Title

Transcriptomic profiling of primary alveolar epithelial cell differentiation in human and rat

Authors

Marconett, Crystal N

Zhou, Beiyun

Siegmund, Kimberly D

et al.

Publication Date

2014-12-01

DOI

10.1016/j.gdata.2014.05.011

Peer reviewed

Genomics Data 2 (2014) 105–109

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Transcriptomic profiling of primary alveolar epithelial cell differentiation in human and rat

Crystal N. Marconett [a,b,c,*], Beiyun Zhou [c,d], Kimberly D. Siegmund [c,e], Zea Borok [b,c,d], Ite A. Laird-Offringa [a,b,c]

[a] Department of Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[b] Department of Biochemistry and Molecular Biology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[c] Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[d] Will Rogers Institute Pulmonary Research Center and Division of Pulmonary, Critical Care and Sleep Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[e] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 30 April 2014
Accepted 2 May 2014
Available online 29 May 2014

Keywords:
Alveolar epithelial cells
Differentiation
Transcriptomic analysis

ABSTRACT

Cell-type specific gene regulation is a key to gaining a full understanding of how the distinct phenotypes of differentiated cells are achieved and maintained. Here we examined how changes in transcriptional activation during alveolar epithelial cell (AEC) differentiation determine phenotype. We performed transcriptomic profiling using *in vitro* differentiation of human and rat primary AEC. This model recapitulates *in vitro* an *in vivo* process in which AEC transition from alveolar type 2 (AT2) cells to alveolar type 1 (AT1) cells during normal maintenance and regeneration following lung injury. Here we describe in detail the quality control, preprocessing, and normalization of microarray data presented within the associated study (Marconett et al., 2013). We also include R code for reproducibility of the referenced data and easily accessible processed data tables.

Specifications		
	Human gene expression	Rat gene expression
Organism	*Homo sapiens*	*Rattus norvegicus*
Tissue	Primary alveolar epithelial cells	Primary alveolar epithelial cells
Platform	Illumina HT12v4	Illumina RatRef-12
GEO accession ID	GSE38569	GSE38570

Direct link to deposited data SuperSeries (containing both datasets)

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38571

Experimental design, materials & methods

Human remnant lung selection and alveolar epithelial type 2 cell purification

Remnant human transplant lungs were obtained in compliance with Institutional Review Board—approved protocols for the use of human source material in research (HS-07-00660) and processed within 3 days of death. Rat AT2 cells were isolated in compliance with IACUC protocol #11360 from Sprague-Dawley male rats. Lungs were accepted from donors between 18 and 75 years of age with no history of smoking, negative serologies and negative cultures with the exception of CMV, EBV, and hepatitis (with confirmed vaccination record), not a current drug user, and a $pO_2 > 200$ on 100% FIO_2. Additionally, donor lungs were rejected for: heavy marijuana usage, any cancer present within the patient, and chest X-ray indicating pneumonia, asthma, emphysema, or chronic obstructive pulmonary disease (COPD). Also rejected were lungs from donors on ventilator greater than 4 days, any presence of bacterial or viral meningitis, or the presence of MRSA. Human lung tissue was processed as previously described [2] and detailed cell purification techniques have also been described previously [1].

RNA isolation

One microgram of RNA was converted into cRNA using the Illumina TotalPrep RNA amplification kit (Life Technologies, USA) and used for human (Illumina HT12v4) or rat (Rat-Ref-12) expression analysis at the Southern California Genotyping Consortium, University of California Los Angeles.

* Corresponding author at: Department of Surgery, Keck School of Medicine, University of Southern California, 1441 Eastlake Ave, NOR 6420, Los Angeles, CA 91352, USA. Tel.: +1 3238650654.
E-mail address: cmarcone@usc.edu (C.N. Marconett).

http://dx.doi.org/10.1016/j.gdata.2014.05.011

Table 1
Meta data for human AEC.

