Diagnostic applications of Clonal Sequencing (NGS)

Prof. Graham Taylor PhD FRCPath
Leeds Teaching Hospitals &
Leeds Institute of Molecular Medicine

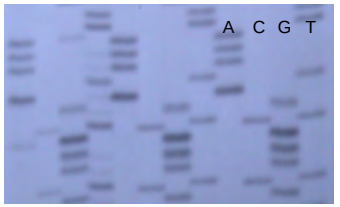
Diagnostic applications of NGS

- Background
- Workflow
- Diagnostic Applications
  - Resequencing
    - PCR
    - Capture
  - CNV-seq
- Summary & Next Steps

Sequencing Goes Large

Roche 454	Illumina Solexa	ABI SOLiD
Helicos	Pacific Bioscence	Oxford Nanopore

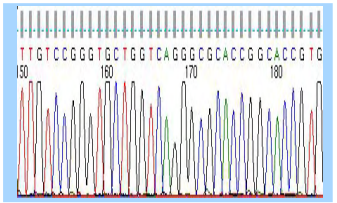
Last Gen



10^1 sequences per run

10^3 bases per run

Current Gen



10^2 sequences per run

10^4 bases per run

Next Gen

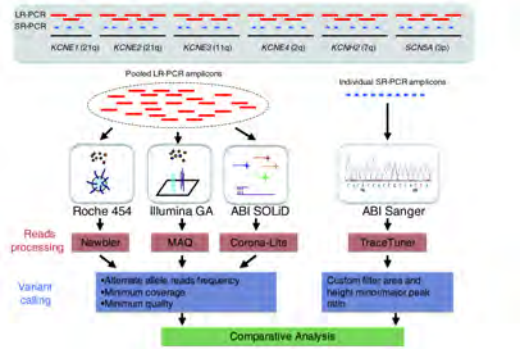
```
HWI-EAS318 11    2    17    453   1023 0    2     ATGAAGATGATGCTGGAGAGCCGGTTCCGCGTGCCCATCCAGCTGCTGTCC    a]VZ]V]
ZVVFMT_BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB      hs_ref_chr7.fa        288283      F     16T1C15T5T1A2G3T1     0
      56              22    R
HWI-EAS318 11    2    48    1258  339  0    2     GCATCCTGCAGGACTTCAGCTCCGACCCCTCCTCCAACCTCTCGTCCAACT    __
\_`a`aaXV_]GI^\^\_\[`\PVUF_FXNV\\T`VBBBBBBBBBBBBB      hs_ref_chr7.fa        288548      F     47C3 31    31
      62    R
HWI-EAS318 11    2    74    770   1871 0    2     GGTCTTAGGGCGCCGGGATCCCGCCCCCGCAAGGCGCGCTCGGCCGGCTCG    _ZI
[R_T__a\``_`]_`SXUXTS\Z[`^aF[\\UX]BBBBBBBBBBBBBB hs_ref_chr7.fa        288624      R     51    31    119
                -107  F
HWI-EAS318 11    2    68    799   774  0    2     CACCAGGCTCAGCCCCAAAGCGGCGGAGAACCCGGACTGGTGAGTGGGGGC
      ``aaaZZa`a\Z_```^``a^]V]```X]]\^VJWRQ\OJUS^XKRTF\BBB    hs_ref_chr7.fa        288849      F     51    78
      169             25    R
HWI-EAS318 11    2    93    1049  1842 0    2     ACTCACGGGCCTGTTCTTTCCATGCTGATGTTCGTCCTCGCTGCTCCCTGC    _
\a`aaaa`_X[[RT^a_a_`^^`[^`W\XJQ[^UU\U^[VPTTLXVVRNS      hs_ref_chr7.fa        290584      F     51    92    172
      27    R
```

10^6 -10^7 sequences per run

10^8 -10^9 bases per run

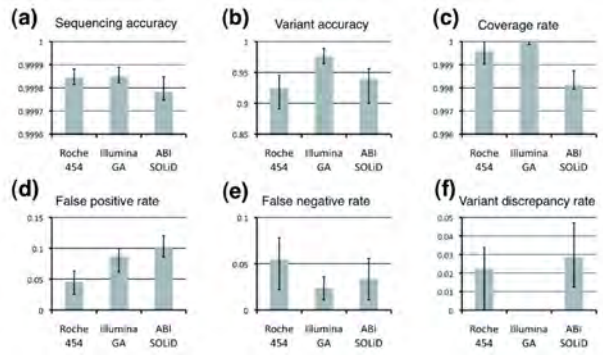
10^4 – 10^5 increase in amount of data generated

Comparison of 454, SOLiD & GAII



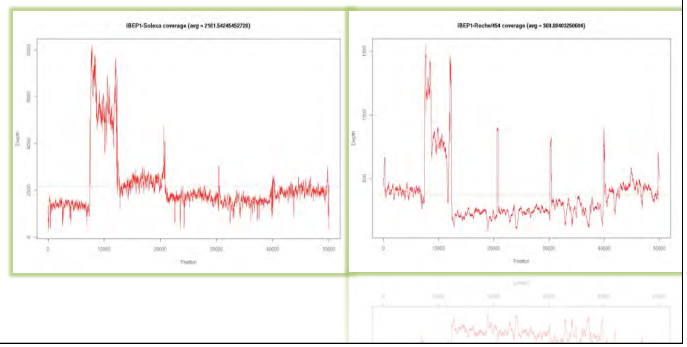
Harismendy *et al. Genome Biology 2009* ***10:R32***

Accuracy & Coverage



Harismendy *et al. Genome Biology 2009* ***10:R32***

Sequencing 8q24 from long PCR products



Lessons from the 8q24 study

- Short reads (32 base) can cover long genomic regions because repeats are imperfect and can be aligned
- Long PCR is an effective way to produce template for sequencing
- Coverage is uneven (why?)
- Point mutations are detectable and consistent with conventional sequencing results
- Indels are missing or under-represented
- Cost per base about **50 fold less** with Illumina than Roche

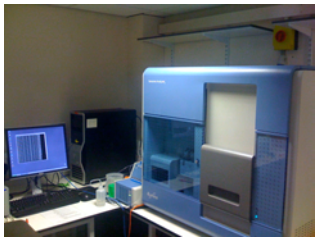
Application of Illumina GAII

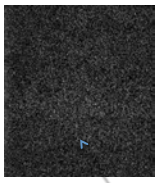
Cost: 0.002p per base

Capacity: >7.0 Gigabase per run

Simplicity:

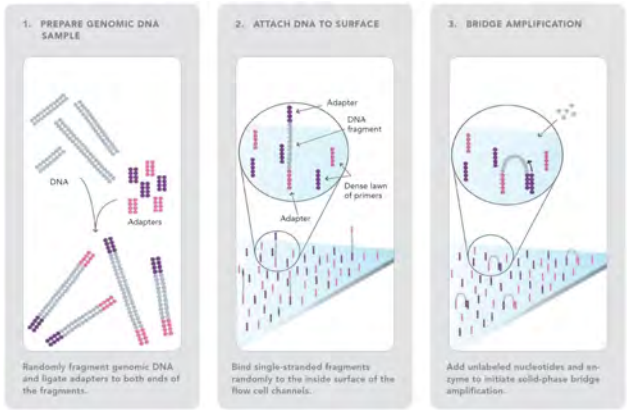
library>cluster station>sequence>data



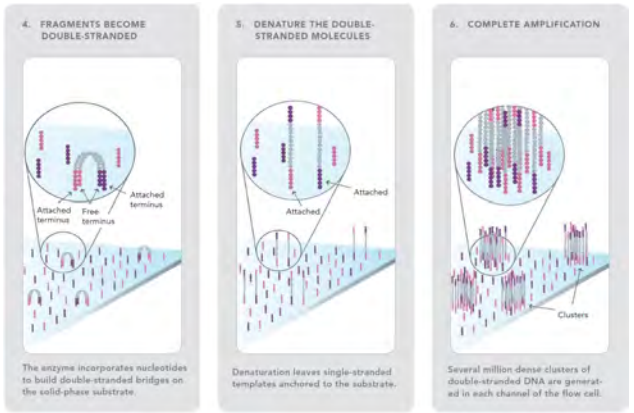


100,000 spots (1/100th of one flow-cell channel)

