Supplementary File I

Integrated analysis of transcript level regulation of metabolism reveals disease relevant nodes of the human metabolic network

Mafalda Galhardo[1*], Lasse Sinkkonen[1*], Philipp Berninger[2], Jake Lin[3,4], Thomas Sauter[1#] and Merja Heinäniemi[1,5#]

[1]Life Sciences Research Unit, University of Luxembourg, 162a Avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg

[2]Biozentrum, Universität Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, 4056 Basel, Switzerland

[3]Institute for Systems Biology, 401 Terry Avenue North, 98109-5234, Seattle, Washington, USA

[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, House of Biomedicine, 7 Avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg

[5]A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, FI-7120 Kuopio, Finland

[*]These authors contributed equally

[#]To whom correspondence should be addressed

Table of Content

Supplementary Table Legends

Table S1: Primer sequences for RT-qPCR and ChIP-qPCR. The primer sequences used in PCR reactions of the validation experiments are listed here.

Table S2: ChIP-seq peaks identified for PPARγ, CEBPα and LXR in day 10 differentiated SGBS cells and in a comparable analysis of SRX032890 and SRX019521data sets. The number of reads that passed through each processing step is indicated for the SGBS data obtained here and the public raw read data processed from SRX032890 and SRX019521. The ChIP-seq peaks together with enrichment quantification and statistical significance values as identified using QuEST tool (35) are presented for each data set. The same analysis settings were applied to generate highly comparable data. The peaks that pass the enrichment threshold >30, chosen here to distinguish high-occupancy binding, are highlighted in bold. Notice that a separate sheet exists in the xls file for each data set.

Table S3: Peak to gene association and ontology term enrichment analysis for ChIP-seq data sets. The ChIP-seq peak coordinates from Table S2 were used as input for the GREAT tool (36) that first associates the peaks to putative target genes listed. Significant ontology terms were collected and highlight the role of these TFs in lipid and carbohydrate metabolism.

Table S4: HUVEC TF and disease association result. Recon1 metabolic genes are shown in context of the number of associated diseases and TFs (from 10 ChIP-Seq studies on HUVEC), detailed analysis description is found on Materials and Methods. Data were sorted by gene relevance for endothelial disease, by the number of associated TFs and by the H3K4me3 active transcription mark. The number of diseases the gene is associated to is based on DisGeNET database (32).

Table S5: Differentially expressed metabolic genes during SGBS differentiation. The average logarithmic fold change values and statistical analysis including t-test for individual time points and F-test results across all time points is presented for differentially expressed metabolic genes (based on Recon1 (2)).

Table S6: Comparison on reaction activity predictions for pre-adipocytes and adipocytes. Metabolic changes resulting from human SGBS pre-adipocyte cell differentiation were qualitatively predicted from gene expression data using an implementation of the constraint-based method from (6). The 323 reactions with predicted reaction activity change are highlighted.

Table S7: Naming of metabolic genes and enzymes from selected pathways. The complete names for the metabolites and enzymes included in the pathway figures are presented.

Table S8: Differentially expressed TF genes during SGBS differentiation. The average logarithmic fold change values and statistical analysis including t-test for individual time points and F-test results across all time points is presented for differentially expressed TF genes.

Table S9: Genes with altered H3K4me3 status during SGBS differentiation. The H3K4me3 histone marker quantification from -1250 to +750 bp around gene TSS is presented for metabolic genes that changed their H3K4me3 status in SGBS cells, including the respective data from primary adipocytes (10).

Table S10: Target gene associations for miR-27a, miR-29a and miR-222 based on combined microarray target profiling and heptamer motif analysis. Genes identified to be responsive to miRNA mediated regulation from SGBS array profiling experiments following miRNA over-expression and 3'-UTR motif analysis are listed (see Methods for details). Notice the separate data sheets for each miRNA.

Table S11: Differentially expressed genes in 4 h LXR agonist T0901317 stimulated SGBS adipocytes. The average logarithmic fold change values and statistical analysis is presented for differentially expressed genes upon ligand activation of LXRs.

Table S12: Additional data supporting LXR peak to gene associations. Data in support of LXR mediated regulation of the genes associated with LXR peaks is presented collected from own microarrays (see Table S11) and GSE35262.

Table S13: Hypergeometric test for enriched pathway terms among genes regulated by SREBF1. Recon1 pathway enrichment results from a hypergeometric test on genes reported as SREBF1 targets on muscle. Cholesterol pathway ranked first, followed by oxidative phosphorylation and fatty acid activation and elongation.

Supplementary Figure Legends

Fig. S1: Integrated metabolic pathways of arginine and proline metabolism in HUVECs. The complete argine-proline metabolism pathway that contains the top disease associated gene *NOS3* is shown. Regulatory associations from ten ChIP-seq studies (as in Fig. 1) are displayed in the gene metanodes. Among genes involved the initial steps of the pathway, *ALDH4A1, ALDH2* and *MTAP*, represent genes associated with endothelial relevant disease and with multiple TFs. TF association is indicated with filled circles. The genes discussed further in the text are shown as larger metanodes for clarity.

Fig. S2: Time series expression profile of metabolic genes during SGBS differentiation. The average logarithmic fold change values from 4, 8 and 12 h and days 1, 3 and 12 are displayed in color using GEDI maps (25) to cluster metabolic genes with similar expression profiles. Initially, the responses seen are modest shifting to more prominent up- and down-regulation by day 3 with the largest changes observed at day 12. The number of genes in each cluster is displayed in the Gene density panel below. To distinguish pathway dynamics, overrepresented metabolic pathways among significantly regulated genes are listed beside each map. The clustering of sample replicates and the separation between the time points using AutoSOME (26) is illustrated in the figure inset, in agreement the day 12 samples separate most from the other time points.

Fig. S3: Time series expression profile of selected GO categories during SGBS differentiation. The average logarithmic fold change values from 4, 8 and 12 h and days 1, 3 and 12 are displayed in color using GEDI maps (25) as in Fig. S2 from other functionally related genes for comparison. All genes in the HT12 Illumina array (**A**), or genes from the GO categories cell projection, envelope, locomotion and receptor activity (respectively **B**, **C**, **D** and **E**) having similar number of genes as Recon1 were selected to show gene expression changes. Focusing on day 12, several up- and downregulated clusters relative to 4 h can be observed, however not as prominent as observed for metabolic genes based on color intensity or the percentage of significantly differentially expressed genes (adjusted F-test p-value <0.01, absolute log2 fold change >1) indicated below the panels.

Fig. S4: MiRNA expression profiling by microarrays reveals down-regulation of several miRNA clusters during adipocyte differentiation. Total RNA samples from time points day 0, day 1, day 3 and day 12 of SGBS differentiation time series were used to profile miRNAs using miChip arrays (v.11.0) arrays (27) containing probes for all miRNAs from miRBase version 11.0. In order to identify miRNAs that could contribute to prevailing upregulation of mRNAs during adipogenesis, the analysis is focused on down-regulated miRNAs. Bar graph depicts all miRNAs that have a normalized expression signal of > 50 at time point day 0, that become early down-regulated > 1.25-fold on day 1 and day 3 of differentiation and that remain down-regulated > 1.5-fold on day 12 of differentiation. The measured expression values were median normalized and are shown relative to undifferentiated cells, value of which was set to 1 (light grey bars). Data points indicate the mean expression values of triplicate experiments and the error bars represent SD. No statistical analysis was applied due to large variation between separate array hybridizations following the median normalization. The miRNA clusters with multiple downregulated mature miRNAs and early downregulation profile were selected for further analysis and are indicated with a black bar.

Fig. S5: To illustrate the separation of the measured values between the signal and noise distributions the model fits are shown for SGBS pre-adipocytes (**A**) and adipocytes (**B**). At the overlapping region, genes remain unassigned.

Fig. S6: Negative and positive control regions for ChIP-seq validation experiments. The ChIP-seq signal tracks from PPARγ studies in SGBS cells and primary adipocytes (10, 41), CEBPα and LXR from SGBS adipocytes and H3K4me3 from primary and SGBS cells comparing pre-adipocytes and adipocytes are shown from a 200 kb region centered at TSS regions of *CDH1* (negative control) (**A**) and *FABP4* (positive control) (**B**) regions. Regions with high enrichment for one or several TFs from each locus were selected for validation by ChIP-qPCR (indicated by a lined box and numbered for each locus). Each region was tested for enrichment using antibodies against all three TFs and IgG as a control as is shown in adjacent plots for the regions indicated on the ChIP-seq tracks. The enrichment values are shown relative to the enrichment of IgG and indicate the mean enrichment values of triplicate experiments and the error bars represent SEM. One sample t-test was performed to determine the significance of TF enrichment compared to IgG (*, $p < 0.05$; **, $p < 0.01$).

Fig. S7: Feed-forward loops based on ChIP-seq data. The regulatory connections to the SREBF1 locus as identified from high-occupancy ChIP-seq regions and qPCR validation experiments are shown. The arrows represent the directionality of regulation (regulated by), and the sign of regulation is indicated if inferred from data (shown here only for LXR based on the microarray data).

Fig. S8: Integrated metabolic pathways of fatty acid oxidation and activation. The fatty acid oxidation (A) and activation (B) pathways from Recon1 (2) are shown. The metanodes composition and edge color are identical to those in Figure 4. **A**) Fatty acids are broken down in the mitochondria to acetyl-CoA and a two-carbons shorter acyl-CoA, through β-oxidation. The figure represents the fatty acid oxidation in a simplified manner, where each fatty acyl-CoA is directly or via octanoyl-CoA (occoa[m]) oxidized to acetyl-CoA (accoa[m]). The pathway is largely predicted to shift to active in adipocytes (red edges). Two genes are controlling these reactions, *ACADS* and *ACADM*, that encode acyl-CoA dehydrogenases for short and medium chain fatty acids, respectively. *ACADM* is associated to PPARγ, CEBPα and miR-222. **B**) The fatty acid activation is shown. Fatty acids need to be esterified to coenzyme A (CoA) in order to be metabolically processed (oxidative degradation, elongation into complex lipids or attached to proteins as lipid anchors), catalyzed by fatty acyl CoA synthetases (ACSs). The pathway is

largely predicted to be activated in adipocytes (red edges) and *ACSL1* is among the top genes associated with high-occupancy binding sites for all three TFs.

Fig. S9: Integrated metabolic pathway of branched chain amino acid metabolism with HUVEC data. The BCAA metabolism pathway as in Figure 6 is shown for HUVEC data. Association with 8 or more TFs is highlighted and these nodes include five of the transporters including *SLC7A5* and genes from upstream reactions catalyzed by *BCAT1* and the branched-chain α-keto acid dehydrogenase complex that overlap highly regulated nodes in SGBS.

