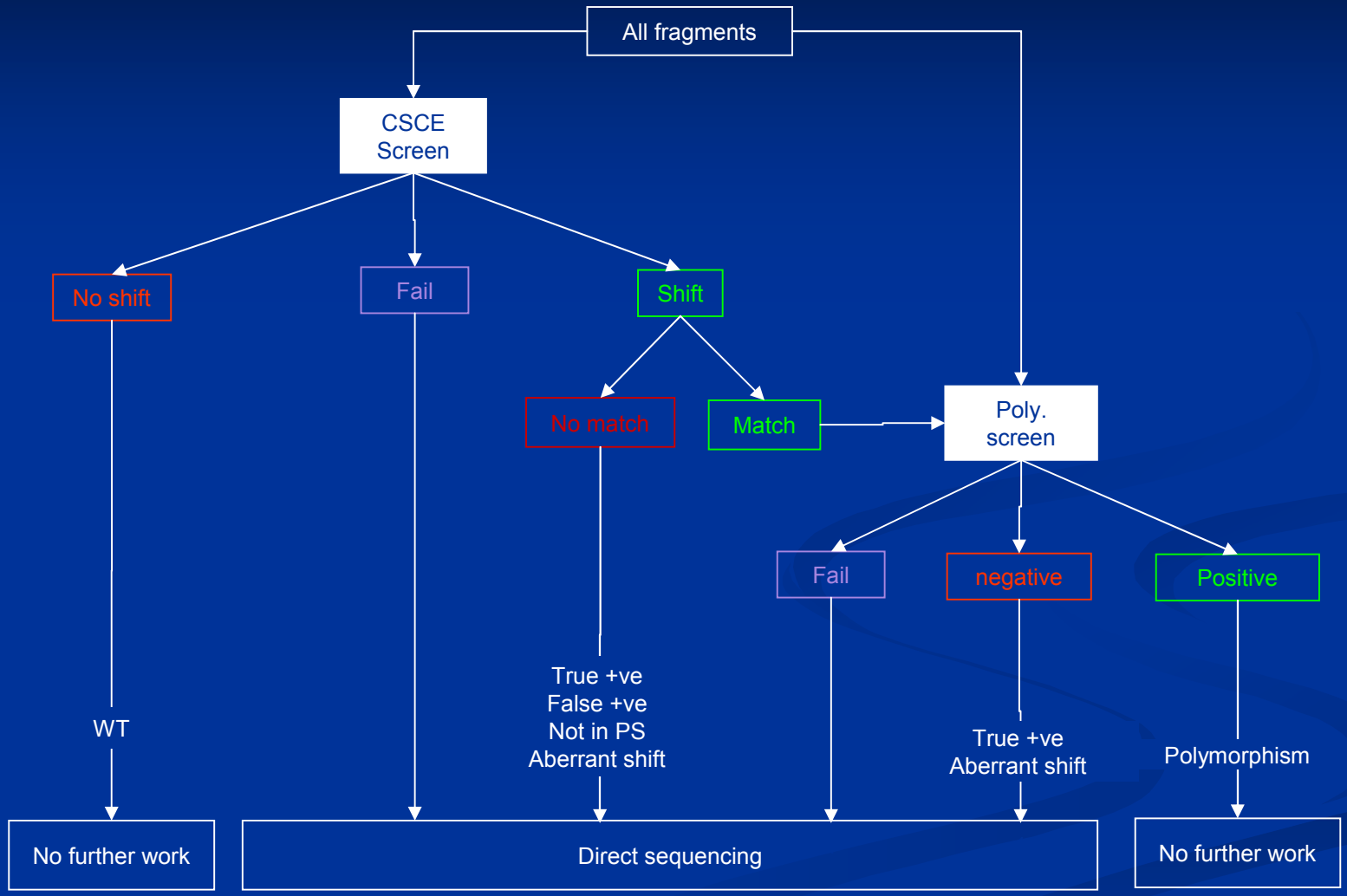
- Diagnostic service for BRCA1 & 2 april 2006

CSCE


a
b
wt
mut
Denature +
Re-anneal
Capillary
electrophoresis
heteroduplexes
homoduplexes


wt
mut
2000
1500
1000
500
0
280
290
300
Signals are Considered Different: Failed on No. of Peaks(DN)

Testing algorithm



HTSF to date

Tests	No.	%
Tested	1235	
CSCE	78309	
Sequenced	7499	9.6
Mutation	165	13.4
UV	1302	18.2
Polymorphism	7525	6.1 /sample

