

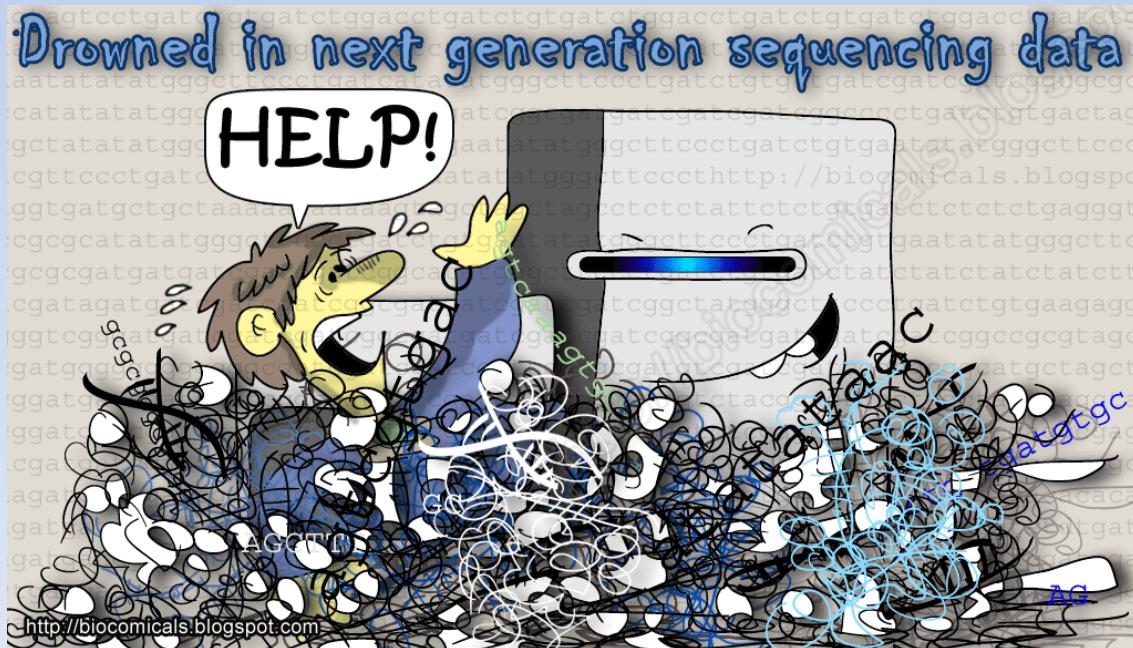
Current Methodologies for Targeted Sequencing

Mike Quail

WGC AC: Next Gen Sequencing

What is Targeted Sequencing??

“Targeted resequencing is a variation of re-sequencing where only a small subset of the genome is sequenced, such as the exome, a particular chromosome, a set of genes or a region of interest”

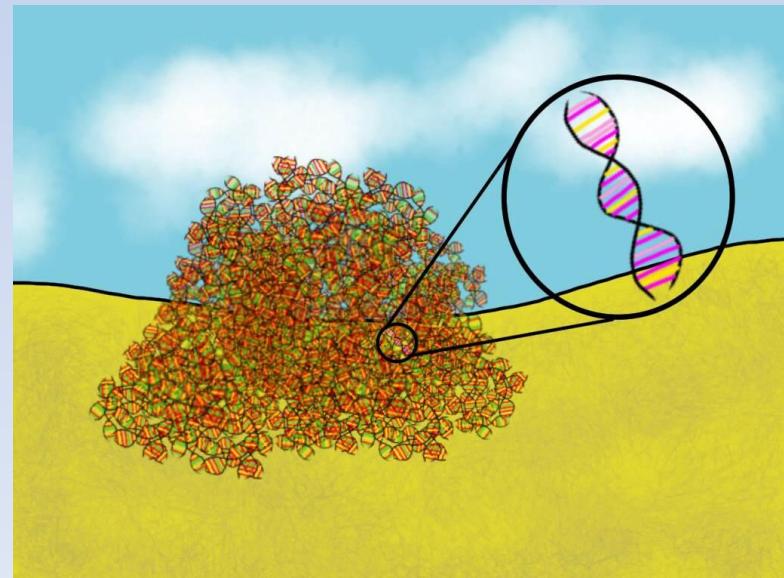


Overview

- Why??
- How??
 - Hybridisation capture
 - Amplicon
- Applicability depends upon project
- Examples

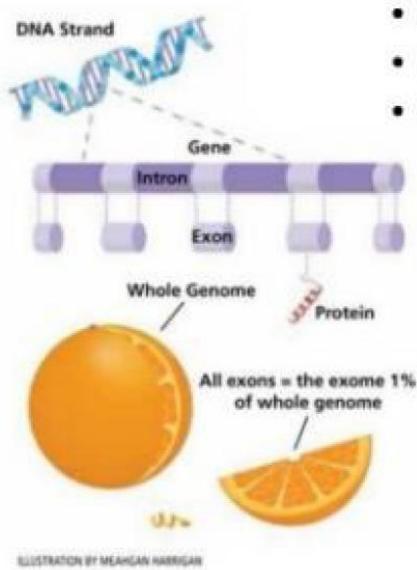
Why Targeted Sequencing?

- Cheaper £££££
- Diagnostic – particular diseases
- Detect low levels of variation
- Genome-wide association studies (GWAS) instead of arrays
 - Rare alleles
- Genotyping



Exome sequencing

- Each person has 23 pairs of chromosomes
- Human genome contains $\sim 3 \times 10^9$ bases (20 000 genes)
- DNA sequencing of all bases = whole genome sequencing (WGS)



- Approximately 1% is coding sequence – the exome
- ~85% of disease-causing mutations reside in the exome
- Whole exome sequencing (WES) has provided a cost-effective and practical means of identifying pathogenic variants

WGS	WES
$\sim 3 \times 10^9$ bases	$\sim 3 \times 10^7$ bases
£100-£200	£50
90 Gb	4-5 Gb
30X	>50X

Why is targeted sequencing popular?

- Lower costs
 - £100-£200 WGS (90Gb)
 - £50 WES (5Gb)
 - £10 few loci (<0.1Gb)
- Increased sensitivity
- Simplified analysis workflow
- Diagnostics
- A few examples...

Cost

Target	Size	Data	Cost
Human Genome	3Gb	90Gb	£100-£200
Human Exome	50Mb	5Gb	£50
Custom panel	<1Mb	0.5Gb	£100

Applications

Diagnostic

[FOR PATIENTS & FAMILIES](#)[CLINICS AND SERVICES](#)[FOR HEALTH CARE PROFESSIONALS](#)[ABOUT US](#)[GIVE TO CMH](#)[FIND A DOCTOR](#)

[Home](#) > [Health Care Professionals](#) > [Research](#) > [Pediatric Genomic Medicine](#)

PEDIATRIC GENOMIC MEDICINE TaGSCAN Diagnostic Test

Pediatric Genomic Medicine

- Services

[TaGSCAN Diagnostic Test](#)[Exome Sequencing](#)[STAT-Seq](#)[Whole Genome Sequencing](#)

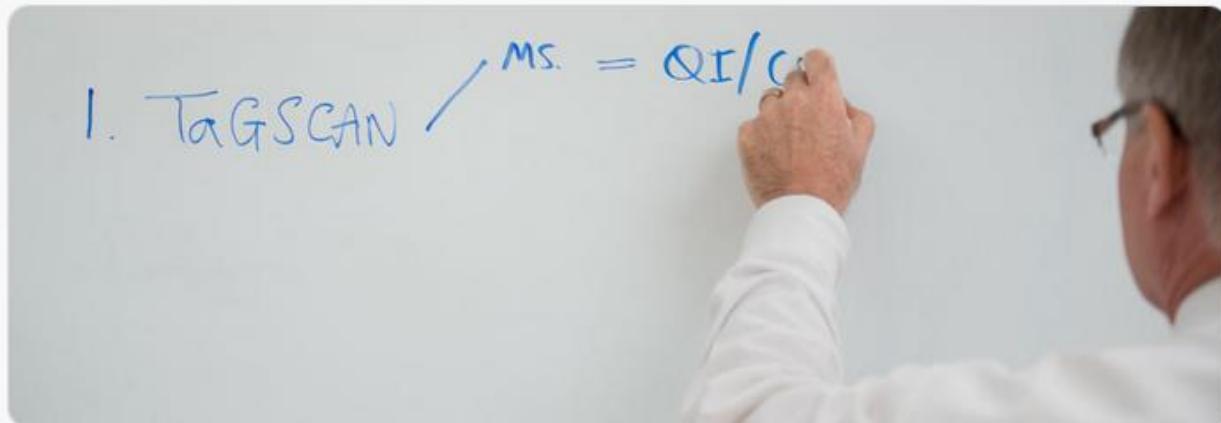
Our Staff

Software Tools

Facilities

Publications

News and Features



The Targeted Gene Sequencing and Custom Analysis (TaGSCAN) test is a novel, targeted screening panel covering the coding region of 514 genes known to cause severe diseases with childhood onset. TaGSCAN was developed by the Center for Pediatric Genomic Medicine at Children's Mercy Hospitals and Clinics.



CONTACT US

Phone: (816) 234-3059

Email

Billing and Reimbursement

The cost is comparable to a single gene test done by Sanger sequencing and is especially cost effective when considering a genetically heterogeneous disorder for which two or more genes are on the panel. Like all molecular genetic testing, insurance coverage may be variable. CPT code: 81407

Diagnostic

GENETICS IN MEDICINE | ORIGINAL RESEARCH ARTICLE



Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing

Mark B. Consugar MS, Daniel Navarro-Gomez BS, Emily M. Place MS, Kinga M. Bujakowska PhD, Maria E. Sousa BA, Zoë D. Fonseca-Kelly PhD, Daniel G. Taub BA, Maria Janessian BA, Dan Yi Wang PhD, Elizabeth D. Au PhD, Katherine B. Sims MD, David A. Sweetser MD, Anne B. Fulton MD, Qin Liu PhD, Janey L. Wiggs MD, PhD, Xiaowu Gai PhD & Eric A. Pierce MD, PhD

Affiliations | Corresponding author

Genetics in Medicine (2015) 17, 253–261 | doi:10.1038/gim.2014.136

Received 08 July 2014 | Accepted 23 October 2014 |

Results:

The GEDi test is highly reproducible and accurate, with sensitivity and specificity of 97.9 and 100%, respectively, for single-nucleotide variant detection. The sensitivity for variant detection was notably better than the 88.3% achieved by whole-exome sequencing using the same metrics, because of better coverage of targeted genes in the GEDi test as compared with a commercially available exome capture set. Prospective testing of 192 patients with inherited retinal degenerations indicated that the clinical sensitivity of the GEDi test is high, with a diagnostic rate of 51%.

Conclusion:

Based on quantified performance metrics, the data suggest that selective targeted enrichment is preferable to whole-exome sequencing for genetic diagnostic testing.

Disease Prediction

ACCEL-AMPLICON® BRCA1 AND BRCA2 PANEL

Targeted Sequencing for Breast Cancer Genes

The Accel-Amplicon BRCA1 and BRCA2 Panel offers comprehensive coverage of the entire coding sequence of *BRCA1* and *BRCA2* genes. Utilizing 246 amplicons with average size of 148 bp, this panel generates targeted libraries compatible with Illumina® and Ion Torrent sequencing platforms. The unique design of the Accel-Amplicon BRCA1 and BRCA2 Panel enables compatibility with FFPE and cfDNA samples. This product is a complete kit that includes all components necessary for generating ready-to-sequence libraries, including primer pairs and indexed sequencing adapters.

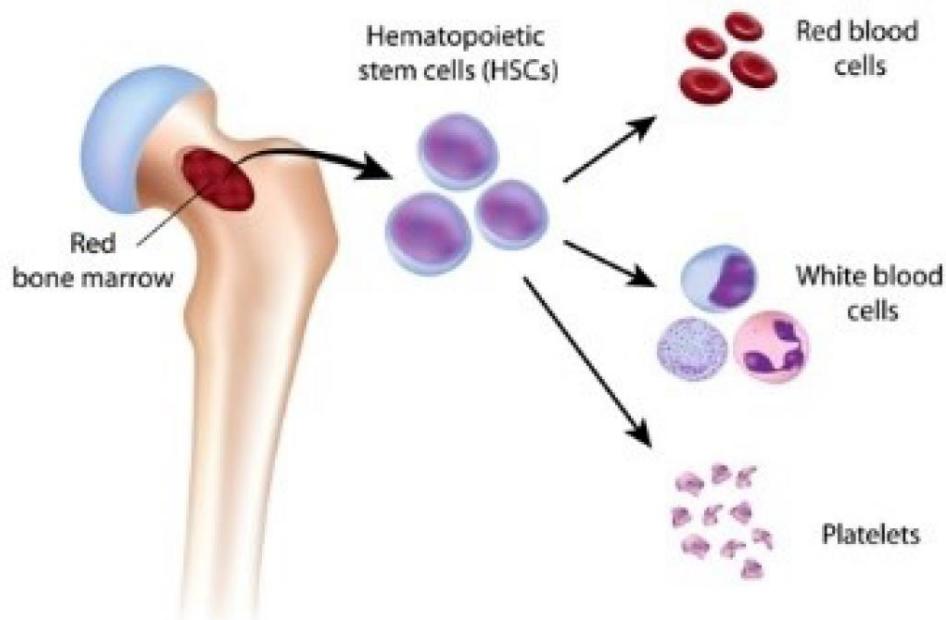


BROCHURE

[DOWNLOAD PDF](#)

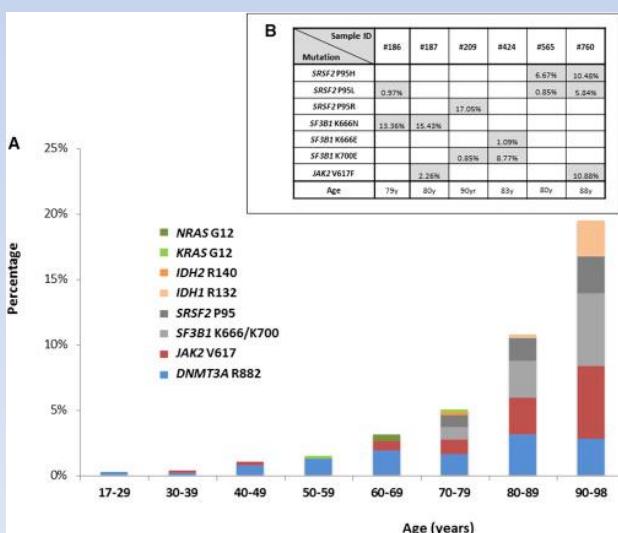
Minimal residual disease

Low Allele Frequencies



Low level variants

The screenshot shows the Cell Reports website interface. At the top, there's a navigation bar with links for 'Explore', 'Online Now', 'Current Issue', 'Archive', 'Journal Information', and 'For Authors'. Below the navigation bar, the journal title 'Cell Reports' is prominently displayed. To the right of the title, there are buttons for 'All Content' and 'Cell Reports'. The main content area displays the article details: 'Volume 10, Issue 8, p1239–1245, 3 March 2015'. The article title is 'Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis'. The authors listed are Thomas McKerrell^{1,3}, Naomi Park^{1,3}, Thaidy Moreno, Carolyn S. Grove, Hannes Pongstingl, Jonathan Stephens, Understanding Society Scientific Group, Charles Crawley, Jenny Craig, Mike A. Scott, Clare Hodkinson, Joanna Baxter, Roland Rad, Duncan R. Forsyth, Michael A. Quail, Eleftheria Zeggini, Willem Ouwehand, Ignacio Varela, George S. Vassiliou. The note '13 Co-first author' is present. The article summary discusses the development and validation of a methodology for detecting mutation-associated circulating blood cell clones.



detailed age distributions). To do this, we developed and validated a robust methodology, employing barcoded multiplex PCR of mutational hot spots followed by next-generation sequencing (MiSeq) and bioinformatic analysis, to extract read counts and allelic fractions for reference and non-reference nucleotides. This reliably detected mutation-associated circulating blood cell clones with a variant allele fraction (VAF) ≥ 0.008 (0.8%; see [Supplemental Experimental Procedures](#) and [Figure S2](#)).