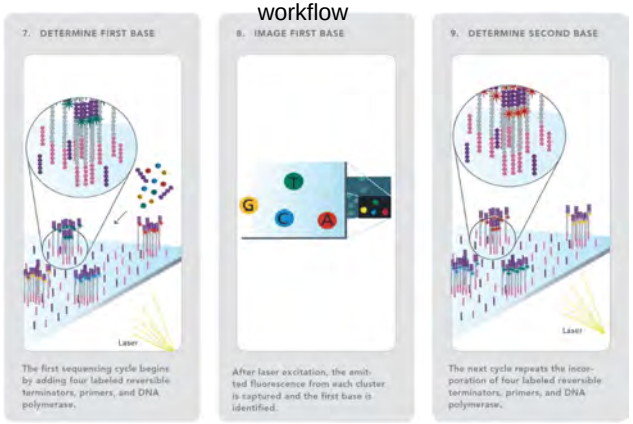
Illumina workflow



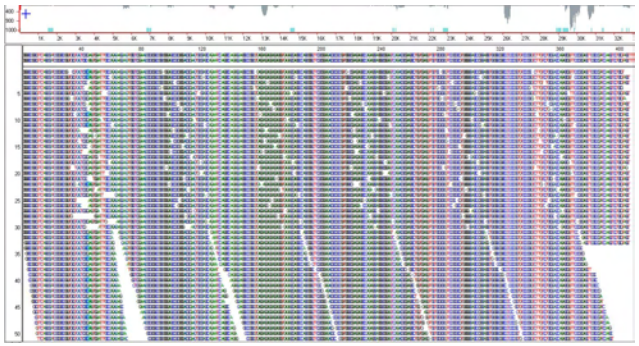
Illumina workflow



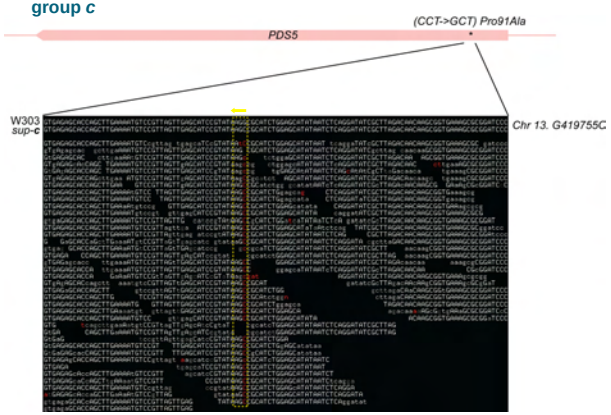
Illumina workflow



The days of manual checking are over..

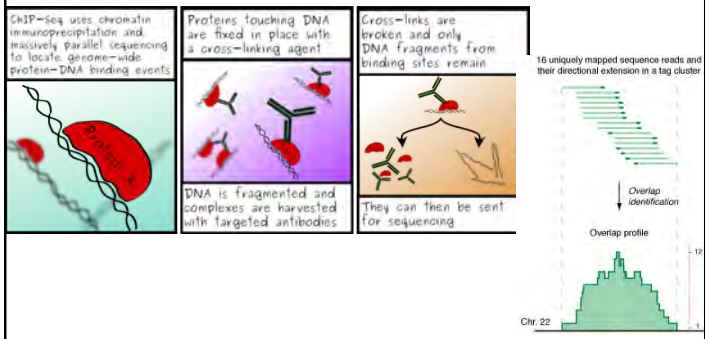


Resequencing: Identification of *eco1-1* suppressor complementation group *c*



Illumina Genome Analyzer (CR-UK Genomics Service, Leeds)

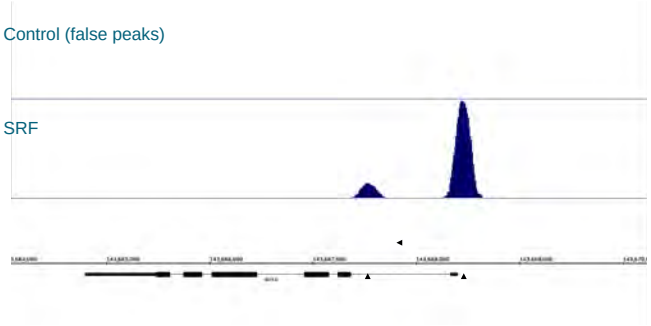
ChIP-seq



ChIP-seq SRF Binding along the gene *Actb*

Control (false peaks)

SRF



CCATAAAAGG

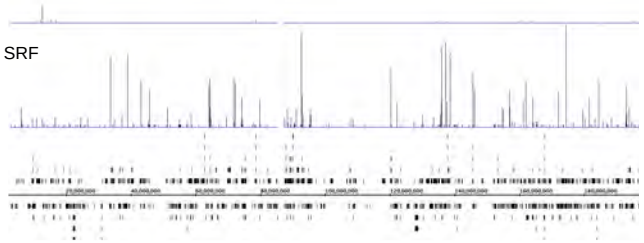
CCATAAAAGG

SRE:CC(A/T)$_6$GG (Known consensus of SRF binding site)

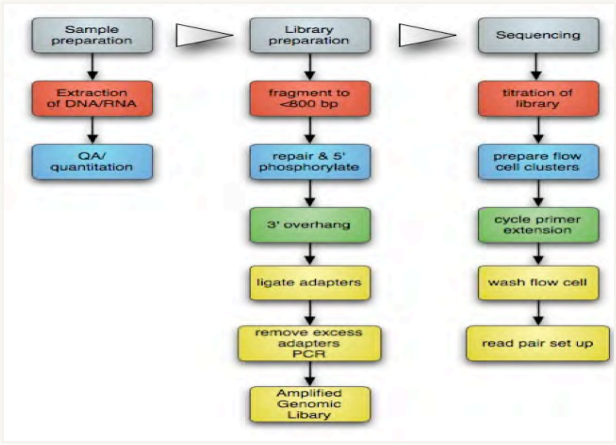
SRF binding along Chr1 (NIH-3T3)

Control (false peaks)

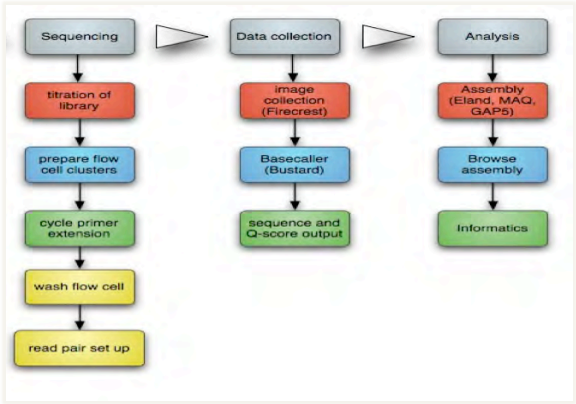
SRF



Workflow



Workflow



Software

- Pipeline: Signal>Base calls
  - Illumina: GOAT, Bustard and Firecrest (Unix)
- Assembly, Mapping & Browsing
  - ELAND (Illumina, Unix)
  - MAQ & Maqview (Sanger, Unix)
  - NextGene (Softgenetics, Windows)

What does this mean for clinical diagnostics?

The currency of genetic analysis will become DNA sequence

- Sequence count
- Sequence variation
- Sequence arrangement
- The currency is dropping to commodity prices
- Skill set required is less lab, more informatics biased

How will we use it?

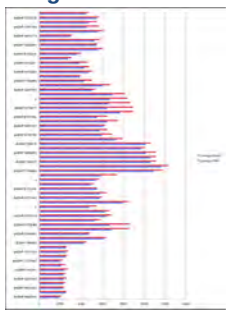
- Test design when DNA sequence costs are trivial
  - Urgent (same day) or non urgent (2 week)
  - Pre and post test costs
  - Syndrome/Gene list driven
    - Hereditary cancers, “Cardiome” “Retinome” “Kinome” : 20 – 200 genes 250-3,000 megabases

Clinical Applications

- Incremental rather than radical
- Mutation Detection in *TP53, BRCA1 & BRCA2*
- Gene capture in autozygosity projects
- Gene dosage & chromosome counting
  - Fixed tissue
- What does it mean for the clinical service?

