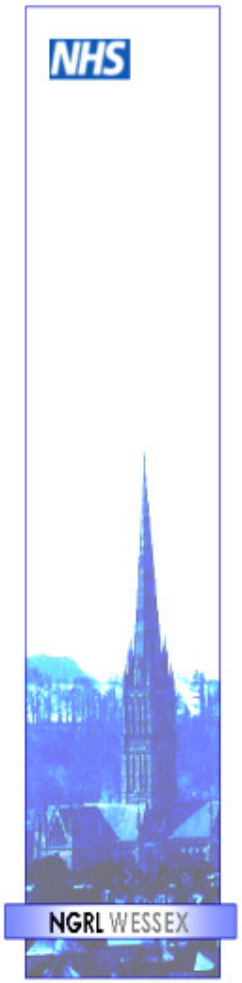
Multi-platform comparison for array-CGH in diagnostic constitutional applications and new high throughput workflow



Shuwen Huang

National Genetics Reference Laboratory (Wessex)

New and Developing Technologies for Genetics Diagnostics 2010

DH Department of Health

Outline of the talk

- Why should we do this?
- How have we carried this out?
- What results have we obtained?
- What will we be doing now?

Why should we do this?

- Copy number variants are associated with developmental delay, mental retardation, congenital anomalies and autistic spectrum disorders (DD/MR/CA)

- Array Comparative Genomic Hybridization (aCGH) is now the first line test for patients with DD/MR/CA in many developed countries

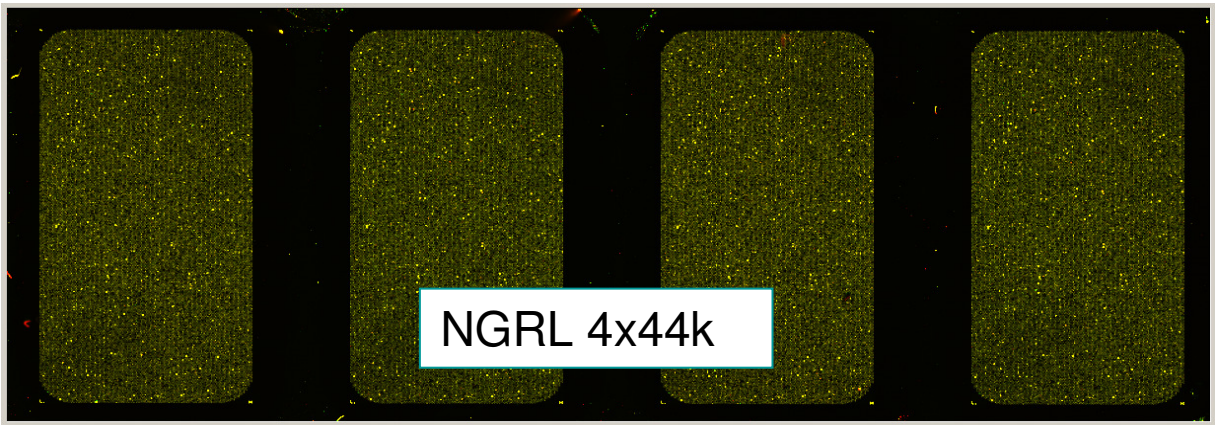
- Several commercial platforms are available which can either be used directly or customised to meet specific diagnostic criteria

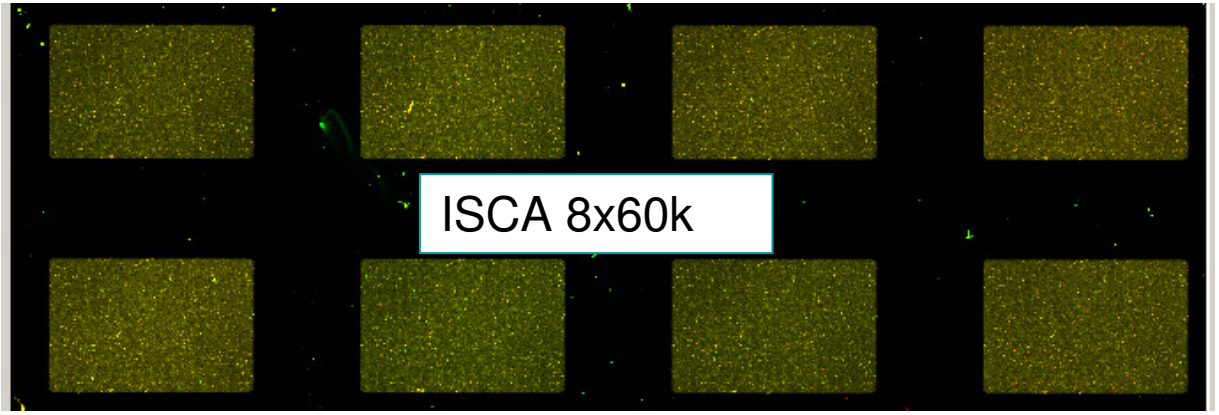
- An evaluation and comparison of these platforms is important so that data are available to determine which platform (s) are most suitable for the clinical cytogenetics diagnostic setting

