Ottawa Health Research Institute
OHRI IRSO
Institut de recherche en santé d'Ottawa

StemBase:
the Database of the Stem Cell Genomics Project

Miguel Andrade
Bioinformatics
Ontario Genomics Innovation Centre
OHRI

AN INSTITUTE OF • UN INSTITUT DE

The Ottawa Hospital
L'Hôpital d'Ottawa

Université d'Ottawa
University of Ottawa

Ontario Genomics Innovation Centre

http://www.ottawagenomecenter.ca/

- Created in 2002 within the **Ottawa Health Research Institute** with the support of **Genome Canada** and the **Ontario Genomics Institute**.

- Mission: accomplish large-scale and technological projects in the field of genomics

- The Stem Cell Genomics Project is our first project.

Stem Cell Network Réseau de cellules souches

http://www.stemcellnetwork.ca/

- Canadian not-for-profit corporation; invest in Stem Cell Research; 65+ scientists, clinicians, engineers, and ethicists.
  - **I. Ethical, Legal, Social, and Policy Issues**
  - **II. Biology of Stem Cells**
  - **III. Bioengineering of Stem Cells**
  - **IV. Clinical Applications**

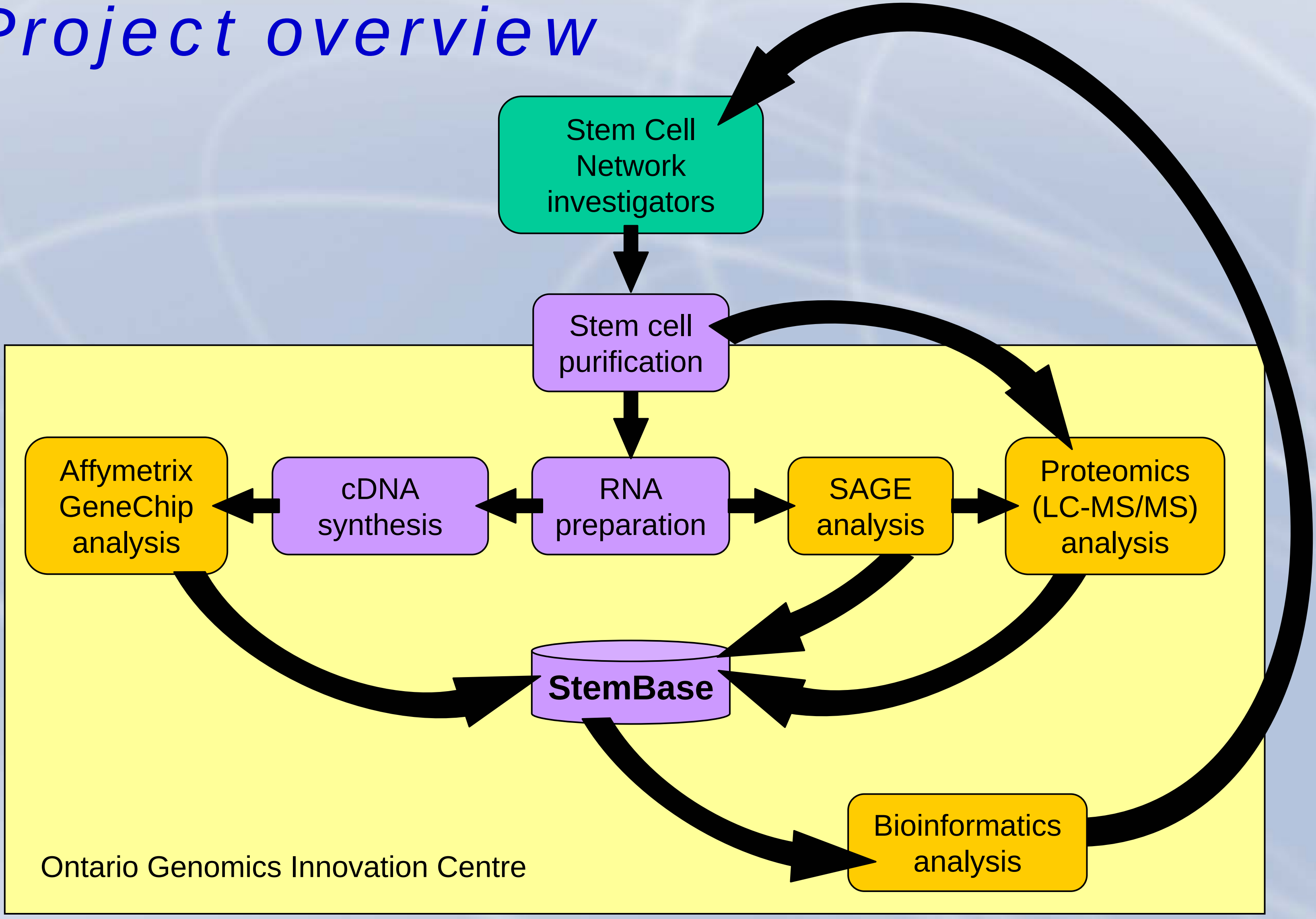
- Not-for-profit corporation; invest in Stem Cell Research; 65+ leading scientists, clinicians, engineers, and ethicists.
  - **I. Ethical, Legal, Social, and Policy Issues**
  - **II. Biology of Stem Cells**
    - ***Stem Cell Plasticity***
    - ***The Stem Cell Genomics Project***
    - ***HOXB4 Target-Genes Specifying Hematopoietic Stem Cell Self-Renewal***
  - **III. Bioengineering of Stem Cells**
  - **IV. Clinical Applications**

The Stem Cell Genomics project

- Objective: acquire a complete understanding of the genetic factors that:
  - ***specify stem cell identity and function; and***
  - ***regulate commitment and differentiation***

- Rationale:
  - **Stem cells play an essential role in the human body as they provide the starting material for every organ and tissue**
  - **Knowledge of regulatory genes acting in and on stem cells is necessary to exploit their full therapeutic potential**

Project overview



Stem Cell Network investigators

Stem cell purification

Affymetrix GeneChip analysis

cDNA synthesis

RNA preparation

SAGE analysis

Proteomics (LC-MS/MS) analysis

StemBase

Bioinformatics analysis

Ontario Genomics Innovation Centre

OHRI IRSO

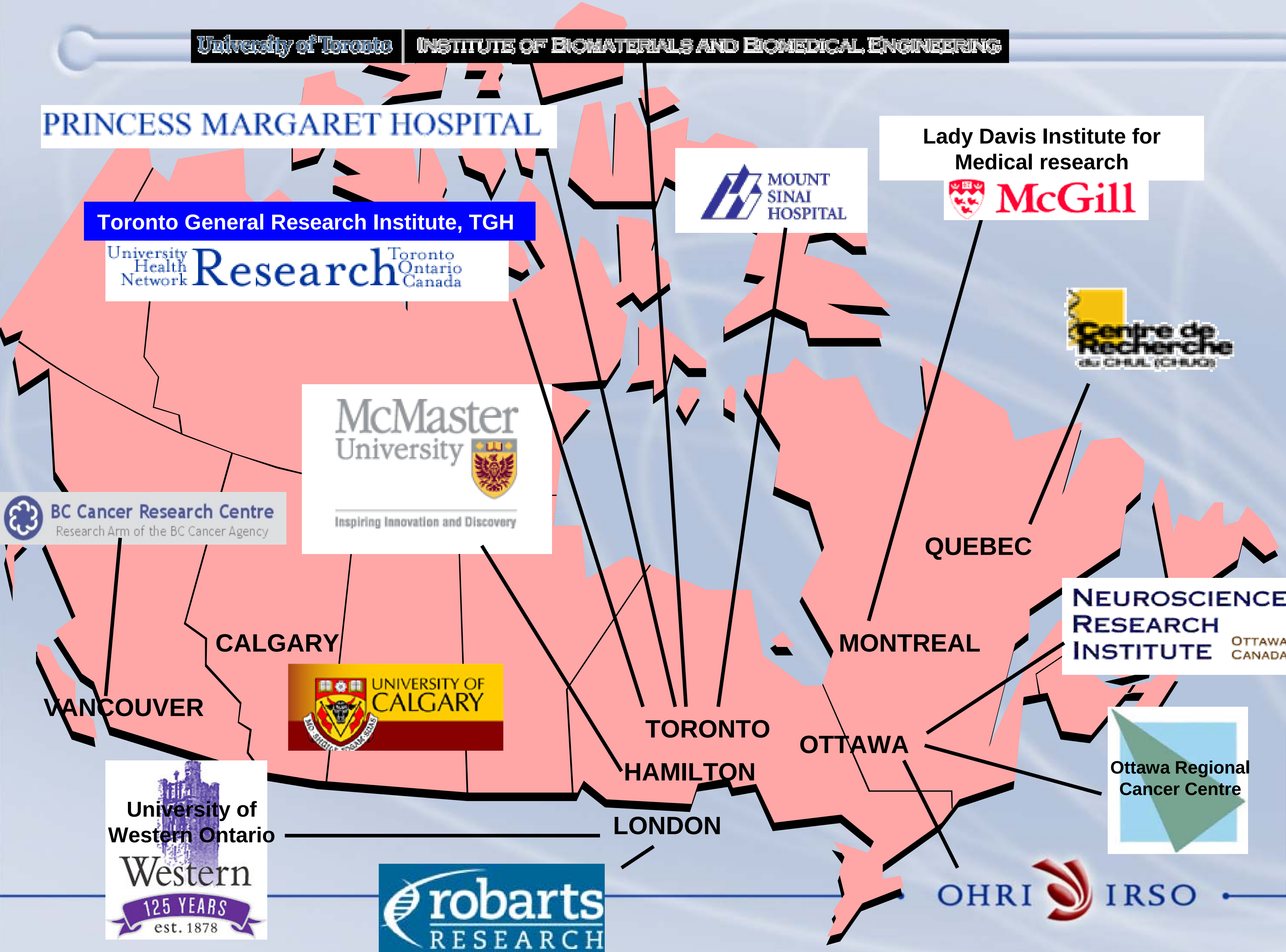
Microarray analysis

(Affymetrix GeneChip)

- **Objective: 250 preparations of stem cells and their downstream derivatives (March ’04 - 112)**
- **Identify common patterns of expression among different stem cell types**
- **Identify genes that:**
  - Maintain cells “stemness” (proliferative, multipotent)
  - Help drive differentiation into more specialized cell types
- **Identify preparations that are most homogeneous and suitable for SAGE analysis**

Sample Contributors

- Jane Aubin
- Mick Bhatia
- John Dick
- Jacques Galipeau
- Alain Garnier
- Marina Gertsentein
- John Hassell
- Keith Humphries
- Norman Iscove
- Michael McBurney
- Lynn Megeney
- James Piret
- Derrick Rancourt
- Janet Rossant
- Michael Rudnicki
- Luc Sabourin
- JP Tremblay
- T. Michael Underhill
- Valery Wallace
- Peter Zandstra



PRINCESS MARGARET HOSPITAL

Toronto General Research Institute, TGH

University Health Network Research Toronto Ontario Canada

MOUNT SINAI HOSPITAL

Lady Davis Institute for Medical research
McGill

Centre de Recherche du CHUL (CHUQ)

McMaster University
Inspiring Innovation and Discovery

BC Cancer Research Centre
Research Arm of the BC Cancer Agency

UNIVERSITY OF CALGARY

NEUROSCIENCE RESEARCH INSTITUTE OTTAWA CANADA

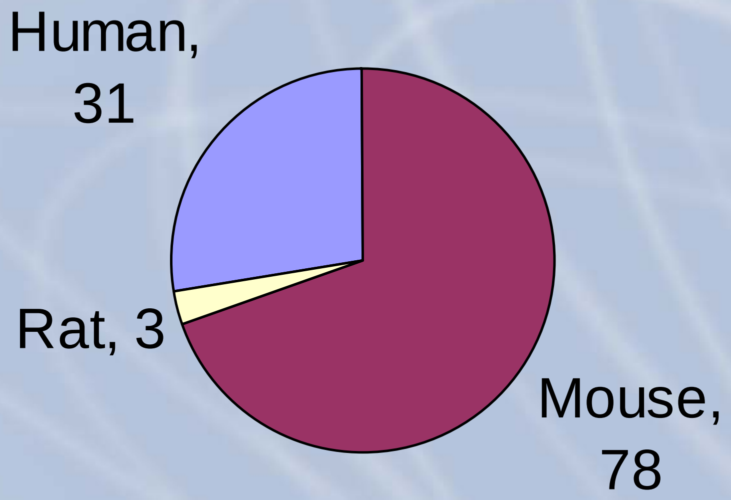
Ottawa Regional Cancer Centre

University of Western Ontario
Western 125 YEARS est. 1878

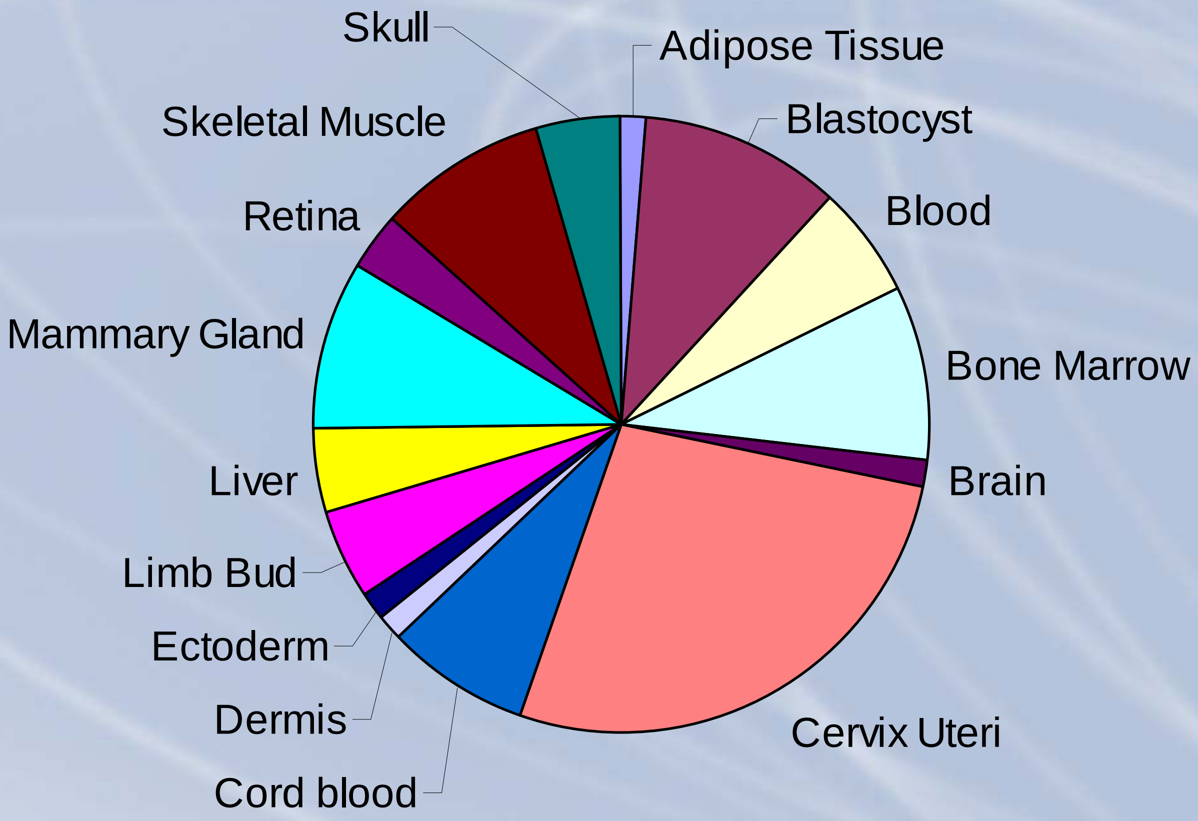
robarts RESEARCH

OHRI IRSO

Sample submissions by species



Sample distribution by tissue type



Serial Analysis of Gene Expression (SAGE)

