

Next generation diagnostic sequencing: means to reduce overall sequencing cost

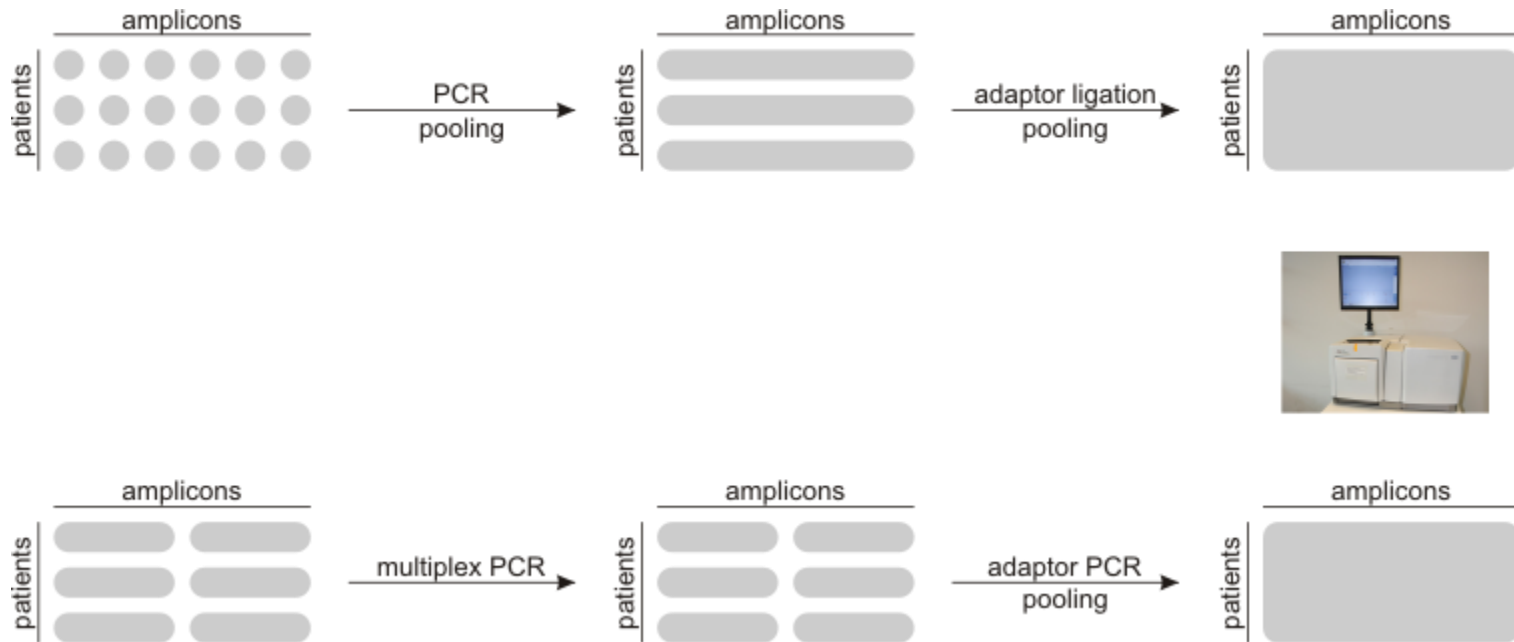
Jan Hellemans
CMGG & nextgnt

Background

- need to move NGS from research to diagnostics
- NXTGNT / CMGG cases
 - BRCA: 2 genes
 - FAA: 3 genes
 - deafness: 15 genes

 - multiple runs per case
 - 10 - >100 samples per case
- 1 run: multiple samples * multiple genes
- PCR based enrichment
 - more affordable than DNA capture if <1000 PCRs
 - amplicon size & GS-FLX read length match

