beadarray: An R Package for Illumina BeadArrays

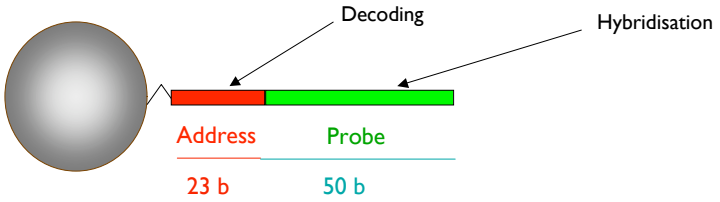
Mark Dunning - md392@cam.ac.uk

PhD Student - Computational Biology Group,
Department of Oncology - University of Cambridge

http://www.bioconductor.org/packages/bioc/1.8/html/beadarray.html

UNIVERSITY OF CAMBRIDGE The Hutchison/MRC Research Center

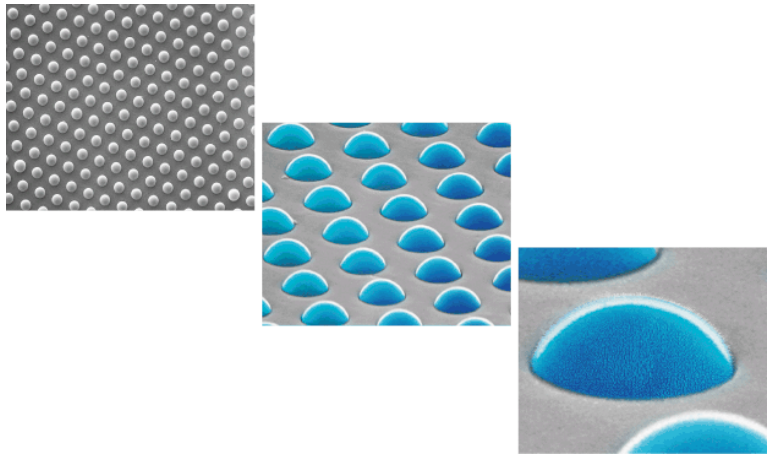
The Bead



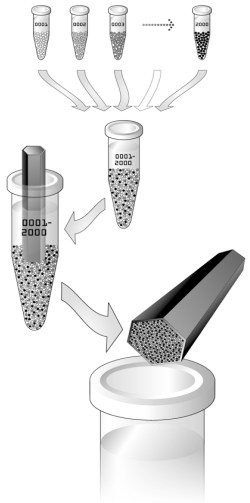
Each silica bead is 3 microns in diameter

700,000 copies of same probe sequence are covalently attached to each bead for hybridisation & decoding

Beads in Wells

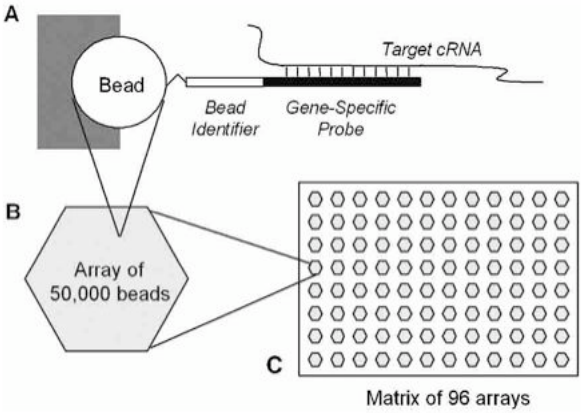


Bead Preparation and Array Production



- Bead pools produced containing 384 to 24,000 bead types
- Wells created in either fibre optic bundle (hexagon) or chip (rectangle) & exposed to array
- Beads self-assemble into wells to form **randomly arranged** array of beads
- *Average of 30 beads of each type*
- Each array produced separately

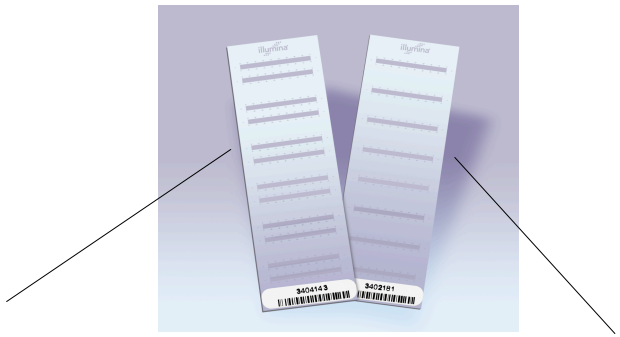
Combining Arrays - The SAM



Beads 6 microns apart
~1500 bead types on array ~30 of each type
1 array = 1 sample or treatment
96 arrays processed in parallel - **High throughput**

