ChIPping Away at Muscle Differentiation

Mark Takahashi
November 12, 2003

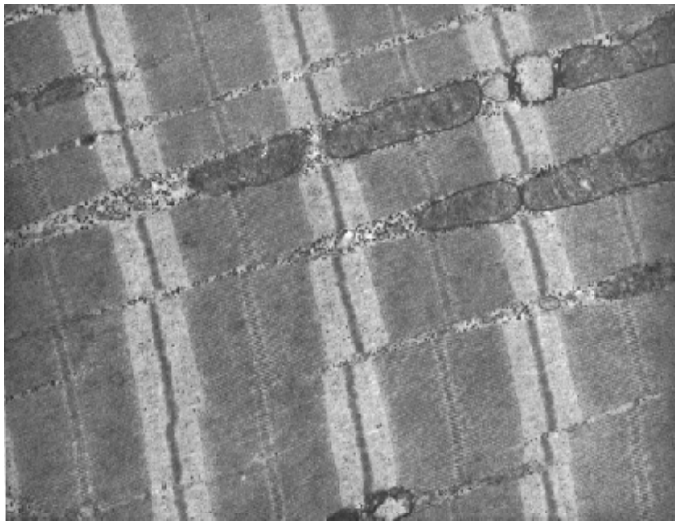
UHN Microarray Centre

Approximately 50-55% of the body is made up of muscle.

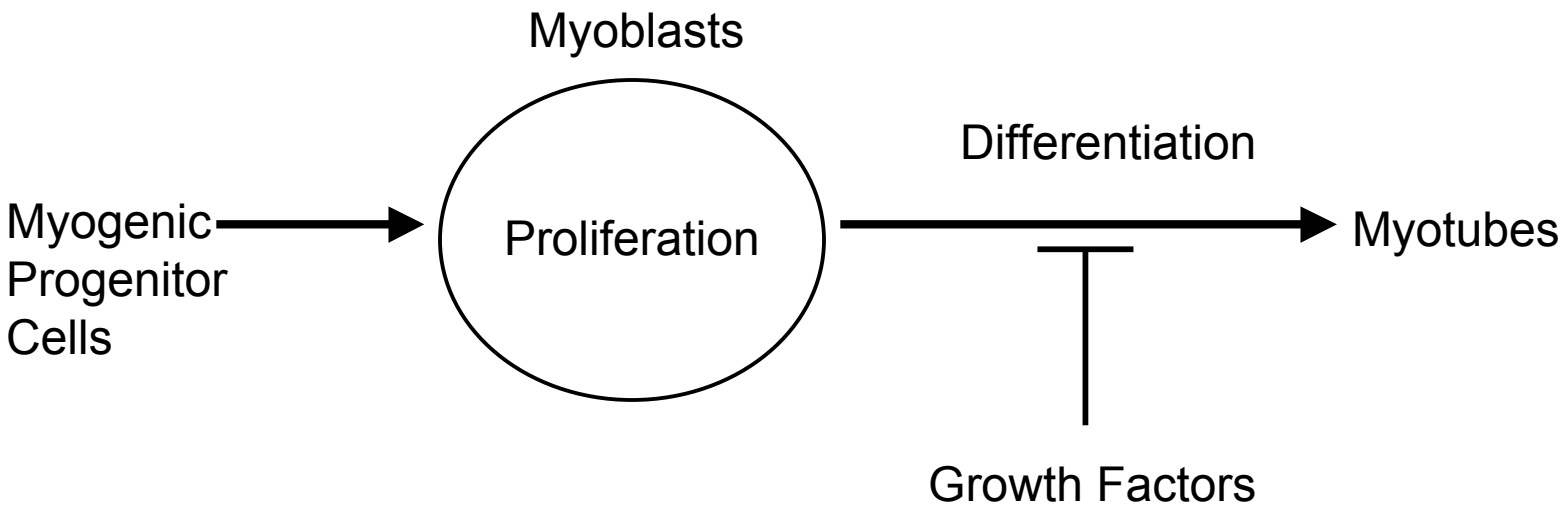
Highly plastic tissue.

Relatively little is understood concerning the mechanisms that regulate adaptive processes such as repair and regeneration.

The potential exists for therapeutic applications
e.g.
Cardiac tissue regeneration
Blood vessel growth
Muscle myopathies

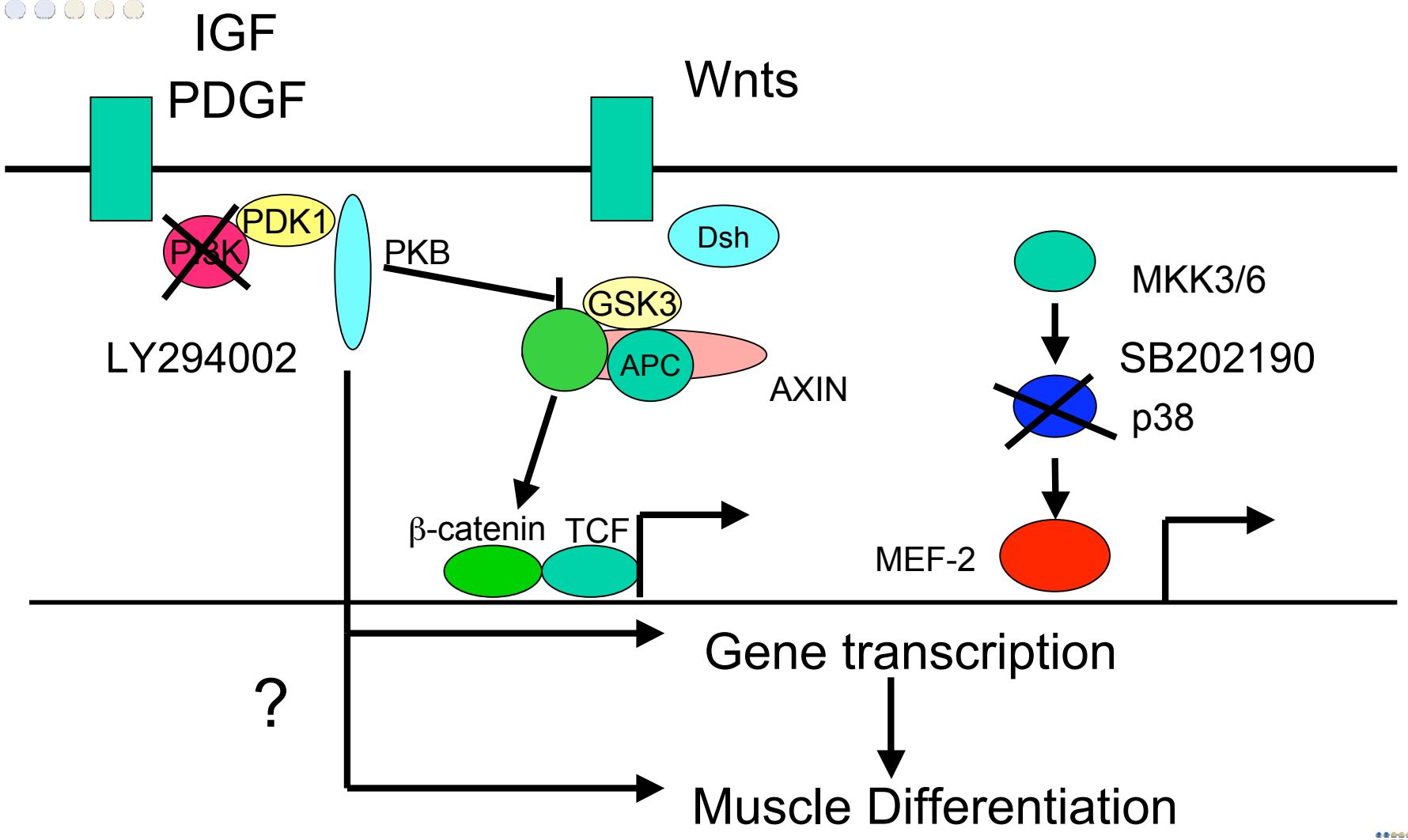


Differentiation of Skeletal Muscle:
myoblast to myotube



Introduction

- Over the past decade a great deal has been elucidated concerning muscle cell differentiation. Much of this information had arisen from the landmark discovery of the key muscle transcription factor, MyoD (Davis et al., 1987).
- Following this finding a number of transcription factors were isolated from muscle: myogenin, myf5, MRF4 and MEF-2 transcription factors A, B, C and D.
- The identification of binding sites for these transcription factors upstream of many muscle genes demonstrates the potential of these factors to regulate gene expression.
- Furthermore, the ability of these factors to force non-muscle cells down the path to muscle differentiation attests to the power of these transcription factors to regulate cellular fate.
- Recent insight has been gained to understand the signaling pathways that impinge upon these muscle transcription factors.
  - phosphatidylinositol 3-kinase (PI3-K), the mitogen activated protein kinase p38 , ERK and the Wnt pathways.
- No clear definition of the gene expression patterns associated with muscle differentiation.