Coverage and costs

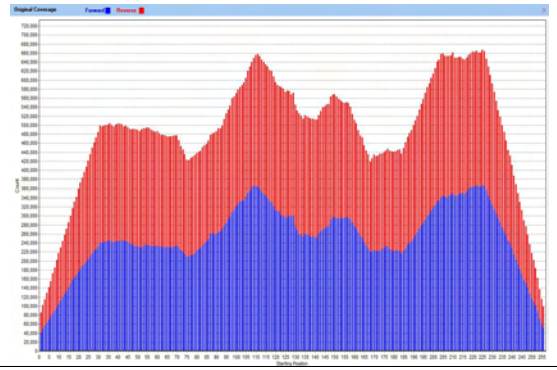
Coverage



Costs

Capital Outlay	220000
Pipeline costs	30000
Total Project Outlay	**250000**
Annual Running Costs:	
Pay	
1 FTE – SSO	37500
1 FTE – SO2	30000
0.5 FTE Bioinformatics	18000
Total Pay Costs	**85500**
Non Pay	
Reagent costs	100000
Maintenance costs	20000
Depreciation	44000
Total Non Pay Costs	**164000**
Total Annual Running Costs	**249500**

Coverage can be wide (*e.g.*exome) or deep (*e.g.* pooled samples)



Sequence wide or deep to achieve cost efficiency

- Sample pooling
  - Ingman & Gyllensten EJHG 17 383-6 2009
- Sample tagging
  - Craig *et al* Nat Methods 5, 887-93, 2008
  - 3 base
  - 6 base tag = 4^5 (1024) variants
  - Illumina kit

The case for gene-centric analysis

Coverage

- 1 flow cell = 7 channels
- 1 channel = 10 million reads
- 1 read = 36 – 200 bases
- 1 channel = 0.3 – 2 Gigabases
- 1 flow cell = 4 haploid genome equivalents
- At mean coverage of 100, one channel can cover 5-20 Megabases

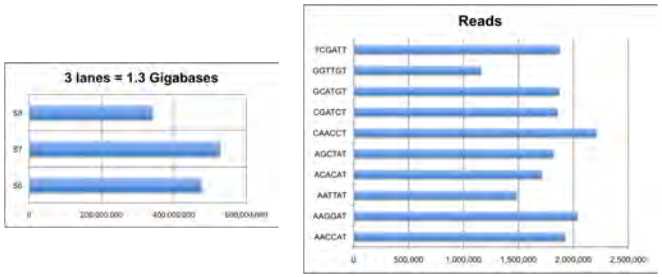
Costs

- 1 channel costs between £600-£1,200 in reagents
- 1 base costs between 2/10,000 pence – 5/100,0000 pence
- At mean coverage of 100, one base costs 1/50 – 5/10,000 pence

Target selection methods

- Sequence everything and select later in software
- OR
- Long PCR
- Gene Collector
- Genome Partitioning

6 base tags 51 base reads



Long PCR

- 8q24
- TP53
- CDKN2A
- BRCA1 and 2 exon 11

Current software releases not ideal

- ELAND/CASAVA/Genome Studio: supplied with the machine.
- Academic
- Commercial
  - NextGene
  - CLC Workbench
  - DNAStar
- Illuminator
  - http://dna.leeds.ac.uk/illuminator/

Additional software used

http://www.softgenetics.com/

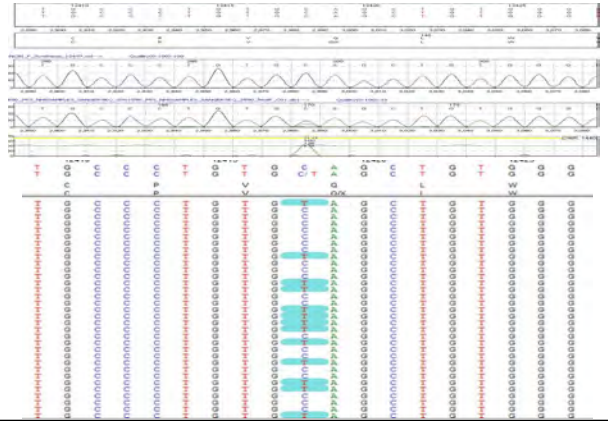
http://dna.leeds.ac.uk/sequencing/

Proof of principle

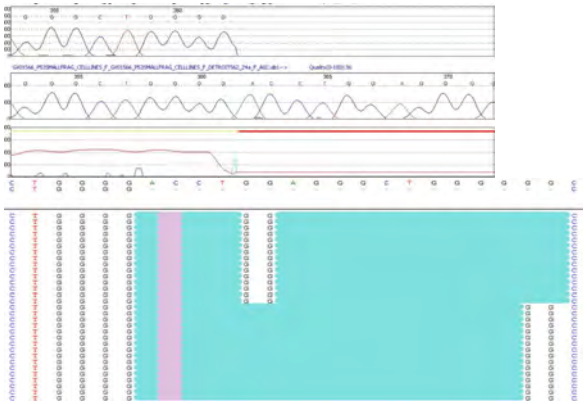
- *TP53*: mutated in 50% somatic cancers
- Small gene (4,510 bases of TP53, including 1,172 bases of coding sequence)
- Cell lines and cases available

Select the entire gene coding regions in two long PCRs

Point mutation (p.144Q>QX)



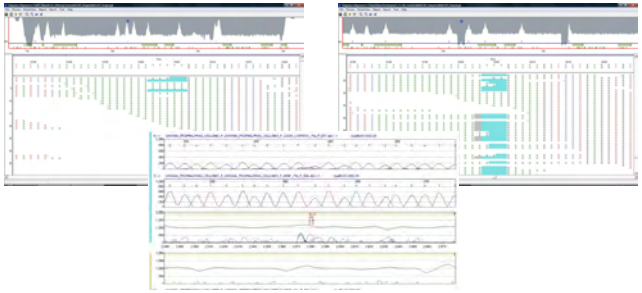
16 base insertion-deletion



False +ves: slippage at $(A)_{18}$
Poor quality Sanger sequencing

Manageable number of false positives, very few in or near exons

30 base reads | 45 base reads



TP53 Results

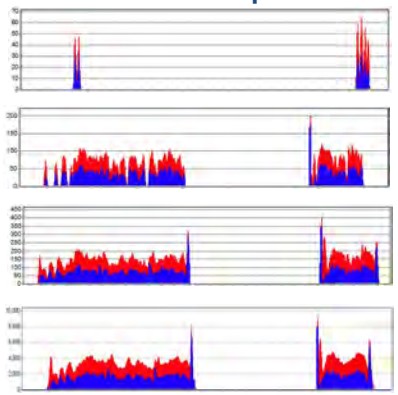
4 cell lines in duplicate
- 5 coding variants reported
- Duplicate runs fully concordant
- No false negatives
- All pathogenic mutations identified in COSMIC or by Sanger sequencing were detected

10 cases in duplicate
- 45 variants reported within 50 bases of exons 2-11
- Duplicate runs fully concordant
- No false negatives
- All pathogenic mutations previously seen by Sanger sequencing were identified

TP53 Coding Variants

Case	Amino Acid Change	Mutation Call	Reported Mutation
1	72P>R	c.[215C>G]+[215C>G]	
	144Cb>QX	c.[430C>T]+[=]	c.430C>T (p.Gln144X)
	377T>TP	c.[1129A>C]+[=]	
2	72P>R	c.[215C>G]+[215C>G]	
	248R>QR	c.[743G>A]+[=]	c.743G>A (p.Arg248Gln)
3	72P>R	c.[215C>G]+[215C>G]	
	248R>QR	c.[743G>A]+[=]	c.743G>A (p.Arg248Gln)
	371S>SS	c.[1113C>C]+[1113C>C]	
	377T>TP	c.[1129A>C]+[=]	
4	72P>R	c.[215C>G]+[215C>G]	
	158R>HR	c.[473G>A]+[=]	c.473G>A (p.Arg158His)
	377T>TP	c.[1129A>C]+[=]	
5	72P>PR	c.[215C>G]+[=]	
	158R>PR	c.[473G>C]+[=]	c.473G>C (p.Arg158Pro)
	72P>R	c.[215C>G]+[215C>G]	
6	256T>TA	c.[766A>G]+[=]	c.766A>G (p.Thr256Ala)
	377T>TP	c.[1129A>C]+[=]	
7	72P>PR	c.[215C>G]+[=]	
	273R>HR	c.[818G>A]+[=]	c.818G>A (p.Arg273His)
	377T>TP	c.[1129A>C]+[=]	
8	72P>R	c.[215C>G]+[215C>G]	
	242C>XC	c.[726C>A]+[=]	c.726C>A (p.Cys242X)
	377T>TP	c.[1129A>C]+[=]	
9	125T>TR	c.[374C>G]+[=]	c.374C>G (p.Thr125Arg)
	377T>TP	c.[1129A>C]+[=]	
10	72P>PR	c.[215C>G]+[=]	
	158R>HR	c.[473G>A]+[=]	c.473G>A (p.Arg158His)