- **Objective: Construct and analyze 20 SAGE libraries. March ‘04 – 4 libraries completed**
- **Perform gene expression analysis at the level of five mRNA copies per cell**
- **Facilitate the discovery of novel genes**

Bioinformatics

- **Organization of results in a database *StemBase***
- **Study of the microarray, SAGE, and proteomics data using different tools and techniques.**

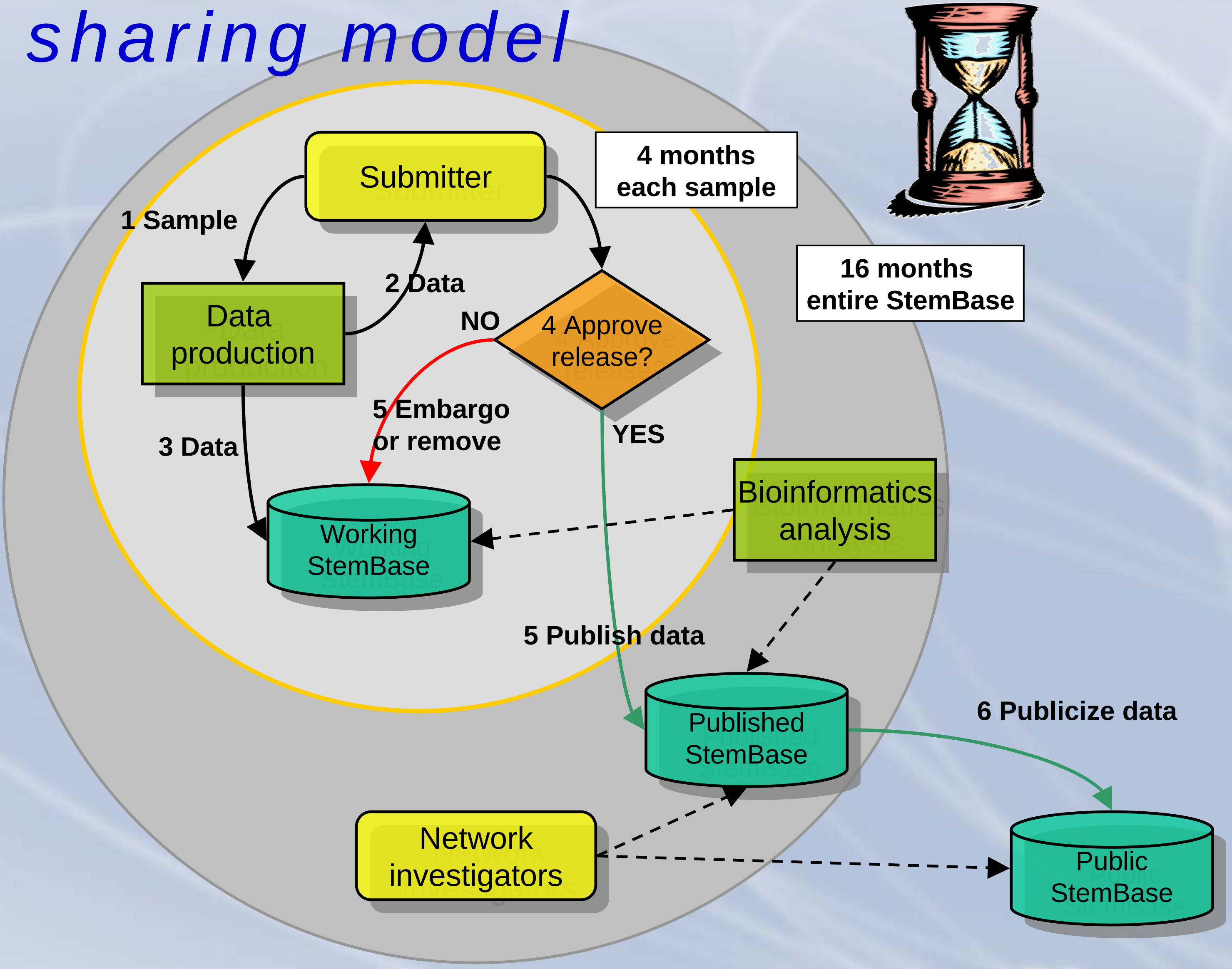
StemBase

StemBase

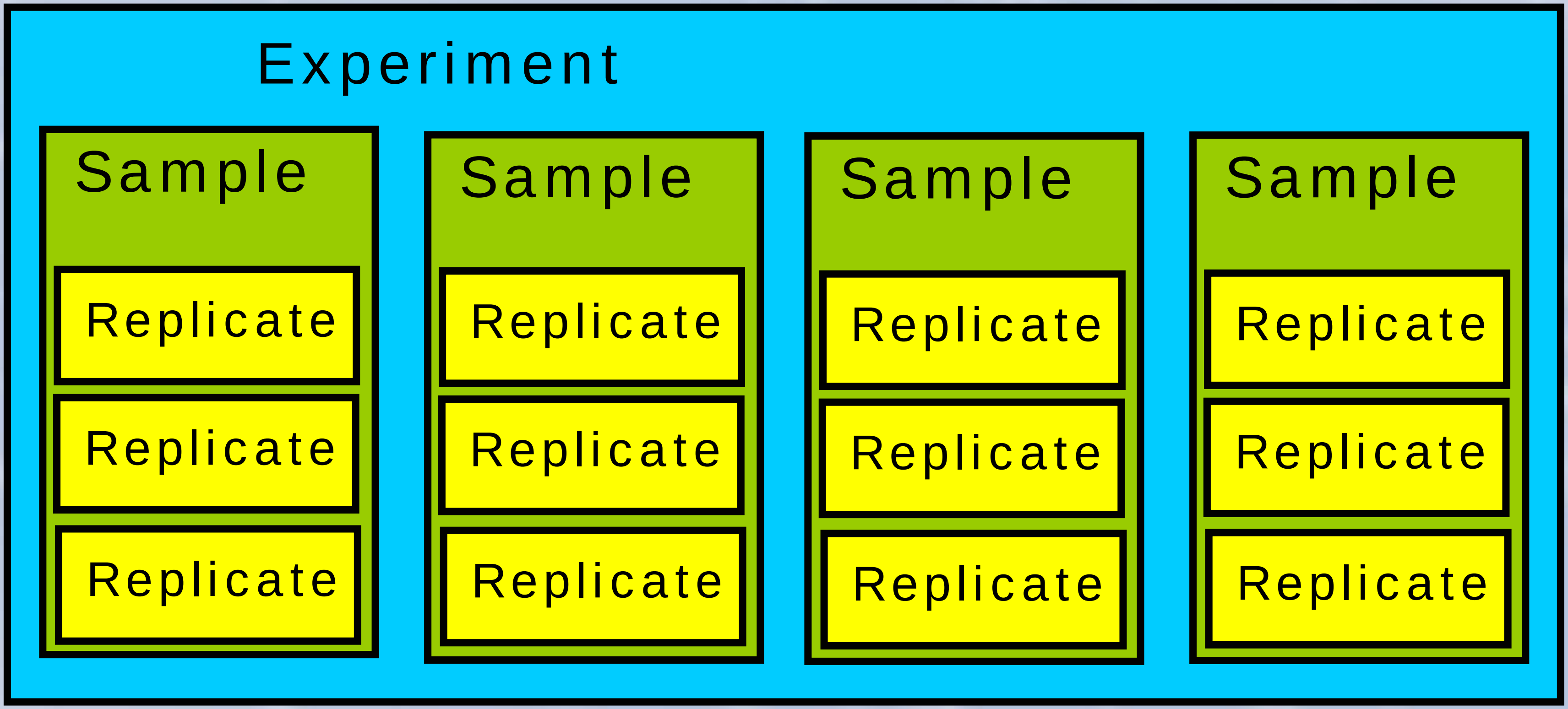
Single database incorporating data from all platforms

- Give access to raw data
- Offer browsing and querying of data
- Allow some level of analysis
- Links to other databases
- Web access to StemBase open to SCN researchers since February ’04. Public opening April ’04 http://www.ottawagenomecenter.ca:8080/StemBase

Data sharing model



Microarray data schema



Experiment

Sample

Replicate

Replicate

Replicate

Sample

Replicate

Replicate

Replicate

Sample

Replicate

Replicate

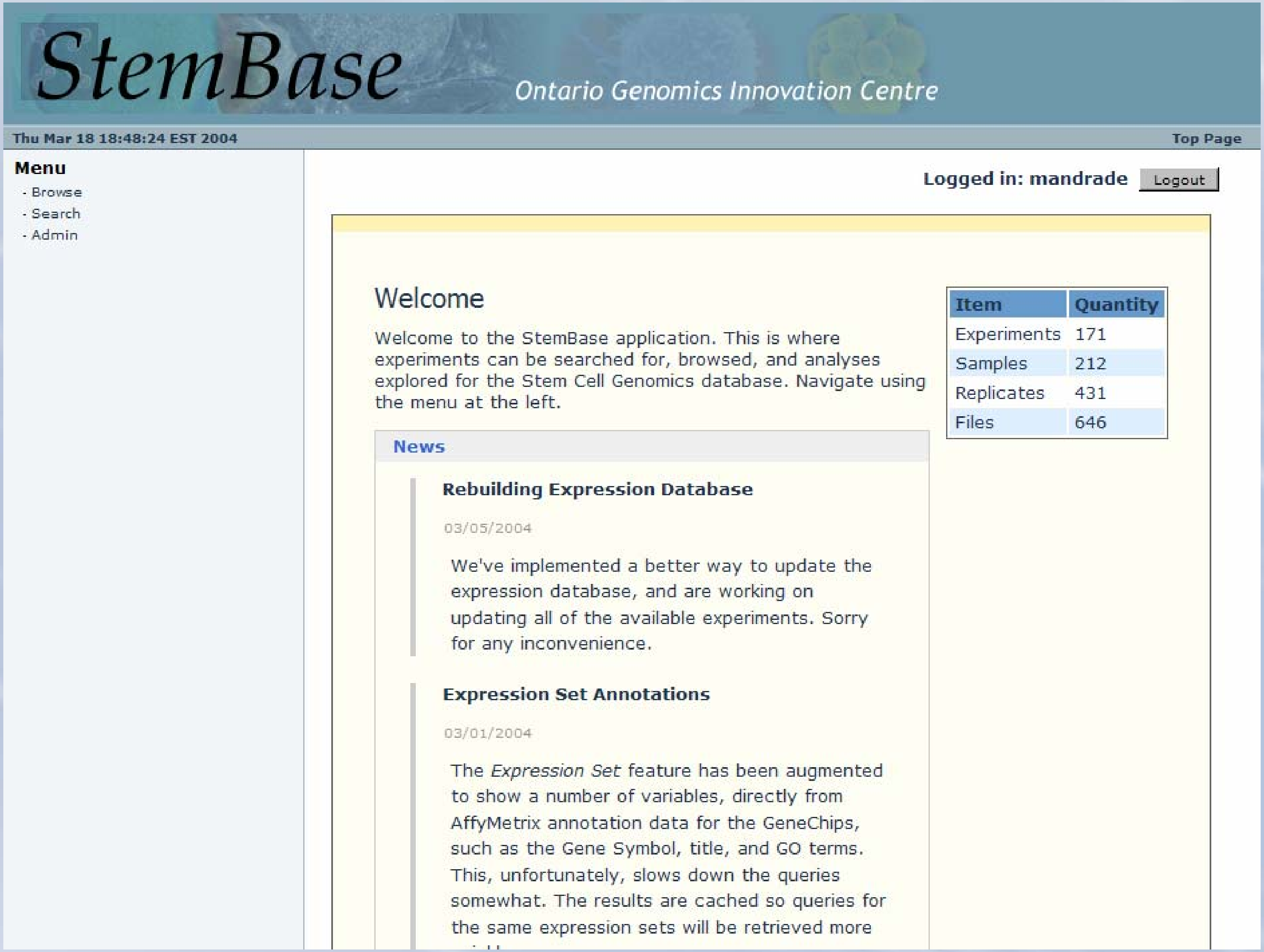
Replicate

Sample

Replicate

Replicate

Replicate



Welcome

Welcome to the StemBase application. This is where experiments can be searched for, browsed, and analyses explored for the Stem Cell Genomics database. Navigate using the menu at the left.

Item	Quantity
Experiments	171
Samples	212
Replicates	431
Files	646

News

Rebuilding Expression Database

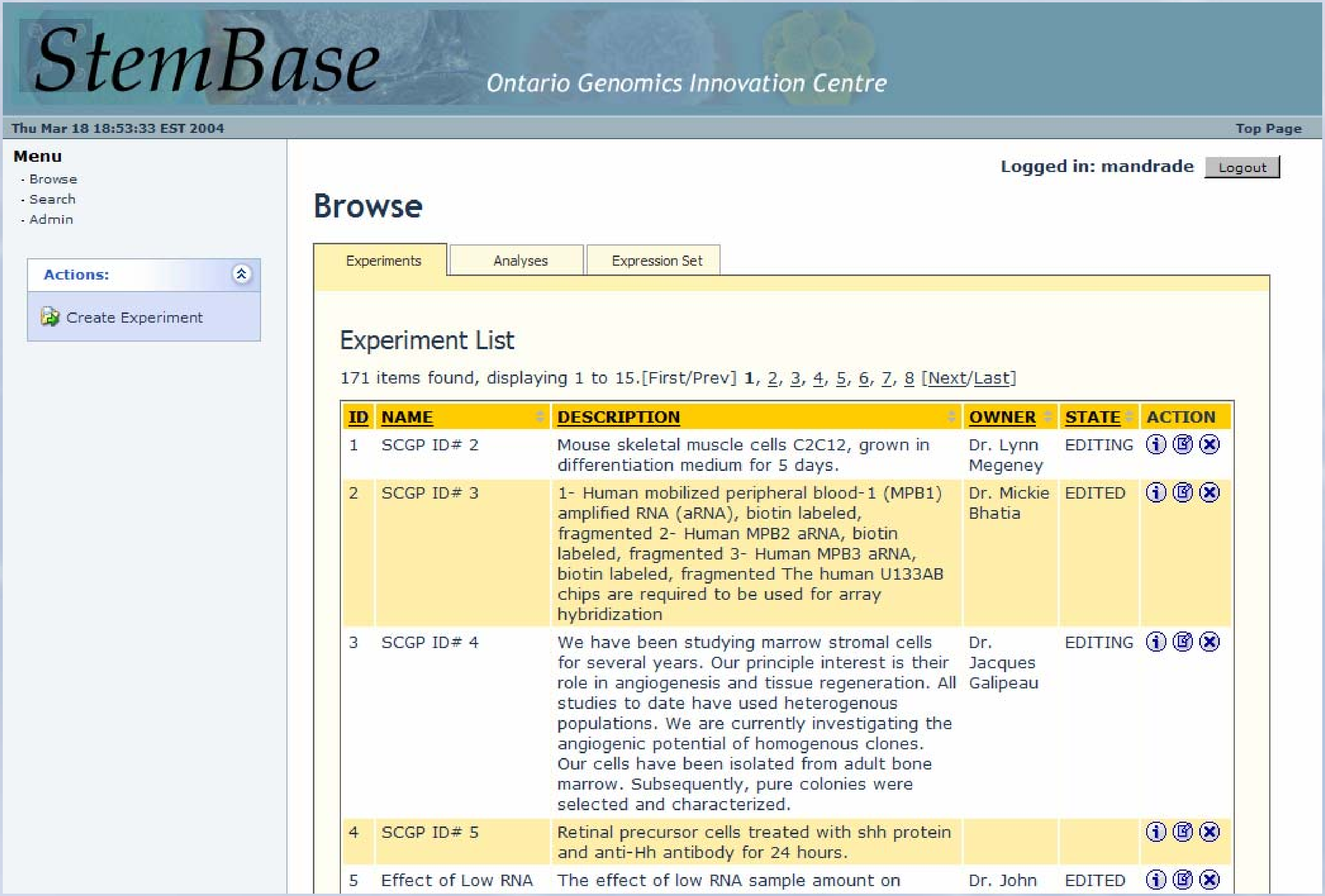
03/05/2004

We've implemented a better way to update the expression database, and are working on updating all of the available experiments. Sorry for any inconvenience.