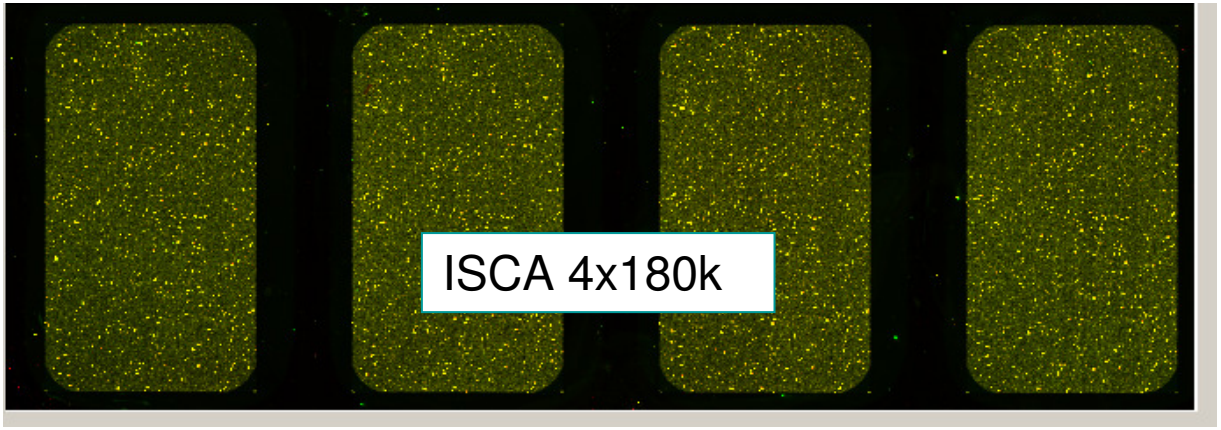
The basic comparison elements:

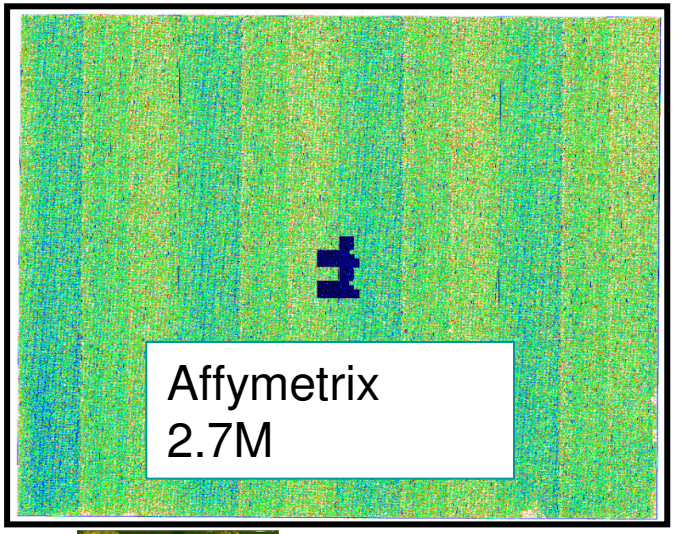
- Core sample set
  - 12 samples with known and varied chromosomal abnormalities
  - Enough genomic DNA to be run on different platforms
- Platforms
  - Agilent
    - NGRL 4x44k
    - ISCA 8x60k (OGT)
    - ISCA 4x180k
  - Affymetrix — Affymetrix 2.7M Cytogenetics Research Solution
  - NimbleGen — NimbleGen CGX-12
- Software
  - For Agilent array
    - Agilent DNA Analytics
    - BlueGnome Bluefuse Multi
    - OGT CytoSure Interpret
  - For Affymetrix array — ChAS
  - For NimbleGen array — NimbleScan v2.5 & Genoglyphix

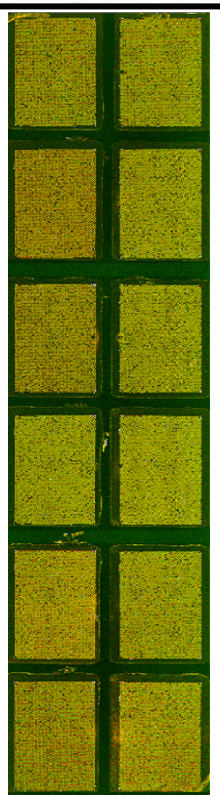
Examples of scan images of the platforms