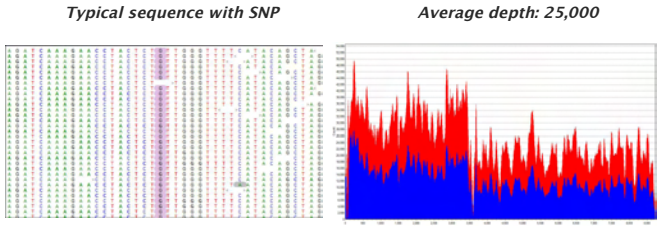
Read Depth *vs.* Detection



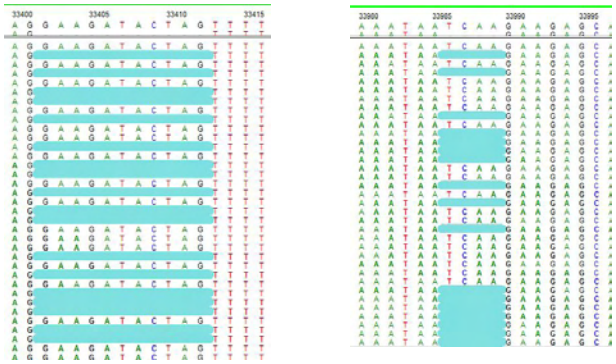
Reads	Samples	Detection
5,000	1,000	0/11
10,000	500	7/11
20,000	250	11/11
500,000	10	11/11

Mutation Call	Amino Acid Change
c.[74+38C>G]+[74+38C>G]	
c.96+41delACCTAGGGCTGGGGGGG	
c.[215C>G]+[215C>G]	72P>R
c.[375-283T>C]+[375-283T>C]	
c.375-160delAAA	
-	
-	
c.[375-91G>A]+[375-91G>A]	
c.[451C>T]+[451C>T]	151P>S
c.[672+62A>G]+[672+62A>G]	
c.[672-66T>C]+[=]	
c.[672-36G>C]+[672-36G>C]	

BRCA1 and *BRCA2* exon 11 control



BRCA1 exon 11



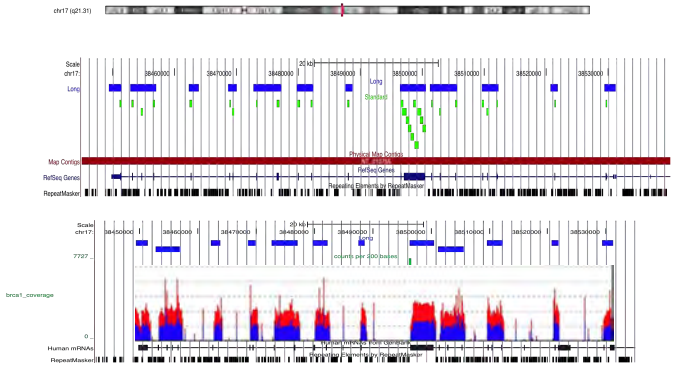
BRCA1 exon 11 and *BRCA2* exon 11 summary

- 13 of 13 mutations plus known SNPs were detected and reported automatically, even though present as 1/8 or 1/6 of a sample mixture
- Even at this crude pilot level, the cost for complete sequence screening is competitive with existing techniques, and faster

BRCA1 and *BRCA2*

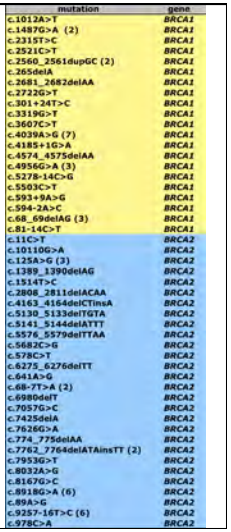
- 65,159 bases from the BRCA1 and BRCA2 genes in 22 amplicons
- 16,032 coding bases
- 55 previously sequenced cases
- Average sequencing coverage 592
- One run repeated 6 times
- One control sample included in all runs

BRCA1 primers in UCSC



BRCA1 & BRCA2

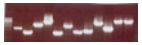
- 671 variants (326 *BRCA1*, 345 *BRCA2*)
- 77 variants identified previously were all detected
- 49 distinct mutations seen
- All pathogenic variants were detected
- 23 frame shifting mutations (14 variants)
- 10 nonsense mutations (7 variants)
- 16 missense mutations



mutation	gene
c.1012A>T	BRCA1
c.1487G>A (2)	BRCA1
c.2315T>C	BRCA1
c.2521C>T	BRCA1
c.2560_2561dupGC (2)	BRCA1
c.265delA	BRCA1
c.2681_2682delAA	BRCA1
c.2722G>T	BRCA1
c.301+24T>C	BRCA1
c.3319G>T	BRCA1
c.3607C>T	BRCA1
c.4039A>G (7)	BRCA1
c.4185+1G>A	BRCA1
c.4574_4575delAA	BRCA1
c.4956G>A (3)	BRCA1
c.5278-14C>G	BRCA1
c.5503C>T	BRCA1
c.593+9A>G	BRCA1
c.594-2A>C	BRCA1
c.68_69delAG (3)	BRCA1
c.81-14C>T	BRCA1
c.11C>T	BRCA2
c.10110G>A	BRCA2
c.125A>G (3)	BRCA2
c.1389_1390delAG	BRCA2
c.1514T>C	BRCA2
c.2808_2811delACAA	BRCA2
c.4163_4164delCTinsA	BRCA2
c.5130_5133delTGTA	BRCA2
c.5141_5144delATTT	BRCA2
c.5576_5579delTTAA	BRCA2
c.5682C>G	BRCA2
c.578C>T	BRCA2
c.6275_6276delTT	BRCA2
c.641A>G	BRCA2
c.68-7T>A (2)	BRCA2
c.6980delT	BRCA2
c.7057G>C	BRCA2
c.7425delA	BRCA2
c.7626G>A	BRCA2
c.774_775delAA	BRCA2
c.7762_7764delATAinsTT (2)	BRCA2
c.7953G>T	BRCA2
c.8032A>G	BRCA2
c.8167G>C	BRCA2
c.8918G>A (6)	BRCA2
c.89A>G	BRCA2
c.9257-16T>C (6)	BRCA2
c.978C>A	BRCA2

Possible Diagnostic Workflow

- Clinical referral
- Molecular Genetics lab
  - Long PCR works well with commercial saliva kit
- Clonal Sequencer "satellite"
- Data only report to lab
- Confirm sequence
- Clinical report



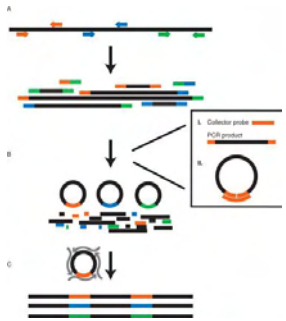
Lab extracted DNA

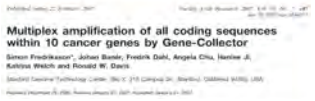
Saliva kit extracted DNA

Mutation report

coverage

Gene collector





Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector

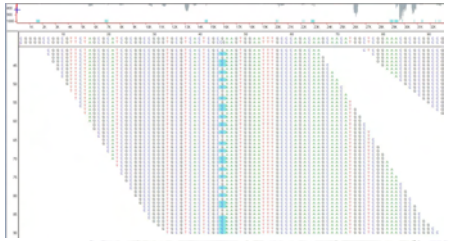
- Multiplex PCR (pfu polymerase), blunt-ended products suitable for ligation by circularization
- Collector probes guide circularization, closed circles formed by thermostable ligase
- Enrichment of circular DNA by exonuclease treatment and rolling circle amplification
- Risk of artefacts from RCA

Gene collector

Addresses the analysis of short fragments by Illumina

- *XPC*
- *KRAS/BRAF in fixed tissue*
- *BRCA 1 and 2: towards complete analysis of BRCA sequence by long PCR and gene collector*