Sequencing	No.	%
CSCE fail	1302	1.7
SNPlex fail	3158	4.0
False +ve	2649	3.4
Mutation / UV	390	0.5

- Ave TAT <40working days since Sept 2006
- >141,000 sequencing reactions (~1500 plates),
- 52 Mbases sequencing analysis

Current validation status

Sample set	All variations		20-60% GC		All variations - Blind		20-60% GC - Blind	
	T	D	T	D	T	D	T	D
BRCA (SCOBEC set)	38	38	38	38	38	38	38	38
Marfans (internal)	48	48	48	48	48	48	48	48
Generic Mutation controls	48	47	36	36	48	47	36	36
BRCA Mutation controls	72	72	72	72				
BRCA poly controls	36	36	36	36				
Totals	242	241	230	230	134	133	122	122
Sensitivity (95% CI)	>98.35		>98.7		>97.03		**>97.54**	

Inter-laboratory Validation

- To determine the nature and affect of the key factors influencing sensitivity (Se) and specificity (Sp) and build a general model for the Se and Sp of CSCE.
- To determine Se and Sp of mutation scanning by CSCE for the BRCA genes under laboratory conditions.
- To provide a framework to enable test conditions to be set up and maintained in any interested lab for diagnostic testing.

Validation setup

- Phase I - Determine and evaluate parameters that are likely to affect the Se and Sp of CSCE (control variables)
  - Focus on CSCE alone i.e. central analyte preparation
  - Blind full factorial experiment
  - Systematic panel of artificial controls
- Phase II – Determination of Se and Sp
  - Full process set up in each participating laboratory
  - Blinded analysis
  - Panel of mutations ascertained by methods other than CSCE and confirmed by sequencing.

Phase I - Physical variables

Generic mutation detection controls - plasmid controls for systematic evaluation factors that are of general importance for mutation scanning technologies (48 mutant + 4 wt)

- GC content of the amplicon - 20%, 40%, 60% and 80%
- Location of the mutation in the fragment.

- Type of base substitution

Mutation	Heteroduplexes formed
A>C	C:T & G:A
A>T	T:T & A:A
G>A	A:C & T:G
G>C	C:C & G:G

Phase I - Process variables

Heteroduplexing regime (3)
Urea content of polymer (3)
Run temperature (3)
Run voltage (3)

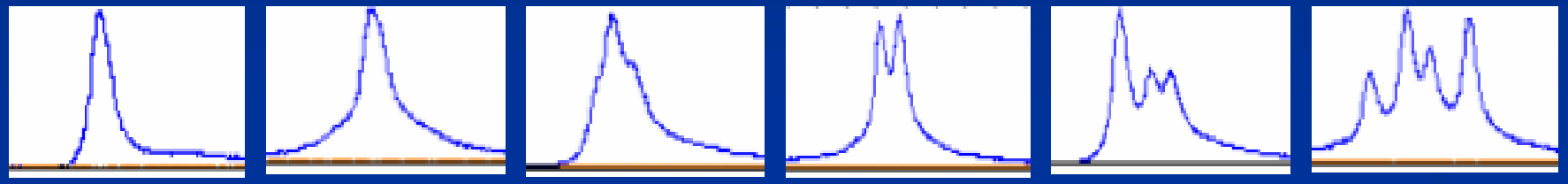
} Labs (3)

- Each variable used at two extreme values (+ and -) and one centre point (0) to enable determination of curvature in the response.
- Full factorial experiment = 3^4 =81 experiments (runs)
- Each run comprises 48 GMD mutant and 48 wt controls on 96 well plate.
- Full experiment blocked (split) over 3 labs