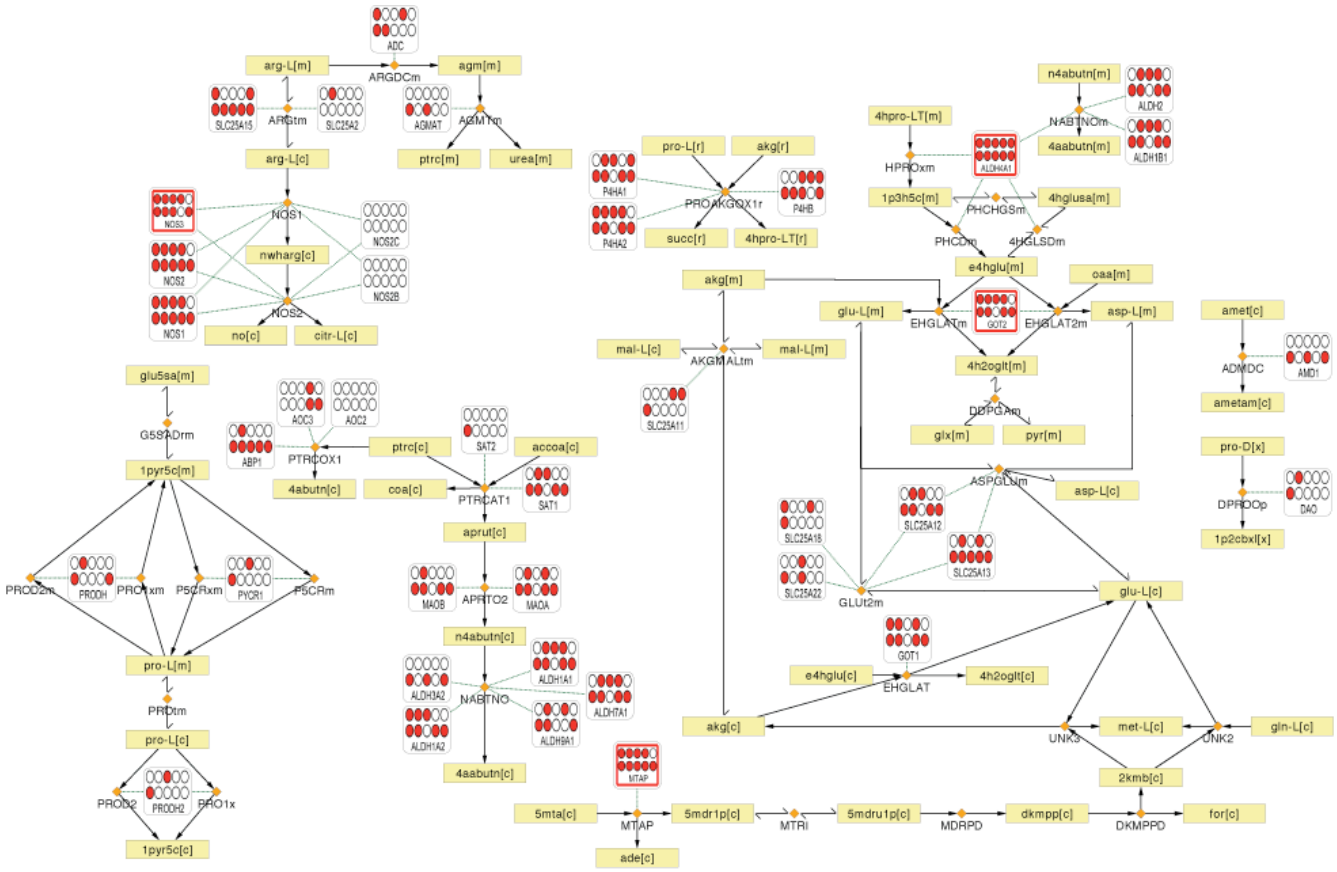

ADC
arg-L[m]
ARGDCm
agm[m]
SLC25A15
ARGtm
SLC25A2
AGMAT
AGMTm
ptrc[m]
urea[m]
arg-L[c]
NOS3
NOS1
NOS2C
NOS2
NOS2B
NOS1
nwharg[c]
NOS2
no[c]
citr-L[c]
pro-L[r]
akg[r]
P4HA1
P4HA2
PROAKGOX1r
P4HB
succ[r]
4hpro-LT[r]
4hpro-LT[m]
HPROxm
ALDHA1
n4abutn[m]
ALDH2
NABTNOm
4aabutn[m]
ALDH1B1
1p3h5c[m]
PHCHGSm
4hglusa[m]
PHCDm
4HGLSDm
e4hglu[m]
caa[m]
akg[m]
glu-L[m]
EHGLATm
GOT2
EHGLAT2m
asp-L[m]
mal-L[c]
AKGMALtm
mal-L[m]
SLC25A11
4h2oglt[m]
DBPGAm
glx[m]
pyr[m]
ASPGLUm
asp-L[c]
amet[c]
ADMDC
AMD1
ametam[c]
pro-D[x]
DPROOp
DAO
1p2cbxl[x]
SLC25A18
SLC25A22
GLUt2m
SLC25A12
SLC25A13
glu-L[c]
GOT1
e4hglu[c]
EHGLAT
4h2oglt[c]
akg[c]
UNK3
met-L[c]
UNK2
gln-L[c]
2kmb[c]
glu5sa[m]
G5SADrm
1pyr5c[m]
PROD2m
PRODH
PROTxm
P5CRxm
PYCR1
P5CRm
pro-L[m]
PRDtm
pro-L[c]
PROD2
PRODH2
PRO1x
1pyr5c[c]
AOC3
AOC2
ABP1
PTRCOX1
4abutn[c]
ptrc[c]
coa[c]
SAT2
accoa[c]
PTRCAT1
SAT1
aprut[c]
MAOB
APRTO2
MAOA
n4abutn[c]
ALDH3A2
ALDH1A2
NABTNO
ALDH1A1
ALDH7A1
ALDH9A1
4aabutn[c]
MTAP
5mta[c]
MTAP
ade[c]
5mdr1p[c]
MTRI
5mdru1p[c]
MDRPD
dkmpp[c]
DKMPPD
for[c]

Fig. S1

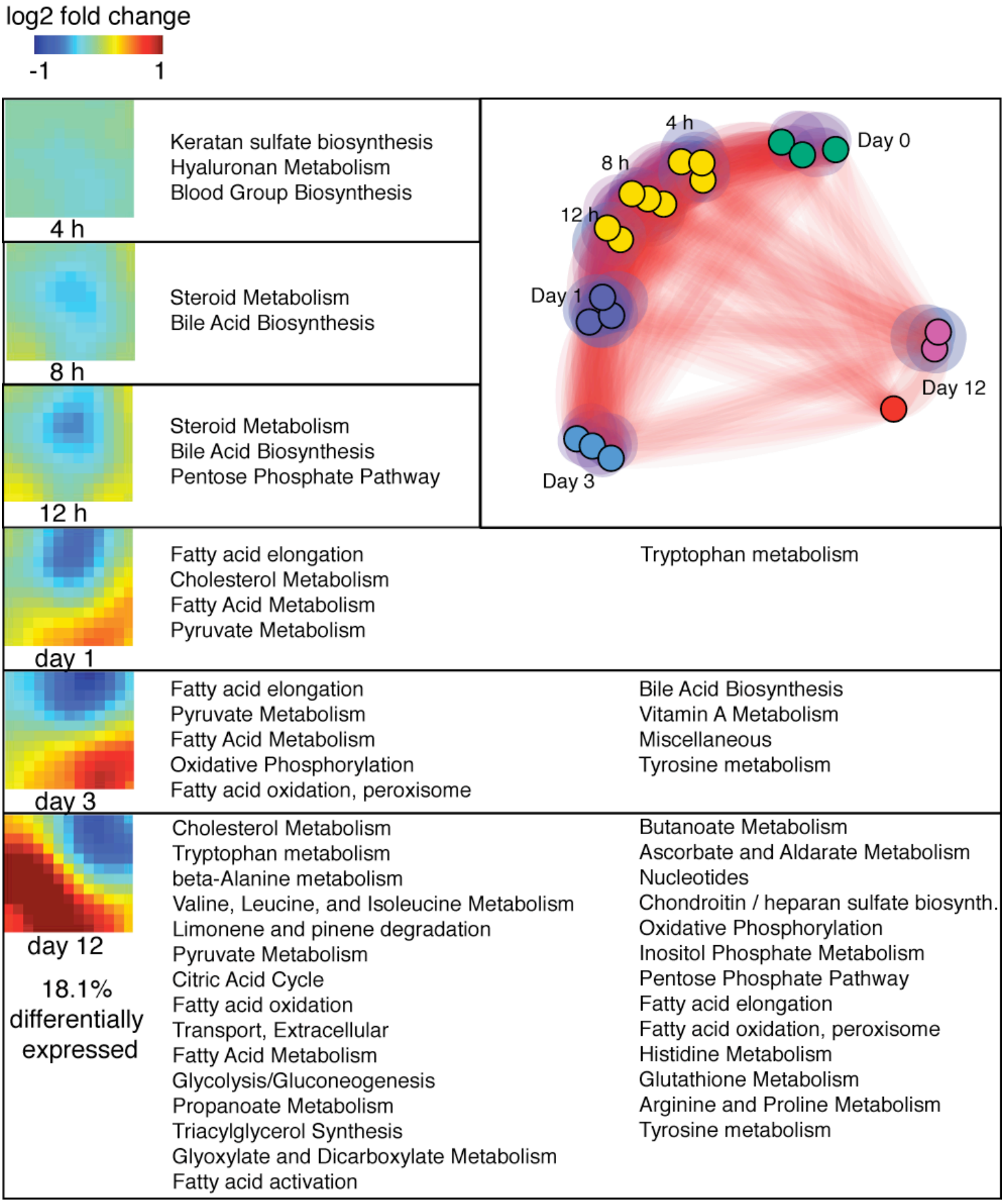

log2 fold change
-1
1
4 h
Keratan sulfate biosynthesis
Hyaluronan Metabolism
Blood Group Biosynthesis
8 h
Steroid Metabolism
Bile Acid Biosynthesis
12 h
Steroid Metabolism
Bile Acid Biosynthesis
Pentose Phosphate Pathway
4 h
8 h
12 h
Day 0
Day 1
Day 3
Day 12
day 1
Fatty acid elongation
Cholesterol Metabolism
Fatty Acid Metabolism
Pyruvate Metabolism
Tryptophan metabolism
day 3
Fatty acid elongation
Pyruvate Metabolism
Fatty Acid Metabolism
Oxidative Phosphorylation
Fatty acid oxidation, peroxisome
Bile Acid Biosynthesis
Vitamin A Metabolism
Miscellaneous
Tyrosine metabolism
day 12
18.1%
differentially
expressed
Cholesterol Metabolism
Tryptophan metabolism
beta-Alanine metabolism
Valine, Leucine, and Isoleucine Metabolism
Limonene and pinene degradation
Pyruvate Metabolism
Citric Acid Cycle
Fatty acid oxidation
Transport, Extracellular
Fatty Acid Metabolism
Glycolysis/Gluconeogenesis
Propanoate Metabolism
Triacylglycerol Synthesis
Glyoxylate and Dicarboxylate Metabolism
Fatty acid activation
Butanoate Metabolism
Ascorbate and Aldarate Metabolism
Nucleotides
Chondroitin / heparan sulfate biosynth.
Oxidative Phosphorylation
Inositol Phosphate Metabolism
Pentose Phosphate Pathway
Fatty acid elongation
Fatty acid oxidation, peroxisome
Histidine Metabolism
Glutathione Metabolism
Arginine and Proline Metabolism
Tyrosine metabolism
50
0
Gene density