Mutation Detection

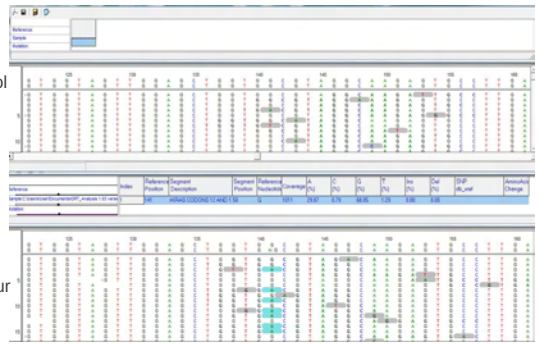


Reference Position	Segment Description	Segment Position	Reference Nucleotide	Coverage	A (%)	C (%)	G (%)	T (%)	Ins (%)	Del (%)
50	>XPC exon1	50	C	28	0.00	0.00	0.00	0.00	78.57	0.00
94	>XPC exon1	94	G	58	0.00	56.90	-56.90	0.00	0.00	0.00
1150	>XPC exon 6	184	T	45	0.00	0.00	0.00	0.00	100.00	0.00
1863	>XPC exon 9_10_11	379	G	44	0.00	9.09	-100.00	0.00	0.00	90.91
1935	>XPC exon 9_10_11	451	G	20	0.00	100.00	-100.00	0.00	0.00	0.00
1936	>XPC exon 9_10_11	452	A	20	-100.00	0.00	0.00	0.00	0.00	100.00
2017	>XPC exon 9_10_11	533	C	111	0.00	-45.05	0.00	45.05	0.00	0.00
2452	>XPC exon 12_13	34	T	35	100.00	0.00	0.00	-100.00	0.00	0.00
2564	>XPC exon 12_13	146	C	42	0.00	-100.00	0.00	0.00	0.00	100.00

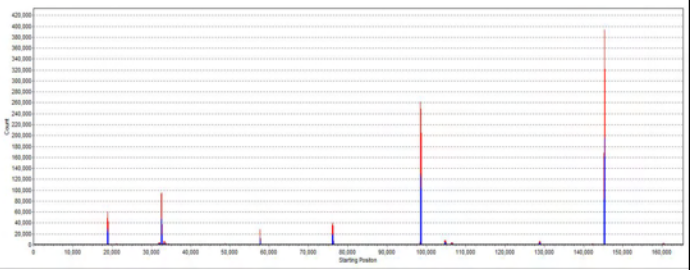
Ras Raf Mutation Report (c.38G>A)

Control

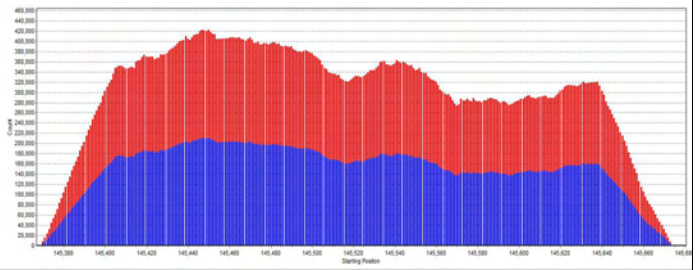
Tumour



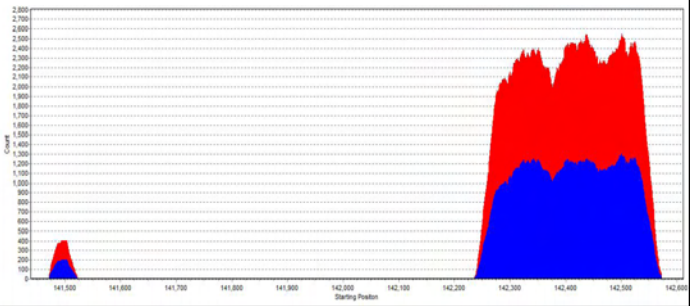
BRCA gene collector



BRCA gene collector



BRCA gene collector



BRCA Mutation Report

Segment Description	Position	Ref Base	Coverage	A(%)	C(%)	G(%)	T(%)	a a change
BRCA1-NCBI-NP_009225.2.gbk	18858	G	55240	0	0	98.2	1.8	
BRCA1-NCBI-NP_009225.2.gbk	19077	T	44629	0	2.22	0	97.8	
BRCA1-NCBI-NP_009225.2.gbk	19118	G	41437	0	0	99.4	0.65	
BRCA1-NCBI-NP_009225.2.gbk	33789	A	3043	91.3	0	8.68	0	1290K>KE
BRCA1-NCBI-NP_009225.2.gbk	33806	G	3341	21.1	0	78.9	0	1295L>LL
BRCA1-NCBI-NP_009225.2.gbk	33913	A	2679	84.9	0	0	15.1	1331Q>QL
BRCA1-NCBI-NP_009225.2.gbk	57763	G	29505	2.25	0	97.8	0	1665V>MV

Matched *vs.* Unmatched reads

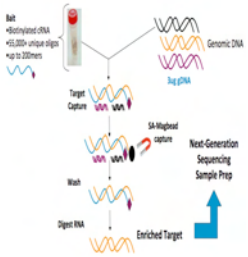
Sample	Matched	Unmatched
TP53 long PCR	20,763	465 (2%)
TP53 long PCR + templiphi	17,097	31,646 (65%)
XPC gene collector	15,411	49,860 (76%)
Kras/Braf (FFPE) gene collector	6,368	26,669 (81%)
BRCA1 & *BRCA2*	21,276	12,432 (37%)

Summary of Gene Collector

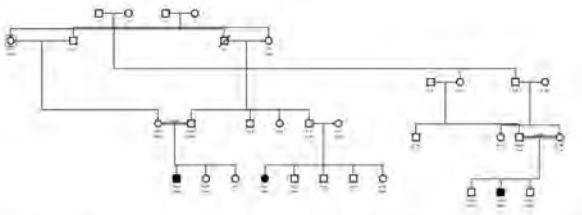
- Converts short amplicons to Illumina-readable format
- Identifies mutations
- Some loss of useful reads compared with long PCR
- Difficult to achieve even amplification

Genome partitioning

•No PCR
•Genome Fragmentation
•Linker and Tag Ligation
•Oligo Capture
•Sequence



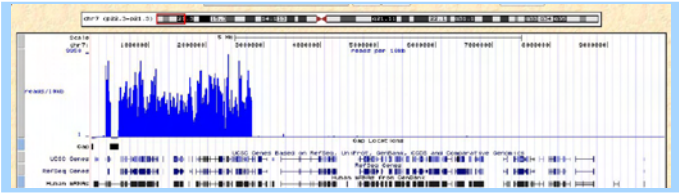
Primary Ciliary Dyskinesis



Defect of the cilia affecting lungs, sinuses, ears and nose.
Can affect organ position- situs invertus.

Autosomal recessive inheritance.
SNP analysis revealed homozygous target region on tip of chr7p.
Contains 32 genes with 300 exons.
Designed custom Agilent capture chip for 2.7Mb region.

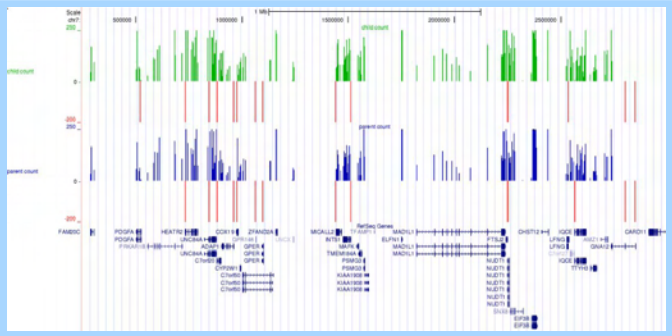
Chr7p target region enrichment



17M reads aligned to genome, 2.2M aligned to target region

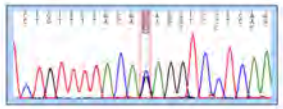
Target region is ~3Mb or ~0.1% of genome, giving enrichment of 130-fold

Read depth across the exons

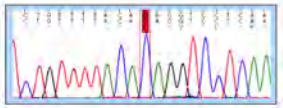


Identification of a splice site mutation

Parent (heterozygous carrier)
Boundary change at 786115 seen in 3 reads of 37
Boundary change at 786307 seen in 2 reads of 54
Boundary change at 791679 seen in 7 reads of 15
Boundary change at 838687 seen in 1 reads of 63
Boundary change at 838764 seen in 2 reads of 62

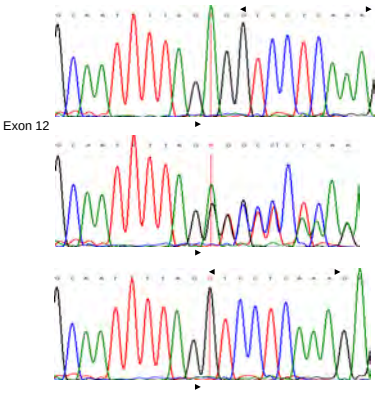


Child (homozygous affected)
Boundary change at 786115 seen in 2 reads of 68
Boundary change at 786307 seen in 1 reads of 96
Boundary change at 791679 seen in 36 reads of 36
Boundary change at 838687 seen in 2 reads of 116
Boundary change at 838764 seen in 5 reads of 170