The SAM

Combining Arrays - BeadChips



Whole Genome

6 arrays per chip: 2 strips = 1 array

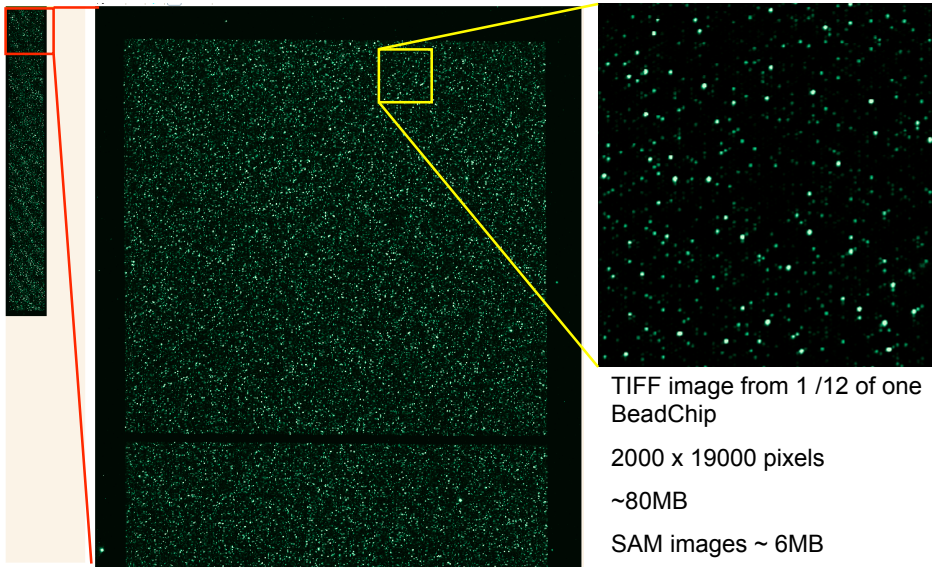
48,000 bead types (24,000 RefSeq + 24,000 supplemental) on each array

RefSeq BeadChip

8 arrays per chip 1 strip = 1 array

24,000 bead types from RefSeq database x 30 reps on each array

Whole Genome TIFF images



TIFF image from 1 /12 of one BeadChip

2000 x 19000 pixels

~80MB

SAM images ~ 6MB

Data Formats - Bead Level

Bead Level = information about each **bead** on an array

One TIFF for each array - 12 for BeadChip, 96 for SAM

The latest version of Illumina **scanning software** will give information for each bead on an array (BeadStudio will not give this)

Output is a csv (Excel) file with 50,000 rows for SAM ~ 1.1 million for BeadChip

	A	B	C	D	E	F	G
1	Index	Code	Grn	Red	Outlier	x	y
2	1	859	395	0	0	888	852
3	2	56	546	0	0	883	849
4	3	2256	1208	0	0	888	846
5	4	1002	586	0	0	893	849
6	5	1009	393	0	0	893	855
7	6	1099	685	0	0	887	858
8	7	3520	2857	0	0	882	855
9	8	2022	573	0	0	878	845
10	9	5788	4794	0	0	883	843
11	10	4848	596	0	0	889	840
12	11	3780	543	0	0	894	843
13	12	1335	570	0	0	899	846
14	13	4268	15775	0	0	898	852
15	14	2363	435	0	0	898	859
16	15	5633	559	0	0	893	861
17	16	840	321	0	1	887	864
18	17	2907	581	0	0	882	861
19	18	1032	593	0	0	877	858
20	19	83	513	0	0	877	851

Data Formats - Bead Summary

Illumina provide software (BeadStudio) to read raw data and produce a single foreground intensity value for each **bead type** after outliers have been excluded and background has been removed

A single file may be generated describing **all arrays** in the experiment with arrays listed along the page

One row for each gene in the experiment

TargetID	AVG_Signal-Array1	BEAD_STDEV-Array1	Avg_NBEADS-Array1	Detection-Array1	AVG_Signal-Array2
GI_100470	122.6	7.1	38	0.30850363	149.6
GI_100470	172.6	13.6	31	0.98879367	165.6
GI_100470	594.3	24.4	42	1	849.3
GI_100470	599.2	31.9	36	1	264.5
GI_100471	1123.8	55	49	1	1840.5
GI_100471	175.8	13.7	35	0.99077126	145.7
GI_100471	139.9	9.9	39	0.71786421	128.2
GI_100471	882	32.4	38	1	1131.5
GI_100471	115.1	9.2	42	0.17073171	132.8
GI_100471	158.2	9.4	34	0.94660514	124.9
GI_100925	110.7	5.5	43	0.10678972	126.4
GI_100925	194	11.1	46	0.9993408	184.6