Fig. S2

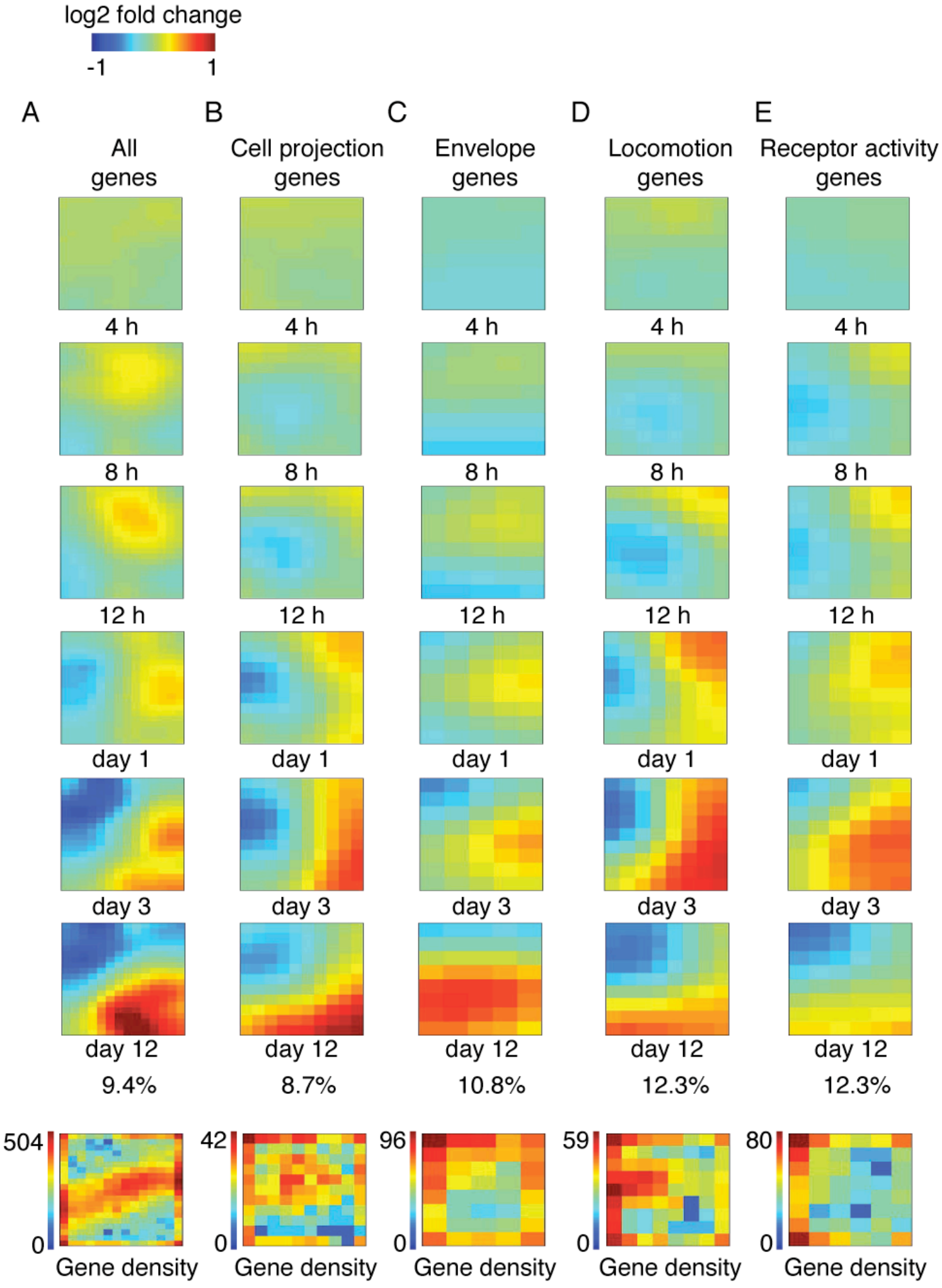

log2 fold change
-1
1
A
All genes
B
Cell projection genes
C
Envelope genes
D
Locomotion genes
E
Receptor activity genes
4 h
8 h
12 h
day 1
day 3
day 12
9.4%
8.7%
10.8%
12.3%
12.3%
504
0
42
0
96
0
59
0
80
0
Gene density
Gene density
Gene density
Gene density
Gene density

Fig. S3

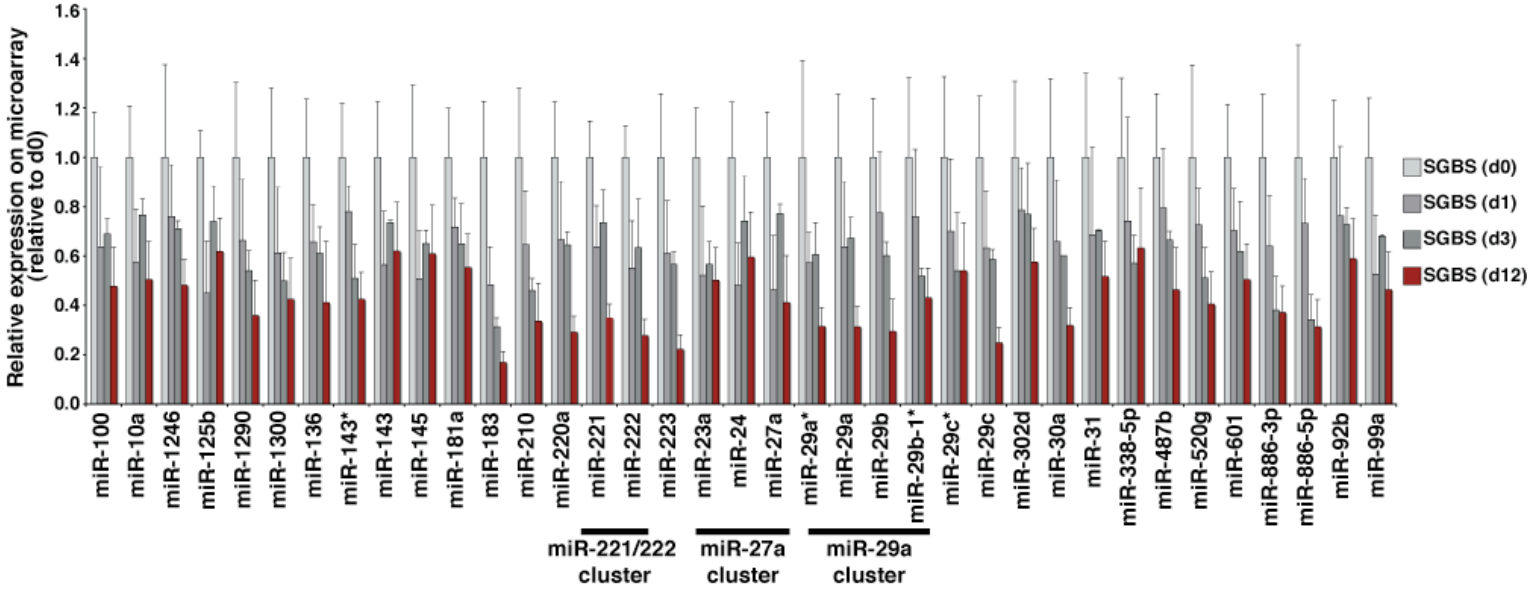

Relative expression on microarray (relative to d0)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
SGBS (d0)
SGBS (d1)
SGBS (d3)
SGBS (d12)
miR-100
miR-10a
miR-1246
miR-125b
miR-1290
miR-1300
miR-136
miR-143*
miR-143
miR-145
miR-181a
miR-183
miR-210
miR-220a
miR-221
miR-222
miR-223
miR-23a
miR-24
miR-27a
miR-29a*
miR-29a
miR-29b
miR-29b-1*
miR-29c*
miR-29c
miR-302d
miR-30a
miR-31
miR-338-5p
miR-487b
miR-520g
miR-601
miR-886-3p
miR-886-5p
miR-92b
miR-99a
miR-221/222 cluster
miR-27a cluster
miR-29a cluster