Target	Sex	Prepdate	Day	Race	Age	Ter	Smoker
5626686051_A	Female	2010Dec	D6	Caucasian	49	Yes	No
5626686051_C	Female	2010Dec	D4	Caucasian	49	Yes	No
5626686051_F	Female	2010Dec	D0	Caucasian	49	Yes	No
5626686051_H	Female	2010Dec	D2	Caucasian	49	Yes	No
5626686051_L	Female	2010Dec	D8	Caucasian	49	Yes	No
5626686013_A	Female	2010Nov	D0	Caucasian	61	Yes	No
5626686013_B	Female	2010Nov	D6	Caucasian	61	Yes	No
5626686013_C	Female	2009Dec	D4	Caucasian	66	Yes	No
5626686013_D	Female	2010Nov	D4	Caucasian	61	Yes	No
5626686013_E	Female	2009Dec	D6	Caucasian	66	Yes	No
5626686013_F	Female	2009Dec	D2	Caucasian	66	Yes	No
5626686013_G	Female	2010Nov	D0	Caucasian	61	Yes	No
5626686013_H	Female	2010Nov	D2	Caucasian	61	Yes	No
5626686013_I	Female	2009Dec	D8	Caucasian	66	Yes	No
5626686013_J	Female	2010Nov	D4	Caucasian	61	Yes	No
5626686013_K	Female	2009Dec	D0	Caucasian	66	Yes	No
5626686013_L	Female	2010Nov	D8	Caucasian	61	Yes	No

Basic analysis

BeadStudio was used to convert images to raw signal data. Data files from BeadStudio were analyzed in R (version 2.11.1). Code for human expression analysis is included in Appendix A, code for expression analysis of rat is included in Appendix B. Briefly, the data was compiled into an eSet using the LUMI package [3] using metadata from Table 1 (human) and Table 5 (rat). Unique lumiIDs based on probe hybridization sequence were assigned to each probe using lumiHumanIDMapping. For the rat expression data, probes were filtered based on quality using the reMoat reannotation pipeline, available online at: http://www.compbio.group.cam.ac.uk/Resources/Annotation/ [4]. Raw data was checked for enrichment of p-values of less than 0.05, indicating significance above background false discovery using a matrix design (Table 2 for human and Table 6 for rat). Variant stabilization and normalization (VSN) was performed using the VSN package [5] to allow for a large number of differentially expressed genes. Statistical analyses were performed using LIMMA [6] with the technical replicates removed (Table 3, rat had no technical replicates). A linear regression model was fitted over the time-course of differentiation using lmFit, and t-tests performed between D0 and D8. False-discovery rate was controlled using the Benjamini–Hochberg (BH) correction [7]. R was used for principal component analysis and heatmap generation. Heatmaps were generated using Heatmap.plus in R by selecting the top 5% of probes most variant across the whole dataset and clustering with "average" linkage method. Clustering with different linkages, for example "ward" (Fig. 1) and "complete" (Fig. 2), resulted in comparable sample dendrograms. The list of significant differentially expressed genes is included in Table 4 (human) and Table 7 (rat). Pathway analysis was performed on genes with statistically

Table 2
LIMMA design matrix human raw (includes technical replicates).

Target	D0	D2	D4	D6	D8
5626686051_A	0	0	0	1	0
5626686051_C	0	0	1	0	0
5626686051_F	1	0	0	0	0
5626686051_H	0	1	0	0	0
5626686051_L	0	0	0	0	1
5626686013_A	1	0	0	0	0
5626686013_B	0	0	0	1	0
5626686013_C	0	0	1	0	0
5626686013_D	0	0	1	0	0
5626686013_E	0	0	0	1	0
5626686013_F	0	1	0	0	0
5626686013_G	1	0	0	0	0
5626686013_H	0	1	0	0	0
5626686013_J	0	0	1	0	0
5626686013_I	0	0	0	0	1
5626686013_K	1	0	0	0	0
5626686013_L	0	0	0	0	1

Table 3
LIMMA design matrix human clean (excludes technical replicates).

Target	D0	D2	D4	D6	D8
5626686051_A	0	0	0	1	0
5626686051_C	0	0	1	0	0
5626686051_F	1	0	0	0	0
5626686051_H	0	1	0	0	0
5626686051_L	0	0	0	0	1
5626686013_A	1	0	0	0	0
5626686013_B	0	0	0	1	0
5626686013_C	0	0	1	0	0
5626686013_D	0	0	1	0	0
5626686013_E	0	0	0	1	0
5626686013_F	0	1	0	0	0
5626686013_H	0	1	0	0	0
5626686013_I	0	0	0	0	1
5626686013_K	1	0	0	0	0
5626686013_L	0	0	0	0	1

Table 5
Meta data for rat AEC.