Response variables

- Manual analysis

  Resolution score

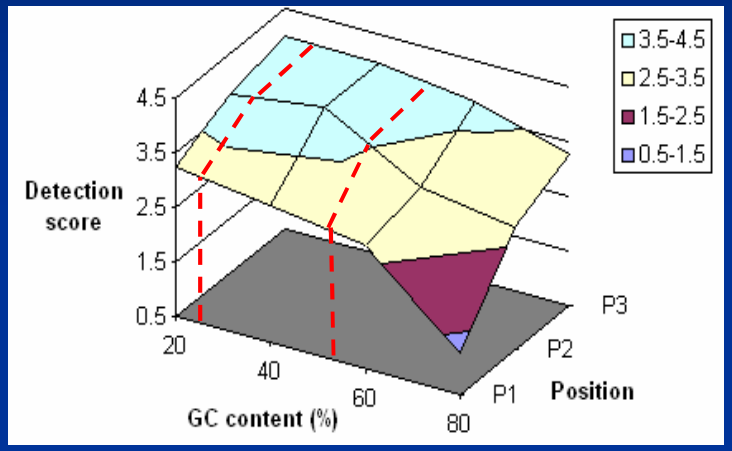


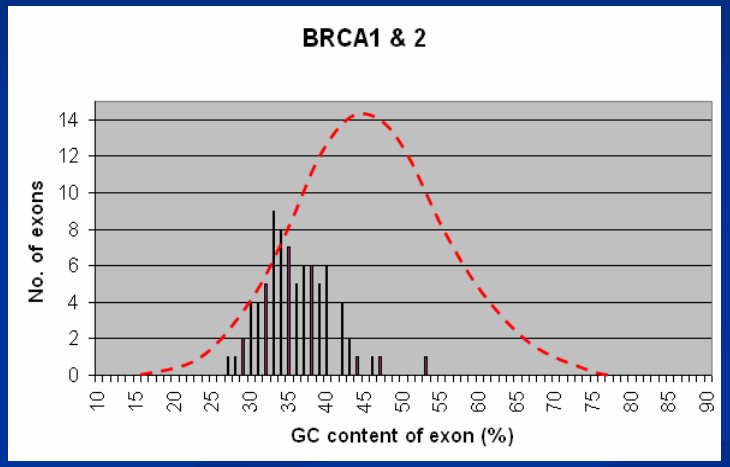
0 1 2 3 4 5

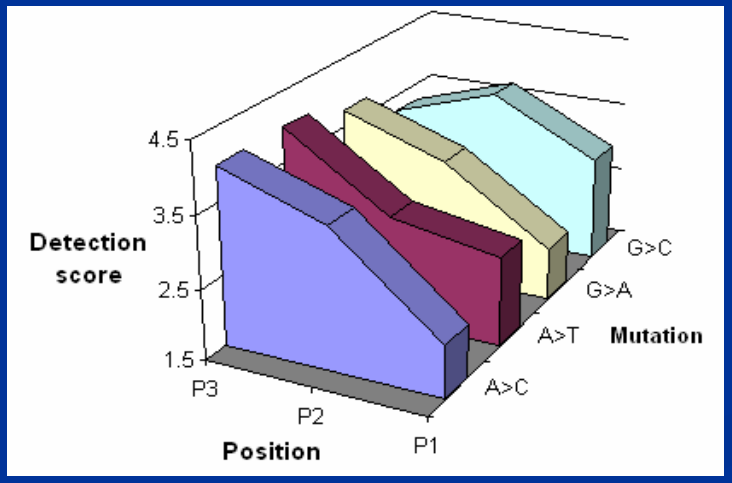
- Automated analysis

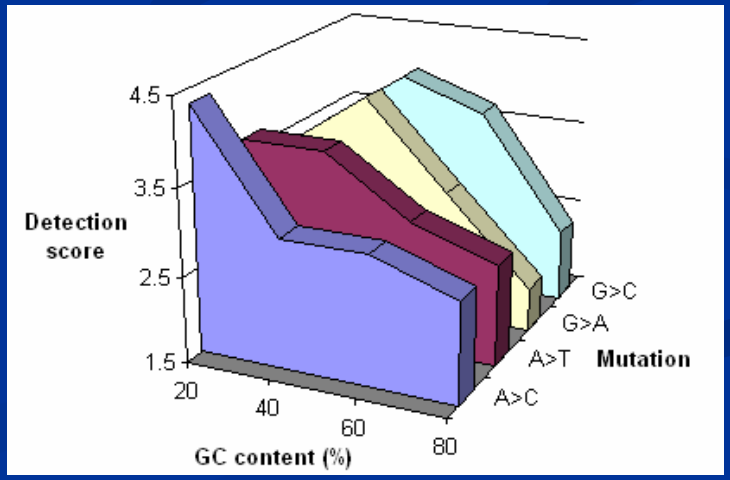
  5 different mathematical parameters that measure differences in peak shapes with respect to a wt control (Bionumerics – Applied Maths)

Outputs


3.5-4.5
2.5-3.5
1.5-2.5
0.5-1.5
Detection score
4.5
3.5
2.5
1.5
0.5
20
40
60
80
GC content (%)
P3
P2
P1
Position