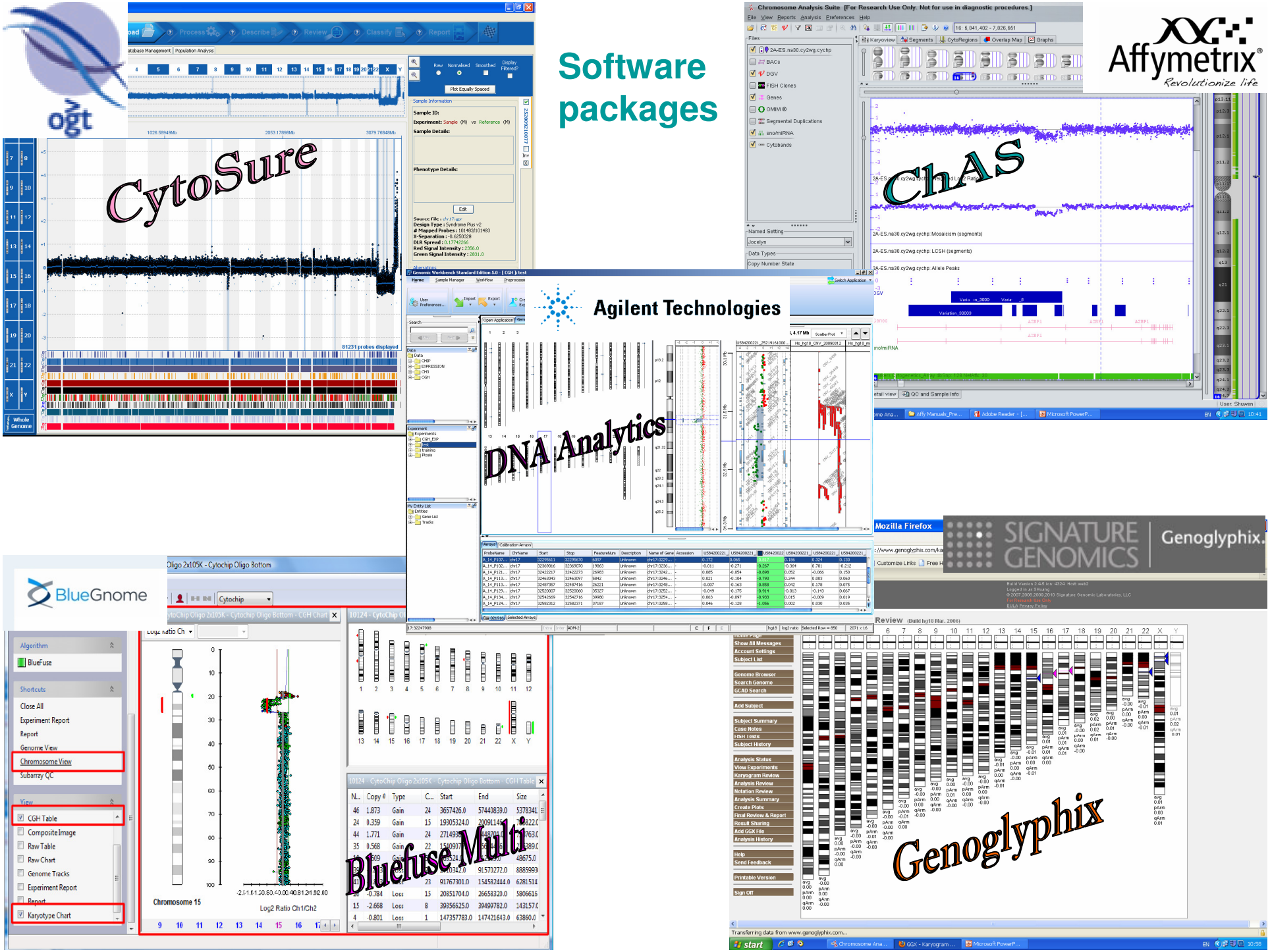






NimbleGen
CGX-12

Software packages

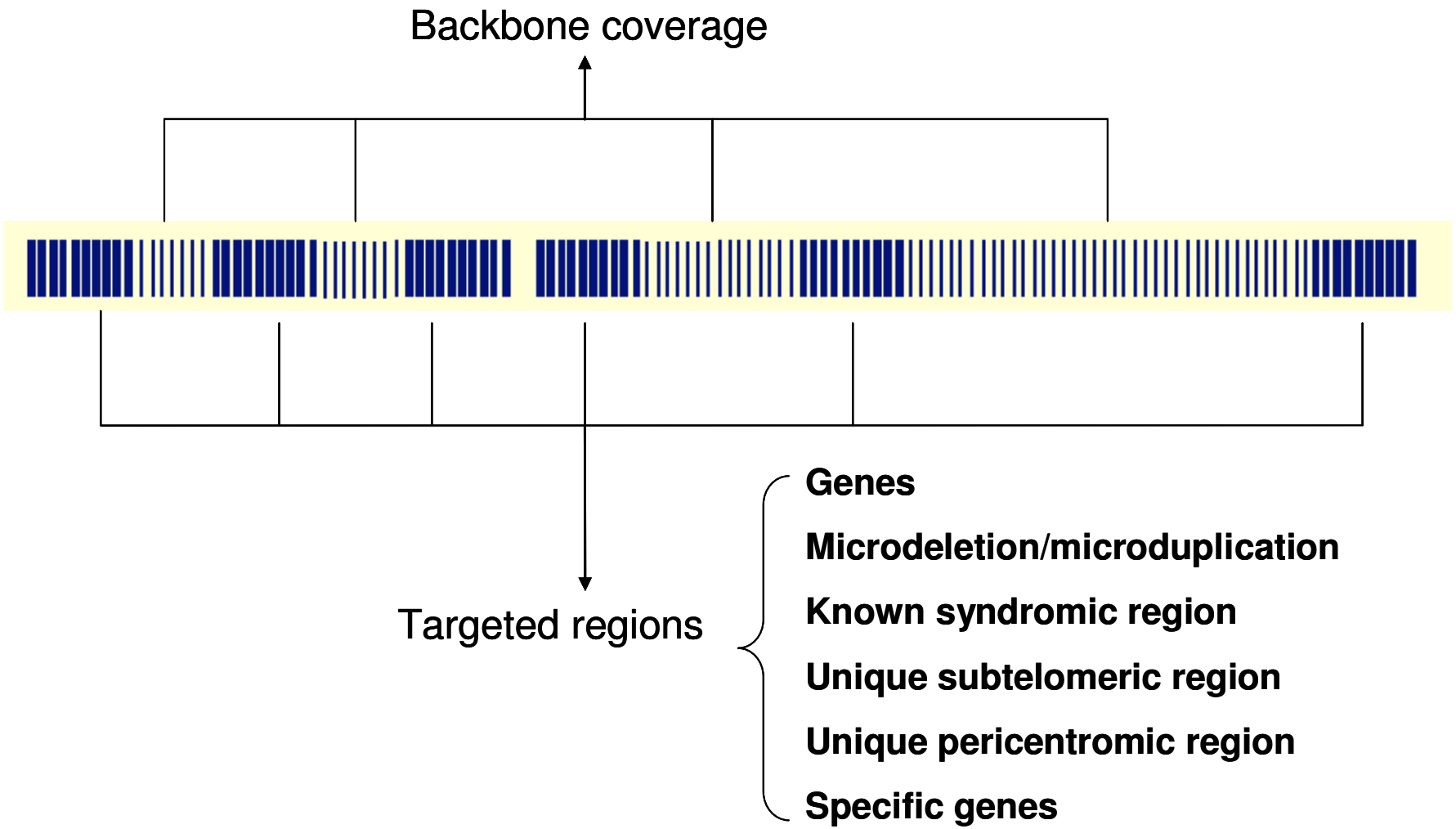


Comparisons

- Design strategy
- Laboratory processing
- Array quality metrics
- Software and analysis processing
- Detection rate
- Breakpoint resolution
- Overall consumable cost

