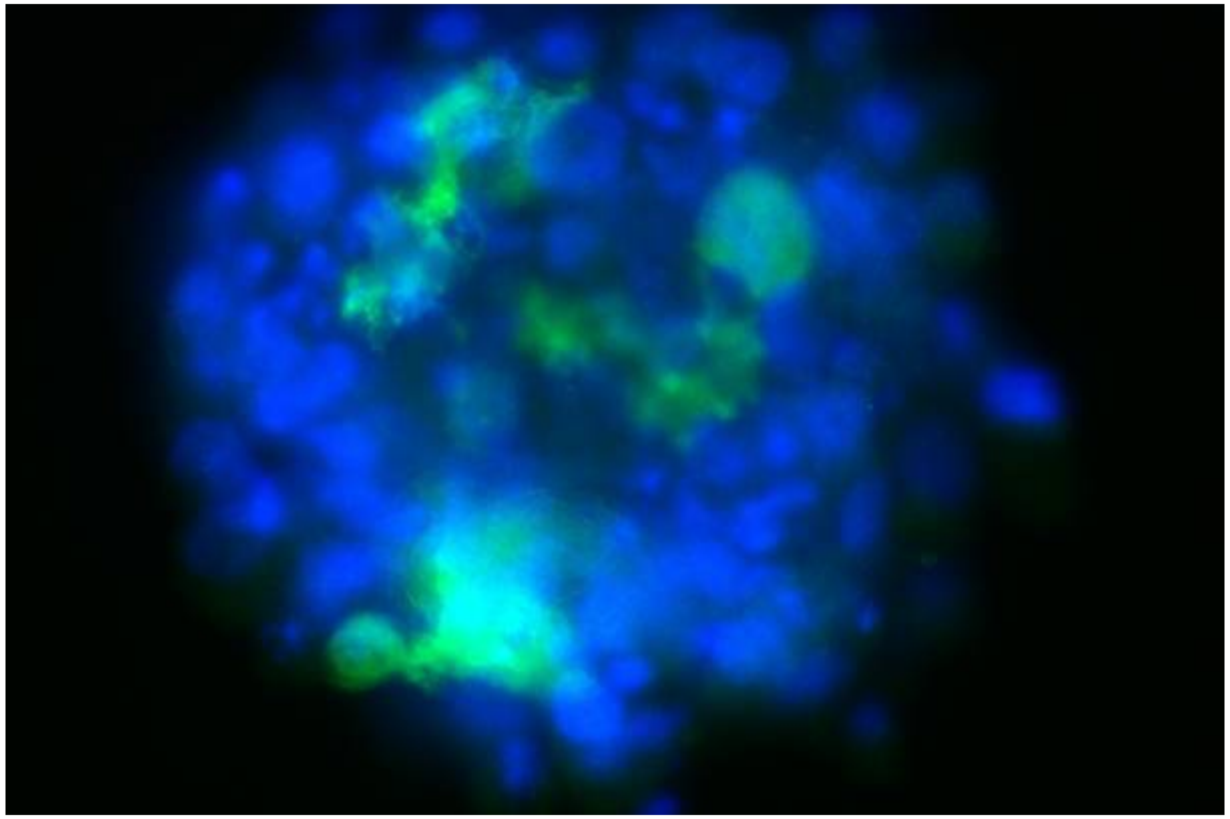


HARVARD
School of Public Health

HSCI
HARVARD STEM CELL
INSTITUTE

The Data Sharing Challenge: Learning Community From Consortia

Winston Hide

National Institute of Genetics International Symposium
Tokyo. 11 October 2010

Relationships, Data Sources & Standards

Genomic Information

Molecular & Cellular Phenotype

Clinical Phenotype

Molecules

Alleles

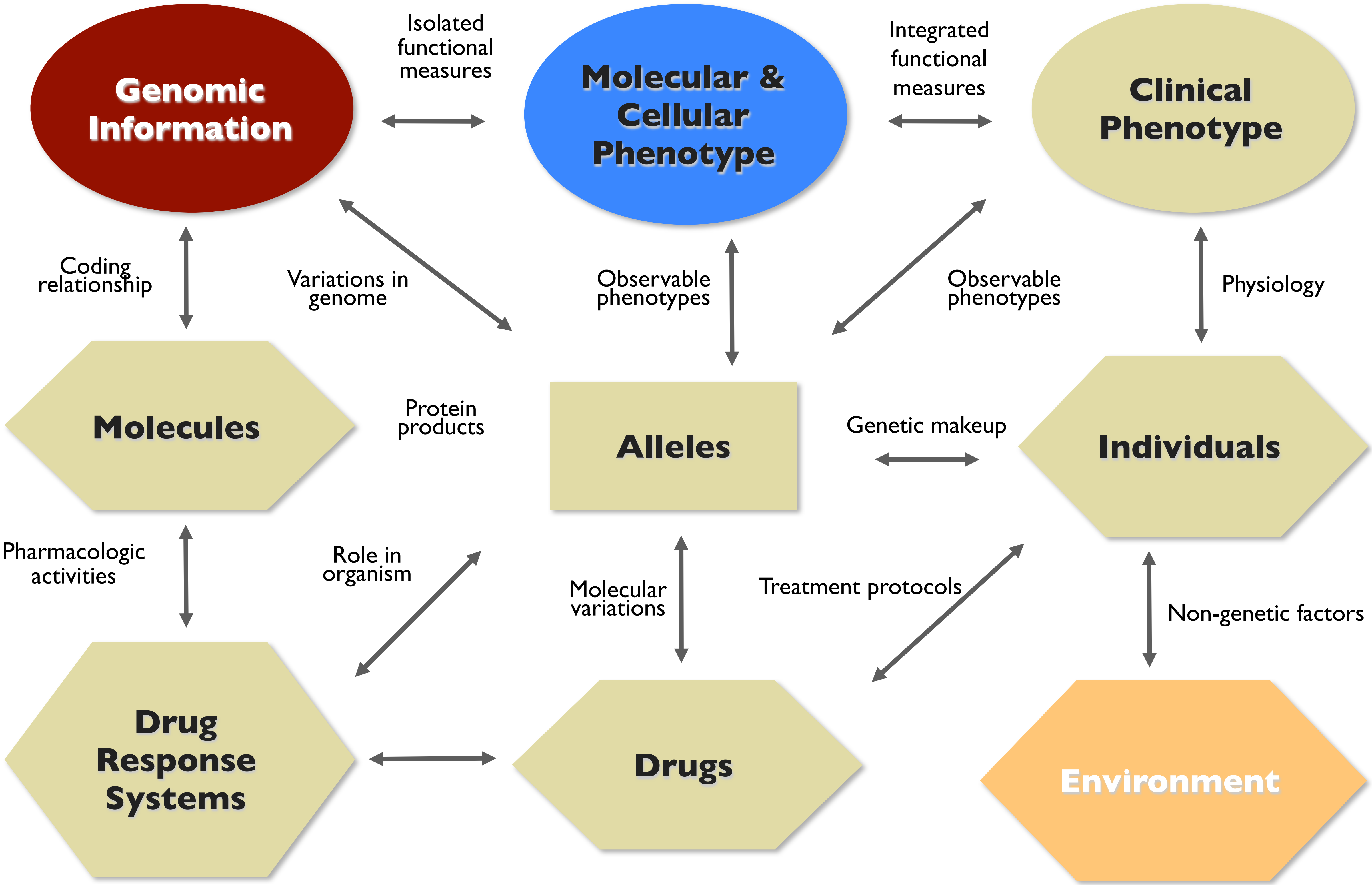
Individuals

Drug Response Systems

Drugs

Environment

Relationships, Data Sources & Standards



Genomic Information

Isolated functional measures

Molecular & Cellular Phenotype

Integrated functional measures

Clinical Phenotype

Coding relationship

Variations in genome

Observable phenotypes

Observable phenotypes

Physiology

Molecules

Protein products

Alleles

Genetic makeup

Individuals

Pharmacologic activities

Role in organism

Molecular variations

Treatment protocols

Non-genetic factors

Drug Response Systems

Drugs

Environment

Open

> rapid research advance

> power of context

> the contribution of the collective

Standards

> value of a technology can be measured by its

  . relative impact

  . global impact

> comparison and reproducibility of results

Peering

The transformation of hierarchy

> a horizontal organization to create information based services

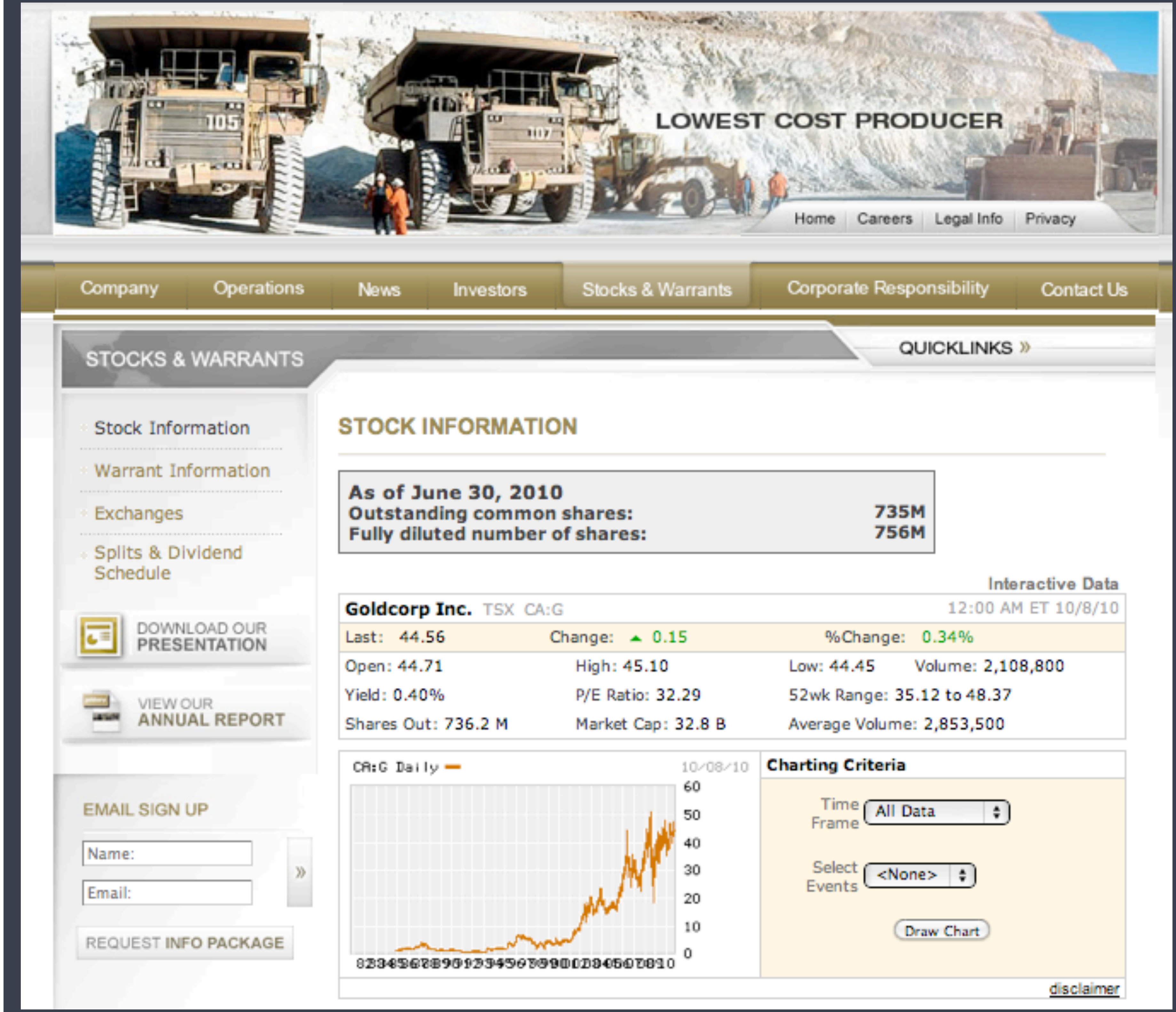
Sharing

Control of data and ideas in a competitive research environment

> sharing of ideas and data

> pharmaceutical investment

- SNP consortium
- Human Genome Project Sanger/NIH



STOCK INFORMATION

As of June 30, 2010
Outstanding common shares: 735M
Fully diluted number of shares: 756M

Interactive Data

Goldcorp Inc. TSX CA:G 12:00 AM ET 10/8/10

Last: 44.56 Change: ▲ 0.15 %Change: 0.34%

Open: 44.71 High: 45.10 Low: 44.45 Volume: 2,108,800

Yield: 0.40% P/E Ratio: 32.29 52wk Range: 35.12 to 48.37

Shares Out: 736.2 M Market Cap: 32.8 B Average Volume: 2,853,500

Gold Corp.

$575 000

80% new gold

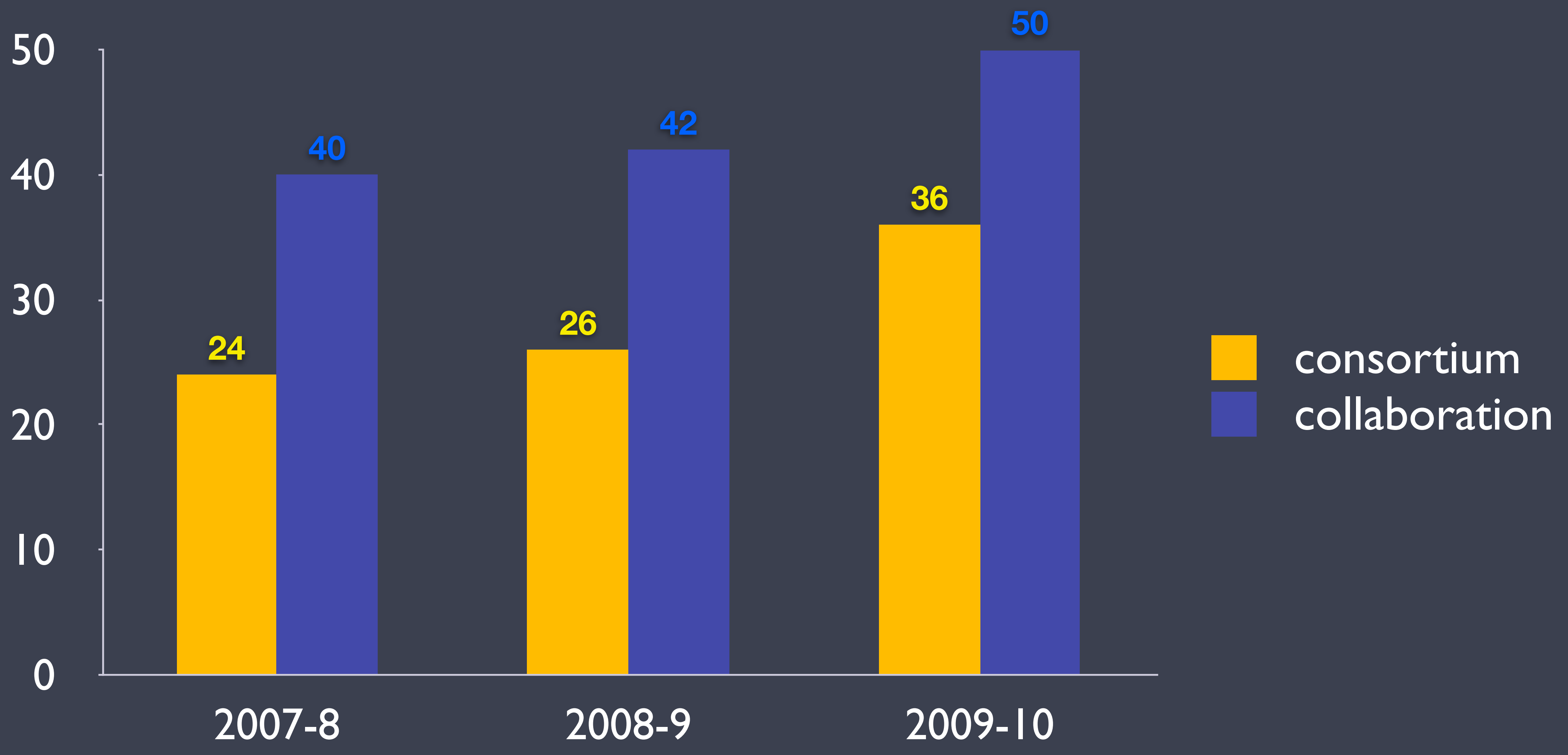
$100M

$ 9 Bn

“creativity is no longer about which companies have the most visionary executives, but who has the most compelling **architecture of participation** ... which companies make it easy, interesting and rewarding for a wide range of contributors to offer ideas, solve problems and improve products?”

Tim O’Reilly

PubMed terms Stem Cell Collaborations



Challenges

. stem cell researchers seek understanding of pluripotency

. need to standardize understanding of cellular phenotypes that define stem cells and their progeny

. independent laboratories and protocols

. independent discoveries

How do develop the benefits of open sharing in closed consortia?

Multi-disciplinary collaborations

- Project proposal
  - Attention to standards, communication, data sharing
- Project rollout
  - Concept of a “database”
  - **Curation**
- Individual research groups vs the collective
  - The post-doc challenge

Biocurator as social engineer

> can the consortium be encouraged to become open?

. internally

. collectively

Facilitate the collaboratory

> challenges of a central shared resource

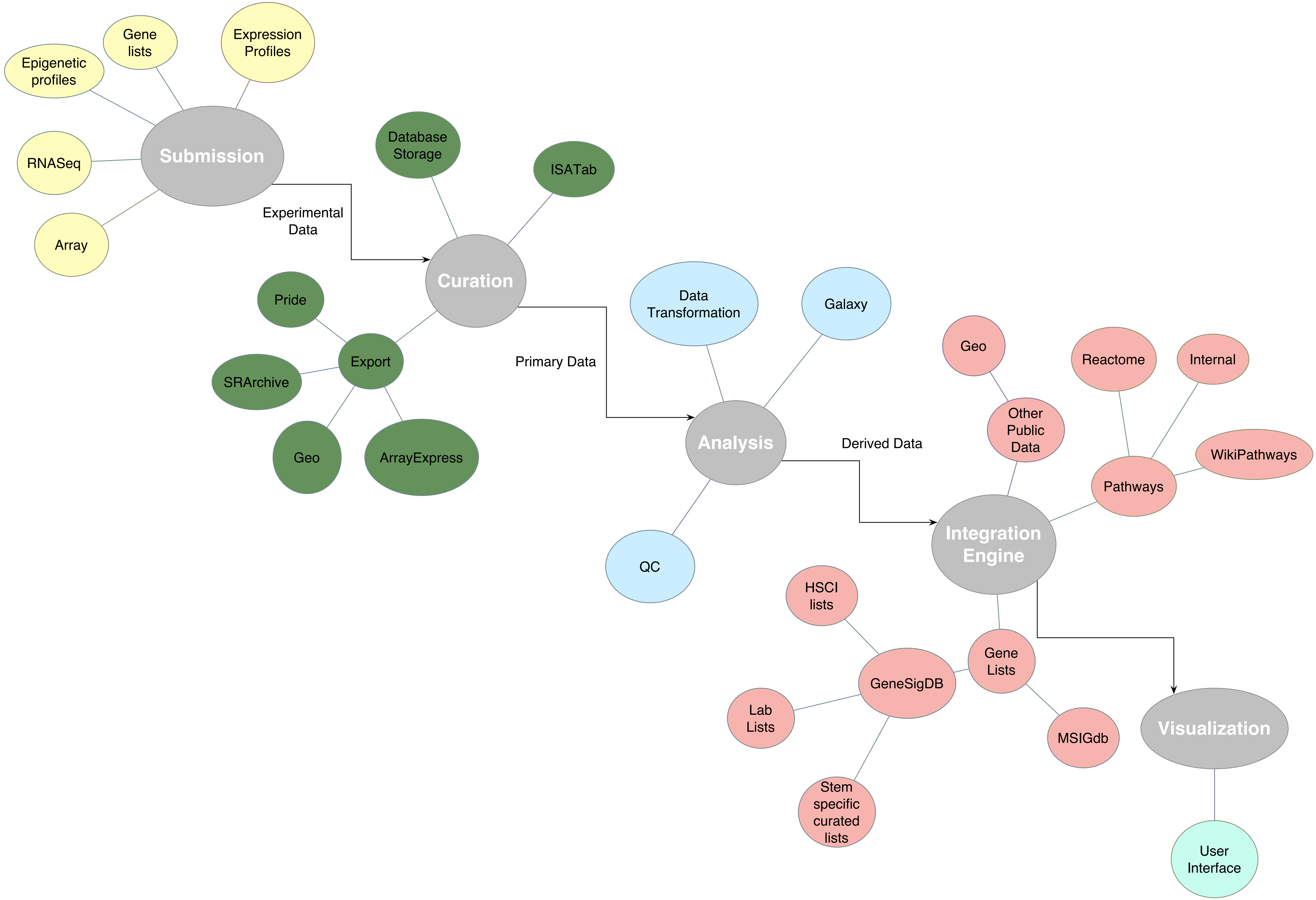
> groups focus on common yet heterogenous themes