Current Analysis Methods

Illumina application BeadStudio gives **average** value for each bead type on the un-logged scale and provides various normalisation and visualisation tools

Lose information about 30 replicates of each bead type

Data is automatically background corrected. ie No control over image processing

The 'beadarray' Library

Collection of BeadArray analysis functions written using **R**

Functions for reading SAM and BeadChip data in bead summary or bead level format

Options for image processing

Also quality control, diagnostic checks and normalisation

Compatible with *limma*, *affy* packages (uses objects similar to 'RGList')

http://www.bioconductor.org/packages/bioc/1.8/html/beadarray.html

Bead Level Analysis

TIFF Images + bead level csv files

BeadLevelList

R ProbeID x y

1 value per bead per array
Columns = arrays, 1 row is NOT same probe

Image processing inc. background correction
Analyse position and intensity of 30 replicates
Analysis of outliers
Look for spatial effects
Normalisation using all beads → Normalisation of bead summary data

Many ways to use bead replicates for DE statistics and other analyses. → DE and downstream analysis across arrays based on summary data only.

Bead Summary Analysis

BeadStudio output

BeadSummaryList → eSet

R BeadStDev NoBeads

1 value per bead type per array
Columns = arrays, rows = probes

The 'beadarray' Library

Computationally expensive tasks are written in C for efficiency

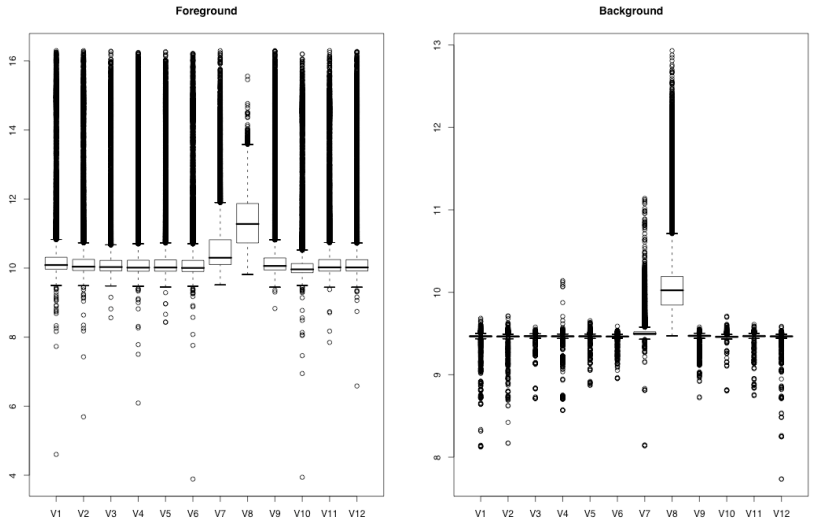
Eg Creating BeadLevelList from TIFF and csv files takes around 1 minute* for each strip on a BeadChip - including time taken for image processing

Converting from BeadLevelList to BeadSummaryList takes around 2 seconds* for each each array on a BeadChip.