Design strategy

Whole genome plus targeted array design



Design strategy comparison

	Affymetrix 2.7M	NimbleGen CGX-12	ISCA 4x180k (Agilent)	ISCA 8x60k (OGT/Agilent)	NGRL 4x44k (Agilent)
Backbone coverage	1kb/probe	35k/probe	25kb/probe	60kb/probe	75kb/probe
Target region coverage	690bp/probe	10kb/probe	5kb/probe or 20 probes per targeted region (or on average, 50 probes per targeted region)	On average, 40 probes per targeted region	Two to five extra probes per targeted regions
Backbone resolution	Depend on number of markers, size and confidence	175kb (5 contiguous probes to make a call)	100kb (4 contiguous probes to make a call)	240kb (4 contiguous probes to make a call)	225kb (3 contiguous probes to make a call)
Target region resolution	If the filter set as 20 markers within 50kb, the resolution should be 20x690 for the targeted region	50kb (5 contiguous probes to make a call)	20kb (4 contiguous probes to make a call)	~48kb (4 contiguous probes to make a call)	~60kb (3 contiguous probes to make a call)
Probe types	2.3 million non-polymorphic markers for CNV, and 400,000 SNP markers for LOH, UPD, and regions identical-by descent	Non-polymorphic markers for CNV	Non-polymorphic markers for CNV	Non-polymorphic markers for CNV	Non-polymorphic markers for CNV
Probe size	49 mer	60 mer	60 mer	60 mer	60 mer
How many targeted regions are included	Coverage (%) if using filter of 20 markers with minimum size of 50 kb RefSeq genes (18,701) 18,533 (99.1%) Cancer genes (318) 318 (100%) Cytogenetics relevant/haploinsufficiency genes (559) 548 (98.0%) X chromosome genes (801) 786 (98.1%) OMIM genes (12,341) 12,242 (99.2%)	Over 700 genes, 200 recognized genetic syndromes, 41 unique subtelomeric regions, 43 unique pericentromeric regions. Regions tested could be viewed at http://www.signaturegenomics.com/disorders_tested.html	501 targeted regions	498 targeted regions	155 targeted regions (http://www.ngrl.org.uk/Wessex/microdel_collection.htm)
Customer self design availability	No	No	Yes	Yes	Yes

Design strategy comparison

	Affymetrix 2.7M	NimbleGen CGX-12	ISCA 4x180k (Agilent)	ISCA 8x60k (OGT/Agilent)	NGRL 4x44k (Agilent)
Backbone coverage	1kb/probe	35k/probe	25kb/probe	60kb/probe	75kb/probe
Target region coverage	690bp/probe	10kb/probe	5kb/probe or 20 probes per targeted region (or on average, 50 probes per targeted region)	On average, 40 probes per targeted region	Two to five extra probes per targeted regions
Backbone resolution	Depend on number of markers, size and confidence	175kb (5 contiguous probes to make a call)	100kb (4 contiguous probes to make a call)	240kb (4 contiguous probes to make a call)	225kb (3 contiguous probes to make a call)
Target region r	If the filter set as 20 markers within 50kb, the resolution should be 20x690 for	50kb (5 contiguous probes	20kb (4 contiguous	~48kb (4 contiguous	~60kb (3 contiguous a call)
Pr					phic CNV
P					
How region					egions rg.uk/W collectio
Cus design availability					

- Different platforms have different backbone and targeted region resolution
- Targeted regions vary between different platforms
- Probe distributions (BED files) will be published on the NGRL(W) website.
- Agilent based array platforms provide facilities for the customers to be directly involved in the design

Laboratory processing

Platform	Affymetrix	NimbleGen CGX	Agilent		
Array	**2.7M Cytogenetics Research Solution**	**CGX-12**	**ISCA 4x180k**	**ISCA 8x60k (OGT)**	**NGRL 4x44k**
Amount of DNA needed	0.1µg	1ug	0.5-1.5µg	0.5-1.0ug	0.5-1.5µg
DNA amplification procedure needed?	Yes	No	No	No	No
Overall time to complete procedure	3 days	3-4 days	2-3 days	2-3 days	2-3 days
Hands-on time	~3 hours for 8 samples	~7 hours	~7 hours for 8 samples	~7 hours for 8 samples**	~7 hours for 8 samples
Scanner resolution	2.5µm	5µm***	3µm	3µm	5µm
Volume of data generated per sample	455Mb	250Mb	756Mb	710Mb	239Mb
Scan time (per slide)	15mins	12min	15mins	15mins	8 mins
Extraction of data time (per sample)	14mins	3mins	3.5mins	1.9mins	1.2mins
Overall time* spent processing data (per sample)	29min	15mins	18.5mins	16.9mins	9.2mins