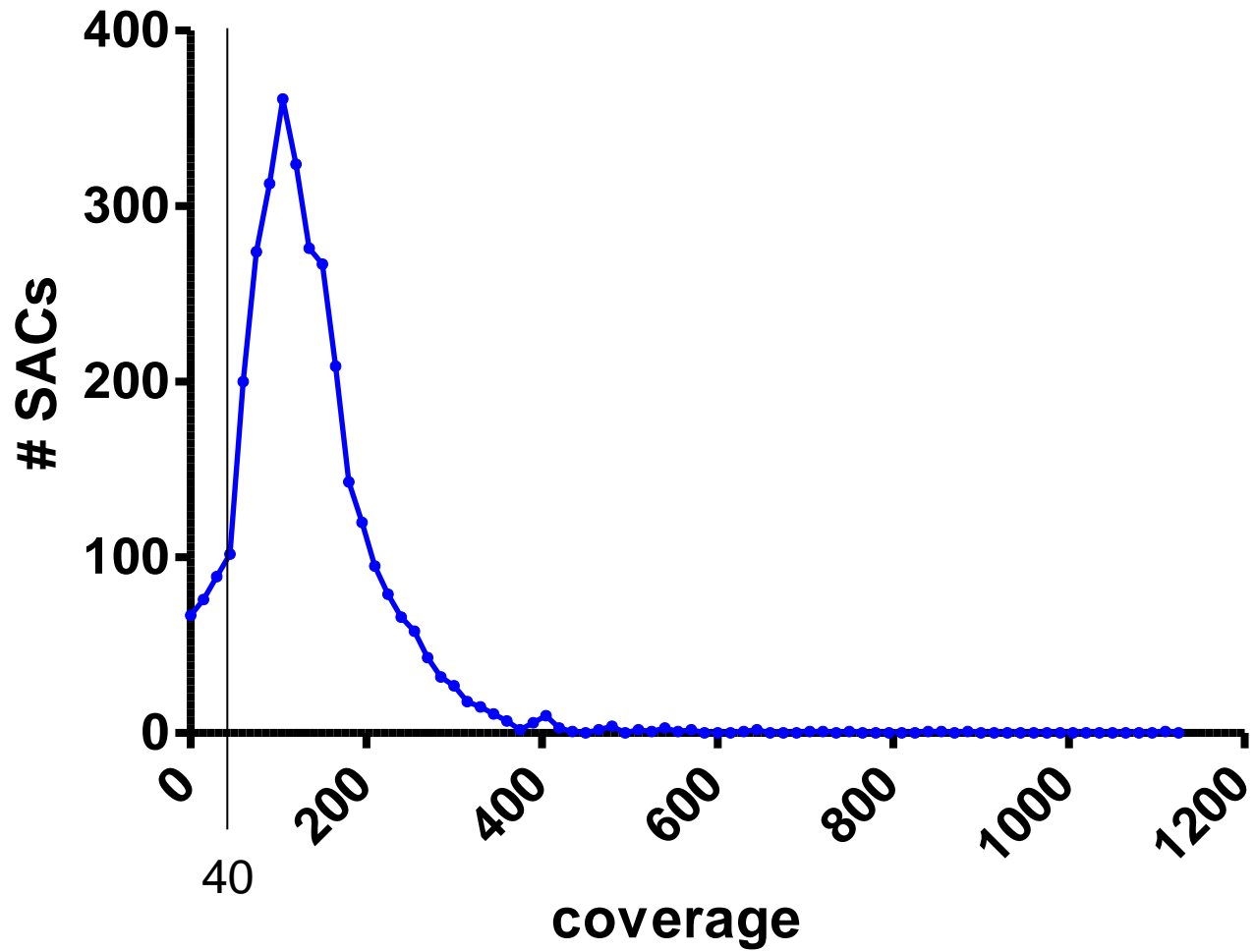
approaches



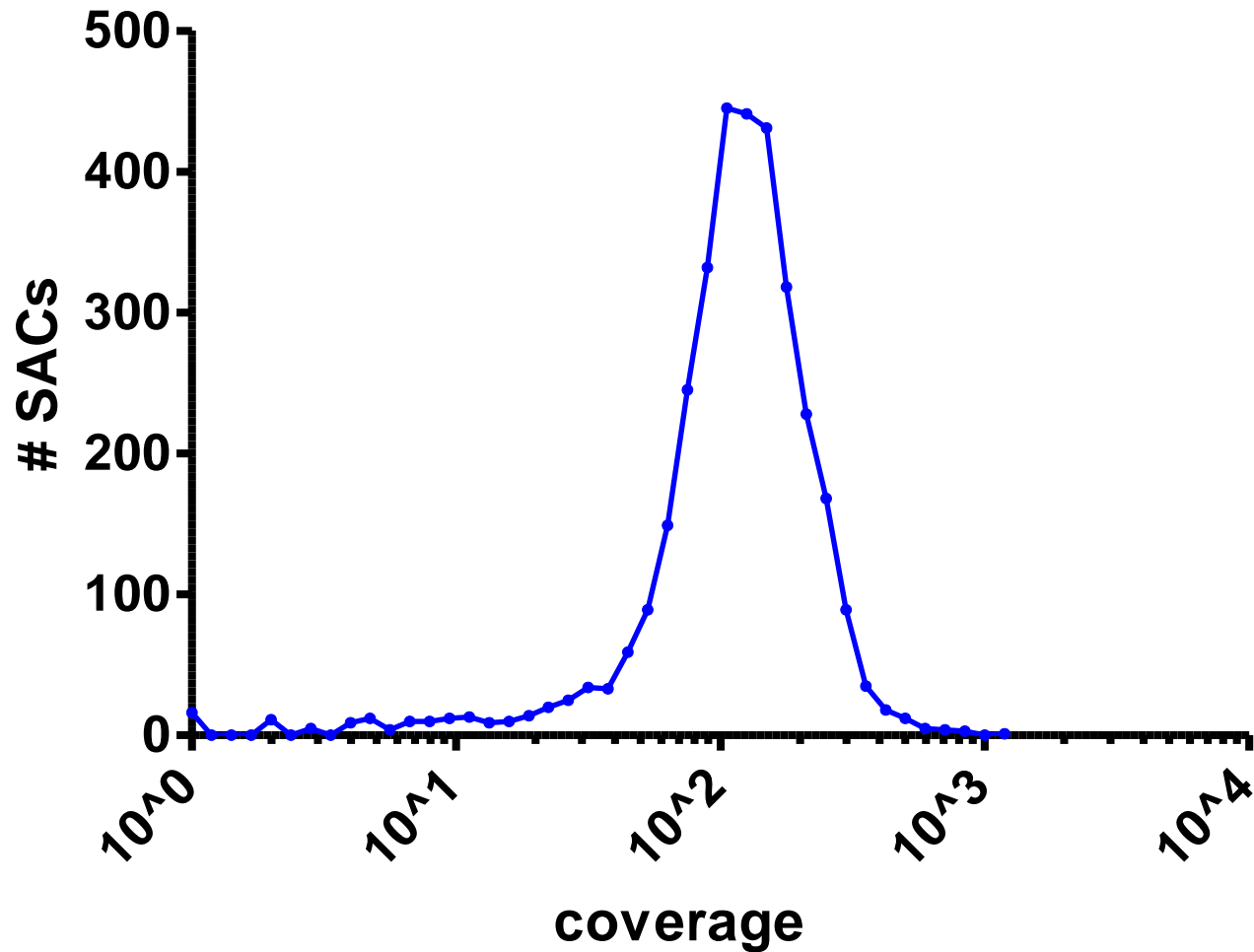
coverage

- x fold coverage required to compensate for
 - sampling effects
 - sequencing errors (especially for homopolymer stretches)
- typically 40-fold coverage for diagnostic detection of heterozygous variants
- ideally the same coverage for all SACs, in practice coverage will vary between
 - samples
 - amplicons

