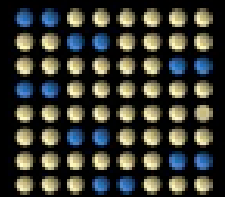


CpG Island Microarrays: Targeting Gene Expression

2nd Annual Canadian Gene
Expression Conference

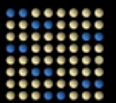
Vancouver BC, March 25th, 2004



UHN Microarray Centre

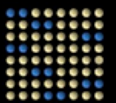
Seminar Overview

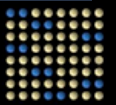
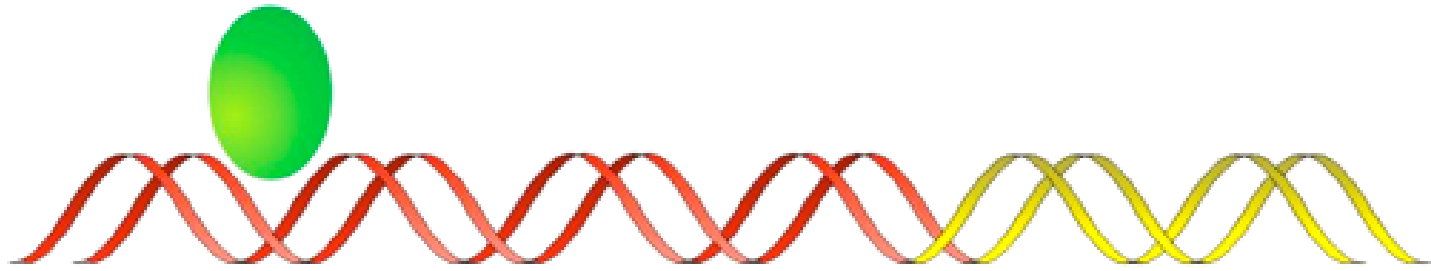
- ChIP (Chromatin Immunoprecipitation)
- CpG and Promoter Microarrays for parallel analysis
- Production Issues
- Informatics
- Epigenetics
 - Methylation analysis using CpG and cDNA microarrays

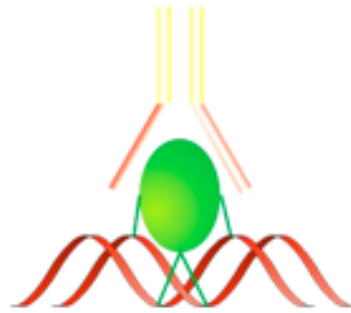


Chromatin Immunoprecipitation (ChIp)

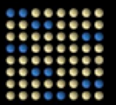
- Typically used to interrogate the occupancy of a particular promoter
- Potentially have the capability to identify novel gene targets associated with specific transcription factors





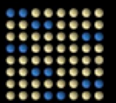


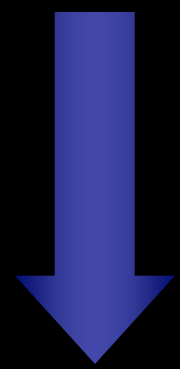
Incubate with specific
1° Antibody



Example: MEF-2 (Myocyte Enhancing Factor 2)

- Muscle transcription factor known to be a key player in the regulation of muscle differentiation
- Regulates the expression of a number of genes such as muscle creatine kinase, skeletal α -actin, myosin light chain, and myoglobin
- Binding sequence: YTA(A/T)₄TAR
- Design primers to flank this region for several regulated genes

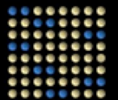




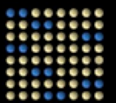
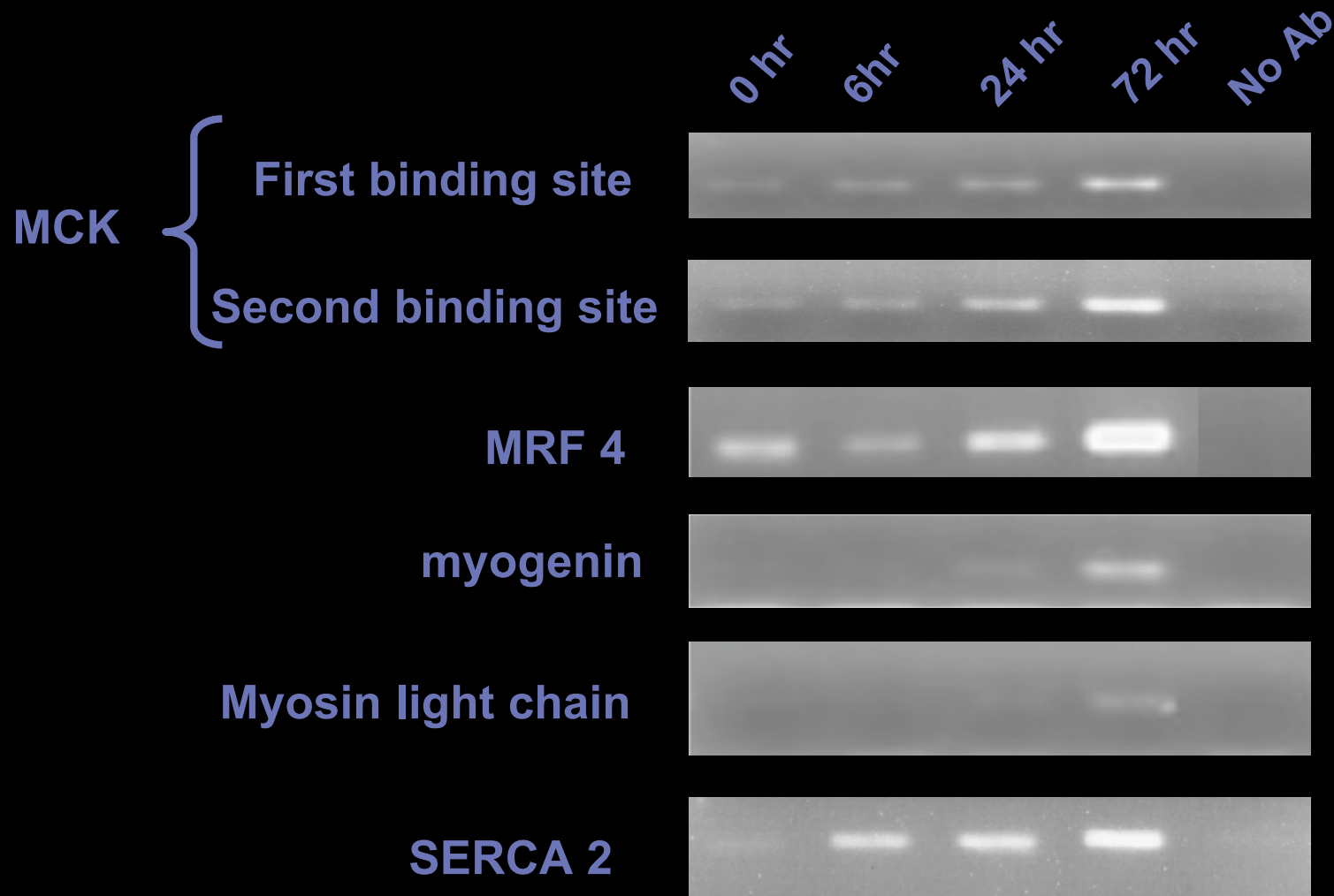
PCR amplify



~ 100-200 bp



ChIP Analysis of MEF-2 Associated Genes

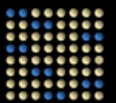


How can we apply this to a high throughput screening method?

Need a method of capturing the regulatory sites of genes

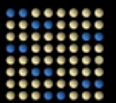
- CpG Island Microarrays

- CpG islands are unmethylated GC rich regions of the genome that are associated with the 5' ends of most house-keeping genes and many regulated genes
- About 80% of CpG islands are common between human and mouse
- About 56% of human genes and 47% of mouse genes are associated with CpG islands



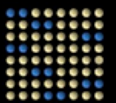
The Sanger 12k Human CpG Set

- 12,288 redundant clones (128 96-well plates)
- Originally there were more clones but due to phage contamination, those plates are no longer available
- Plate 105 also had some phage contamination so it has been excluded from our set
- The Sanger Institute performed 3' and 5' terminal sequencing of each clone
 - Information available at <http://www.sanger.ac.uk/HGP/cgi.shtml>



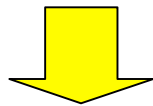
Technical Issues

- Due to the relatively low quantities of DNA recovered from ChIP, we require a method of amplification to be able to visualize the immunoprecipitated DNA
 - Employed a modified method from the Pat Brown lab (www.microarrays.org). Our modifications are available on our website at (www.microarrays.ca)
- Annotation of the CpG clones to identify location in the genome and potential genes in proximity with the CpG island.
 - In conjunction with the NCI, we are having all 12,000 clones resequenced.
 - Once data is available, we will use BLAT to align all sequences to the genome and pull out putative regulatory targets.