Expression Set Annotations

03/01/2004

The *Expression Set* feature has been augmented to show a number of variables, directly from AffyMetrix annotation data for the GeneChips, such as the Gene Symbol, title, and GO terms. This, unfortunately, slows down the queries somewhat. The results are cached so queries for the same expression sets will be retrieved more



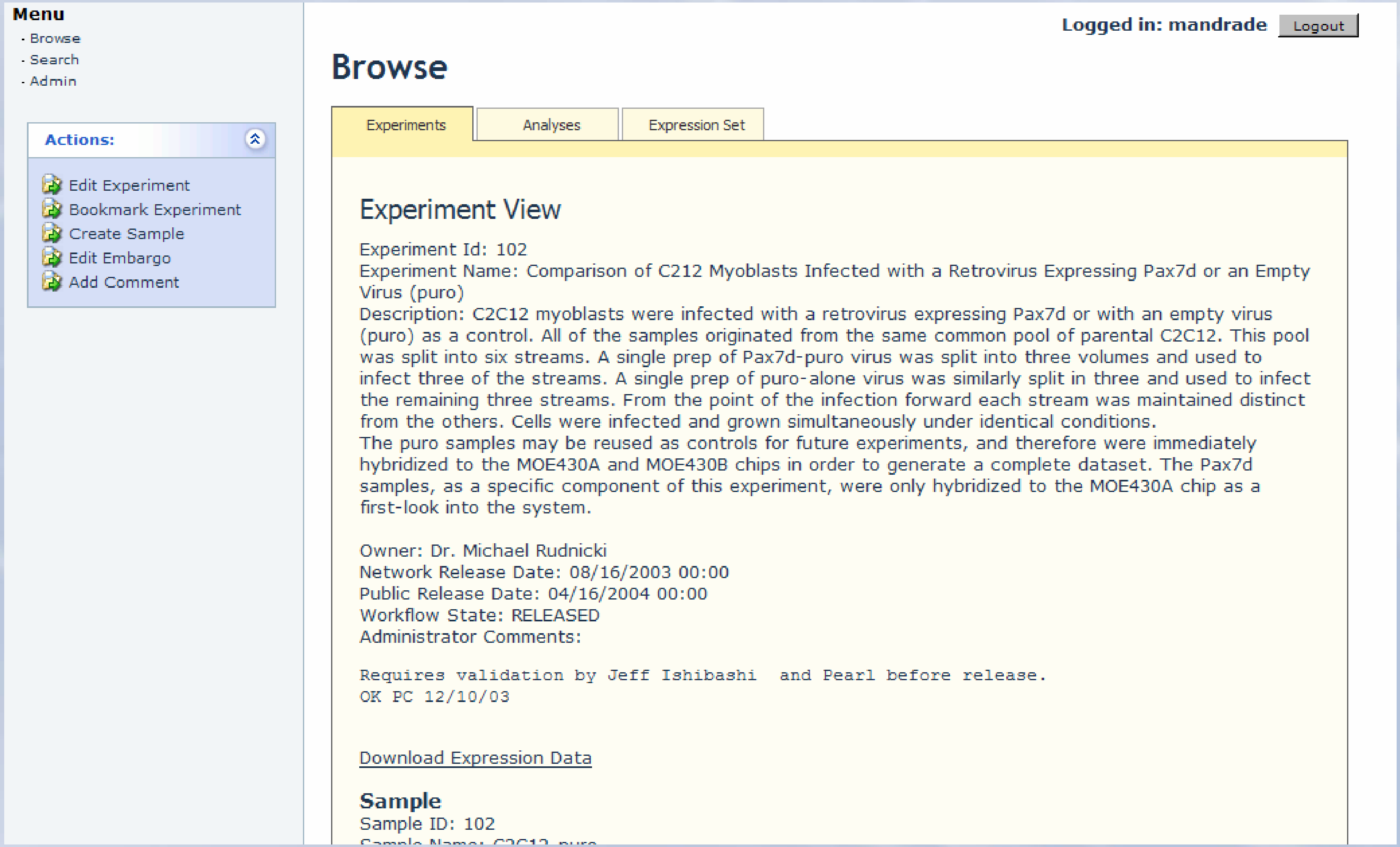
Browse

Experiment List

171 items found, displaying 1 to 15.[First/Prev] **1**, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

ID	NAME	DESCRIPTION	OWNER	STATE	ACTION
1	SCGP ID# 2	Mouse skeletal muscle cells C2C12, grown in differentiation medium for 5 days.	Dr. Lynn Megeney	EDITING	
2	SCGP ID# 3	1- Human mobilized peripheral blood-1 (MPB1) amplified RNA (aRNA), biotin labeled, fragmented 2- Human MPB2 aRNA, biotin labeled, fragmented 3- Human MPB3 aRNA, biotin labeled, fragmented The human U133AB chips are required to be used for array hybridization	Dr. Mickie Bhatia	EDITED	
3	SCGP ID# 4	We have been studying marrow stromal cells for several years. Our principle interest is their role in angiogenesis and tissue regeneration. All studies to date have used heterogenous populations. We are currently investigating the angiogenic potential of homogenous clones. Our cells have been isolated from adult bone marrow. Subsequently, pure colonies were selected and characterized.	Dr. Jacques Galipeau	EDITING	
4	SCGP ID# 5	Retinal precursor cells treated with shh protein and anti-Hh antibody for 24 hours.			
5	Effect of Low RNA	The effect of low RNA sample amount on	Dr. John	EDITED	

OHRI IRSO



Experiment View

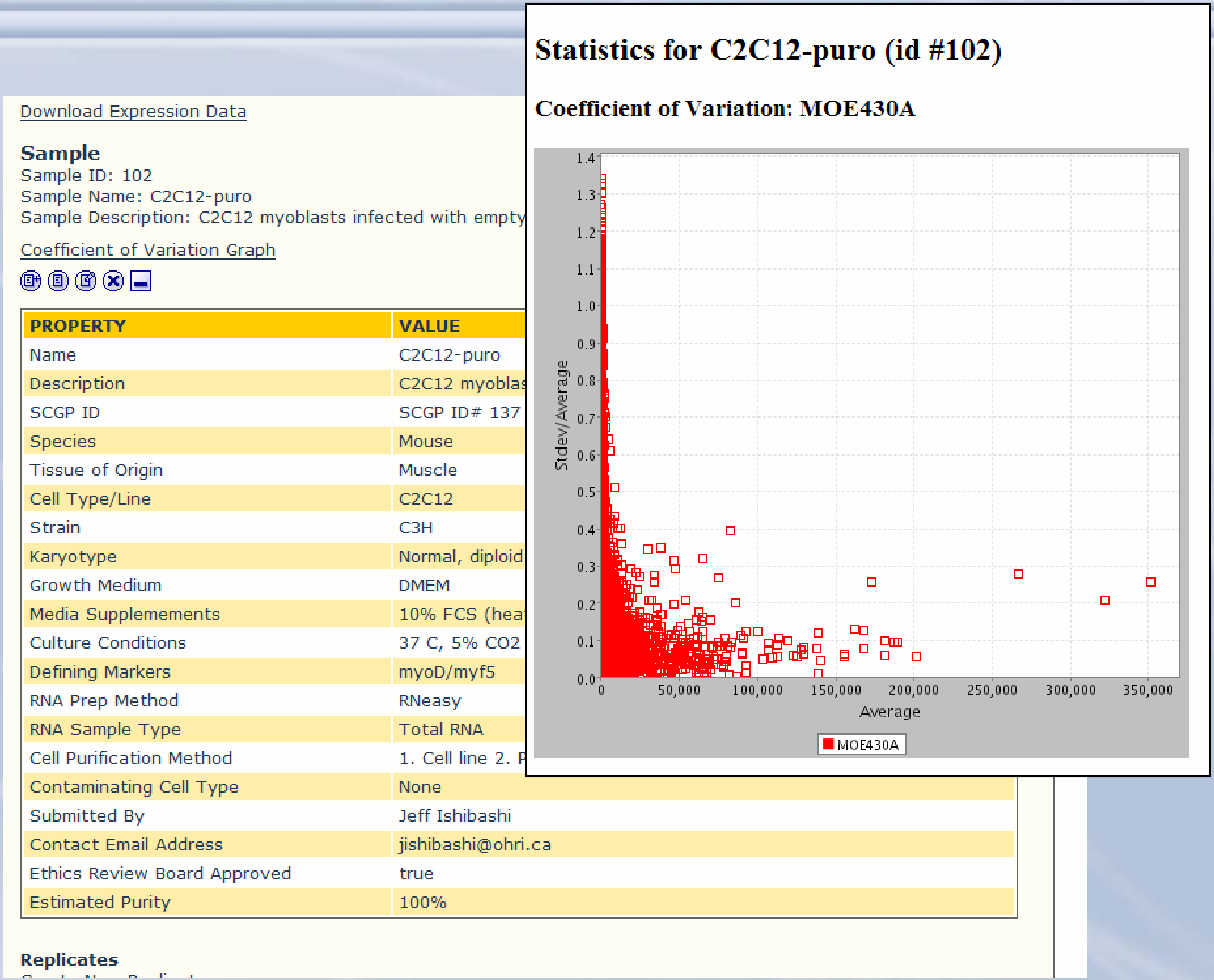
Experiment Id: 102
Experiment Name: Comparison of C212 Myoblasts Infected with a Retrovirus Expressing Pax7d or an Empty Virus (puro)
Description: C2C12 myoblasts were infected with a retrovirus expressing Pax7d or with an empty virus (puro) as a control. All of the samples originated from the same common pool of parental C2C12. This pool was split into six streams. A single prep of Pax7d-puro virus was split into three volumes and used to infect three of the streams. A single prep of puro-alone virus was similarly split in three and used to infect the remaining three streams. From the point of the infection forward each stream was maintained distinct from the others. Cells were infected and grown simultaneously under identical conditions.
The puro samples may be reused as controls for future experiments, and therefore were immediately hybridized to the MOE430A and MOE430B chips in order to generate a complete dataset. The Pax7d samples, as a specific component of this experiment, were only hybridized to the MOE430A chip as a first-look into the system.

Owner: Dr. Michael Rudnicki
Network Release Date: 08/16/2003 00:00
Public Release Date: 04/16/2004 00:00
Workflow State: RELEASED
Administrator Comments:

```
Requires validation by Jeff Ishibashi  and Pearl before release.
OK PC 12/10/03
```

Download Expression Data

Sample
Sample ID: 102
Sample Name: C2C12-puro



Download Expression Data

Sample
Sample ID: 102
Sample Name: C2C12-puro
Sample Description: C2C12 myoblasts infected with empty

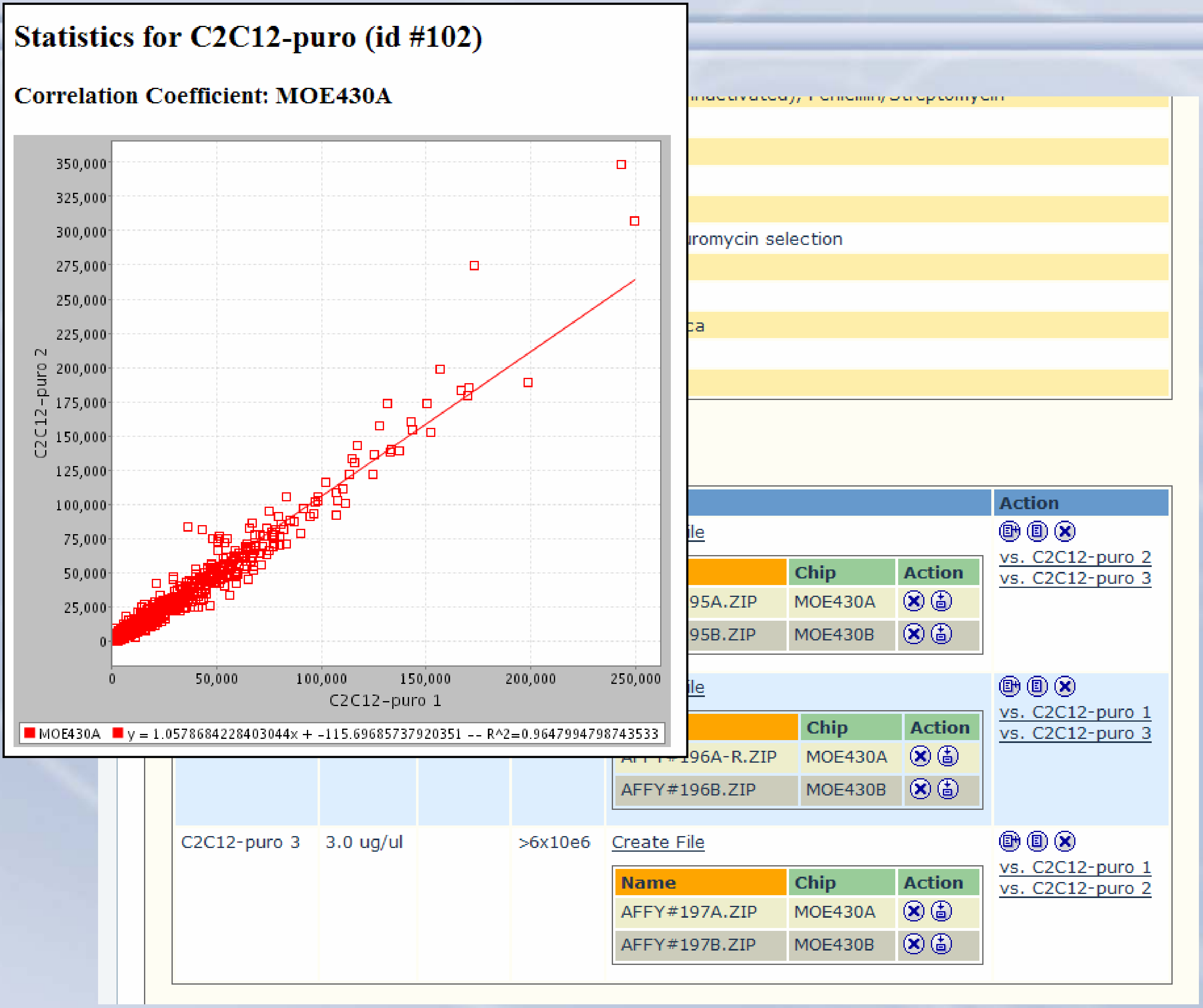
Coefficient of Variation Graph

PROPERTY	VALUE
Name	C2C12-puro
Description	C2C12 myoblas
SCGP ID	SCGP ID# 137
Species	Mouse
Tissue of Origin	Muscle
Cell Type/Line	C2C12
Strain	C3H
Karyotype	Normal, diploid
Growth Medium	DMEM
Media Supplements	10% FCS (hea
Culture Conditions	37 C, 5% CO2
Defining Markers	myoD/myf5
RNA Prep Method	RNeasy
RNA Sample Type	Total RNA
Cell Purification Method	1. Cell line 2. P
Contaminating Cell Type	None
Submitted By	Jeff Ishibashi
Contact Email Address	jishibashi@ohri.ca
Ethics Review Board Approved	true
Estimated Purity	100%

Replicates

Statistics for C2C12-puro (id #102)

Coefficient of Variation: MOE430A



Statistics for C2C12-puro (id #102)