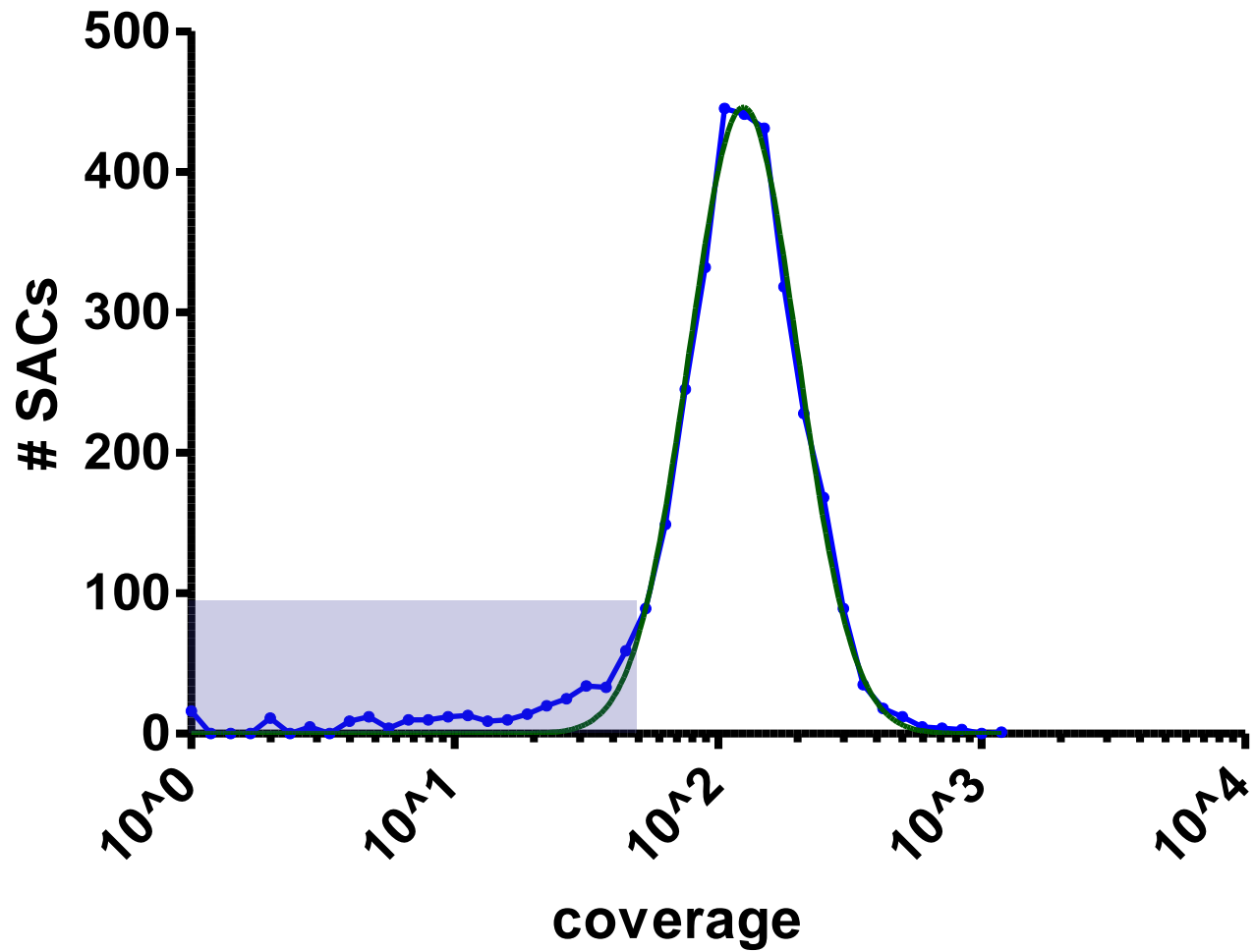
coverage distribution



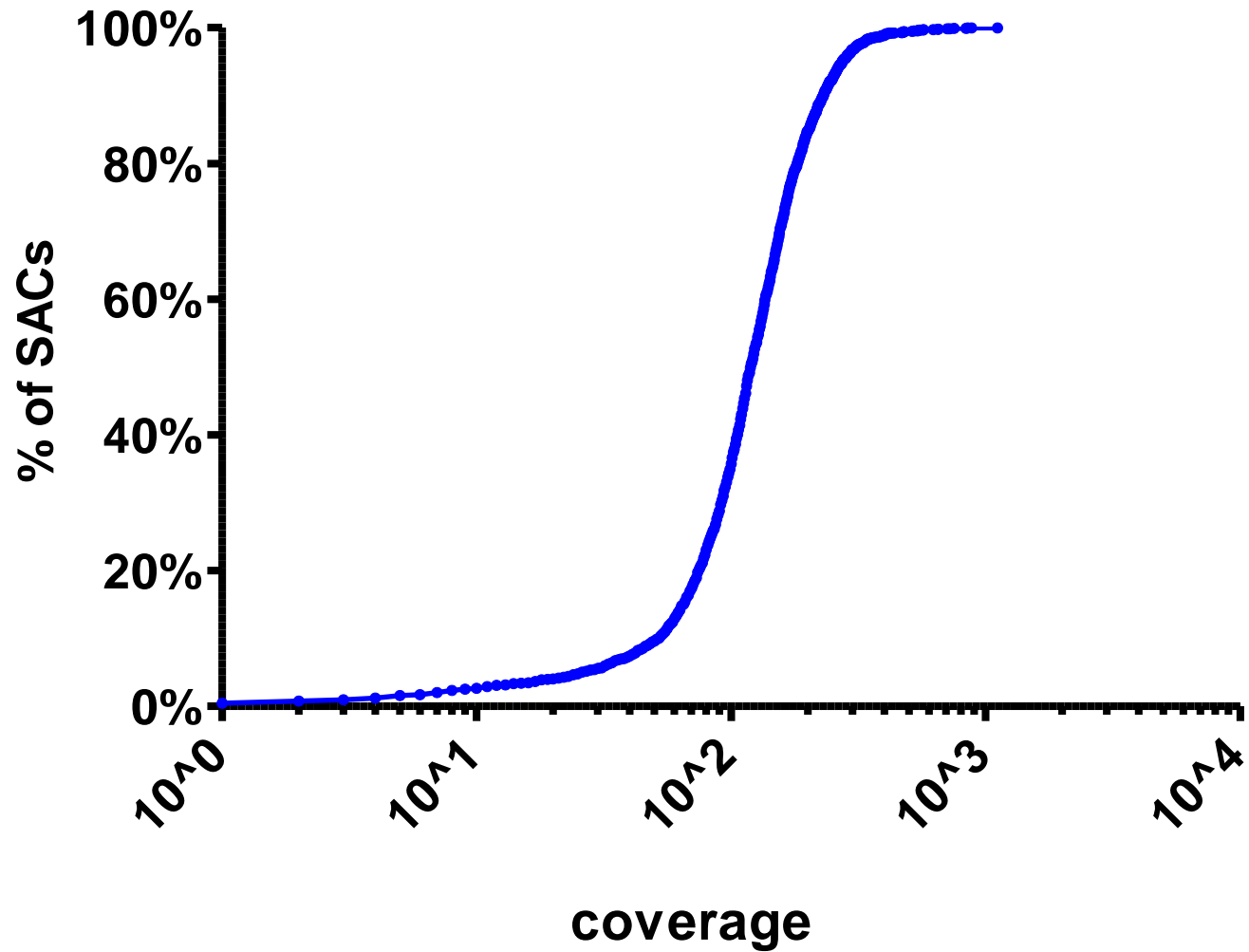
coverage distribution



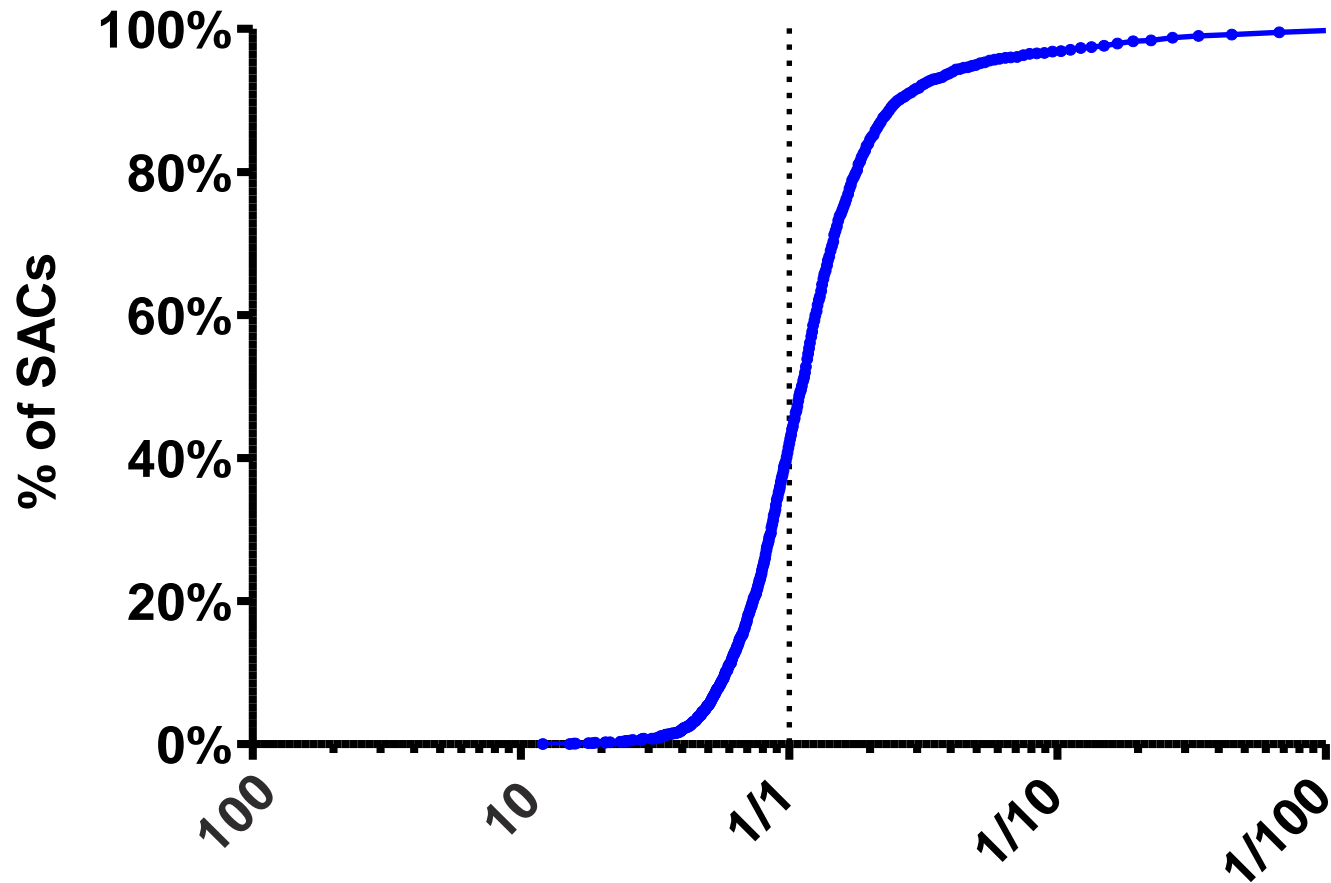
coverage distribution



coverage distribution

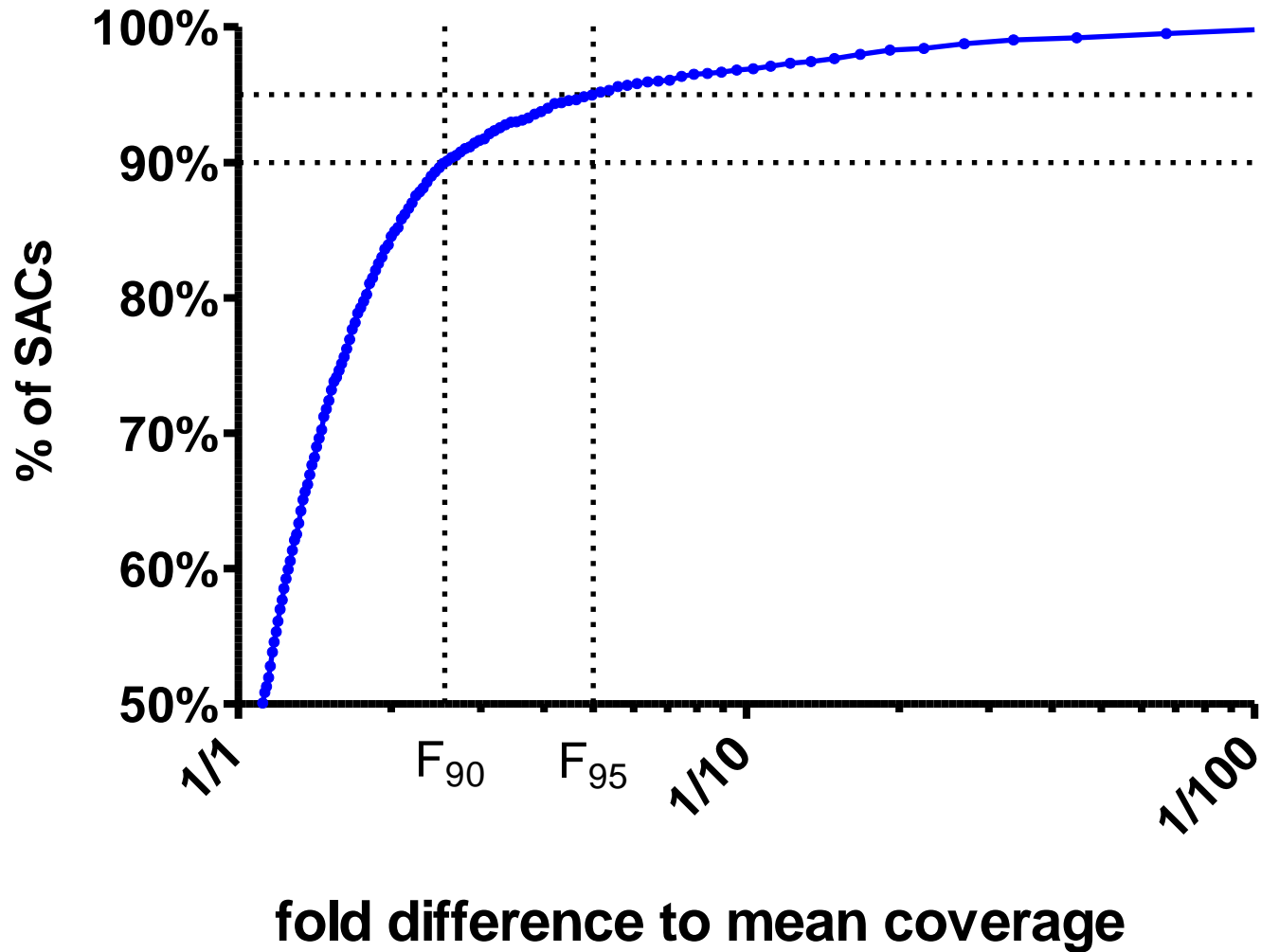