Sample	ChIP lane	ChIP	Sample name	Prepdate	DAY
5665175063_A	A	RAT v1.0	AEC TII D6	Round3	D6
5665175063_B	B	RAT v1.0	AEC TII D6	Round2	D6
5665175063_C	C	RAT v1.0	AEC TII D2	Round3	D2
5665175063_D	D	RAT v1.0	AEC TII D2	Round2	D2
5665175063_E	E	RAT v1.0	AEC TII D4	Round3	D4
5665175063_G	G	RAT v1.0	AEC TII D8	Round3	D8
5665175063_H	H	RAT v1.0	AEC TII D0	Round3	D0
5665175063_I	I	RAT v1.0	AEC TII D0	Round2	D0
5665175063_J	J	RAT v1.0	AEC TII D8	Round2	D8
5665175063_K	K	RAT v1.0	AEC TII D4	Round2	D4
5700760018_A	A	RAT v1.0	AEC TII D2	Round1	D2
5700760018_F	F	RAT v1.0	AEC TII D4	Round1	D4
5700760021_A	A	RAT v1.0	AEC TII D0	Round1	D0
5700760021_D	D	RAT v1.0	AEC TII D6	Round1	D6
5700760021_I	I	RAT v1.0	AEC TII D8	Round1	D8

Table 6
LIMMA design matrix rat.

Target	D0	D2	D4	D6	D8
5665175063_A	0	0	0	1	0
5665175063_B	0	0	0	1	0
5665175063_C	0	1	0	0	0
5665175063_D	0	1	0	0	0
5665175063_E	0	0	1	0	0
5665175063_G	0	0	0	0	1
5665175063_H	1	0	0	0	0
5665175063_I	1	0	0	0	0
5665175063_J	0	0	0	0	1
5665175063_K	0	0	1	0	0
5700760018_A	0	1	0	0	0
5700760018_F	0	0	1	0	0
5700760021_A	1	0	0	0	0
5700760021_D	0	0	0	1	0
5700760021_I	0	0	0	0	1