Genome Wide Association Studies (GWAS)

- GWAS based on common SNPs
 - contribution of 1000's of loci to risk for common diseases and to variation in quantitative traits.

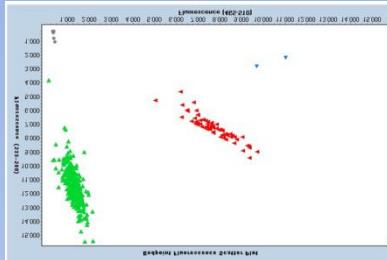
RESEARCH

Open Access

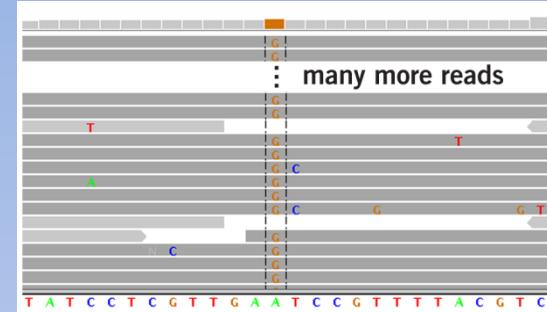


Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants

Anthony J. Griswold¹, Nicole D. Dueker¹, Derek Van Booven¹, Joseph A. Rantus¹, James M. Jaworski¹, Susan H. Slifer¹, Michael A. Schmidt¹, William Hulme¹, Ioanna Konidari¹, Patrice L. Whitehead¹, Michael L. Cuccaro^{1,2}, Eden R. Martin^{1,2}, Jonathan L. Haines³, John R. Gilbert^{1,2}, John P. Hussman⁴ and Margaret A. Pericak-Vance^{1,2*}



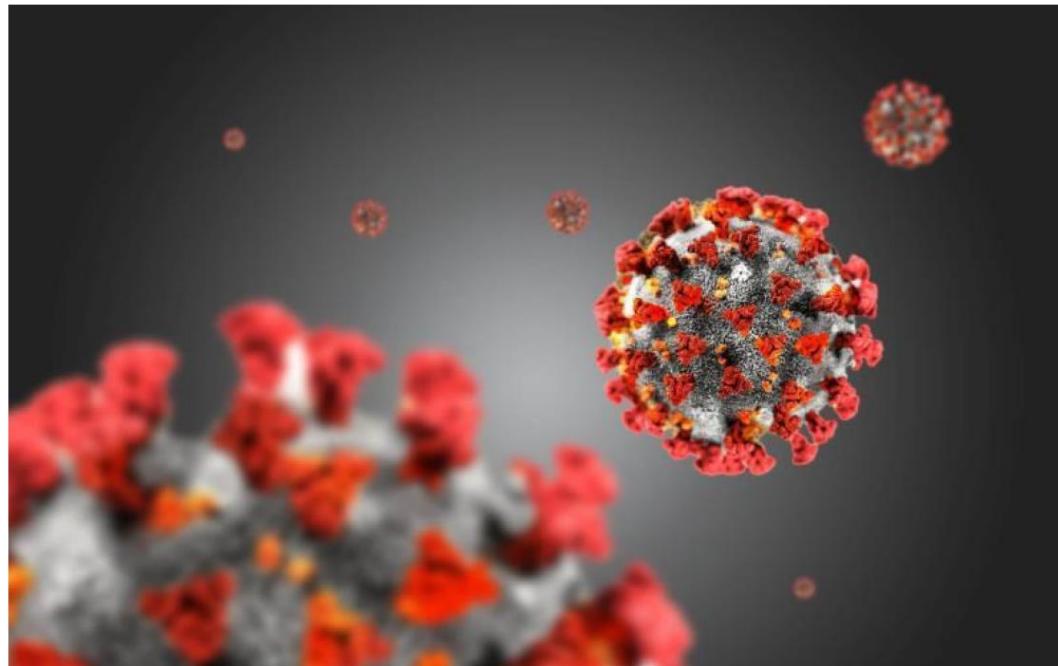
Genotyping

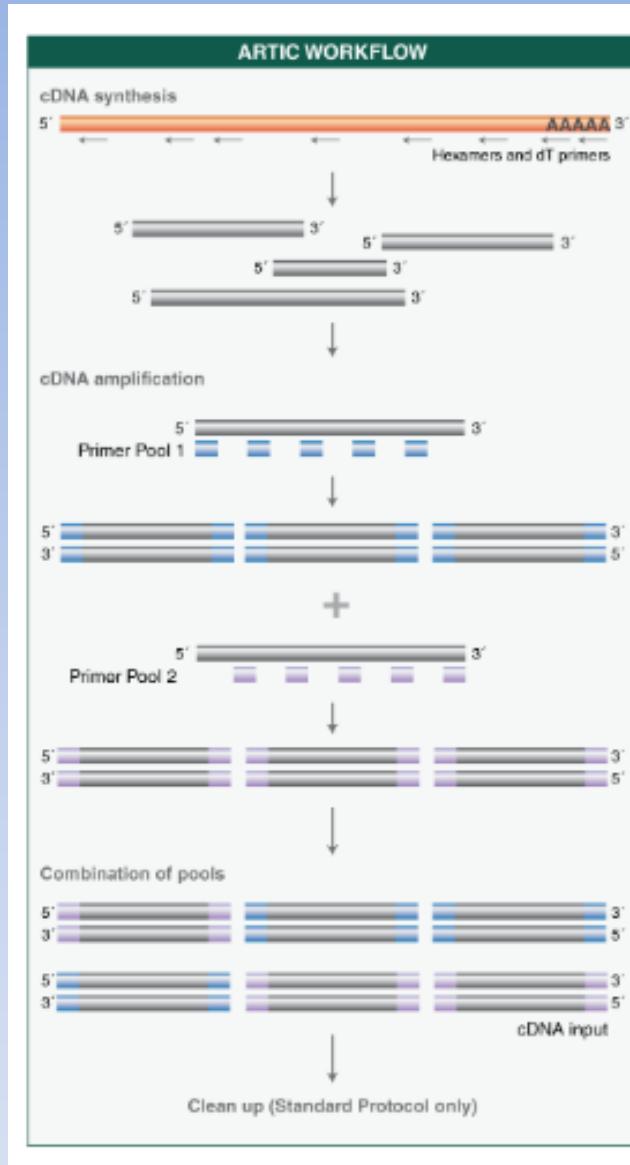


- Multiplex enrichment PCR
 - PCR primers designed to amplify the targets of interest (known SNPs)
- Restriction enzyme (RE) digestion
 - Complexity of genome reduced by restriction digestion
 - little genome sequence is available or no prior SNP variants have been identified