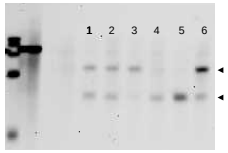


Mutation segregates with the condition
Affects the splicing of the mRNA
Is not present in 200 ethnically matched controls
Is the only pathogenic change identified in the target region
Highly conserved

Splice site mutation activates cryptic site resulting in 2b deletion in mRNA



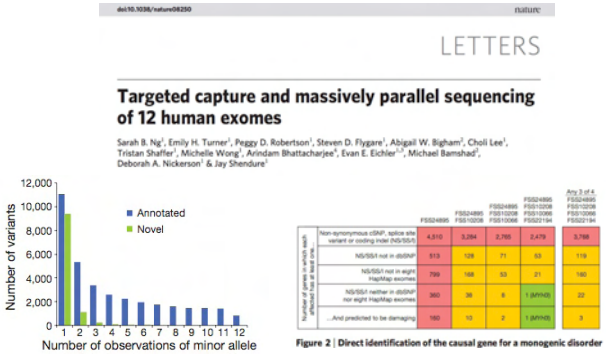
RNAse Protection Assay



Exons 12+13 protected (~245b)
Exon 13 protected (~160b)

1. Heterozygote + ΔAG probe
2. Heterozygote + normal probe
3. Homozygote + ΔAG probe
4. Homozygote + normal probe
5. Normal + ΔAG probe
6. Normal + normal probe

Exome resequencing



LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[3], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

Figure 2 | Direct identification of the causal gene for a monogenic disorder

Exome Sequencing

- "Targetted capture and massively parallel sequencing of 12 human exomes" Ng *et al.* Nature Sept 2009
  - Captured the entire human exome using Agilent arrays (1% of genome)
  - Generated 30Mb sequence per individual spread over 180000 exons
  - Sequenced 8 HapMap individuals
  - Sequenced 4 unrelated individuals with Freeman-Sheldon syndrome, an autosomal dominant disorder due to mutations in MYH3.
  - Averaged 17272 cSNPs per individual

Exome data

	1	2	3	4
Number of shared variants	4510	3284	2765	2479
Variants not present in dbSNP	513	128	71	53
Variants not present in 8 exomes	799	168	53	21
Variants not present in 8 exomes and dbSNP	360	38	8	**1 (MYH3)**

But: 4% of exons (7000) were not sequenced deeply enough for analysis

Sub-Genomic Targetting

- Focus on regions of interest
- Higher coverage
- Easier informatics
- Fewer “unwanted” genotypes
- Target on the basis of phenotype or pathways
- Compare with exome targetting
- *Work in progress…*

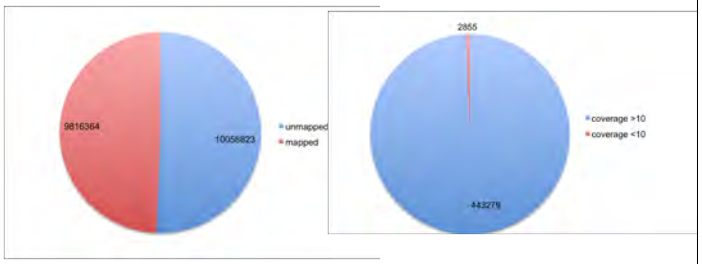
Cardiopathy

- 55 genes
- 600Kb coding region
- Simple repeats were retained in the baits

Class of repeat	Number of each repeat type identified	Percent (%)	Average size (bp)
SINE	15	13.6	201
LINE	12	10.9	198
LTR	3	2.7	319
Simple_repeat	33	30.0	49
DNA	7	6.4	166
Low_complexity	40	36.4	47
snRNA	0	0.0	0
Other	0	0.0	0
tRNA	0	0.0	0
Satellite	0	0.0	0
rRNA	0	0.0	0
scRNA	0	0.0	0
Unknown	0	0.0	0
srpRNA	0	0.0	0
RNA	0	0.0	0
Total	**110**	**100.0**	

Control samples

- Well charactorised case:
- e.g. Craig Venter (see: http://ccr.coriell.org sections/collections/HuRef/?SsId=78). He's available for $150 for 50ug to academic institutions.
- Settled for a CEPH trio that is included in the 1,000 genomes project

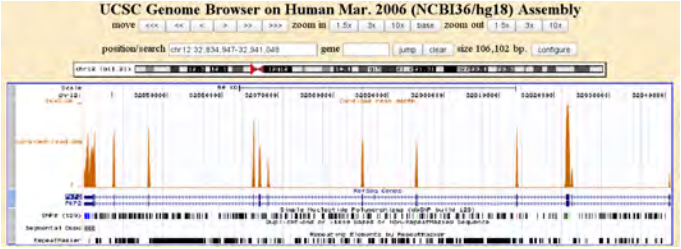


Exons analysed 1,325
Total reads 19,875,187
Reads not mapping to an exon 10,058823
Cumulative exon length 446134 bases
981 bases had zero coverage (as unique reads)
1,874 had coverage <10

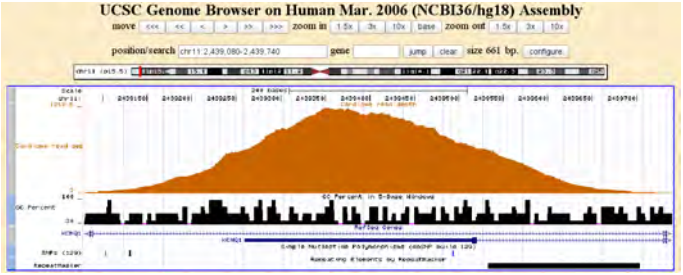
SNP Concordance

DBSNP	Target Array	Exome
identical	230	76
different	3	12
not found	0	89
not in CCDS		56
Total	233	233

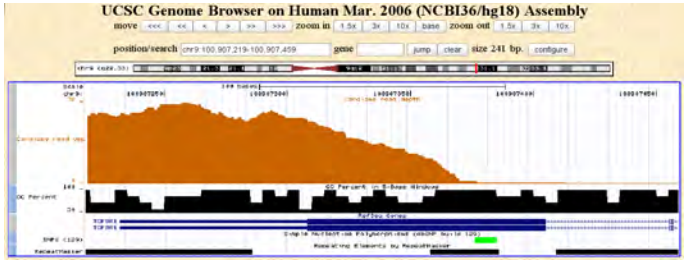
Coverage in a well-behaved region



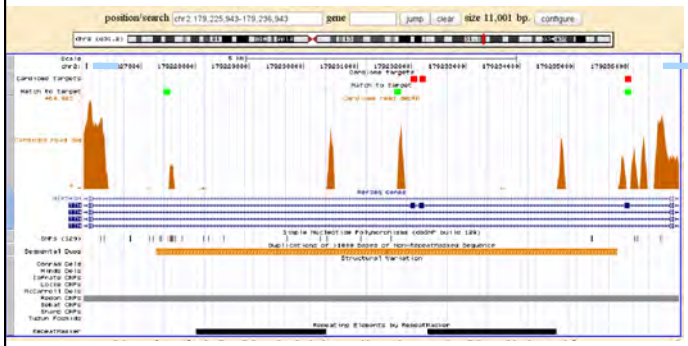
Good coverage extended by 150bp on either side



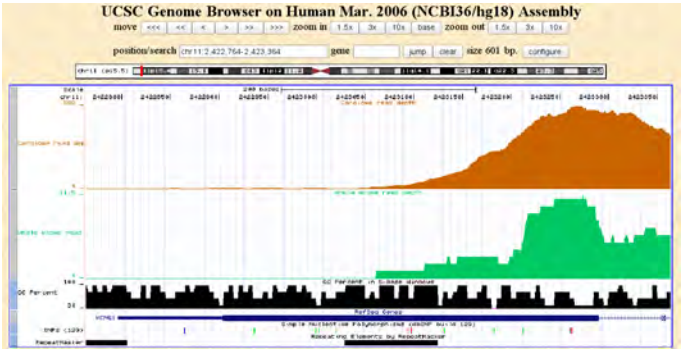
Inadequate coverage -why?