* Overall time = scan time + data extraction time

** Significantly reduced by semi-automation (see page 14)

*** Scanner resolutions of 2, 2.5 and 3µm are also available.

Array quality metrics

Platform	**Affymetrix**	**NimbleGen**	**Agilent**		
Software	**ChAS*** (2.7M)	**NimbleScan v2.5 & Genoglyphix** (12x135k)	**DNA Analytics*** (4x180k, 8x60k, 4x44k)	**Bluefuse Multi*** (4x44k)	**CytoSure Interpret*** (4x44k)
Major parameters and high quality array reference values	**MAPD < 0.27**	**SD<0.14**	**DLRS <0.2**	**SD<0.1**	**DLRS <0.2**
	SNP QC >1.1	Mad.1dr <0.23	Signal to noise green >100	**DLRS <0.15**	Green Signal to Noise Ratio > 100
	Interquantile range (iqr)	Interquartile density 0.5-1.8	Signal to noise red >100	Mean Ch1 Spot Amplitude: 400-700	Red Signal to Noise Ratio > 100
	Median-raw-intensity >2000	Ratio range <1.0	Signal intensity green >150	Mean Ch2 Spot Amplitude: 400-700	Green Signal Intensity > 150
	Antigenomic ratio <0.4	Signal range <1.5	Signal intensity red >150	SBR Ch1: 5-20	Red Signal Intensity > 150
	Waviness-seg-count <100		BG noise Green <5	SBR Ch2: 5-20	Green Background Noise < 5
	Waviness-sd <0.1		BG Noise Red <5		Red Background Noise < 5
			Reproducibility Green <0.05		Signal Intensity Ratio > 0.7
			Reproducibility Red <0.05		Non-Uniform Features < 0.005 (0.5%)
					Saturated Features < 0.005 (0.5%)
Value comparison of major parameter for the 12 core samples	**MAPD=0.18** **SNP QC=2.28**	**SD=0.15**	**DLRS** **4x44k = 0.16** **8x60k = 0.12** **4x180k = 0.15**	**SD=0.1** **DLRS=0.14**	**DLRS=0.15**

* These are obtained under our laboratory conditions and so may vary when used locally.

Software and analysis processing

Array Company	**Affymetrix**	**NimbleGen**	**Agilent**				
Software	**ChAS v1.0.1**	**Genoglyphix v2.4 (Signature Genomics)**	**DNA Analytics v4.0**			**BlueFuse Multi v2.1 (BlueGnome)**	**CytoSure Interpret v3.0.6 (OGT)**
Array platform	**2.7M**	**CGX-12**	**4x180k**	**8x60k**	**4x44k**	**4x44k**	**4x44k**
Total analytical time per case* (Average) in minutes	40	30	45	35	30	45	30
Aberration filters	100kb with 85 markers for amp + del, 50kb with 20 markers for cyto relevant regions 5Mb for LCSH$^{\Psi}$ with 3 markers and mosaicism with 500 markers, 85% Confidence	5 contiguous probes	4 contiguous probes	4 contiguous probes	3 contiguous probes	3 contiguous probes	3 contiguous probes
Algorithms		NG packager segmentation algorithm	ADM-2			Multi v1.0	CBS
Tracks included	BACs, DGV, FISH Clones, Genes, OMIM, Segmental Duplications, sno/miRNA	FISH probes, Sequence Gaps, Segmental Duplications, GC Content, SignatureSelect Clones, SignatureSelect OS 105K Probes, NimbleGen CGX Probes, SignatureSelect OS 44K Probes, Abnormal Region(s), MyGCAD, Community GCAD, GCAD, Benign CNVs, Genes, RefSeq Genes, SGL GPS, SGL CNVs, DGV	Genes, DGV, CpGIsland, miRNA, PAR			Disease, Genes, BlueFISH, BAC Gain/ BAC Loss, Oligo Gain/ Oligo Loss, DGV Gain/ DGV Loss, BG Gain/ BG Loss	Syndrome, Gene, Exon, CHOP CNV, ECARUCA, Recombination hotspot, DGV, Confirmation (FISH and MLPA probes), DECIPHER, Redon CNV
Possibility of adding custom tracks	Yes	Yes	Yes	Yes	Yes	No	Yes

* Total analytical time including analysing, checking and authorising

$^{\Psi}$ LCSH= Long contiguous stretches of homozygosity

Software and analysis processing

Array Company	Affymetrix	NimbleGen	Agilent				
Software	ChAS v1.0.1	Genoglyphix v2.4 (Signature Genomics)	DNA Analytics v4.0			BlueFuse Multi v2.1 (BlueGnome)	CytoSure Interpret v3.0.6 (OGT)
Array platform	2.7M	CGX-12	4x180k	8x60k	4x44k	4x44k	4x44k
Total analytical time per case* (Average) in minutes	40	30	45	35	30	45	30

- Every software platform has positive features
- No single platform can perform all tasks required
- Human interaction (and interpretation) is still needed
- Different settings or annotation files can generate different results
- Annotation tracks included in the software are very useful
- Customers can add their own tracks on most software platforms

* Total analytical time including analysing, checking and authorising

$^{\Psi}$ LCSH= Long contiguous stretches of homozygosity

Detection rate and breakpoint resolution

Platform	**Affymetrix**	**NimbleGen**	**Agilent**		
Array	**2.7M Cytogenetics Research Solution**	**CGX-12**	**ISCA 4x180k**	**ISCA 8x60k (OGT)**	**NGRL 4x44k**
What can be detected	Copy Number Changes	Copy Number Changes	Copy Number Changes	Copy Number Changes	Copy Number Changes
	Mosaicism	Mosaicism	Mosaicism	Mosaicism	Mosaicism
	LCSH*				
Overall detection rate	100%	100%	100%	93% Missing dup (6)(q25.1) on case 11	93% Missing A2BP1 gene deletion on case 1)
Average breakpoint resolution (not mediated by known segmental duplications) (bp)	7665	27670	18307	43255	73073
Average breakpoint resolution (mediated by known segmental duplications) (bp)	43952	510138	24228	74030	272029
Overall breakpoint resolution (bp)	25809	268904	21268	58643	172551
Average number of calls per patient	3.5	9.67	34.2	7.3	2.67
Calls on genic regions	3.2	7.17	25.5	5.5	2.41
Calls on non-genic regions	0.3	2.5	8.7	2	0.16