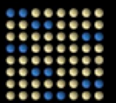
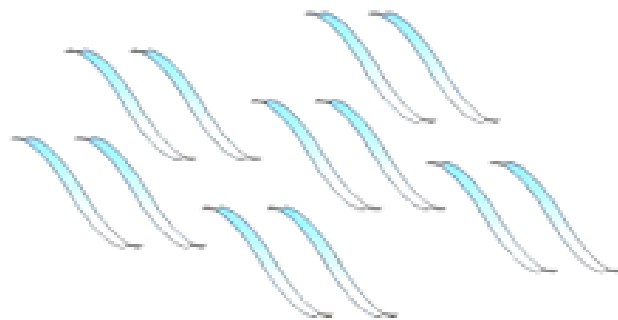




Degenerate Primers



PCR Amplify ChIp'ed DNA



The UCSC Genome Browser

<http://genome.ucsc.edu>

UCSC Genome Browser on Human July 2003 Freeze

move zoom in dense zoom out

position size 10,974 image width

Base Position: 26888967 (Chromosome Bands Localized by FISH Mapping Clones)

Chromosome Band

GeneScan Genes

Human genes from Genes

CpG Islands

move start Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options

move end

Guidelines Labels: left center

Note: Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing Tracks

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb. Rate
<input type="button" value="on"/>	<input type="button" value="dense"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>
Forward End Pairs	GC Percent			
<input type="button" value="hide"/>	<input type="button" value="hide"/>			

Genes and Gene Prediction Tracks

RefSeq Genes	MGC Genes	GenScan Genes
<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="dense"/>

mRNA and EST Tracks

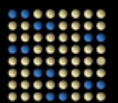
Human mRNAs	Soliced ESTs	Human ESTs	NonHuman mRNAs	NonHuman ESTs
<input type="button" value="pack"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>	<input type="button" value="hide"/>

Expression and Regulation

CpG Islands
<input type="button" value="dense"/>

Variation and Repeats

RepeatMasker	Simple Repeats
<input type="button" value="hide"/>	<input type="button" value="hide"/>



CpG Island Info from UCSC

CpG Island Info

CpG Island Info

Chromosome: 7

Band: 7p15.2

Begin in Chromosome: 26877616

End in Chromosome: 26879008

Genomic Size: 1393

[View DNA for this feature](#)

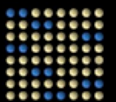
Size: 1393

CpG count: 116

C count plus G count: 911

Percentage CpG: 16.7%

Percentage C or G: 65.4%



Neighbouring Gene Info

AK022839

Information on mRNA [AK022839](#)

Description: Homo sapiens cDNA FLJ12777 fa, clone NT28P2001720

Gene: n/a

Product: n/a

Author: [Iwata T, Ota T, Hirotsu E, Saitama T, Oishi T, Saito Y, Hishikawa T, Hama K, Tsunoda S, Aoyama E, Yoshizawa Y, Matsumura H, Iida S, Kawa Y, Sato K, Yamamoto J, Wakayama A, Nakamura Y, Nishikubo K, Masuda Y and Suzuki H](#)

Organism: [Homo sapiens](#)

Tissue: n/a

Development stage: n/a

Cell type: INT2

Sex: n/a

Library: NT28P2

Clone: NT28P2001720

CDS: n/a

Date: 2002-07-01

GeneLink: [AK022839](#)

Standard SOURCE: [AK022839](#)

mRNA/Genomic Alignments

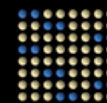
SIZE	IDENTITY	CDS/GENOME	STRAND	START	END	QUERY	START	END	TOTAL
2342	98.9%		+	18678220	18680948	AK022839	1	2342	2342

Description

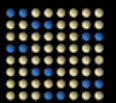
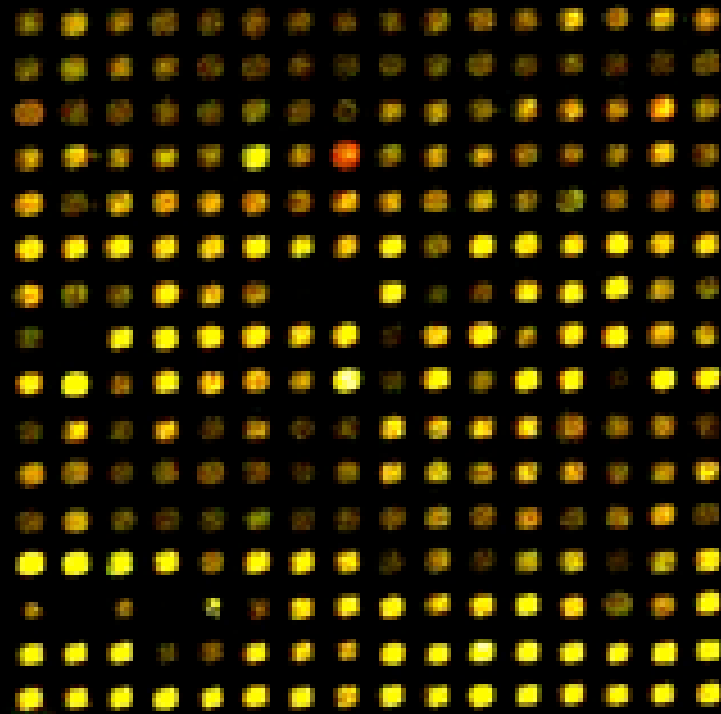
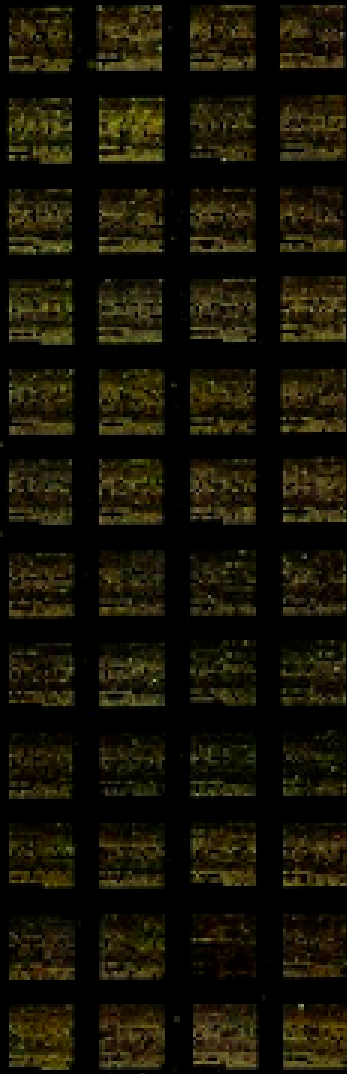
The Human mRNA track shows alignments between human mRNAs in Genbank and the genome. Aligning regions (usually exons) are shown as black boxes connected by lines for gaps (spliced out introns usually). In 3D display, arrows on the introns indicate the direction of transcription.

Method

Genbank human mRNAs are aligned against the genome using the [tblastx](#) program. When a single mRNA aligns in multiple places, the alignment having the highest base identity is found. Only alignments that have a base identity level within 1% of the best are kept. Alignments must also have at least 20% base identity to be kept.

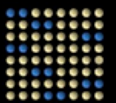


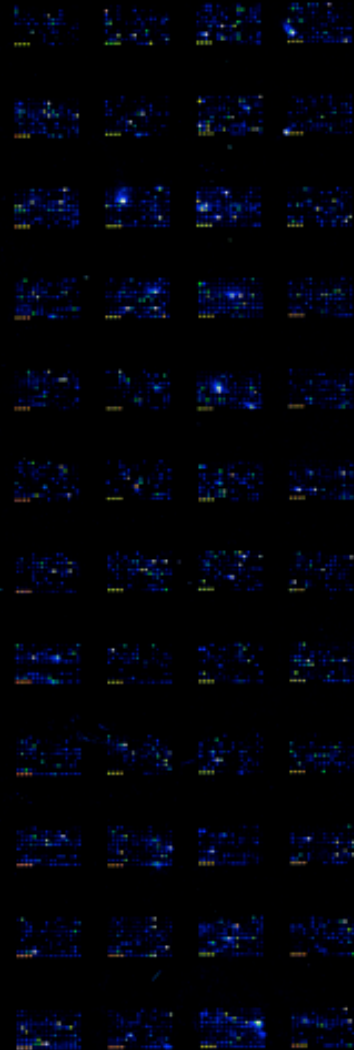
12k Mk1 CpG Array



Mouse 7k CpG Island Array

- Mouse CGI library obtained from Sanger Institute (UK)
 - CpG island clones in pGEM-5Zf vector
- Used a 100 μ l aliquot (10^7 cells)
- Plated library - approx. 300 colonies/plate
- Used colony picker (Genetix Q-Pix2) to pull clones
- Total of 24 plates
- 7000 viable colonies picked
- Amplified inserts using T7/SP6 primers
- Purified, transferred to 384 well plates and prepared for arraying

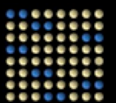
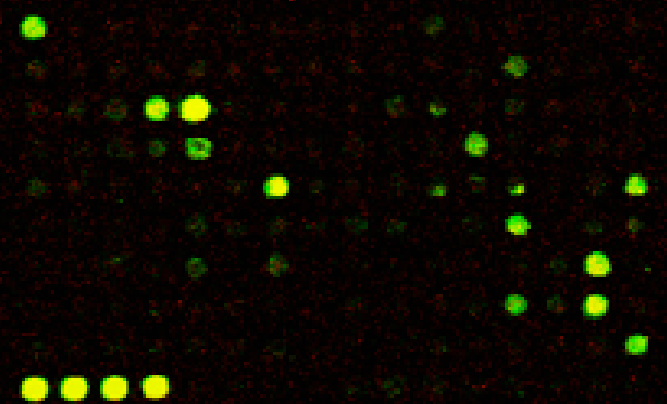




Sample mouse CGI array v.1

Total number of spots: 7680

Arabidopsis controls included
in each subarray for
normalization purposes



Methodology (MEF-2 Example)

C_2C_{12} cells were grown to 80% confluence



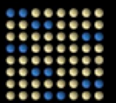
Media was changed to differentiation media
(1% horse serum and 50 ng/ml IGF-1)



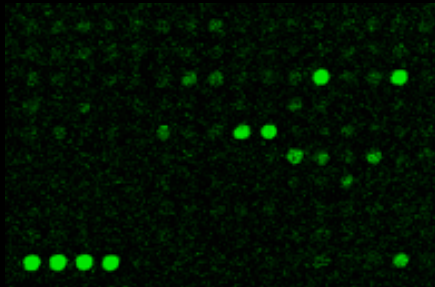
Following 0, 6, 24 and 72 hours, cells were treated and
harvested