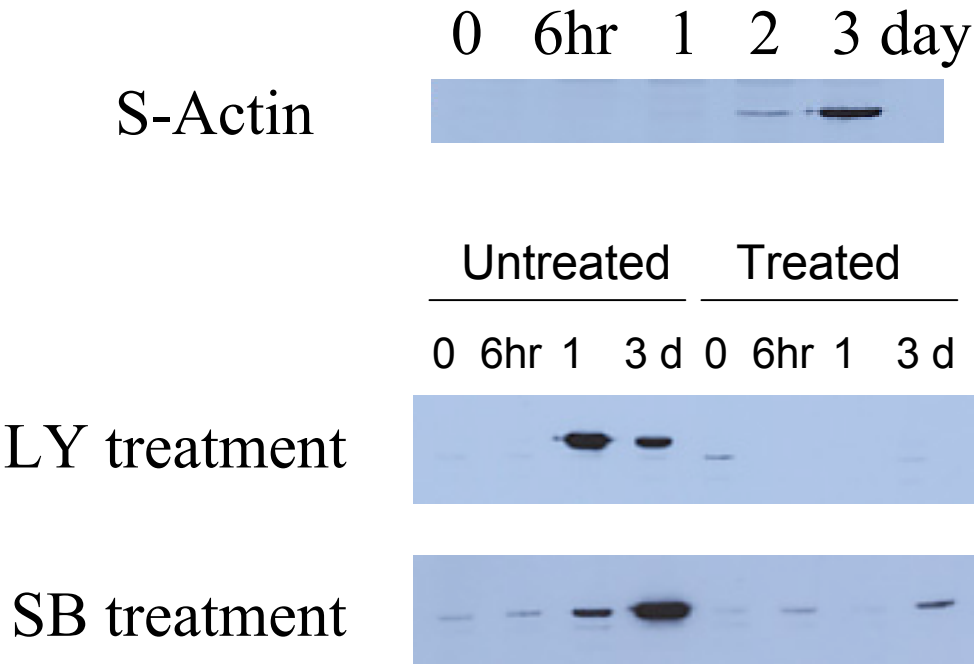

IGF
PDGF
Wnts
PDK1
PI3K
PKB
Dsh
GSK3
APC
AXIN
LY294002
MKK3/6
SB202190
p38
β-catenin
TCF
MEF-2
?
Gene transcription
Muscle Differentiation

Methods

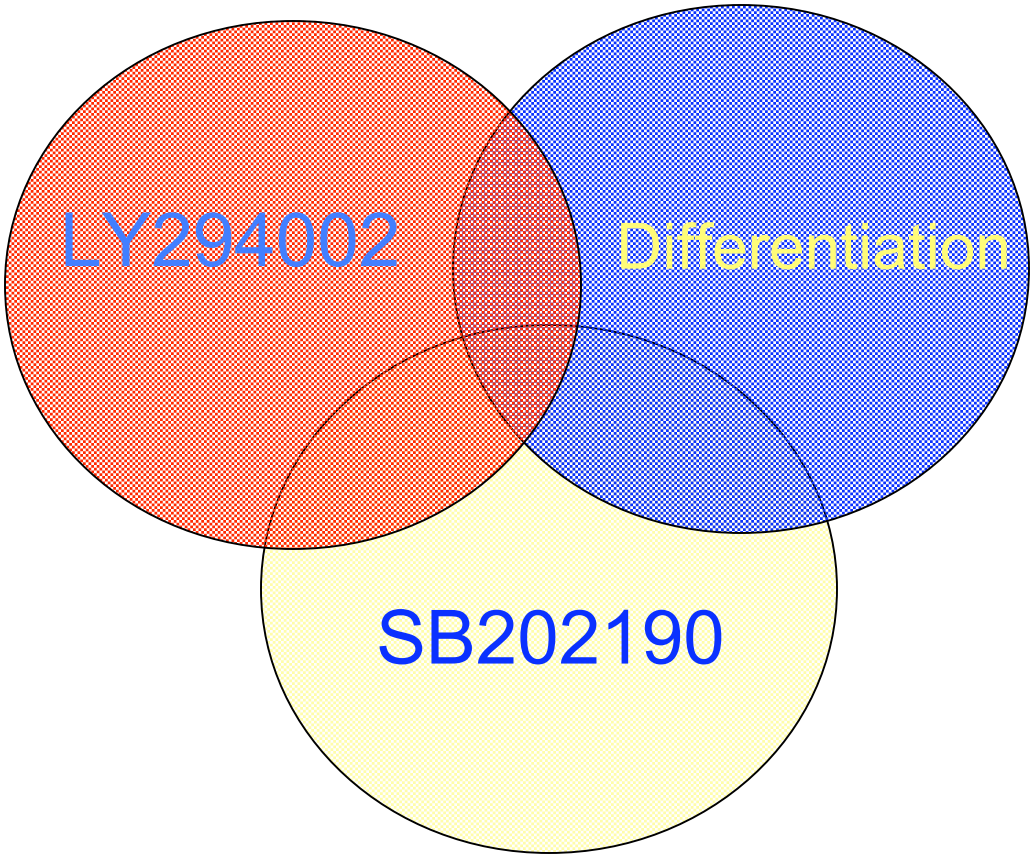
- C2C12 mouse skeletal muscle myoblast cell line
  - Grown to 80% confluence
  - Differentiation initiated by serum withdrawal and addition of 10 _g/ml IGF-1
  - Harvest cells and isolate total RNA

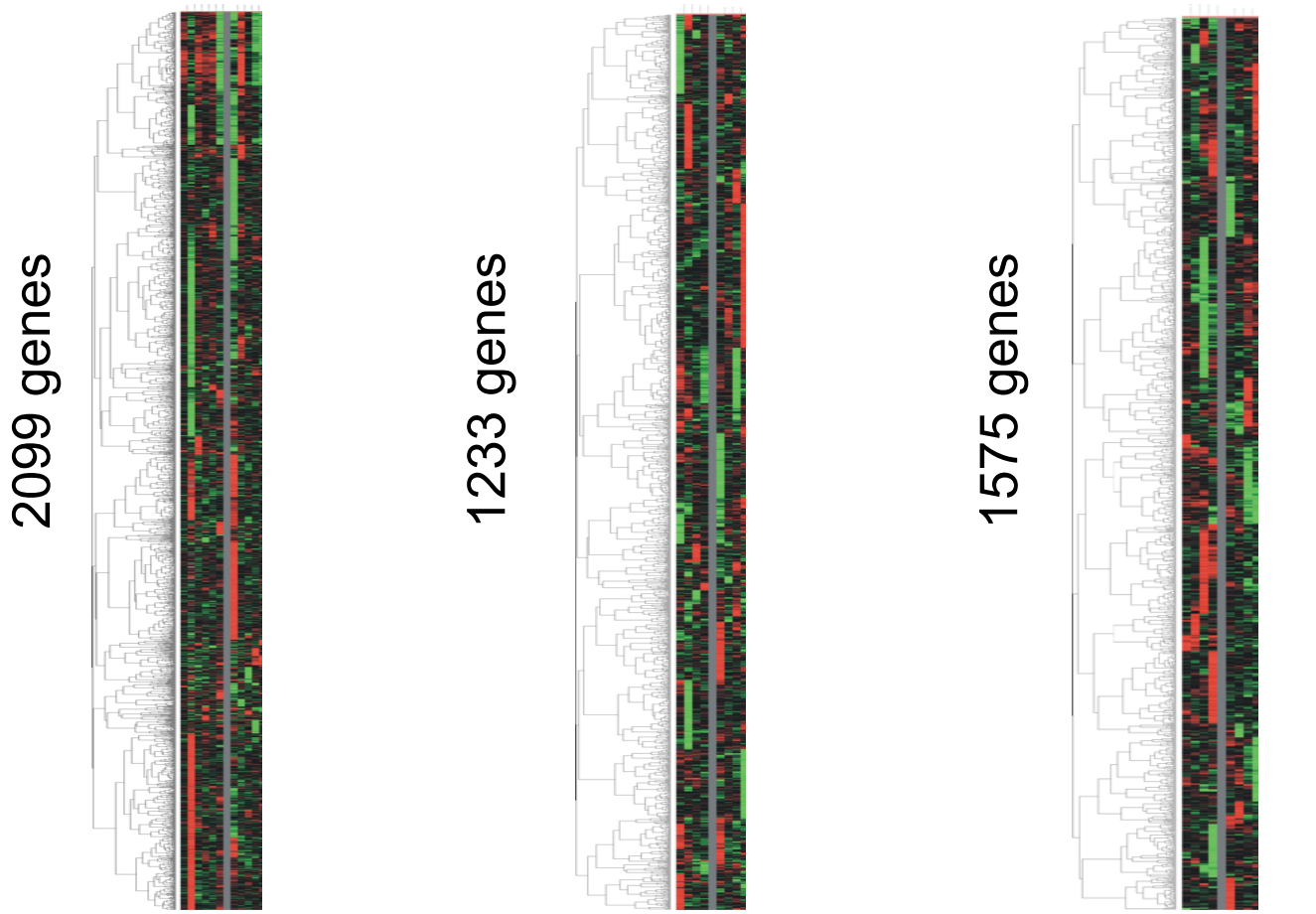
- Time course: 0, 6 hr, 1, 2, and 3 days