Fig. S4

Mixture model distributions for log2 score from H3K4me3 ChIP-seq

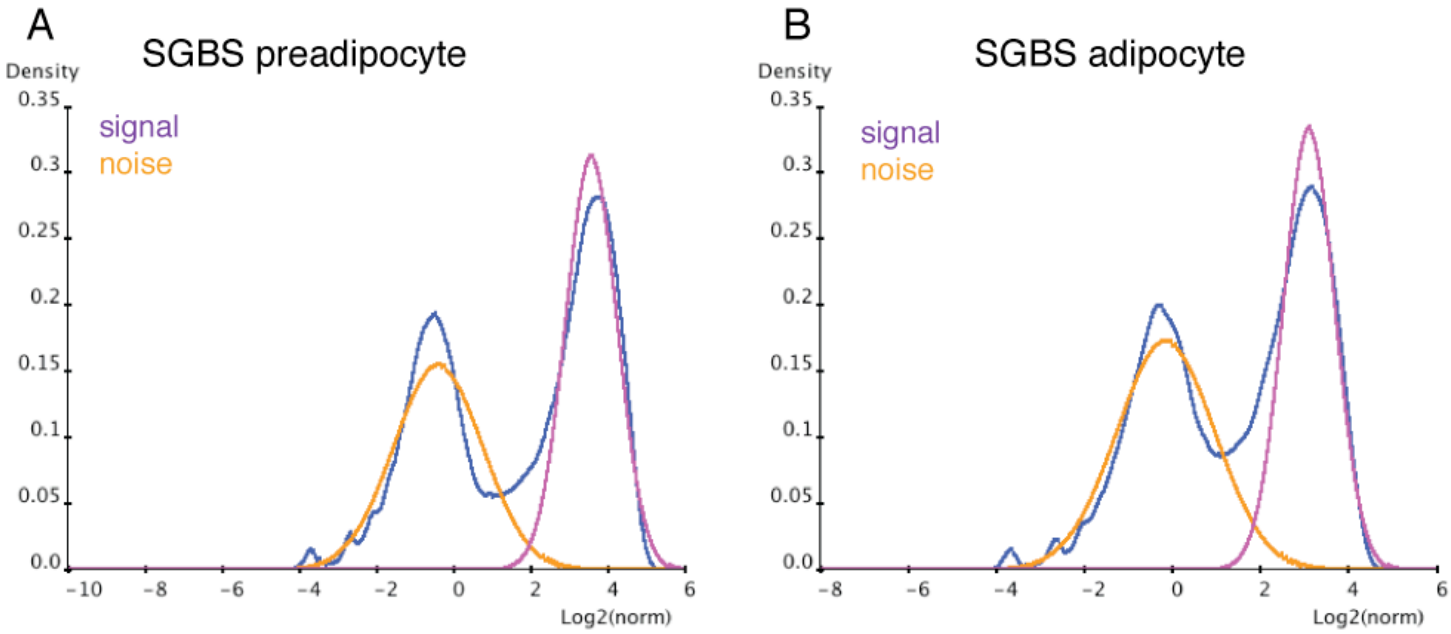


Fig. S5

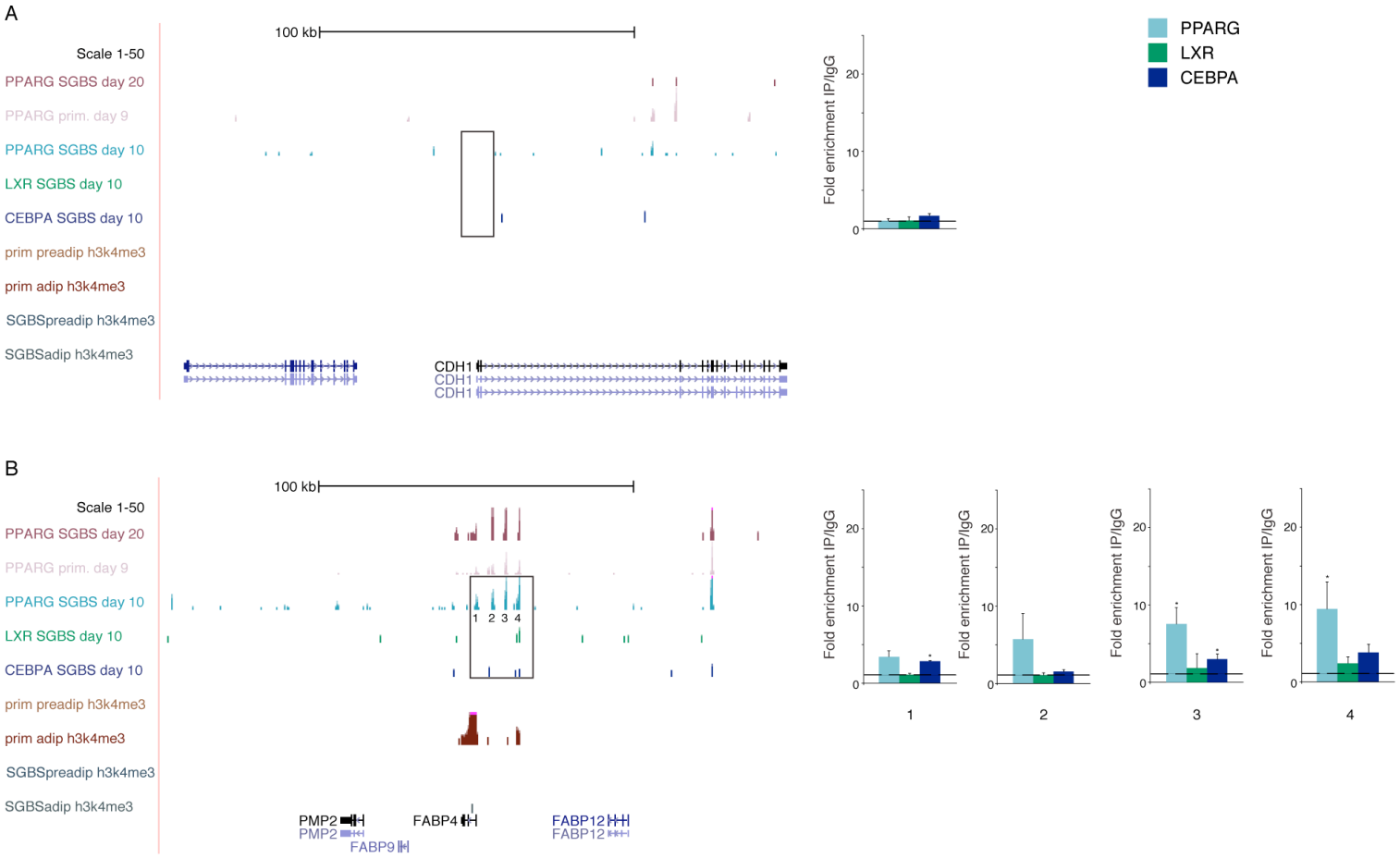
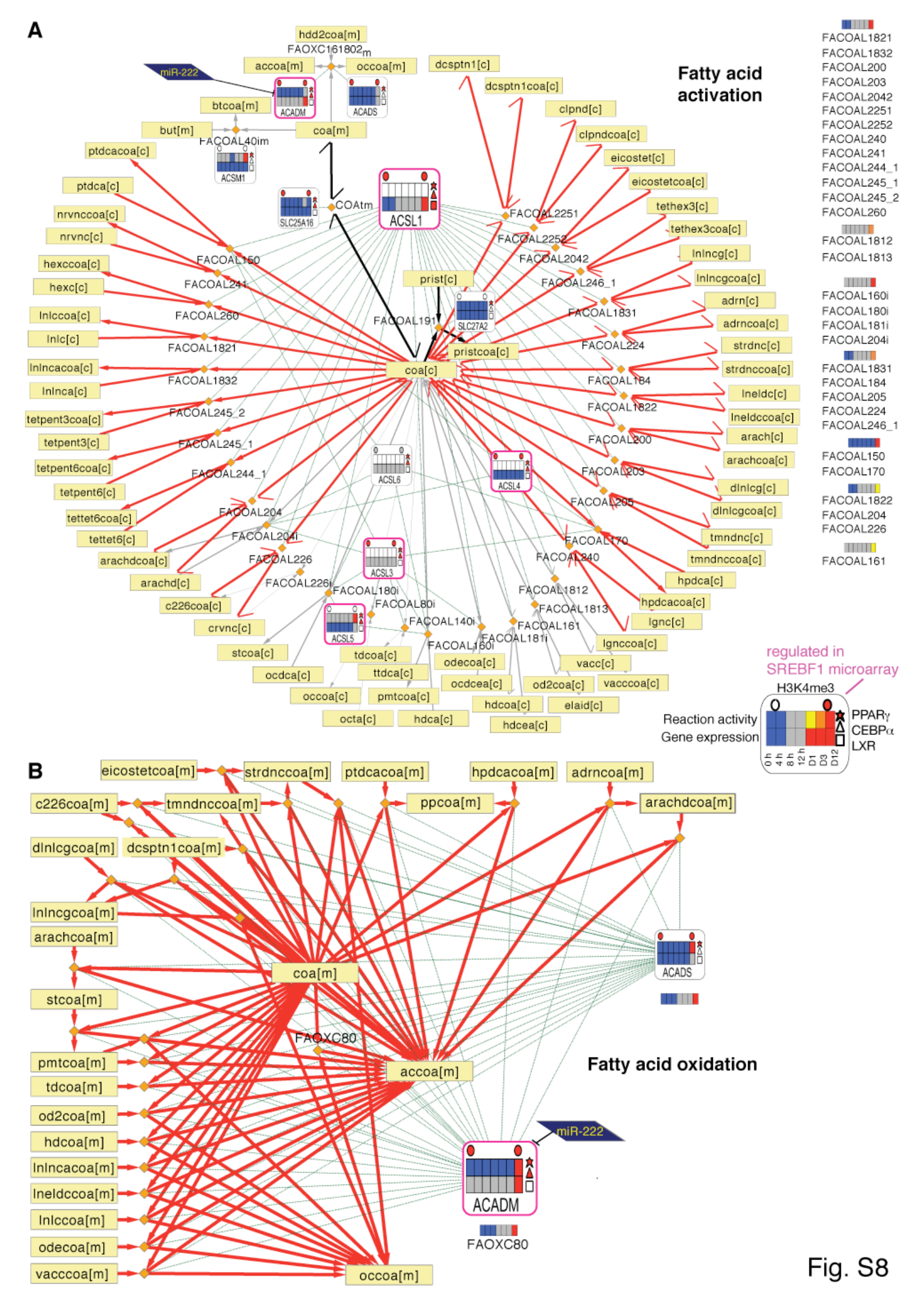

A
100 kb
Scale 1-50
PPARG SGBS day 20
PPARG prim. day 9
PPARG SGBS day 10
LXR SGBS day 10
CEBPA SGBS day 10
prim preadip h3k4me3
prim adip h3k4me3
SGBSpreadip h3k4me3
SGBSadip h3k4me3
CDH1
CDH1
CDH1
Fold enrichment IP/IgG
20
10
0
PPARG
LXR
CEBPA
B
100 kb
Scale 1-50
PPARG SGBS day 20
PPARG prim. day 9
PPARG SGBS day 10
LXR SGBS day 10
CEBPA SGBS day 10
prim preadip h3k4me3
prim adip h3k4me3
SGBSpreadip h3k4me3
SGBSadip h3k4me3
1 2 3 4
PMP2
PMP2
FABP9
FABP4
FABP12
FABP12
Fold enrichment IP/IgG
1
2
3
4

Fig. S6


PPARG
LXRα
CEBPα
+
SREBF1
regulates
(activation / inhibition not inferred)
+ activation

Fig. S7


A
Fatty acid activation
hdd2coa[m]
FAOXC161802m
accoa[m]
occoa[m]
miR-222
ACADM
ACADS
btcoa[m]
but[m]
FACOAL40im
ACSM1
coa[m]
ptdcacoa[c]
ptdca[c]
nrvnccoa[c]
nrvnc[c]
hexccoa[c]
hexc[c]
lnlccoa[c]
lnlc[c]
lnlncacoa[c]
lnlnca[c]
tetpent3coa[c]
tetpent3[c]
tetpent6coa[c]
tetpent6[c]
tettet6coa[c]
tettet6[c]
arachdcoa[c]
arachd[c]
c226coa[c]
crvnc[c]
stcoa[c]
ocdca[c]
occoa[c]
octa[c]
pmtcoa[c]
ttdca[c]
tdcoa[c]
hdca[c]
od2coa[c]
odecoa[c]
ocdcea[c]
hdcea[c]
hdcoa[c]
elaid[c]
vacc[c]
vacccoa[c]
lgnccoa[c]
lgnc[c]
hpdcacoa[c]
hpdca[c]
tmndnccoa[c]
tmndnc[c]
dlnlcgcoa[c]
dlnlcg[c]
arachcoa[c]
arach[c]
lneldccoa[c]
lneldc[c]
adrncoa[c]
adrn[c]
lnlncgcoa[c]
lnlncg[c]
tethex3coa[c]
tethex3[c]
eicostetcoa[c]
eicostet[c]
clpndcoa[c]
clpnd[c]
dcsptn1coa[c]
dcsptn1[c]
pristcoa[c]
prist[c]
strdnccoa[c]
strdnc[c]
COAtm
SLC25A16
ACSL1
SLC27A2
ACSL6
ACSL4
ACSL3
ACSL5
FACOAL150
FACOAL241
FACOAL260
FACOAL1821
FACOAL1832
FACOAL245_2
FACOAL245_1
FACOAL244_1
FACOAL204
FACOAL204i
FACOAL226
FACOAL226i
FACOAL180i
FACOAL80i
FACOAL140i
FACOAL160i
FACOAL181i
FACOAL161
FACOAL1813
FACOAL1812
FACOAL240
FACOAL170
FACOAL205
FACOAL203
FACOAL200
FACOAL1822
FACOAL184
FACOAL224
FACOAL1831
FACOAL246_1
FACOAL2042
FACOAL2252
FACOAL2251
FACOAL191
FACOAL1821
FACOAL1832
FACOAL200
FACOAL203
FACOAL2042
FACOAL2251
FACOAL2252
FACOAL240
FACOAL241
FACOAL244_1
FACOAL245_1
FACOAL245_2
FACOAL260
FACOAL1812
FACOAL1813
FACOAL160i
FACOAL180i
FACOAL181i
FACOAL204i
FACOAL1831
FACOAL184
FACOAL205
FACOAL224
FACOAL246_1
FACOAL150
FACOAL170
FACOAL1822
FACOAL204
FACOAL226
FACOAL161
regulated in SREBF1 microarray
H3K4me3
Reaction activity
Gene expression
PPARγ
CEBPα
LXR
0 h
4 h
8 h
12 h
D1
D3
D12
B
eicostetcoa[m]
strdnccoa[m]
ptdcacoa[m]
hpdcacoa[m]
adrncoa[m]
c226coa[m]
tmndnccoa[m]
ppcoa[m]
arachdcoa[m]
dlnlcgcoa[m]
dcsptn1coa[m]
lnlncgcoa[m]
arachcoa[m]
coa[m]
stcoa[m]
FAOXC80
pmtcoa[m]
tdcoa[m]
od2coa[m]
hdcoa[m]
lnlncacoa[m]
lneldccoa[m]
lnlccoa[m]
odecoa[m]
vacccoa[m]
occoa[m]
accoa[m]
ACADS
ACADM
miR-222
FAOXC80
Fatty acid oxidation
Fig. S8