PCR based approaches can be very sensitive eg Covid-19 ARTIC method

Targeted genome

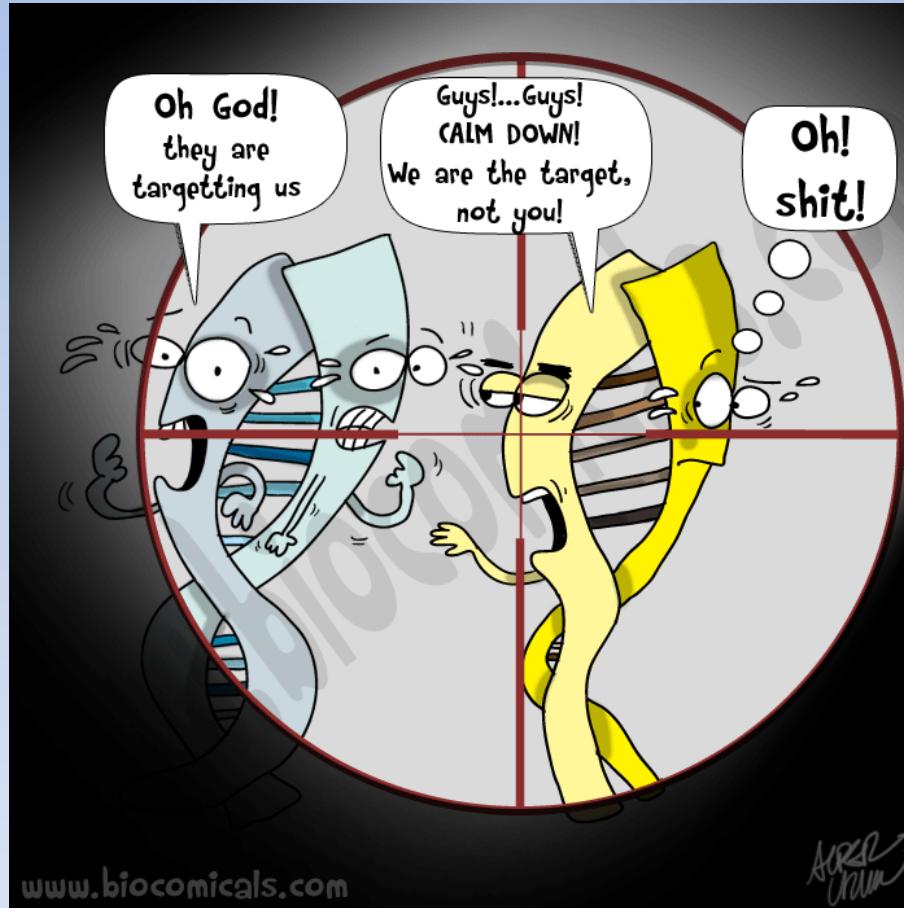




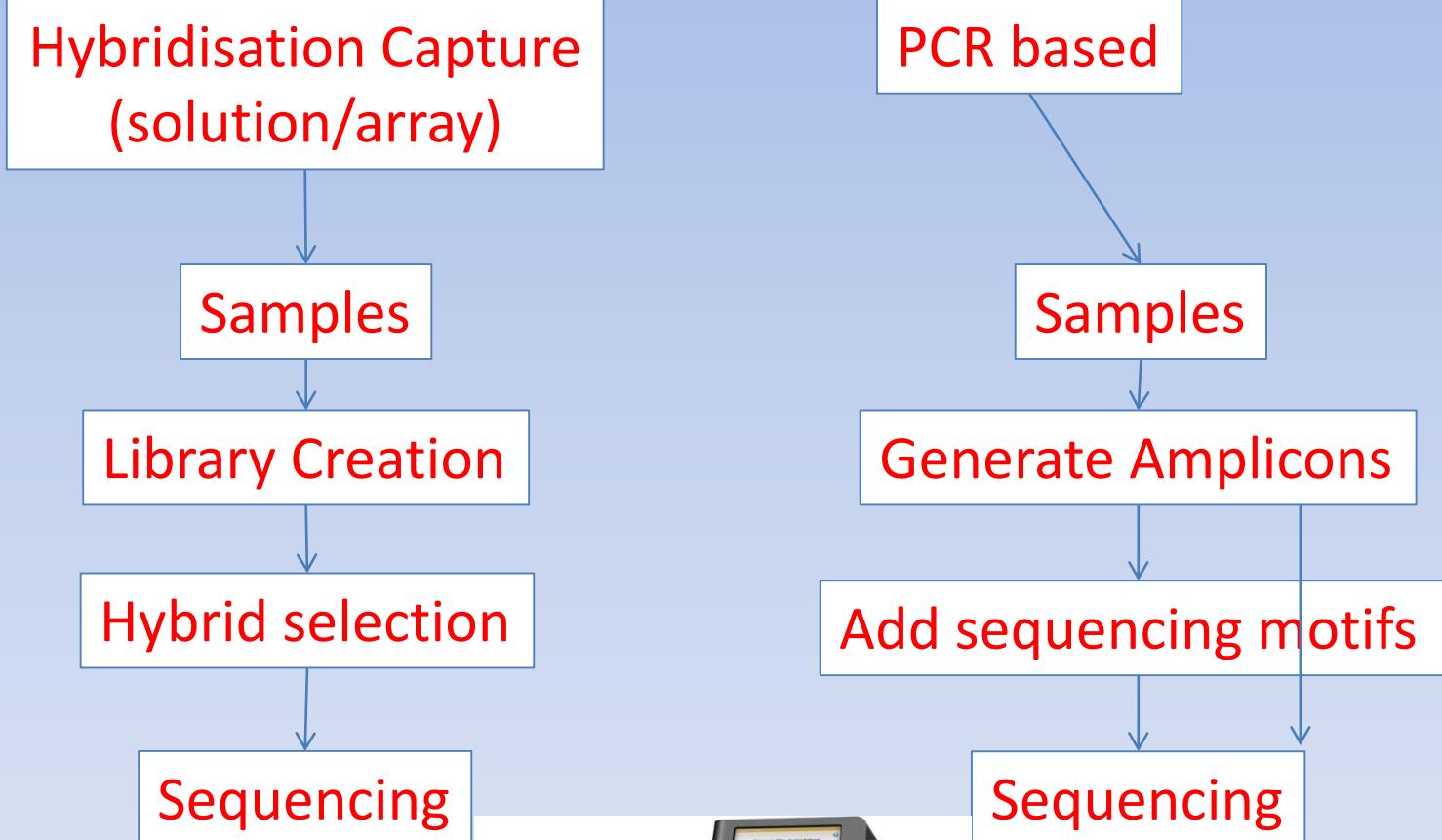
* Can also be used with single barcodes



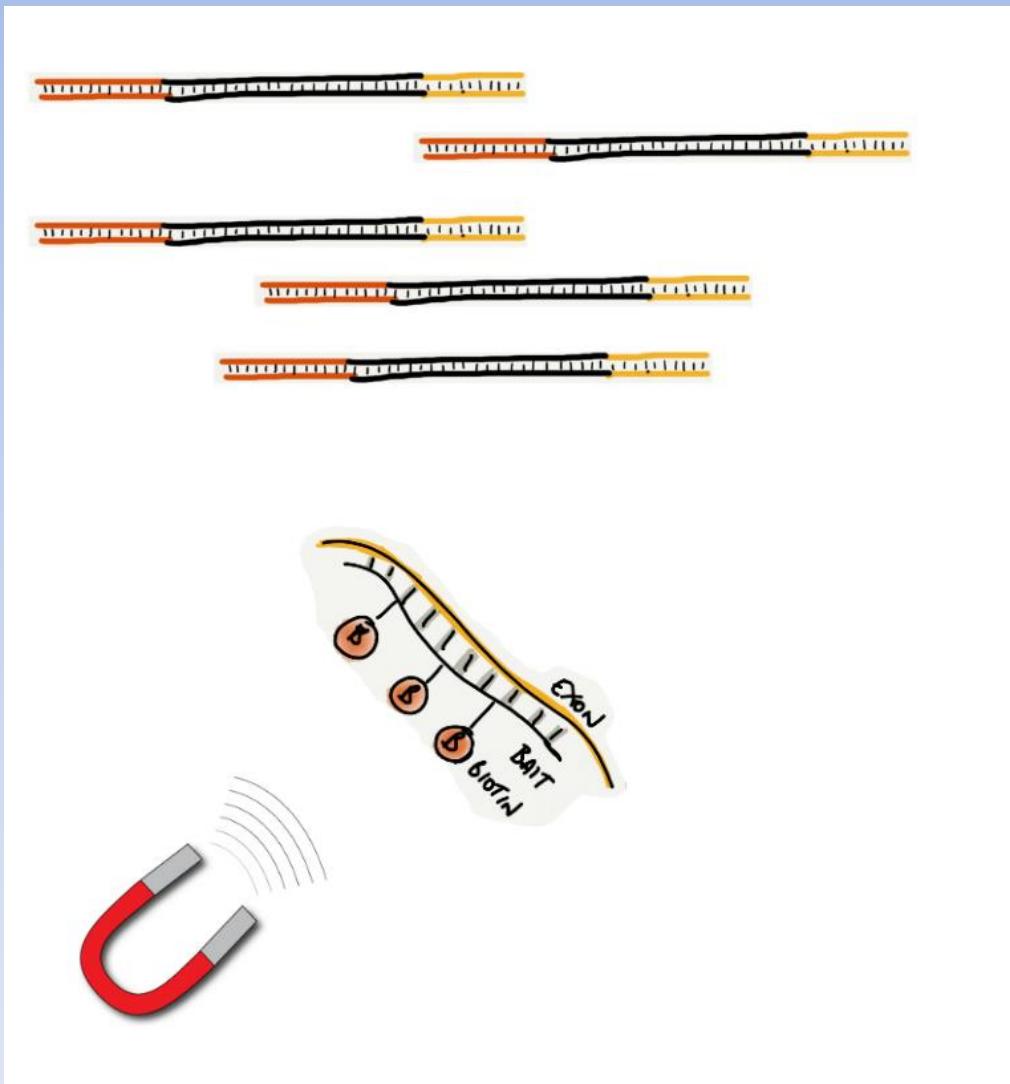
Targeted Sequencing?



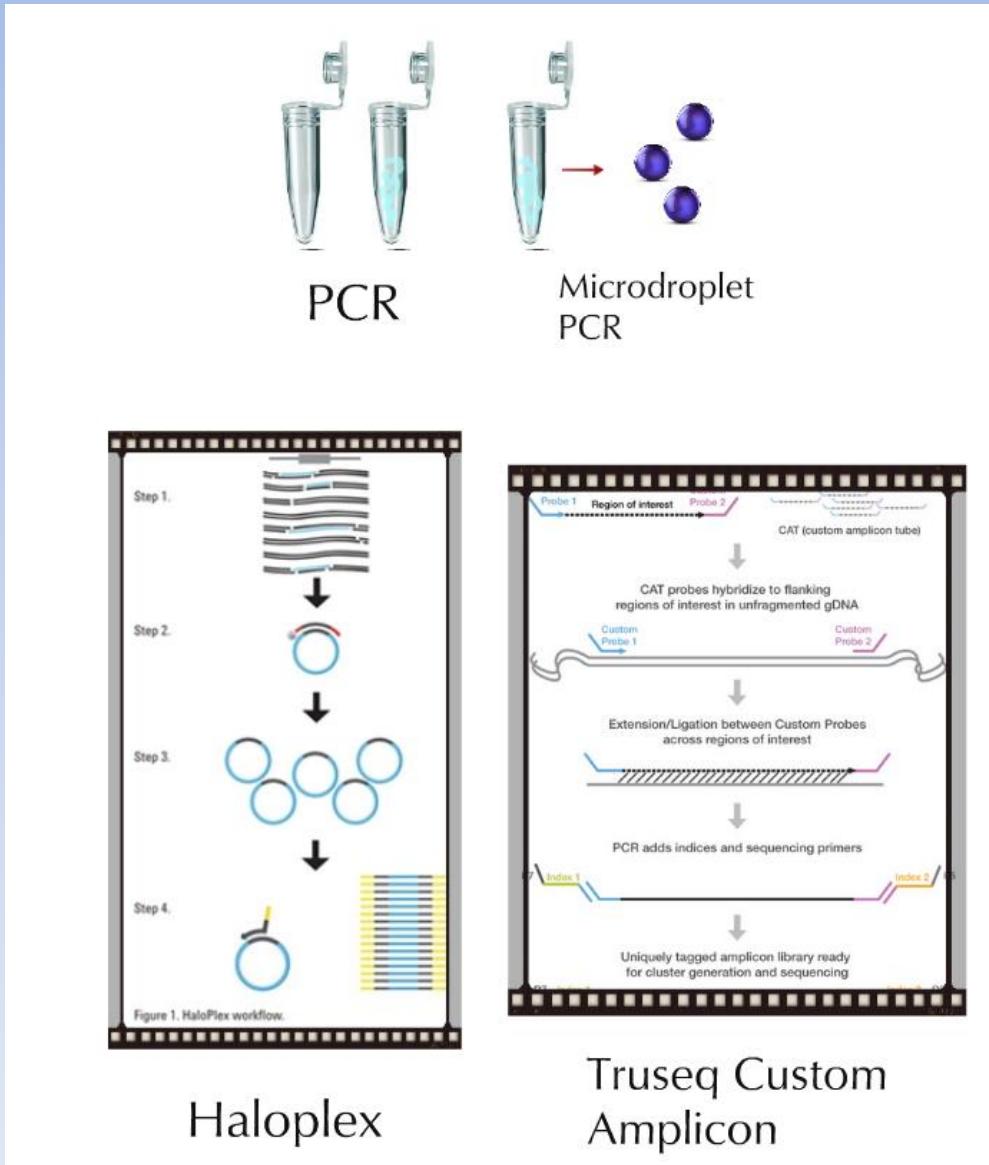
How?



Hybrid Capture



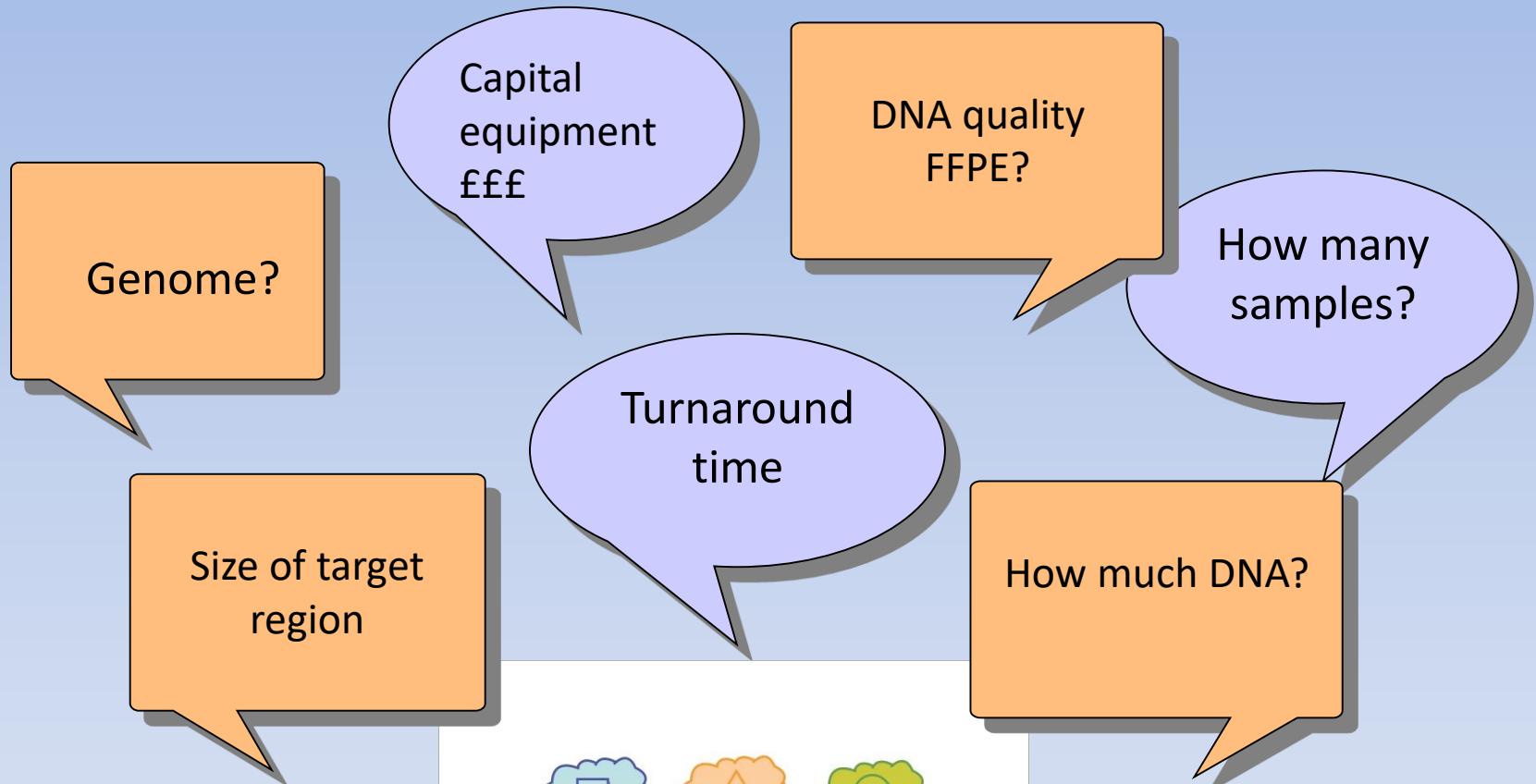
PCR based capture



Which Technology?

Method	Platform	Company	Capital equipment?	Target Size
Non-commercial	Multiplex PCR	n/a	N	<0.02Mb
	Molecular Inversion Probes	n/a	N	0.02Mb-1.5Mb
Commercial Amplicon	rhAmpSeq	IDT	N	Upto 0.2Mb
				upto 5Mb (custom)
	Haloplex	Agilent	N	32Mb (exome)
	Ampliseq	Ion Torrent	N	upto 33Mb
Hybridisation	NEB Next Direct	NEB	N	Upto 0.3Mb
	Sure Select	Agilent	N	0.001-100Mb
	TruSeq	Illumina	N	upto 67Mb
	EZ-Cap	Nimblegen	N	upto 200Mb
	X-Gen Lockdown	IDT	N	upto 60Mb
	Target enrichment	Twist	N	Upto 50Mb
	Target enrichment	Celemics	N	Upto 10Mb

Which Technology?

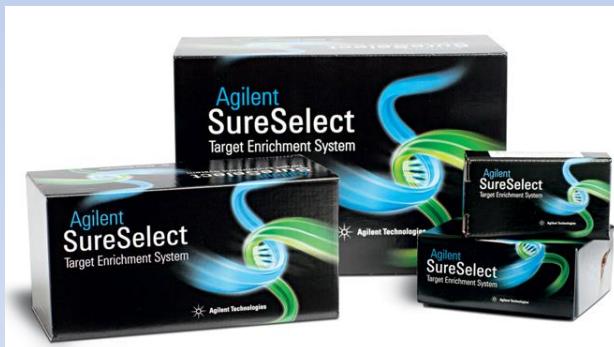


We are all in agreement then.

Size of target region

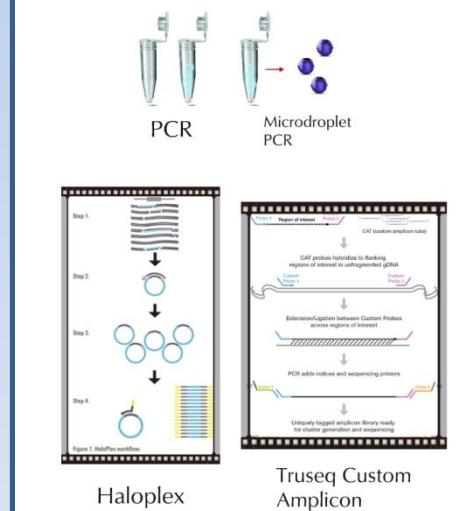
Whole exome
~60Mb (X30)

Focused Exome
(i.e. disease associated target panel) ~16Mb (X200)



- Limited sample multiplexing
- Coverage per target
- Bioinformatics

Custom Small region (X1000)



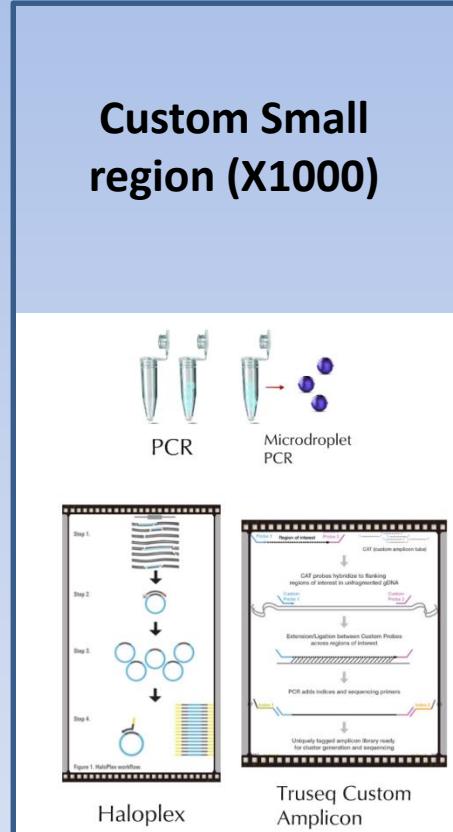
Which Technology?