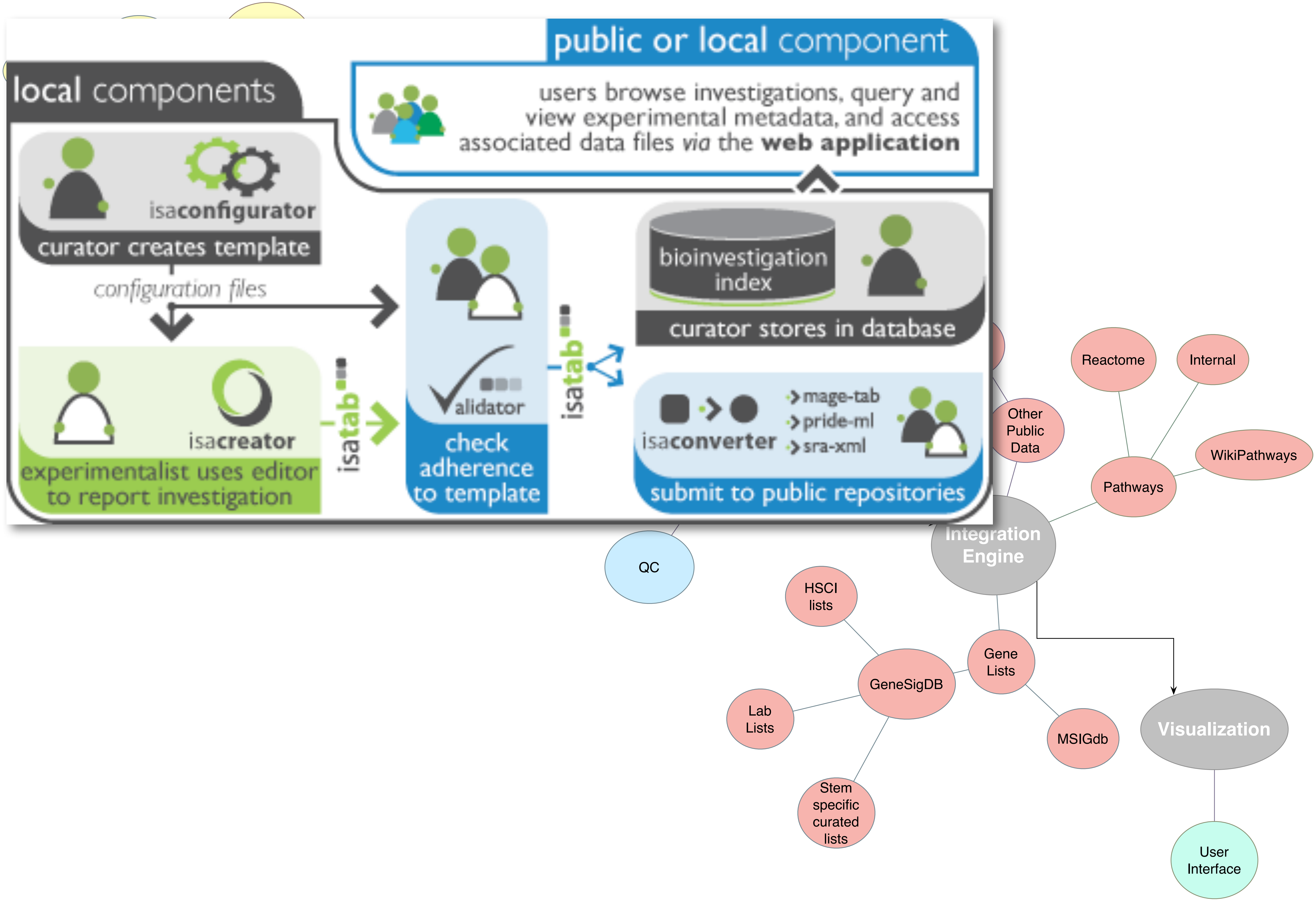
. haematopoesis

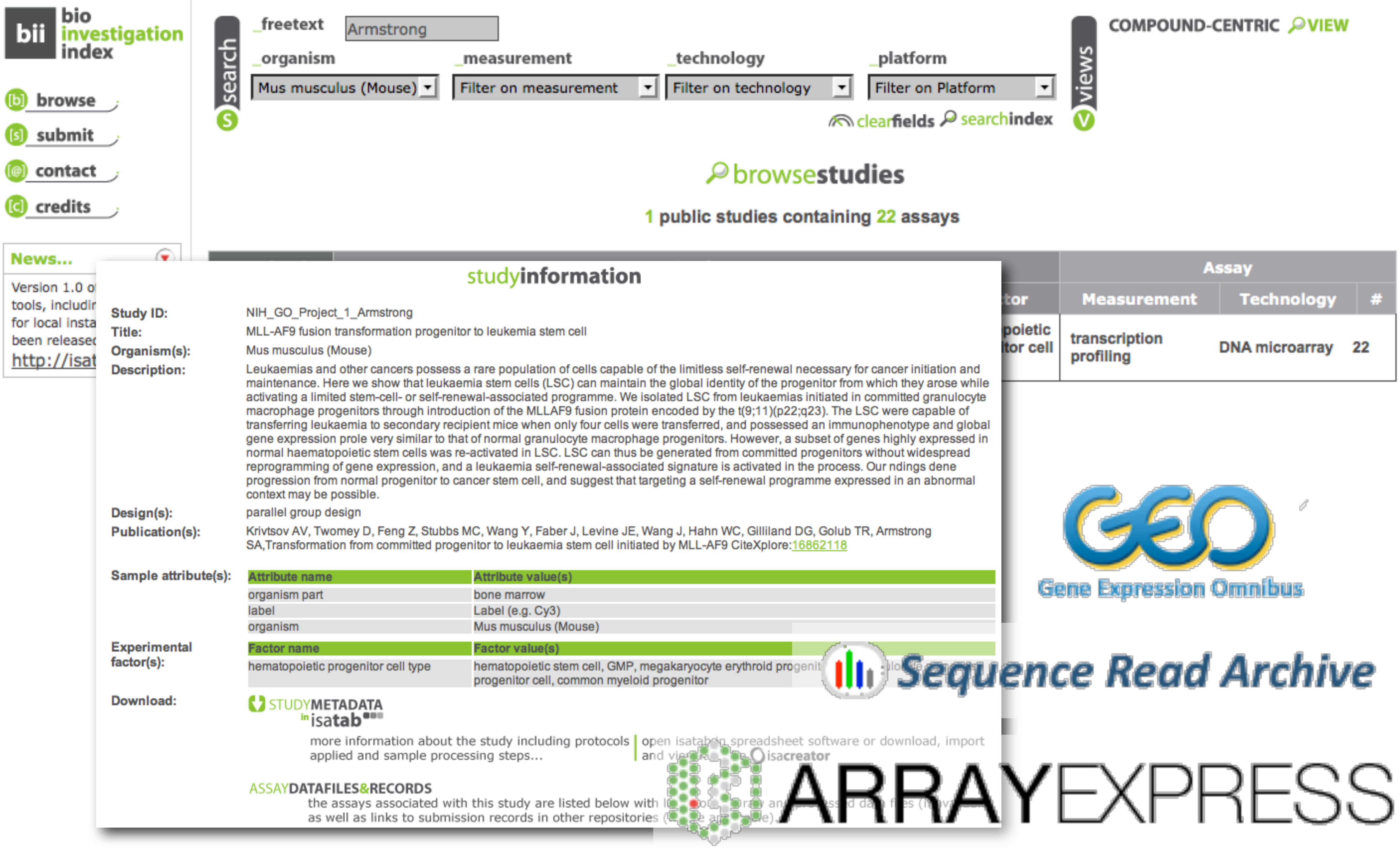
    . mouse
    . human
    . zebrafish
    . chlp-seq
    . array

Implementation

Making it easy to share.....


Gene lists
Expression Profiles
Epigenetic profiles
Submission
RNASeq
Array
Experimental Data
Database Storage
ISATab
Curation
Pride
Export
SRArchive
Geo
ArrayExpress
Primary Data
Data Transformation
Galaxy
Analysis
QC
Derived Data
Geo
Other Public Data
Reactome
Internal
WikiPathways
Pathways
Integration Engine
HSCI lists
GeneSigDB
Gene Lists
Lab Lists
MSIGdb
Stem specific curated lists
Visualization
User Interface


public or local component
users browse investigations, query and view experimental metadata, and access associated data files *via* the **web application**
local components
isaconfigurator
curator creates template
configuration files
isacreator
experimentalist uses editor to report investigation
isatab
validator
check adherence to template
isatab
bioinvestigation index
curator stores in database
isaconverter
mage-tab
pride-ml
sra-xml
submit to public repositories
Reactome
Internal
Other Public Data
WikiPathways
Pathways
Integration Engine
QC
HSCI lists
Gene Lists
GeneSigDB
Lab Lists
MSIGdb
Visualization
Stem specific curated lists
User Interface



studyinformation

Study ID: NIH_GO_Project_1_Armstrong

Title: MLL-AF9 fusion transformation progenitor to leukemia stem cell

Organism(s): Mus musculus (Mouse)

Description: Leukaemias and other cancers possess a rare population of cells capable of the limitless self-renewal necessary for cancer initiation and maintenance. Here we show that leukaemia stem cells (LSC) can maintain the global identity of the progenitor from which they arose while activating a limited stem-cell- or self-renewal-associated programme. We isolated LSC from leukaemias initiated in committed granulocyte macrophage progenitors through introduction of the MLLAF9 fusion protein encoded by the t(9;11)(p22;q23). The LSC were capable of transferring leukaemia to secondary recipient mice when only four cells were transferred, and possessed an immunophenotype and global gene expression prole very similar to that of normal granulocyte macrophage progenitors. However, a subset of genes highly expressed in normal haematopoietic stem cells was re-activated in LSC. LSC can thus be generated from committed progenitors without widespread reprogramming of gene expression, and a leukaemia self-renewal-associated signature is activated in the process. Our ndings dene progression from normal progenitor to cancer stem cell, and suggest that targeting a self-renewal programme expressed in an abnormal context may be possible.

Design(s): parallel group design

Publication(s): Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, Levine JE, Wang J, Hahn WC, Gilliland DG, Golub TR, Armstrong SA,Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9 CiteXplore:16862118

Sample attribute(s):

Attribute name	Attribute value(s)
organism part	bone marrow
label	Label (e.g. Cy3)
organism	Mus musculus (Mouse)

Experimental factor(s):

Factor name	Factor value(s)
hematopoietic progenitor cell type	hematopoietic stem cell, GMP, megakaryocyte erythroid progenit progenitor cell, common myeloid progenitor

Download:

STUDYMETADATA in isatab

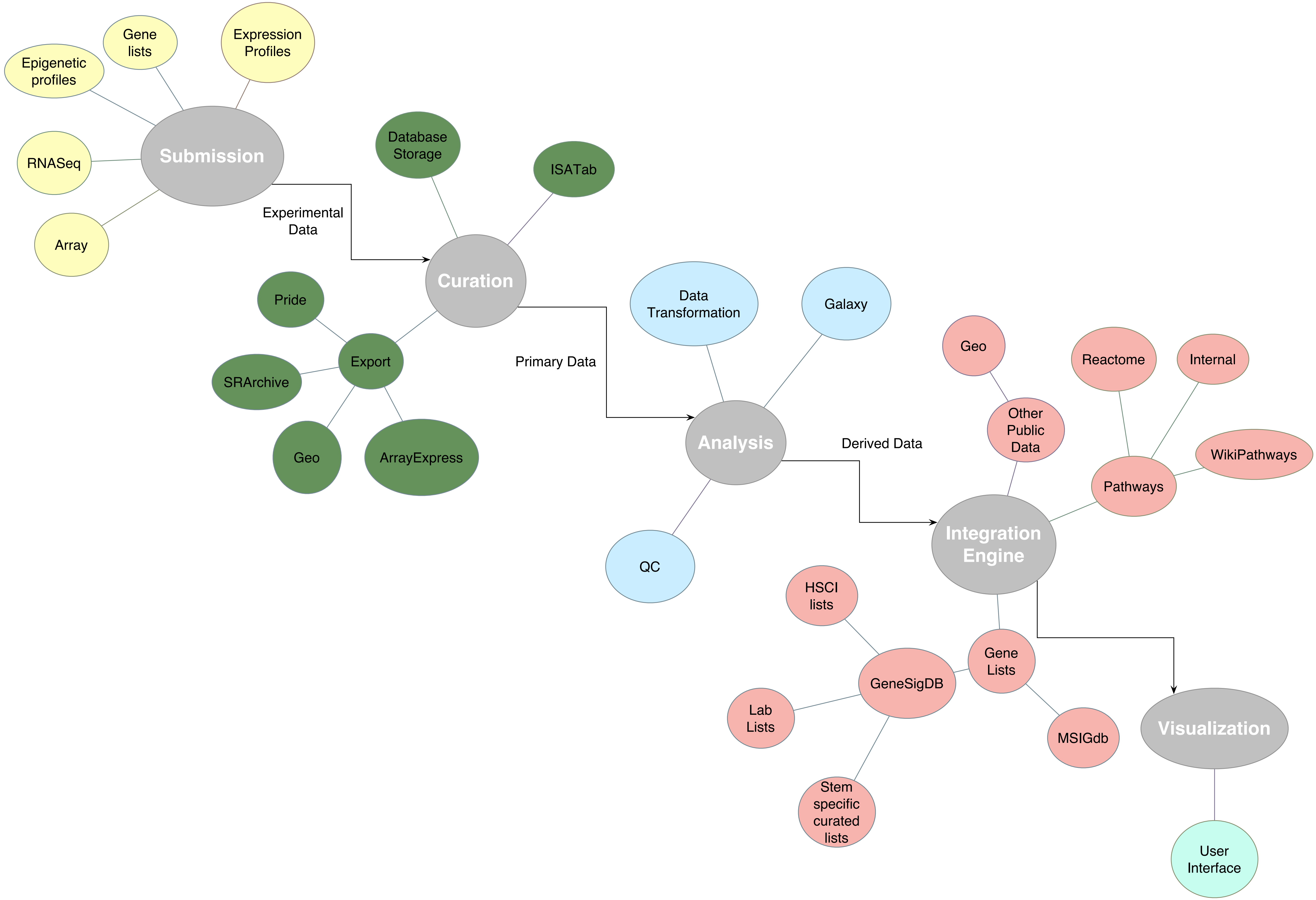
more information about the study including protocols applied and sample processing steps...

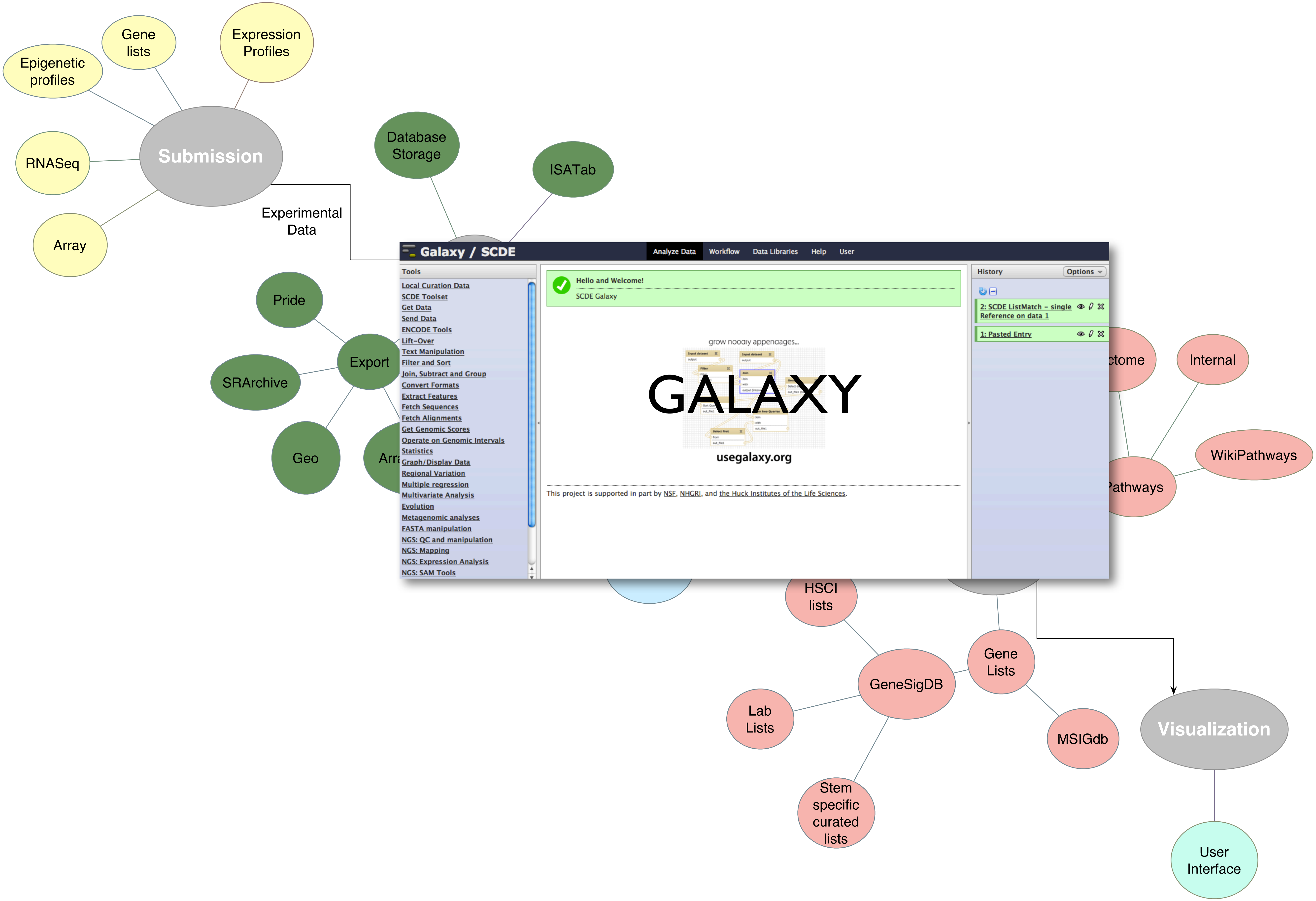
open isatab spreadsheet software or download, import and vie isacreator

ASSAYDATAFILES&RECORDS

the assays associated with this study are listed below with
as well as links to submission records in other repositories

Sharing data and dissemination using BII

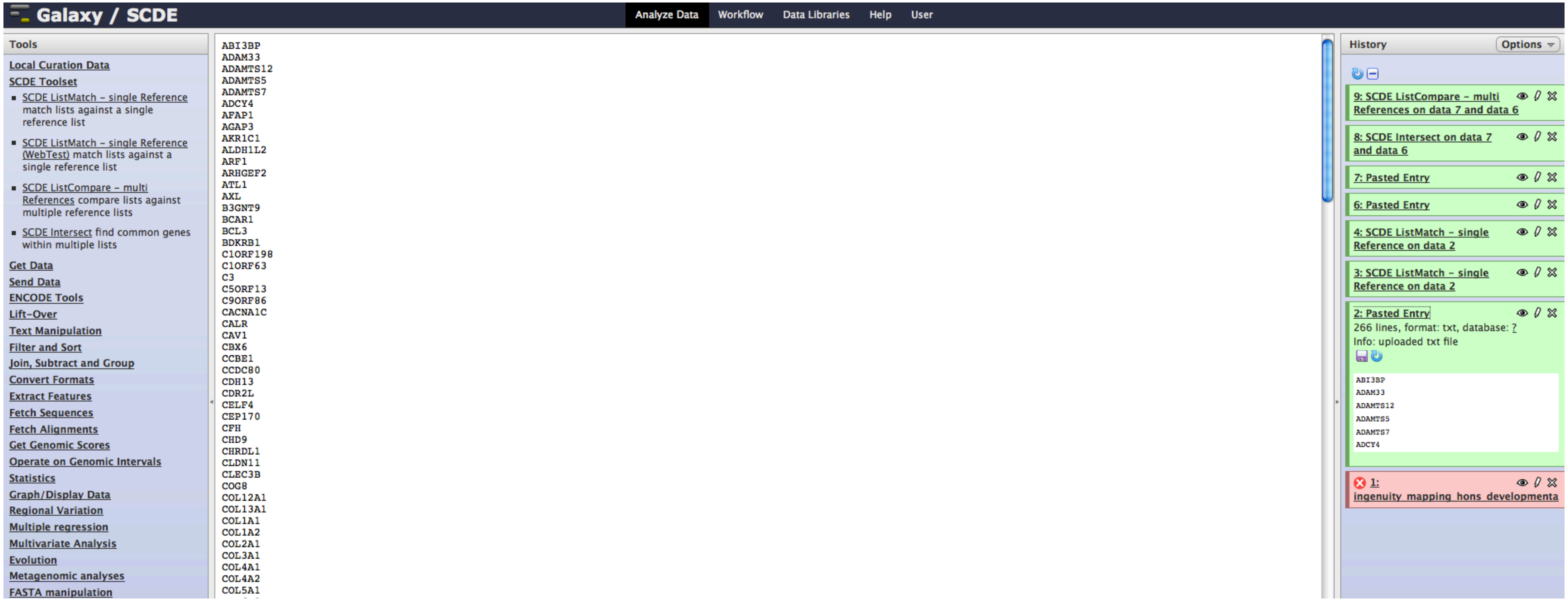

Epigenetic profiles
Gene lists
Expression Profiles
RNASeq
Array
Submission
Experimental Data
Database Storage
ISATab
Curation
Pride
Export
SRArchive
Geo
ArrayExpress
Primary Data
Data Transformation
Galaxy
Analysis
QC
Derived Data
Geo
Other Public Data
Reactome
Internal
WikiPathways
Pathways
Integration Engine
HSCI lists
GeneSigDB
Gene Lists
Lab Lists
Stem specific curated lists
MSIGdb
Visualization
User Interface



Hello and Welcome!