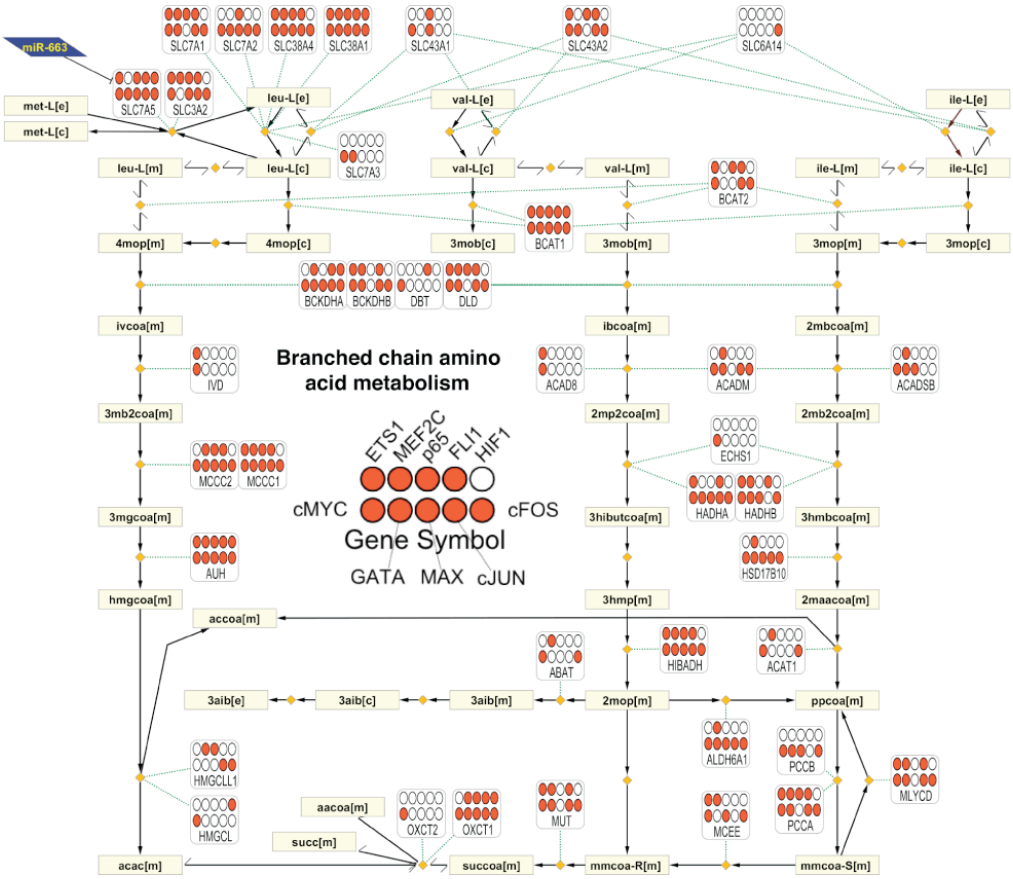
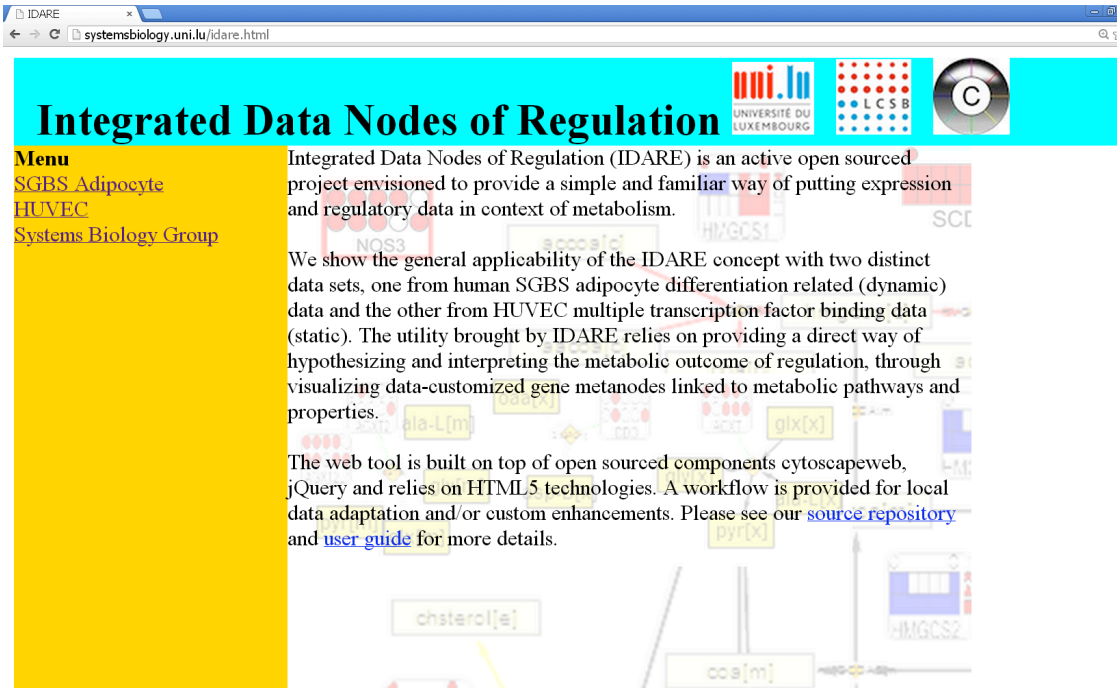

miR-663
SLC7A1
SLC7A2
SLC38A4
SLC38A1
SLC43A1
SLC43A2
SLC6A14
SLC7A5
SLC3A2
met-L[e]
met-L[c]
leu-L[e]
val-L[e]
ile-L[e]
leu-L[m]
leu-L[c]
SLC7A3
val-L[c]
val-L[m]
BCAT2
ile-L[m]
ile-L[c]
4mop[m]
4mop[c]
3mob[c]
BCAT1
3mob[m]
3mop[m]
3mop[c]
BCKDHA
BCKDHB
DBT
DLD
ivcoa[m]
ibcoa[m]
2mbcoa[m]
IVD
ACAD8
ACADM
ACADSB
Branched chain amino acid metabolism
3mb2coa[m]
2mp2coa[m]
2mb2coa[m]
ETS1
MEF2C
p65
FLI1
HIF1
cMYC
cFOS
Gene Symbol
GATA
MAX
cJUN
ECHS1
MCCC2
MCCC1
HADHA
HADHB
3mgcoa[m]
3hibutcoa[m]
3hmbcoa[m]
AUH
HSD17B10
hmgcoa[m]
3hmp[m]
2maacoa[m]
accoa[m]
ABAT
HIBADH
ACAT1
3aib[e]
3aib[c]
3aib[m]
2mop[m]
ppcoa[m]
ALDH6A1
PCCB
HMGCLL1
MLYCD
aacoa[m]
OXCT2
OXCT1
MUT
MCEE
PCCA
HMGCL
succ[m]
acac[m]
succoa[m]
mmcoa-R[m]
mmcoa-S[m]

Fig. S9

IDARE (Integrated DAta nodes or REgulation)

Visualizing regulatory data in context of metabolic pathways

http://systemsbiology.uni.lu/idare.html
http://systemsbiology.uni.lu/adipoflux.html
http://systemsbiology.uni.lu/huvec.html



User guide

IDARE relies in two separate components:
1) The generation of gene 'metanode' image files on Matlab®;
2) Web interactivity through open sourced Html and CytoscapeWeb.

Web address: http://systemsbiology.uni.lu/IDARE
Contact: mafalda.galhardo@uni.lu and jlin@systemsbiology.org
Updated: June 4th, 2013

Table of content:

1. General purpose

IDARE was envisioned to provide a simple and familiar way of showing expression and regulatory data in context of metabolism. Using metabolic maps it provides easy links to biochemical knowledge and extends from current representations by introducing gene metanodes in association to the metabolic pathways.
We show the general applicability of the **IDARE** concept with two distinct data sets, one from HUVEC multiple transcription factor binding data (static) and the other from human SGBS adipocyte differentiation related (dynamic) data.
The utility brought by **IDARE** relies on providing a direct way of hypothesizing and interpreting the metabolic outcome of regulation, through visualizing data-customized gene metanodes linked to metabolic pathways and properties.

2. Available metabolic pathways and datasets

Currently, two datasets are available for exploring with **IDARE**:

1) **AdipoFlux** : human SGBS adipocyte differentiation dynamic data;
2) **Huvec**: human endothelial cell static multiple transcription factor binding data.

The basis for **IDARE** pathway representations is Recon1 (Duarte et al., 2007), a general human metabolic network reconstruction containing metabolic reactions (3742) and associated enzymes, genes and metabolites.
All Recon1 metabolic pathways are available, of which 5 were manually laid out due to their relevance in context of the adipocyte dataset (http://systemsbiology.uni.lu/adipoflux.html) we first analyzed:

- Cholesterol metabolism;
- Fatty acid activation;
- Fatty acid oxidation;
- Triacylglycerol synthesis;
- Valine, leucine and isoleucine metabolism.

These pathways were initially selected based on highest predicted metabolic activity difference between pre-adipocyte and adipocyte stages, as supported by our data and analysis. The networks were manually arranged on Cytoscape and saved as xgmml files containing xy node coordinates. The Cobra toolbox for Matlab® was used to extract network files from the Recon1 model that were imported into Cytoscape (sif).
Additionally, the Arginine and Proline metabolism pathway has been manually laid out in context of the HUVEC dataset (http://systemsbiology.uni.lu/huvec.html).

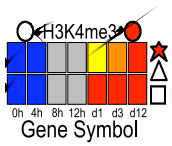
3. IDARE high level components

IDARE relies in two separate components:
1) Generation of gene 'metanode' image files on Matlab®;
2) Web interactivity (Html and Cytoscape Web).

1) Generation of gene 'metanode' image files on Matlab®

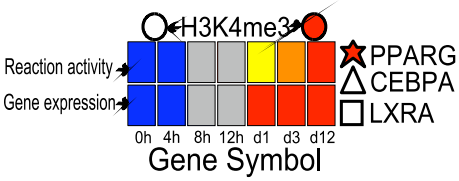
IDARE is currently supporting 2 metanode types, customized in context of SGBS adipocyte differentiation data (CASE 1) and HUVEC transcription factor data (CASE 2).

Summary table of metanode types and properties:

Metanode class	Dataset	Type	Input data	Icon
Heterogeneous	SGBS	dynamic (7 time points)	discrete gene expression (-1, 0, 1)	
			reaction activity prediction (-1, 0, 1, 2, 3)	
			TF putative binding of 3 TFs (gene list)	
			H3K4me3 presence or absence (0 or 1) on 2 time points	
Homogenous	HUVEC	static	10 TFs putative binding (0 or 1)	

CASE 1: SGBS data – dynamic gene expression, reaction activity predictions and few regulators.

We exemplify this metanode type with the SGBS adipocyte differentiation dataset.
Based in our dataset properties, we defined a gene metanode as containing:



- one lower line with discrete gene expression;
- one upper line with reaction activity prediction based on the gene expression;

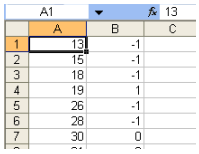
(both lines contain slots representing seven adipocyte differentiation time points)

- three polygons on the right side of the lines, representing the putative binding of three transcription factors (TFs: PPARγ, CEBPα, LXR);
- two circles on top of the lines, aligned according to the time point they belong to, which represent the presence of a histone modification mark associated with active transcription start sites (TSSs) – H3K4me3.

Matlab®: using mainly Entrez gene IDs for mappings, reads in data files, collects arrays for each data type and associates them to each metanode component. For each gene, colors the metanode based on those component arrays.

The following input data files were used (tabular text or excel files):

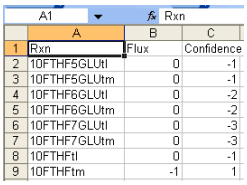
i. Discrete gene expression data for coloring the bottom line:
   - 1st column – Recon1 Entrez gene IDs;
   - Remaining columns – discrete gene expression values:
     - -1 – lowly expressed gene;



- 0 – moderately expressed gene;
- 1 – highly expressed gene.