Turnaround time

Design

Optimisation

Commercial / non-commercial



Capital equipment £££

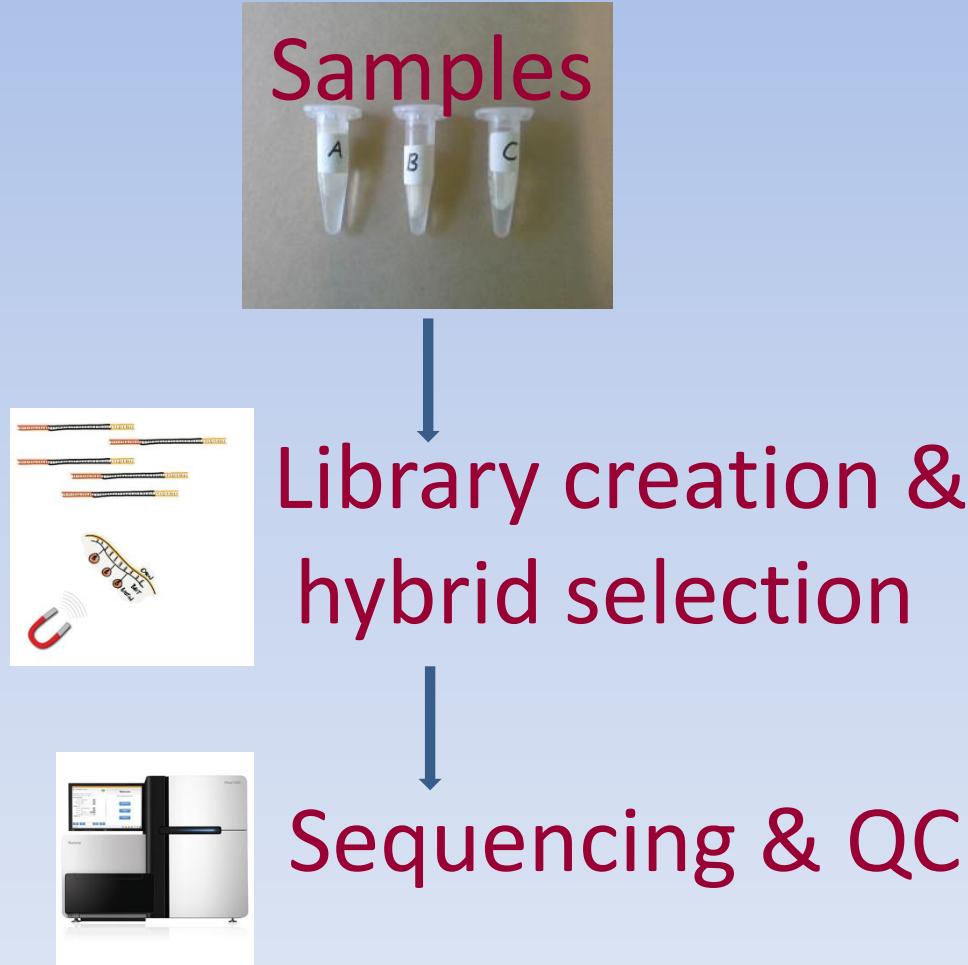
How many samples?
Kits X16/X96

1000's samples????

Which Technology?

Method	Platform	Company	Capital equipment?	Target Size
Non-commercial	Multiplex PCR	n/a	N	<0.02Mb
	Molecular Inversion Probes	n/a	N	0.02Mb-1.5Mb
Commercial Amplicon	Ampliseq	Ion Torrent	N	upto 33Mb
				upto 5Mb (custom) 32Mb (exome)
Hybridisation	Sure Select	Agilent	N	0.001-100Mb
	TruSeq	Illumina	N	upto 67Mb
	HyperExome	Roche	N	upto 200Mb
	X-Gen Lockdown	IDT	N	upto 60Mb
	Target enrichment	Twist	N	Upto 50Mb
	Target enrichment	Celemics	N	Upto 10Mb

@Sanger – Agilent sureselect or Twist





Hybridisation Capture at Sanger

Post-cap PCR

Captured library is
seq ready . . .

Use p5/p7 → 10-16× PCR

OFF-BEAD

Agilent Bravo for hybrid capture automation

Liquid handling



Agilent NGS Workstation (x6)

Hybridisation Capture - Examples

- Whole Exome
- Pan Cancer
- Heterogeneous disorders
- Hearing loss with many causal genes
 - 200 to 250 genes may be involved in hereditary HL
- Intellectual disabilities
- Autism spectrum disorders
- Cardiovascular disease
- Obesity, diabetes
- Minor allele frequencies
 - enriched for functional mutations, potentially have medical relevance

Why Twist

- Fast turnaround
- Greater uniformity
- Cheaper custom panels

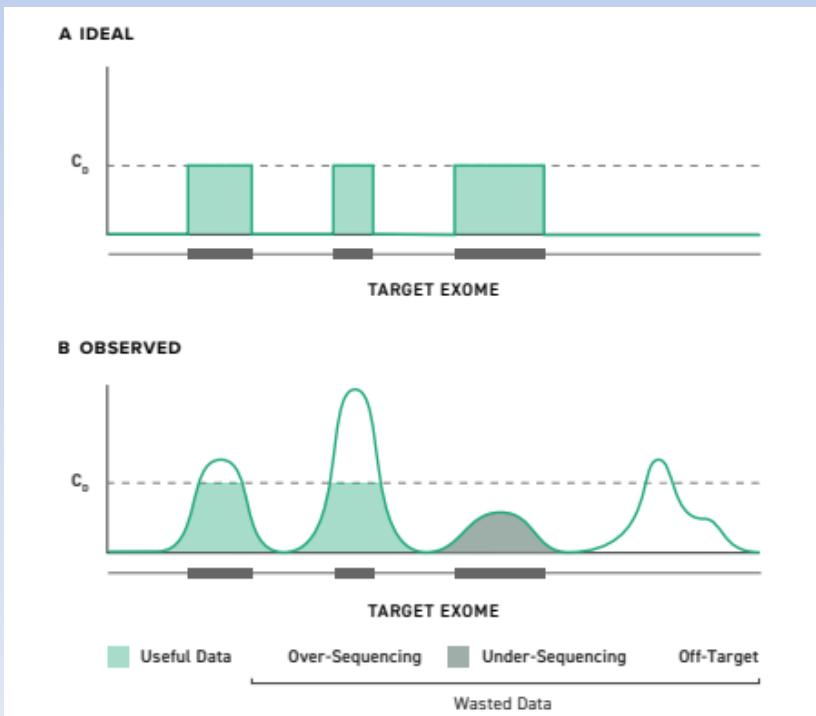


FIGURE 2

$$\text{Fold-80} = C_M / Q_{20}$$

where C_M is the mean depth
and Q_{20} is the 20th percentile

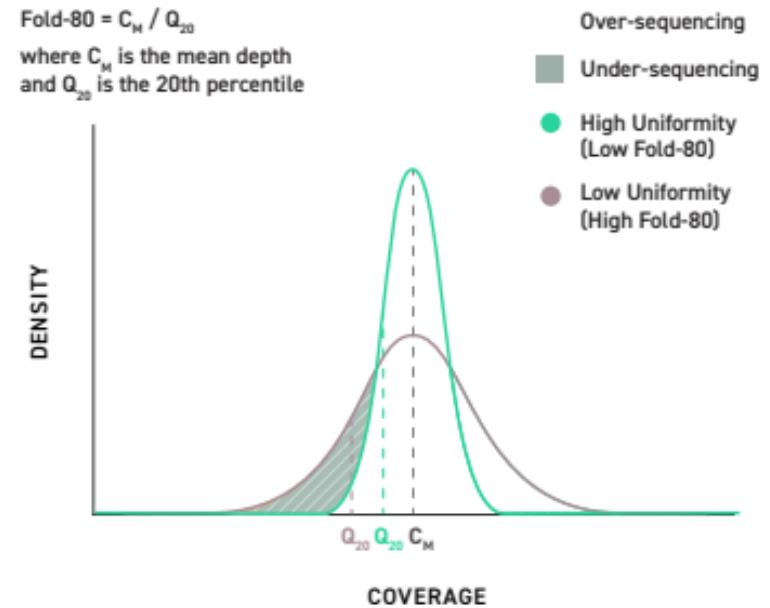
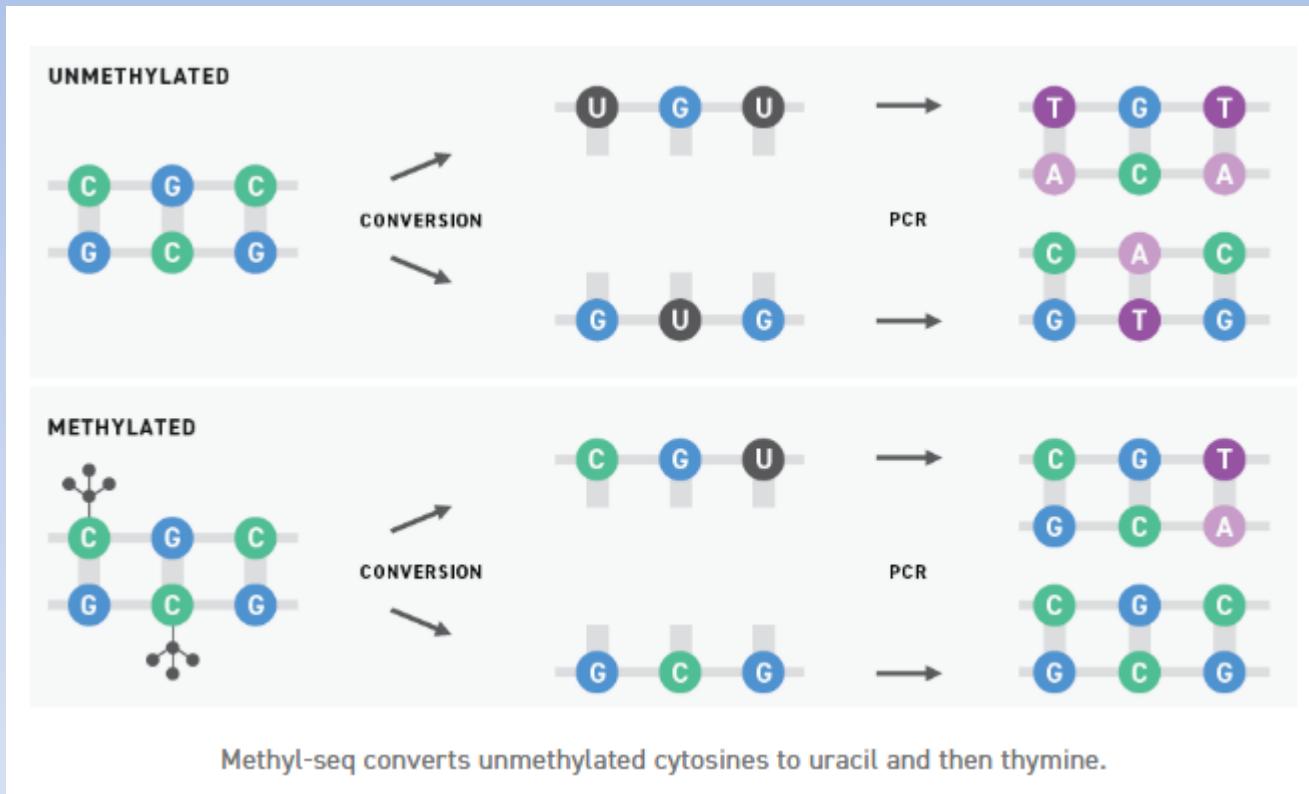


Figure 2. Uniformity reflects distribution shape. Two different hypothetical read distribution profiles showing low (green) and high (gray) fold-80 scores and the relative abundance of reads mapping back to over- and under-sequenced regions. Lowering the fold-80 score (gray curve to green curve) both rescues under-sequenced regions and reduces the fraction of over-sequenced regions for more efficient read utilization. In reality, poor uniformity often shows less symmetric distributions.

Twist hybrid capture for methylation

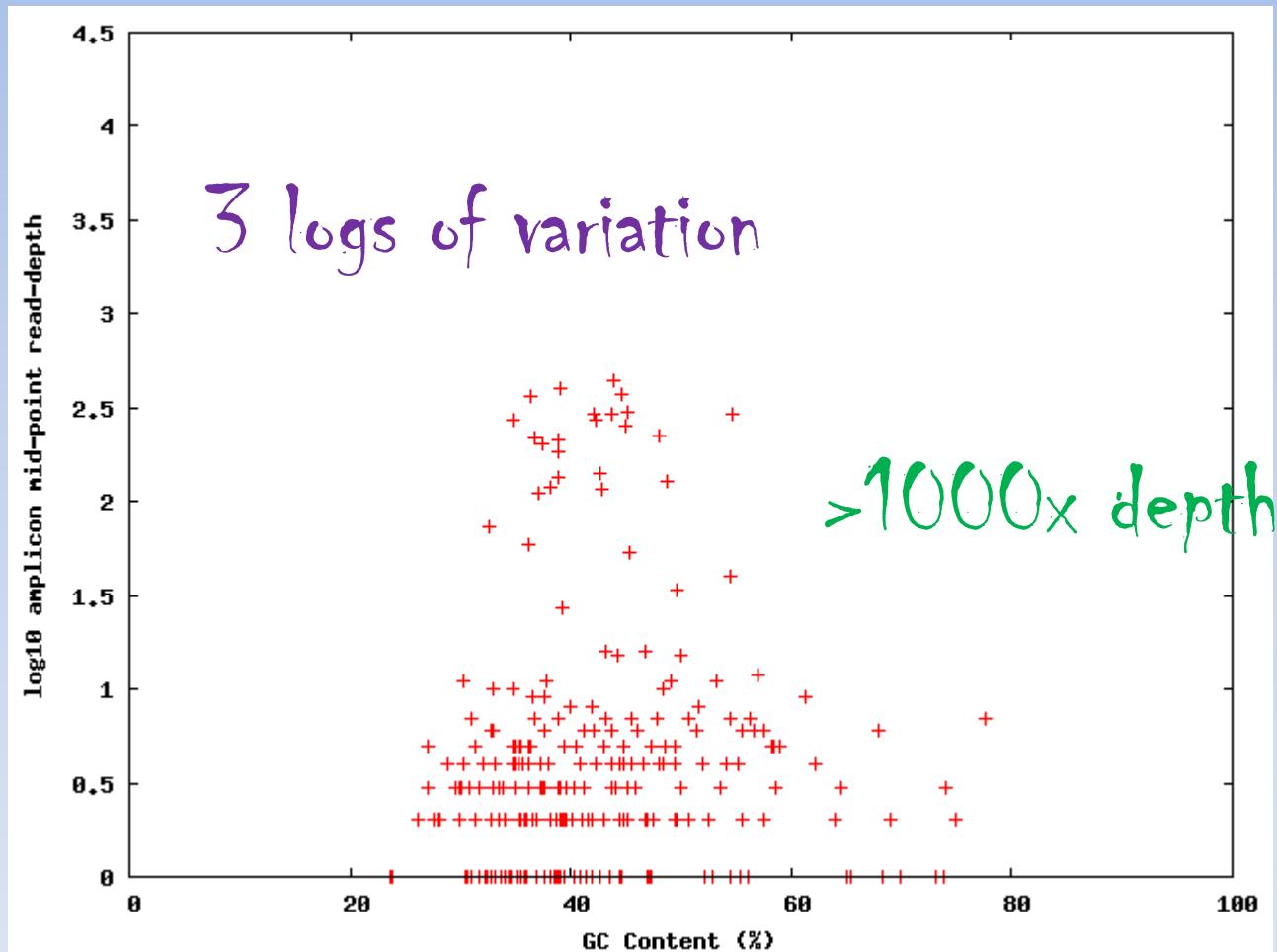


Twist has probes for both strands

Twist can produce more complex probe pools to cover all sequence combinations

Method	Platform	Company	Capital equipment?	Target Size
Non-commercial	Multiplex PCR	n/a	N	<0.02Mb
	Molecular Inversion Probes	n/a	N	0.02Mb-1.5Mb
Commercial Amplicon	rhAmpSeq	IDT	N	Upto 0.2Mb upto 5Mb (custom) 32Mb (exome)
	Haloplex	Agilent	N	
	Ampliseq	Ion Torrent	N	upto 33Mb
Hybridisation	NEB Next Direct	NEB	N	Upto 0.3Mb
	Sure Select	Agilent	N	0.001-100Mb
	TruSeq	Illumina	N	upto 67Mb
	EZ-Cap	Nimblegen	N	upto 200Mb
	X-Gen Lockdown	IDT	N	upto 60Mb
	Target enrichment	Twist	N	Upto 50Mb
	Target enrichment	Celemics	N	Upto 10Mb