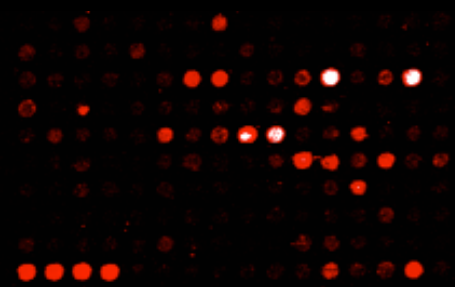
Followed ChIP on CpG microarray protocol



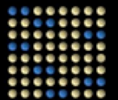
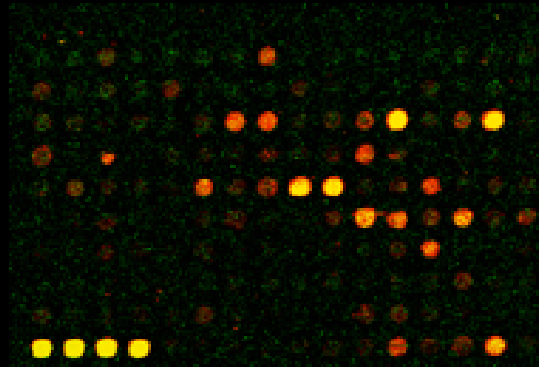
No Antibody



Antibody

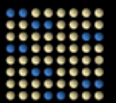
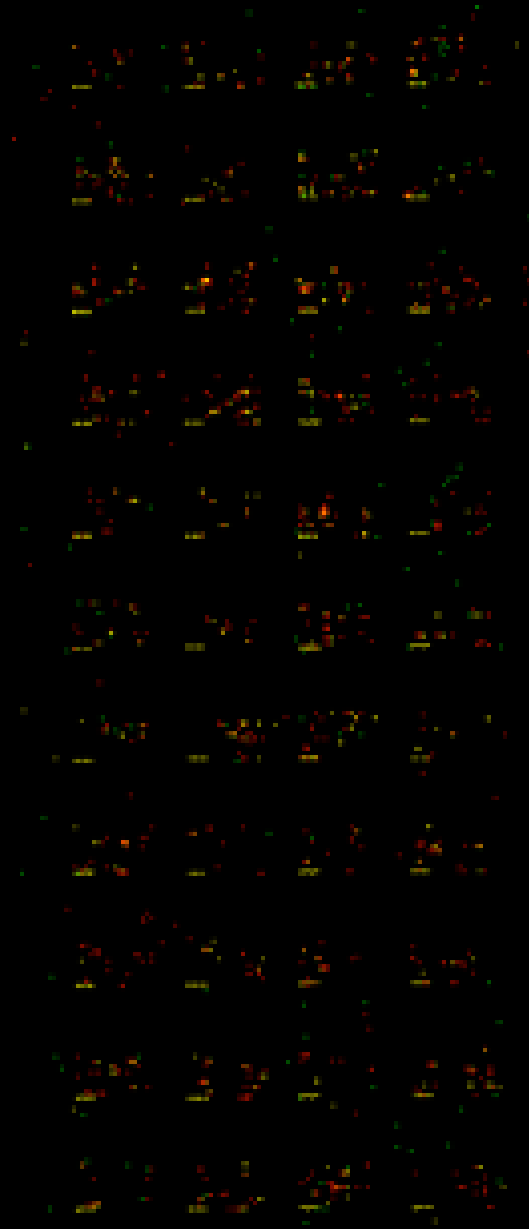
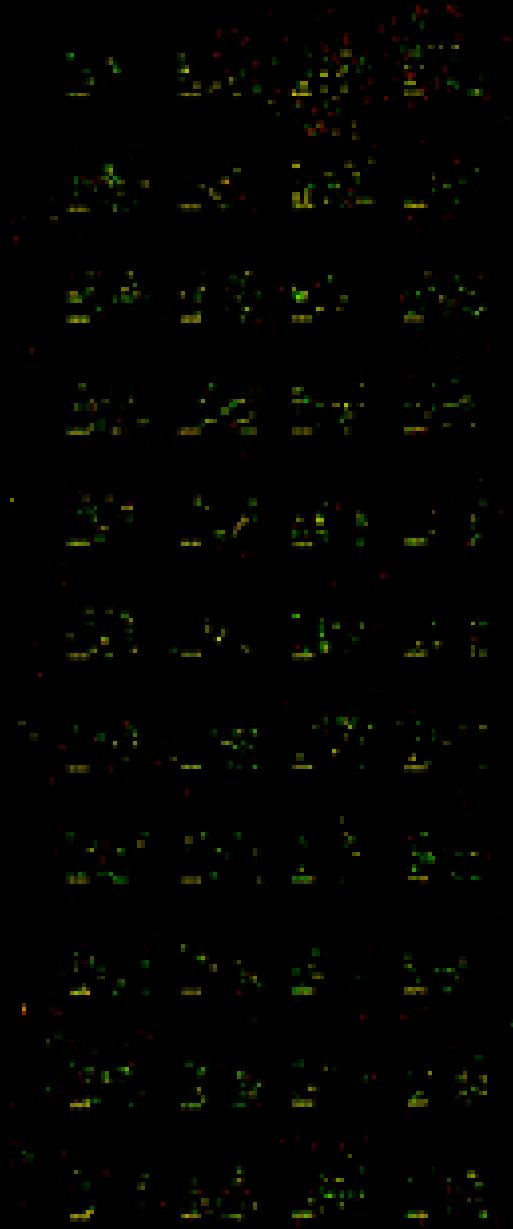


Composite



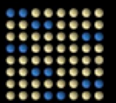
0 hours

1 day



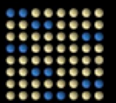
Data Analysis

- All arrays scanned, quantified and entered into GeneTraffic™
- Spot intensities normalised by *Arabidopsis* controls
- Filtered all spots with experimental channel intensity less than 512
- Selected spots with > 2 -fold ratio of +Ab/-AB



Results

- Total number of positive clones:
 - 0 hours: 0
 - 6 hours: 0
 - 24 hours: 260
 - 72 hours: 0
- 20 clones randomly chosen and sequenced (both directions)
- Sequence data queried against genomic sequence



Informatics Pipeline

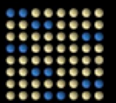
Hits from microarrays → 260 clones

Collect sequences → 20 random

Remove redundancy → 12 remaining

Scan for sequences with repetitive elements → 6 remaining

Submit sequences to BLAT (UCSC)



Find only those with exact matches (no gaps)



Look for any genes (predicted or known) in region



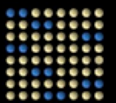
Retrieve large fragment of genomic sequence encompassing the potential promoter, gene and submitted sequence



Use MSCAN to search for potential Mef-2 binding consensus sequence




Extract regulatory region





Use BLAT to locate this region on whole genome



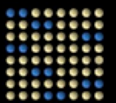
Ensure that this sequence is within a reasonable distance of the gene of interest



Design primers to this region



Run ChIP with these specific primers and confirm microarray result



Sample ID	0 hour FOR	0 hour REV	6 hour FOR	6 hour REV	1 day FOR	1 day REV	3 day FOR	3 day REV
16K01	0.256413942	0.128071129	0.713034687	0.214087724	1.791018318	1.562965508	0.592307149	0.106925116
07H16	0.70023574	-0.103017219	0.9733921	0.194145439	1.70091124	1.26233237	0.795308229	-0.123129723
15G10	0.146715123	0.332615023	0.488551273	0.561857374	1.681257867	1.708030011	0.861299916	0.391230433
03J13	0.473024447	0.381420254	0.737245956	0.567375425	1.676496466	1.519621488	0.52732692	0.383569913
12L24	0.481859942	0.127395987	0.769809207	0.413118907	1.652590398	1.586356043	0.50673634	0.129924325
04B24	0.090128174	0.171141685	0.506018996	0.752638297	1.63963705	1.682808467	0.712043544	0.289281761
02P11	0.37548072	0.410583237	0.388537605	0.577599645	1.63397755	1.691026244	0.445840564	0.410075151
11G22	0.265889717	0.32211013	0.430616757	0.203614646	1.627434286	1.159876808	0.915625876	0.218526916
18G22	0.128237578	0.155283572	0.657384613	0.535431612	1.619981684	1.457295301	0.813558108	0.152732591
14J11	0.302189709	0.326958358	0.449631288	0.622304728	1.617297851	1.787123433	0.343860494	0.430822514
11N01	0.465143407	0.166056566	0.77321046	0.128051847	1.617117254	1.234939723	0.747396268	0.010551712
19K23	0.29633545	0.276918323	0.455015826	0.573781672	1.602261659	1.427559251	0.570072622	0.245716727
13J04	0.75777217	-0.159757551	0.745569792	0.230288466	1.600392978	1.231966923	0.540334694	-0.04314942
05H07	0.146620715	0.434712615	0.697660989	0.36874243	1.563675027	1.593980287	0.755247063	0.138862363
10P15	0.550307985	0.033648846	0.808097901	0.148087902	1.549880514	1.280703094	0.515766324	-0.027965905
19O11	0.139866218	0.528095054	0.117722127	0.654554767	1.532797106	1.390544747	0.797551384	0.114852505
01B02	0.56287742	0.231678632	0.414053947	0.390077023	1.525624877	1.275517159	0.55601813	0.174848198
17F02	0.122234069	0.057376578	0.457740894	0.422462962	1.525354938	1.540102053	0.579401247	0.134821937
11F11	0.347238717	0.011828431	0.588984471	0.216530856	1.524419597	1.485622226	0.84312707	0.130073256