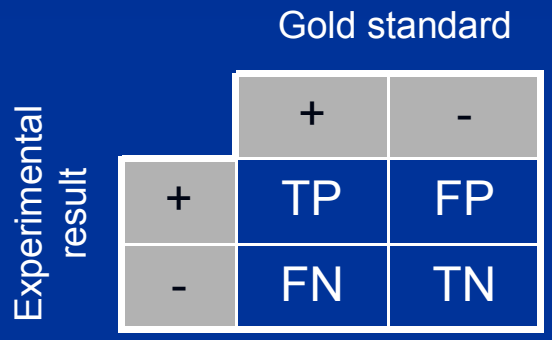

BRCA1 & 2
No. of exons
14
12
10
8
6
4
2
0
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
GC content of exon (%)


Detection score
4.5
3.5
2.5
1.5
P3
P2
P1
Position
G>C
G>A
A>T
A>C
Mutation


Detection score
4.5
3.5
2.5
1.5
20
40
60
80
GC content (%)
G>C
G>A
A>T
A>C
Mutation

Phase II - Sensitivity and specificity

- Measured relative to a chosen 'gold standard' that defines what is known to be present in the chosen sample set
  - Widely considered to be sequencing but should more properly include a range of techniques with different capabilities.
  - NB it is likely that the gold standard in itself is flawed (i.e. <100% sensitive)



TP = True positive
FP = False positive
FN = False negative
TN = True negative

- Sensitivity – Proportion of gold standard +ves correctly identified TP/(TP+FN)
- Specificity – Proportion of gold standard -ves correctly identified TN/(TN+FP)
  - Important factor given the need for confirmation of detected variants

Sensitivity

	Sensitivity
More than 500 mutations have been identified in the CFTR gene, making it an excellent system for testing mutation scanning techniques. To assess the sensitivity of denaturing gradient gel electrophoresis (DGGE), we collected a representative group of 202 CFTR mutations. All mutations analyzed were detected by scanning methods other than the DGGE approach evaluated in this study. DGGE analysis was performed on 24 of the 27 exons and their flanking splice site sequences. After optimization, 201 of the 202 control samples produced an altered migration pattern in the region in which an alteration occurred. The remaining sample was sequenced and found not to have the reported mutation. The ability of DGGE to identify novel mutations was evaluated in three Asian CF patients with four unknown CF alleles. Three novel Asian mutations were detected-K166E, L568X, and 3121-2 A-->G (in homozygosity)-accounting for all CF alleles. These results indicate that an optimized DGGE scanning strategy is highly sensitive and specific and can detect 100% of mutations.	100%
...A larger set of 32 mutant DNA specimens was then analyzed using these optimized tandem CAE-SSCP/HA protocols and materials and yielded 100% sensitivity of mutation detection…	100%

Rule of three

NB this is an approximation but remarkably accurate

We can say with 95% confidence that the probability of a false negative, given a study of **n** samples with no false negatives, is **3/n**.

$\therefore$ for n = 201 (with no false negatives)

probability of a false negative = 3/201
= 0.015
= 1.5%

$\therefore$ With 95% confidence sensitivity $\geq$ 98.5%

(if you want to be 99% confident the rule is 4.6/n)