spread factor



fold difference to mean coverage

spread factor



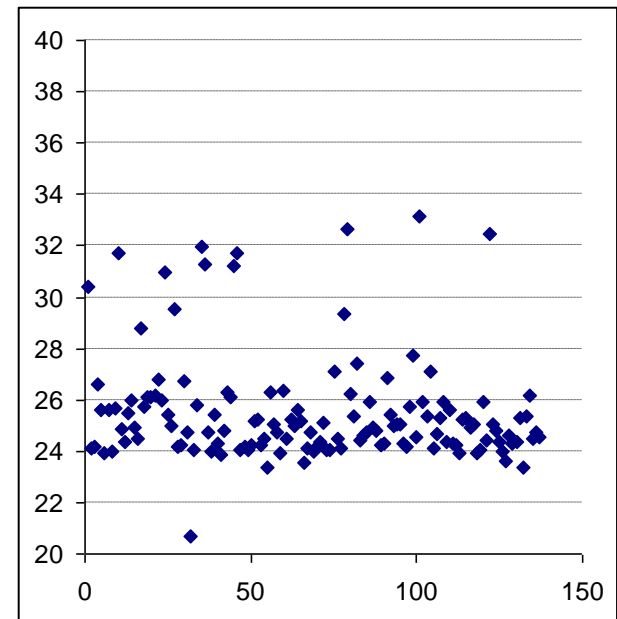
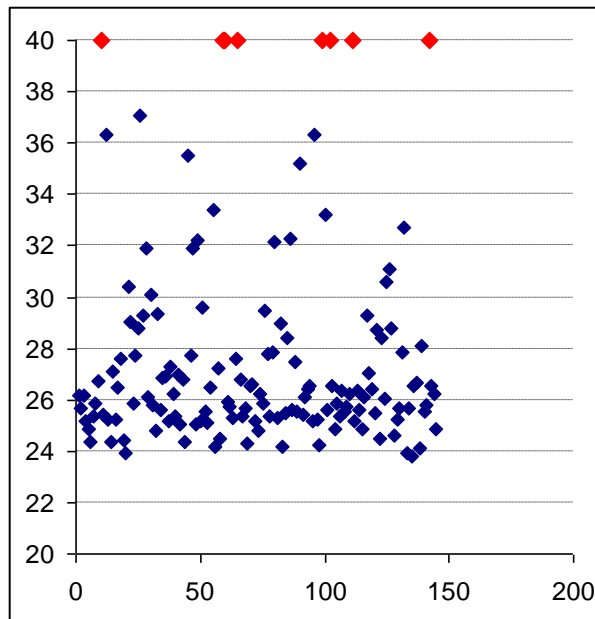
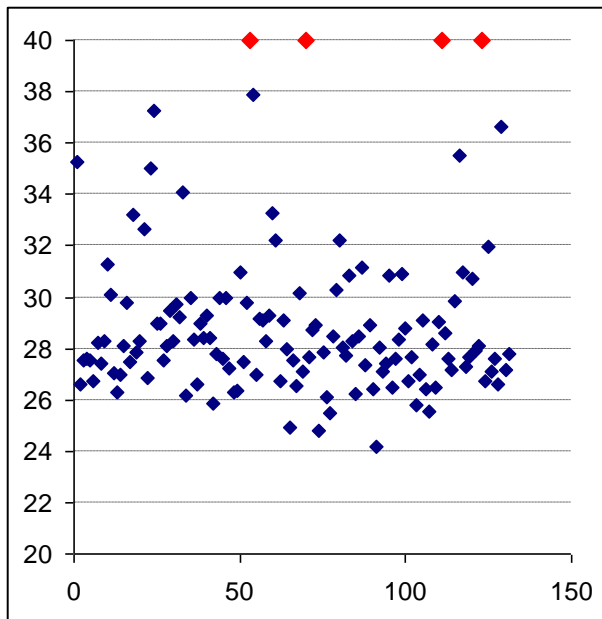
reducing cost

- required average coverage
= required minimum coverage * spread factor
= $40 * F_{90}$
- reduce cost → reduce average coverage
 - → reduce required minimum coverage
 - → reduce spread factor
- reduce required minimum coverage
 - (lower sensitivity)
 - reduce error rate by improving data analysis
 - V.I.P. (http://athos.ugent.be/VIP_pipeline/)