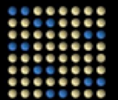


0 6hr 1 3 day



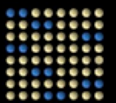
Courtesy Mark Takahashi, CGC

UHN Microarray Centre



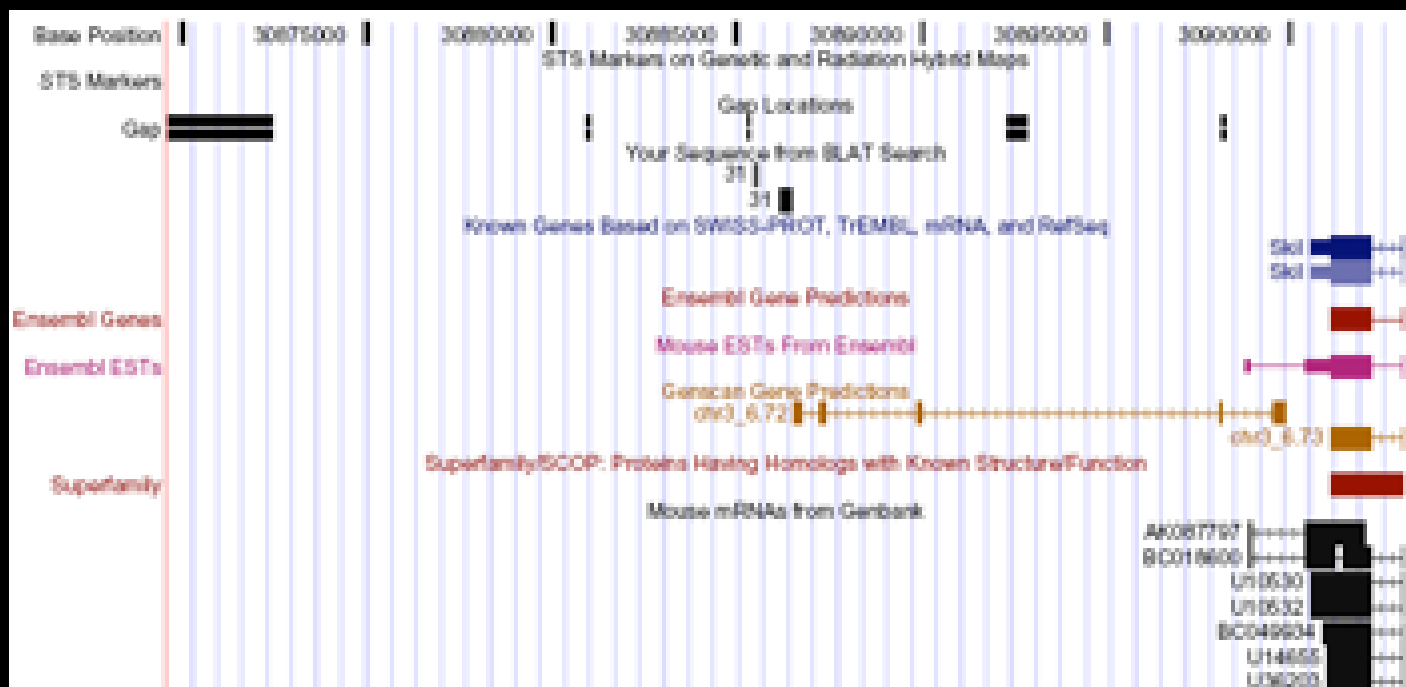
Search Results | Current Page: 1 | Total Pages: 1 | Total Results: 8

	Hyb. Group	Hybridization	Gene ID	Sample ID	LEX.E - BG	LEX.R Norm.	Fold Change	Flag	Image
1	0 hours	0 hour FOR	10P15	10P15	39645	27072	1.46		
2	0 hours	0 hour REV	10P15	10P15	16088	15717	1.02		
3	1 day	1 day FOR	10P15	10P15	35601	12159	2.93		
4	1 day	1 day REV	10P15	10P15	28524	11740	2.43		
5	3 day	3 day FOR	10P15	10P15	38542	26957	1.43		
6	3 day	3 day REV	10P15	10P15	24751	25235	-1.02		
7	6 hours	6 hour FOR	10P15	10P15	18083	10328	1.75		
8	6 hours	6 hour REV	10P15	10P15	21749	19627	1.11		



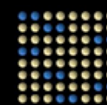
Courtesy Mark Takahashi, CGC

UHN Microarray Centre

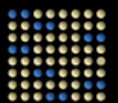


Courtesy Mark Takahashi, CGC

UHN Microarray Centre



SOURCE GeneReport <i>M. musculus</i>		Ski
Sloan-Kettering viral oncogene homolog UniGene , LocusLink		
Alarms		
<ul style="list-style-type: none"> 2710012028a, 2610001A118a, BC004032, MOC200, MOC 2300 		
Chromosomal Location		
Chromosome/Cytoband	4 73.9 cM	
Microarray Gene Expression Data		
Data available	Show Gene Expression Data	
SwissProt Information		
SwissProt Accession No.	Q68698 Ski oncogene (<i>Mus musculus</i>), 99% similarity over 325 a.a.	
Function	may play a role in terminal differentiation of skeletal muscle cells but not in the determination of cells to the myogenic lineage	
Developmental Stage	is expressed in a uniform pattern in all embryonic cells prior to skeletal muscle cell formation in the myotomes of somites. expression is first upregulated in skeletal muscle at 12 dpc, this upregulation is evident first in body wall muscle and one day later in limb muscles. at 13.5 dpc a most prominent expression is seen in all skeletal muscles. at this stage expression is seen in all other cells and tissues but at lower levels than in skeletal muscle.	
Subcellular Location	nucleus	
Similarity	to csk protein of sloan-kettering virus and to src oncogenes	
SwissProt Copyright	<p>The SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL, consortium - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions so long as its content is not reproduced and the source is acknowledged. Usage by and for commercial entities requires a license agreement. (See http://www.ebi.ac.uk/Ensembl/entry.html#copyright)</p>	

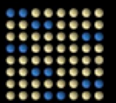


1Day non-amplified ChIP analysis

Reconfirm the results from the amplified data.

Plated cells onto thirty 150 mm dishes and grew to 80% confluence.

Following 1 day of differentiation cells were fixed and harvested for ChIP analysis.



Amplified DNA

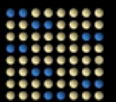
Search Results | Current Page: 1 | Total Pages: 1 | Total Results: 8

	Hyb. Group	Hybridization	Gene ID	Sample ID	LEX.E - BG	LEX.R Norm.	Fold Change	Flag	Image
1	0 hours	0 hour FOR	10P15	10P15	39645	27072	1.46		
2	0 hours	0 hour REV	10P15	10P15	16088	15717	1.02		
3	1 day	1 day FOR	10P15	10P15	35601	12159	2.93		
4	1 day	1 day REV	10P15	10P15	28524	11740	2.43		
5	3 day	3 day FOR	10P15	10P15	38542	26957	1.43		
6	3 day	3 day REV	10P15	10P15	24751	25235	-1.02		
7	6 hours	6 hour FOR	10P15	10P15	18083	10328	1.75		
8	6 hours	6 hour REV	10P15	10P15	21749	19627	1.11		

Non-amplified DNA

Search Results | Current Page: 1 | Total Pages: 1 | Total Results: 1

	Hyb. Group	Hybridization	Gene ID	Sample ID	LEX.E - BG	LEX.R Norm.	Fold Change	Flag	Image
1	hyb1	Hyb1	10P15	10P15	18668	6490	2.88		

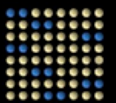


Courtesy Mark Takahashi, CGC

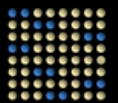
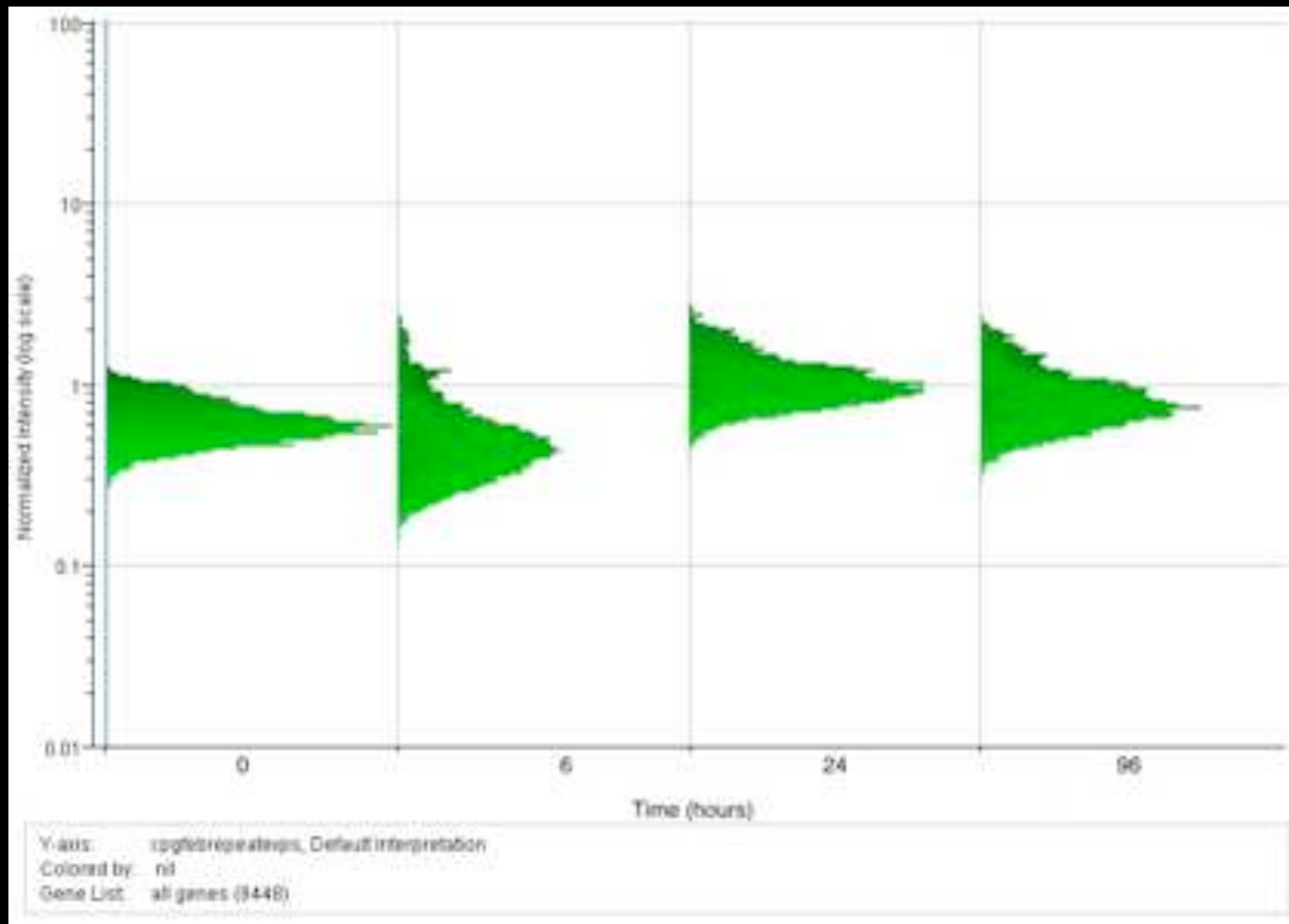
UHN Microarray Centre