Correlation Coefficient: MOE430A

🟥 MOE430A 🟥 y = 1.0578684228403044x + -115.69685737920351 -- R^2=0.9647994798743533

uromycin selection

ca

	Action
ile (Chip / Action table: 95A.ZIP, MOE430A; 95B.ZIP, MOE430B)	vs. C2C12-puro 2 vs. C2C12-puro 3
ile (Chip / Action table: 196A-R.ZIP, MOE430A; AFFY#196B.ZIP, MOE430B)	vs. C2C12-puro 1 vs. C2C12-puro 3

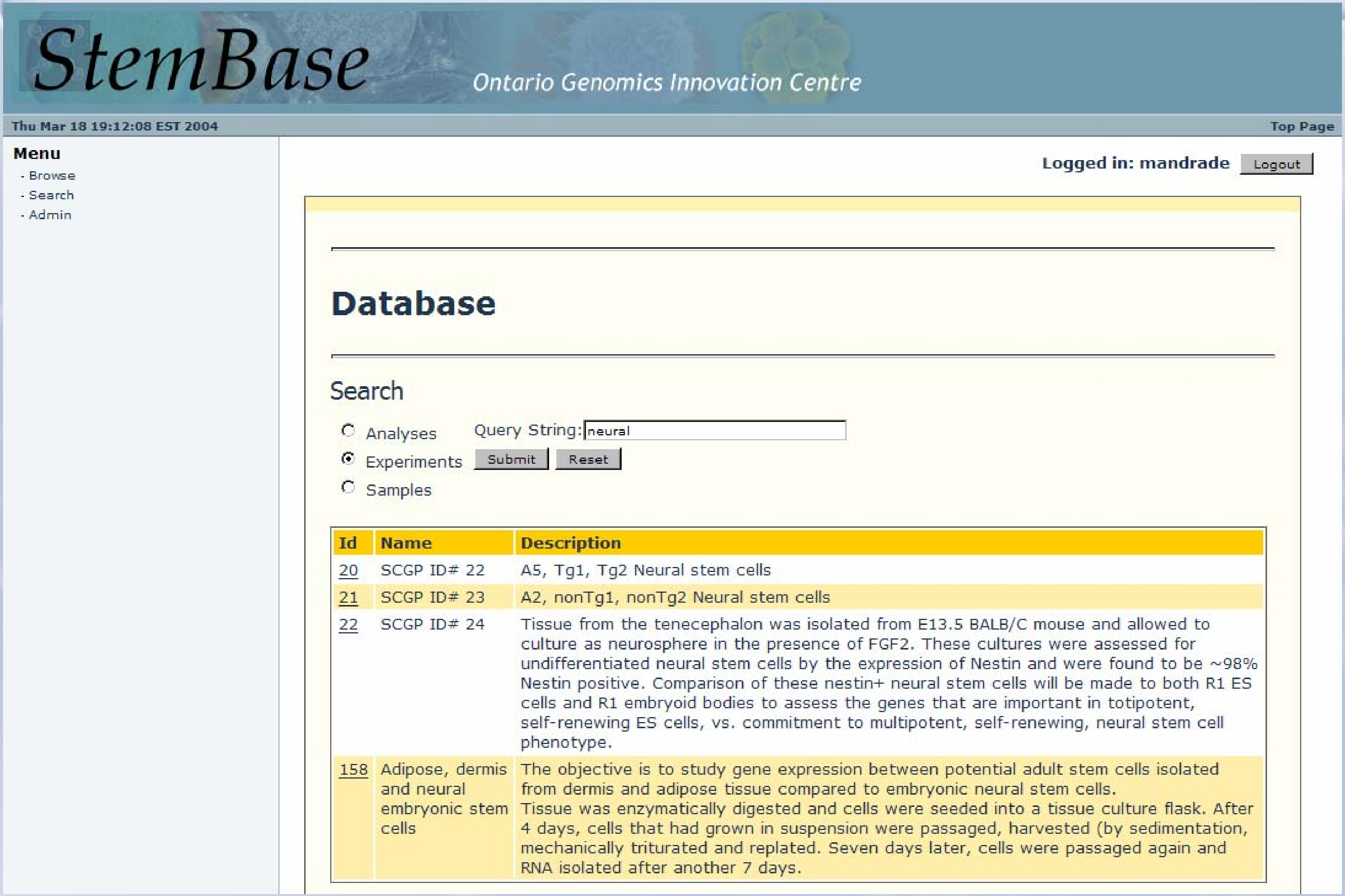
Chip	Action
MOE430A	⊗
MOE430B	⊗

Chip	Action
MOE430A	⊗
AFFY#196B.ZIP MOE430B	⊗

C2C12-puro 3 3.0 ug/ul >6x10e6 Create File

Name	Chip	Action
AFFY#197A.ZIP	MOE430A	⊗
AFFY#197B.ZIP	MOE430B	⊗

vs. C2C12-puro 1
vs. C2C12-puro 2



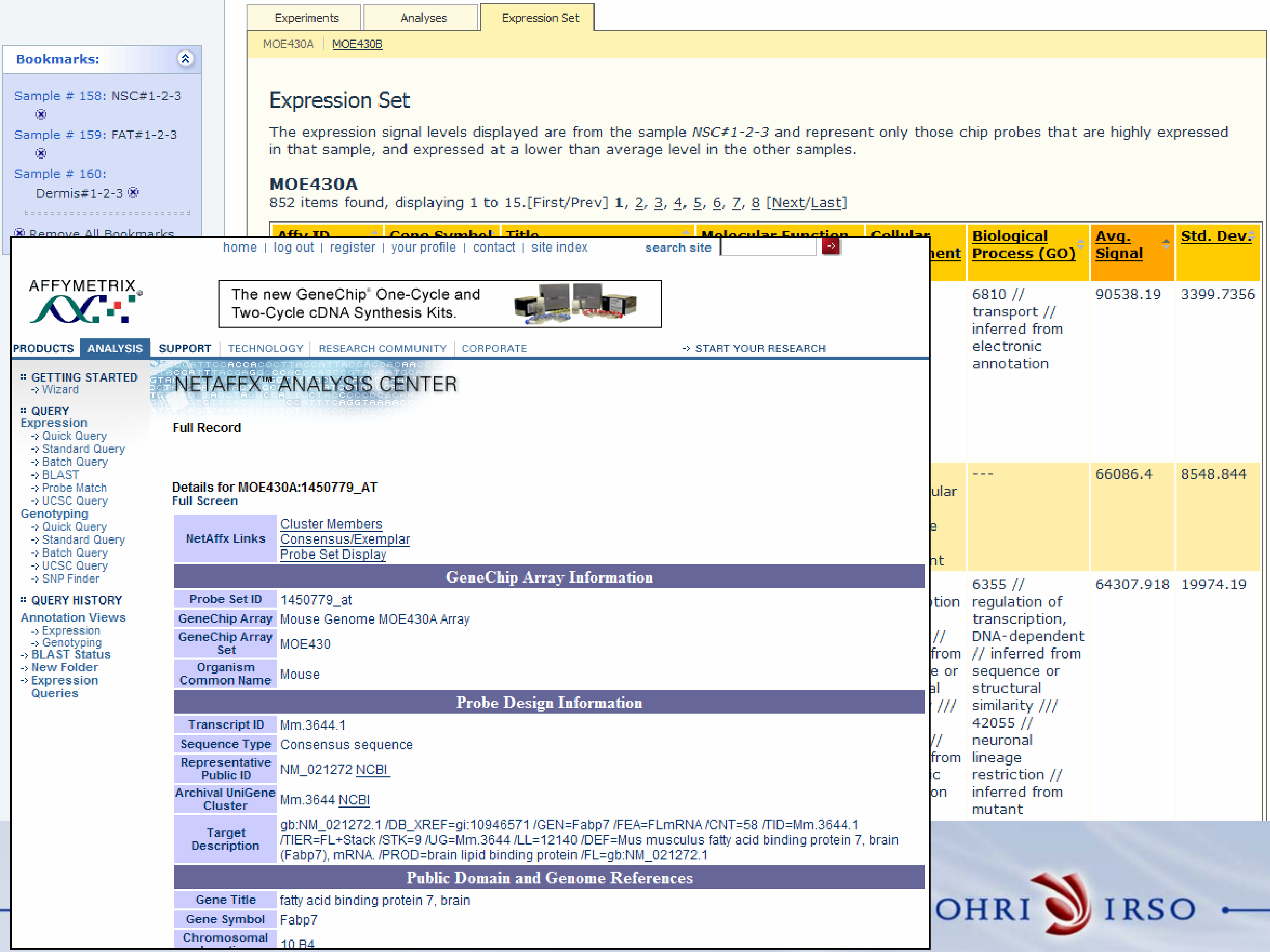
Database

Search

○ Analyses
● Experiments
○ Samples

Query String: neural

Id	Name	Description
20	SCGP ID# 22	A5, Tg1, Tg2 Neural stem cells
21	SCGP ID# 23	A2, nonTg1, nonTg2 Neural stem cells
22	SCGP ID# 24	Tissue from the tenecephalon was isolated from E13.5 BALB/C mouse and allowed to culture as neurosphere in the presence of FGF2. These cultures were assessed for undifferentiated neural stem cells by the expression of Nestin and were found to be ~98% Nestin positive. Comparison of these nestin+ neural stem cells will be made to both R1 ES cells and R1 embryoid bodies to assess the genes that are important in totipotent, self-renewing ES cells, vs. commitment to multipotent, self-renewing, neural stem cell phenotype.
158	Adipose, dermis and neural embryonic stem cells	The objective is to study gene expression between potential adult stem cells isolated from dermis and adipose tissue compared to embryonic neural stem cells. Tissue was enzymatically digested and cells were seeded into a tissue culture flask. After 4 days, cells that had grown in suspension were passaged, harvested (by sedimentation, mechanically triturated and replated. Seven days later, cells were passaged again and RNA isolated after another 7 days.



Bookmarks:

Sample # 158: NSC#1-2-3

Sample # 159: FAT#1-2-3

Sample # 160: Dermis#1-2-3

Remove All Bookmarks

Expression Set

The expression signal levels displayed are from the sample *NSC#1-2-3* and represent only those chip probes that are highly expressed in that sample, and expressed at a lower than average level in the other samples.

MOE430A

852 items found, displaying 1 to 15.[First/Prev] **1**, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Affy ID	Gene Symbol	Title	Molecular Function	Cellular Component	Biological Process (GO)	Avg. Signal	Std. Dev.
					6810 // transport // inferred from electronic annotation	90538.19	3399.7356
				ular e nt	---	66086.4	8548.844
				tion // from e or al /// // from c on	6355 // regulation of transcription, DNA-dependent // inferred from sequence or structural similarity /// 42055 // neuronal lineage restriction // inferred from mutant	64307.918	19974.19

NETAFFX™ ANALYSIS CENTER

Full Record

Details for MOE430A:1450779_AT
Full Screen

NetAffx Links	Cluster Members Consensus/Exemplar Probe Set Display
GeneChip Array Information	
Probe Set ID	1450779_at
GeneChip Array	Mouse Genome MOE430A Array
GeneChip Array Set	MOE430
Organism Common Name	Mouse
Probe Design Information	
Transcript ID	Mm.3644.1
Sequence Type	Consensus sequence
Representative Public ID	NM_021272 NCBI
Archival UniGene Cluster	Mm.3644 NCBI
Target Description	gb:NM_021272.1 /DB_XREF=gi:10946571 /GEN=Fabp7 /FEA=FLmRNA /CNT=58 /TID=Mm.3644.1 /TIER=FL+Stack /STK=9 /UG=Mm.3644 /LL=12140 /DEF=Mus musculus fatty acid binding protein 7, brain (Fabp7), mRNA. /PROD=brain lipid binding protein /FL=gb:NM_021272.1
Public Domain and Genome References	
Gene Title	fatty acid binding protein 7, brain
Gene Symbol	Fabp7
Chromosomal	10 B4

Comparing Samples

Gene in one sample: Present or Absent?

Gene X

		Replicates	
MOE430A	probe	PPP	
	probe	PAP	#P 7 > #A 1
			#P > 50%
MOE430B	probe	MPP	P

Gene in one sample: Present or Absent?

Gene Y

Replicates

MOE430A

probe AMP

probe MMP

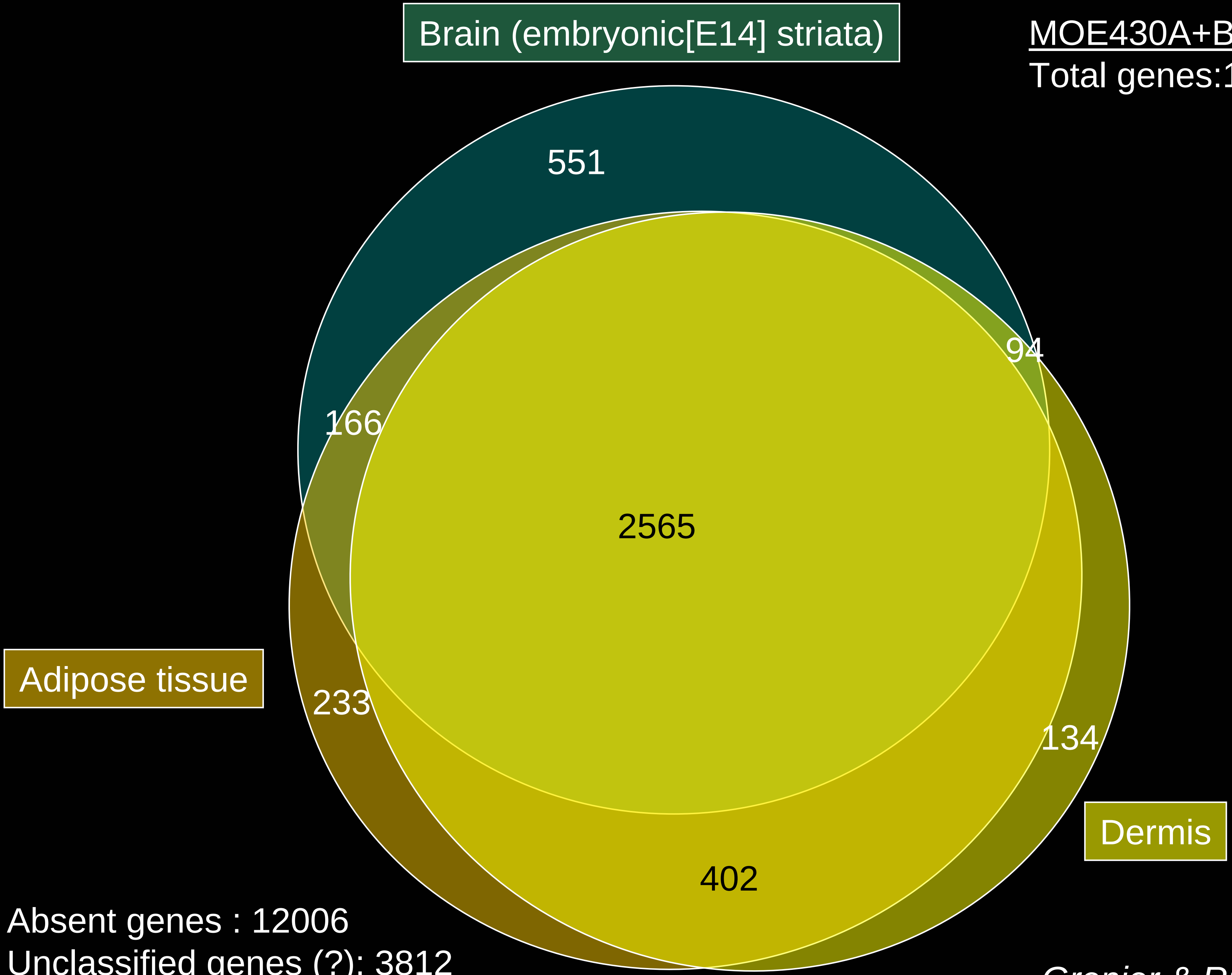
#P 2 > #A 1

#P < 50%

MOE430B

?