Bias

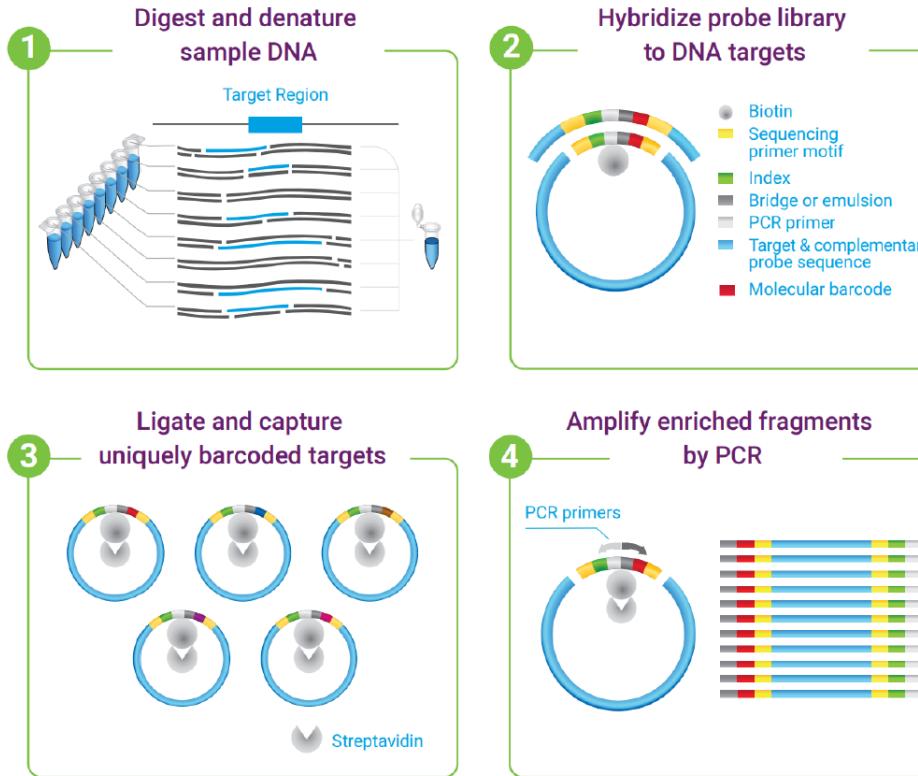


Agilent Haloplex

High sensitivity. Adds UMIs. Good for deep sequencing of small capture areas

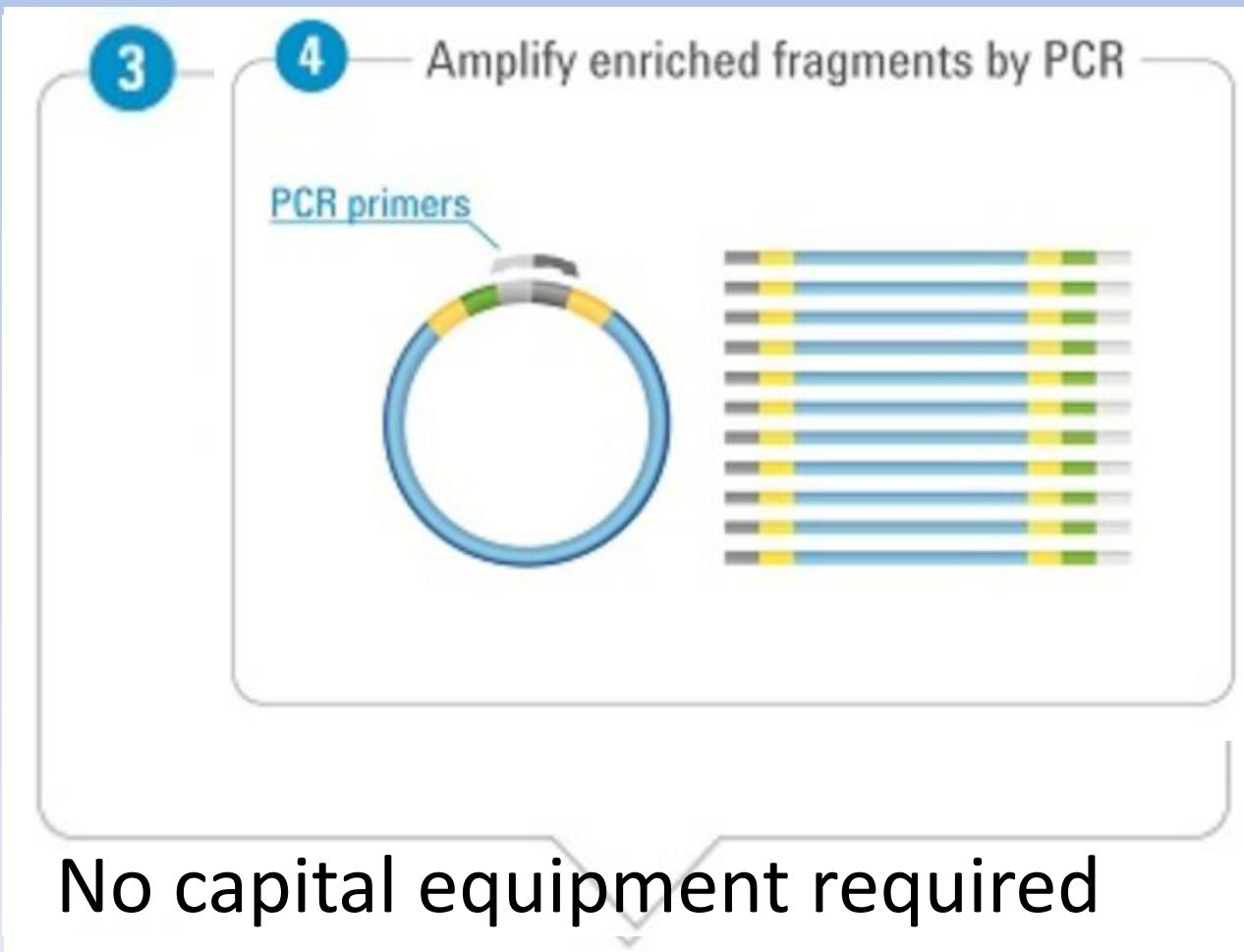
Commercial Amplicon

No capital equipment required

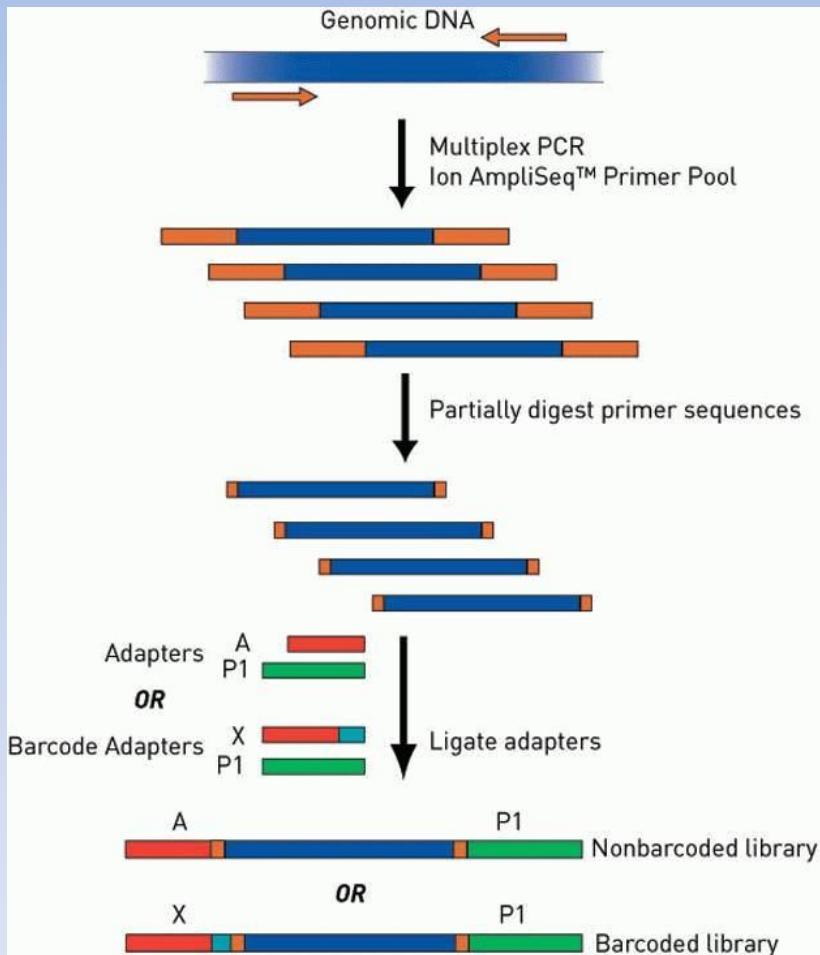


Commercial Amplicon

- Haloplex (16/48 samples)

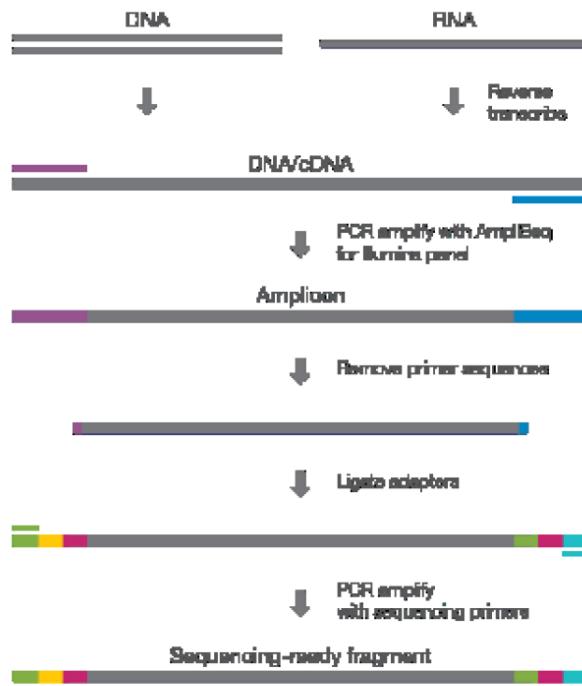


Ampliseq (Ion Torrent)



Commercial Amplicon

AmpliSeq for Illumina Custom DNA Panel

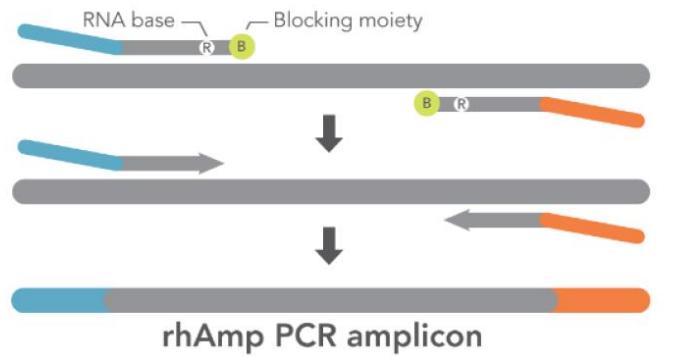


Commercial Amplicon - IDT

Targeted rhAmp PCR 1

Activation of rhAmp primers by RNase H2 cleavage

Amplification



Indexing PCR 2

Amplification with indexing primers

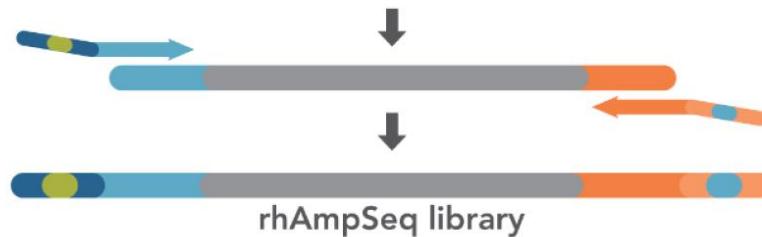
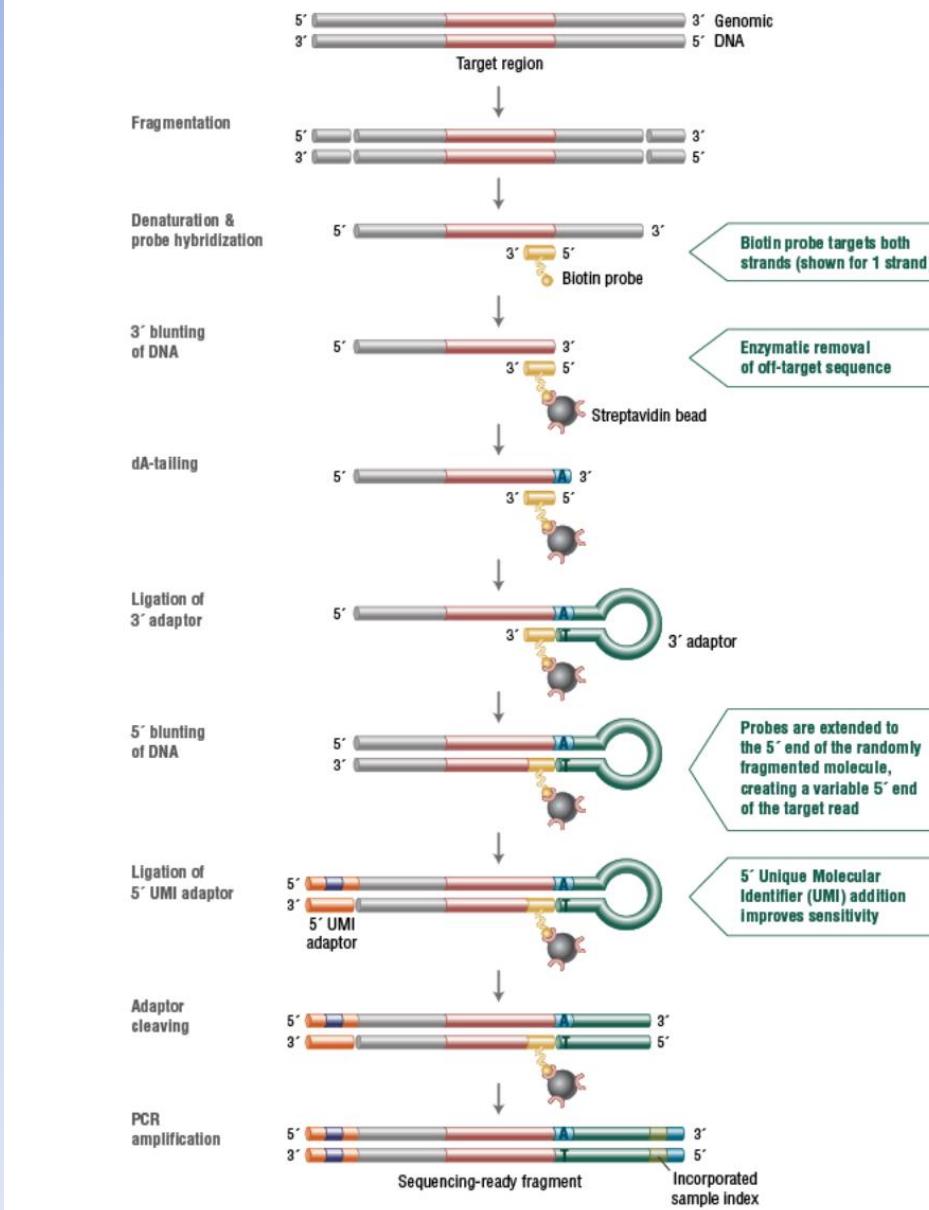