Let's Try that Again

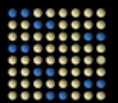
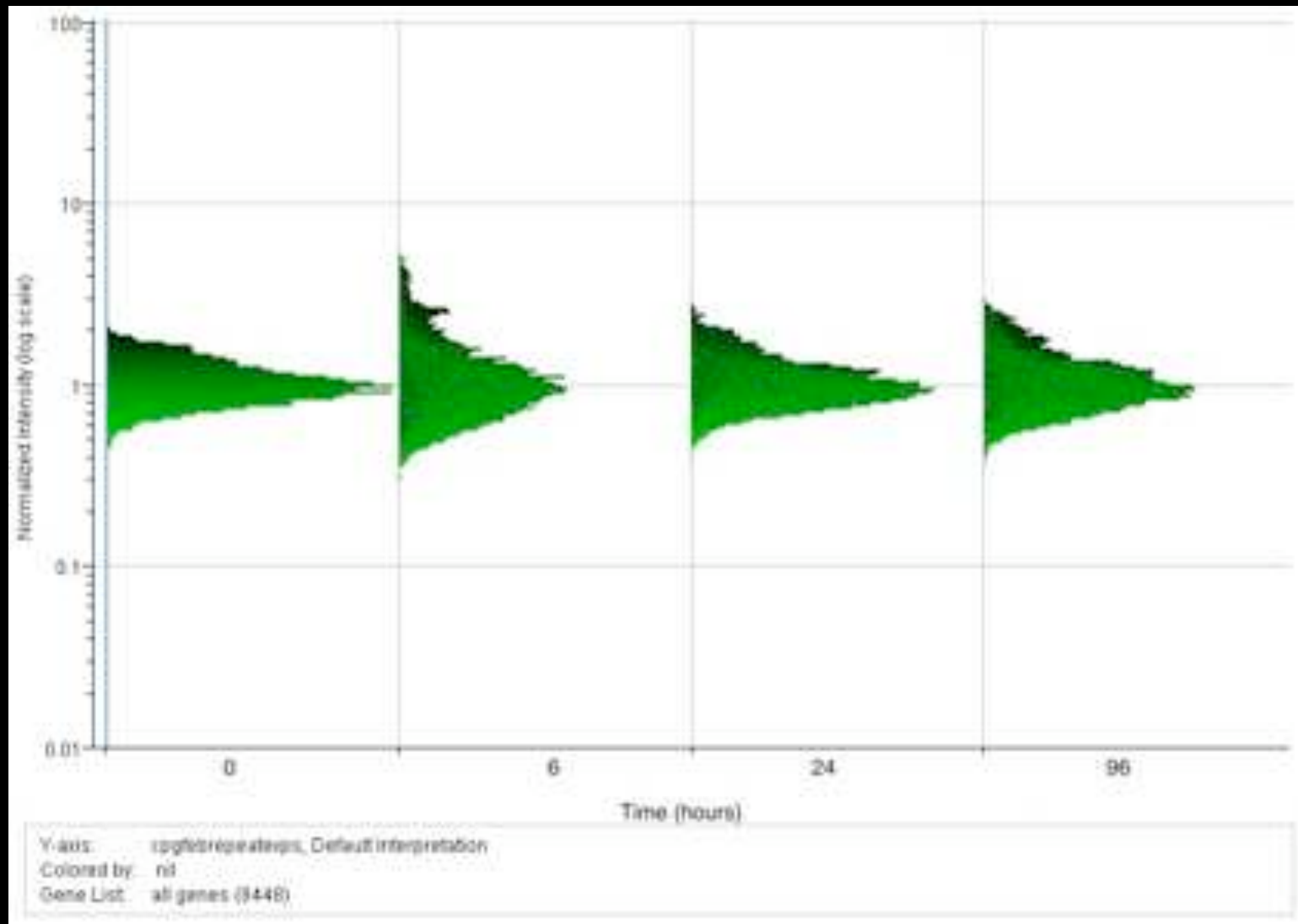
- Repeat experiment, and alter experimental design
- More samples
- Reduce variability at seeding stage
- Reduce variability at labelling, storage stages
- Enhance bioinformatics approaches



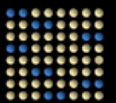
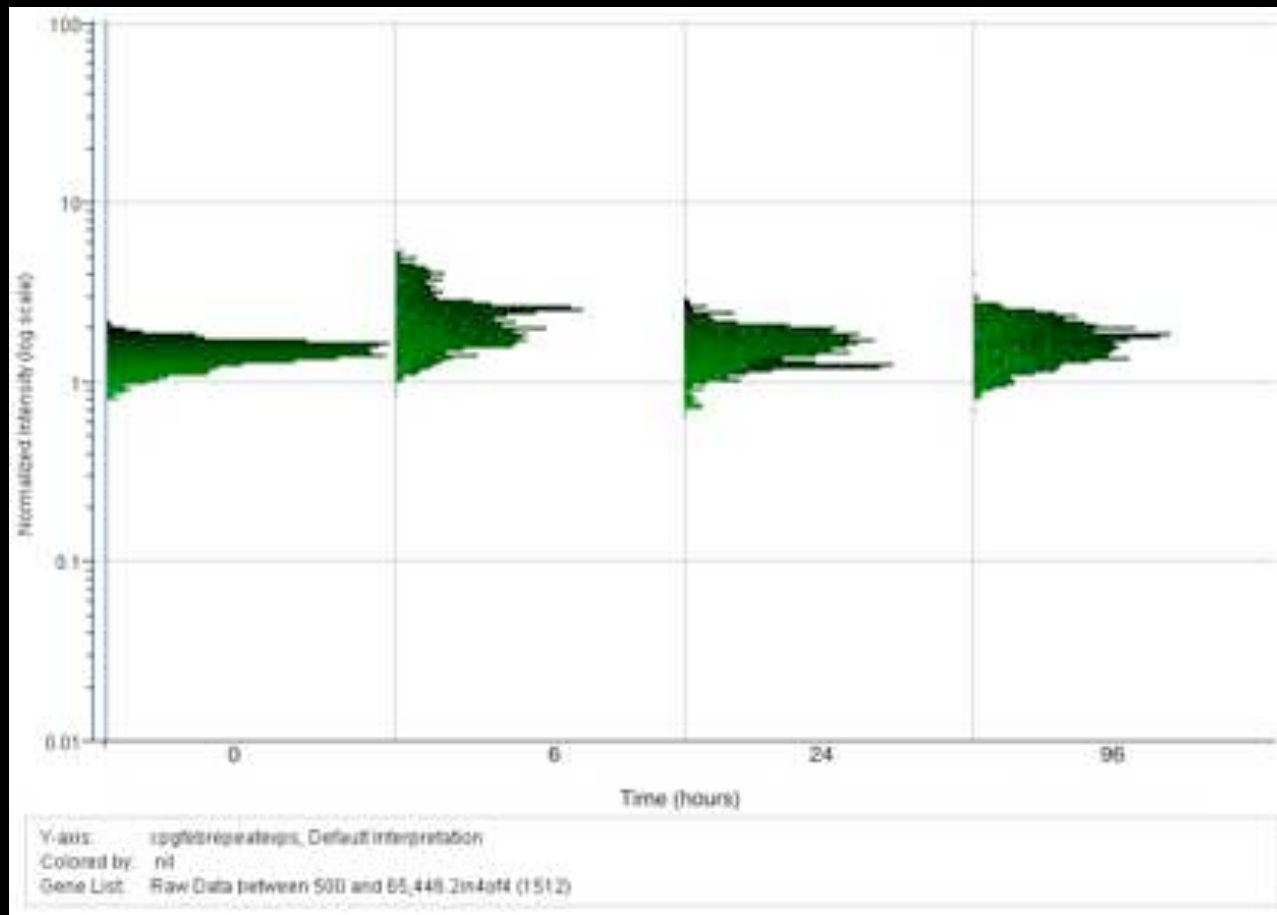
Pre-normalised Data