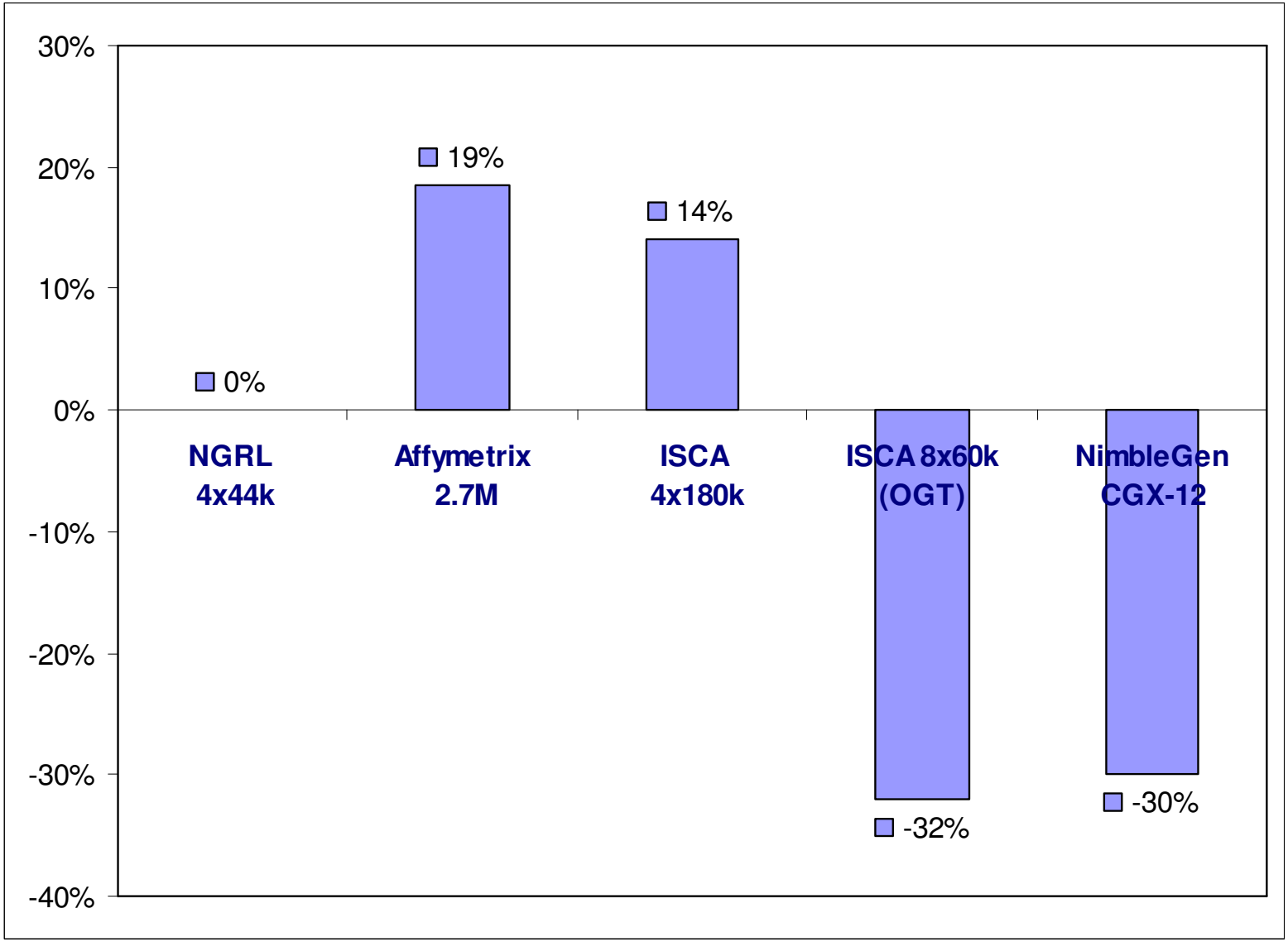
* LCSH= long contiguous stretches of homozygosity

Detection rate and breakpoint resolution

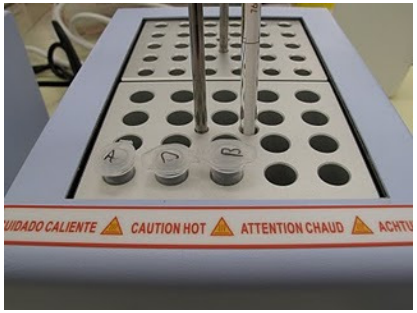
Platform	Affymetrix	NimbleGen	Agilent		
Array	**2.7M Cytogenetics Research Solution**	**CGX-12**	**ISCA 4x180k**	**ISCA 8x60k (OGT)**	**NGRL 4x44k**
What can be detected	Copy Number Changes	Copy Number Changes	Copy Number Changes	Copy Number Changes	Copy Number Changes
	Mosaicism	Mosaicism	Mosaicism	Mosaicism	Mosaicism
	LCSH*				
Overall detection rate	100%	100%	100%	93% Missing dup (6)(q25.1) on case 11	93% Missing A2BP1 gene deletion on case 1)

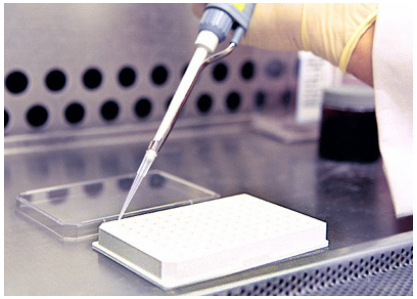
- The main elements are: what can be detected and how well they can be detected.
- All these arrays can detect copy number change and mosaicism. Affymetrix array can also detect LCSH
- The detection rate is related to the array resolution, but other factor (s) (e.g. reference DNA used) also need to be considered
- The breakpoint resolution is consistent with the array resolution, but is also related to the settings, algorithms and/or annotation files used