SCDE Galaxy

This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.



- SCDE ListMatch - single Reference match lists against a single reference list
- SCDE ListMatch - single Reference (WebTest) match lists against a single reference list
- SCDE ListCompare - multi References compare lists against multiple reference lists
- SCDE Intersect find common genes within multiple lists

```
ABI3BP
ADAM33
ADAMTS12
ADAMTS5
ADAMTS7
ADCY4
AFAP1
AGAP3
AKR1C1
ALDH1L2
ARF1
ARHGEF2
ATL1
AXL
B3GNT9
BCAR1
BCL3
BDKRB1
C1ORF198
C1ORF63
C3
C5ORF13
C9ORF86
CACNA1C
CALR
CAV1
CBX6
CCBE1
CCDC80
CDH13
CDR2L
CELF4
CEP170
CFH
CHD9
CHRDL1
CLDN11
CLEC3B
COG8
COL12A1
COL13A1
COL1A1
COL1A2
COL2A1
COL3A1
COL4A1
COL4A2
COL5A1
```

9: SCDE ListCompare - multi References on data 7 and data 6

8: SCDE Intersect on data 7 and data 6

7: Pasted Entry

6: Pasted Entry

4: SCDE ListMatch - single Reference on data 2

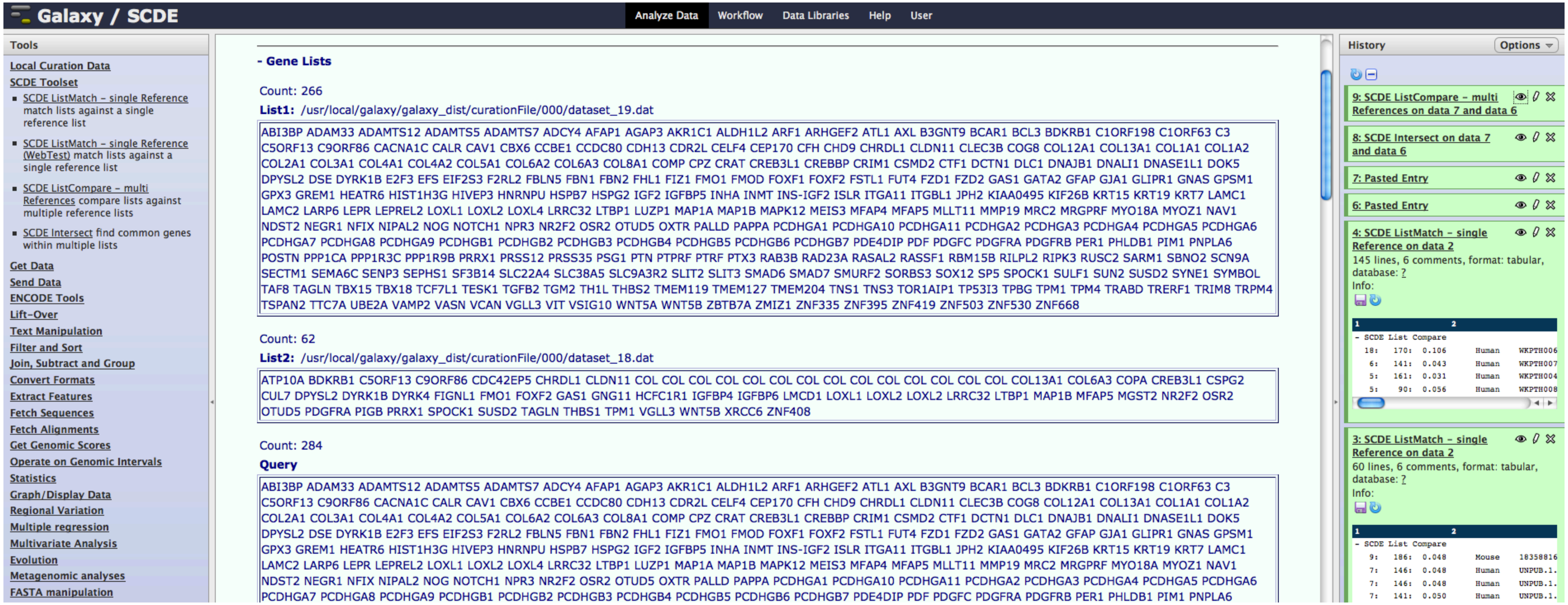
3: SCDE ListMatch - single Reference on data 2

2: Pasted Entry
266 lines, format: txt, database: ?
Info: uploaded txt file

```
ABI3BP
ADAM33
ADAMTS12
ADAMTS5
ADAMTS7
ADCY4
```

1: ingenuity_mapping_hons_developmenta

method chains: History/sharing



- SCDE ListMatch - single Reference match lists against a single reference list
- SCDE ListMatch - single Reference (WebTest) match lists against a single reference list
- SCDE ListCompare - multi References compare lists against multiple reference lists
- SCDE Intersect find common genes within multiple lists

- Gene Lists

Count: 266

List1: /usr/local/galaxy/galaxy_dist/curationFile/000/dataset_19.dat

ABI3BP ADAM33 ADAMTS12 ADAMTS5 ADAMTS7 ADCY4 AFAP1 AGAP3 AKR1C1 ALDH1L2 ARF1 ARHGEF2 ATL1 AXL B3GNT9 BCAR1 BCL3 BDKRB1 C1ORF198 C1ORF63 C3 C5ORF13 C9ORF86 CACNA1C CALR CAV1 CBX6 CCBE1 CCDC80 CDH13 CDR2L CELF4 CEP170 CFH CHD9 CHRDL1 CLDN11 CLEC3B COG8 COL12A1 COL13A1 COL1A1 COL1A2 COL2A1 COL3A1 COL4A1 COL4A2 COL5A1 COL6A2 COL6A3 COL8A1 COMP CPZ CRAT CREB3L1 CREBBP CRIM1 CSMD2 CTF1 DCTN1 DLC1 DNAJB1 DNALI1 DNASE1L1 DOK5 DPYSL2 DSE DYRK1B E2F3 EFS EIF2S3 F2RL2 FBLN5 FBN1 FBN2 FHL1 FIZ1 FMO1 FMOD FOXF1 FOXF2 FSTL1 FUT4 FZD1 FZD2 GAS1 GATA2 GFAP GJA1 GLIPR1 GNAS GPSM1 GPX3 GREM1 HEATR6 HIST1H3G HIVEP3 HNRNPU HSPB7 HSPG2 IGF2 IGFBP5 INHA INMT INS-IGF2 ISLR ITGA11 ITGBL1 JPH2 KIAA0495 KIF26B KRT15 KRT19 KRT7 LAMC1 LAMC2 LARP6 LEPR LEPREL2 LOXL1 LOXL2 LOXL4 LRRC32 LTBP1 LUZP1 MAP1A MAP1B MAPK12 MEIS3 MFAP4 MFAP5 MLLT11 MMP19 MRC2 MRGPRF MYO18A MYOZ1 NAV1 NDST2 NEGR1 NFIX NIPAL2 NOG NOTCH1 NPR3 NR2F2 OSR2 OTUD5 OXTR PALLD PAPPA PCDHGA1 PCDHGA10 PCDHGA11 PCDHGA2 PCDHGA3 PCDHGA4 PCDHGA5 PCDHGA6 PCDHGA7 PCDHGA8 PCDHGA9 PCDHGB1 PCDHGB2 PCDHGB3 PCDHGB4 PCDHGB5 PCDHGB6 PCDHGB7 PDE4DIP PDF PDGFC PDGFRA PDGFRB PER1 PHLDB1 PIM1 PNPLA6 POSTN PPP1CA PPP1R3C PPP1R9B PRRX1 PRSS12 PRSS35 PSG1 PTN PTPRF PTRF PTX3 RAB3B RAD23A RASAL2 RASSF1 RBM15B RILPL2 RIPK3 RUSC2 SARM1 SBNO2 SCN9A SECTM1 SEMA6C SENP3 SEPHS1 SF3B14 SLC22A4 SLC38A5 SLC9A3R2 SLIT2 SLIT3 SMAD6 SMAD7 SMURF2 SORBS3 SOX12 SP5 SPOCK1 SULF1 SUN2 SUSD2 SYNE1 SYMBOL TAF8 TAGLN TBX15 TBX18 TCF7L1 TESK1 TGFB2 TGM2 TH1L THBS2 TMEM119 TMEM127 TMEM204 TNS1 TNS3 TOR1AIP1 TP53I3 TPBG TPM1 TPM4 TRABD TRERF1 TRIM8 TRPM4 TSPAN2 TTC7A UBE2A VAMP2 VASN VCAN VGLL3 VIT VSIG10 WNT5A WNT5B ZBTB7A ZMIZ1 ZNF335 ZNF395 ZNF419 ZNF503 ZNF530 ZNF668

Count: 62

List2: /usr/local/galaxy/galaxy_dist/curationFile/000/dataset_18.dat

ATP10A BDKRB1 C5ORF13 C9ORF86 CDC42EP5 CHRDL1 CLDN11 COL COL COL COL COL COL COL COL COL COL COL COL COL COL COL13A1 COL6A3 COPA CREB3L1 CSPG2 CUL7 DPYSL2 DYRK1B DYRK4 FIGNL1 FMO1 FOXF2 GAS1 GNG11 HCFC1R1 IGFBP4 IGFBP6 LMCD1 LOXL1 LOXL2 LOXL2 LRRC32 LTBP1 MAP1B MFAP5 MGST2 NR2F2 OSR2 OTUD5 PDGFRA PIGB PRRX1 SPOCK1 SUSD2 TAGLN THBS1 TPM1 VGLL3 WNT5B XRCC6 ZNF408

Count: 284

Query

ABI3BP ADAM33 ADAMTS12 ADAMTS5 ADAMTS7 ADCY4 AFAP1 AGAP3 AKR1C1 ALDH1L2 ARF1 ARHGEF2 ATL1 AXL B3GNT9 BCAR1 BCL3 BDKRB1 C1ORF198 C1ORF63 C3 C5ORF13 C9ORF86 CACNA1C CALR CAV1 CBX6 CCBE1 CCDC80 CDH13 CDR2L CELF4 CEP170 CFH CHD9 CHRDL1 CLDN11 CLEC3B COG8 COL12A1 COL13A1 COL1A1 COL1A2 COL2A1 COL3A1 COL4A1 COL4A2 COL5A1 COL6A2 COL6A3 COL8A1 COMP CPZ CRAT CREB3L1 CREBBP CRIM1 CSMD2 CTF1 DCTN1 DLC1 DNAJB1 DNALI1 DNASE1L1 DOK5 DPYSL2 DSE DYRK1B E2F3 EFS EIF2S3 F2RL2 FBLN5 FBN1 FBN2 FHL1 FIZ1 FMO1 FMOD FOXF1 FOXF2 FSTL1 FUT4 FZD1 FZD2 GAS1 GATA2 GFAP GJA1 GLIPR1 GNAS GPSM1 GPX3 GREM1 HEATR6 HIST1H3G HIVEP3 HNRNPU HSPB7 HSPG2 IGF2 IGFBP5 INHA INMT INS-IGF2 ISLR ITGA11 ITGBL1 JPH2 KIAA0495 KIF26B KRT15 KRT19 KRT7 LAMC1 LAMC2 LARP6 LEPR LEPREL2 LOXL1 LOXL2 LOXL4 LRRC32 LTBP1 LUZP1 MAP1A MAP1B MAPK12 MEIS3 MFAP4 MFAP5 MLLT11 MMP19 MRC2 MRGPRF MYO18A MYOZ1 NAV1 NDST2 NEGR1 NFIX NIPAL2 NOG NOTCH1 NPR3 NR2F2 OSR2 OTUD5 OXTR PALLD PAPPA PCDHGA1 PCDHGA10 PCDHGA11 PCDHGA2 PCDHGA3 PCDHGA4 PCDHGA5 PCDHGA6 PCDHGA7 PCDHGA8 PCDHGA9 PCDHGB1 PCDHGB2 PCDHGB3 PCDHGB4 PCDHGB5 PCDHGB6 PCDHGB7 PDE4DIP PDF PDGFC PDGFRA PDGFRB PER1 PHLDB1 PIM1 PNPLA6

9: SCDE ListCompare - multi References on data 7 and data 6

8: SCDE Intersect on data 7 and data 6

7: Pasted Entry

6: Pasted Entry

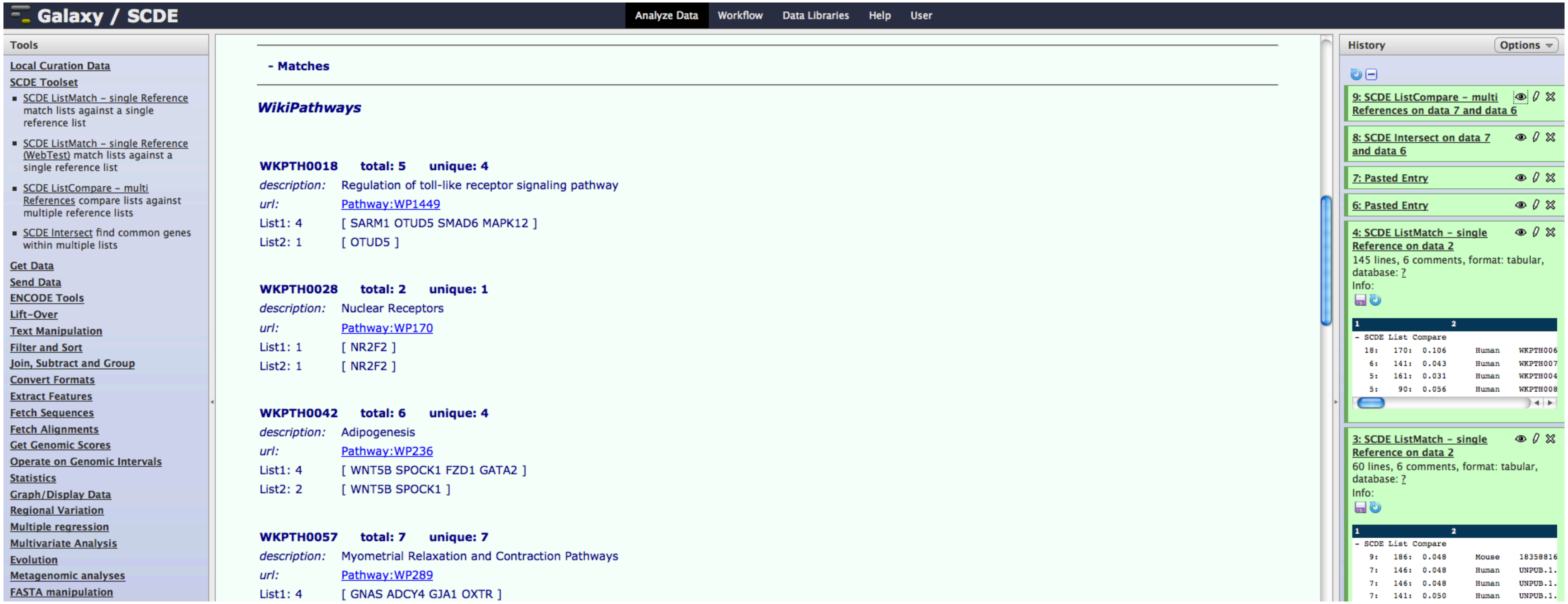
4: SCDE ListMatch - single Reference on data 2
145 lines, 6 comments, format: tabular, database: ?
Info:

```
1                  2
- SCDE List Compare
  18:   170:  0.106      Human    WKPTH006
   6:   141:  0.043      Human    WKPTH007
   5:   161:  0.031      Human    WKPTH004
   5:    90:  0.056      Human    WKPTH008
```

3: SCDE ListMatch - single Reference on data 2
60 lines, 6 comments, format: tabular, database: ?
Info:

```
1                  2
- SCDE List Compare
   9:   186:  0.048      Mouse    18358816
   7:   146:  0.048      Human    UNPUB.1.
   7:   146:  0.048      Human    UNPUB.1.
   7:   141:  0.050      Human    UNPUB.1.
```

Shared genelists



- Matches

WikiPathways

WKPTH0018 **total: 5** **unique: 4**

description:	Regulation of toll-like receptor signaling pathway
url:	Pathway:WP1449
List1: 4	[SARM1 OTUD5 SMAD6 MAPK12]
List2: 1	[OTUD5]

WKPTH0028 **total: 2** **unique: 1**

description:	Nuclear Receptors
url:	Pathway:WP170
List1: 1	[NR2F2]
List2: 1	[NR2F2]

WKPTH0042 **total: 6** **unique: 4**

description:	Adipogenesis
url:	Pathway:WP236
List1: 4	[WNT5B SPOCK1 FZD1 GATA2]
List2: 2	[WNT5B SPOCK1]

WKPTH0057 **total: 7** **unique: 7**

description:	Myometrial Relaxation and Contraction Pathways
url:	Pathway:WP289
List1: 4	[GNAS ADCY4 GJA1 OXTR]

9: SCDE ListCompare - multi References on data 7 and data 6

8: SCDE Intersect on data 7 and data 6

7: Pasted Entry

6: Pasted Entry

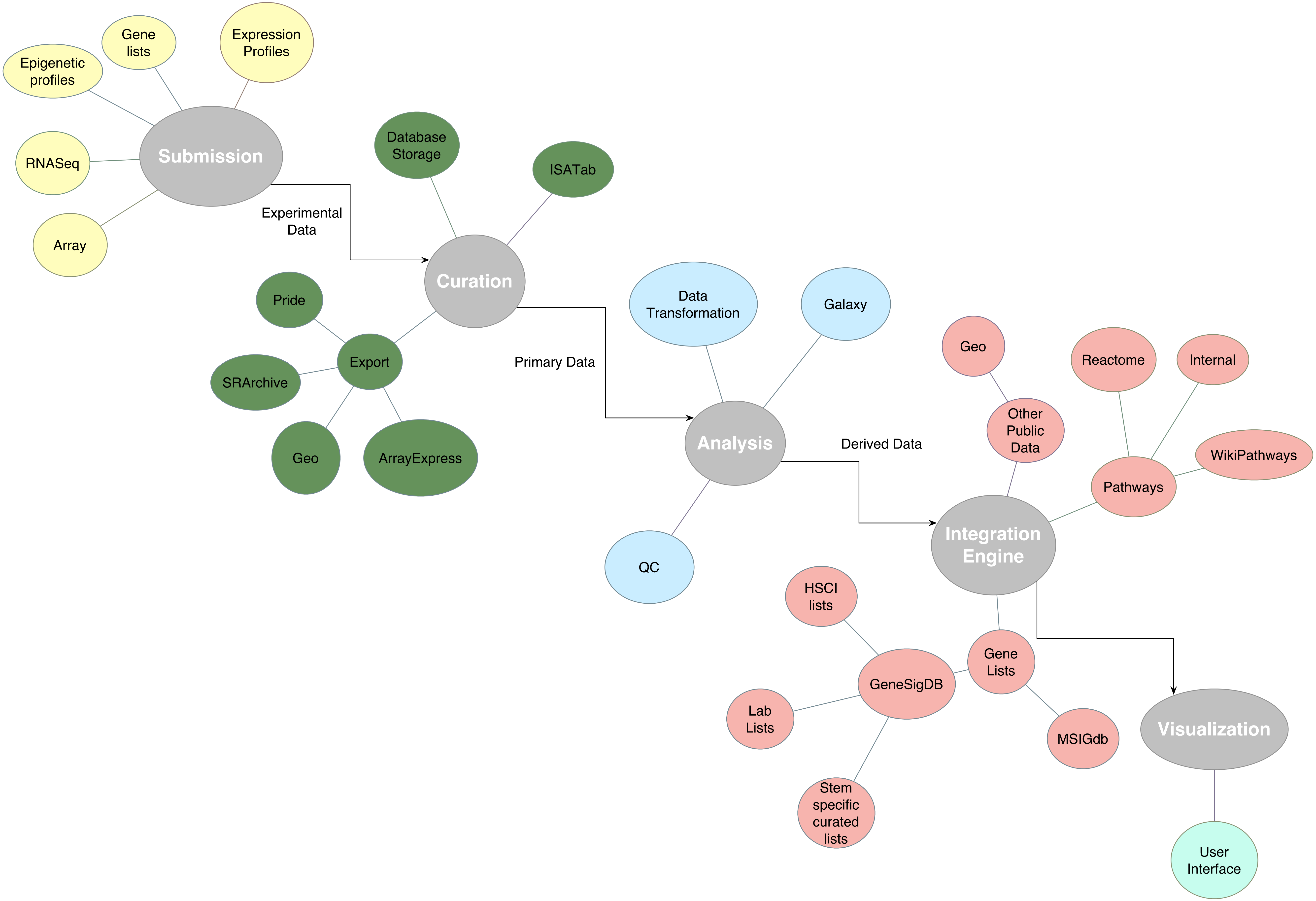
4: SCDE ListMatch - single Reference on data 2
145 lines, 6 comments, format: tabular, database: ?
Info:

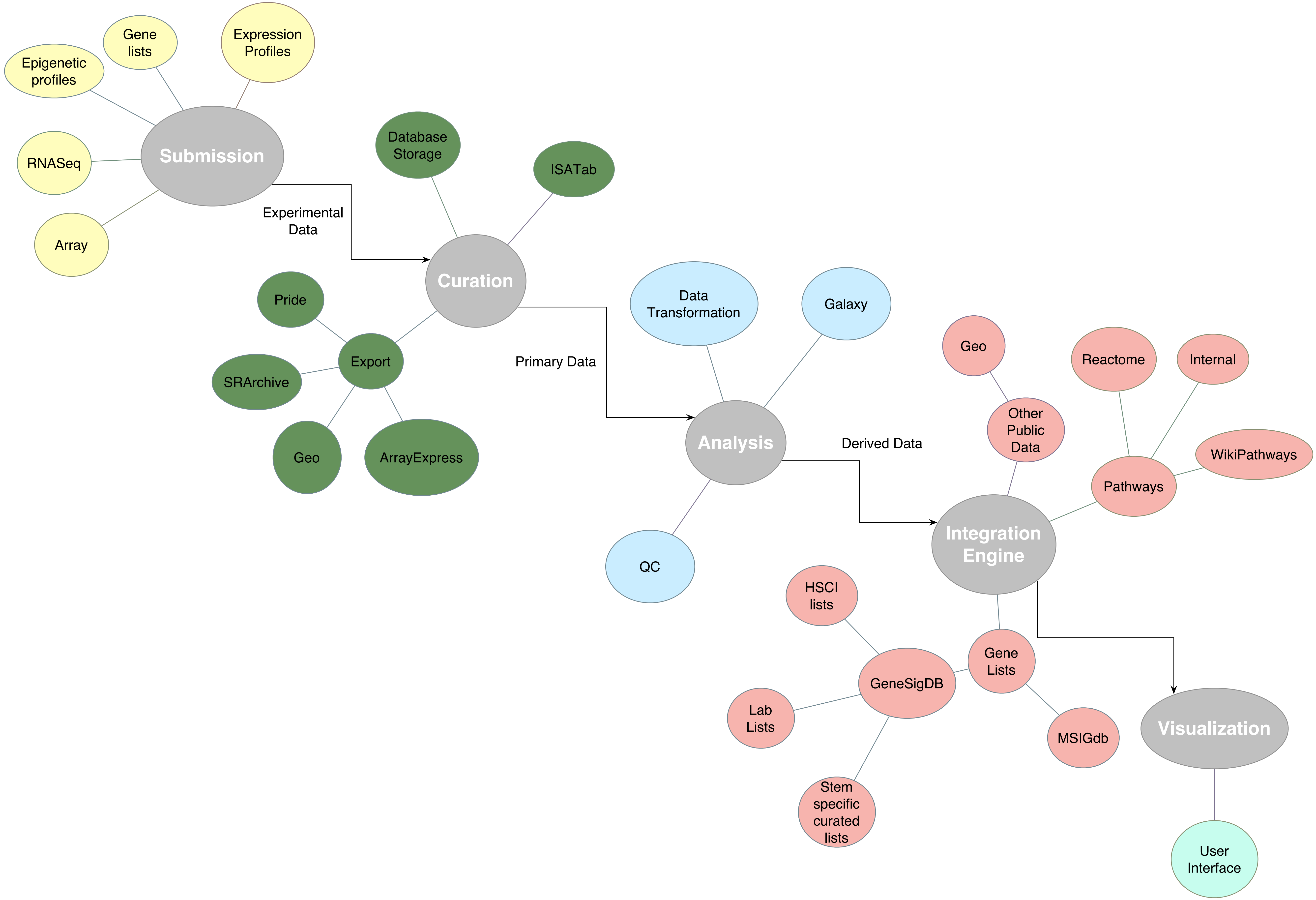
```
1                  2
- SCDE List Compare
  18:   170:  0.106      Human    WKPTH006
   6:   141:  0.043      Human    WKPTH007
   5:   161:  0.031      Human    WKPTH004
   5:    90:  0.056      Human    WKPTH008
```

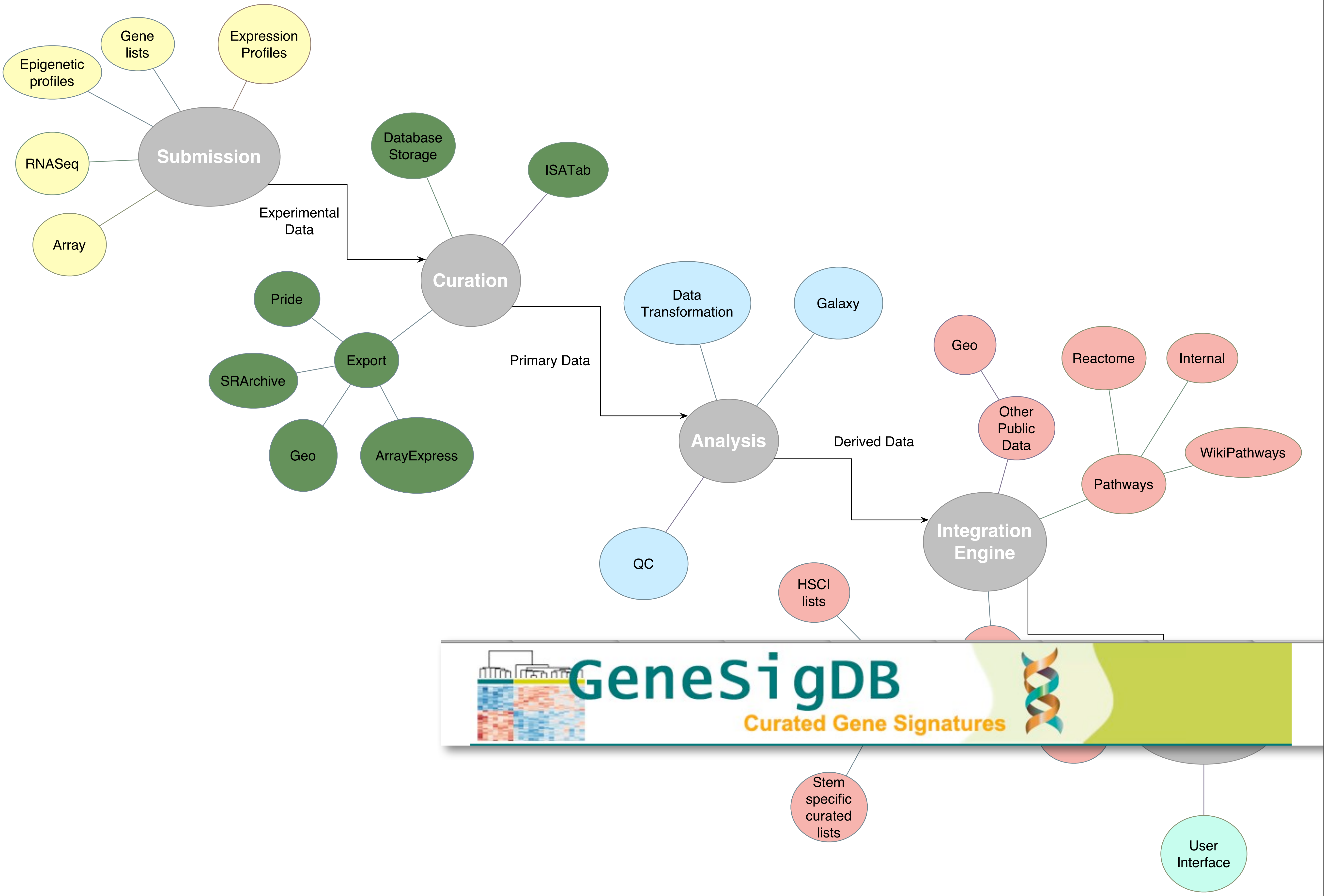
3: SCDE ListMatch - single Reference on data 2
60 lines, 6 comments, format: tabular, database: ?
Info:

```
1                  2
- SCDE List Compare
   9:   186:  0.048      Mouse    18358816
   7:   146:  0.048      Human    UNPUB.1.
   7:   146:  0.048      Human    UNPUB.1.
   7:   141:  0.050      Human    UNPUB.1.
```

Projecting to Wikipathways


Epigenetic profiles
Gene lists
Expression Profiles
RNASeq
Submission
Array
Experimental Data
Database Storage
ISATab
Curation
Pride
Export
SRArchive
Geo
ArrayExpress
Primary Data
Data Transformation
Galaxy
Analysis
QC
Derived Data
Geo
Other Public Data
Reactome
Internal
WikiPathways
Pathways
Integration Engine
HSCI lists
GeneSigDB
Gene Lists
Lab Lists
MSIGdb
Stem specific curated lists
Visualization
User Interface


Gene lists
Expression Profiles
Epigenetic profiles
Submission
RNASeq
Array
Experimental Data
Database Storage
ISATab
Curation
Pride
Export
SRArchive
Geo
ArrayExpress
Primary Data
Data Transformation
Galaxy
Analysis
QC
Derived Data
Geo
Other Public Data
Reactome
Internal
WikiPathways
Pathways
Integration Engine
HSCI lists
GeneSigDB
Gene Lists
Lab Lists
MSIGdb
Stem specific curated lists
Visualization
User Interface