- Treatments: A) Differentiation
  B) LY294002-PI_3K inhibitor
  C) SB202190-P38 inhibitor



UHN Microarray Centre


LY294002
Differentiation
SB202190



Require another means of focusing in on the potential genes associated with muscle differentiation.

UHN Microarray Centre

Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF

Vishwanath R. Iyer*†‖, Christine E. Horak‖§, Charles S. Scafe*‖, David Botstein‖, Michael Snyder§ & Patrick O. Brown*‖¶

*Department of Biochemistry and ¶Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305, USA
‖Department of Genetics, Stanford University Medical Center, Stanford, California 94305, USA
§Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA
†These authors contributed equally to this work

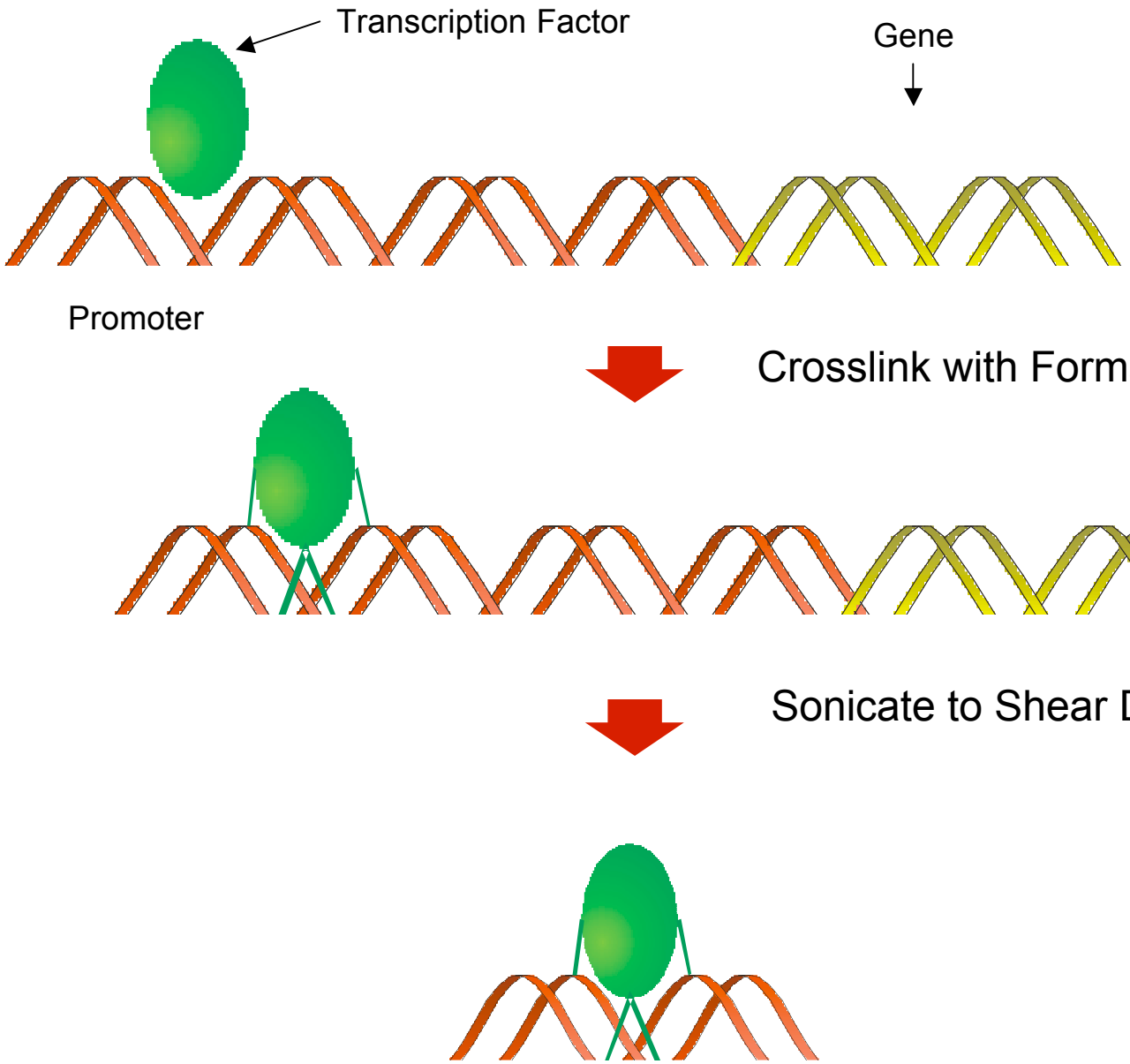
Nature January, 2001

ChIP (Chromatin Immunoprecipitation)

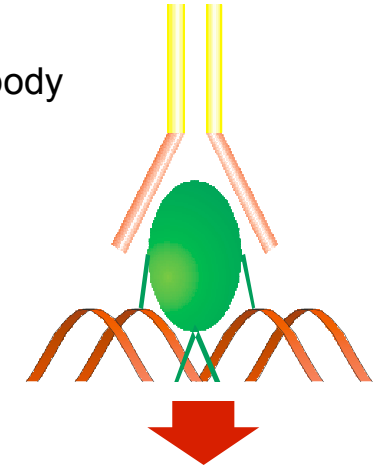
Provides us with the ability to identify genomic DNA associated with specific proteins (transcription factors)

Potentially have the capability to identify gene targets associated with specific transcription factors

How does ChIP work?

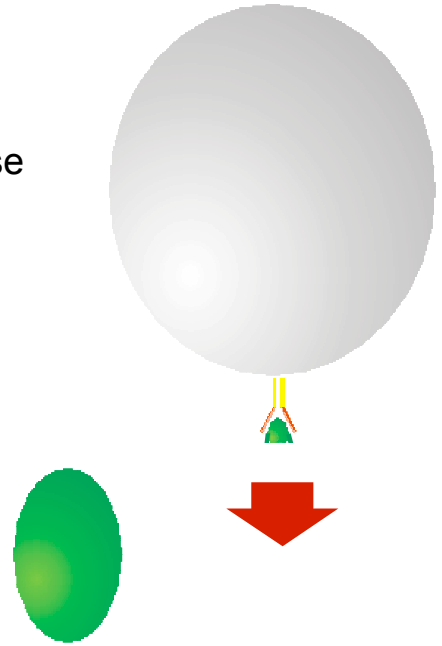

Transcription Factor
Gene
Promoter
Crosslink with Formaldehyde
Sonicate to Shear DNA