Matlab® : first step is to build empty (white) rectangles, as many as time points (7), on defined positions (x,y) that are colored based on the data from correspondent time point.

ii. Reaction activity prediction for coloring the top line:

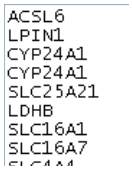


	A	B	C
1	Rxn	Flux	Confidence
2	10FTHF5GLUtl	0	-1
3	10FTHF5GLUtm	0	-1
4	10FTHF6GLUtl	0	-2
5	10FTHF6GLUtm	0	-2
6	10FTHF7GLUtl	0	-3
7	10FTHF7GLUtm	0	-3
8	10FTHFtl	0	-1
9	10FTHFtm	-1	1

- 1st column – Recon1 reaction abbreviation;
- Remaining columns – reaction prediction results (including confidence):
  - 0 – inactive;
  - 1 – active (direct way);
  - -1 – active (reverse way);
  - 2 – active (unknown direction);
  - 3 – undetermined.

Matlab® : second step is to build empty (white) rectangles, as many as time points (7), on defined positions (x,y) on top of the gene expression line; each rectangle is colored based on the data from correspondent time point.

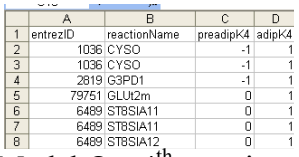
iii. Transcription factor Recon1 associated genes (list of gene symbols):



ACSL6
LPIN1
CYP24A1
CYP24A1
SLC25A21
LDHB
SLC16A1
SLC16A7

Matlab® : third step is to build empty (white) polygons, one for each TF (3), on defined positions (x,y) on the right side of the expression line; from a discrete array (0 or 1) colors polygons in red (1) when gene is associated with a TF.

iv. H3K4me3 data for each gene in Recon1:



	A	B	C	D
1	entrezID	reactionName	preadipK4	adipK4
2	1036	CYSO	-1	1
3	1036	CYSO	-1	1
4	2819	G3PD1	-1	1
5	79751	GLUt2m	0	1
6	6489	ST8SIA11	0	1
7	6489	ST8SIA11	0	1
8	6489	ST8SIA12	0	1

- 1st and 2nd columns: IDs for mapping (Recon1 Entrez gene IDs and reaction abbreviations);
- Remaining columns: discrete values for the presence or absence of the histone mark.

Matlab® : 4th step is to build empty (white) circles, on defined positions (x,y) on top of the reaction activity prediction line, aligned accordingly to the time point they represent; from reading the H3K4me3 data file, forms a discrete array (-1, 0, 1) and colors circles in red (1) or grey (0) accordingly.

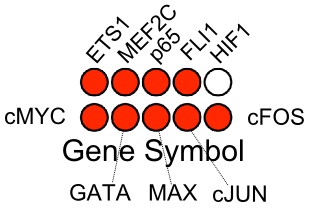
Metanodes are generated per pathway and within each pathway, per gene.
In cases when one gene is associated with multiple reactions, the following occurs:

- generate one metanode with white reaction activity prediction line (defines multiple reactions associated to that gene);
- for each reaction, plots that individual reaction prediction line which is shown on the left side panel when clicking on the gene metanode.

<u>CASE 2: HUVEC data – regulator data only (TFs, static metanode).</u>

We exemplify this metanode type with the HUVEC transcription factor dataset.

Based in the HUVEC dataset properties, we defined a gene metanode as:



- two lines of circles (5 each) representing the putative binding of a total of 10 TFs:
  - Bottom line – cMYC, GATA, MAX, cJUN and cFOS (ENCODE data);
  - Top line – ETS1, MEF2C, p65, FLI1 and HIF1 (own data).
- Circles are filled in red when data supports the binding of correspondent TF to current gene.

Example input data file (tabular text or excel):

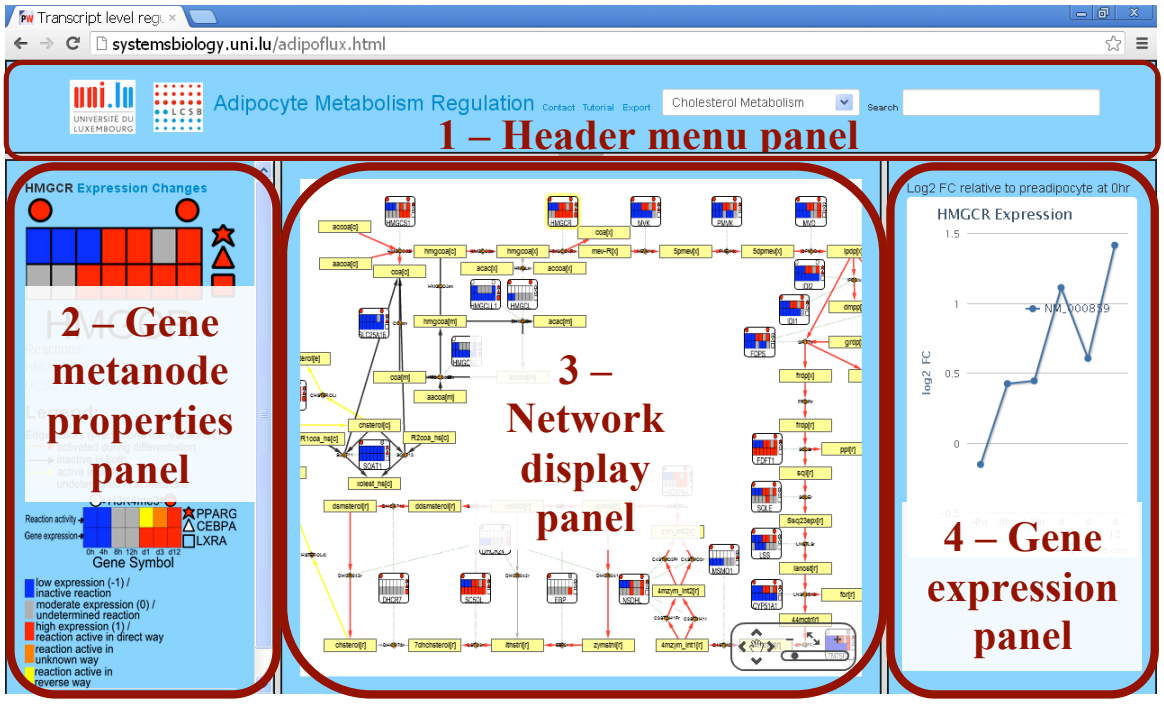
P6

	A	B	C	D	E	F	G	H	I	J	K	L
1	Entrez ID	pathway	ETS1	MEF2C	p65	FLI1	HIF1	cMYC	GATA	MAX	cJUN	cFOS
2	26	beta-Alanine metabolism	0	1	0	0	0	1	1	1	1	1
3	314	beta-Alanine metabolism	0	0	0	0	0	0	0	0	0	0
4	8639	beta-Alanine metabolism	0	0	0	1	0	0	0	0	1	1
5	1591	Vitamin D	0	1	0	1	0	1	0	0	0	1
6	1594	Vitamin D	0	0	1	0	0	0	1	0	0	0
7	89874	Lysine Metabolism	0	1	1	0	0	1	1	0	1	1
8	160287	Propanoate Metabolism	0	0	1	0	0	1	1	0	1	1
9	3939	Propanoate Metabolism	1	0	1	0	1	1	0	1	0	1
10	3945	Propanoate Metabolism	0	0	0	0	0	1	1	1	1	1

Matlab®: using Entrez gene IDs and pathways for mappings, reads in data file with discrete values for the presence or absence (1, 0) of the putative binding of a TF on a gene; draws ten white circles on defined coordinates (x,y), and colors them red when finding a data point "1".

4. IDARE display

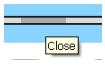
On the following, we exemplify **IDARE** details using as example the adipocyte dataset first analyzed (AdipoFlux instance). The same general characteristics apply to the Huvec dataset.



General view of a **IDARE** instance (AdipoFlux):

Panel overview

1 – **Header menu panel**: here the user can select which metabolic pathway to display as well as access the search function. Links to the associated publication, network export, this user guide and our group's homepage are also available.
2 – **Left side panel** – Gene metanode properties: opens by clicking on a gene metanode and displays details.
The first line on the panel shows the selected gene symbol and a link ('Expression Changes') to the right side panel where gene expression values are plotted.
On the second line, 'UCSC Genome browser' link opens a pop-up 'Adipocyte ChipSeq UCSC Genome Tracks' with a link to a UCSC Genome Browser Adipoflux hub containing ChIP-Seq tracks associated to this work (TFs and H3K4me3 modification) and the selected gene's position that the user should copy and paste to the genome browser in order to visualize the tracks in the selected gene location. Please refer to Chapter 6, section 2 for more details.
A large gene metanode is shown below the two first lines followed by the 'Reactions' associated to the gene. Clicking on a reaction name opens a pop-up with reaction static details.
On the bottom of the left side panel, a legend for edge colors and metanode is provided, so that the user can keep track of what is being represented. Please refer to 'section 3' for more details.
3 – **Central panel** – CytoscapeWeb metabolic network display: this panel contains the metabolic network which can be re-arranged in accordance to user's preference and embeds click-on functions for the nodes' additional details.
4 – **Right side panel** – gene expression: dynamically plots for a selected gene the log2 FC values of each differentiation time point relative to control pre-adipocyte values (microarray data from SGBS cell differentiation time course).



Panels 1, 2 and 4 can be expanded or collapsed by clicking on the center of the grey bar next to them.

5. IDARE interactive elements and functions

This chapter is a walk through **IDARE** interactive elements and functions, using as example 'Cholesterol metabolism' pathway. The metanodes exemplified are in context of the SGBS adipocyte differentiation data. All descriptions apply too for the Huvec data, except that the metanodes have a different visual display, as previously described on Chapter 4.1, case 2.
Below we exemplify how to 'read' the cholesterol synthesis metabolic pathway, which starts with the condensation of acetyl-CoA (accoa[c]) and acetoacetyl-CoA (aacoa[c]) to form 3-hydroxy-3-methylglutaryl-CoA (hmgcoa[c]) catalyzed by the enzyme HMG-CoA synthase (gene *HMGCS1*). The end-point metabolite is cholesterol (chsterol[r][m][c][e], r – endoplasmic reticulum, m – mitochondria, c – cytosol, e - extracellular).