Submission

Epigenetic profiles

Gene lists

Expression Profiles

RNASeq

Array

Experimental Data

Curation

Database Storage

ISATab

Export

Pride

SRArchive

Geo

ArrayExpress

Primary Data

Analysis

Data Transformation

Galaxy

QC

Derived Data

Integration Engine

Other Public Data

Geo

Pathways

Reactome

Internal

WikiPathways

HSCI lists

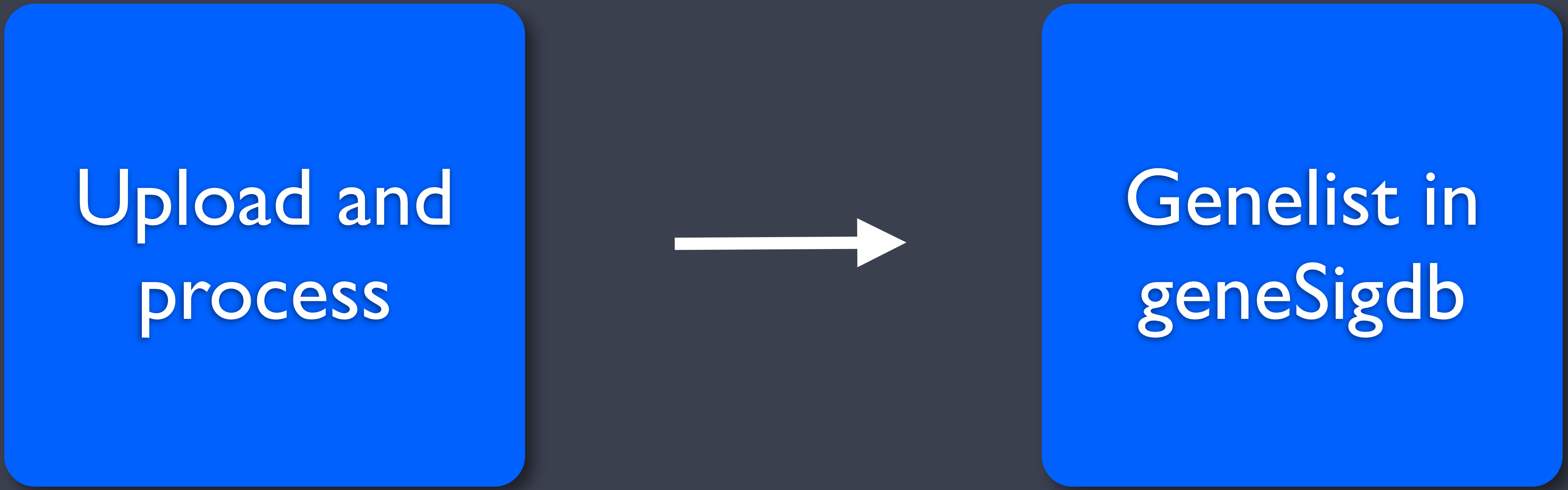
Stem specific curated lists

User Interface

GeneSigDB

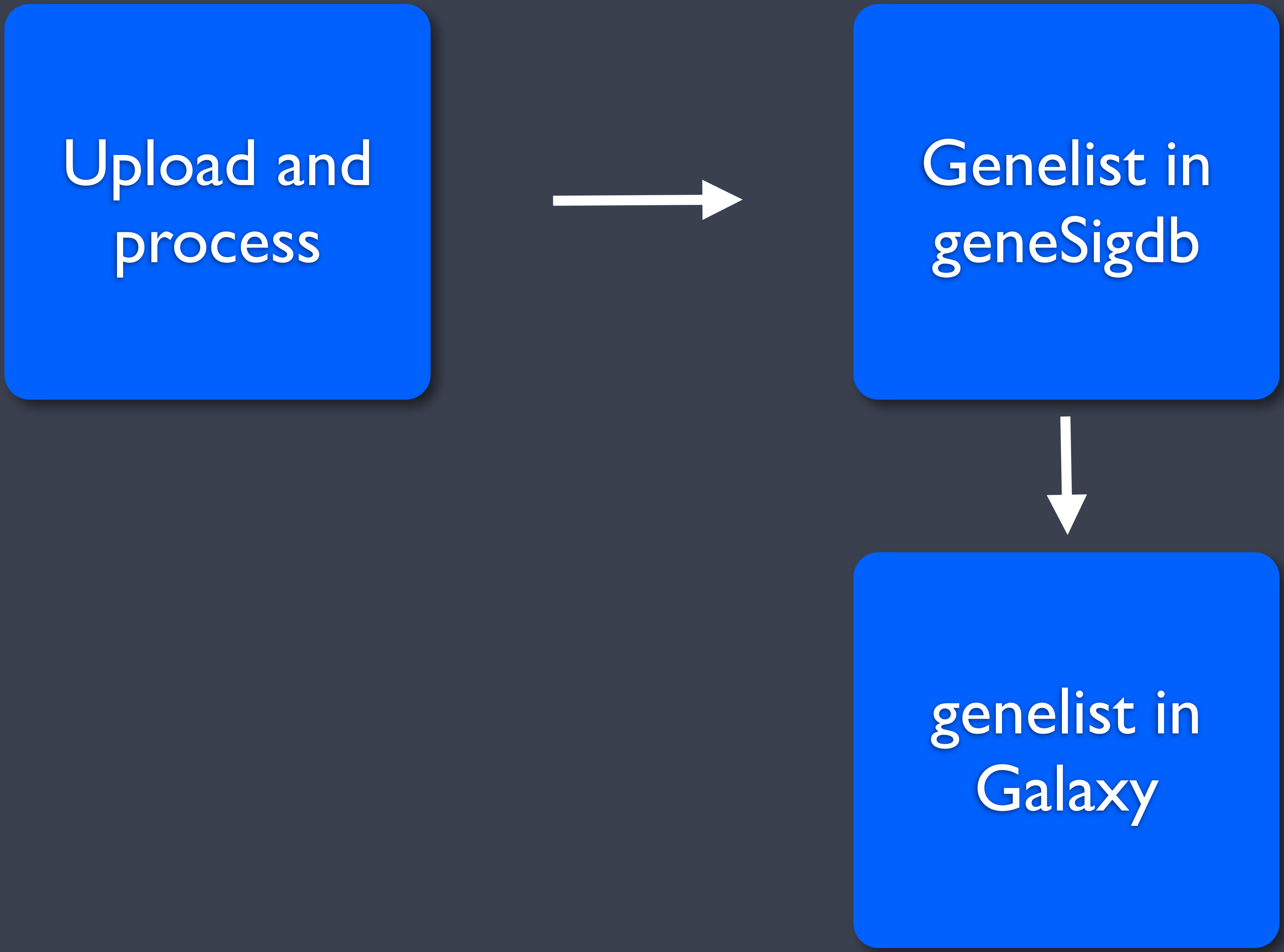
Curated Gene Signatures

Upload and process



Upload and process

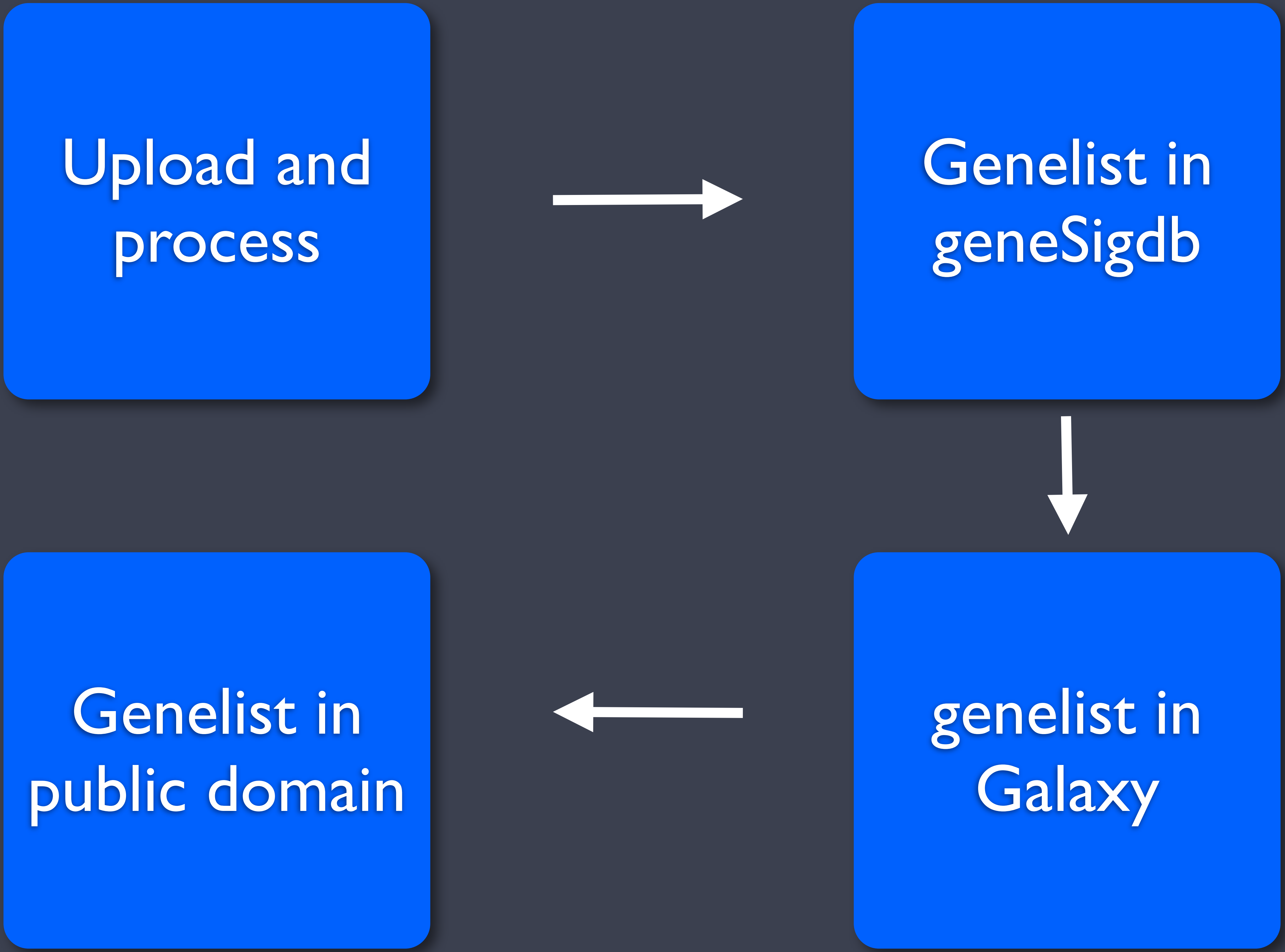
Genelist in geneSigdb



Upload and process

Genelist in geneSigdb

genelist in Galaxy



Upload and process

Genelist in geneSigdb

genelist in Galaxy

Genelist in public domain

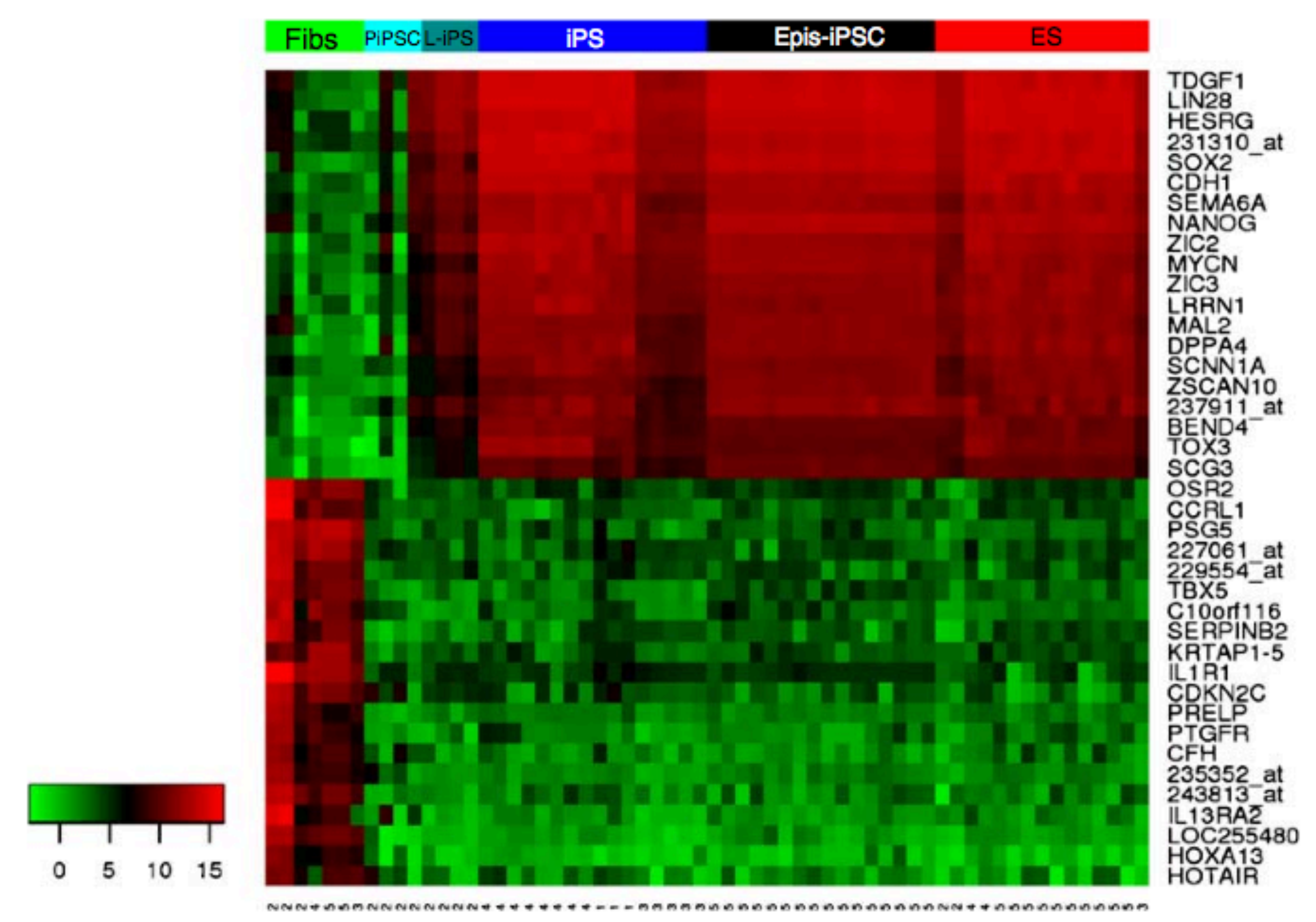
Sharing

. **How to share molecular profiles meaningfully?**

Gene Level Comparisons

Array

Multi-study, single species;
5 different human ES/iPS gene expression studies combined to produce list of 20 differentially regulated genes in the transition from fibroblasts to iPS/ES cells



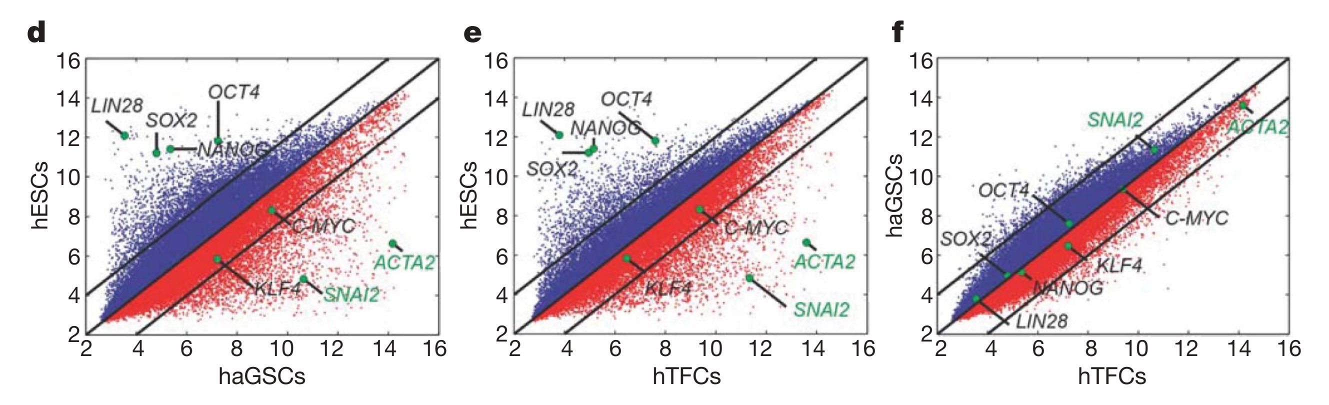
Wang et al Stem Cell Rev. 2010

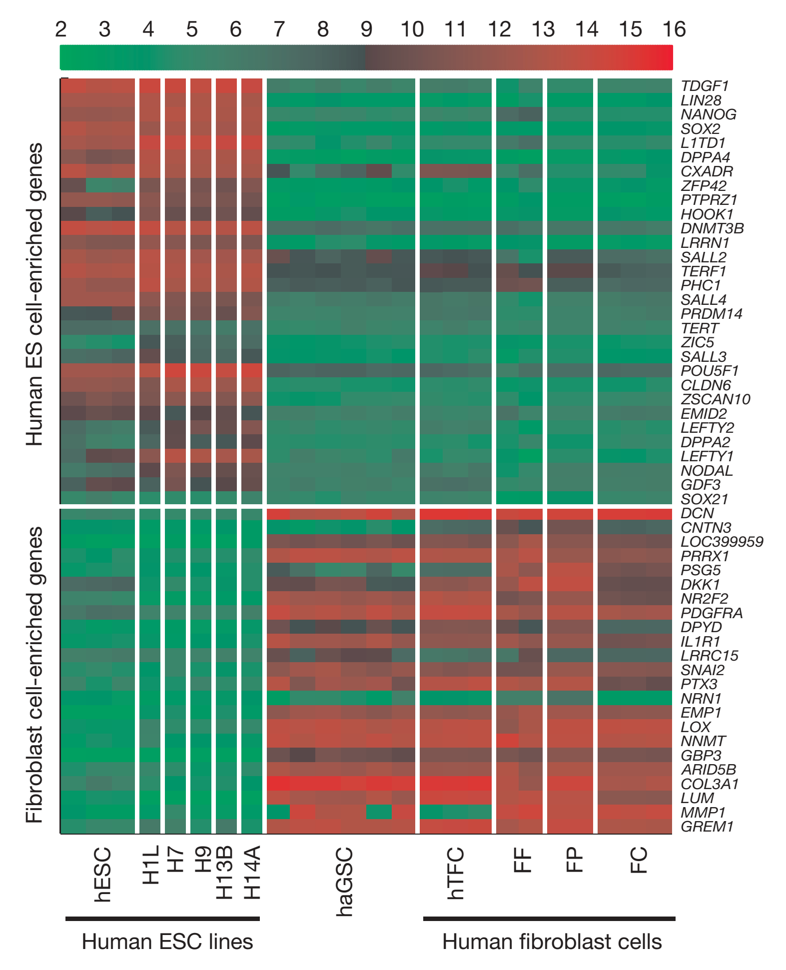
Array II

Ko et al, Nature 2010, dispute Conrad claims based on gene level analysis

Global gene expression analysis comparing the expression of pluripotency marker genes in **haGSCs**, human testicular fibroblasts (**hTFCs**) and **ES** and **fibroblast** samples derived from different laboratories

Demonstrates **haGSCs** are more similar to **fibrobalsts** that ES cells





Pathways - canonical view

KEGG Kyoto Encyclopedia of Genes and Genomes

PC Pathway Commons

BioGRID

Memorial Sloan-Kettering Cancer Center

Human Protein Reference Database

HumanCyc

SBCNY

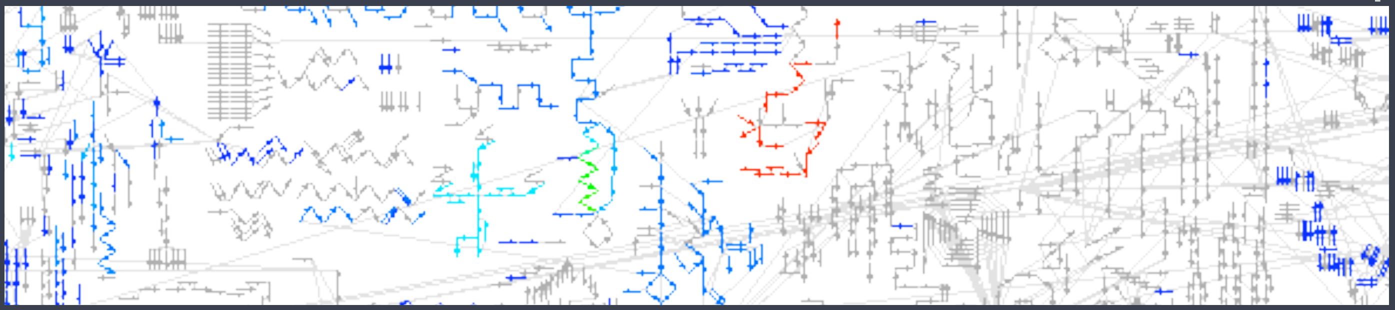
IntAct

MINT

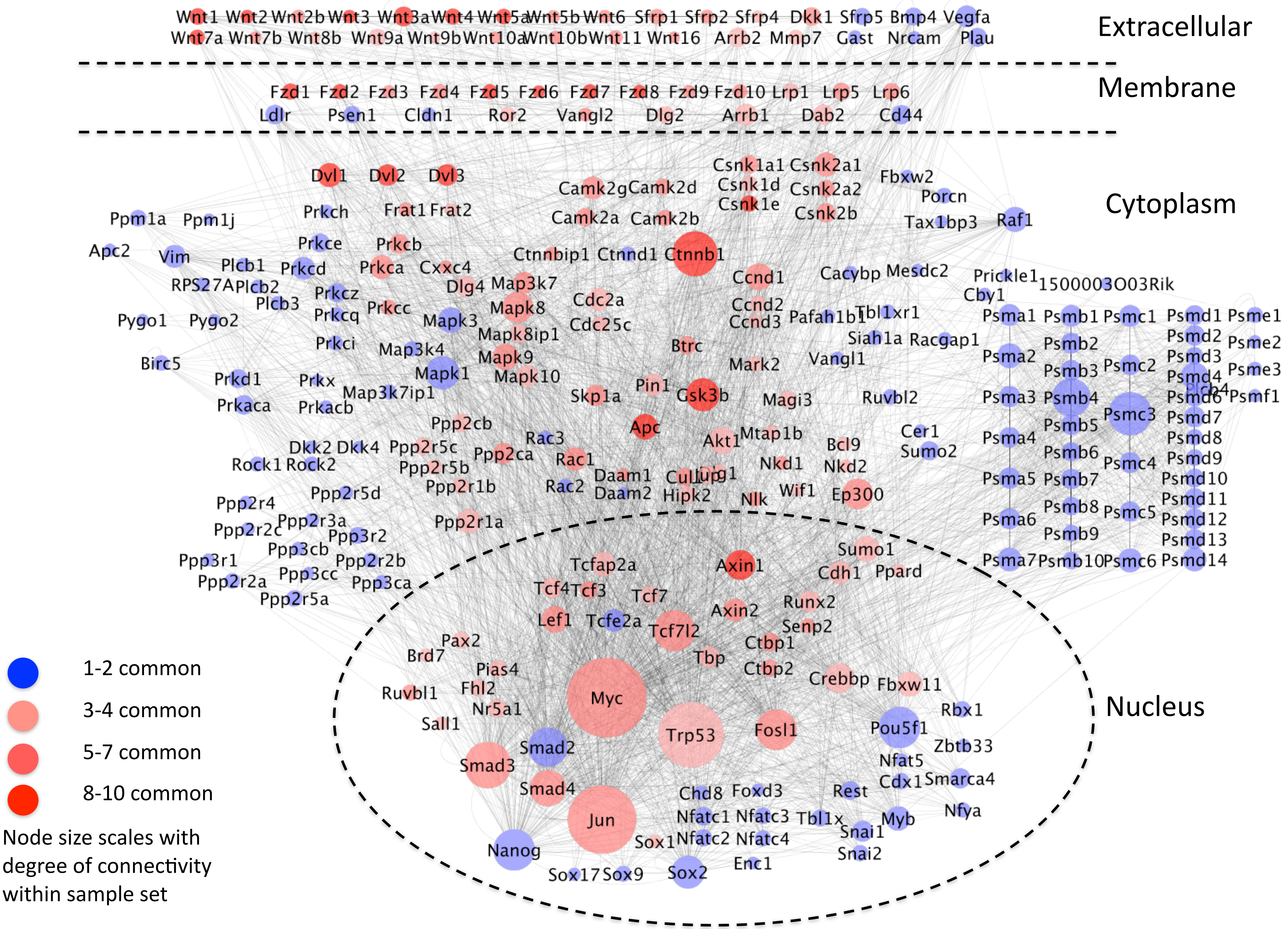
National Cancer Institute nature

Reactome

BETA
WikiPathways
Pathways for the People



Wnt consensus pathway



Node size scales with degree of connectivity within sample set

Nature of pathway annotation

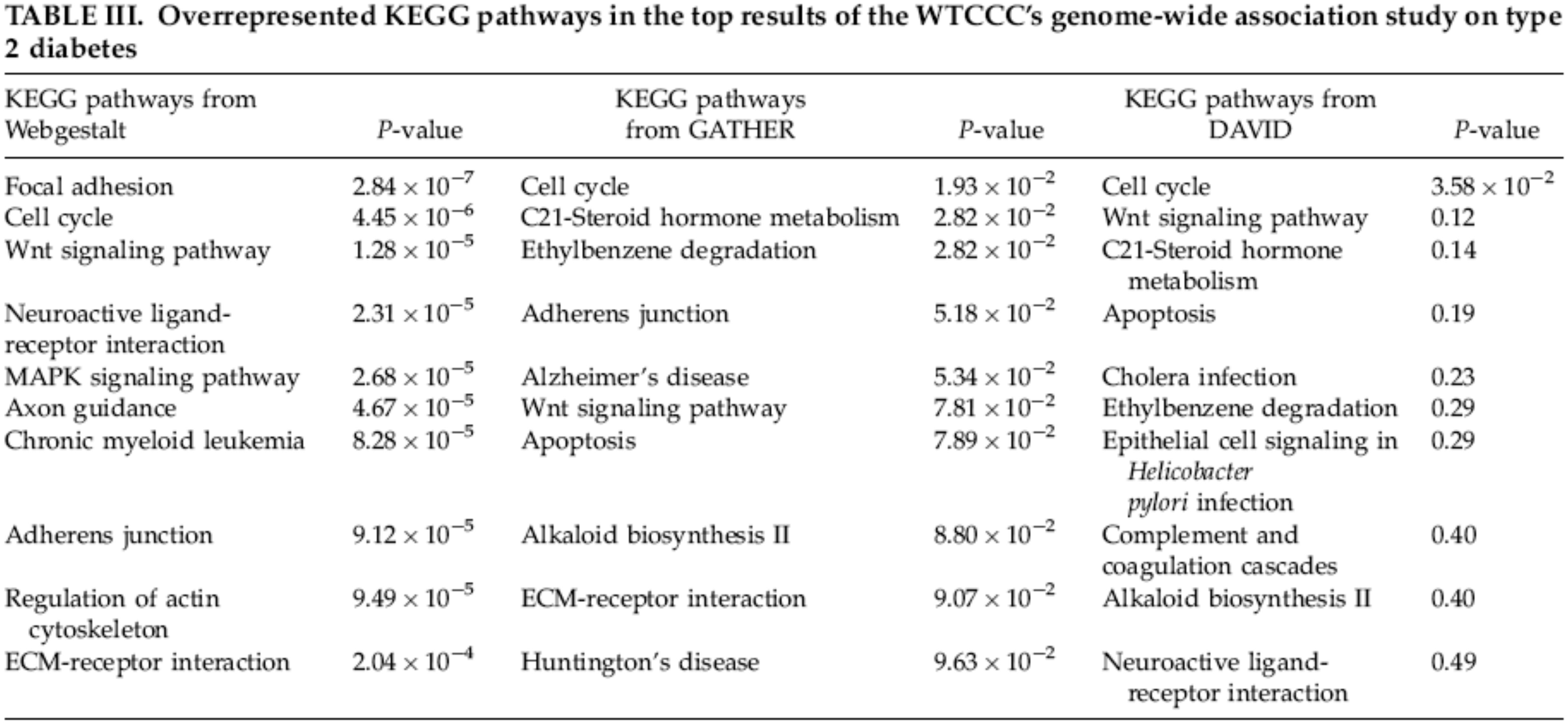


TABLE III. Overrepresented KEGG pathways in the top results of the WTCCC's genome-wide association study on type 2 diabetes

KEGG pathways from Webgestalt	*P*-value	KEGG pathways from GATHER	*P*-value	KEGG pathways from DAVID	*P*-value
Focal adhesion	2.84×10^{-7}	Cell cycle	1.93×10^{-2}	Cell cycle	3.58×10^{-2}
Cell cycle	4.45×10^{-6}	C21-Steroid hormone metabolism	2.82×10^{-2}	Wnt signaling pathway	0.12
Wnt signaling pathway	1.28×10^{-5}	Ethylbenzene degradation	2.82×10^{-2}	C21-Steroid hormone metabolism	0.14
Neuroactive ligand-receptor interaction	2.31×10^{-5}	Adherens junction	5.18×10^{-2}	Apoptosis	0.19
MAPK signaling pathway	2.68×10^{-5}	Alzheimer's disease	5.34×10^{-2}	Cholera infection	0.23
Axon guidance	4.67×10^{-5}	Wnt signaling pathway	7.81×10^{-2}	Ethylbenzene degradation	0.29
Chronic myeloid leukemia	8.28×10^{-5}	Apoptosis	7.89×10^{-2}	Epithelial cell signaling in *Helicobacter pylori* infection	0.29
Adherens junction	9.12×10^{-5}	Alkaloid biosynthesis II	8.80×10^{-2}	Complement and coagulation cascades	0.40
Regulation of actin cytoskeleton	9.49×10^{-5}	ECM-receptor interaction	9.07×10^{-2}	Alkaloid biosynthesis II	0.40
ECM-receptor interaction	2.04×10^{-4}	Huntington's disease	9.63×10^{-2}	Neuroactive ligand-receptor interaction	0.49

Elbers et al 2009 Genetic Epidemiology

Nature of pathway annotation

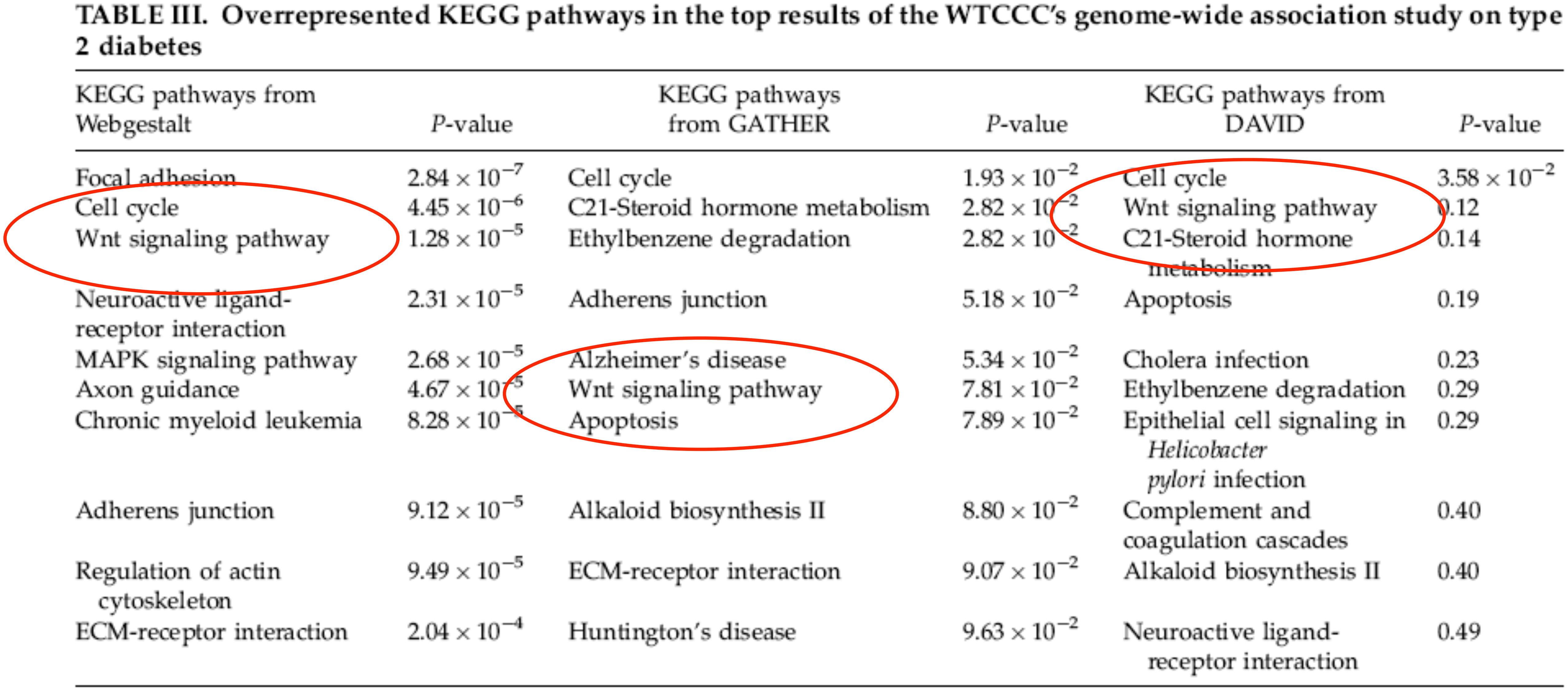


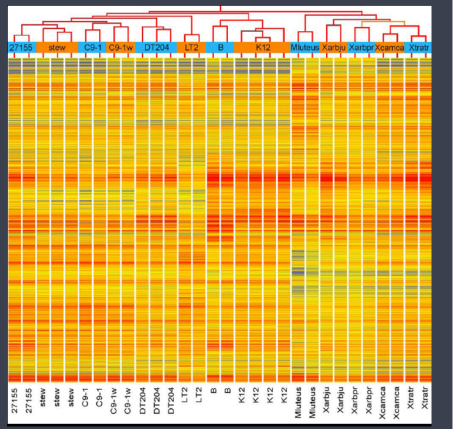
TABLE III. Overrepresented KEGG pathways in the top results of the WTCCC's genome-wide association study on type 2 diabetes