Antibody



Incubate With Antibody Raised Against Transcription Factor

Protein A Sepharose

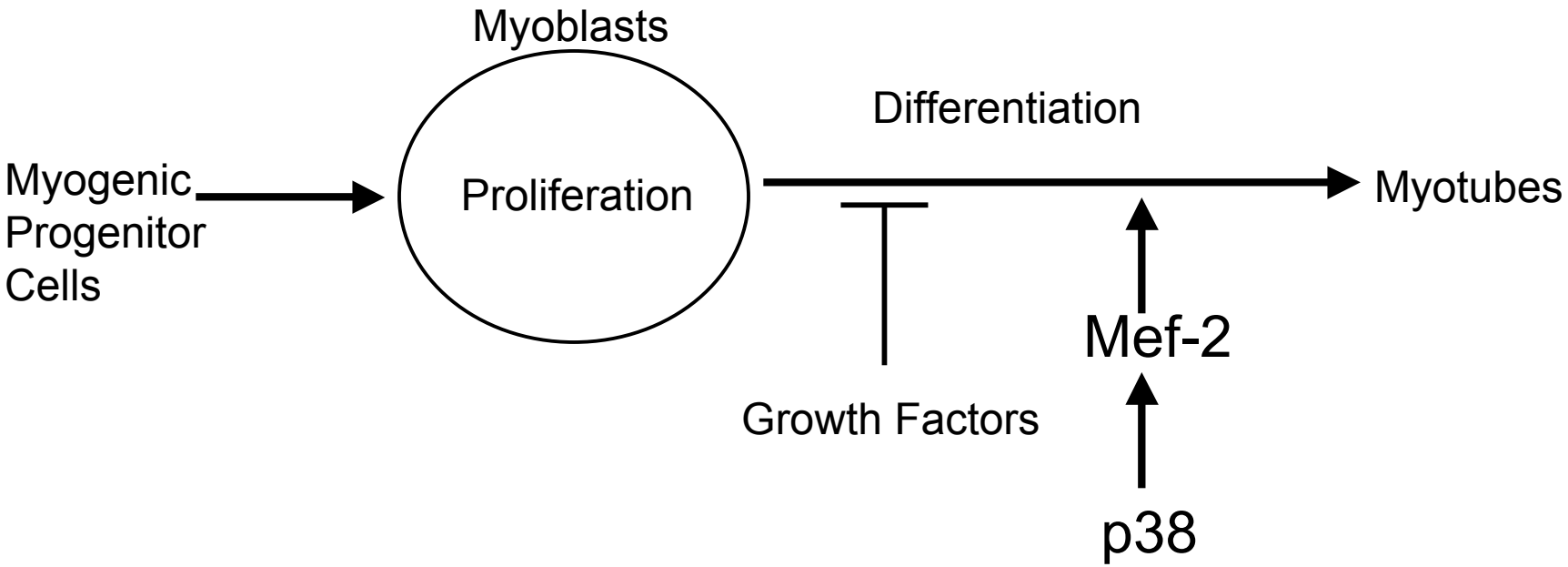


Incubate With Protein A Sepharose Beads and Precipitate

Reverse Crosslink by Incubating at High Temperature and High Salt

UHN Microarray Centre

How can we apply ChIP to understanding muscle differentiation?



MEF-2 (Myocyte Enhancing Factor)

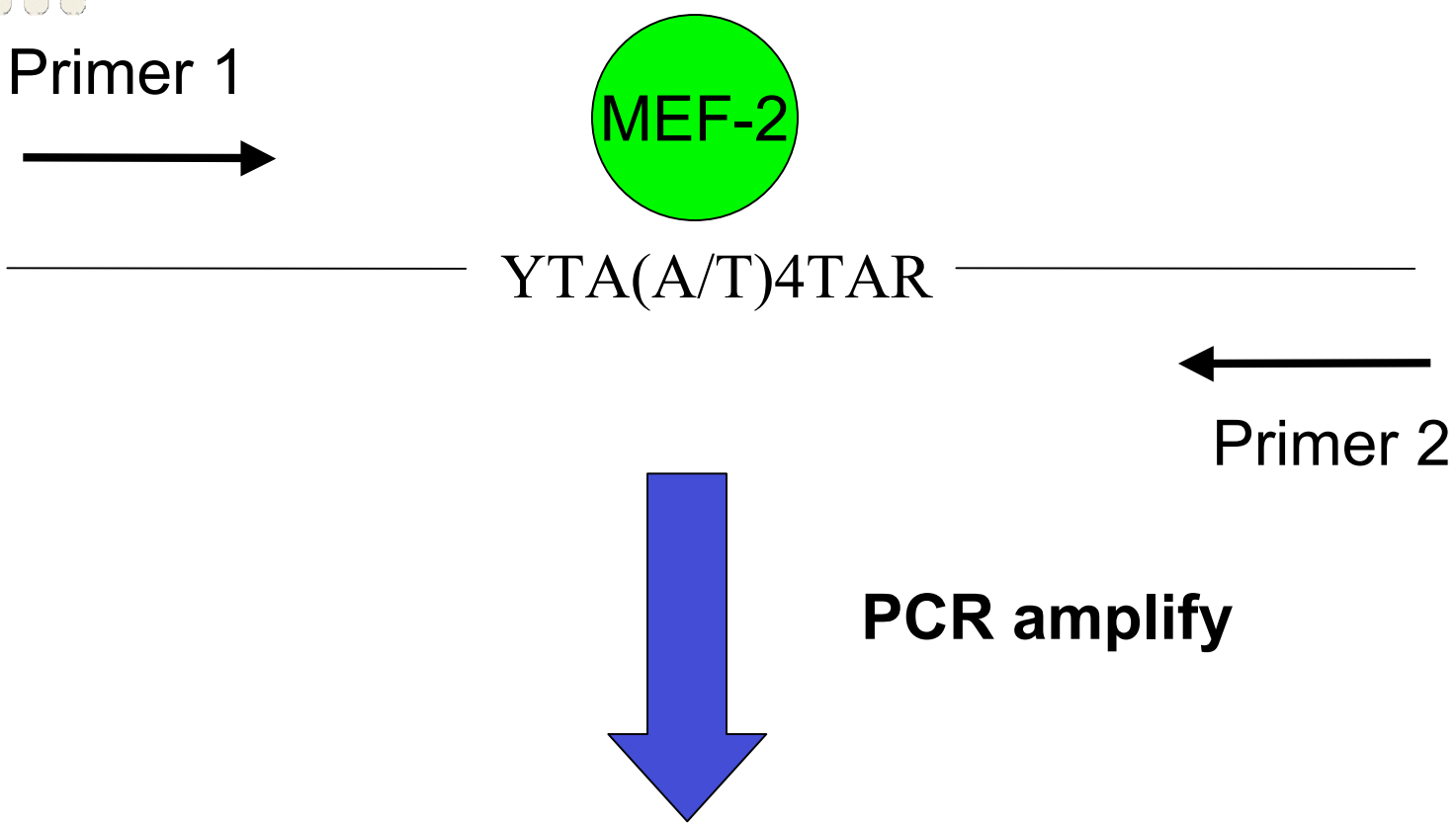
Muscle transcription factor known to be a key player in the regulation of muscle differentiation

Expression of a number of genes are known to be regulated by MEF-2
e.g. muscle creatine kinase, skeletal _-actin, myosin light chain, and myoglobin