MOE430A+B
Total genes:19867



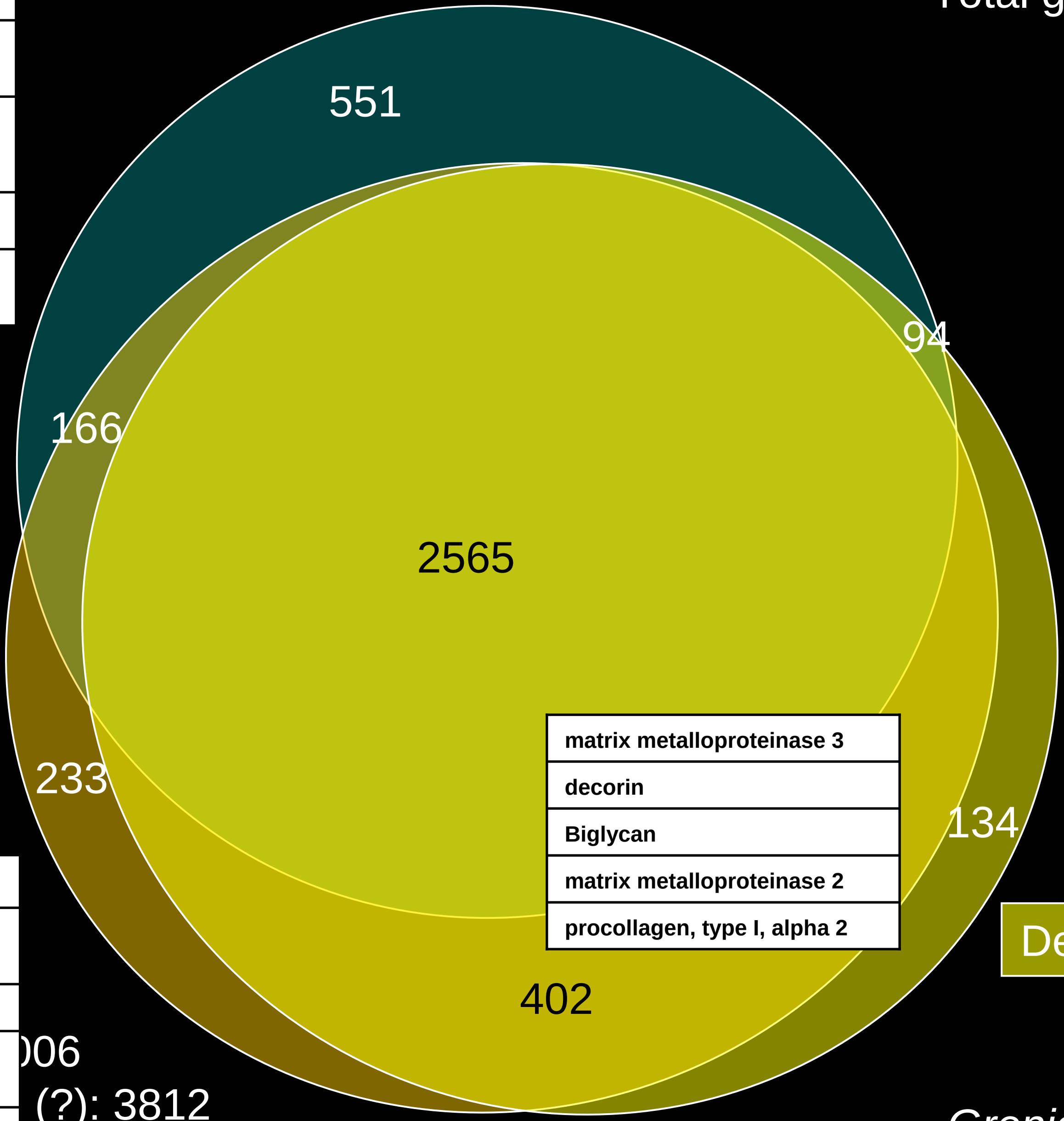
Absent genes : 12006
Unclassified genes (?): 3812

Grenier & Rudnicki

Brain (embryonic[E14] striata)

MOE430A+B
Total genes:19867

- fatty acid binding protein 7, brain
- oligodendrocyte transcription factor 1
- monocyte to macrophage differentiation-associated 2
- myelin basic protein
- discoidin domain receptor family, member 1



- keratin complex 2, basic, gene 6b
- small proline-rich protein 1A
- keratin complex 2, basic, gene 6a
- keratin complex 1, acidic, gene 14
- gi:12843543

- matrix metalloproteinase 3
- decorin
- Biglycan
- matrix metalloproteinase 2
- procollagen, type I, alpha 2

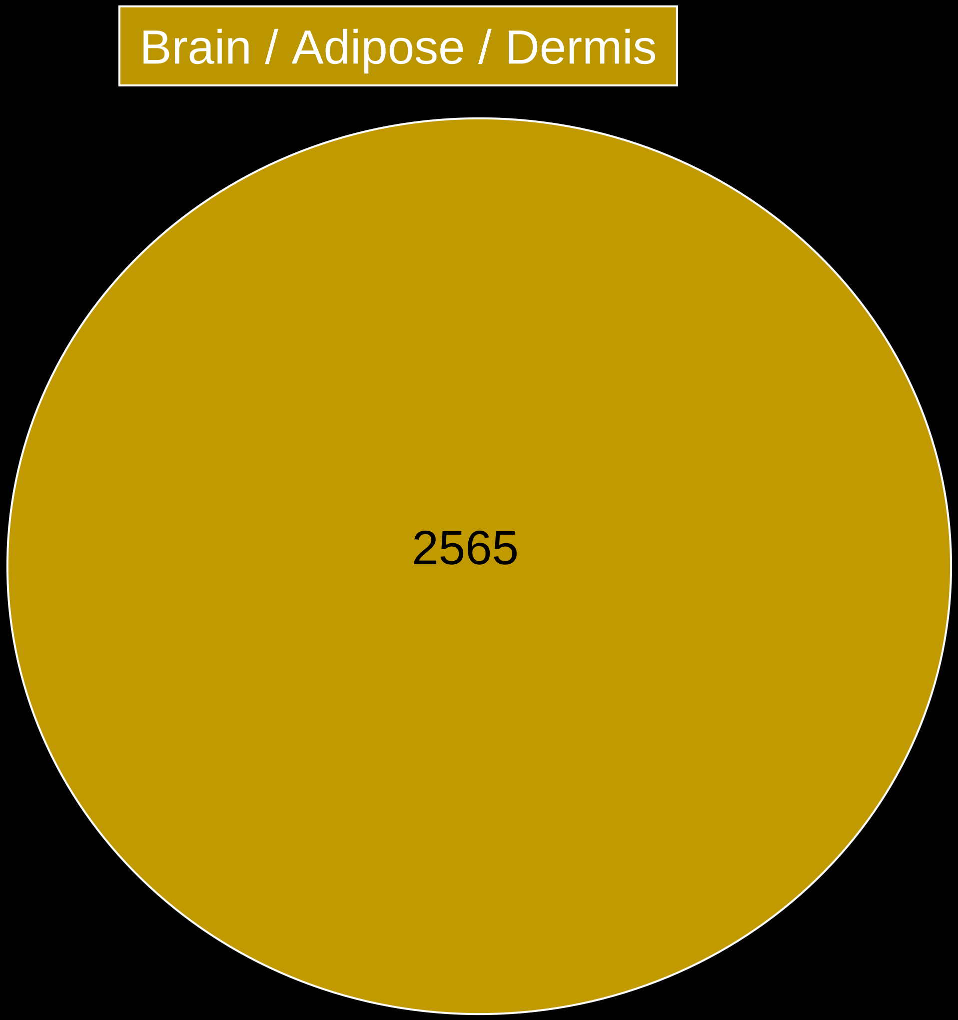
Adipose tissue

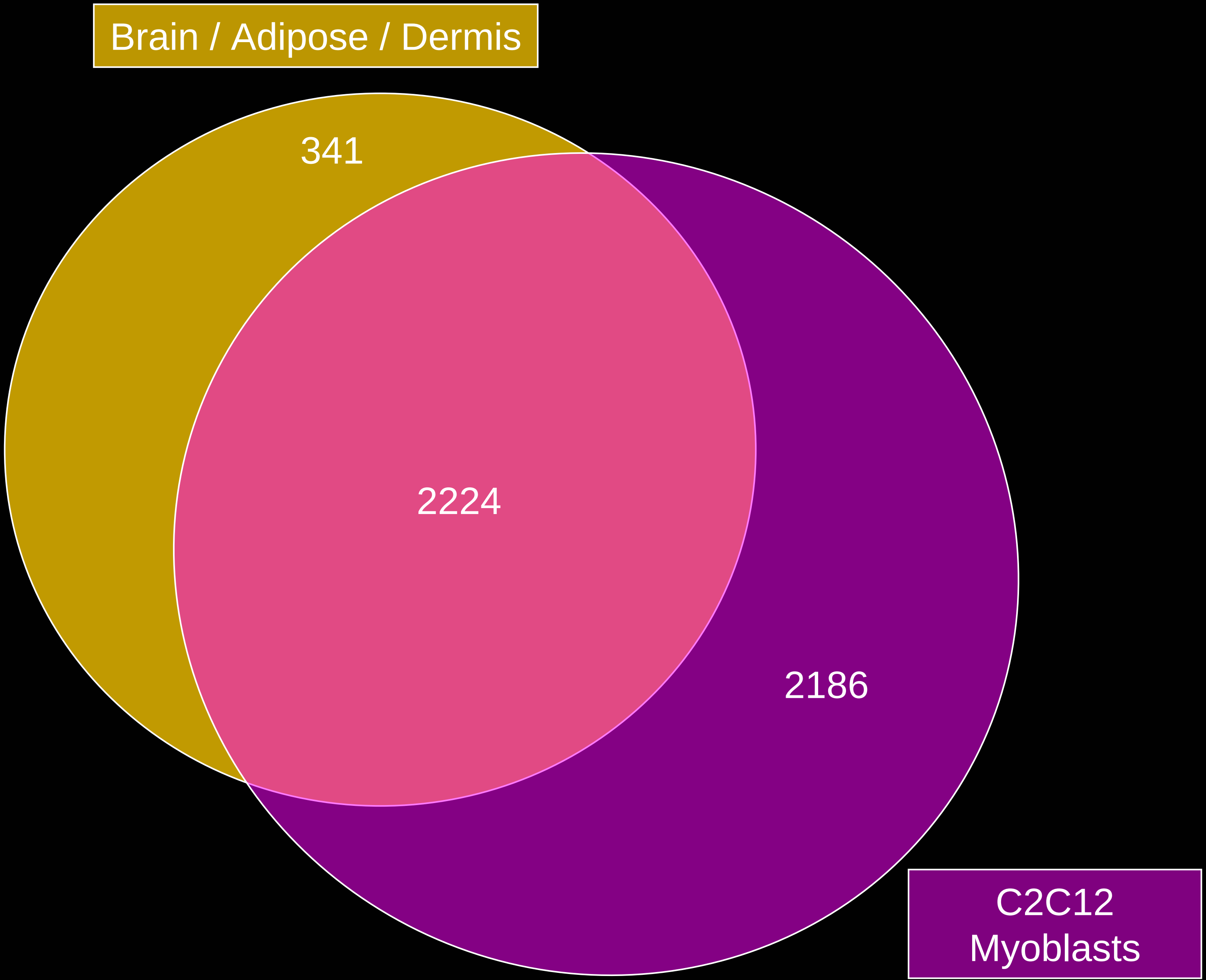
Dermis

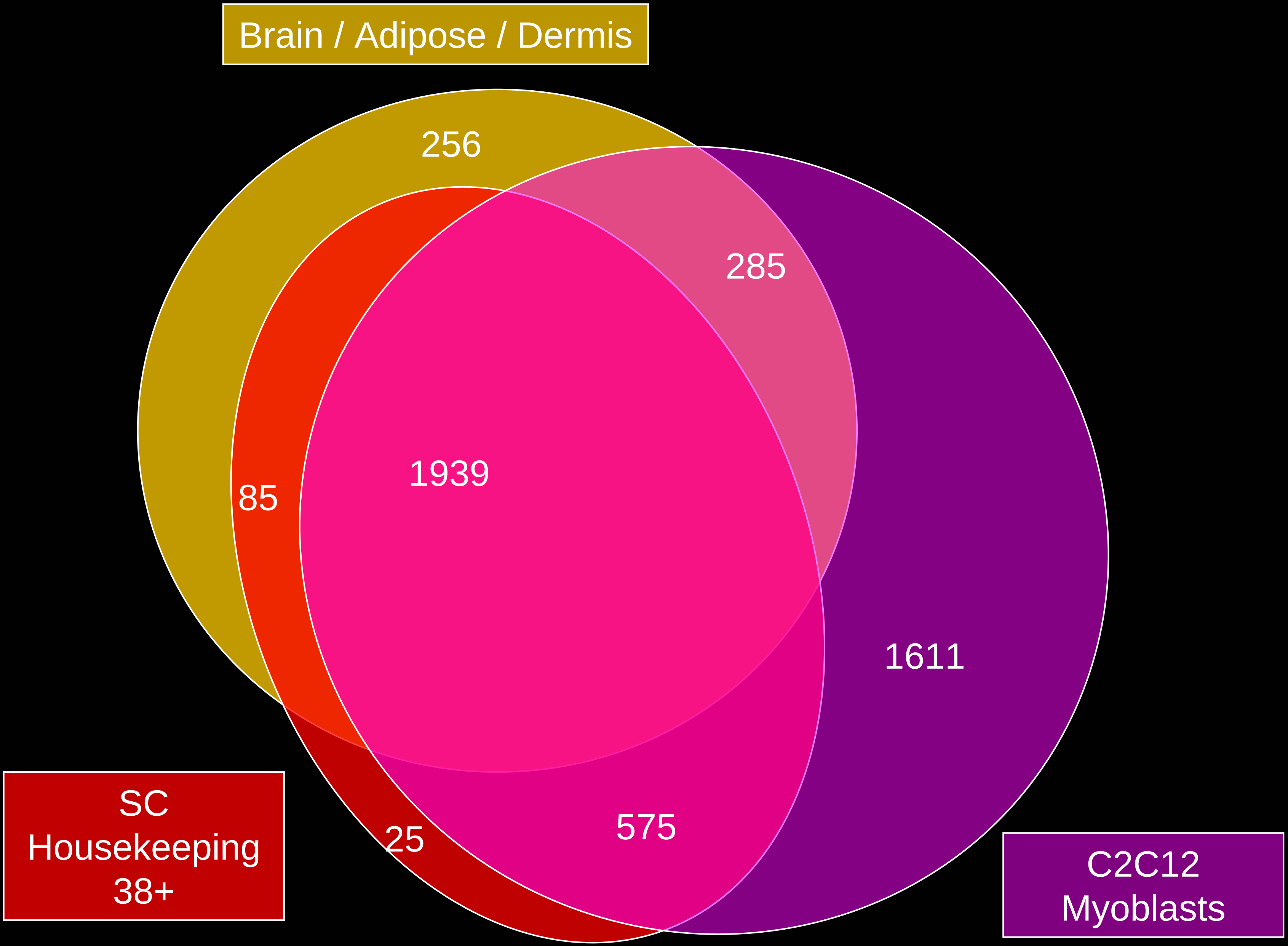
- lysozyme
- chemokine (C-C motif) ligand 6
- gi:5216352
- lysosomal-associated protein transmembrane 5
- glycoprotein 49 B

Abs 006
Un (?): 3812

Grenier & Rudnicki


Brain / Adipose / Dermis
2565


Brain / Adipose / Dermis
341
2224
2186
C2C12
Myoblasts


Brain / Adipose / Dermis
256
285
85
1939
1611
SC
Housekeeping
38+
25
575
C2C12
Myoblasts

SC Housekeeping 38+	-
Brain / Adipose / Dermis	+
C2C12 Myoblasts	-

256

Kdr: kinase domain receptor, also known as VEGF (Vascular endothelial growth factor) receptor-2

Egf: epidermal growth factor

Ephb1: tyrosine kinase receptor

Gapd: glyceraldehyde 3-phosphate dehydrogenase?