reduce spread factor

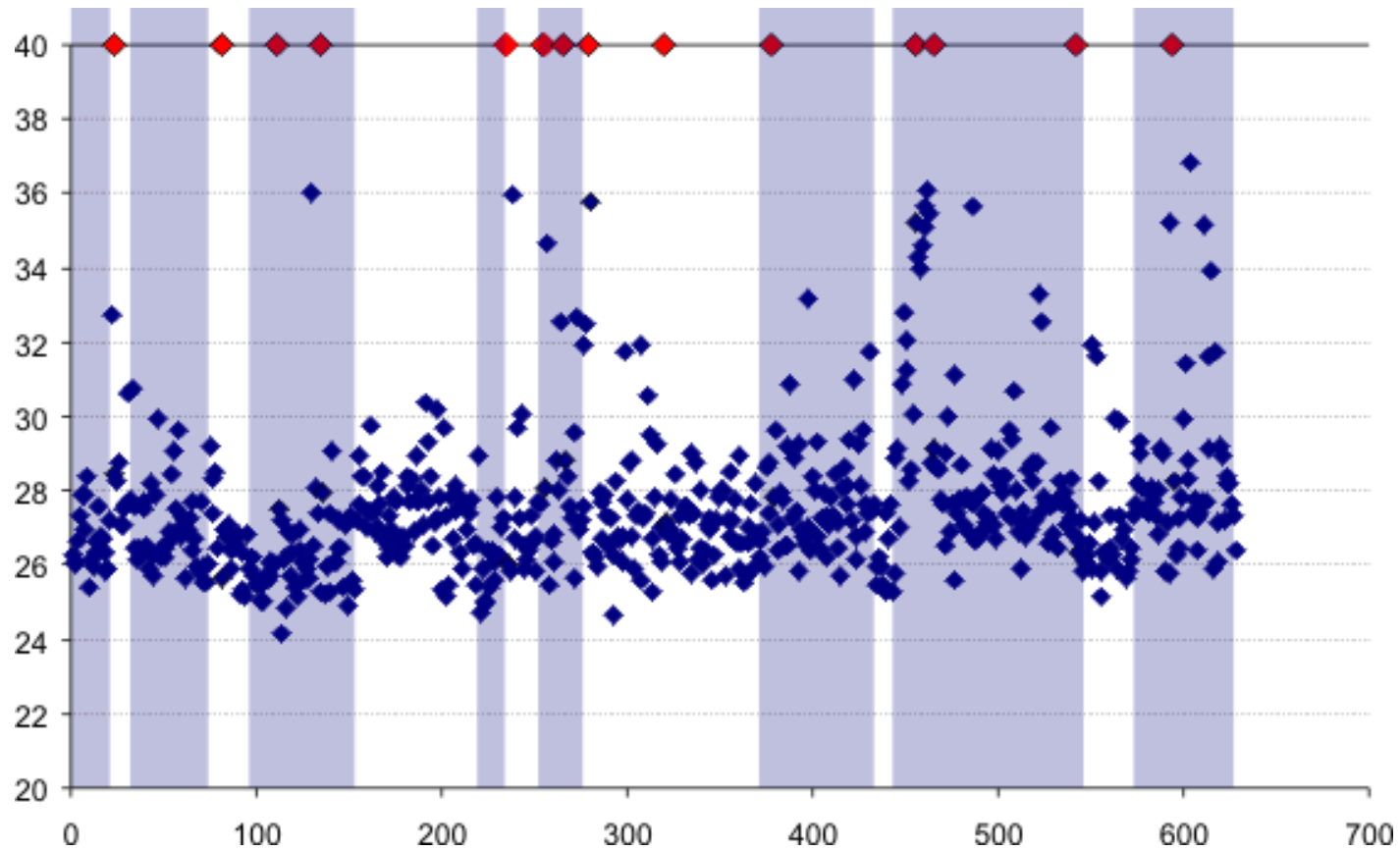
- reducing spread = improving equimolarity
- develop assays with little variation in performance
 - NXTGNT / CMGG primer desing pipeline

