Functional Annotation of RIKEN Mouse cDNA Clones Using GNF Expression Atlas

Serge Batalov, Andrew I. Su, Keith A. Ching, et al.

References	This article cites 2 articles, 1 of which can be accessed free at: **http://genome.cshlp.org/content/13/6b/1556.full.html#ref-list-1**
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

Abstract

Functional Annotation of RIKEN Mouse cDNA Clones Using GNF Expression Atlas

Serge Batalov,[1] Andrew I. Su, Keith A. Ching, and Colin F. Fletcher

Genomics Institute of the Novartis Research Foundation (GNF), San Diego, California 92121, USA

Annotation of novel cDNAs relies on multiple lines of evidence, including gene structure, predicted ORFs, and predicted protein motifs. Insight into function also relies on expression profiling, through querying EST databases of direct experimentation.

High-throughput gene expression profiling, by use of spotted arrays or oligo chip methods, has become an important tool for investigating transcriptional activity. Among cDNAs predicted to encode related proteins, different expression patterns can suggest functional specialization. Additionally, common expression patterns can serve to group otherwise unrelated cDNAs. Expression profiling may be particularly useful for difficult to annotate clones, for which other lines of evidence are not available.

GNF has generated and analyzed gene expression previously from a set of 45 murine samples across a diverse list of normal tissues, organs, and cell lines probed with the Affymetrix MG-U74A chip. These data have been published (Su et al. 2002) and are available at the GNF's free and publicly accessible Web site (Gene Expression Atlas, http://expression.gnf.org), which integrates data visualization and curation of current gene annotations.

We identified the relationship between the RIKEN collection and MG-U74A/B/C targets by SIM4-polished BLAST alignments between 60,770 RIKEN clone sequences and 36,701 target sequences arrayed on the U74A/B/C chips. The results of this analysis are summarized in Table 1. There are 25,831 RIKEN clones represented on this array set. Of these, 2994 have lower quality similarity hits at the DNA or protein level (i.e., are not automatically or easily annotated by homology).

Ten additional tissues were profiled using the Affymetrix MG-U74Bv2 and MG-U74Cv2 chips to allow for inferred functional annotation in the absence of other evidence. These tissues were chosen to give the greatest representation of gene expression by extrapolation from the results obtained with U74A. Five of the tissues are also in common with those used to probe the RIKEN microarray. The combined resources should provide an extremely comprehensive characterization of the expression data.

Figure 1 shows just two arbitrarily taken cases in which tentative annotation may be inferred from the expression profile.

Table 1. The Counts of RIKEN Clones Represented on MG-U74 A, B, and C Chips

Chip	Probe sets	RIKEN hits	and not U74A	and not U74B	and not easy to annotate[a]
U74A	12,422	12,199	–	–	454
U74B	12,411	11,165	9,965	–	1,669
U74C	11,868	4,138	3,667	2,926	821

[a]Not counting clones with evidence class 0–4 (The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I and II Team 2002).
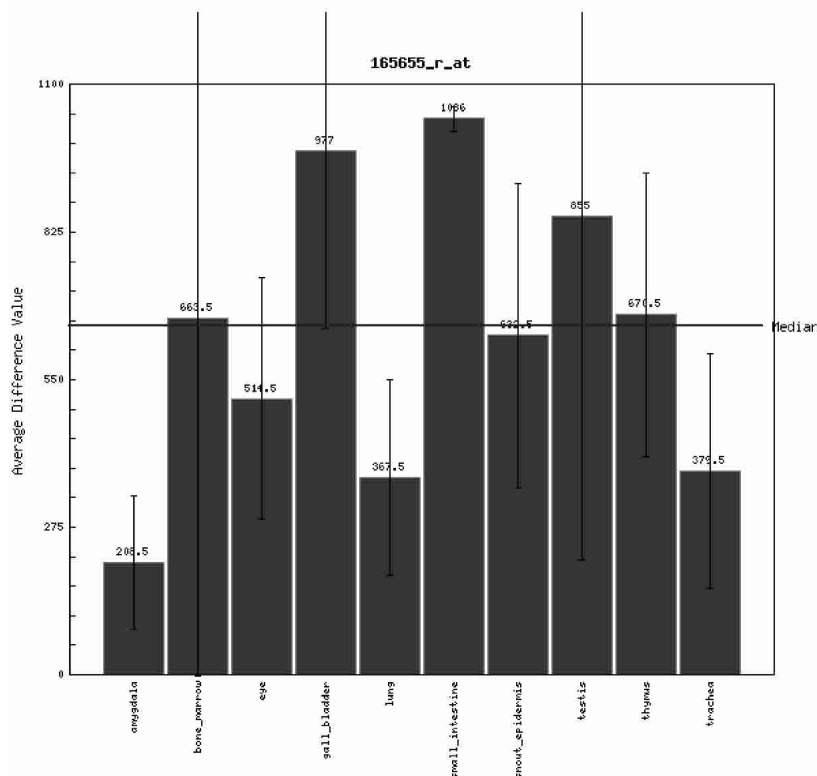
REFERENCES

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I and Phase II Team. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420:** 563–573.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci.* **99:** 4465–4470.