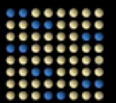
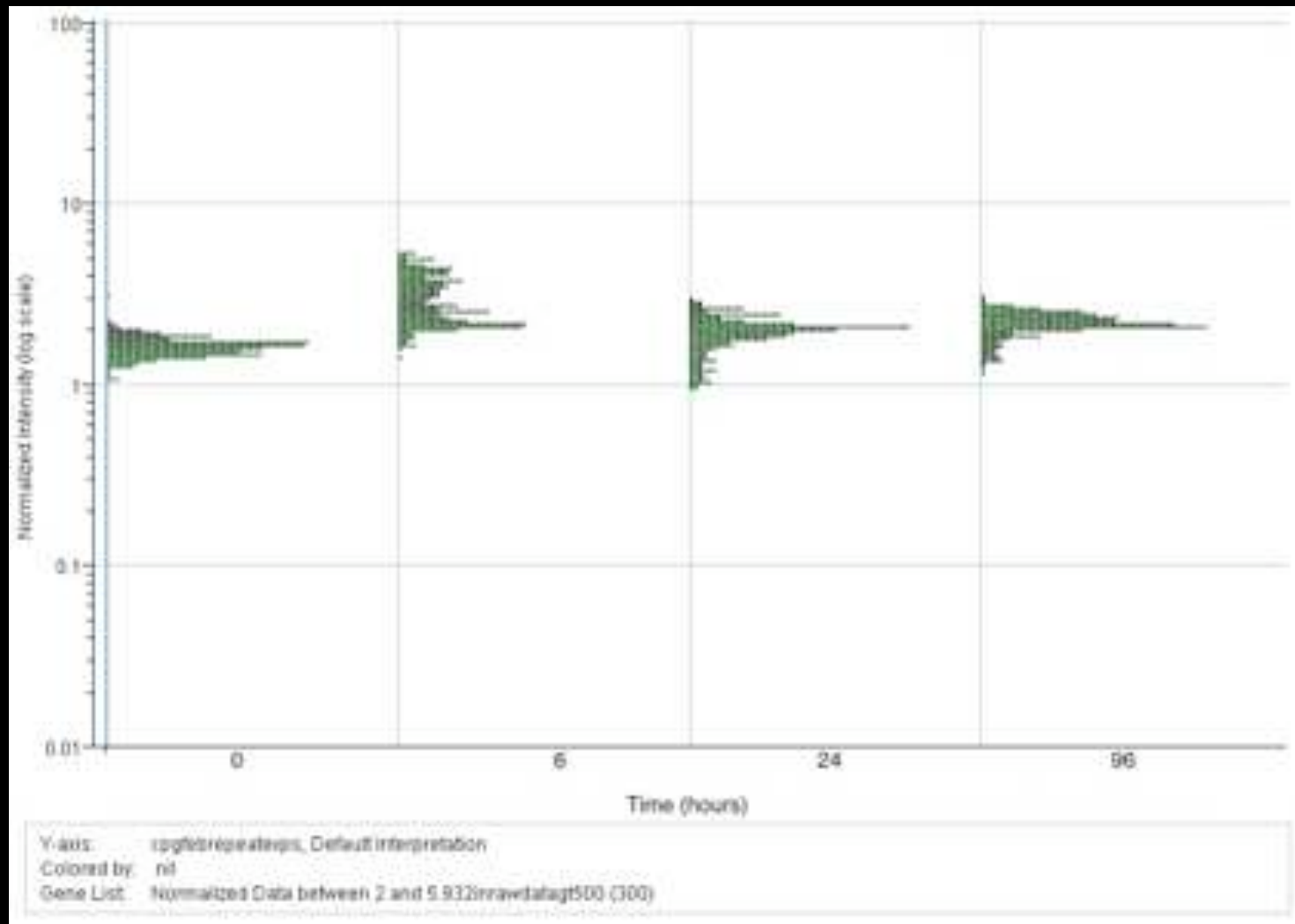
Normalised (Median Intensity)



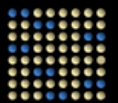
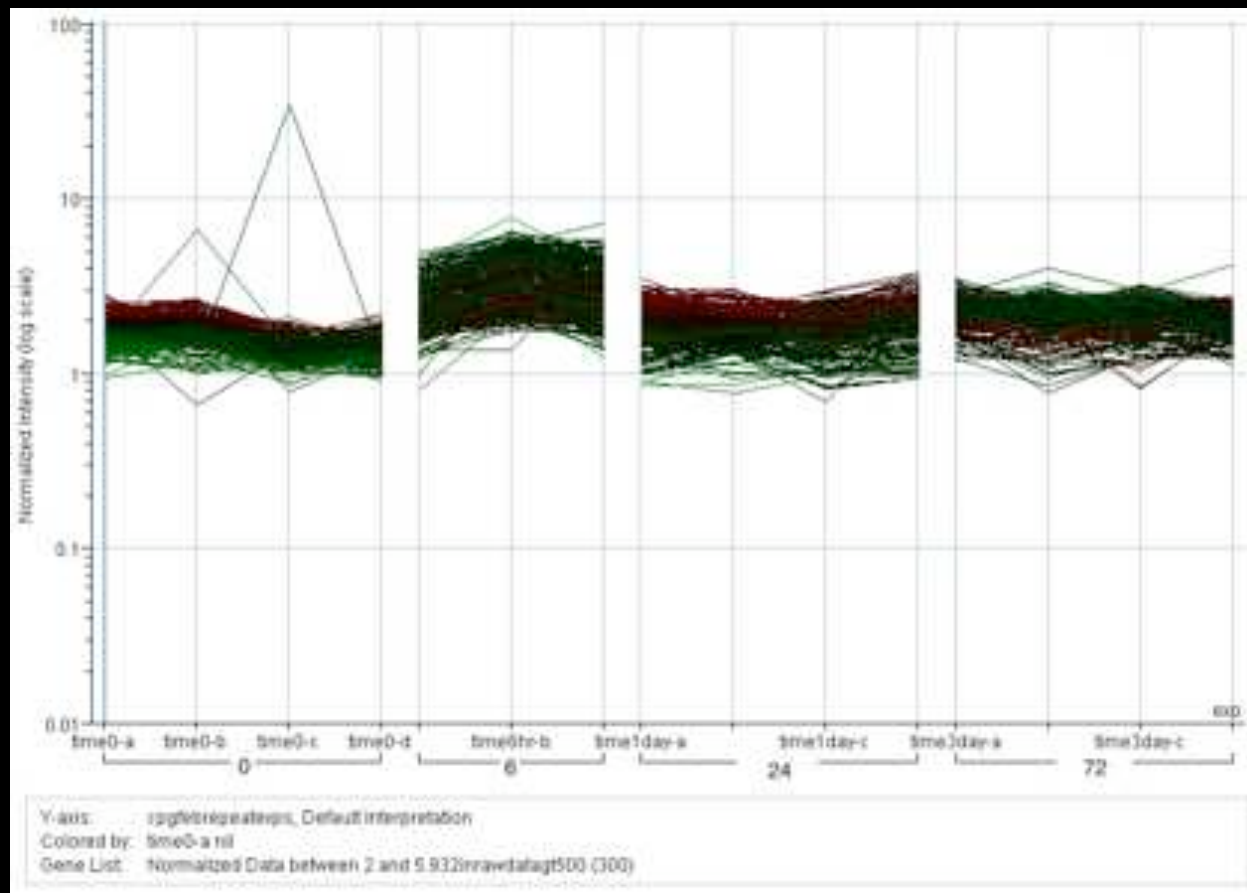
Remove Low Intensity Clones



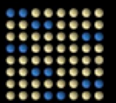
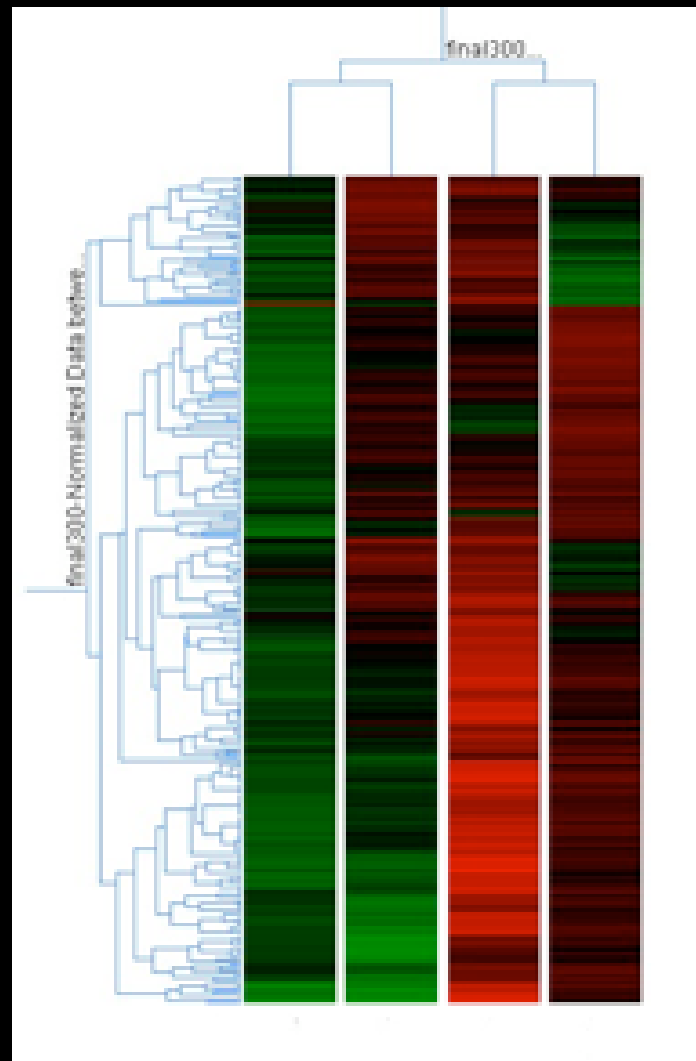
Include only >2 fold ratio



Consistency of Replicates



Clustering Positive Clones



Selecting Positive Clones

System Help Search Results | Current Page: 1 | Total Pages: 2 | Total Results: 32 |

Hyb. Group	Gene ID	Num. Valid Spots	Mean Log ₂ Ratio	STDEV	COV	Norm. STDEV
------------	---------	------------------	-----------------------------	-------	-----	-------------

Gene Table Options - Microsoft Internet Explorer

Information required for data identification:

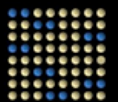
Name: Regulated Genes

Description: Changers, low variance, high expression change

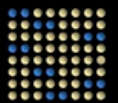
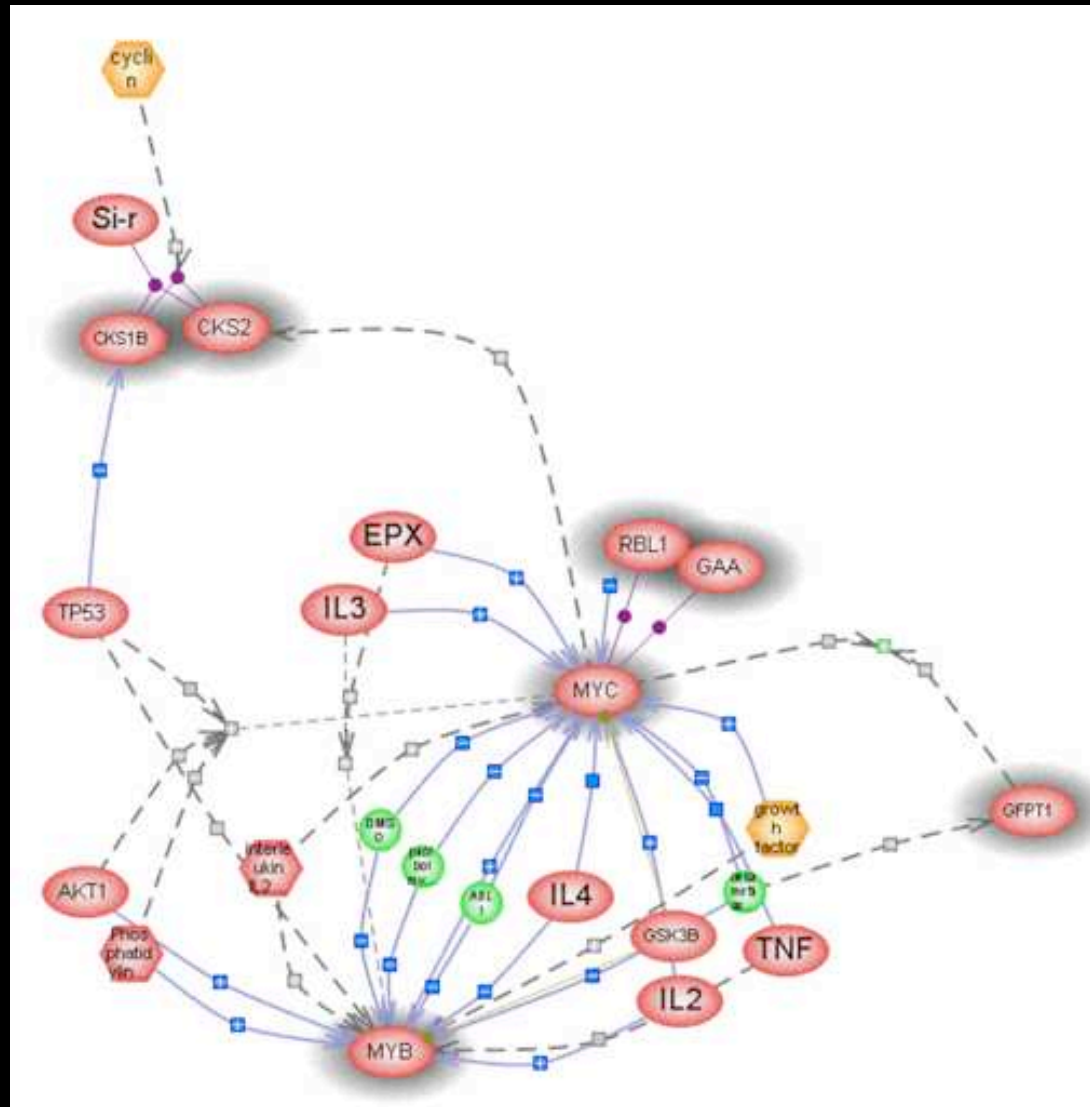
Gene Table
A Gene Table stores one row of values for each gene in the set and one column for each hybridization group. Normalized log-space expression ratio for a particular gene in a particular hybridization group is shown at the intersection of the corresponding row and column. In addition to unique gene and hybridization identifiers, the table may also contain columns and rows storing gene description and annotation.

OK Cancel

Gene Se. Par
Mean Lo
2
AND
Hyb. Gr
equals
Sorting C
Hyb. Group
Ascending
Items per Page
20
STDEV
COV
Norm. STDEV

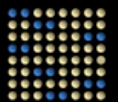
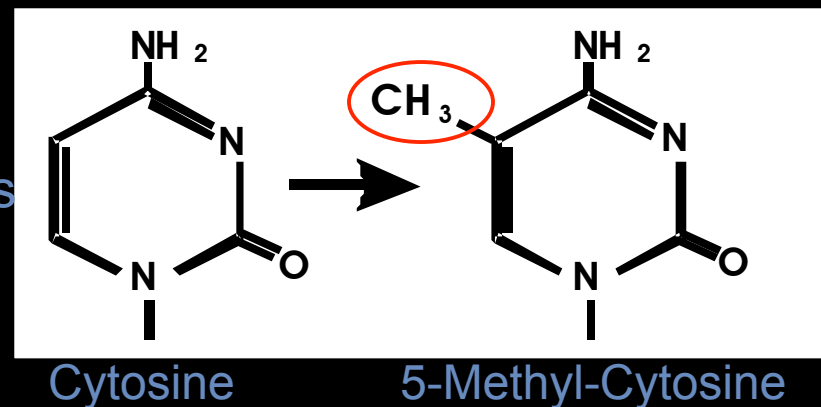


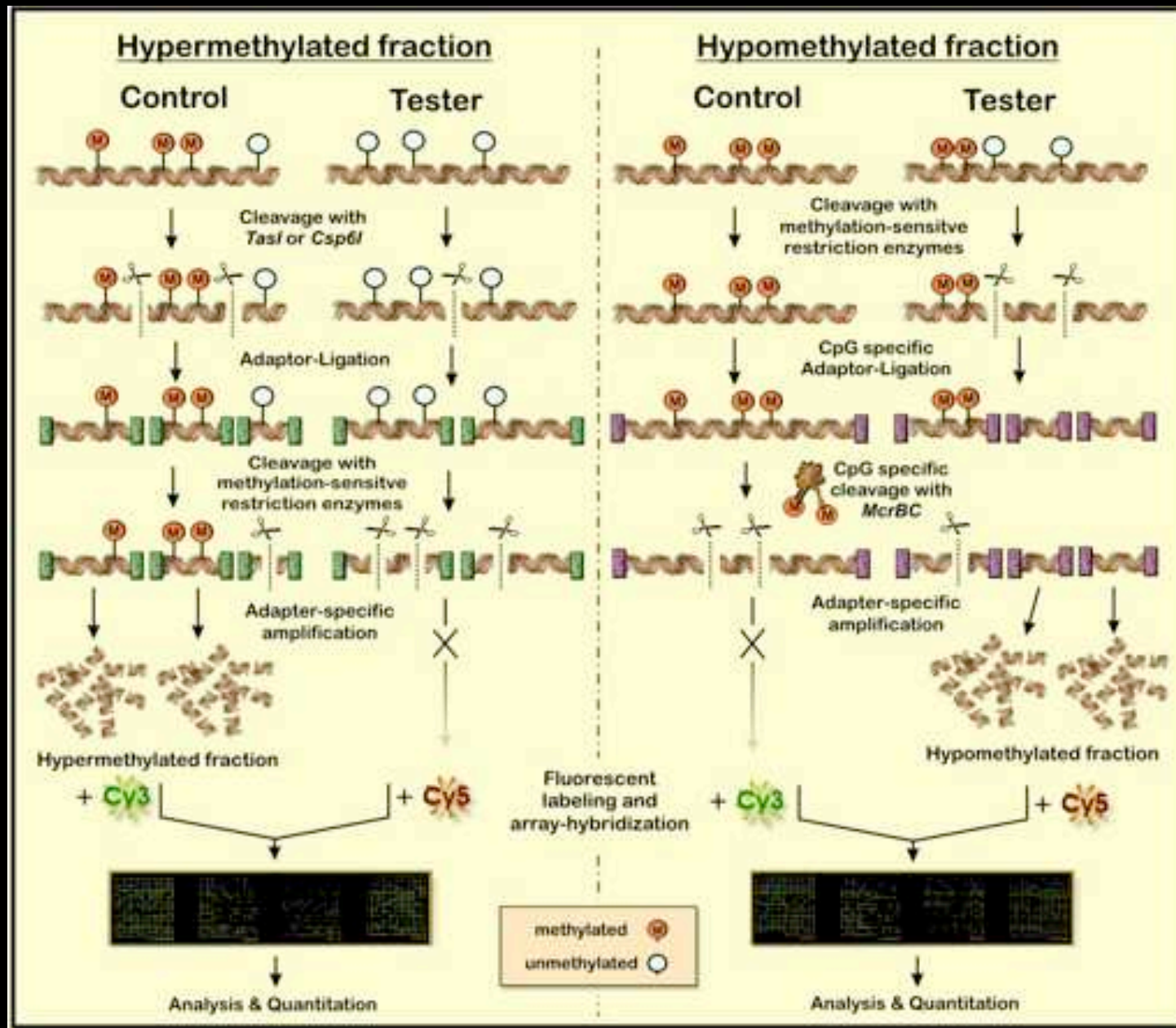
Pathway Analysis (with PathwayAssist™)



Epigenetics

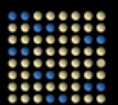
- **Epigenomics** is the genomwide study of methylation and other epigenetic phenomena - that is, the study of the epigenome.
- The **epigenome** is the collection of biochemical modifications to chromatin that indexes genetic information.
- This collection of modifications includes DNA and protein modifications like:
 - Histone-Acetylation
 - Histone-Methylation
 - Matrix attachment sites
 - **DNA-Methylation**