However, large amounts of memory (> 1 Gb) are required for these operations

*Running on 3Ghz Pentium IV PC

Bead Level Analysis - Foreground and Background

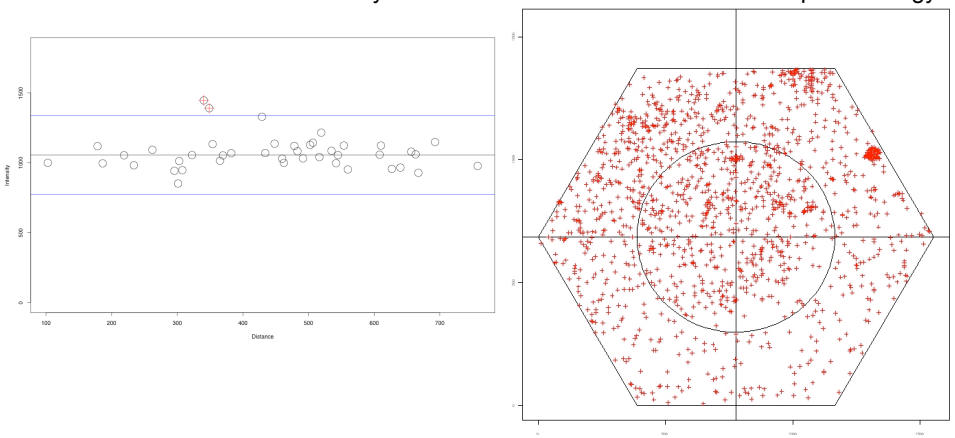


Bead Level Analysis - Outlier Analysis

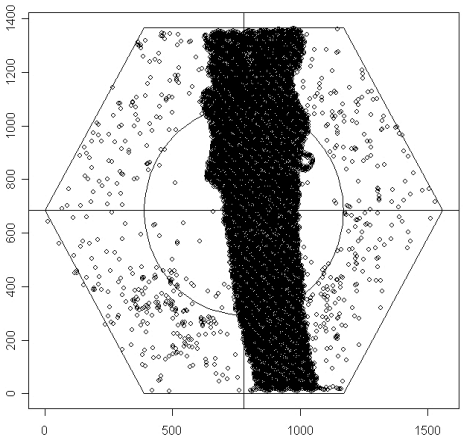
Illumina say outliers are beads > 3 M.A.D from the mean for their bead type

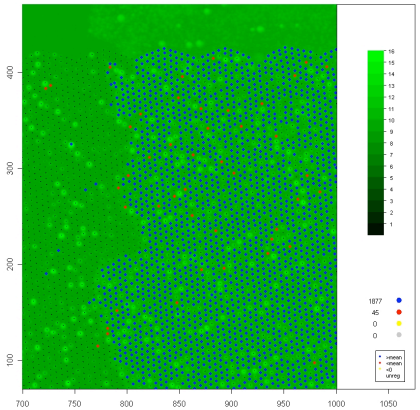
Can plot the position of particular beads or beads of the same type

Around 5% total beads on an array are outliers on both SAM and BeadChip technology



Bead Level Analysis - When BeadArrays go wrong



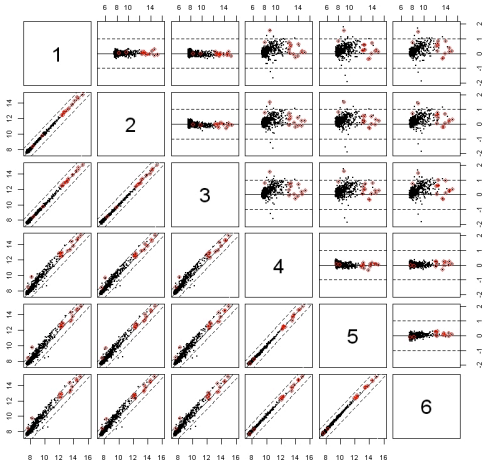


Array with 12,000 outliers, nearly 25% of beads

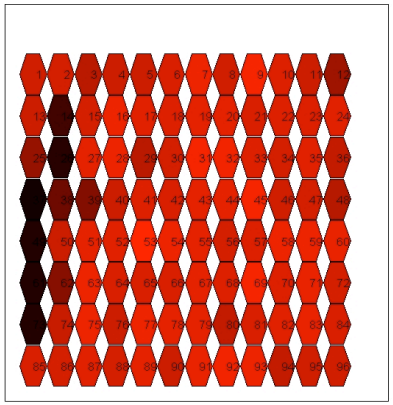
This array is rare example roughly 1 in 100 arrays are “bad”

Bead Summary Analysis - Comparing Arrays

“MAXY” plot for comparing multiple arrays



SAM summary plot for comparing a measured quantity across all 96 arrays



Further Analysis

Since we have an expression matrix, further analysis can proceed as for other microarray technologies

Normalisation can be done using affy package or limma

limma provides tools for linear modeling

Also clustering, PCA methods can be easily applied

We will investigate methods for detecting DE and normalising using the bead level data

Acknowledgements

Computational Biology Group (Cambridge)
Natalie Thorne
Mike Smith
Isabelle Camilier
Simon Tavaré

Illumina (San Diego)
Brenda Kahl
Semyon Kruglyak
Gary Nunn

Dermitzakis group (Sanger Institute - Cambridge)
Manolis Dermitzakis
Barbara Stranger
Matthew Forrest

UCSD(San Diego)
Roman Sasik

References

http://www.bioconductor.org/packages/bioc/1.8/html/beadarray.html

1. MJ Dunning, NP Thorne, I Camilier, M Smith, S Tavaré.
Quality Control and Low-Level Statistical Analysis of Illumina BeadArrays. *Revstat* 4:1-30

2. BE Stranger, M. Forrest, … Genome-wide Associations of Gene Expression Variation in Humans. *PLoS Genet*, 1:695-704, 2005

3. KL Gunderson, S Kruglyak, … Decoding randomly ordered DNA arrays. *Genome Res*, 14:870-877, 2004

4. K Kuhn, SC Baker, … A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res*, 14:2347-2356, 2004

5. KL Gunderson, FJ Steemers, … A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*, 27:549-554, 2005

6. M Barnes, J Freudenberg, … Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic Acids Res*, 33:5914-5923, 2005