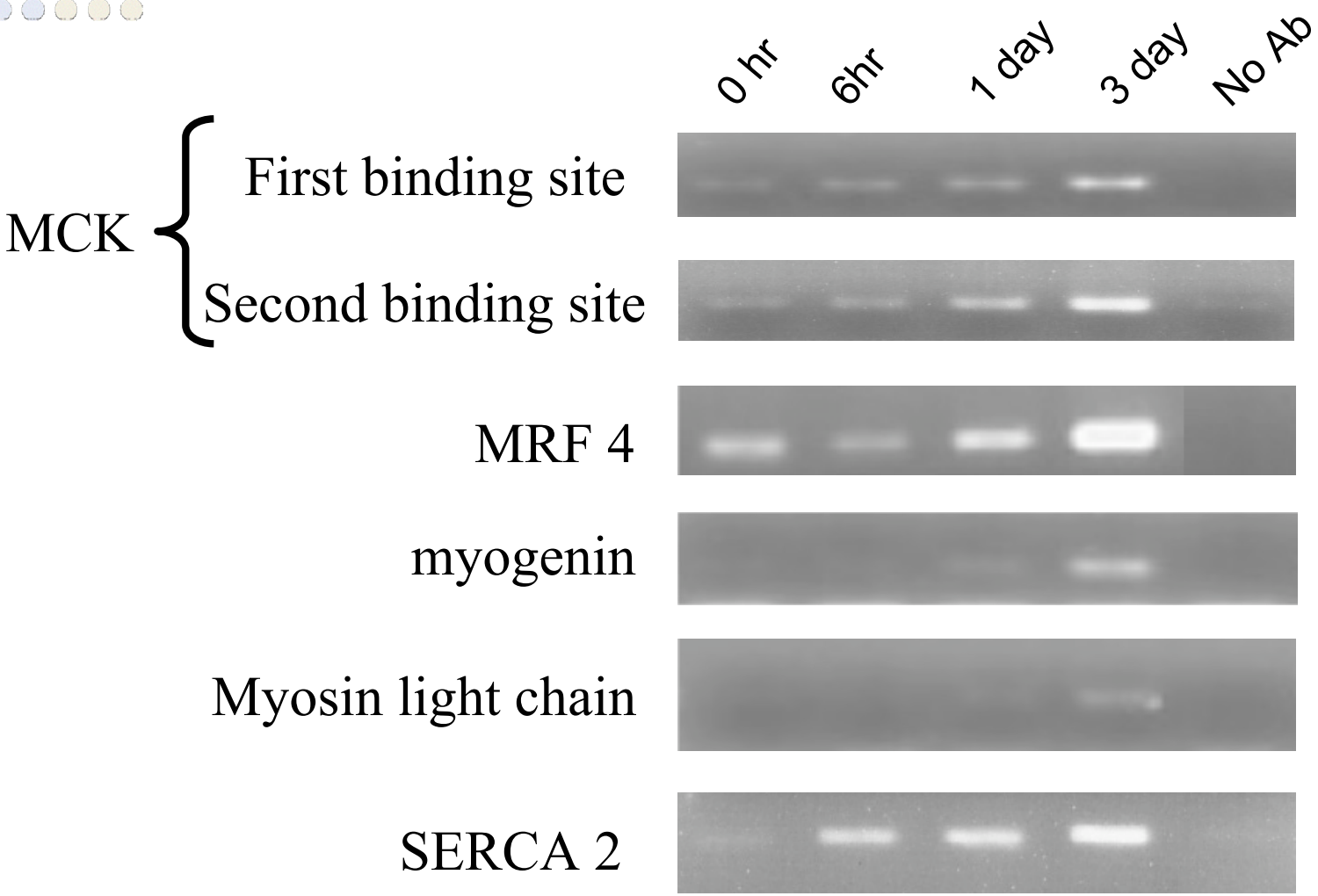
Binding sequence: $ATA(A/T)_4TAR$

Identify potential sites upstream of genes containing this sequence

Design primers to flank this region


Primer 1
MEF-2
YTA(A/T)4TAR
Primer 2
PCR amplify
~ 100-200 bp

ChIP Analysis of MEF-2 Associated Genes



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

Ideally, we need a means of profiling all of the regulatory sites/promoter regions of all genes.

Microarrays have been used to profile a large number of cDNA/genes simultaneously.

In order to use a microarray based method we need to have specific targets that can be arrayed.

CpG islands

•CpG islands are unmethylated C-G rich (60-70%) regions of the genome

•Account for approximately 2% of the genome

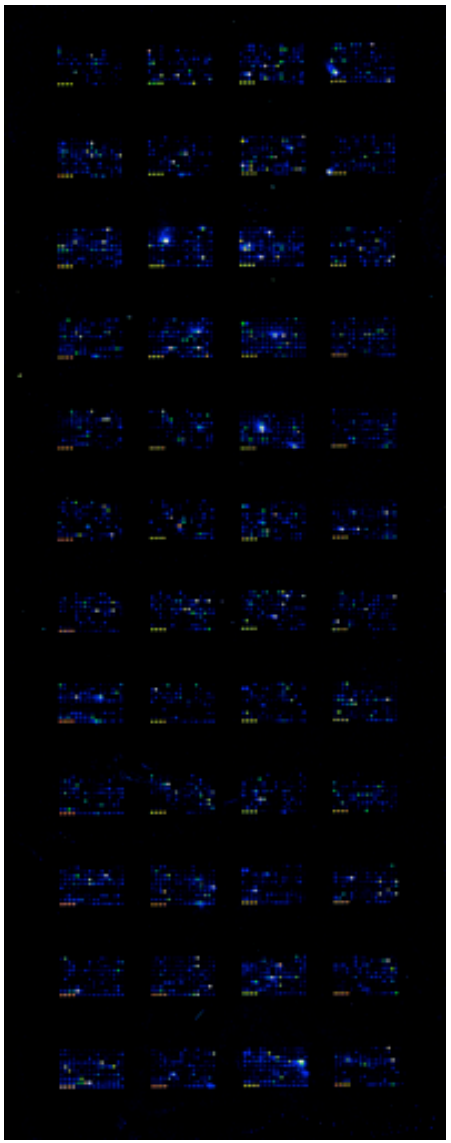
•Associated with the 5' ends of all house-keeping genes and a large number of regulated genes

•About 60% of human genes and 47% of mouse genes are associated with CpG islands

•About 80% of CpG islands are common between human and mouse

Construction of the 7K mouse CpG island microarray

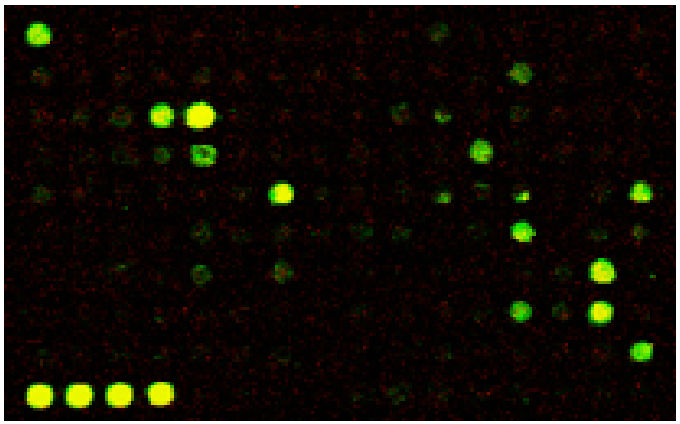
- Obtained mouse CGI library from the Sanger Institute UK
- All CpG islands were cloned into pGEM-5Zf vectors
- 100ul aliquot representing a total of 10^7 cells
- Plated library out on plates; approximately 300 colonies/plate
- Used colony picker (Genetix Q-pix2)
- Plated a total of 24 plates; approx. 7,000 viable colonies
- 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



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
www.microarrays.ca

Annotation of the CpG clones to identify location in the genome and potential genes in proximity with the CpG island.