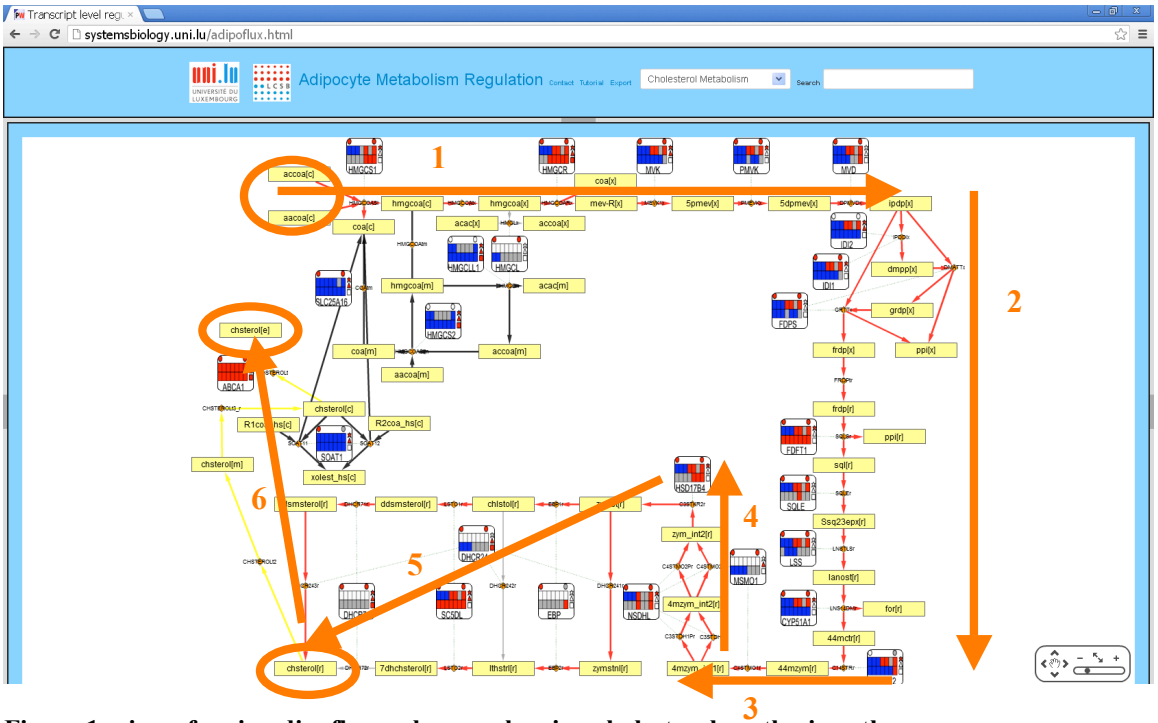


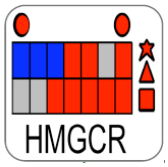
Figure 1: view of main adipoflux webpage, showing cholesterol synthesis pathway.

1. Pathway components:

a) Nodes:

accoa[c] - Metabolites (yellow boxes).

- Reactions (orange diamonds), appear with reaction abbreviation (as of Recon1) on top.



- Gene metanodes representing 4 data levels: gene expression (bottom line rectangles), predicted reaction activity (upper line rectangles), TF association (right side polygons) and marker for active TSS (upper circles) – detailed metanode legend on the left side panel of the webtool and below. Gene metanodes link to reaction nodes (orange diamonds) representing gene-protein-reaction associations contained in Recon1.

miR-222 - miRNA nodes that link to target genes (gene metanodes). Data from miRs -27a, -29a and -222 are included, all the three consistently down-regulated during adipocyte differentiation.

b) Edges:

- Metabolic edges (solid lines): link substrate and product metabolites (yellow nodes) via reactions (orange nodes) that are catalyzed by enzymes.

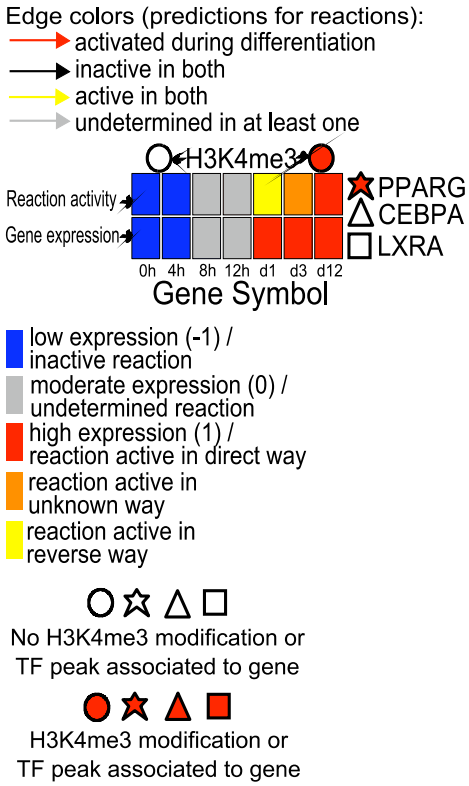
Edge color represents predicted reaction activity based on a constraint-based method (Shlomi et al., 2008), the general human metabolic model Recon1 (Duarte et al., 2007) and gene expression data from a differentiation time course experiment on human SGBS cells. See description below for each edge color.

Edge width represents prediction confidence, with thinner lines for reactions undetermined in the pre-adipocyte and/or the adipocyte stage (colored in grey) and thicker lines for confident reaction activity prediction on both stages (please refer to Shlomi's method for the concept underlying prediction confidence).

- Gene-protein-reaction (GPR) edges (dashed green lines): link metabolic reactions back to the gene(s) encoding the enzymes that catalize them; this info is contained in Recon1.

- miRNA target-inhibition edges (black solid lines in 'T' shape on target interface): link miRNAs with target genes, based on own experimental data.

c) Edge color and gene metanode legend: can be found on the left side panel from within the webtool (click to show/hide)



Red – reactions predicted inactive in pre-adipocytes and active in adipocytes.
Black – reactions predicted inactive in both pre-adipocyte and adipocyte stages.
Yellow - reactions predicted active in both pre-adipocyte and adipocyte stages.
Grey – reactions undetermined in at least one of pre-adipocyte or adipocyte stage.

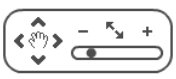
Metanode H3K4me3 panel – represents whether a tri-methylated Lisine-4 residue of Histone 3 was associated to the specific gene (red) or not (white). Left circle represents data on pre-adipocytes and right circle on adipocytes.

Metanode TF panel – represents whether at least one peak from PPARG (star), CEBPA (triangle) or LXR (square) was associated with the specific gene (red) or not (white). Peak-gene associations were obtained from the GREAT tool by providing a list of TF-peak genomic coordinates.

Metanode gene expression (bottom-line rectangles) and predicted reaction activity (upper-line rectangles) per differentiation time-point are represented by color with legend below the gene metanode icon.

2. Data interactivity and integration

Our web tool provides interactive access to the discussed five metabolic pathways combined with several *omics* data, future releases will incorporate many more pathways.



The network itself can be re-arranged by moving nodes as preferred by the user. This can be done by clicking on the 'hand' icon on the panel to

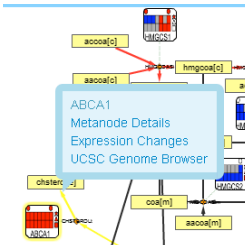
the lower right corner of the screen.
Once the ‘hand’ is selected, one can move each node to a desired position and the edges.

The network can be exported as an svg file, by hitting the link ‘export network’ on the ‘header menu panel’.

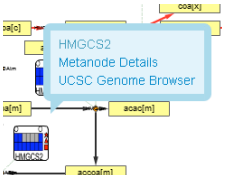
Data interactivity is provided through the network nodes, via click-in functions that open the left side panel or pop-up tool tips on mouse over events.

Mousing-over metabolite (yellow boxes), reaction (orange diamonds) or gene (white) nodes displays a callout with additional info/links.

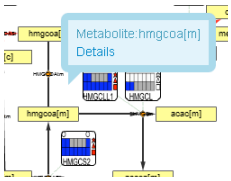
Gene mouse-over (blue callout):



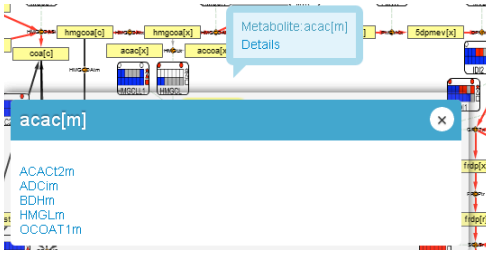
Clicking on ‘Metanode Details’ opens the left side panel, while clicking on ‘Expression changes’ opens the right side panel and dynamically plots the gene expression log2 FC of each differentiation time point relative to control pre-adipocytes. The ‘UCSC Genome Browser’ link opens a pop-up that can re-direct to UCSC GB to visualize ChIP-Seq data tracks on the gene’s position.
Genes whose expression was too low (not detected on the microarray) do not contain the ‘Expression changes’ link to the right side panel (e.g. HMGCS2 on the right).
Direct click on a gene metanode opens both side panels (when expression values are available).



Metabolite mouse-over:

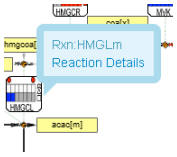


Clicking on ‘Details’ opens a pop-up containing the reactions the metabolite is associated to (either being consumed or produced). In the case of the reaction name containing ‘:’, that has been replaced with ‘_’ to conform to HTML5 Java Script Object Notation syntaxes.

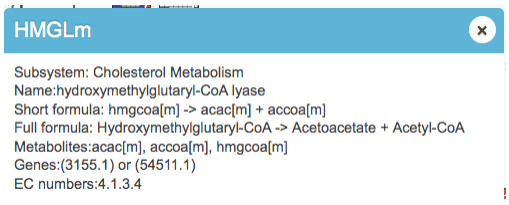


Further clicking on the reaction link opens a new pop-up with reaction details (see below for reaction details description). Direct click on metabolite node opens same pop-up showing which reactions the metabolite associates with.

Reaction mouse-over:



Clicking on ‘Reaction Details’ opens a pop-up with static info on the current reaction. This info is contained in Recon1 and includes:

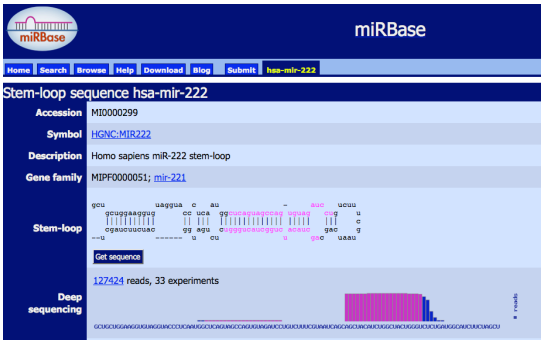


- Subsystem (Recon1 pathway the reaction belongs to);

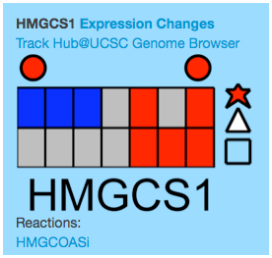
- Name (full name of selected reaction);
- Short formula (biochemical reaction equation with metabolite abbreviations);
- Full formula (biochemical reaction equation with metabolite full names);
- Metabolites (abbreviations of the metabolites involved in the reactions);
- Genes (Recon1 gene-reaction rule – if multiple genes are associated with the reaction, shows the Boolean rules 'and'/'or' that characterize it; the notation was kept as of Recon1 and it represents as outdated version of Entrez Gene IDs; 'undefined' stands for reactions without associated genes as of Recon1 (e.g. transport reactions);
- EC numbers: Enzyme commission numbers for the enzyme(s) catalyzing the reaction, as of Recon1; 'undefined' is shown for reactions for which no EC number was available.

MiR node mouse over shows the miRBase accession number and ID for the selected human microRNA (hsa-id).

On click of the miRNA node redirects to miRBase page for the specific human miRNA.