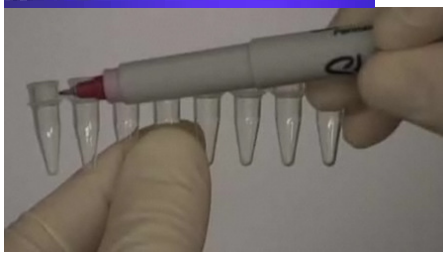
Overall consumable costs



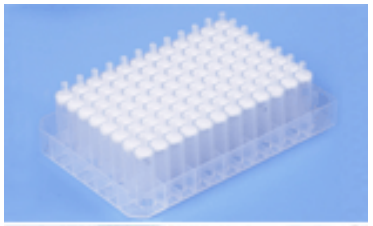
Wet lab development and semi-automated workflow



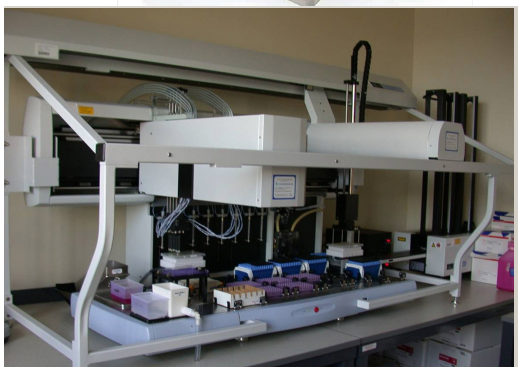


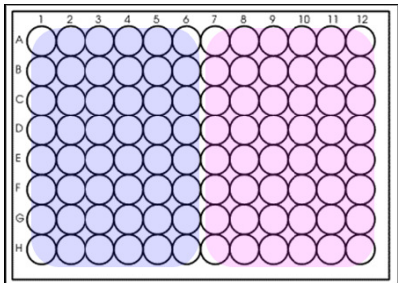


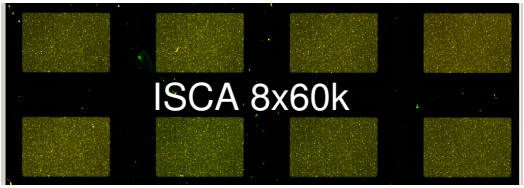
Manual method



96-well Purification System







semi-automation

Pre No	Position	DNA No	Sample (First) Name	Sample Surname	DNA Conc (ng/ul)	Volume of DNA (ul)	Volume of dH20 (ul)	DNA Used (ng)	Array No	Slide No
1	**A1**				148.2	3.37	6.73	500		
2	**A2**				100.2	4.99	5.11	500		
3	**A3**				171.0	2.92	7.18	500		
4	**A4**				97.9	5.11	4.99	500		
5	**A5**				157.0	3.18	6.92	500		
6	**A6**				108.0	4.63	5.47	500		
7	**A7**				175.0	2.86	7.24	500		
8	**A8**				112.5	4.44	5.66	500		
9	**B1**				177.3	2.82	7.28	500		
10	**B2**				184.7	2.71	7.39	500		
11	**B3**				205.2	2.44	7.66	500		
12	**B4**				146.5	3.41	6.69	500		

Testing female

Testing male

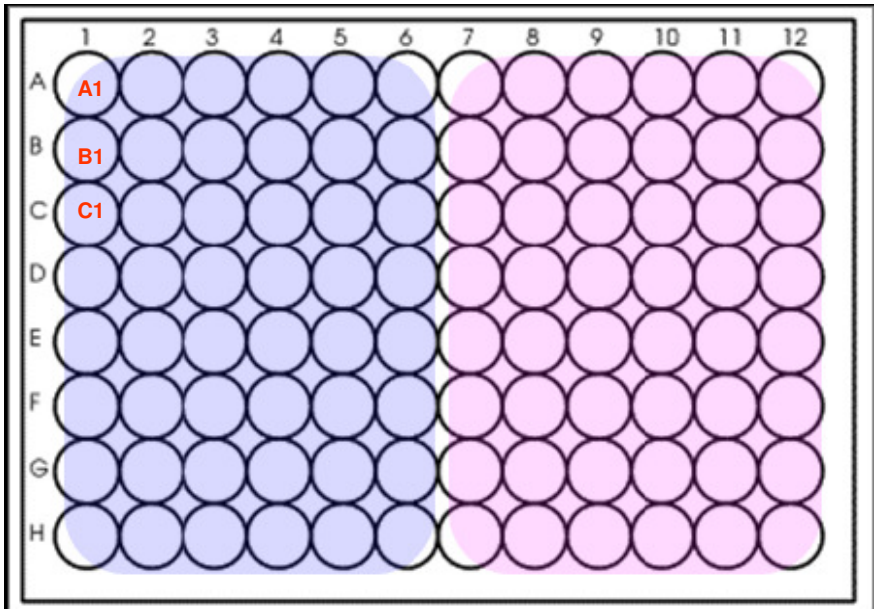
Control female

Control male

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	51	59	67	75	81	89
B	2	10	18	26	34	42	52	60	68	76	82	90
C	3	11	19	27	35	43	53	61	69	77	83	91
D	4	12	20	28	36	44	54	62	70	78	84	92
E	5	12	21	29	37	45	55	63	71	79	85	93
F	6	14	22	30	38	46	56	64	72	80	86	94
G	7	15	23	31	39	47	57	65	73	81	87	95
H	8	16	24	32	40	48	58	66	74	82	88	96