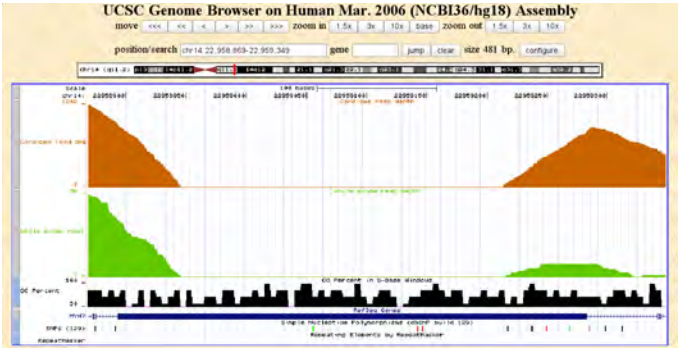
Off- Target Effects



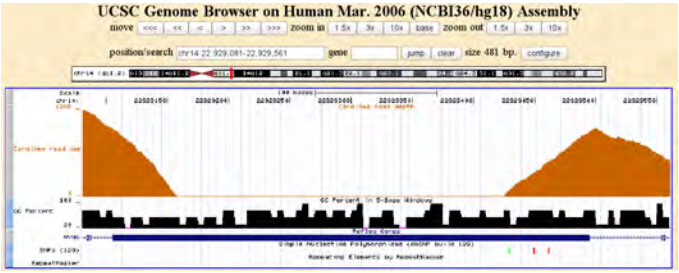
Comparison with exome



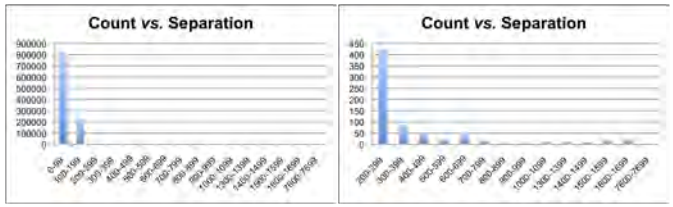
Both exome and target missed the same coding regions



Duplicate reads on *MYH6* & *MYH7*

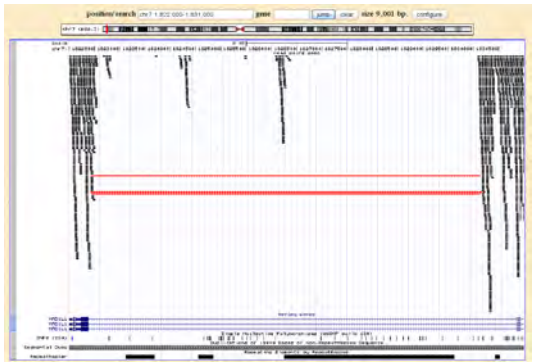


Genomic rearrangements by paired read analysis



Size of fragment sequenced was ~200b
Separations > 200b may indicate large changes

A novel homozygous deletion?



Partitioning Summary

- A practical proposition
- Needs continuous review against whole genome sequencing (currently @ around £5,000)
- Some off target effects
- Some problems with Eland and multiple alignments (BWA is OK)
- Genome browsing and reporting still to be finalised

Copy number by sequence read depth

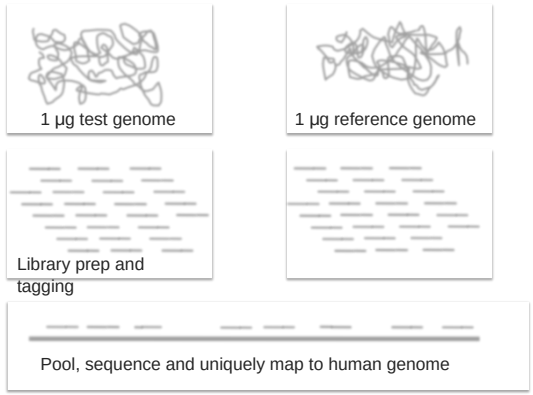
or – aCGH by next generation sequencing

Limitations of array CGH

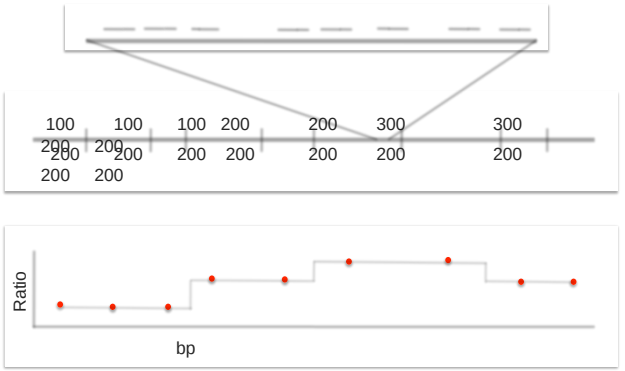
- Reliance on nucleic-acid hybridisation
  - DNA quality effects
  - Prior sequence knowledge required
  - "Analog" nature of fluorescence signal
- Reproducibility

CNV-Seq

- Massively-parallel DNA sequencing using Illumina Genome II analyser
- 'Tagging' allows multiplex experiments per lane
- Standard and non-standard data analysis
- Sequence depth of coverage digitally quantified using custom-designed Python and R scripts

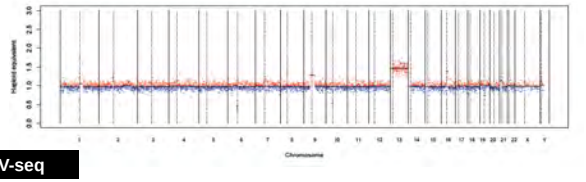


Work done using standard Illumina pipeline

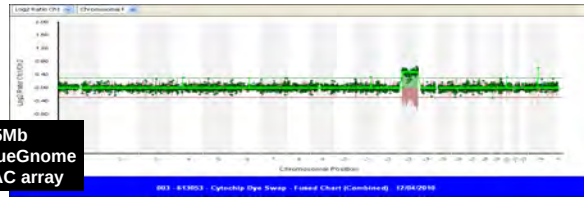


Non – standard post sequencing analysis: splitting the genome into windows, counting reads in each window, calling breakpoints. Custom written scripts and aCGH based packages.