Figure 1: NEBNext Direct Workflow



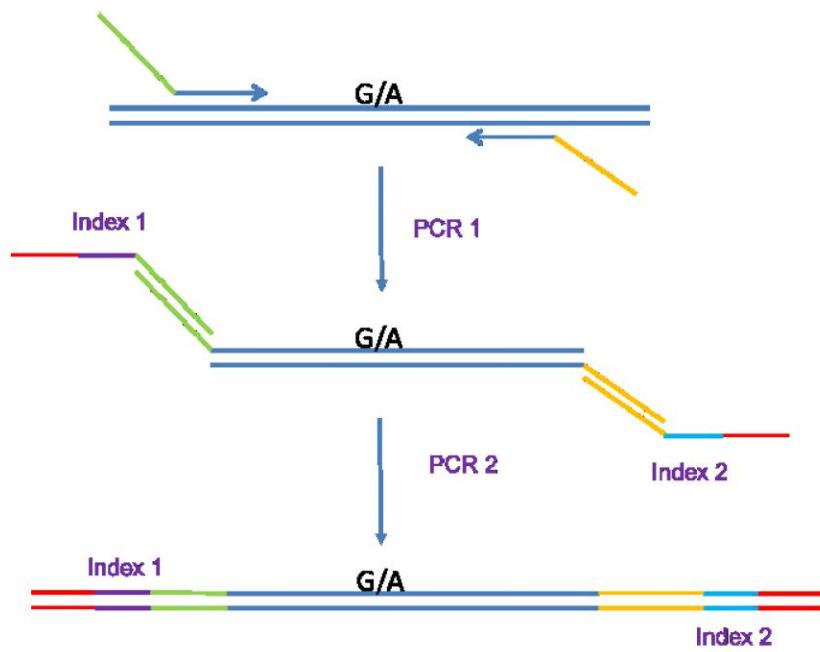
Non-commercial

Method	Platform	Company	Capital equipment?	Target Size
Non-commercial	Multiplex PCR	n/a	N	<0.02Mb
	Molecular Inversion Probes	n/a	N	0.02Mb-1.5Mb

FAST, CHEAP, EFFECTIVE



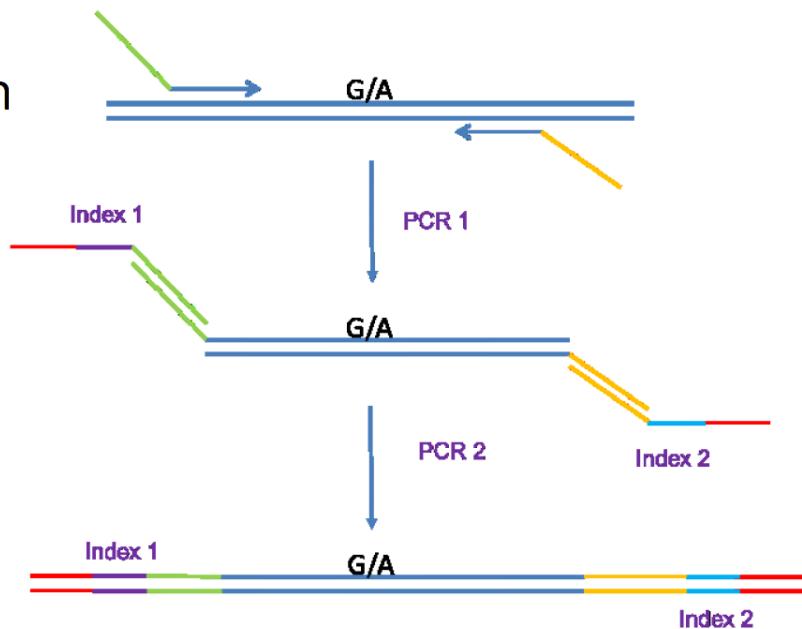
Multiplex PCR



- High-multiplex two step PCR
 - PCR 1 captures surrounding regions
 - PCR 2 incorporates dual-index barcodes and Illumina sequencing motifs
 - Genotyping and low VAF sensitivity

Multiplex PCR

- Multiple amplicons per reaction
 - Simple workflow
 - Cost effective approach
- Design Challenges
 - Define targets
 - Define size of amplicons
 - Optimal annealing temperatures
 - mPrimer – design software



Truseq Illumina Library and Primer Schema

.....--i5 index read-->.....--Read 01-->.....--i7 index read-->
.....ACACTCTTCCCTACACGACGCTCTCCGATCT →GATCGGAAGAGCACACGTCTGAACCTCCAGTCAC →
5'...AATGATAACGGCGACCACCGAGATCTACAC[i5]ACACTCTTCCCTACACGACGCTCTCCGATCT**xxxxxxxxx**AGATCGGAAGAGCACACGTCTGAACCTCCAGTCAC[i7]ATCTCGTATGCCGTCTCTGCTTG...3'
3'...TTACTATGCCGCTGGTGGCTCTAGATGTG[i5]TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA**xxxxxxxxx**TCTAGCCTCTCGTGTGCAGACTTGAGGTCAGTG[i7]TAGAGCATACGGCAGAACGAAAC...5'
.....<-TCTAGCCTCTCGTCTGCAGACTTGAGGTCAGTG.....5'
.....<-Illumina PE Read 2 Sequencing Primer
.....←----P5----→[i5]←-----SBS-----→<**insert**>←-----SBS-----→[i7]←-----P7-----→
.....<--TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA -5' i5 index primer

Truseq style PCR indexing primers

The IDT and other available UDI barcoding primers will have sequence:

P7

CAAGCAGAAGACGGCATACGAGAT[barcode]GTGACTGGAGTTCAGACGTGTGCTCTCCGATC*T

P5 indexing oligo

AATGATACGGCGACCACCGAGATCTACAC[barcode]ACACTTTCCCTACACGACGCTCTCCGATC*T

Your PCR primers will need Illumina flanking sequences at 5' end compatible with these sequences so should have general design

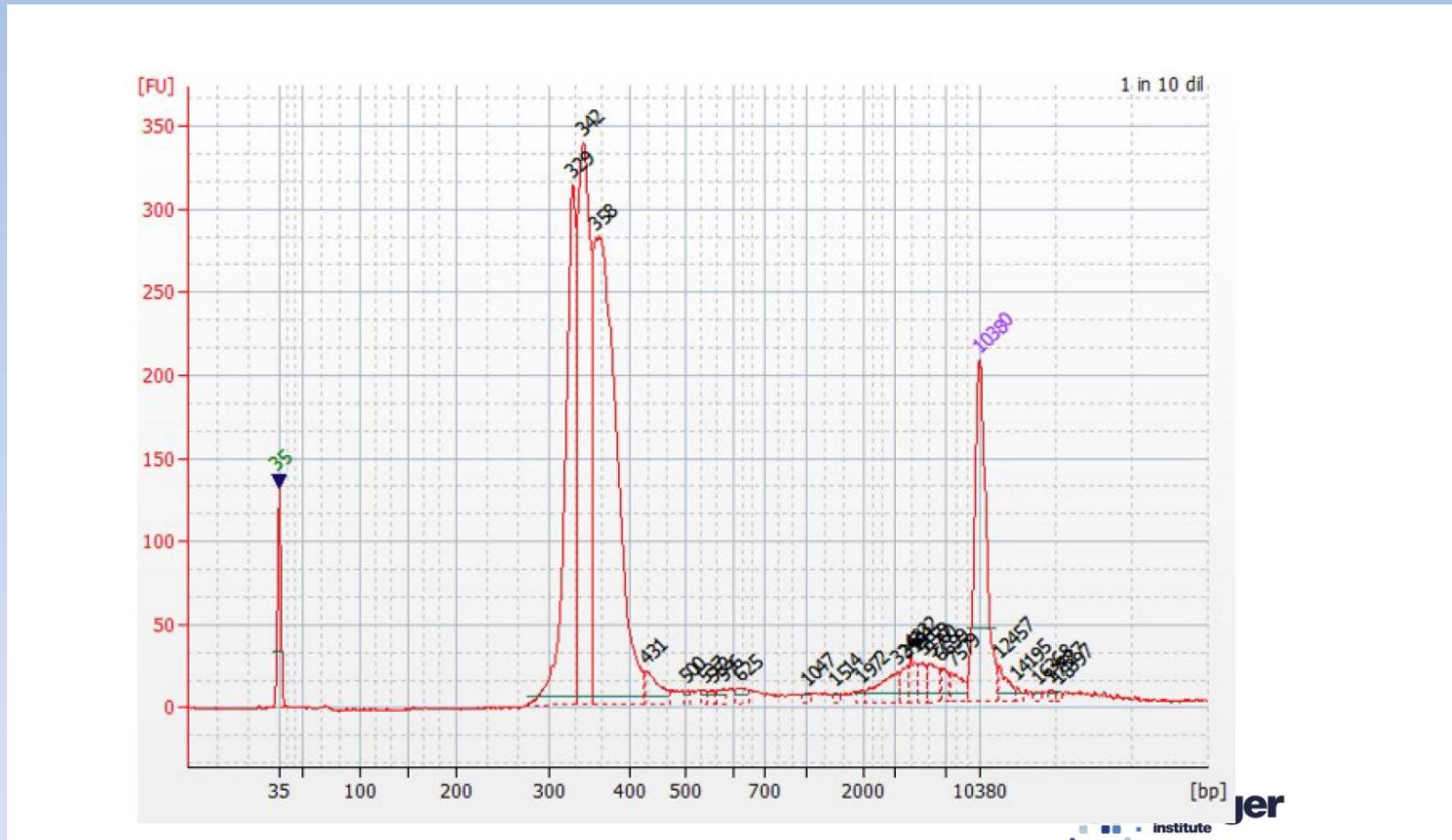
Forward primer would have design

5' ACACTTTCCCTACACGACGCTCTCCGATCT gene specific sequence 3'

Reverse primer would have design

5' GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT gene specific sequence 3'

Using these sequences means that custom sequencing primers are not required.

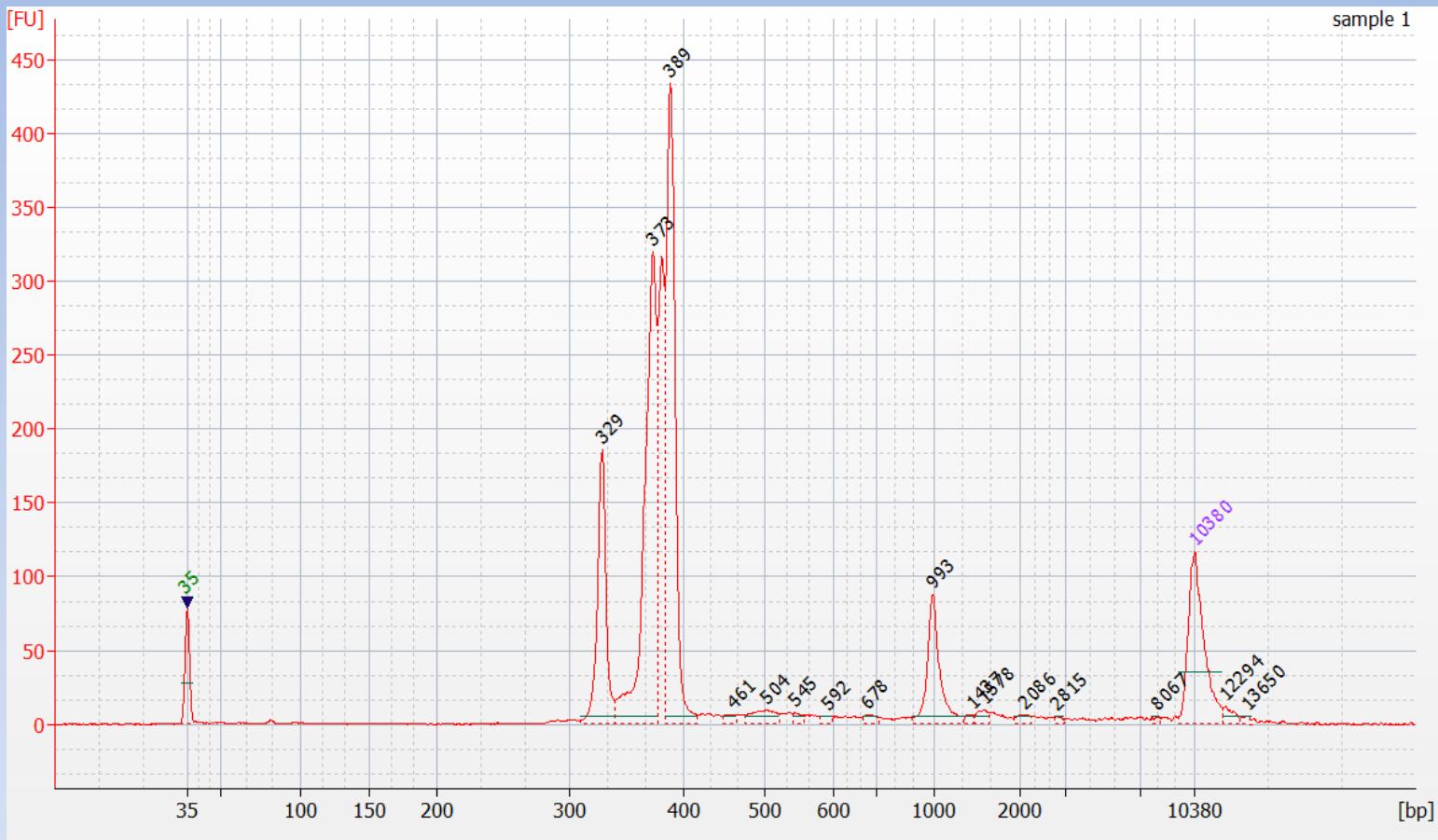


Multiplex PCR design

- Design difficulties
 - Specificity
 - Cross primer interactions (dimers/non specific amplification)
- Need good design software!!
Mprimer
 - Thorough design algorithms
 - Primer 3 input
 - Input target and surrounding sequence
 - Constrain target sizes to 190-250bp
 - Annealing temperature



Zhiyong Shen et. al. MPprimer: a program for reliable multiplex PCR primer design. BMC Bioinformatics, 2010, 11:143.