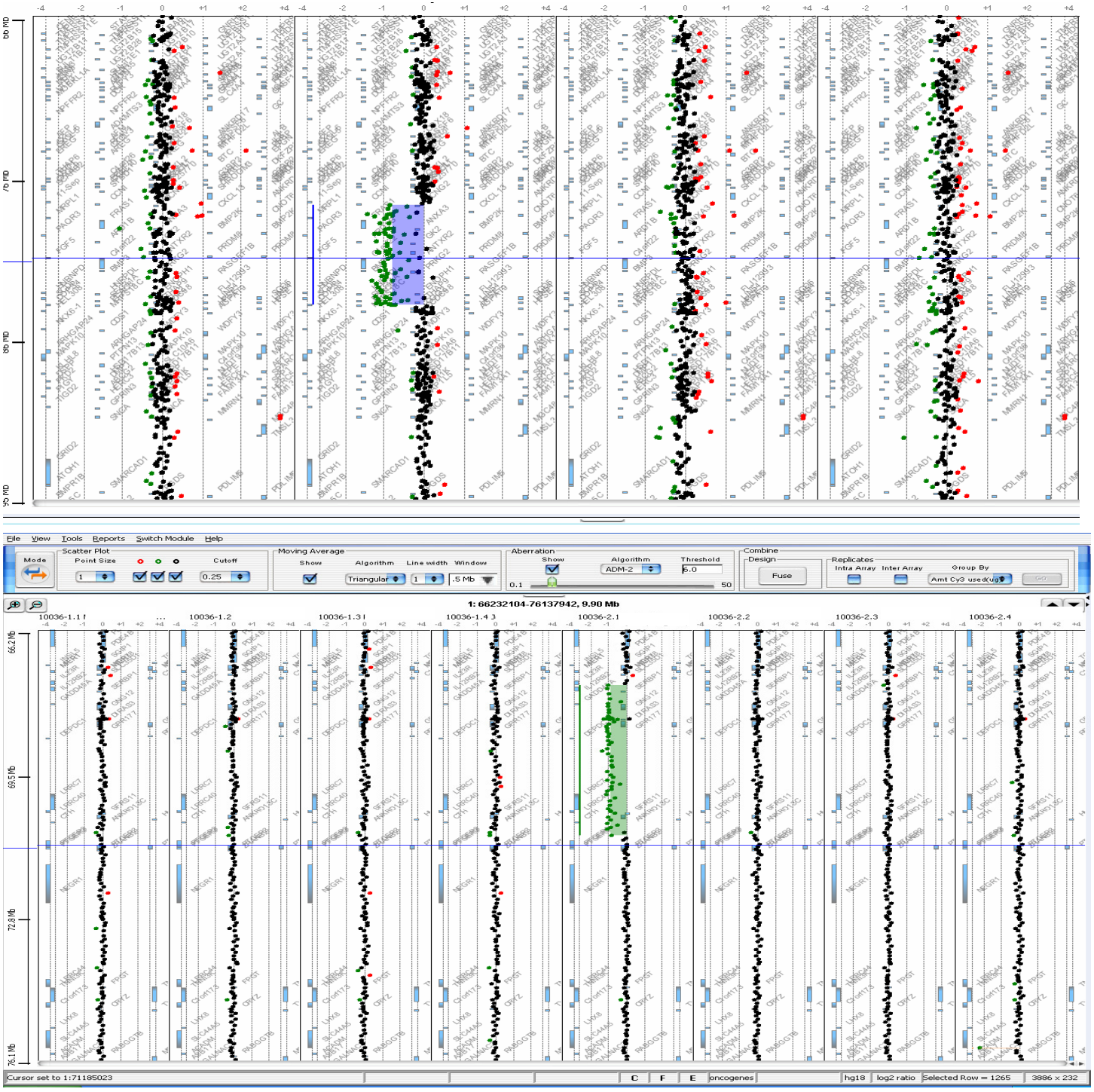
Testing samples

Control samples



Comparison between manual method and semi-automated workflow

Comparisons	Manual	semi-automated
Array platform	NGRL 4x44k	ISCA 8x60k
DNA QA	single channel Nanodrop	8 - channel Nanodrop
Sample handling	1.5ul Microcentrifuge tubes	96-well plate
DNA preparation	maually with single channel pippette	liquid handling robot
ID marking on tube	Yes	No
Each run	8 samples	48 samples
Labeled DNA purification	filter columns	96-well purification module
Dye incorporation checking	single channel Nanodrop	8 - channel Nanodrop
Wet lab hands-on time	7 hours	5 hours
Average hands on time/patient	0.875 hours	0.104 hours
Efficiency	1	8.4
DLRS calculation	160 samples (Dec 2009-Jan 2010)	160 samples (May 2010)
DLRS > 0.25	13	2
Average DLRS (DLRS<0.25)	0.16612	0.12387



4x44k array

8x60k array

Summary

- Twelve core samples were run on 5 platforms and analysed using 5 software packages
- All platforms adopted a whole genome plus targeted design strategy
- Design flexibility, array quality and resolution, wet lab processing, software analysis, and consumable cost, etc. should be considered when a decision is made
- We selected ISCA 8x60k array provided by OGT
- We have developed semi-automated workflows
- More efficient and economic workflows are being investigated

Aknowledgement

NGRL (W) Microarray Laboratory (NWML)

- Agilent Technology
- Affymetrix
- NimbleGen
- Signature Genomics
- Oxford Gene Technology
- BlueGnome

- Wessex Regional Genetics Laboratory (WRGL) high throughput lab (Dan Ward)