high-throughput primer design



	1	2	3
% drop out	3%	6%	0%
stdev Cq	3.1	4.0	2.1
average Cq	29.0	27.7	25.5

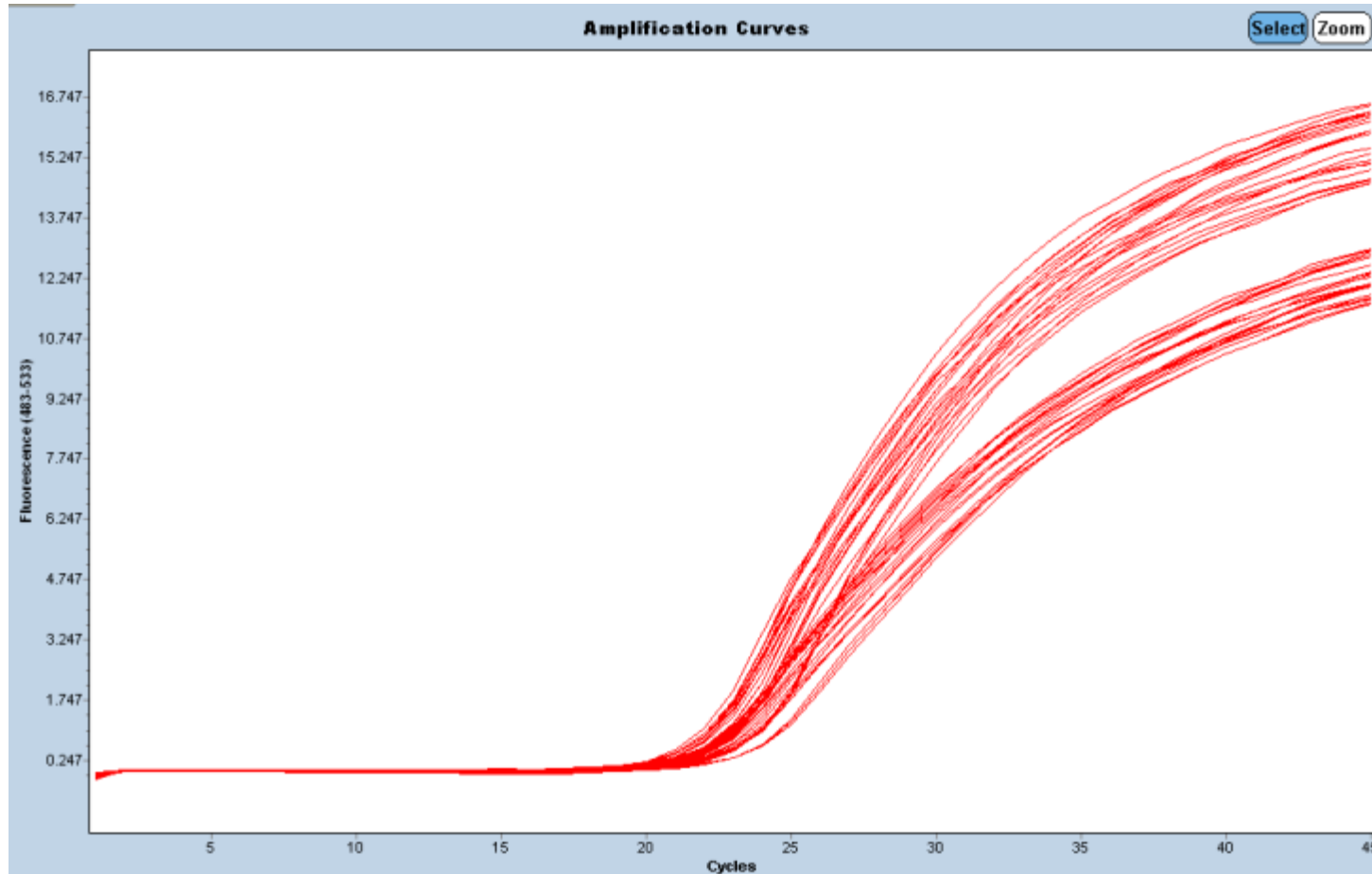
high throughput primer design



reduce spread factor

- reducing spread = improving equimolarity
- develop assays with little variation in performance
 - NXTGNT / CMGG primer desing pipeline
 - automated high-throughput
 - 610/628 succesful PCRs (97 %)
 - 543/610 PCRs with similar PCR quantities (89%)
 - 610 singleplex PCRs → $F_{90} = 2.5$
→ mean coverage = $40 * 2.5 = 100$
- normalize PCR products
 - monitor reactions with qPCR
 - determine end point fluorescence using a saturating dye
 - mix PCR products equimolarly