Methods

C_2C_{12} cells were grown to 80% confluence

↓

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

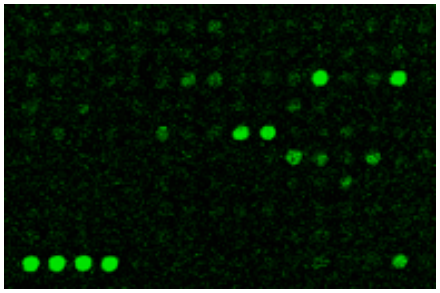
↓

Following 0, 6hours, 1 and 3 days cells were treated and harvested

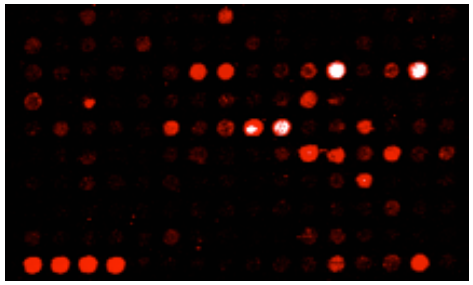
↓

Followed ChIP on CpG microarray protocol

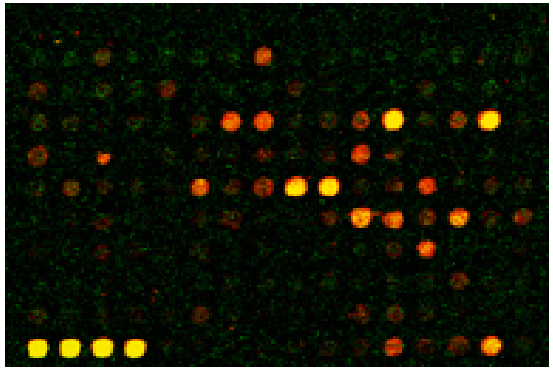
No Antibody



Antibody

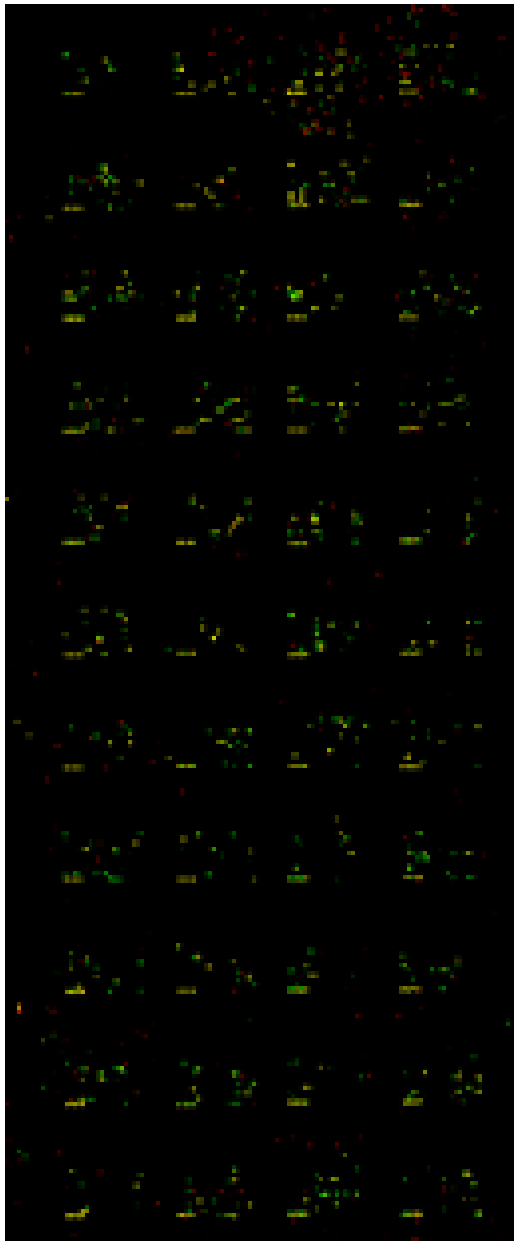


Composite

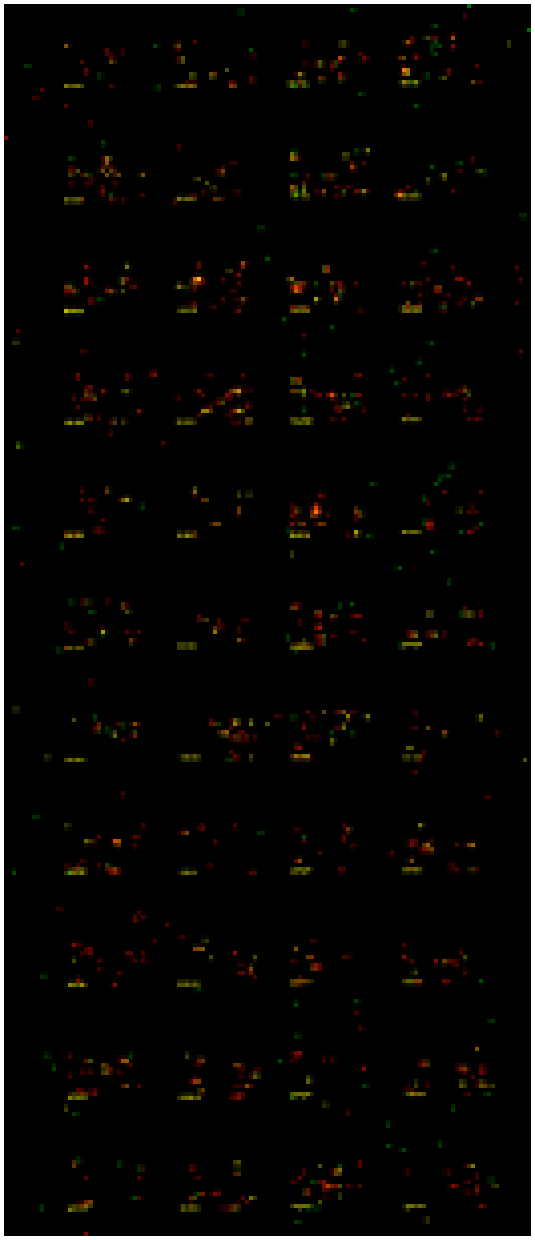


UHN Microarray Centre

0 hours



1 day



UHN Microarray Centre

Data analysis

•All arrays were scanned, quantified and entered into GeneTraffic.

•Normalize spot intensities to spiked Arabidopsis fragment.

•Filter out all spots in the experimental channel with an intensity of less than 512.

•Select only those spots that are > 2-fold in intensity than the no antibody control.

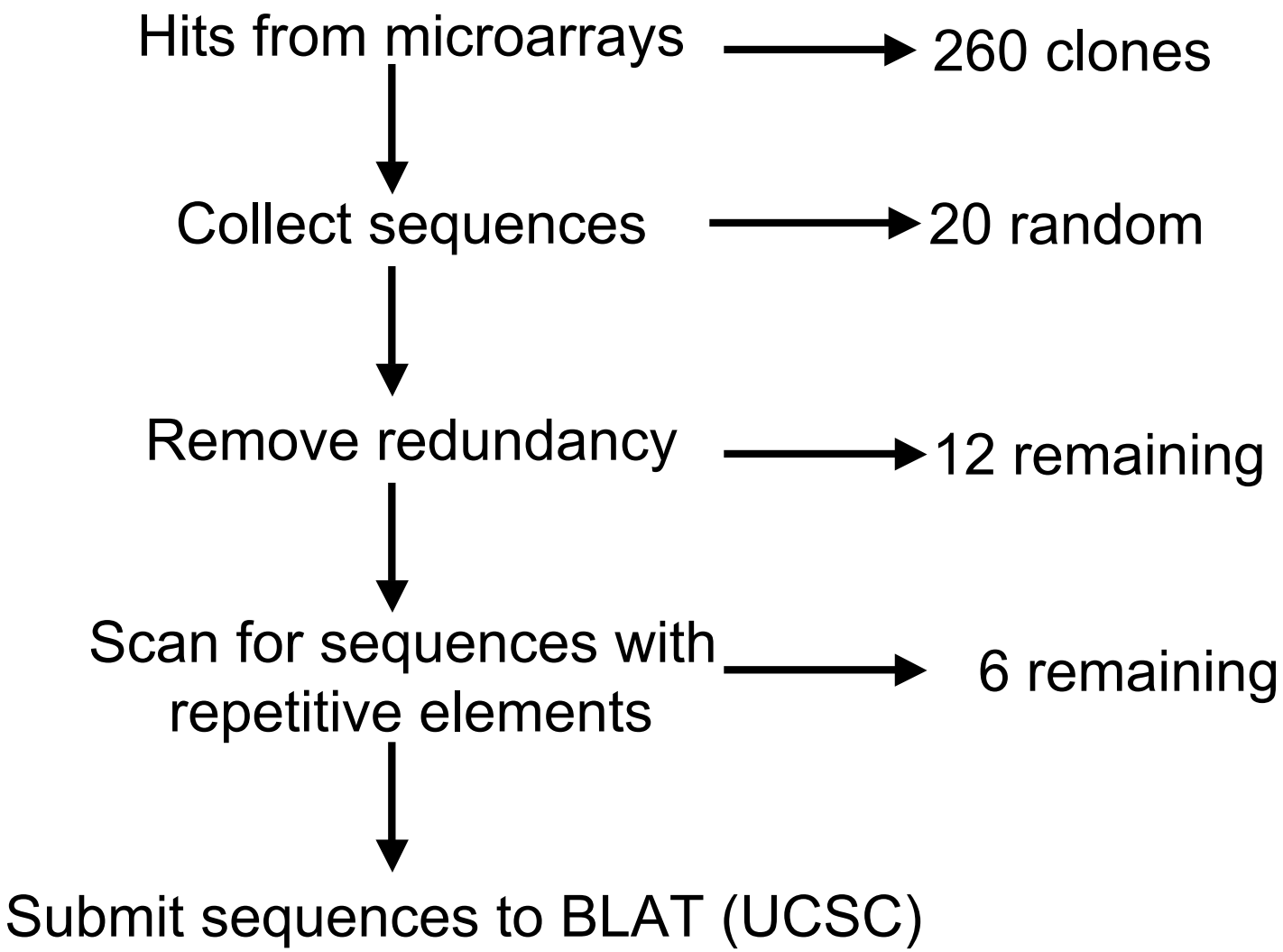
Results

Total number of positive clones;

0 hours: none
6 hours: none
1 day: 260
3 day: none

20 clones were randomly chosen and sequenced in both directions and queried against the mouse genomic sequence.