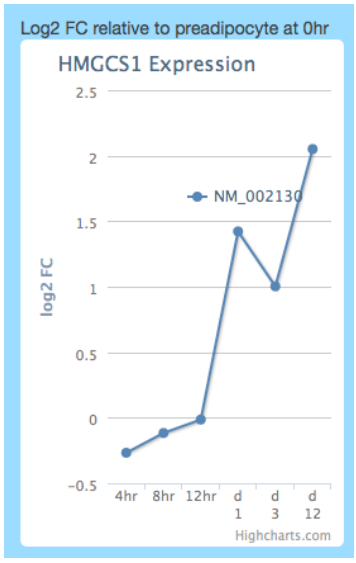


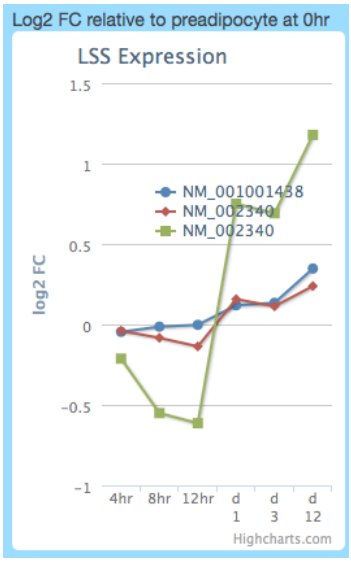
Plotting Gene Expression Changes:



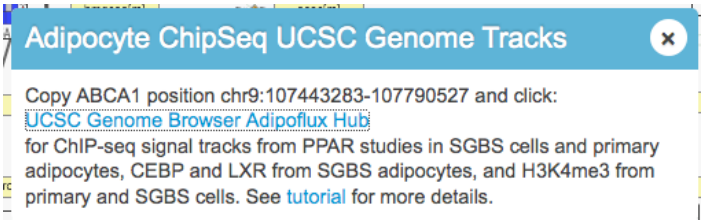
A key feature built in to adipoflux viz is the ability to plot gene expression changes collected over time; these values can serve as confirmatory metrics for the 'Gene expression' bottom line of the metanode graphs. Gene expression data is available for a large majority of genes in the pathways presented. Clicking on gene metanode automatically renders these dynamic plots when the data is available. For more convenience, clicking the hovering tooltip 'Expression Changes' or within the Metanode Details (left panel) 'Expression Change' will dynamically plot the selected gene expression values.

Shown below are HMGCS1 and LSS expression over a differentiation time course (cholesterol metabolism pathway). Notice that LSS gene has multiple microarray probes and therefore expression sets (different lines in the plot). The probe ID and fold change value relative to control pre-adipocytes are displayed on mouse over. Further, these charts are exportable.



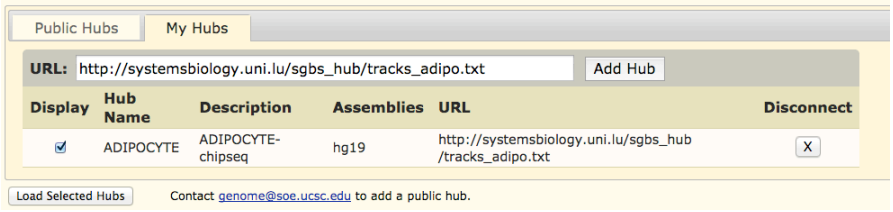


Launch Adipocyte Track Hub@UCSC Genome Browser:

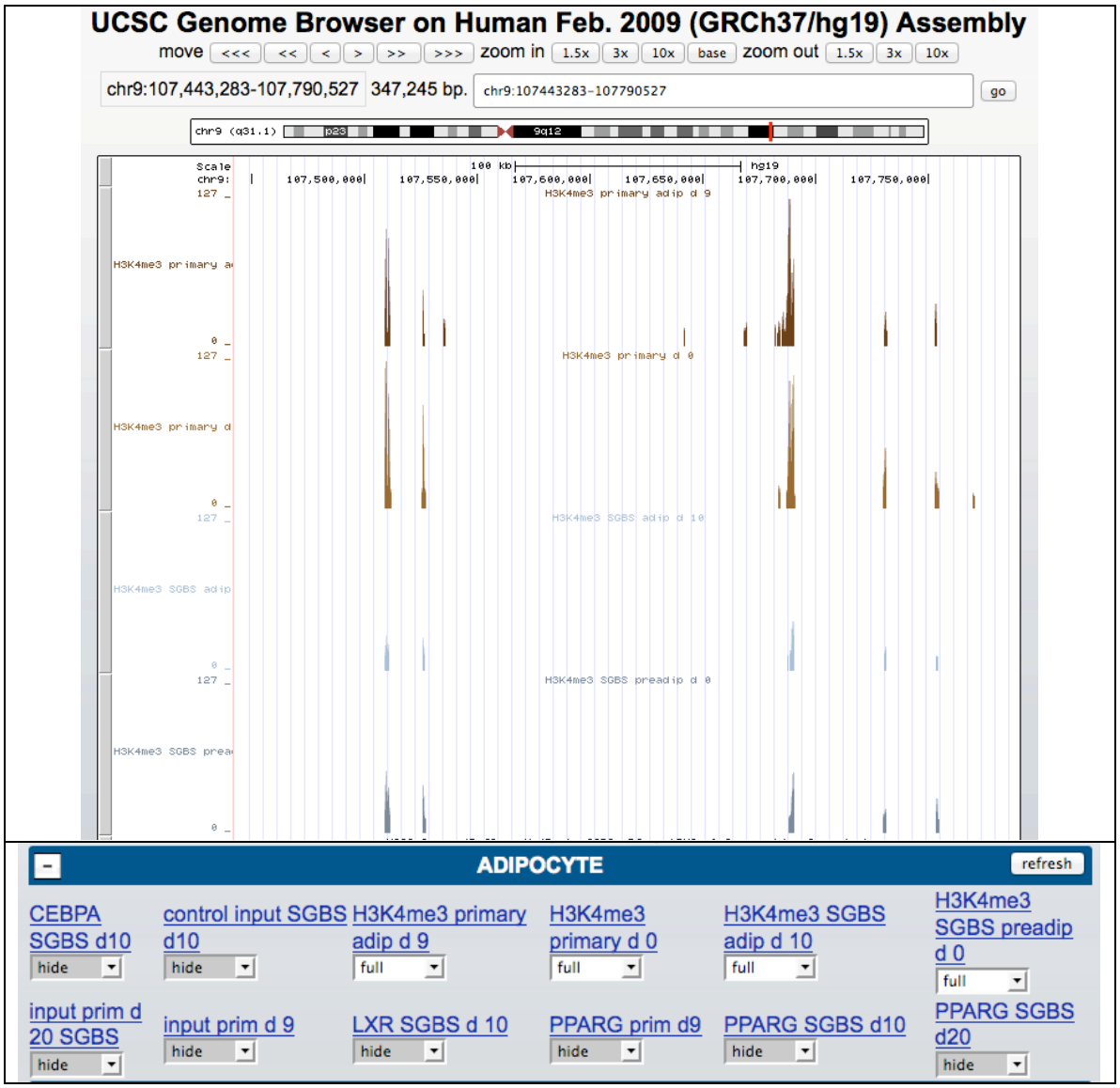


On clicking of the UCSC Genome Browser link within the Metanode details section, a url to connect to the UCSC Adipocyte Track Hub is served, using ABCA1 as an example.

The URL launches to the following screen, then click Load Selected Hubs:



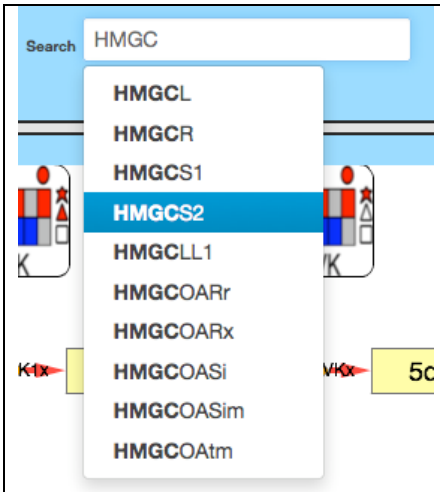
UCSC Genome Browser with 4 H3K4me3 tracks visible, screen below indicates visibility control of Adipocyte Hub custom tracks:



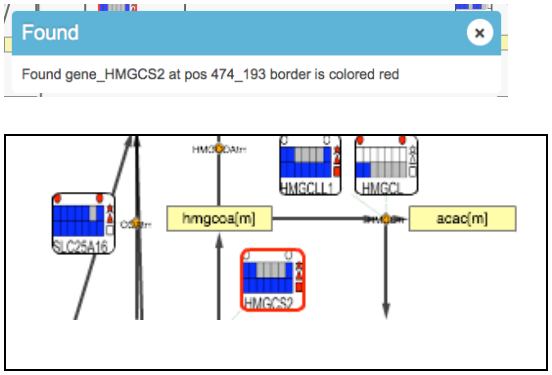
3. Search function

The search function has built in type ahead. A potential list of matches is displayed as a selectable list. This feature augments the power of search and is available for genes, reactions and metabolites; a message is shown on whether the term used is part of the active pathway, along with highlighting of the searched node.

An example scenario of searching in the cholesterol pathway for HMGCS2 is as follows:



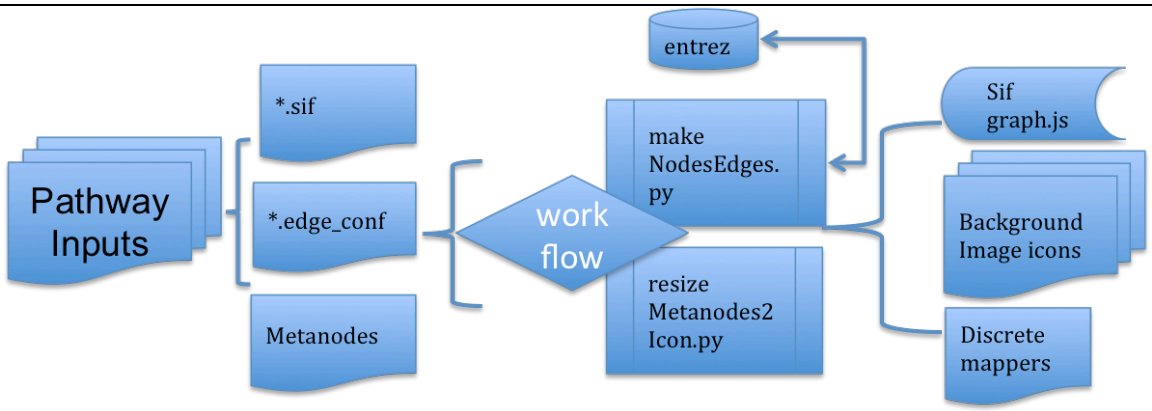
Type ahead automatically shows potential search terms.



A message dialog shows search status; and the found node is highlighted in red.

6. Web graph object automation

We defined a pathway as a set of nodes (genes, reactions, metabolites) connected by edges (tripartite graph). The genesis of AdipoFlux uses xgmml files that contains custom graphs, where the node positions (x,y) and graph attributes (node shapes, colors, edge arrow types…) read in on page load. While this layout scheme is informative and controlled, it is quite manual and only 5 pathways were included. Soon we realized the need for an automated workflow to integrate new pathways. Using the pathway definition – our solution is based on simple interaction (add link) and edge configuration files; in addition, the generated metanodes are also read in for creation of background image icons. All the required scripts are available as part of release.



Workflow inputs are set in a conf file and on workflow invocation, graph and image icons are created and placed into web container paths. Custom discrete mappers are coded as needed.

Because our graph object automation workflow is built using python and bash scripts and designed around principled graph theory, being extensible for other pathway datasets.

7. Web interactivity (HTML and CytoscapeWeb).

IDARE is an open sourced Web 2.0 application, built based on modern HTML specs. The source code is available online on https://code.google.com/p/adipoflux/, including a Matlab® file that can be used for producing the metanode image files (png).
We use Cytoscape Web to embed metabolic pathways as interactive networks.
We are grateful to the following open sourced projects that **IDARE** references:

- JQuery and JQuery plugins
    - Messi
    - JQuery.layout
    - JQuery.tipsy
- Bootstrap.js
- Highchart.js
- UCSC Genome Browser
- ImageMagick

The above tools and libraries were used to render data integration and interactivity and apply to all datasets on **IDARE**. A detailed description of the available functions can be found on the next chapter.