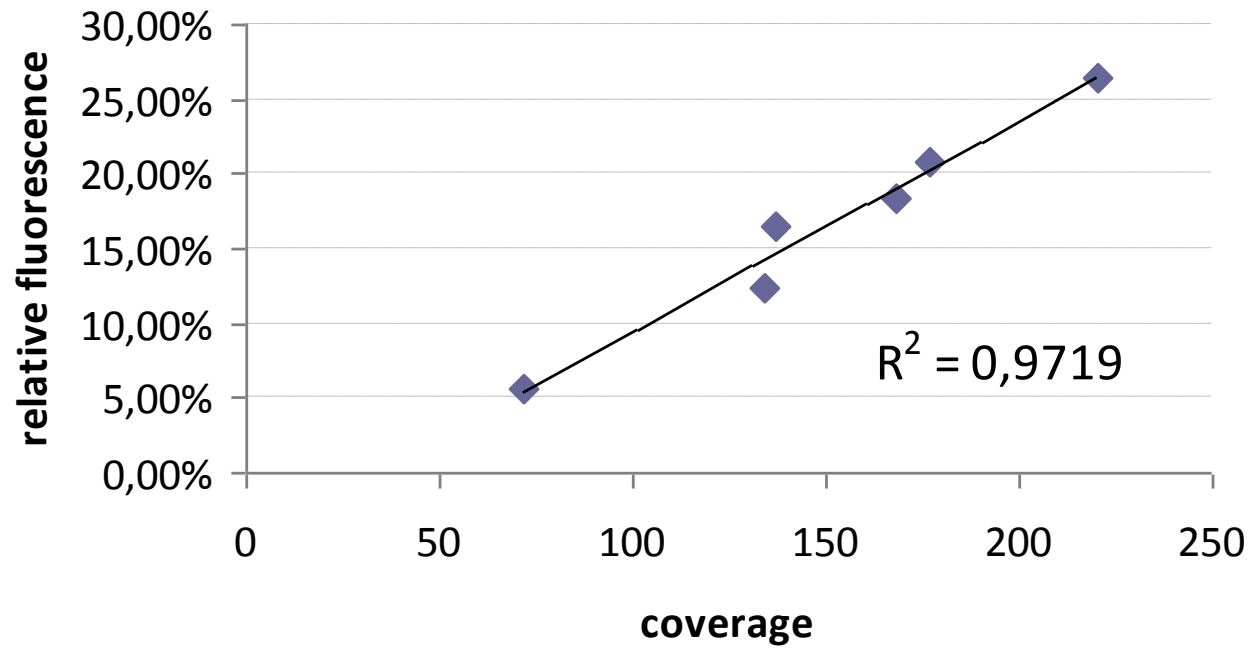
PCR products normalization



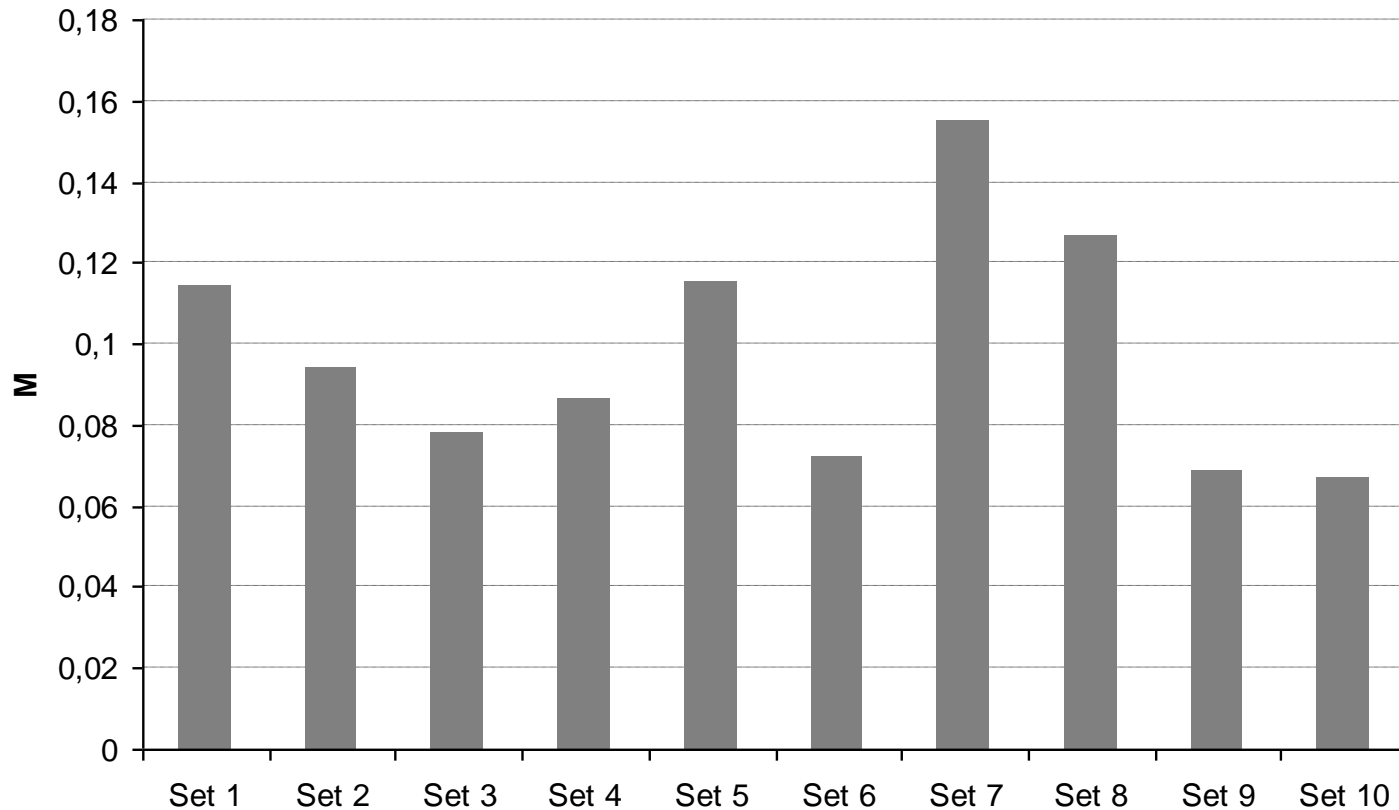
6.5 µl

8 µl

NGS retains equimolarity



NGS retains equimolarity

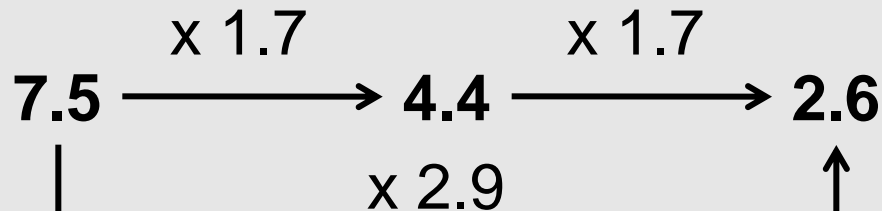
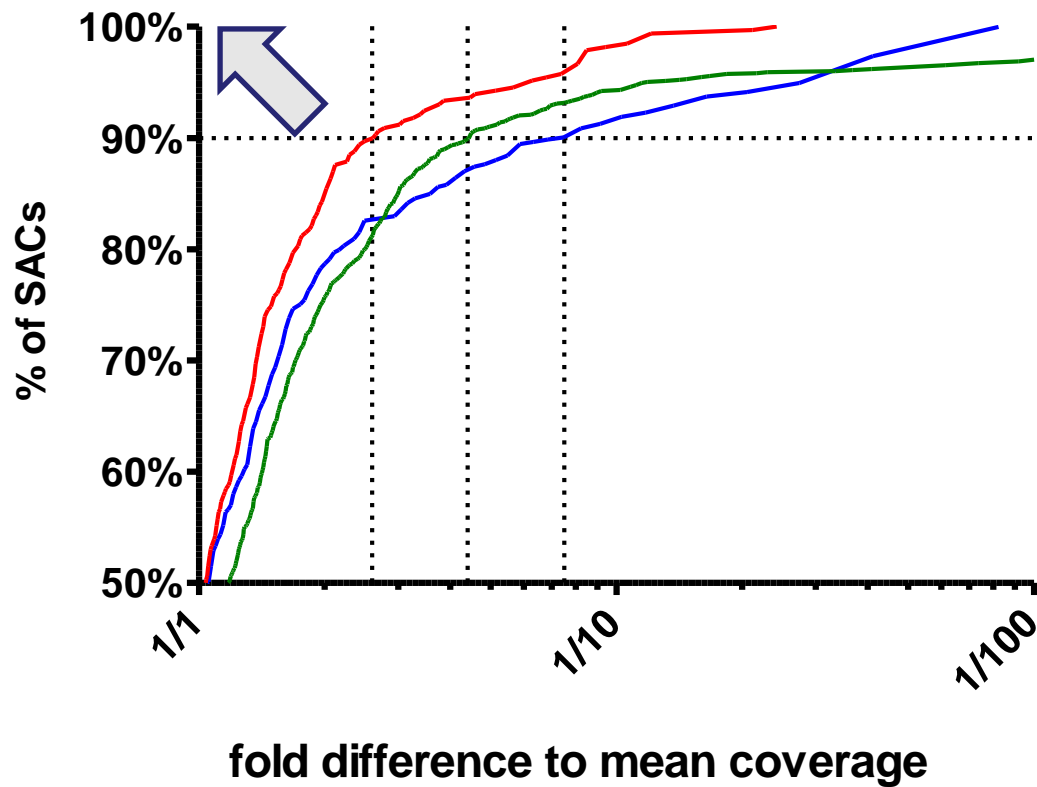


- average $M \approx 0.1$
- similar to random variation between $\pm 25\%$

multiplexing

- Multiplexing increases spread factor?
 - without optimization: $F_{90} = 7.5$
 - after optimization: $F_{90} = 4.4 - 2.5$
 - Focus on outliers and repeat failed reactions
→ spread factor (F_{90}) may reduce to 2

evolution of spread factor



BRCA example

- 16 multiplex PCRs
- required average coverage = $40 * 2.5 = 100$
- required reads per patient = $100 * 112 = 11.200$
- 25% safety margin → 14.000 reads / patient
- GS-FLX
 - standard (400.000 reads): 28 patients
 - Titanium (1.000.000 reads): 71 patients
 - goal: 100 patients/run → reduce required coverage to 30
→ reduce spread factor to 1.8
- TAT (no filling / clean up): <2 weeks
- sensitivity & specificity: 100 % & 98 %
 - homopolymers ≥ 8 can't be called correctly

Conclusions

- next generation molecular diagnostics is feasible
- evaluate the performance of an assay
 - drop out fraction
 - spread factor
 - specificity and sensitivity
- improve cost efficiency
 - all PCRs amplify similarly at the same reaction conditions
 - invest in multiplexing?
 - normalize PCR products
 - reduce spread factor
 - software
 - automate analysis
 - reduce error rate and minimal required coverage
- Remaining problems
 - homopolymer related errors
 - software for routine use

acknowledgements



- Joachim De Schrijver
- Kim De Leeneer
- Machteld Baetens
- Sarah De Keulenaer

- Hendrik Van de Voorde
- Jean-Pierre Renard