Informatics Pipeline



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

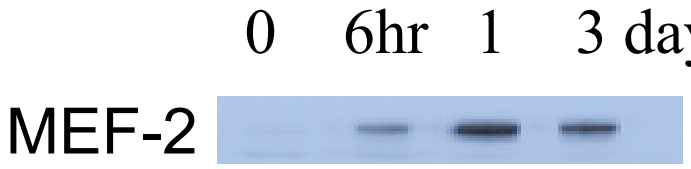
Use BLAT to locate this region on whole genome

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

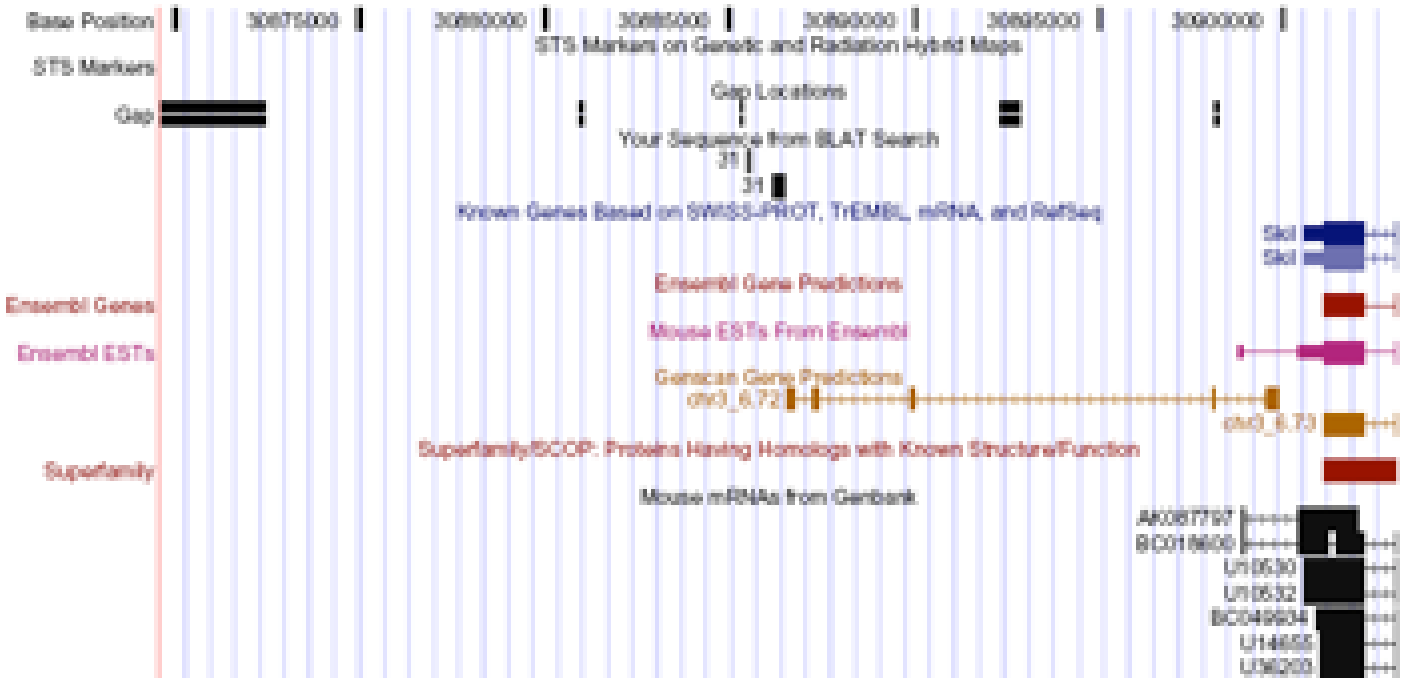
Sample ID	0 hour FOR	0 hour REV	6 hour FOR	6 hour REV	1 day FOR	1 day REV	3 day FOR	3 day REV
16K01	0.256413942	0.128071129	0.713034687	0.214087724	1.791018318	1.562965508	0.592307149	0.106925116
07H16	0.70023574	-0.103017219	0.9733921	0.194145439	1.70091124	1.26233237	0.795308229	-0.123129723
15G10	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

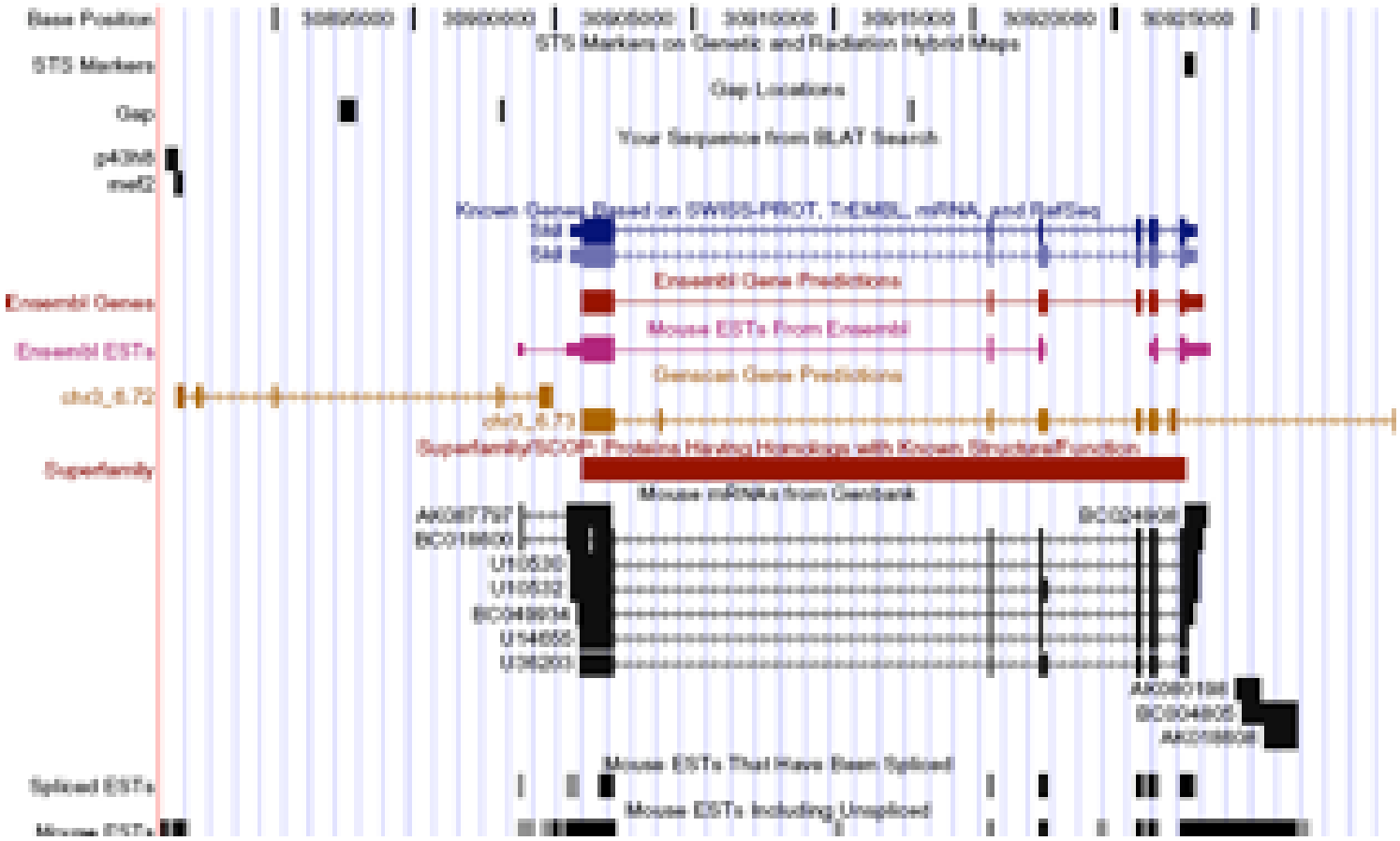


UHN Microarray Centre

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		


Base Position
30875000
30880000
30885000
30890000
30895000
30900000
STS Markers on Genetic and Radiation Hybrid Maps
STS Markers
Gap Locations
Gap
Your Sequence from BLAT Search
31
31
Known Genes Based on SWISS-PROT, TrEMBL, mRNA, and RefSeq
Skil
Skil
Ensembl Gene Predictions
Ensembl Genes
Mouse ESTs From Ensembl
Ensembl ESTs
Genscan Gene Predictions
chr3_6.72
chr3_6.73
SuperfamilySCOP: Proteins Having Homologs with Known Structure/Function
Superfamily
Mouse mRNAs from Genbank
AK087797
BC018600
U10530
U10532
BC049834
U14655
U36203