[1]Corresponding author.
E-MAIL batalov@gnf.org; FAX (858) 812-1570.

A

165655_r_at Original Affymetrix Annotation
AI448459:mq40a07.x1 Mus musculus cDNA, 3 end
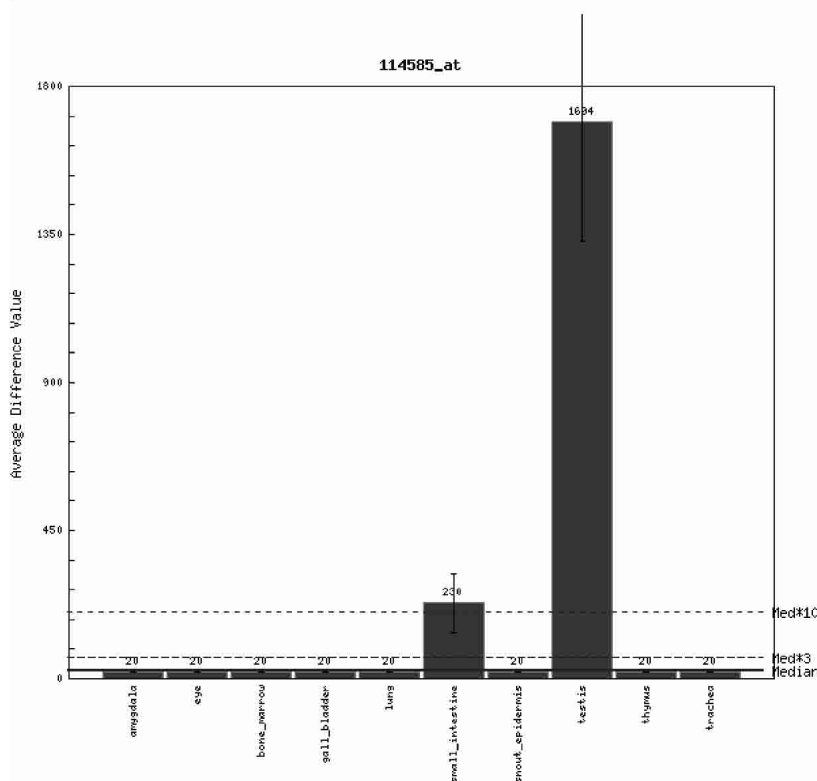
Current Annotation Thu Apr 24 19:15:04 PDT 2003

__Locuslink_id__ 56226
name espin
symbol Espn
product espin
sumfunc none
summary none
go actin bundling activity;ATP binding activity;cilium;brush border;actin cytoskeleton;parallel actin filament bundle formation;
extannot none
phenotype none
map 4 80.1 cM|MGD|G|4 E1
maplink MGD
nm NM_019585 Notes
np NP_062531
pubmed 12207892 11217851 11076861 11042159 10975527 10588661 10349636

__NCBI_MGD Human-Mouse Synteny__

	h_chr	h_id	h_sym	m_chr	m_id	m_sym
1	7161	TP73	4	22062	Trp73
1	49856	WDR8	4	59002	Wdr8
1	83715	DKFZP434A196	4	56226	Espn
1	11332	HBACH	4	70025	2410041A17Rik
1	8514	KCNAB2	4	16498	Kcnab2

__Unigene__ **Expect** 5e-55 **identity** 100% 137 bases
ug Mm.32006
name Espn
title espin
cytoband 4 80.1 cM
chromosome 4
expression eyeball colon lung ...

B

114585_at Original Affymetrix Annotation
AW046432:UI-M-BH1-akp-d-05-0-UI.s1 Mus musculus cDNA, 3 end

Current Annotation Thu Apr 24 19:32:17 PDT 2003

__Locuslink_id__ 207175
name centrin 4
symbol LOC207175
product centrin 4
sumfunc none
summary none
go none
extannot none
phenotype none
map none
maplink none
nm NM_145825 Notes
np NP_665824
pubmed none

__Unigene__ **Expect** 2e-90 **identity** 100%
matchlength 181 bases **Query size** 181 bases
ug Mm.25493
name LOC207175
title centrin 4
cytoband 3 B
chromosome 3
expression visual cortex fetus testis spermatocytes, pooled from multiple mice 18-day leptotene and zygotene spermatocytes Testicles mammary gland bladder olfactory brain bone marrow heart

Figure 1 (*A*) RIKEN clone 0610007F01 is interrogated by the probeset 165655_r_at. (*B*) RIKEN clones 6430560A18 and G430087M20 are associated with the probe set 114585_at with testis-specific pattern.