Time Series

Time series analysis: differentiation of mouse V6.5 ES

Biggs & Rudnicki

Time series: observations are not independent.

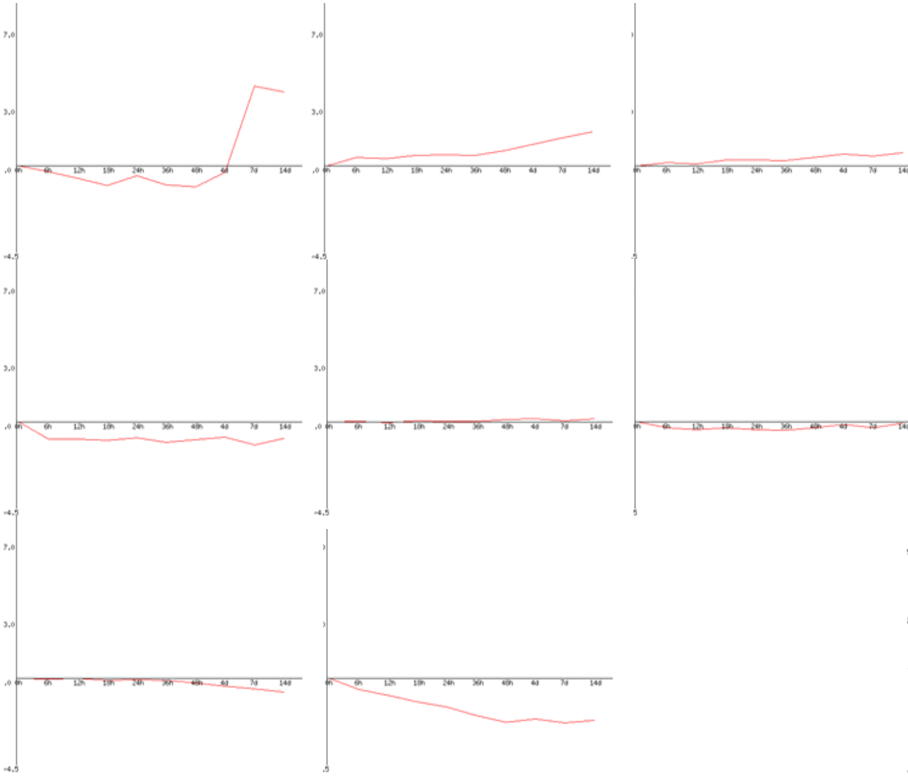
Bayesian clustering by dynamics (BCD)

Ramoni et al., 2002

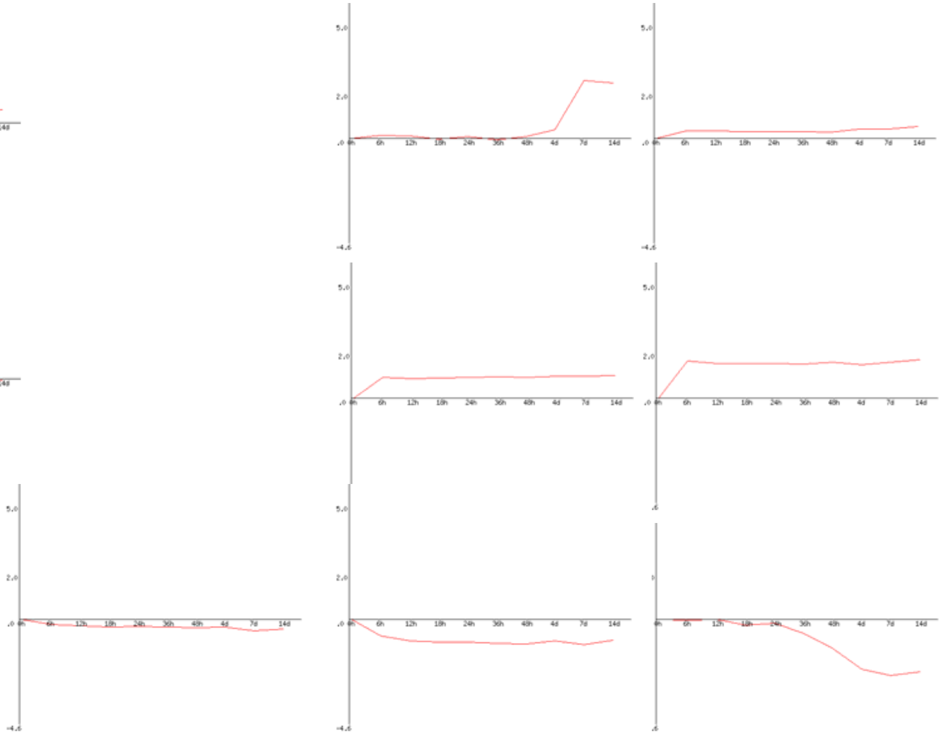
- Logs of expression ratio to the starting point
- “Stemness” genes would be increasingly down-regulated