KEGG pathways from Webgestalt	*P*-value	KEGG pathways from GATHER	*P*-value	KEGG pathways from DAVID	*P*-value
Focal adhesion	2.84×10^{-7}	Cell cycle	1.93×10^{-2}	Cell cycle	3.58×10^{-2}
Cell cycle	4.45×10^{-6}	C21-Steroid hormone metabolism	2.82×10^{-2}	Wnt signaling pathway	0.12
Wnt signaling pathway	1.28×10^{-5}	Ethylbenzene degradation	2.82×10^{-2}	C21-Steroid hormone metabolism	0.14
Neuroactive ligand-receptor interaction	2.31×10^{-5}	Adherens junction	5.18×10^{-2}	Apoptosis	0.19
MAPK signaling pathway	2.68×10^{-5}	Alzheimer's disease	5.34×10^{-2}	Cholera infection	0.23
Axon guidance	4.67×10^{-5}	Wnt signaling pathway	7.81×10^{-2}	Ethylbenzene degradation	0.29
Chronic myeloid leukemia	8.28×10^{-5}	Apoptosis	7.89×10^{-2}	Epithelial cell signaling in *Helicobacter pylori* infection	0.29
Adherens junction	9.12×10^{-5}	Alkaloid biosynthesis II	8.80×10^{-2}	Complement and coagulation cascades	0.40
Regulation of actin cytoskeleton	9.49×10^{-5}	ECM-receptor interaction	9.07×10^{-2}	Alkaloid biosynthesis II	0.40
ECM-receptor interaction	2.04×10^{-4}	Huntington's disease	9.63×10^{-2}	Neuroactive ligand-receptor interaction	0.49

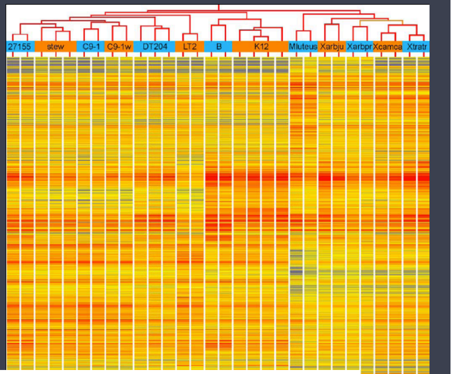
Elbers et al 2009 Genetic Epidemiology

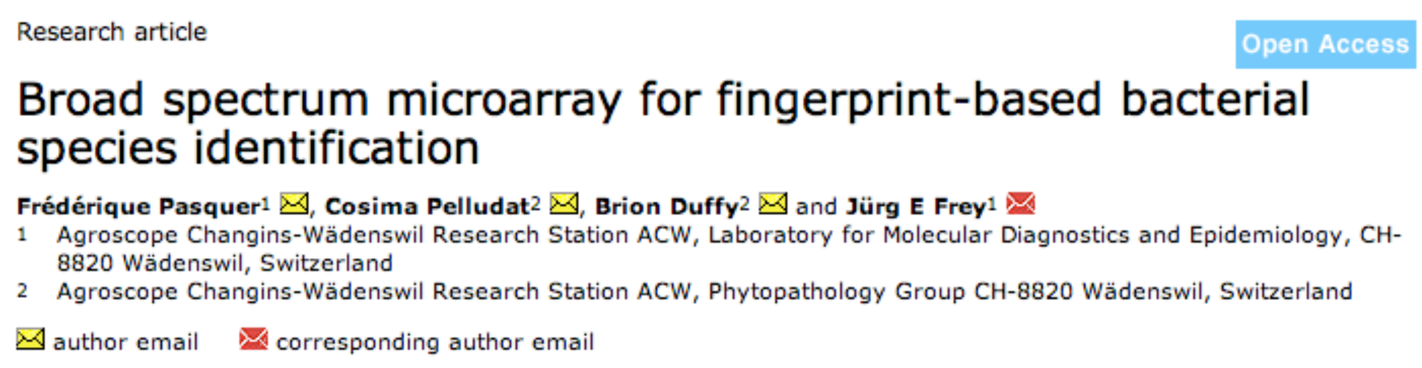
Fingerprints

Fingerprints



Fingerprints





Research article

Open Access

Broad spectrum microarray for fingerprint-based bacterial species identification

Frédérique Pasquer[1] ✉, **Cosima Pelludat**[2] ✉, **Brion Duffy**[2] ✉ and **Jürg E Frey**[1] ✉

1 Agroscope Changins-Wädenswil Research Station ACW, Laboratory for Molecular Diagnostics and Epidemiology, CH-8820 Wädenswil, Switzerland

2 Agroscope Changins-Wädenswil Research Station ACW, Phytopathology Group CH-8820 Wädenswil, Switzerland

✉ author email ✉ corresponding author email

Fingerprints

Research article

Open Access

Broad spectrum microarray for fingerprint-based bacterial species identification

Frédérique Pasquer[1], **Cosima Pelludat**[2], **Brion Duffy**[2] and **Jürg E Frey**[1]

and Epidemiology, CH-

enswil, Switzerland

Metabolic Fingerprints of Altered Brain Growth, Osmoregulation and Neurotransmission in a Rett Syndrome Model

Angèle Viola[1], **Véronique Saywell**[2,3], **Laurent Villard**[2,3], **Patrick J. Cozzone**[1], **Norbert W. Lutz**[1*]

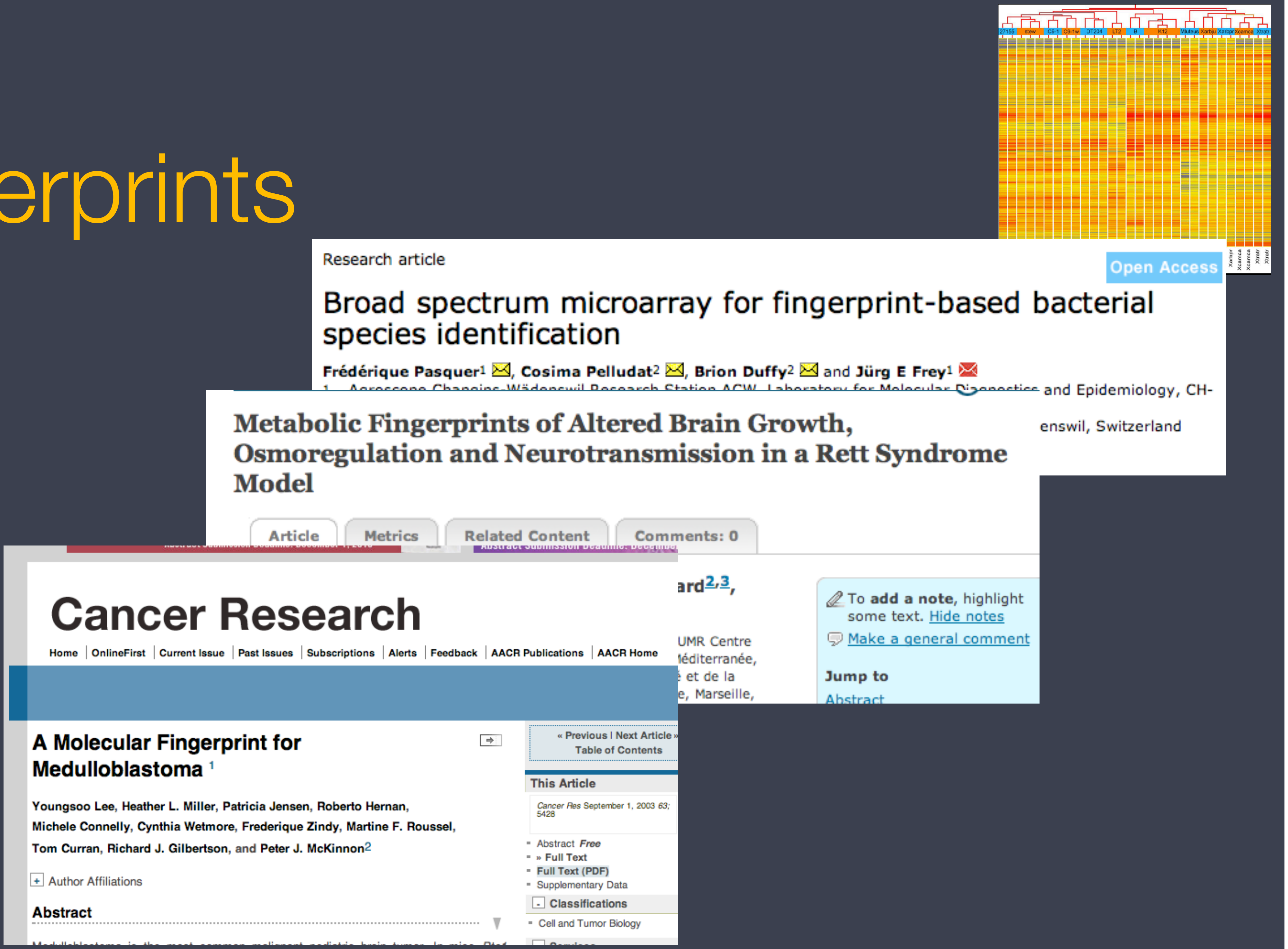
1 Centre de Résonance Magnétique Biologique et Médicale (CRMBM), UMR Centre National de la Recherche Scientifique (CNRS) 6612, Université de la Méditerranée, Faculté de Médecine, Marseille, France, **2** Institut National de la Santé et de la Recherche Médicale (INSERM) U491, Faculté de Médecine de la Timone, Marseille,

To **add a note**, highlight some text. Hide notes

Make a general comment

Jump to

Fingerprints



Research article

Open Access

Broad spectrum microarray for fingerprint-based bacterial species identification

Frédérique Pasquer1, **Cosima Pelludat**2, **Brion Duffy**2 and **Jürg E Frey**1

1 Agroscope Changins-Wädenswil Research Station ACW, Laboratory for Molecular Diagnostics and Epidemiology, CH-

enswil, Switzerland

Metabolic Fingerprints of Altered Brain Growth, Osmoregulation and Neurotransmission in a Rett Syndrome Model

ard2,3,

UMR Centre
éditerranée,
é et de la
e, Marseille,

To add a note, highlight some text. Hide notes

Make a general comment

Jump to

Abstract

Cancer Research

A Molecular Fingerprint for Medulloblastoma 1

Youngsoo Lee, Heather L. Miller, Patricia Jensen, Roberto Hernan, Michele Connelly, Cynthia Wetmore, Frederique Zindy, Martine F. Roussel, Tom Curran, Richard J. Gilbertson, and Peter J. McKinnon2