Aneuploidy detection

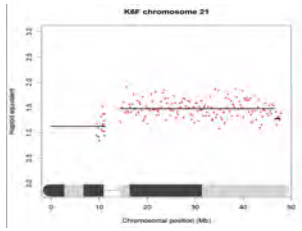


CNV-seq

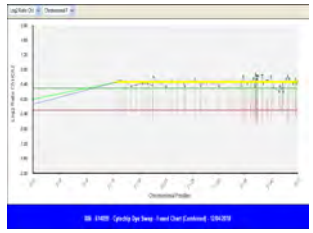


0.5Mb BlueGnome BAC array

Aneuploidy detection



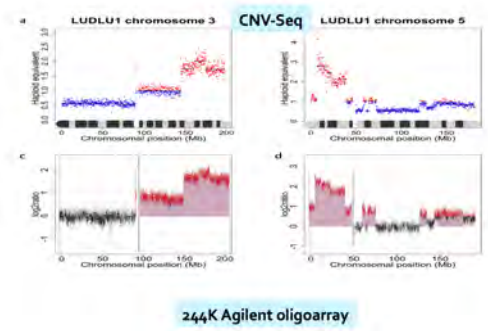
CNV-Seq

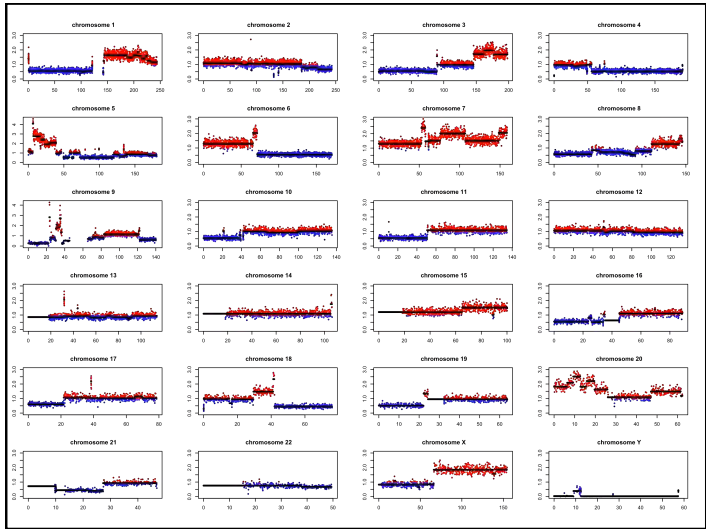


Array CGH

Chromosome 21 view showing trisomy

SCC Lung Tumour Cell lines

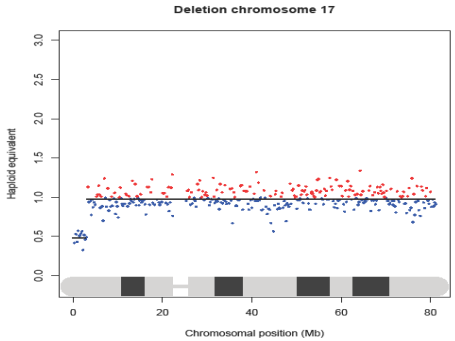




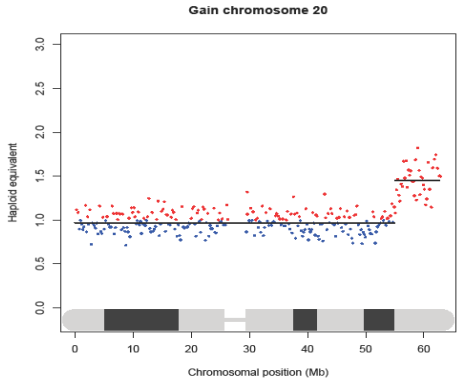
Prenatal CNV-Seq

- Bilateral ventriculomegaly on US at 21 weeks
- Amniocentesis
  - qfPCR normal
  - Long term culture normal karyotype
- CNV-Seq

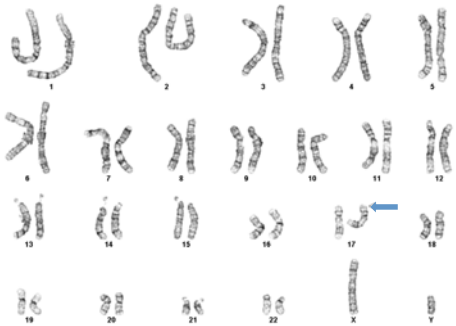
Chr 17: 1- 2 553 180 deletion (2.5Mb)



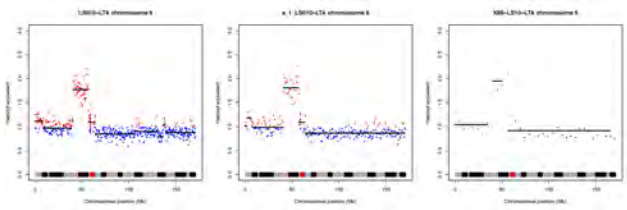
Chr 20: 54 963 376 – 62 332 309 (7.3 Mb gain)



Fetal karyotype
46XY, der(17)t(17:20)(p13.3;q13.3) pat



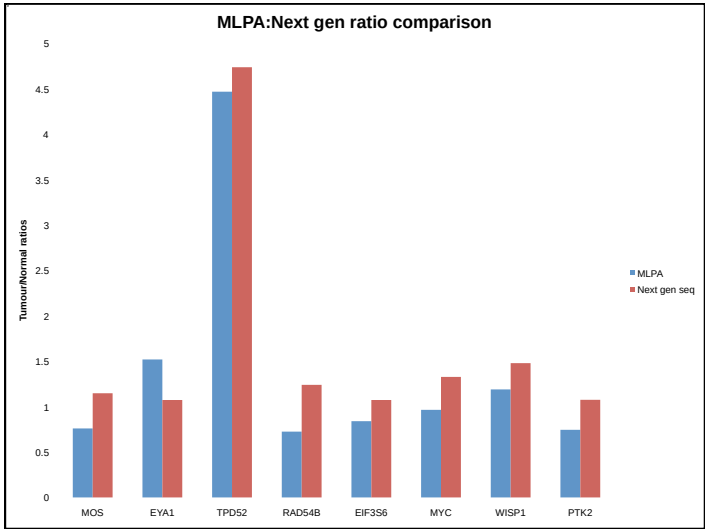
5X multiplex 10X multiplex 80X multiplex



Increased multiplexing decreases resolution, but still allows for basic patterns to be seen at much higher throughput and lower cost

Multiplexing and effect on resolution

Samples per lane (cost)	Number of reads	Mean distance between reads	Mean size of 200 read window
2 (£500)	5 million	620bp	123Kb
5 (£200)	2 million	1165bp	233Kb
10 (£100)	1 million	3062bp	612Kb
80 (£12)	125,000	27Kb	5.5Mb



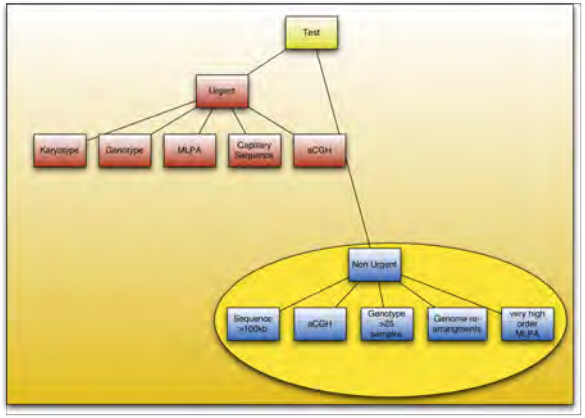
CNV-seq Summary

Next generation sequence platforms can produce copy number data from genomic DNA.

Data is comparable with aCGH data.

Tagging allows levels of multiplexing – hence throughput, cost and resolution can be tailored to the needs of a test.

Service design



Coming soon…

- Structural changes
- Circulating DNA
- Single molecule sequencing

Structural changes

Nature Genetics **40**, 722 - 729 (2008)
Published online: 27 April 2008 | doi:10.1038/ng.128

Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing

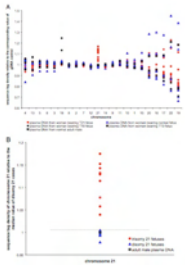
Peter J Campbell[1,4], Philip J Stephens[1,4], Erin D Pleasance[1], Sarah O'Meara[1], Heng Li[1], Thomas Santarius[1,3], Lucy A Stebbings[1], Catherine Leroy[1], Sarah Edkins[1], Claire Hardy[1], Jon W Teague[1], Andrew Menzies[1], Ian Goodhead[1], Daniel J Turner[1], Christopher M Clee[1], Michael A Quail[1], Antony Cox[1], Clive Brown[1], Richard Durbin[1], Matthew E Hurles[1], Paul A W Edwards[2], Graham R Bignell[1], Michael R Stratton[1] & P Andrew Futreal[1]

Circulating DNA

Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan*, Yair J. Blumenfeld*, Usha Chitkara*, Louanne Hudgins*, and Stephen R. Quake*†

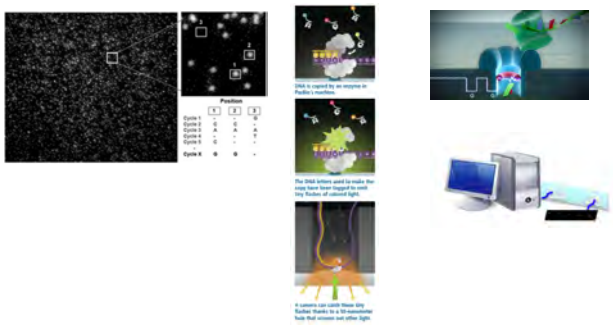
www.pnas.org/cgi/doi/10.1073 pnas.0808319105



Summary & Conclusions

- Clonal sequencing can be applied to diagnostic casework with high sensitivity and specificity
- The cost and efficiency benefits will depend on sequence quality, assay design and scaling
- The capacity is huge: 1 machine delivers over 20 Gigbases per week
- Clonal sequencing will contribute to improved genetic diagnostics

Single Molecule Sequencing



Acknowledgements

Funding

- Emmandjay Trust
- Cancer Research UK
- Dept Health NEAT
- LTHT Challenge Fund

Science

- Jo Morgan
- Aengus Stewart
- Heather Fraser
- Nick Camm
- Helen Lindsay
- Chris Watson
- Kelly Cohen
- Bruce Hayward
- Henry Wood
- Antigone Tzika
- David Bonthron
- Colin Johnson
- Ian Carr
- Constanze Bonifer
- Reuben Tooze
- Carol Chu
- Ruth Charlton
- Pamela Rabbitts
- Sir Alex Markham
- many others…..