Time series analysis: differentiation of mouse V6.5 ES

Biggs & Rudnicki

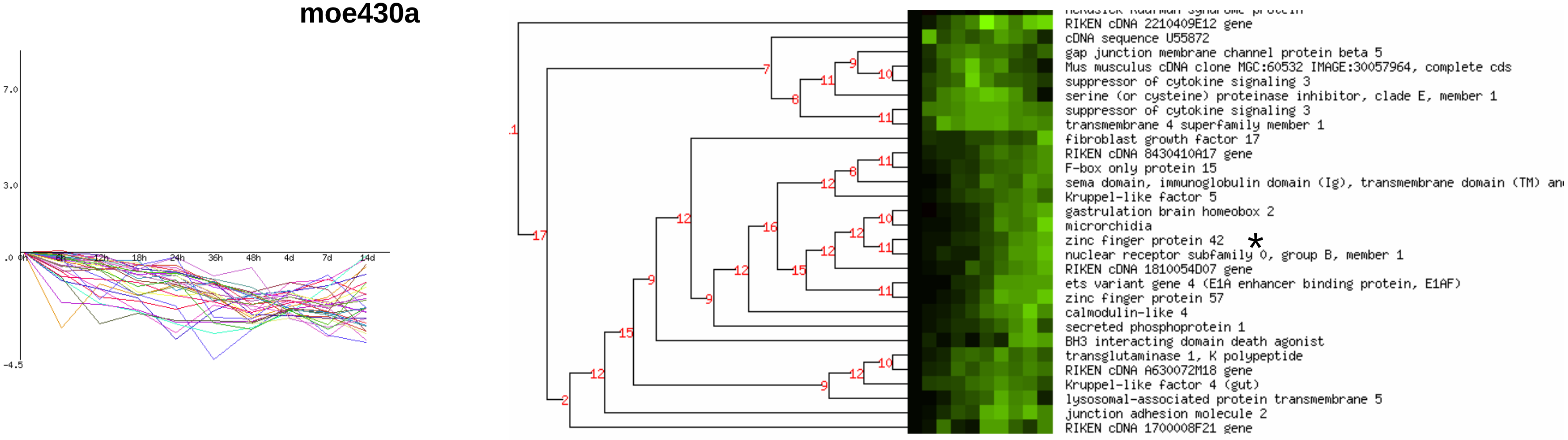


moe430a

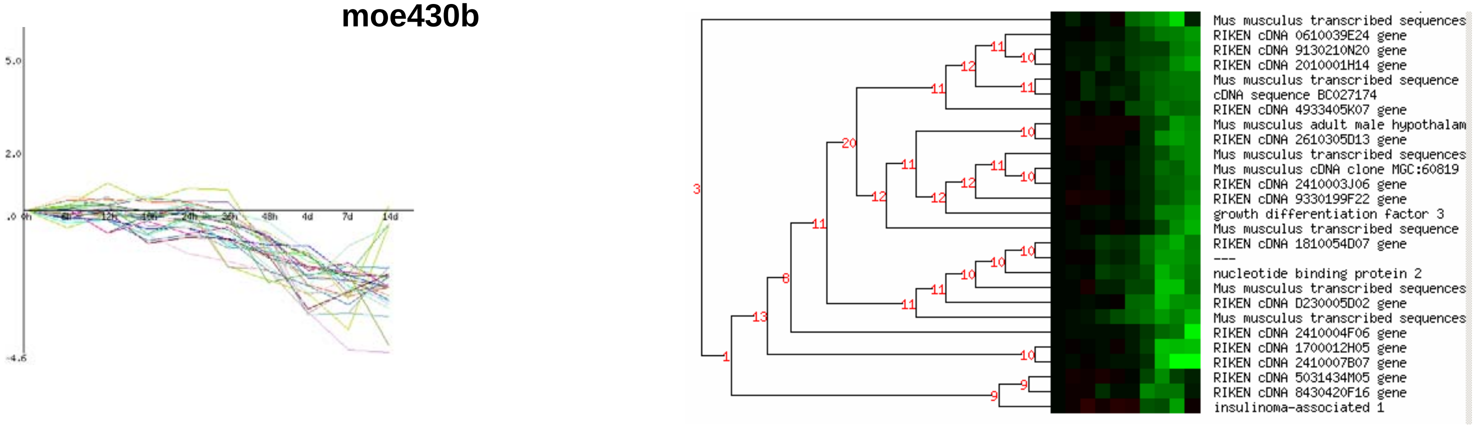


moe430b

moe430a

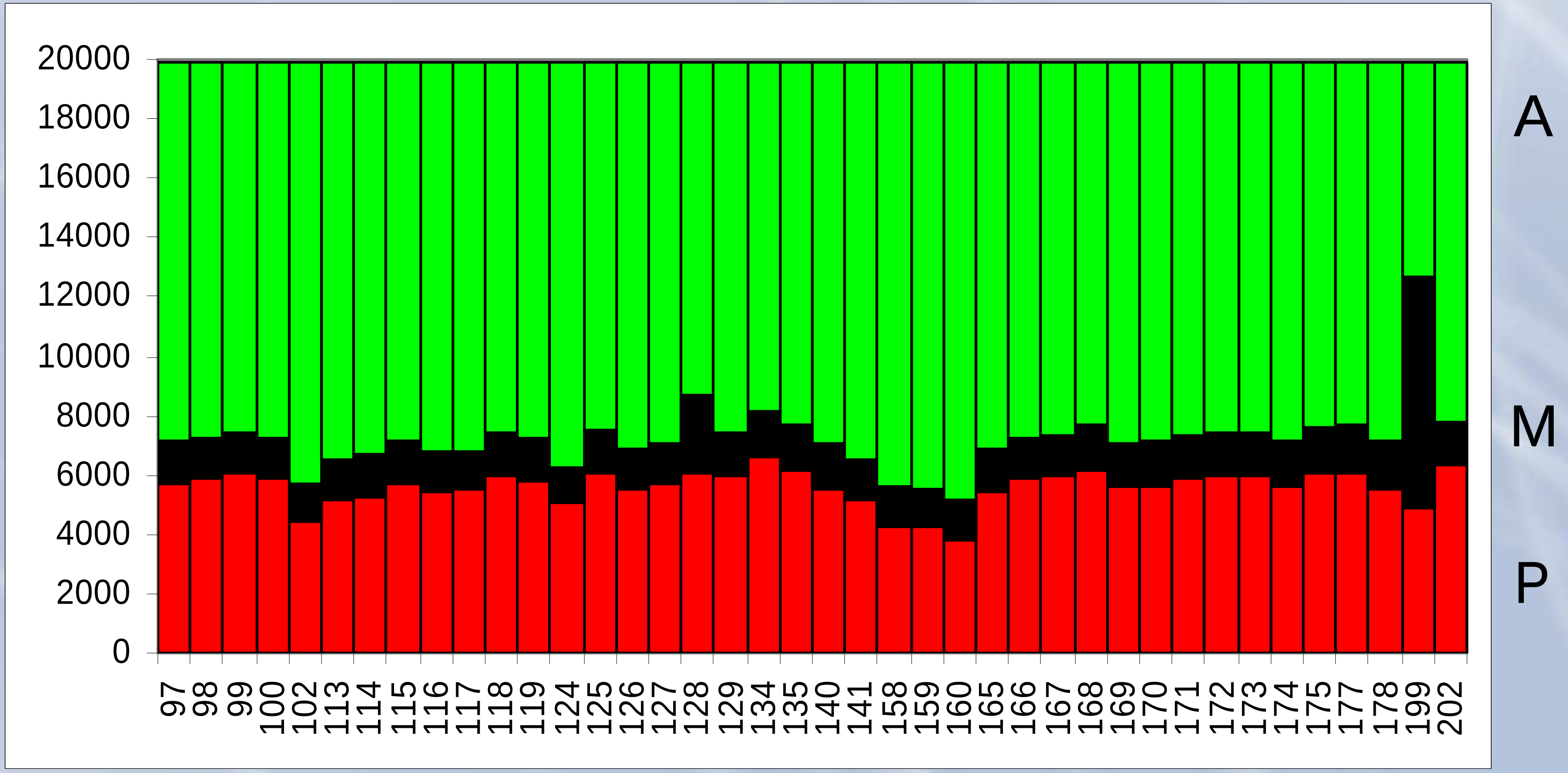


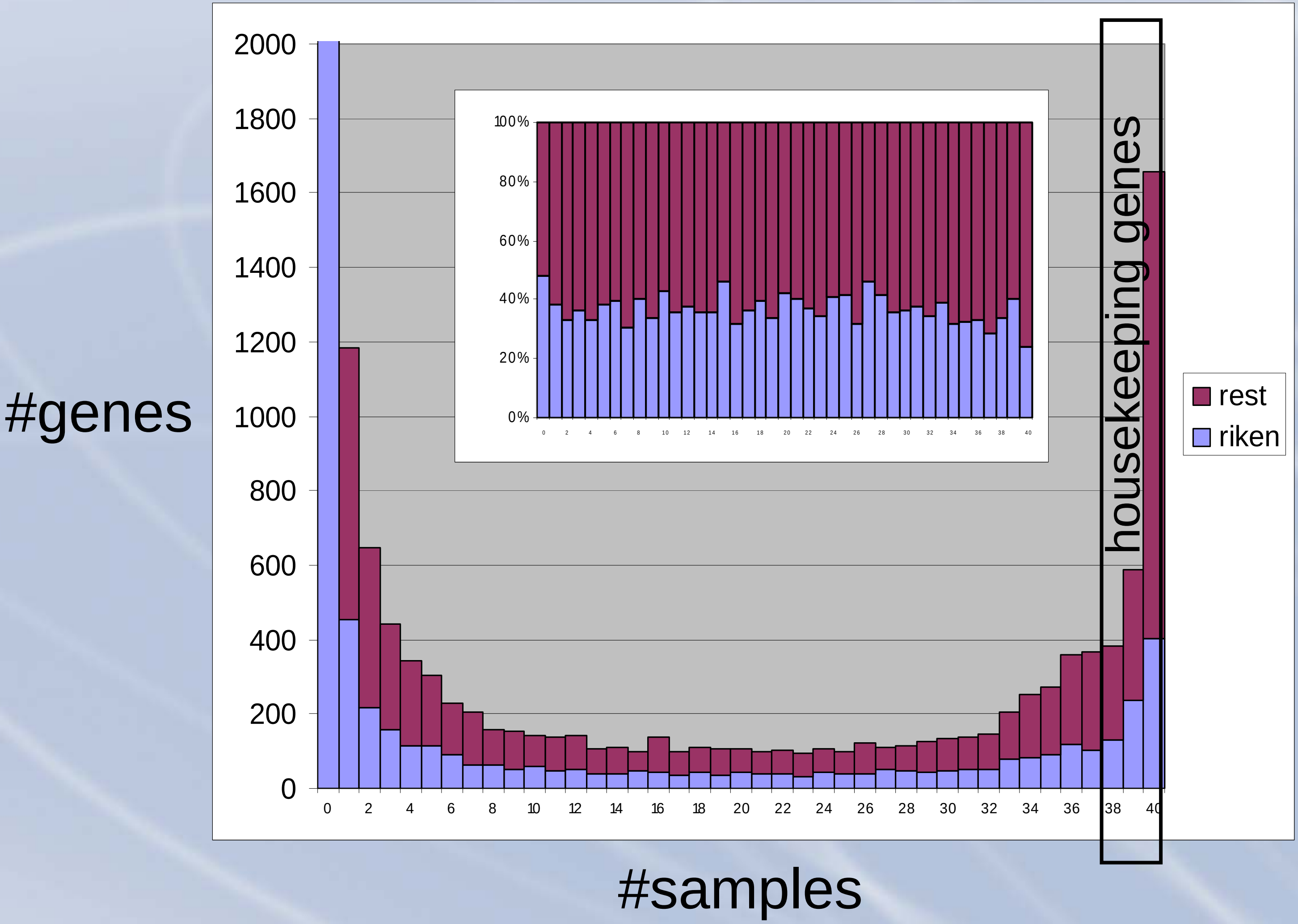
moe430b

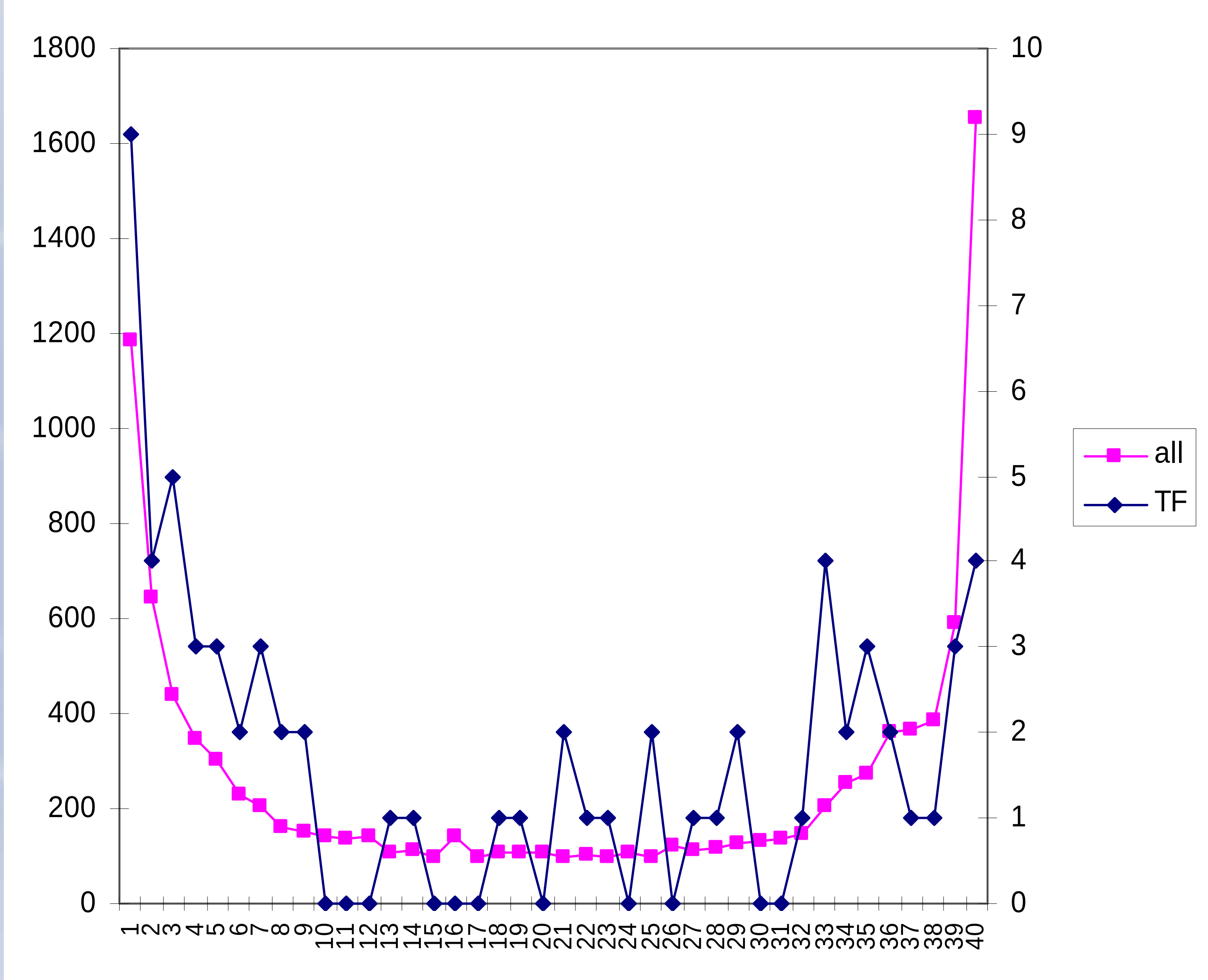


Rex-1 (Zfp42) known to be expressed at high levels in embryonic stem (ES) and F9 teratocarcinoma cells.

Looking at everything


20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
97
98
99
100
102
113
114
115
116
117
118
119
124
125
126
127
128
129
134
135
140
141
158
159
160
165
166
167
168
169
170
171
172
173
174
175
177
178
199
202
A
M
P

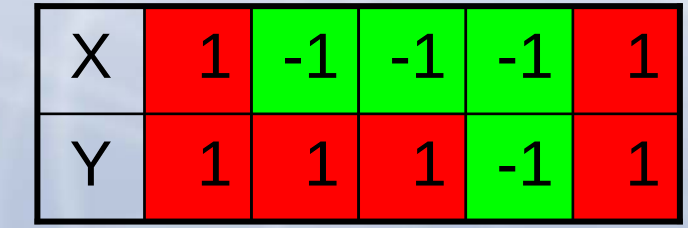

#genes
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
#samples
100%
80%
60%
40%
20%
0%
housekeeping genes
rest
riken


1800
1600
1400
1200
1000
800
600
400
200
0
10
9
8
7
6
5
4
3
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
all
TF

Genetic Networks

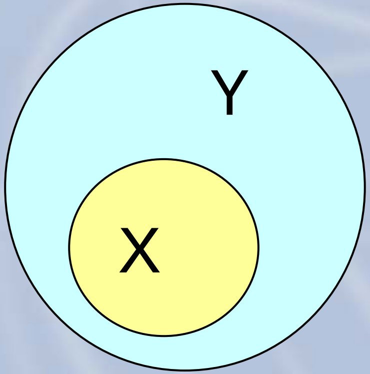
Gene expression values

samples

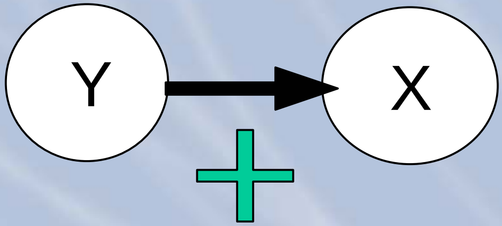


X	1	-1	-1	-1	1
Y	1	1	1	-1	1

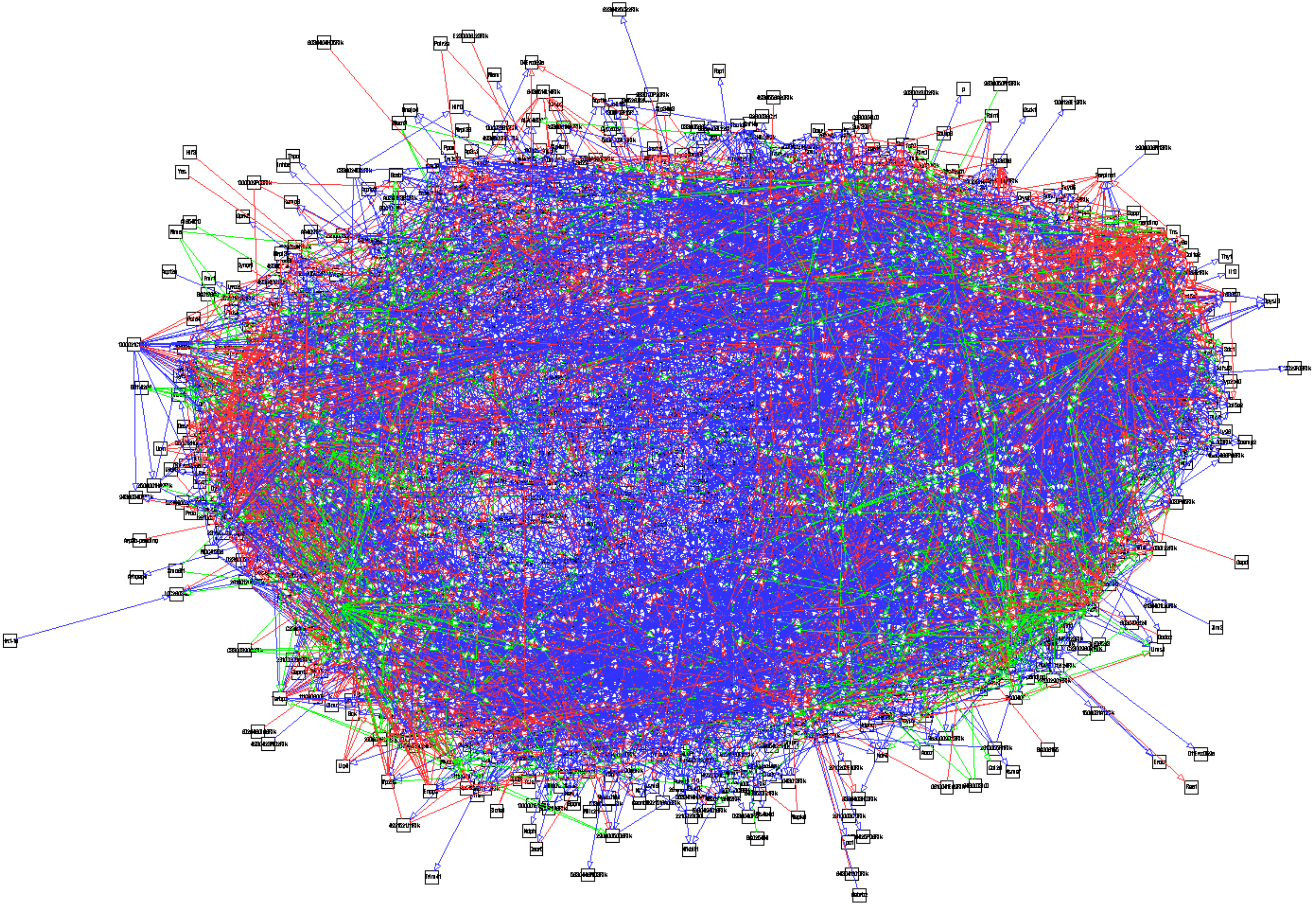
association values



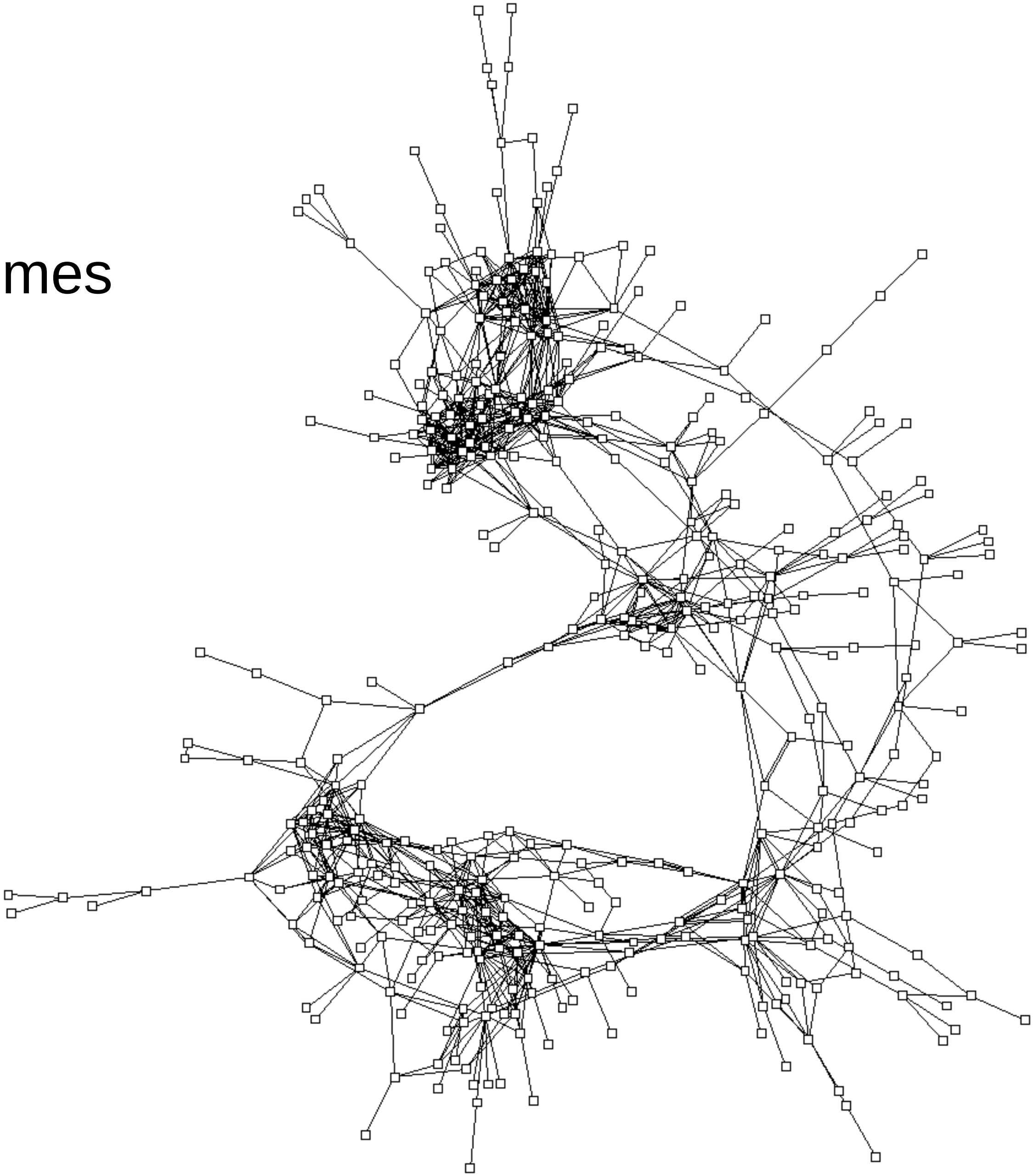
part of a network

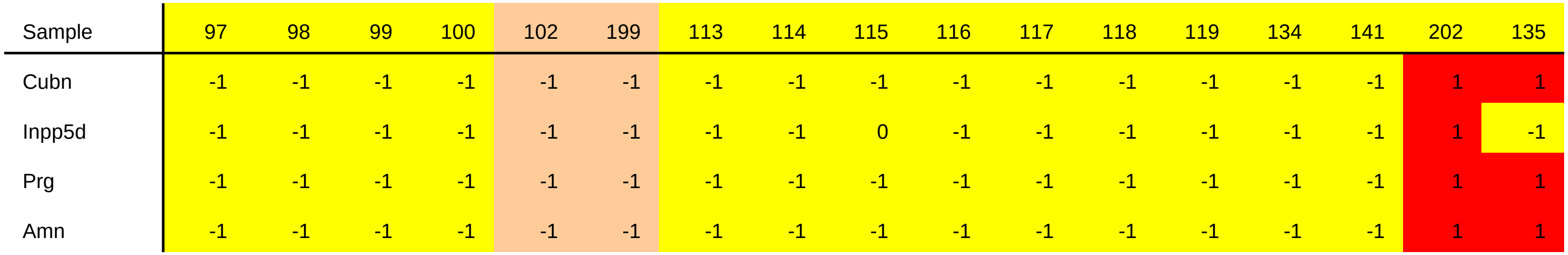


Perez-Iratxeta et al., (2001)