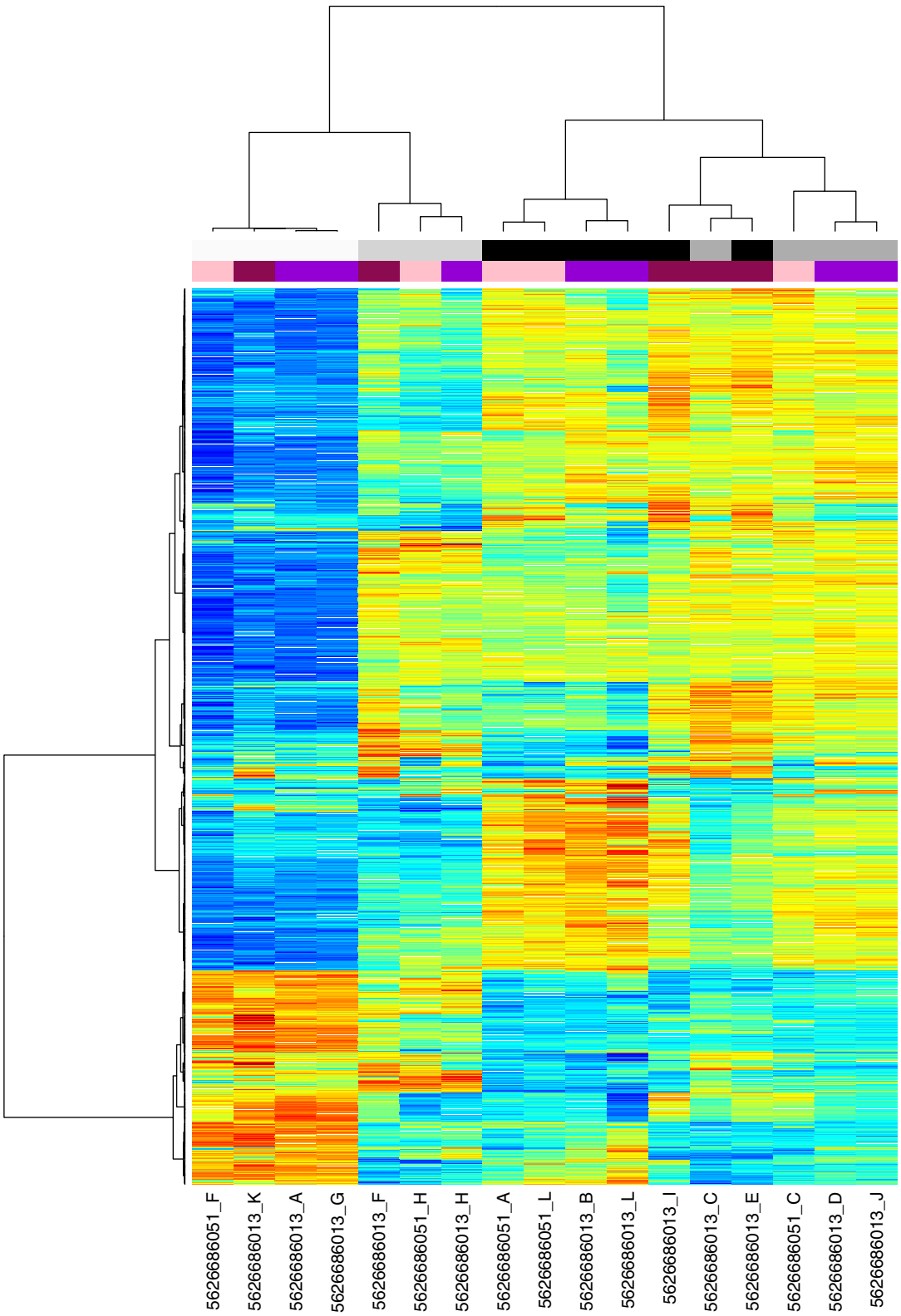

Fig. 1. Unsupervised hierarchical clustering of human HT-12v4 normalized microarray data using the "ward" clustering method. The top 5% of variant probes across the dataset were included.

significant differences in expression using IPA (Ingenuity Systems, www.ingenuity.com) or DAVID [8,9]. Correlation of human and rat gene expression was performed using Entrez identifiers and the Mouse Genome Informatics (MGI) Web database [10], and the correlated microarrays (Table 8) were plotted against each other to reveal genes which were differentially expressed in human, rat, or both. Unique gene symbols were used to calculate overall numbers of genes significantly differentially expressed (Table 9).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.05.011.