Non-commercial Multiplex PCR

- No capital equipment
- Suitable for small target number (1-100 probes)
- Requires design software
- Requires in-house optimisation
- Suitable for few-large sample numbers
- Low cost

Multiplex PCR Targeted Sequencing – Suited to the Miseq



The screenshot shows the homepage of the Cell Reports journal. The header features the journal's name "Cell Reports" in large white letters against a blue background with a microscopic image of cells. On the right side of the header, there are buttons for "All Content", "Cell Reports" (which is selected), and "All J". Below the header, a navigation bar includes links for "Explore", "Online Now", "Current Issue", "Archive", "Journal Information", and "For Authors". The main content area displays the volume and issue information: "Volume 10, Issue 8, p1239–1245, 3 March 2015". There is a "Report" link on the left and a "Switch to Standard View" link on the right. The main article title is "Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis". The authors listed are Thomas McKerrell¹³, Naomi Park¹³, Thaidy Moreno, Carolyn S. Grove, Hannes Ponstingl, Jonathan Stephens, Understanding Society Scientific Group, Charles Crawley, Jenny Craig, Mike A. Scott, Clare Hodkinson, Joanna Baxter, Roland Rad, Duncan R. Forsyth, Michael A. Quail, Eleftheria Zeggini, Willem Ouwehand, Ignacio Varela, George S. Vassiliou, with icons for a pencil and an envelope.

Cell Reports

All Content

Cell Reports All J

Explore Online Now Current Issue Archive Journal Information For Authors

Volume 10, Issue 8, p1239–1245, 3 March 2015

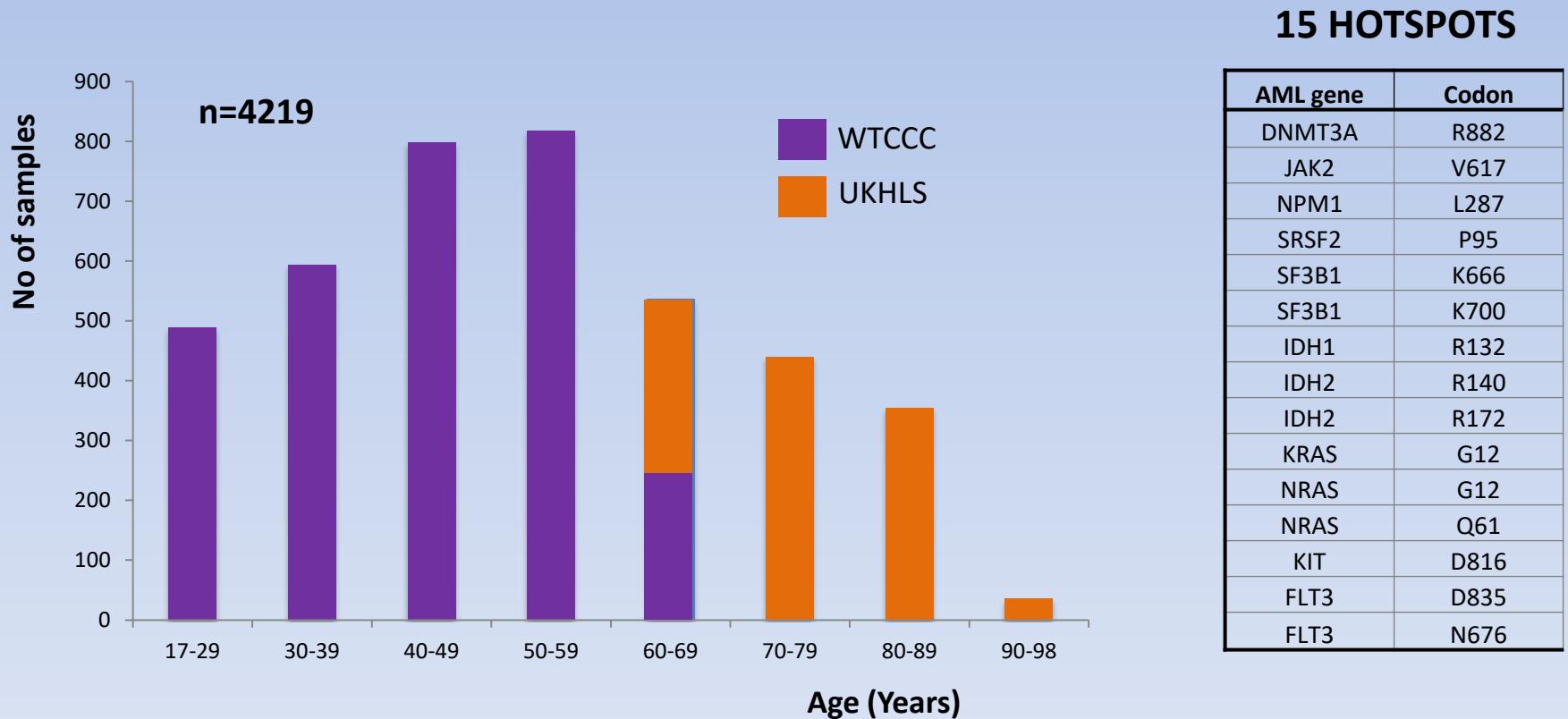
Report Switch to Standard View

Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis

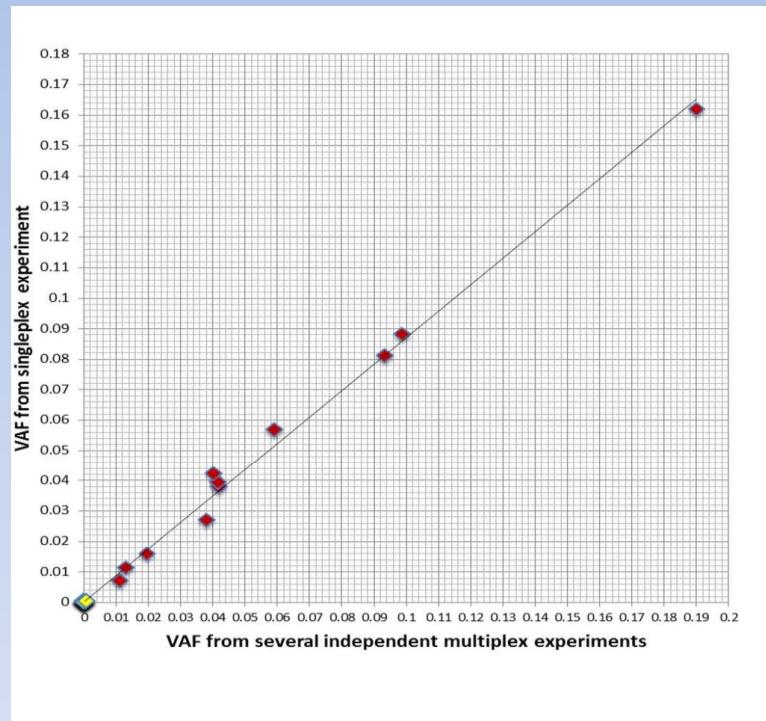
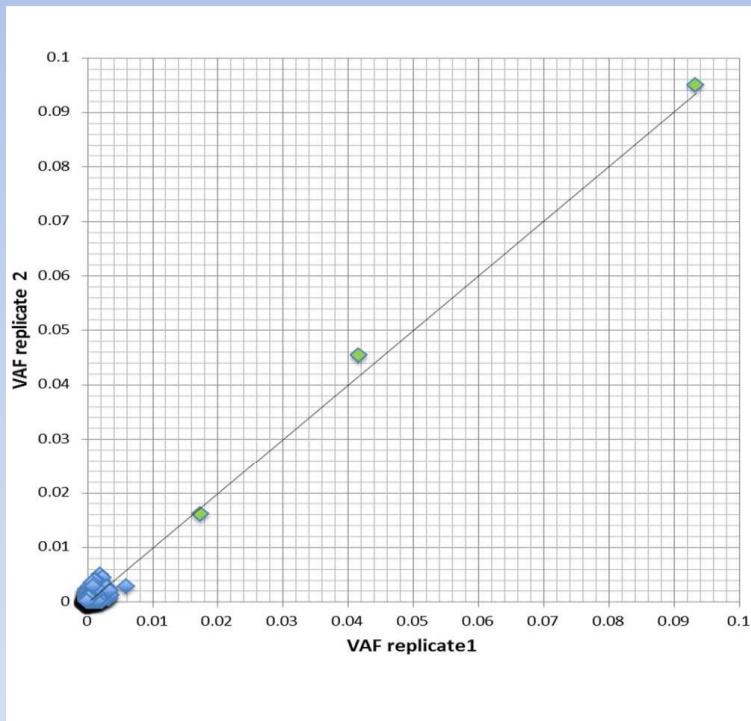
Thomas McKerrell¹³, Naomi Park¹³, Thaidy Moreno, Carolyn S. Grove, Hannes Ponstingl, Jonathan Stephens, Understanding Society Scientific Group, Charles Crawley, Jenny Craig, Mike A. Scott, Clare Hodkinson, Joanna Baxter, Roland Rad, Duncan R. Forsyth, Michael A. Quail, Eleftheria Zeggini, Willem Ouwehand, Ignacio Varela, George S. Vassiliou  

¹³ Co-first author

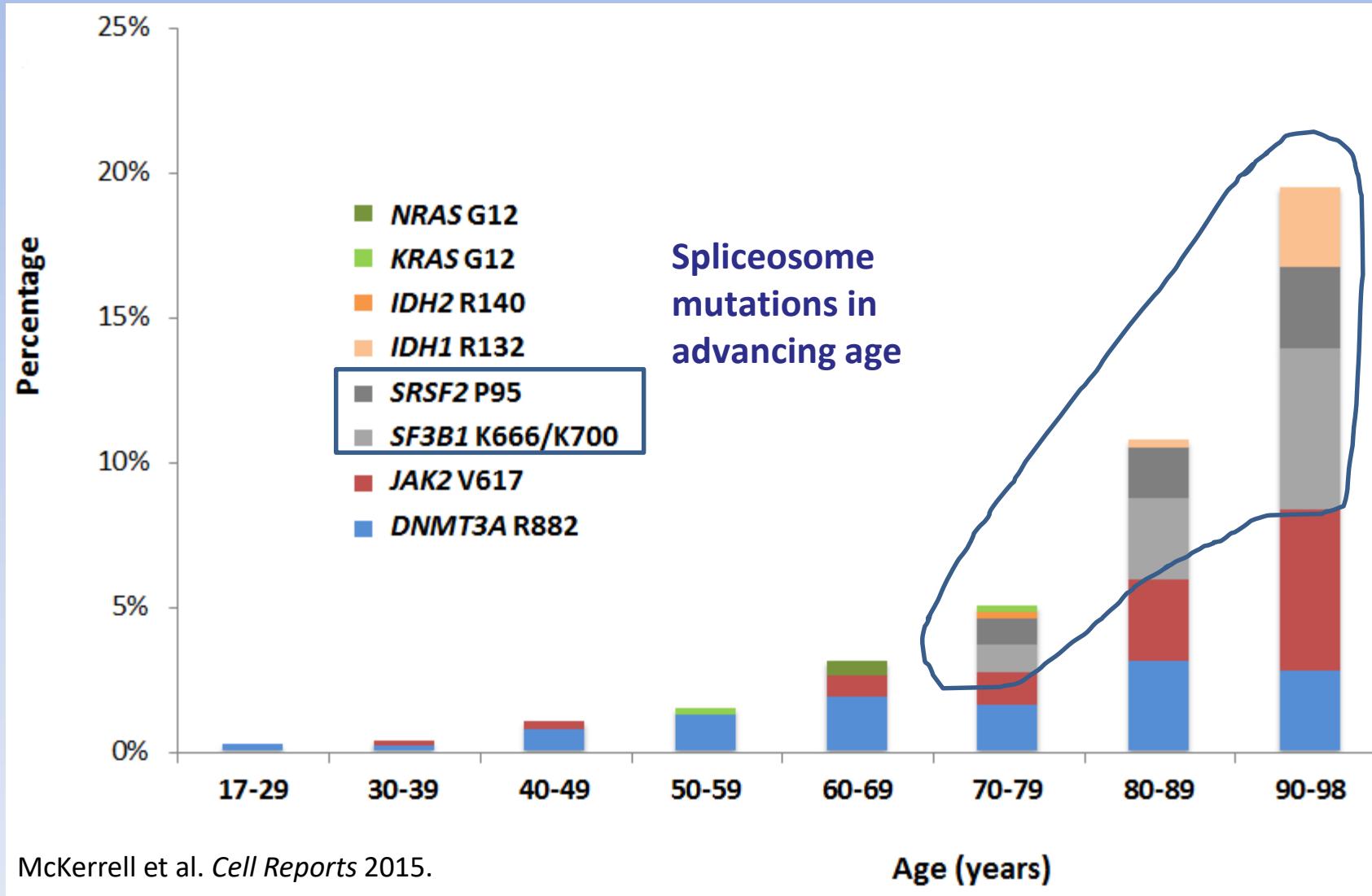
Study Cohort of healthy individuals



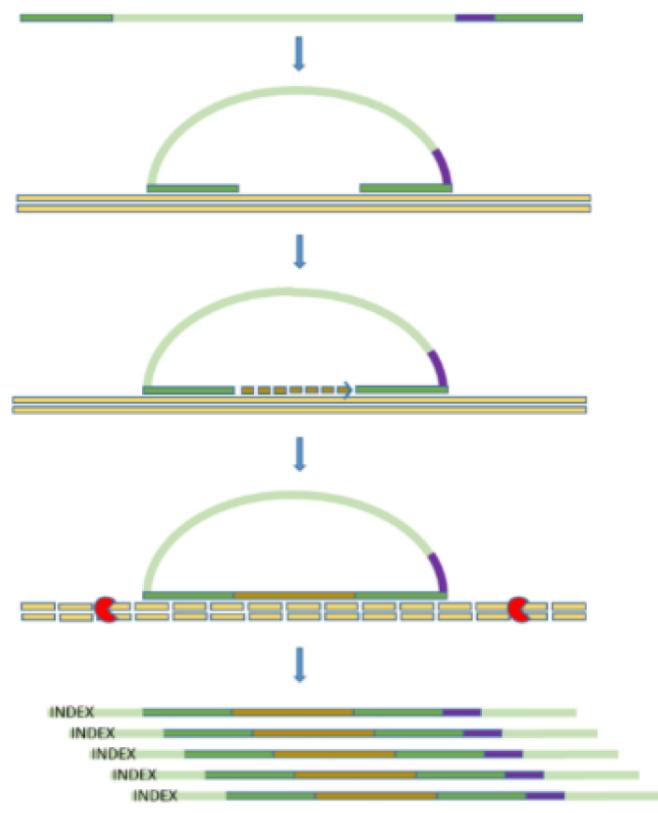
Variant Allele Fraction (VAF) quantitation Reproducibility



Prevalence and patterns of Clonal Haemopoiesis in 4219 healthy individuals



Molecular Inversion Probes (MIPs)



Single oligo per target, designed to span ~160bp

Incorporate UMI (MIPgen)

E.Boyle et al. MIPgen: optimized modelling and design of molecular inversion probes for targeted resequencing. Bioinformatics 2014