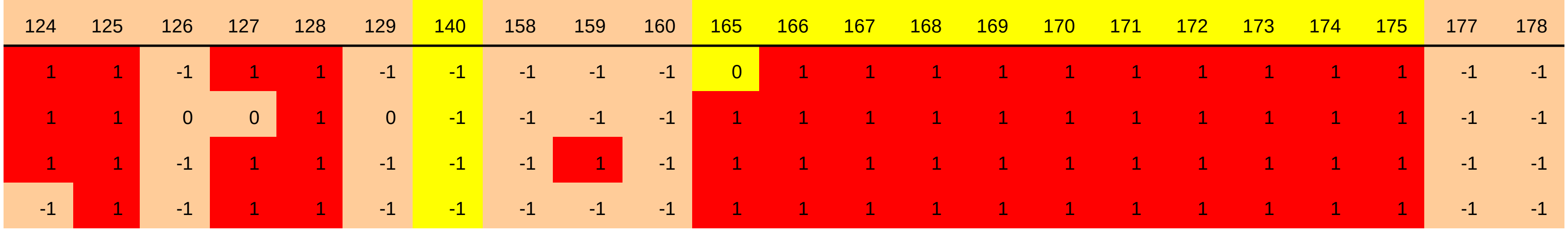


15-20 times





Sample	97	98	99	100	102	199	113	114	115	116	117	118	119	134	141	202	135
Cubn	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
Inpp5d	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	1	-1
Prg	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
Amn	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1



Sample	124	125	126	127	128	129	140	158	159	160	165	166	167	168	169	170	171	172	173	174	175	177	178
Cubn	1	1	-1	1	1	-1	-1	-1	-1	-1	0	1	1	1	1	1	1	1	1	1	1	-1	-1
Inpp5d	1	1	0	0	1	0	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	-1	-1
Prg	1	1	-1	1	1	-1	-1	-1	1	-1	1	1	1	1	1	1	1	1	1	1	1	-1	-1
Amn	-1	1	-1	1	1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	-1	-1

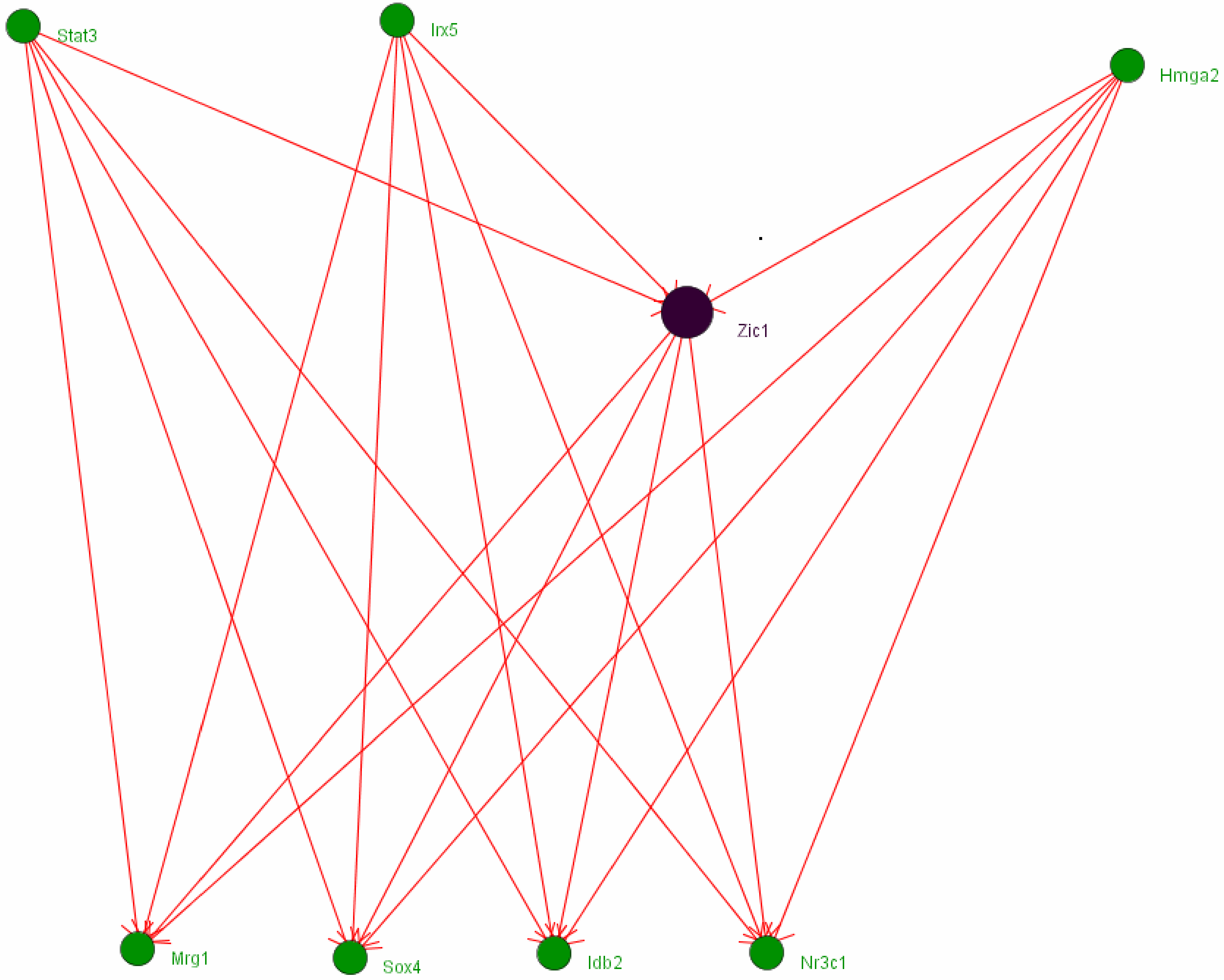
Cubn. intrinsic factor-vitamin B12 receptor (cubilin)

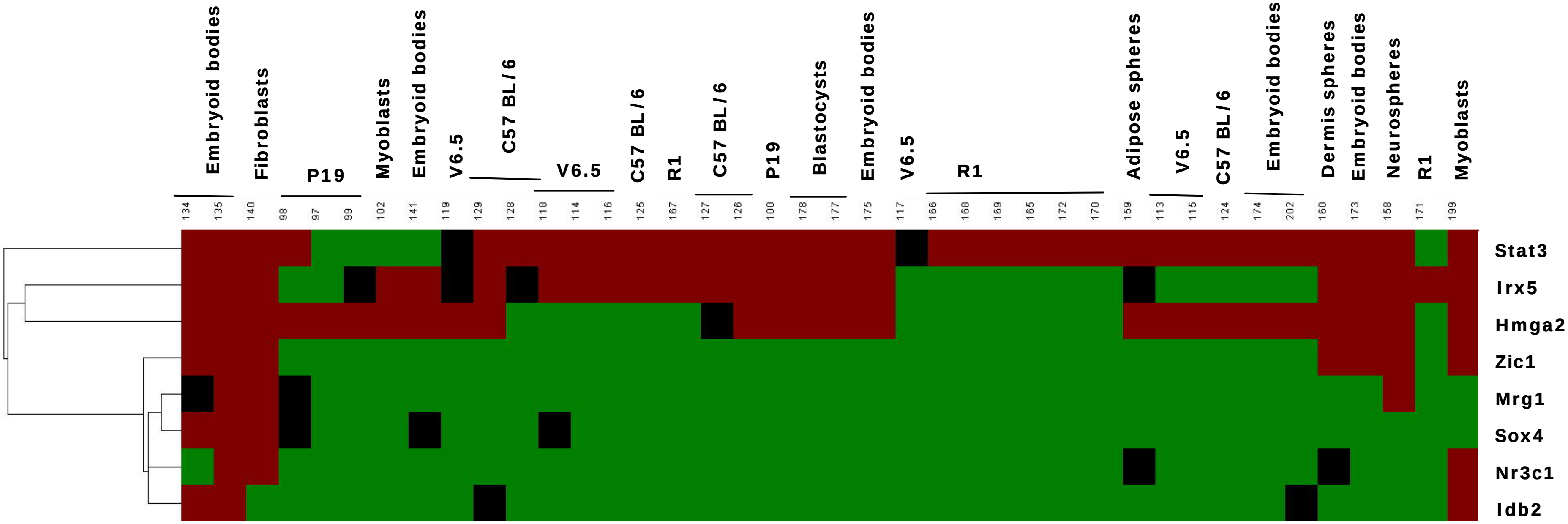
Inpp5d. Aka **SHIP** 145-kDa hemopoietic-restricted SH2-containing inositol 5'- phosphatase / role in cytokine-induced signaling

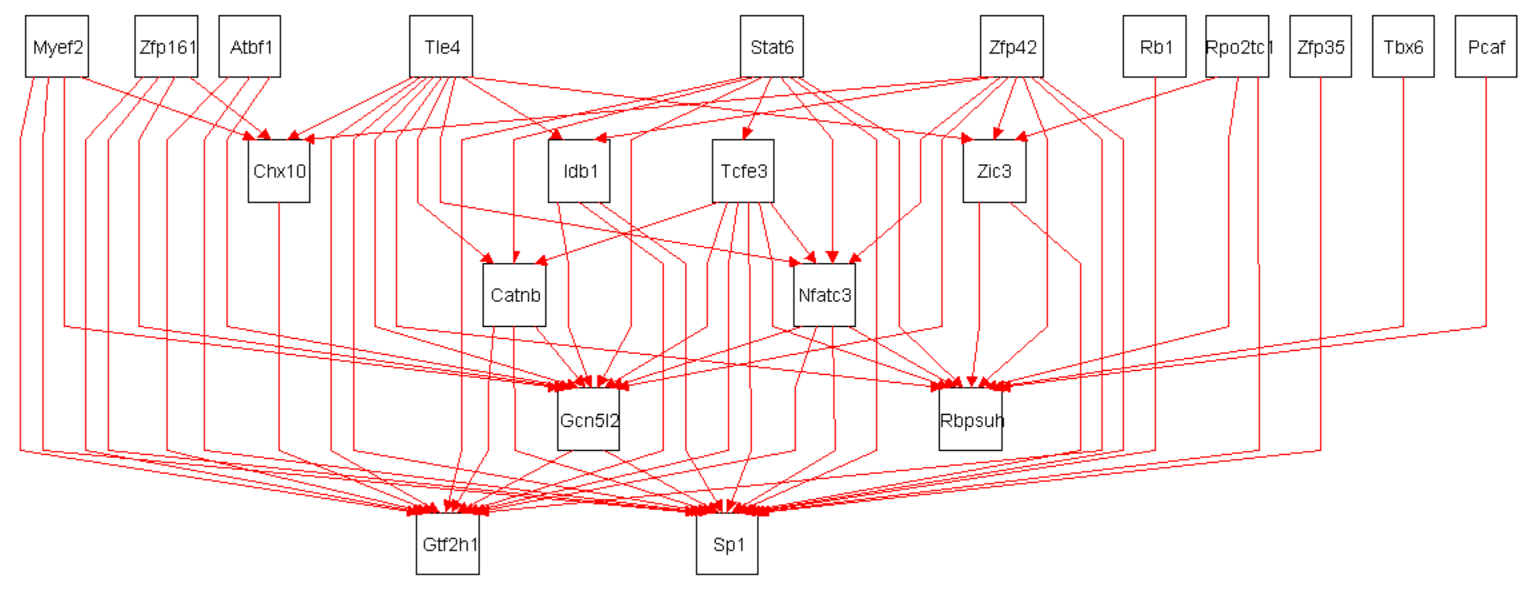
Prg. membrane-associated lipid phosphate phosphatase, plasticity-related gene 1

Amn. The amnionless gene encodes a type I transmembrane protein that is expressed in the extraembryonic visceral layer during gastrulation

March ’04 / The functional cobalamin (vitamin B12)-intrinsic factor receptor is a novel complex of **cubilin** and **amnionless**.


Stat3
Irx5
Hmga2
Zic1
Mrg1
Sox4
Idb2
Nr3c1


Embryoid bodies
Fibroblasts
P19
Myoblasts
Embryoid bodies
V6.5
C57 BL/6
V6.5
C57 BL/6
R1
C57 BL/6
P19
Blastocysts
Embryoid bodies
V6.5
R1
Adipose spheres
V6.5
C57 BL/6
Embryoid bodies
Dermis spheres
Embryoid bodies
Neurospheres
R1
Myoblasts
134
135
140
98
97
99
102
141
119
129
128
118
114
116
125
167
127
126
100
178
177
175
117
166
168
169
165
172
170
159
113
115
124
174
202
160
173
158
171
199
Stat3
Irx5
Hmga2
Zic1
Mrg1
Sox4
Nr3c1
Idb2



25-39 only TFs

Complementing the data

SAGE analysis status (March ’04)

SAGE	MICROARRAY	CONTRIBUTOR LABORATORY
ES R1 embryonic stem cells. Mouse	timepoint 0 for the R1 differentiation timecourse **Exp ID 165 Sample ID 165**	Michael Rudnicki
EB R1 differentiating into Embryoid Bodies (9d). Mouse	timepoint 9d for the R1 differentiation timecourse **Exp ID 165 Sample ID 174**	Michael Rudnicki
TS R1 differentiated into Trophectoderm stem cells. Mouse	GFP-Exe Trophoblast stem cells (WT+GFP transgene) Tissue Extraembryonic ectoderm **Exp ID 11 Sample ID 14**	Janet Rossant
NS Neural Stem Cells. Embyonic (13.5d). Mouse	**Exp ID 22 Sample ID 23**	Luc Sabourin

SAGE analysis status (March ’04)

On tags appearing at least (10-15%)

	all	unidentified
Found only in ES:	2756	567
Found only in EB:	2287	405
Found only in TS:	588	118
Found only in NS:	1463	268
Found in ES, TS and NS, but not in EB:	182	8

Next steps

Integration of SAGE and proteomics data

Data mining

- Stem Cells (MeSH term and 16 children)
- 101,338 papers from +11,000,000 in MEDLINE

Use of GO terms

Development of a Cell Ontology

- Import microarray data
- Point to samples needed

Identify SC bibliography / markers

Microarray data normalization / benchmarking

Guillaume Grenier
Nancy Biggs
Jeff Ishibashi
Michael Rudnicki

Lynn Megeney

Doug Borris
Karine Fayad
Julie Fortin
Kathy Sheikheleslamy
Caroline Vergette
Junnan Xia
Pearl Campbell

Stem Cell Network

Enrique Muro
Gareth Palidwor
Christopher Porter
Carolina Perez-Iratxeta
Neal Sanche
Brian Suomela
Miguel Andrade

Ottawa Health Research Institute
OHRI IRSO
Institut de recherche en santé d'Ottawa

OntarioGenomicsInstitute

Stem Cell
Network Réseau de
cellules souches

Canada Foundation for Innovation
Fondation canadienne pour l'innovation

Ontario
Innovation
Trust

Ontario
Ontario Research
and Development
Challenge Fund
Putting R&D to Work

AN INSTITUTE OF • UN INSTITUT DE

THE OTTAWA HOSPITAL
L'HÔPITAL D'OTTAWA

Université d'Ottawa
University of Ottawa