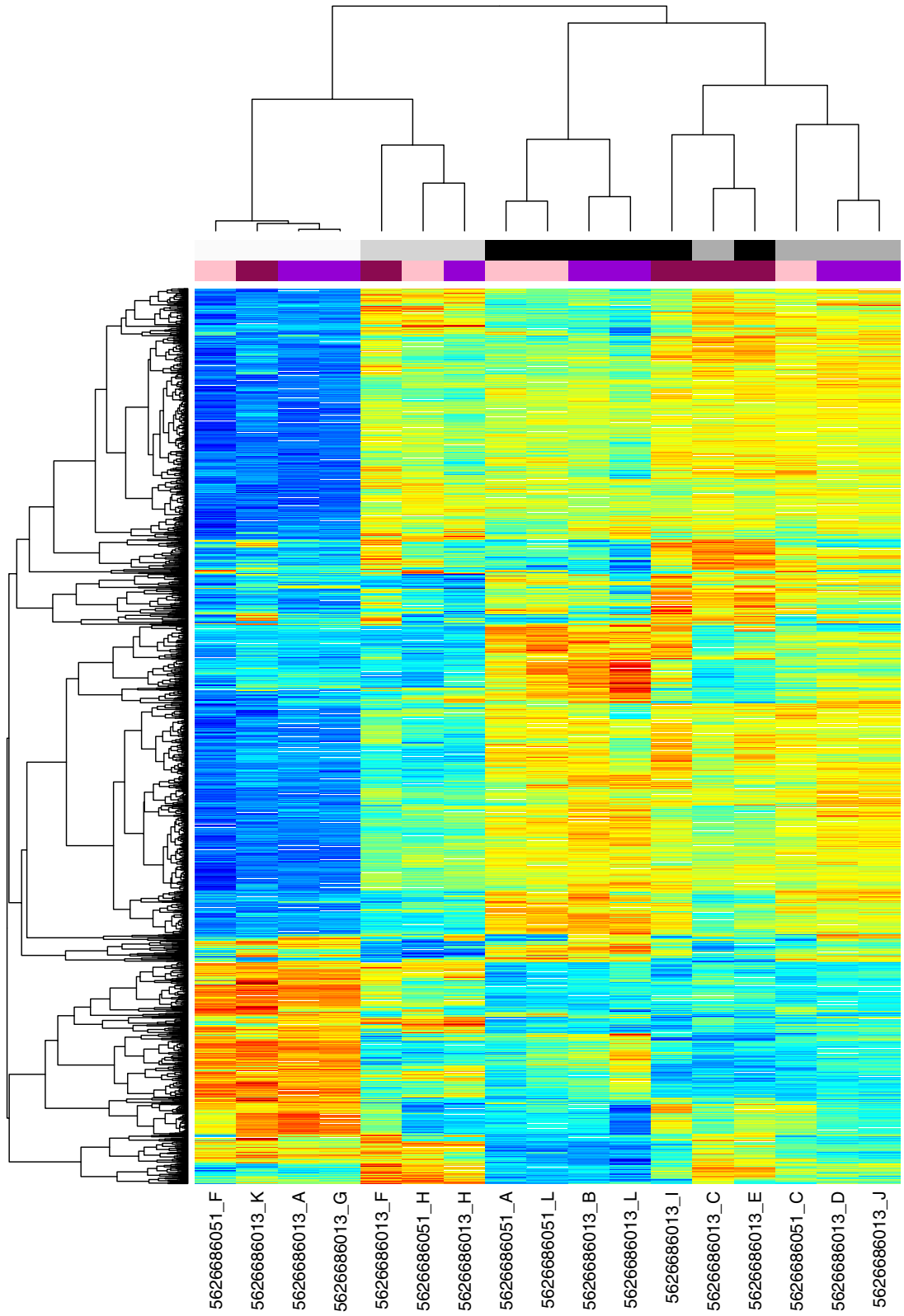

Fig. 2. Unsupervised hierarchical clustering of human HT-12v4 normalized microarray data using the "complete" clustering method. The top 5% of variant probes across the dataset were included.

References

[1] C.N. Marconett, B. Zhou, M.E. Rieger, S.A. Selamat, M. Dubourd, X. Fang, S.K. Lynch, T.R. Stueve, K.D. Siegmund, B.P. Berman, Z. Borok, I.A. Laird-Offringa, Integrated transcriptomic and epigenomic analysis of primary human lung epithelial cell differentiation. PLoS Genet. 9 (2013) e1003513.

[2] P.L. Ballard, J.W. Lee, X. Fang, C. Chapin, L. Allen, M.R. Segal, H. Fischer, B. Illek, L.W. Gonzales, V. Kolla, et al., Regulated gene expression in cultured type II cells of adult human lung. Am. J. Physiol. Lung Cell. Mol. Physiol. 299 (2010) L36–L50.

[3] P. Du, W.A. Kibbe, S.M. Lin, Lumi: a pipeline for processing Illumina microarray. Bioinformatics 24 (2008) 1547–1548.

[4] N.L. Barbosa-Morais, M.J. Dunning, S.A. Samarajiwa, J.F. Darot, M.E. Ritchie, A.G. Lynch, S. Tavaré, A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. Nucleic Acids Res. 38 (2010) e17.
[5] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 (2002) S96–S104.
[6] G.K. Smyth, Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3 (2004) (Article3).
[7] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57 (1995) 289–300.
[8] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (2009) 44–57.
[9] D.W. Huang, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37 (2009) 1–13.
[10] Mouse Genome Informatics (MGI) Web, the Jackson Laboratory, Bar Harbor, Maine. http://www.informatics.jax.org [retrieved 2/2011].