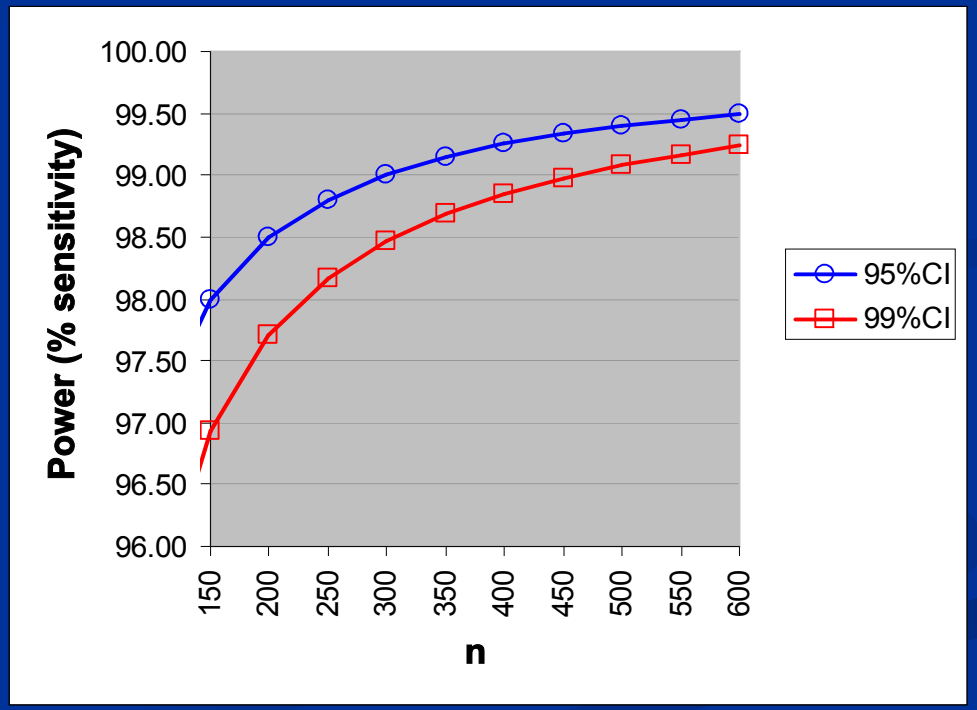
Sensitivity

	Sensitivity with 95% confidence
More than 500 mutations have been identified in the CFTR gene, making it an excellent system for testing mutation scanning techniques. To assess the sensitivity of denaturing gradient gel electrophoresis (DGGE), we collected a representative group of 202 CFTR mutations. All mutations analyzed were detected by scanning methods other than the DGGE approach evaluated in this study. DGGE analysis was performed on 24 of the 27 exons and their flanking splice site sequences. After optimization, 201 of the 202 control samples produced an altered migration pattern in the region in which an alteration occurred. The remaining sample was sequenced and found not to have the reported mutation. The ability of DGGE to identify novel mutations was evaluated in three Asian CF patients with four unknown CF alleles. Three novel Asian mutations were detected-K166E, L568X, and 3121-2 A-->G (in homozygosity)-accounting for all CF alleles. These results indicate that an optimized DGGE scanning strategy is highly sensitive and specific and can detect 100% of mutations.	≥98.5%
...A larger set of 32 mutant DNA specimens was then analyzed using these optimized tandem CAE-SSCP/HA protocols and materials and yielded 100% sensitivity of mutation detection…	≥91.6%

Validation power

- Determined by sample size
- Requirement based on assumption that all mutations will be detected
- Exponentially diminishing return

n	95%CI	99%CI
50	94.00	90.80
100	97.00	95.40
150	98.00	96.93
200	98.50	97.70
250	98.80	98.16
300	99.00	98.47
350	99.14	98.69
400	99.25	98.85
450	99.33	98.98
500	99.40	99.08
550	99.45	99.16
600	99.50	99.23



Phase II - Retrospective validation

- Aim for >300 different (BRCA) mutations previously characterised by sequencing (power for >99% Se and Sp).
- Each variation will be analysed for 3 different exons
  - A balanced study (i.e. $n_{wt} = n_{mut}$) gives equal power to determine Se and Sp – other formats may be more practical
- Whole process to be carried out in 3 different labs
  - Validation of the whole process rather than just the analysis technology – emulates live usage.

Summary and conclusions

- CSCE is a very efficient method for mutation scanning in large genes
- For diagnostics TAT and flexibility w.r.t. batching is essential
- Setting up systems for automating even a simple technology like CSCE is complex and time consuming

Summary and conclusions

- There is a distorted expectation of sensitivity based on numerous reports from companies and in the literature claiming misleading values.
- No technology is (provably) perfect
- No validation is perfect
  - The gold standard is likely to be flawed
  - Not all parameters can be examined e.g. sequence context
- Validation can be applied in a general way but only with the aid of good models of the behaviour of the technology under varying conditions

Summary and conclusions

- Validation should include determination of Se and Sp as well as practical tools to enable replication and maintenance of the technology to that level of accuracy.
- Se and Sp should be determined using a whole process not just the analysis i.e. emulate real life.
- Validation must be practical
  - Size and therefore power may be limited by sample availability
  - There is an exponentially diminishing return in power with increased sample size and someone has to do the work!
  - Inter-laboratory variations are likely to be the largest source of variation in Se for a particular technology
  - There is a need for agreed minimum standards of validation

CSCE validation – Participating labs

- NGRL (Wessex) co-ordinator
- Leuven
- Nijmegen
- WRGL

- Funded By ABI