Probe inverts and binds to one strand of target gDNA

Polymerase extends and ligase closes circle

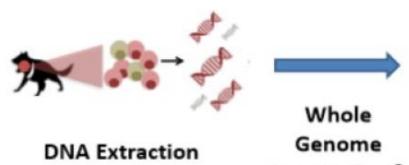
Exonuclease removal remaining linear DNA

PCR incorporating indexed primers

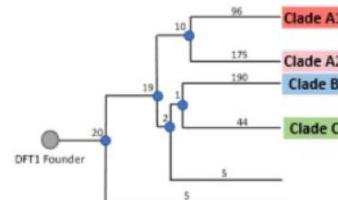
Tasmanian Devil MIP genotyping



Research Overview



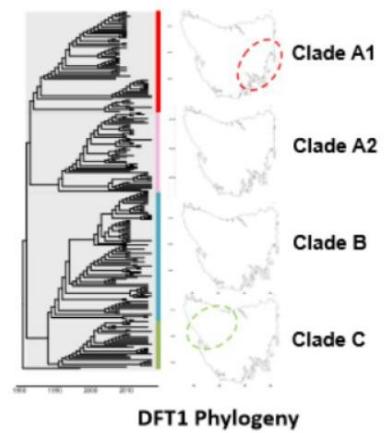
Whole
Genome
Sequencing &
Analysis



Selection of Phylogenetic
Markers & Optimization

May 2018 ... June 2018

Genotype
Across
>1000 samples
using MIPs



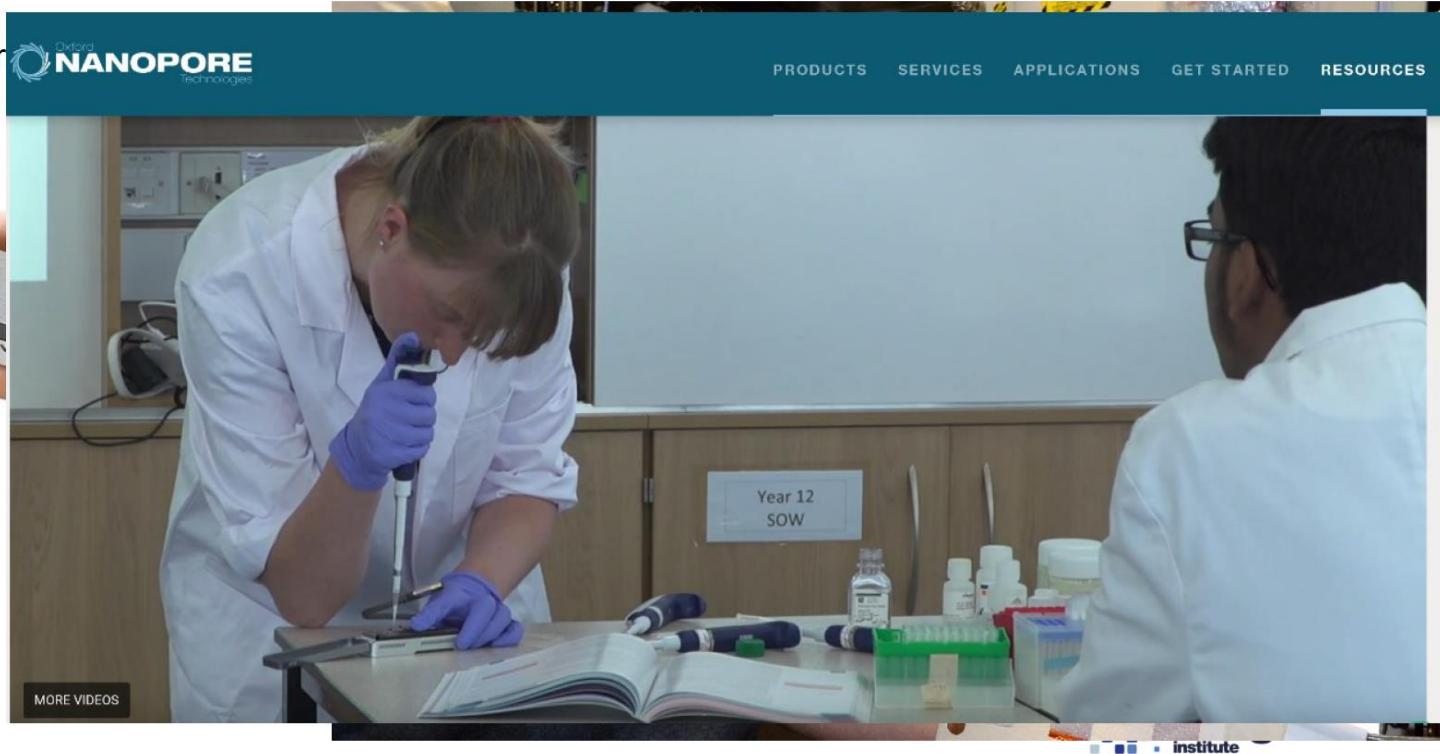
DFT1 Phylogeny

Molecular Inversion Probes

- No capital equipment
- Wide range target sizes (10-1000's probes)
- Incorporates UMI for quantitative variant analysis
- Design software - Mipgen
- Suitable for few-large sample numbers

Long read sequencing – Oxford Nanopore Technologies (ONT)

- Short



Targeted sequencing by:

- PCR
- Hybrid capture
- Adaptive sampling
- cas9

Adaptive sampling

Samples not of interest rejected electronically

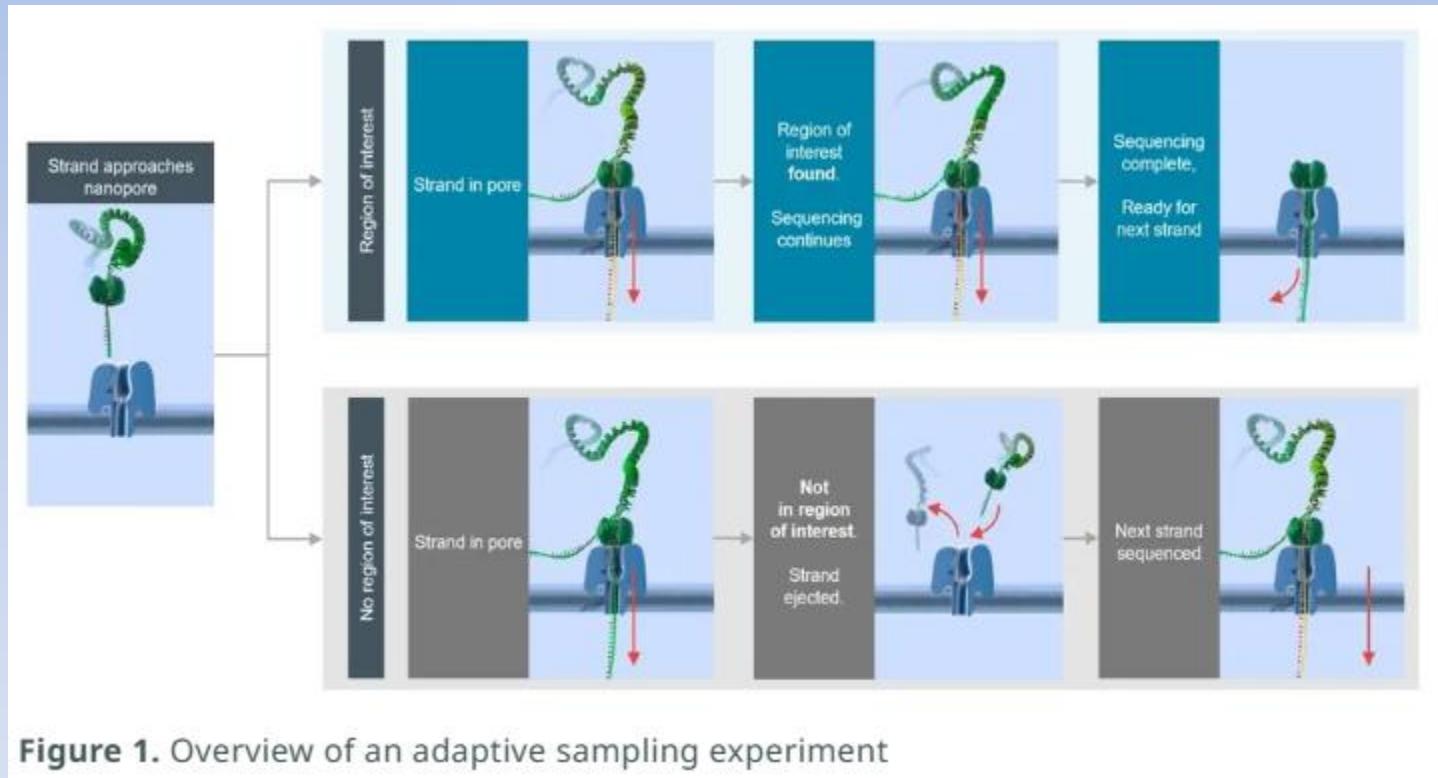
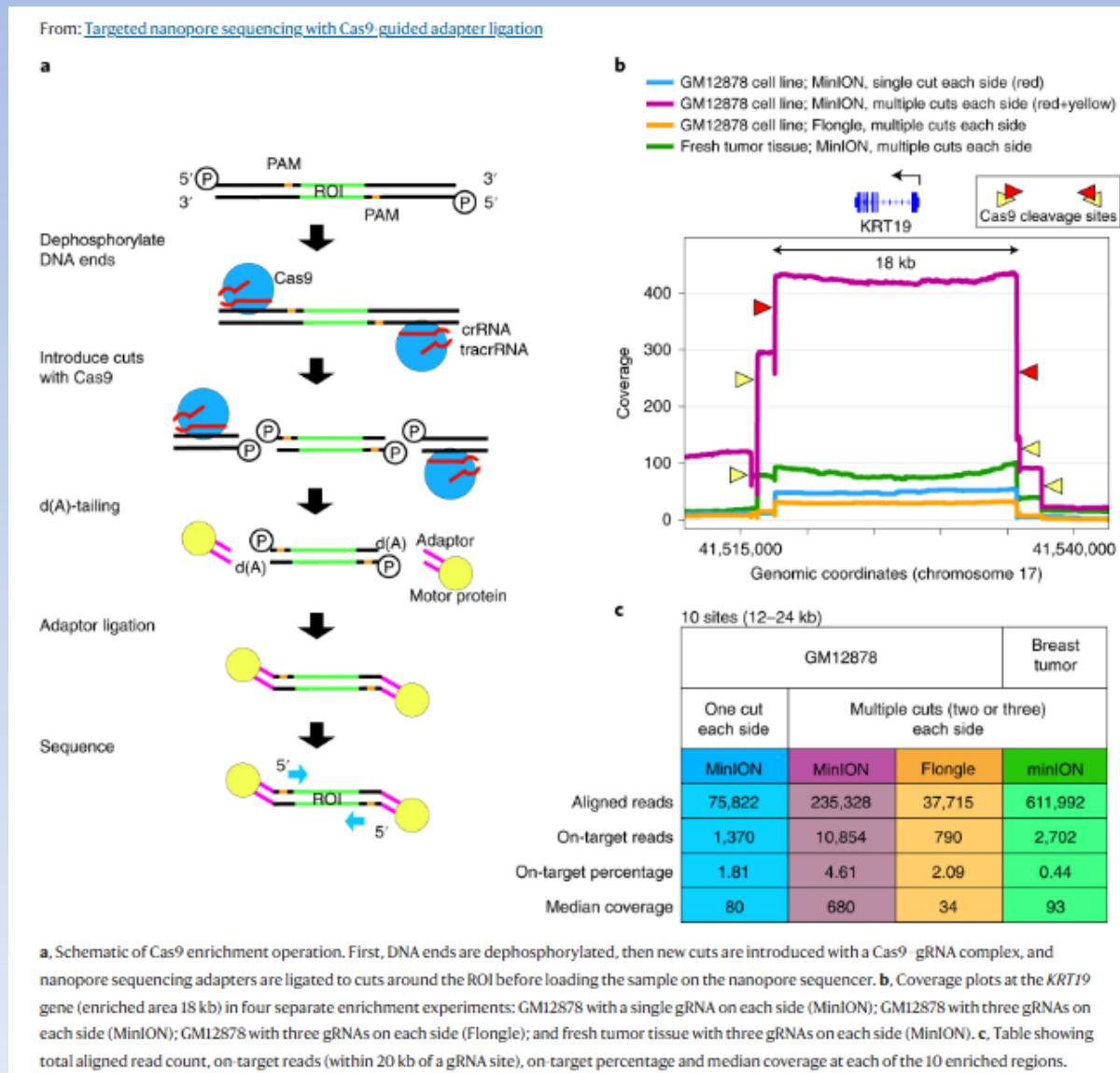


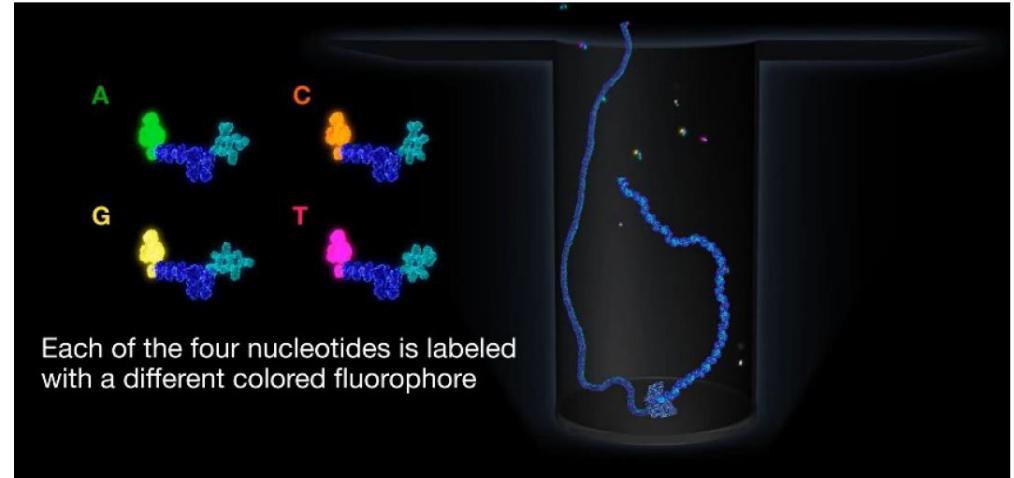
Figure 1. Overview of an adaptive sampling experiment

Cas9 enrichment

<https://www.nature.com/articles/s41587-020-0407-5>



Long read sequencing – Pacbio



Each of the four nucleotides is labeled with a different colored fluorophore

Targeted sequencing by:
-PCR
-Hybrid capture
-cas9

Long read targeted sequencing

Complex genomic regions

- Homopolymeric stretches, repetitive elements, or short tandem repeats (STRs)
- Structural variation
- HLA (immune system)

Targeted capture of long genomic regions

- Long Range PCR (amplification)
- CRISPR/Cas9 (non-amplification)

Conclusions

- **Define** the project/projects
 - Targets
 - How many?
 - Does a panel already exist?
 - What coverage is needed?
 - VAF
 - Sample number
 - Sequencing platform
 - Multiplexing capability
 - How many samples
 - Capital equipment outlay
 - Existing equipment
 - Bioinformatic support