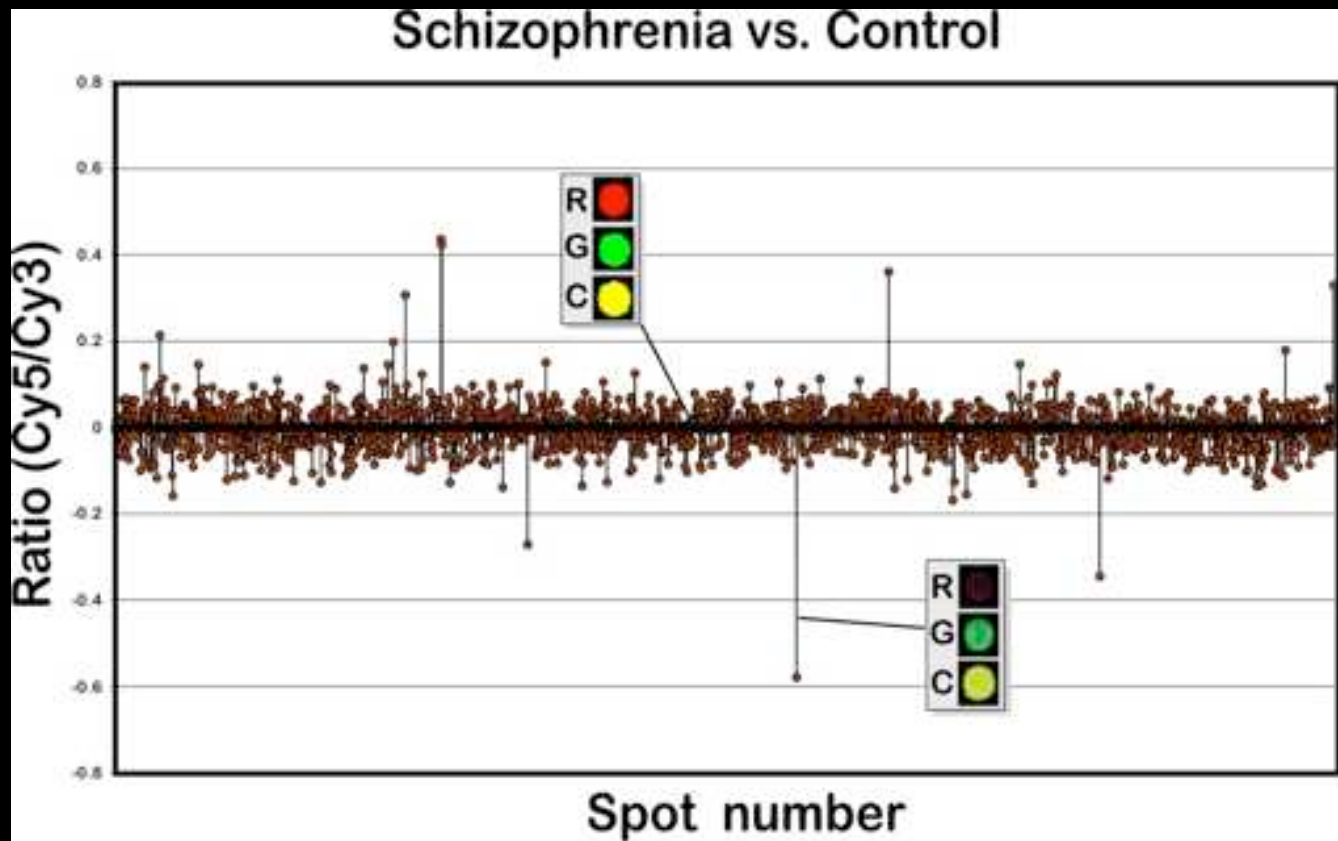


Courtesy Axel Schumacher & Art Petronis, CAMH

UHN Microarray Centre

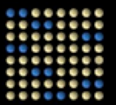


1.7k Human cDNA Microarray

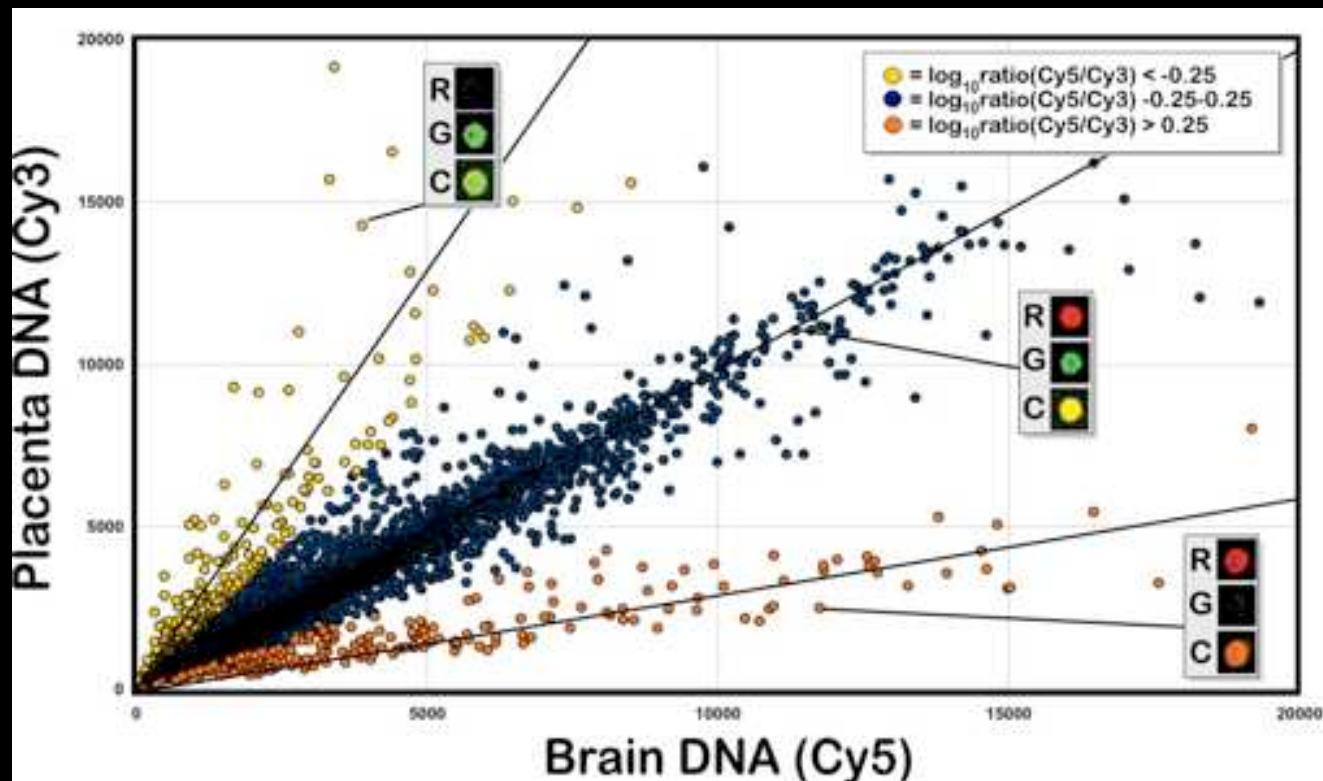


Courtesy Axel Schumacher & Art Petronis, CAMH

UHN Microarray Centre

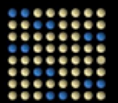


12k CpG Island Microarray



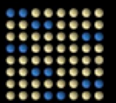
Courtesy Axel Schumacher & Art Petronis, CAMH

UHN Microarray Centre



Plans For the Future

- Sequence entire human and mouse clonesets
 - Currently being done by UBC sequencing facility
- Clean up clones
 - Eliminate contaminated wells if any, chimeric clones etc...
- Eliminate redundancy
 - Bioinformatics
- Develop clones for additional promoter sites
 - Look at Mouse - Human regions of similarity
- Correlate CpG clones to ESTs on expression arrays
 - Bioinformatics



Acknowledgements

Production

- Patrick Yau (Production Manager)
- Eric Ho
- Tuyet Diep
- Christina Johnston
- Robert Kardish
- Shani Mintzberg
- Quyen Tran
- Julissa Tsao
- Stephanie Selders
- Leslie Wyard

Epigenetics (CAMH)

- Axel Schumacher
- Art Petronis

Research and Development

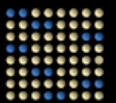
- Mark Takahashi (Associate Scientist)
- Carolyn Modi (Project Manager)
- James Paris

CpG Advisory Board

- Rod Bremner
- Sandy Der
- Linda Penn
- Jim Woodgett

Bioinformatics

- Larry Heisler
- Carl Virtanen
- Zhibin Lu



Thank-You

Please visit us at:

www.microarrays.ca



The screenshot shows the homepage of the Microarray Centre website. The top navigation bar includes links for 'about us', 'products & services', 'support', 'commercial', 'resource links', and 'contact'. The Microarray Centre logo and 'University Health Network' branding are in the top right. A main banner features the text 'A LEADER IN MICROARRAY TECHNOLOGY.' over a background image of laboratory equipment. Below the banner are three columns: 'NEWS' with two articles, 'MICROARRAY CENTRE' with a descriptive paragraph and a 'Read more' link, and 'NEWSLETTER' with a 'SUBSCRIBE TO OUR NEWSLETTER' button.

NEWS

21/07/03
3rd Annual Ontario Microarray Network Symposium Rescheduled.
Our 3rd annual symposium has been rescheduled for November 11 to 13th. The previous event was postponed due to SARS. For more information, [click here](#).

21/07/03
Human CpG "promoter" arrays now available from the Microarray Centre
The UHN Microarray Centre is now distributing 12k Human CpG arrays made from the Sanger Institute's CpG clone set. For more information, [click here](#).

MICROARRAY CENTRE

 The Microarray Centre at The Ontario Cancer Institute, University Health Network is a leader in Canadian microarray technology. We are dedicated to providing high quality microarrays, technical support and service to Canadian researchers. Access to high quality microarrays will allow our Canadian researchers to be on the cutting edge of genetic research.

[Read more ▶](#)

NEWSLETTER

On occasions we may need to distribute new and important information to our users. If you use our microarrays, please sign up to our newsletter, it is the most effective and efficient way to receive information from us.

[SUBSCRIBE TO OUR NEWSLETTER](#)