+ Author Affiliations

Abstract

This Article

Cancer Res September 1, 2003 63; 5428

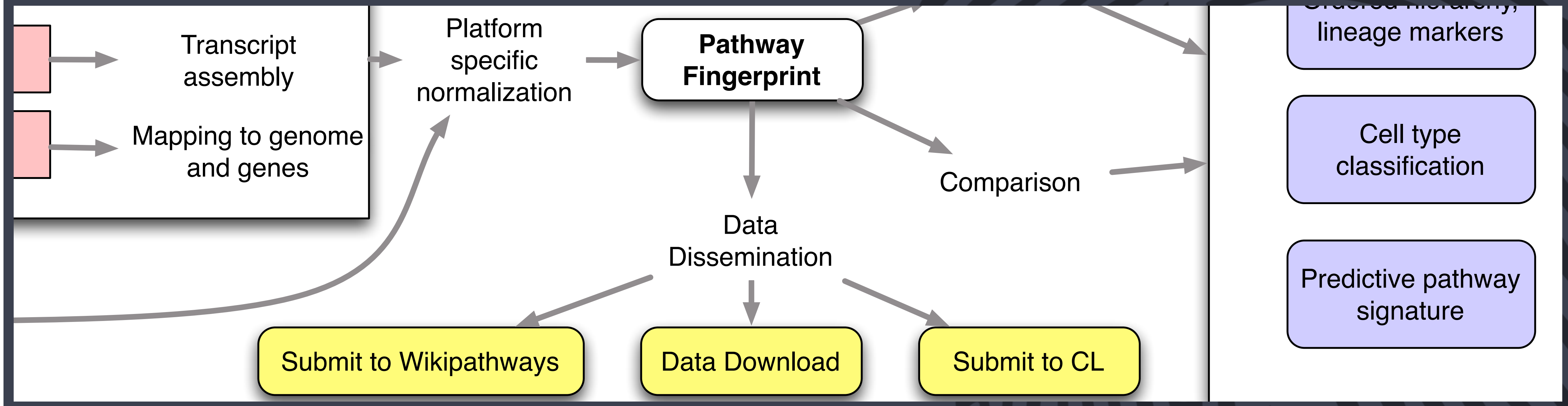
The Pathway Fingerprint

A public resource for classification and discovery of biological function

- common pathways
- common method
- common inerpretation

The Pathway Fingerprint

A public resource for classification and discovery of biological function



Transcript assembly

Mapping to genome and genes

Platform specific normalization

Pathway Fingerprint

Comparison

Data Dissemination

Submit to Wikipathways

Data Download

Submit to CL

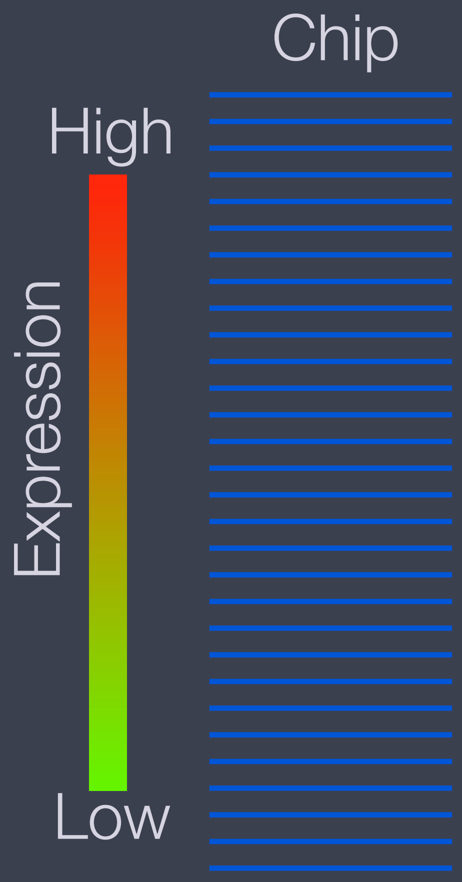
lineage markers

Cell type classification

Predictive pathway signature

Pathway Fingerprint: Method

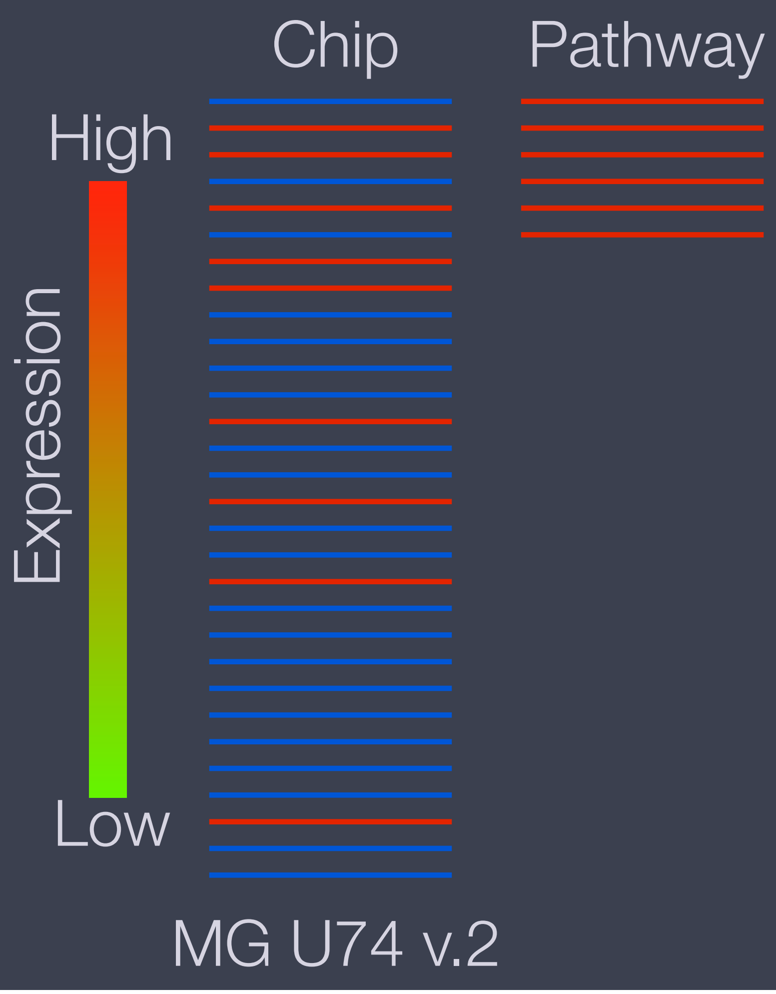
1. Genes ranked by raw expression



e.g. MG U74 v.2

Pathway Fingerprint

2. Identify overlap with pathway



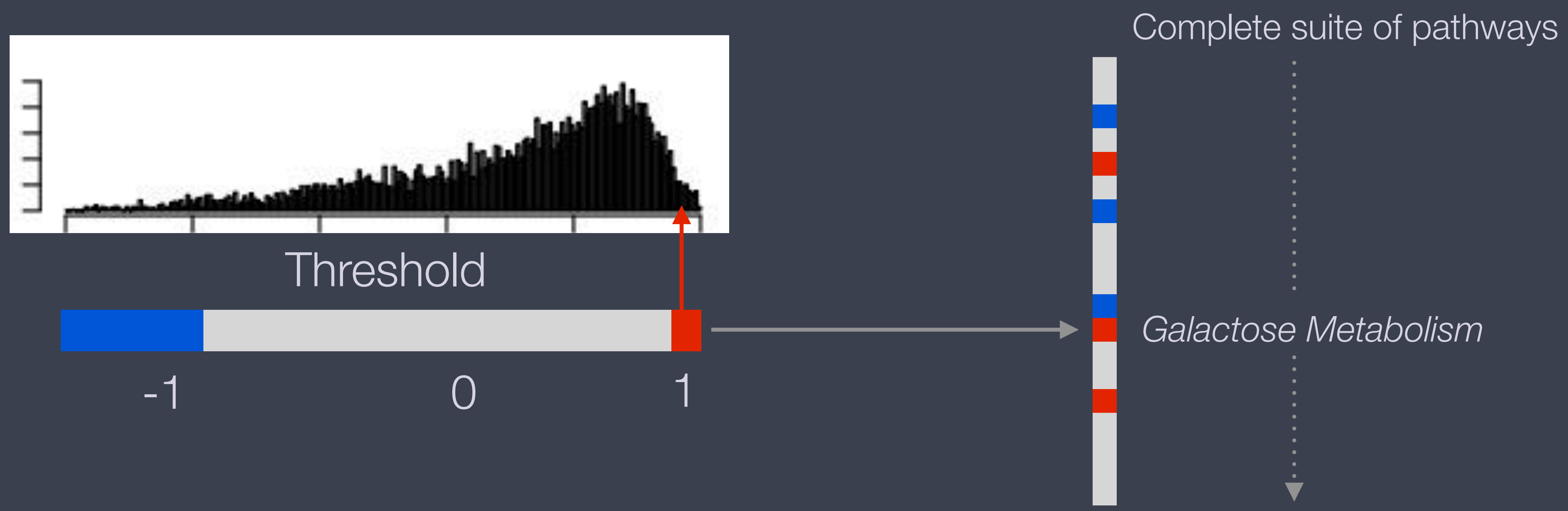
e.g. Galactose metabolism

Pathway Fingerprint

6. Represent as ternary (-1,0,1) vector, compile pathways

MG U74 v.2 - Galactose Metabolism

Pathway fingerprint



Complete suite of pathways

Threshold

-1

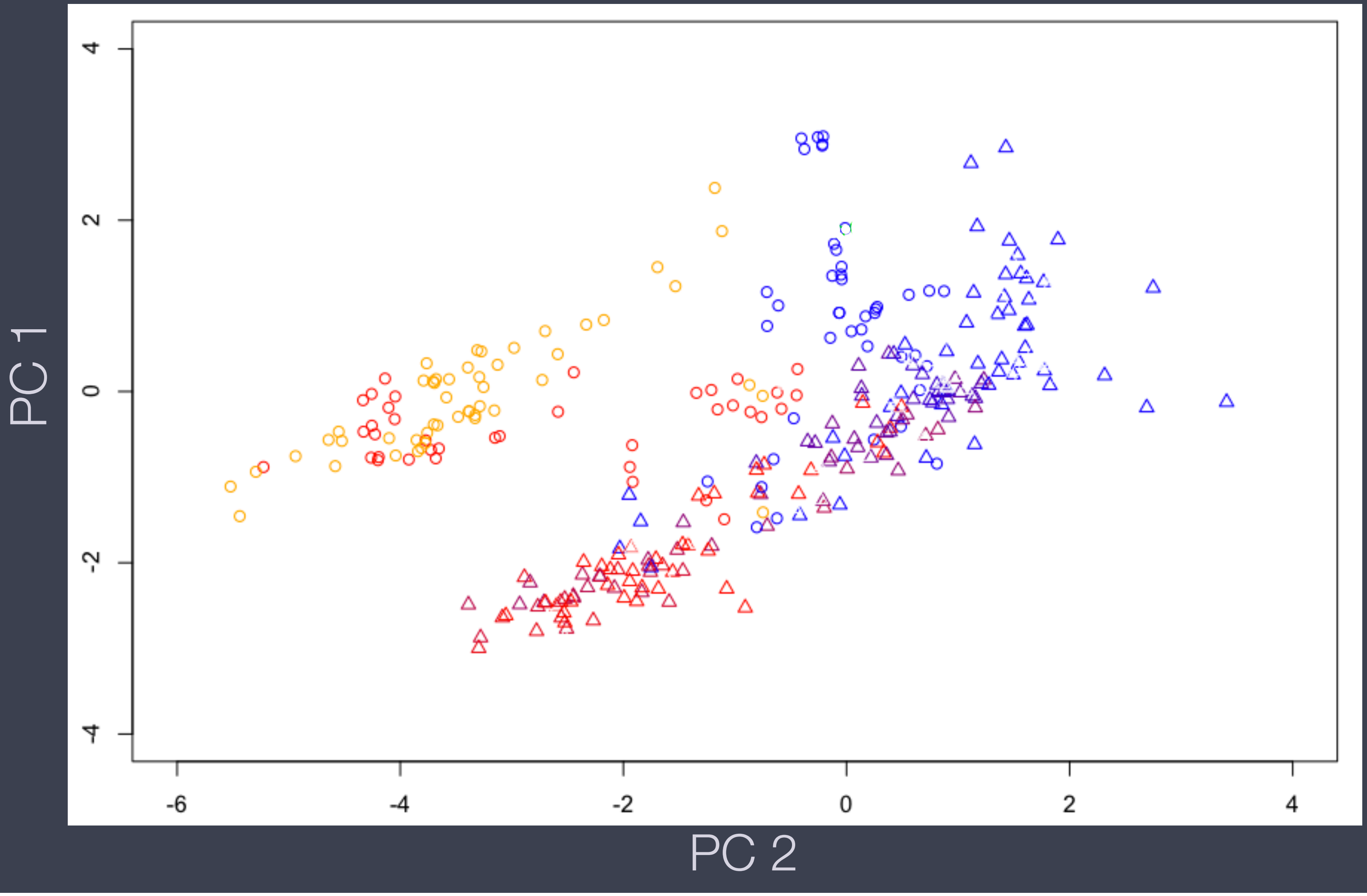
0

1

Galactose Metabolism

Background built from GEO corpus (120,000 chips)