UHN Microarray Centre

SOURCE
GeneReport
M. musculus

Ski

Sloan-Kettering viral oncogene homolog

UniGene, LocusLink

Aliases

- 2310012O20Rik; 2610001A11Rik; BC004855; MGC8300; MGC 8300

Chromosomal Location

Chromosome/Cytoband	4 73.9 cM

Microarray Gene Expression Data

Data available	Show Gene Expression Data

SwissProt Information

SwissProt Accession No.	Q60698 **Ski oncogene** (*Mus musculus*), 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	nuclear
Similarity	to ski protein of sloan-kettering virus and to sno oncogenes
SwissProt Copyright	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See

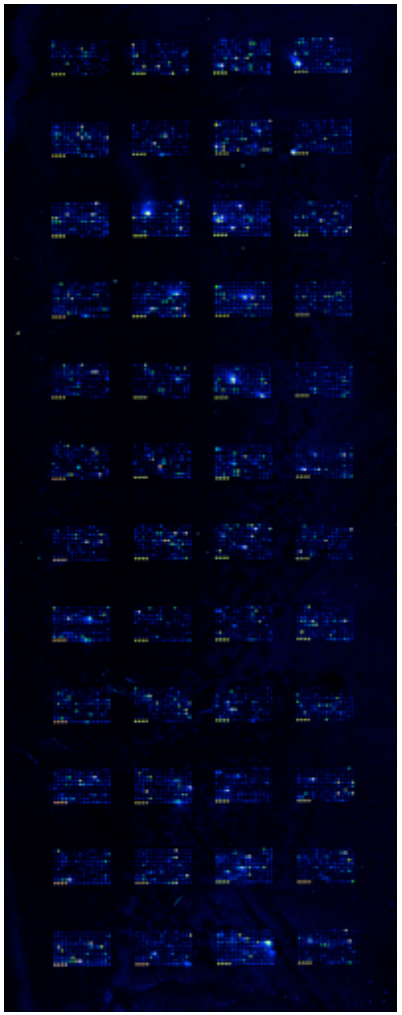
1Day non-amplified ChIP analysis

Reconfirm the results from the amplified data.

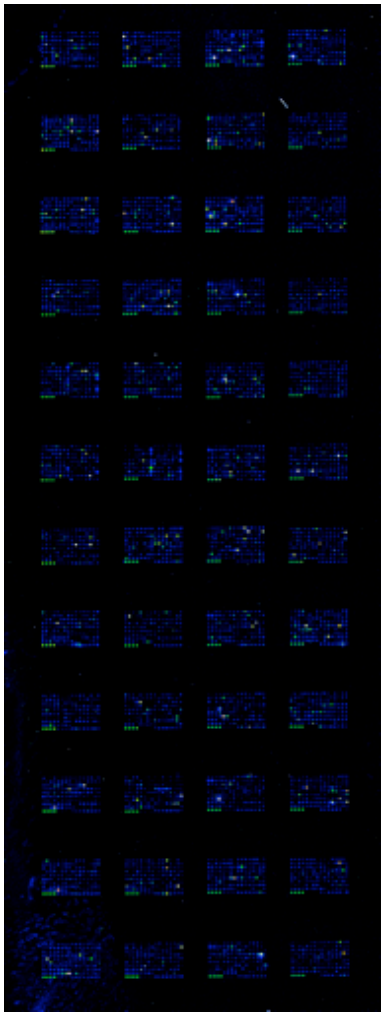
Plated cells onto thirty 150 mm dishes and grew to 80% confluence.

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

Amplified DNA



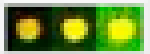
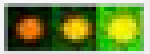
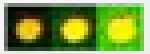
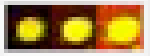
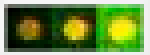
Non-amplified DNA



UHN Microarray Centre

Amplified DNA

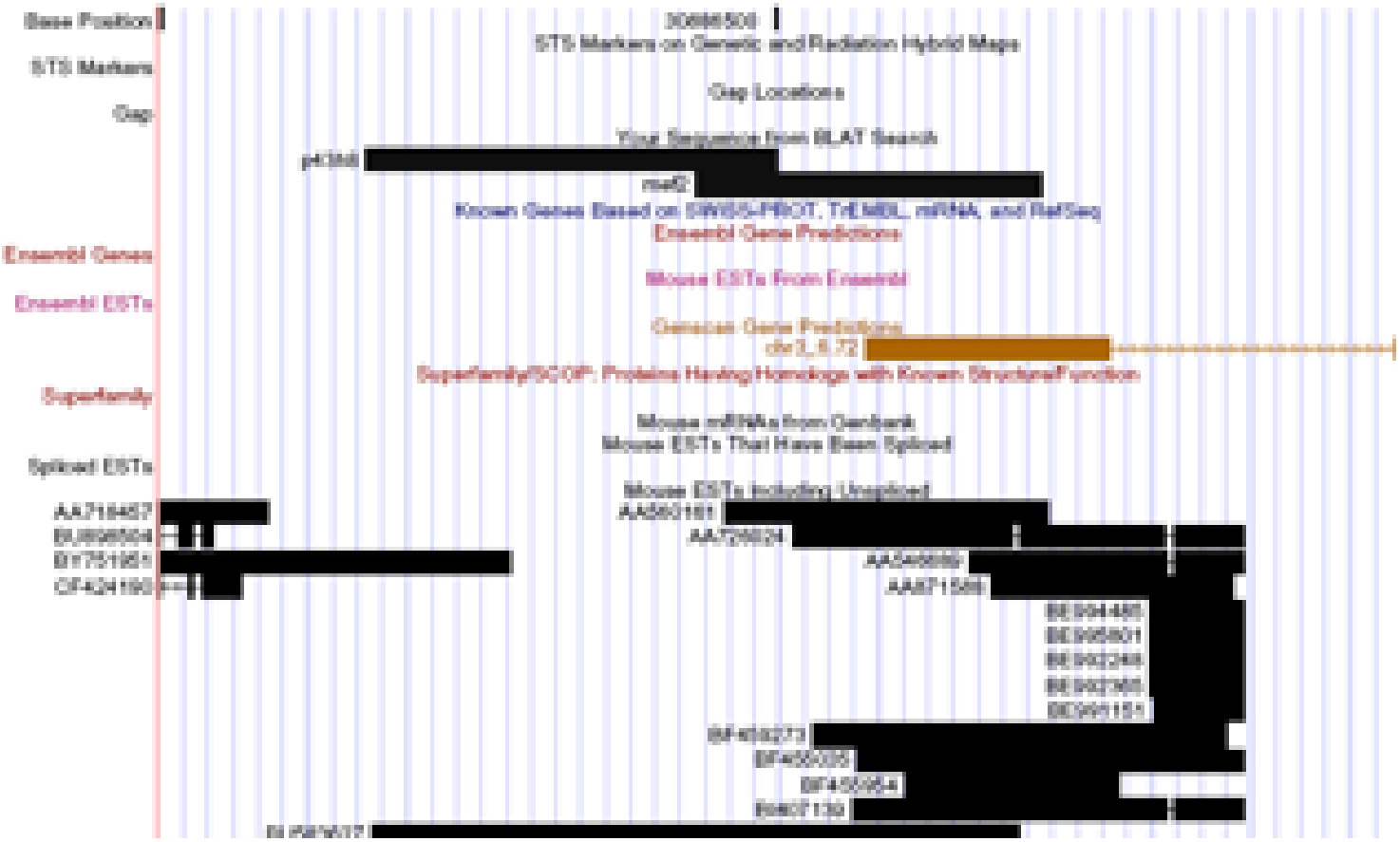
Search Results | Current Page: 1 | Total Pages: 1 | Total Results: 8 |

	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		



UHN Microarray Centre

What to do next?

Informatics

-all clones need to be sequenced (5' and 3' ends)
-use this information to locate the CpG islands in the genomic sequence
-identify any potential novel genes located upstream or downstream of the CpG island
-design specific primers to these potential targets and conduct ChIP analysis to confirm targets
-obtain full length sequence of novel gene targets

Biology

-look for gene expression at the level of mRNA and protein
-experiments to elucidate function of target genes e.g. run RNAi on interesting genes

Acknowledgements

Microarray Centre
James Woodgett

James Paris
Quyen Tran
Tuyet Diep
Carl Virtanen

Neil Winegarden
Eric Ho
Patrick Yau

All members of the Microarray Centre