II Multi-study X-species data



iPS

ES

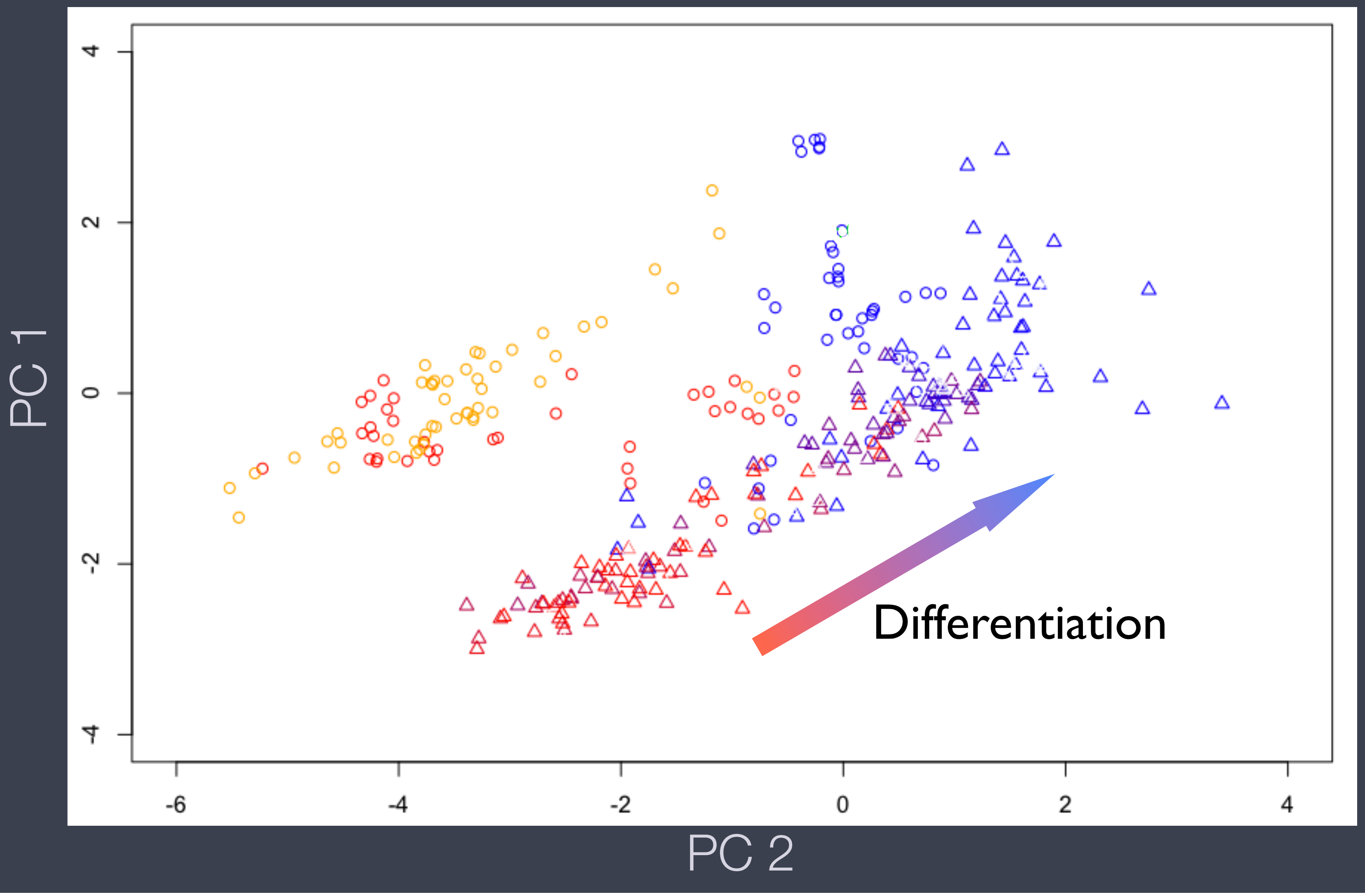
Differentiating ES

Adult differentiated

○ Human

△ Mouse

Cross-species stem cell data



iPS

ES

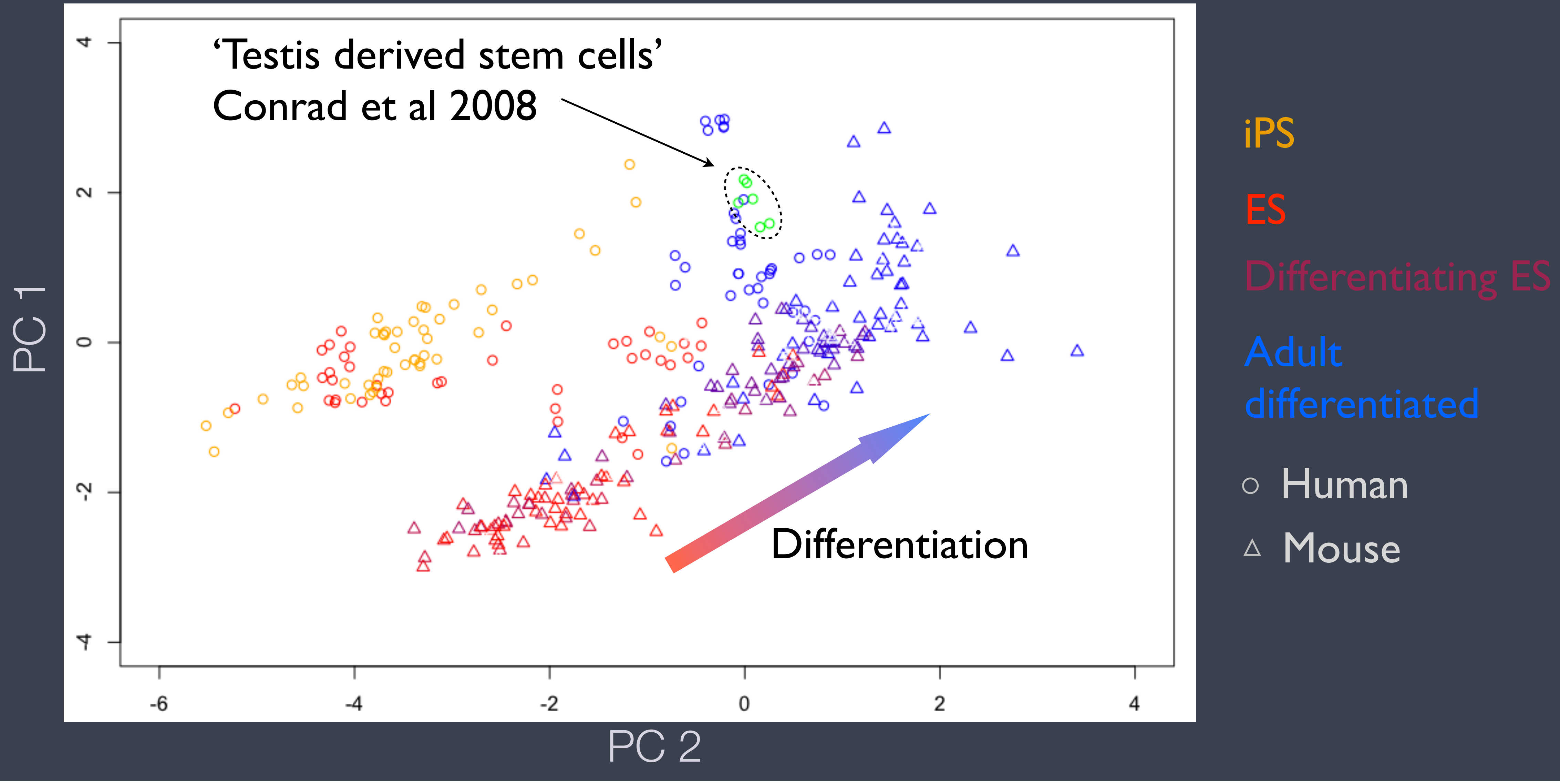
Differentiating ES

Adult differentiated

○ Human

△ Mouse

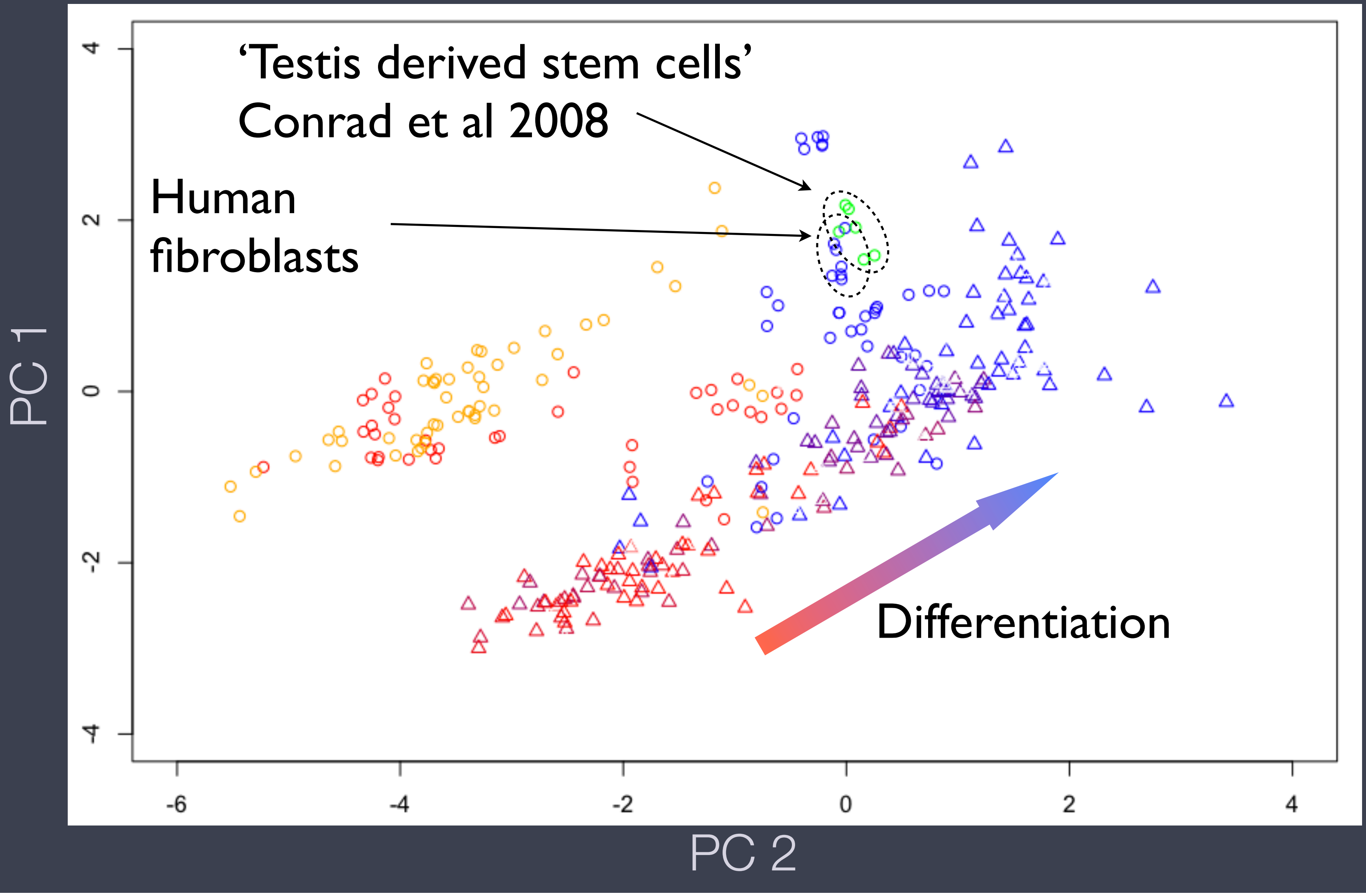
Cross-species stem cell data



‘Testis derived stem cells’
Conrad et al 2008

Differentiation

Cross-species stem cell data



‘Testis derived stem cells’
Conrad et al 2008

Human fibroblasts

Differentiation

iPS

ES

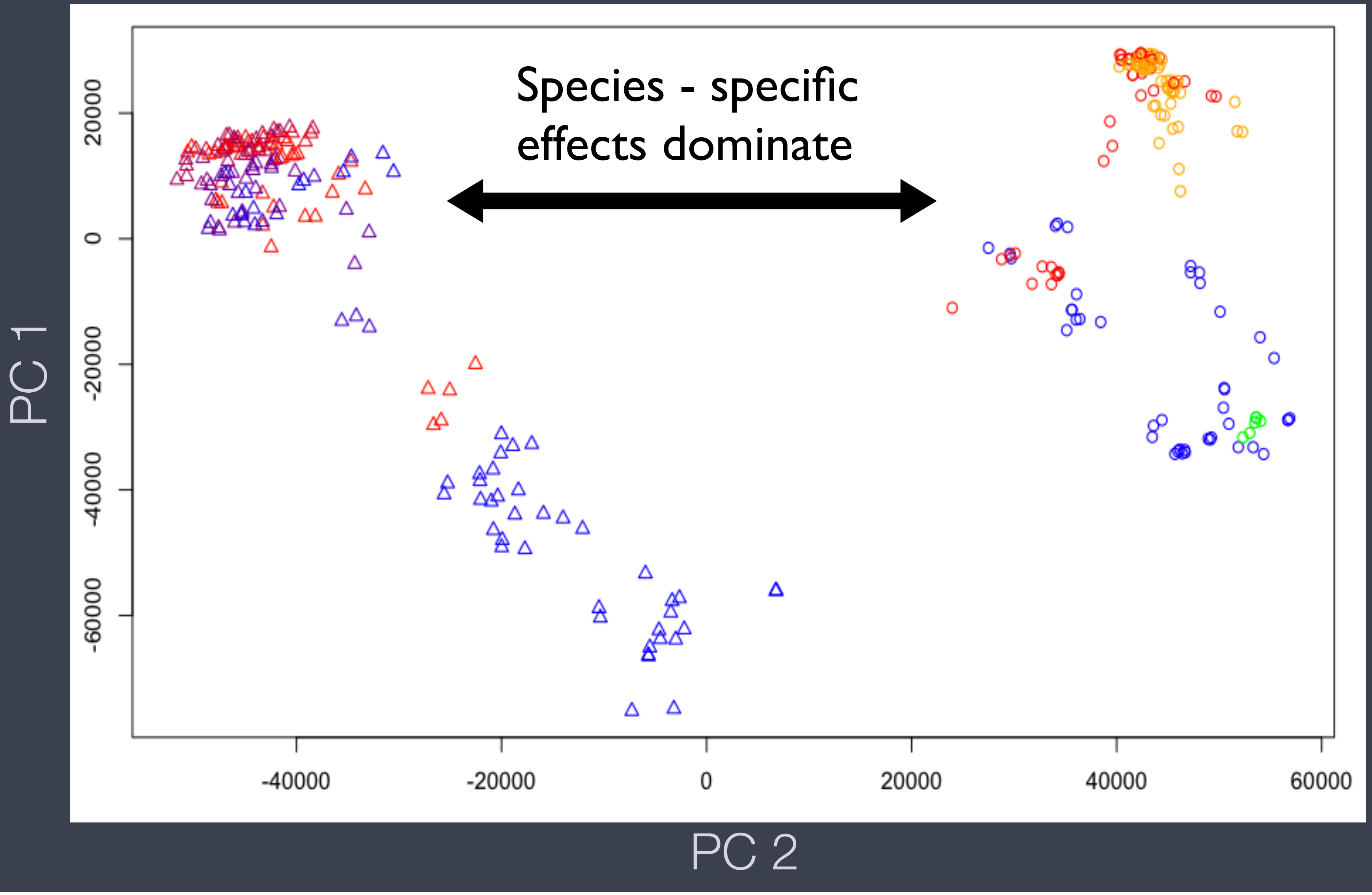
Differentiating ES

Adult differentiated

○ Human

△ Mouse

Comparison with gene clustering



iPS

ES

Differentiating ES

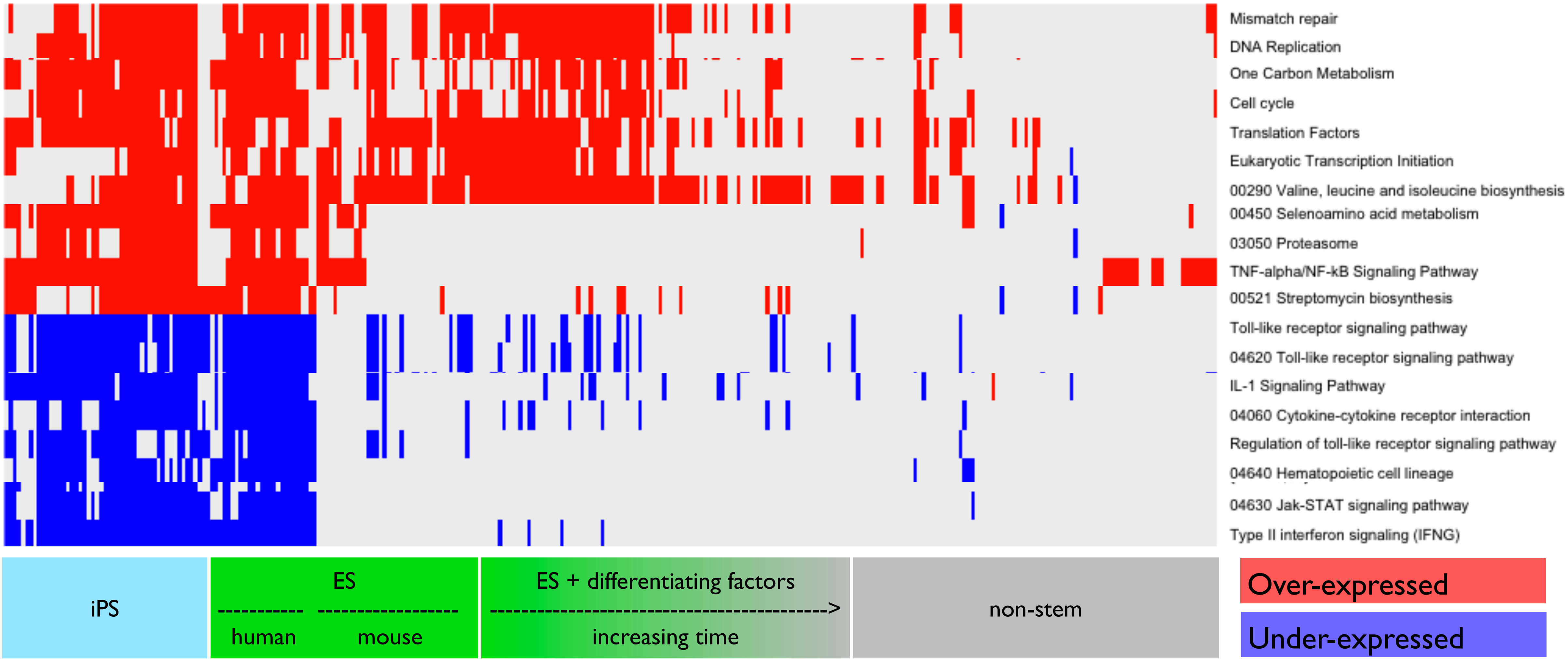
Adult differentiated

○ Human

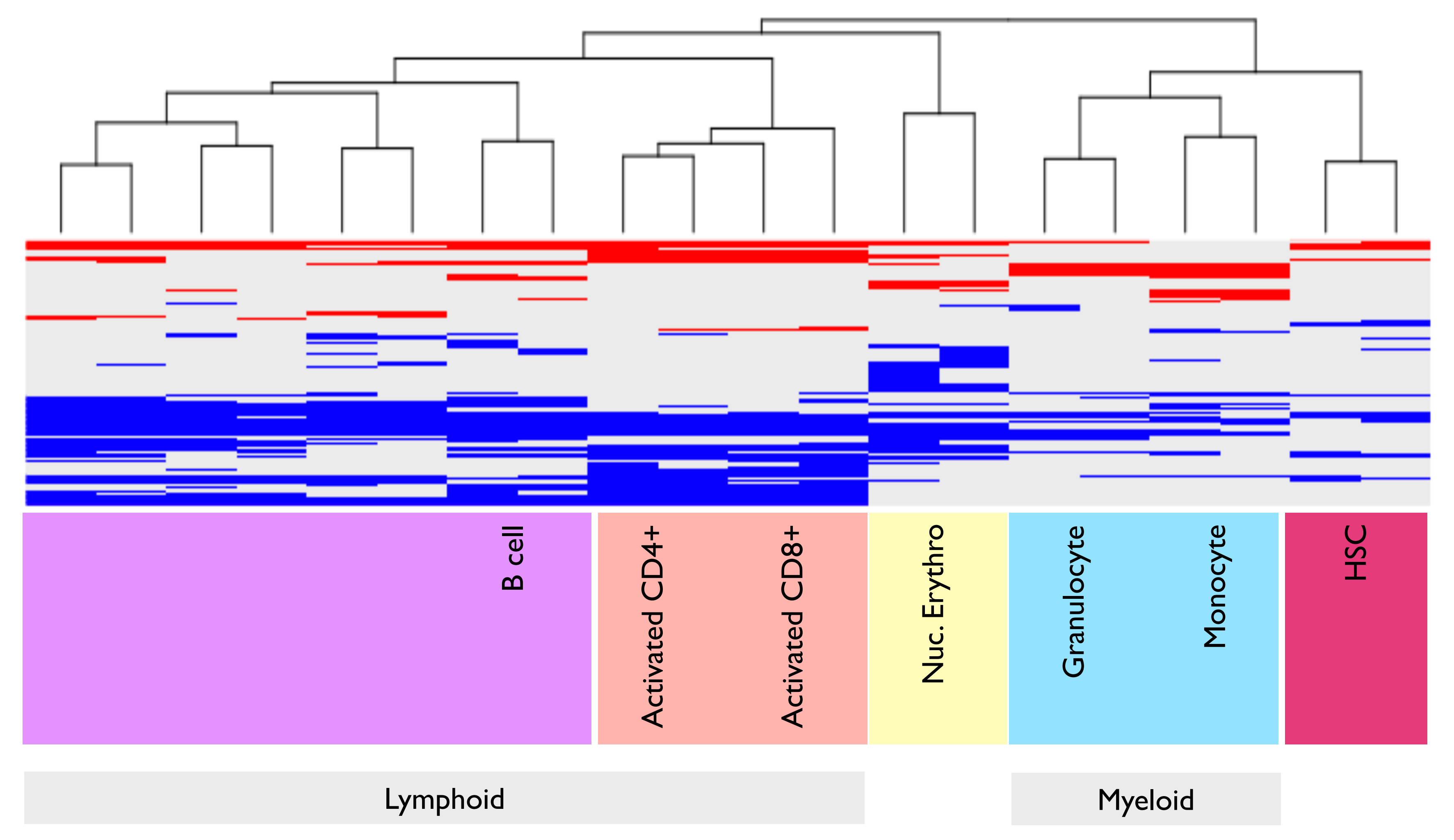
△ Mouse

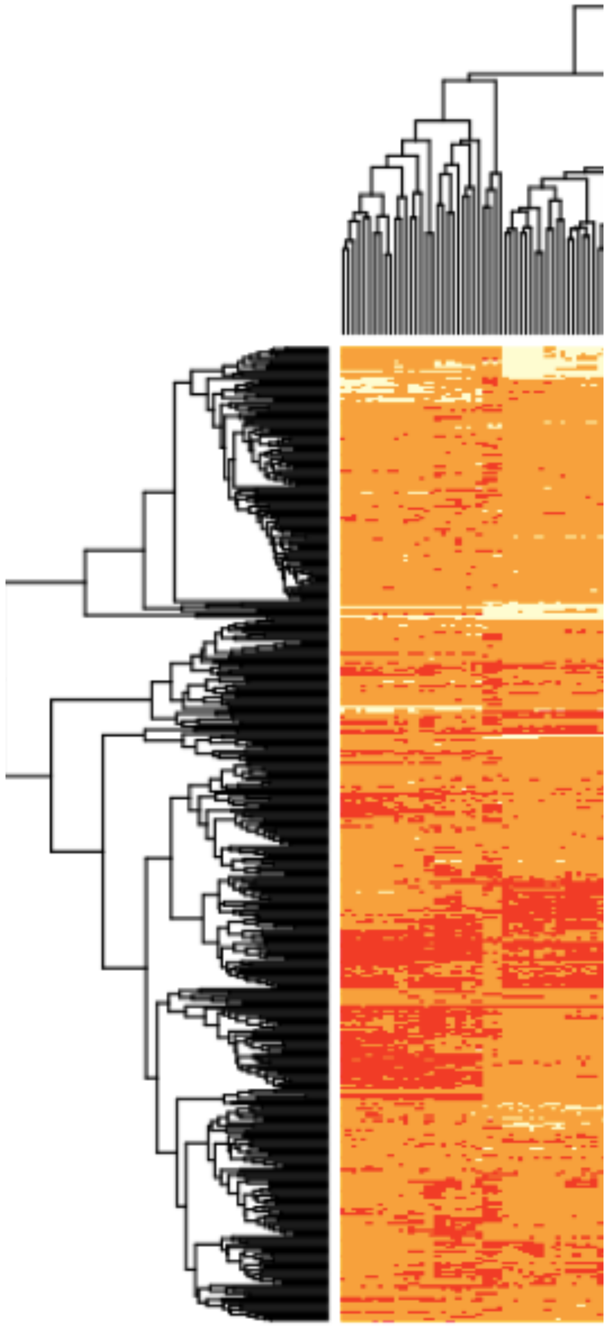
Pluripotent pathway signature

Top 20 pathways



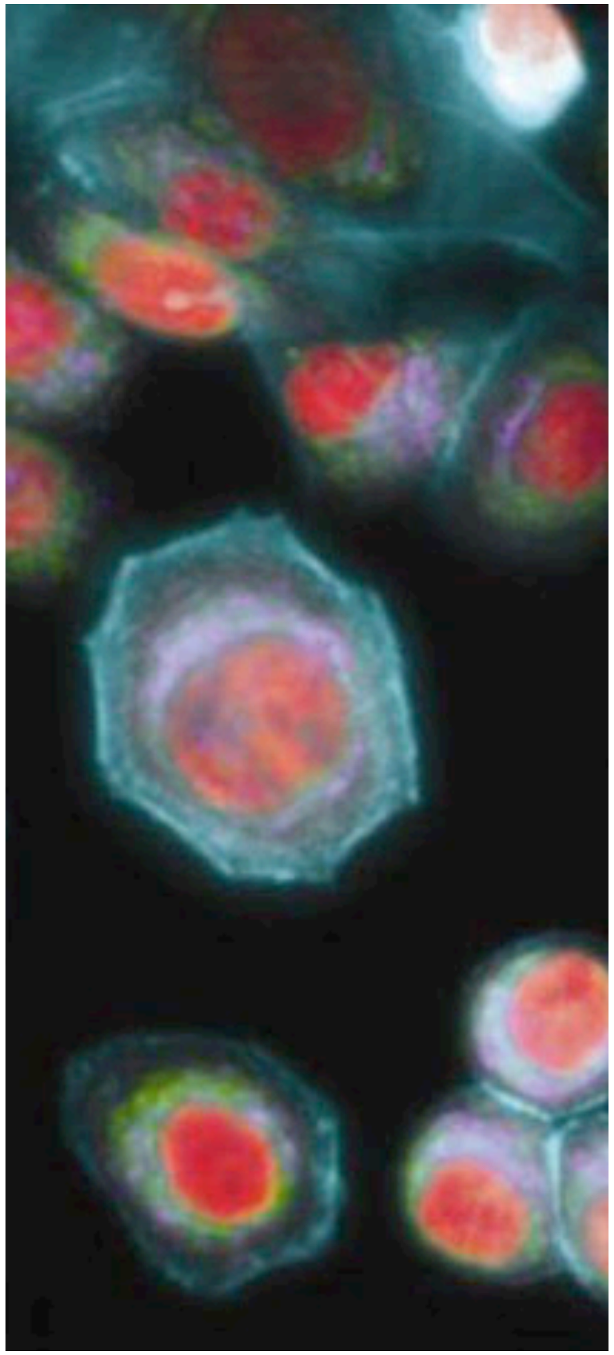
Lineage


B cell
Activated CD4+
Activated CD8+
Nuc. Erythro
Granulocyte
Monocyte
HSC
Lymphoid
Myeloid



From fingerprint to cell definition

Mapping to existing cell descriptions

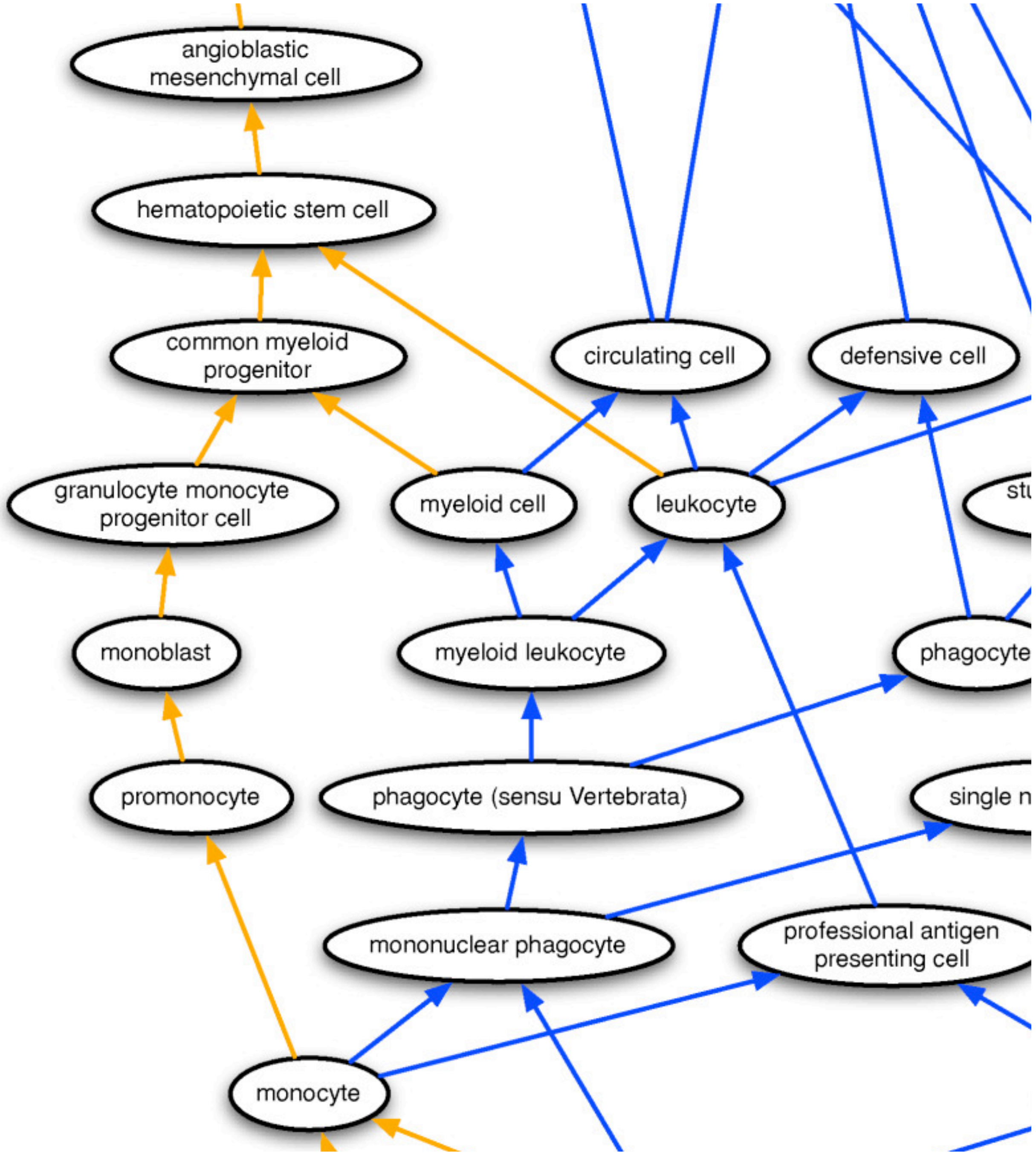


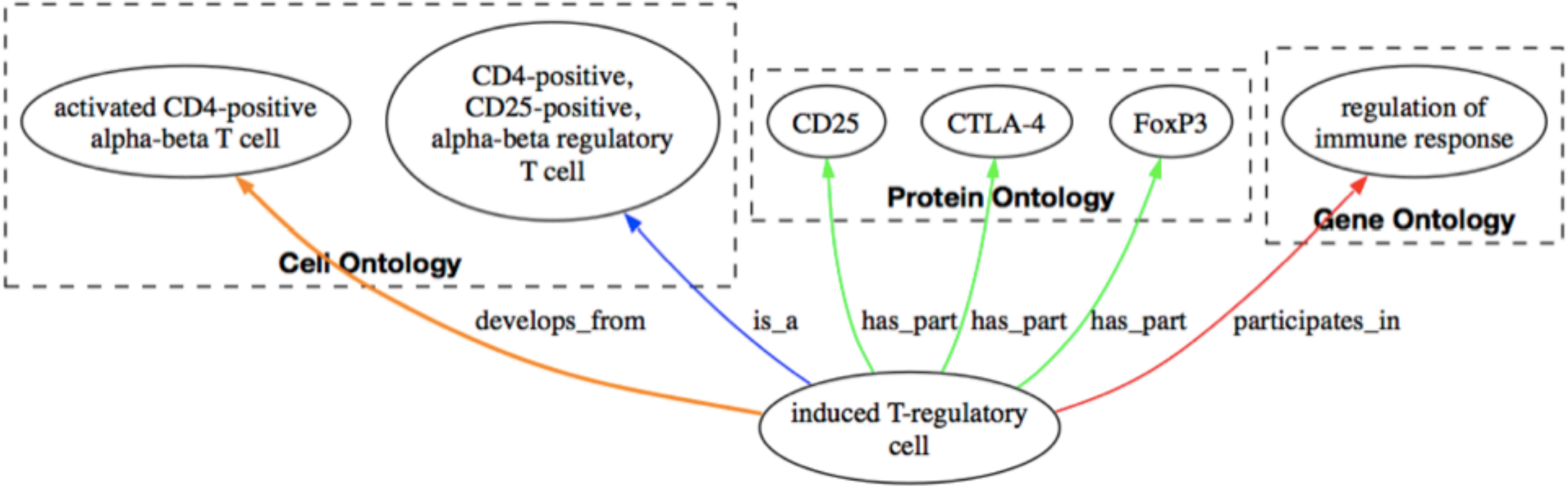
Cell Ontology

Part of OBO Foundry

Actively curated and funded

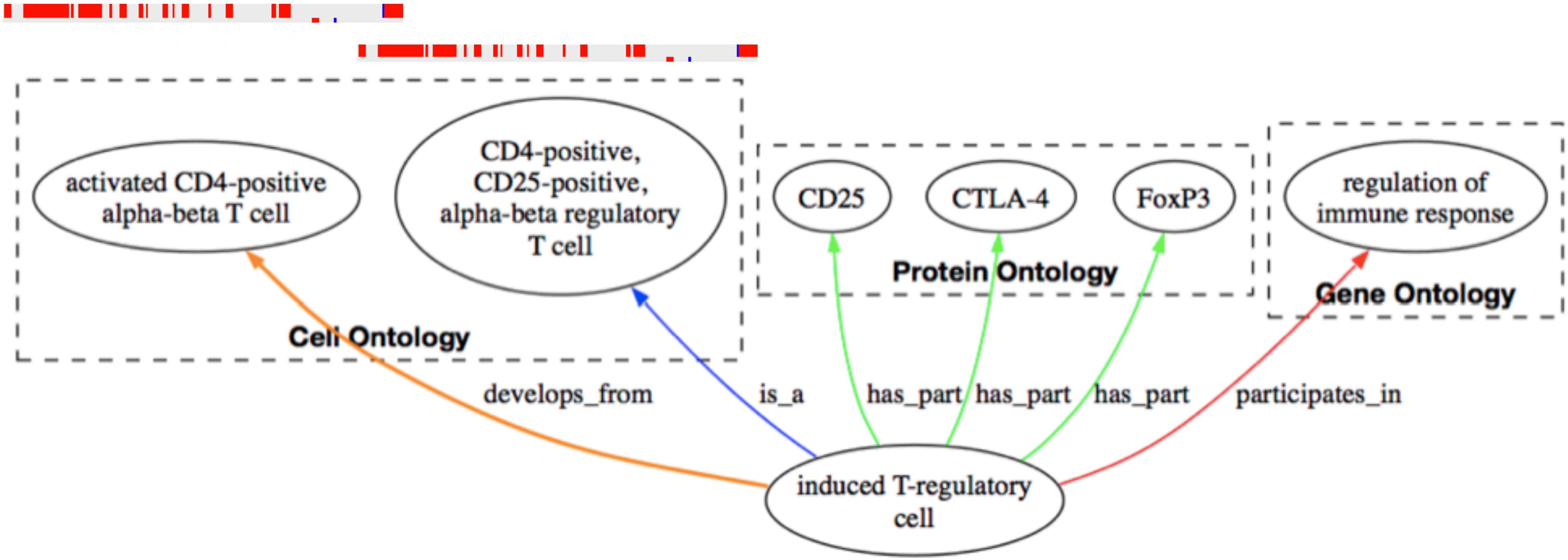
Reorganization to ‘by lineage’ relationships





Logical definitions for cell types

Terry Meehan Alexander Diehl Oliver Hofmann



Logical definitions for cell types

Terry Meehan   Alexander Diehl   Oliver Hofmann

Todos

> private/public expansion

> fingerprint server

> Peering infrastructure

Stem Cell Discovery engine Team

Gabriel Altshuler

Rama Sompellae

Oliver Hofmann

Kimberly Begley