

Gene expression

Affy exon tissues: exon levels in normal tissues in human, mouse and rat

Andrew A. Pohl¹, Charles W. Sugnet², Tyson A. Clark², Kayla Smith¹, Pauline A. Fujita¹ and Melissa S. Cline^{3,*}

¹Center for Biomolecular Science & Engineering, University of California, Santa Cruz, CA 95064,

²Affymetrix, Inc., 3420 Central Expressway, Santa Clara, CA 95051 and ³Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, CA 95064, USA

Received on May 19, 2009; revised on June 26, 2009; accepted on June 29, 2009

Advance Access publication July 1, 2009

Associate Editor: David Rocke

ABSTRACT

Summary: Most genes in human, mouse and rat produce more than one transcript isoform. The Affymetrix Exon Array is a tool for studying the many processes that regulate RNA production, with separate probesets measuring RNA levels at known and putative exons. For insights on how exons levels vary between normal tissues, we constructed the Affy Exon Tissues track from tissue data published by Affymetrix. This track reports exon probeset intensities as log ratios relative to median values across the dataset and renders them as colored heat maps, to yield quick visual identification of exons with intensities that vary between normal tissues.

Availability: Affy Exon Tissues track is freely available under the UCSC Genome Browser (<http://genome.ucsc.edu/>) for human (hg18), mouse (mm8 and mm9), and rat (rn4).

Contact: cline@biology.ucsc.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The mammalian transcriptome is complex. By recent estimates, as many as 94% of human genes undergo alternative splicing (Wang, *et al.*, 2008). Alternative splicing is consequential as well as frequent, with effects ranging from altering protein structure to targeting mRNA for early decay (Hartmann and Valcarcel, 2009). Furthermore, mammalian genomes contain an abundance of non-coding RNA genes (Chu and Rana, 2007). In short, to understand the consequences of transcription, one must look beyond overall expression levels of known genes.

Affymetrix exon arrays facilitate transcriptome analysis with probesets that measure RNA abundance for individual exons, conserved genomic regions, and blocks from syntenic alignments (see http://www.affymetrix.com/support/technical/technotes/exon_array_design_technote.pdf). These arrays have offered new insights on how transcript isoforms may be influenced by a myriad of factors including tissue type (Clark *et al.*, 2007), genetic variation (Kwan *et al.*, 2007; Zhang *et al.*, 2009), differentiation (Yeo *et al.*, 2007), and disease (Soreq *et al.*, 2008; Thorsen *et al.*, 2008).

Alternative splicing can arise through normal, regulated processes; or through abnormalities such as mutation, disease, and environmental stress (Yeo *et al.*, 2005). Before one can understand the abnormal conditions, it is valuable to understand the scope of normal alternative splicing by comparing splicing patterns between normal tissues. To facilitate this, we have provided the Affy Exon Tissues tracks in the UCSC Genome Browser (Kuhn *et al.*, 2009), depicting exon probeset intensities in normal tissues in human, mouse, and rat.

2 DESCRIPTION

The Affy Exon Tissues track consists of two parts: genomic coordinates of the exon array probesets; and a heat map indicating exon probeset intensities in normal tissues, based on data available from http://www.affymetrix.com/support/technical/sample_data/exon_array_data.affx. Briefly, normal tissues were assayed in triplicate, and were analyzed with the Affymetrix Power Tools software (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx) to produce normalized, background-corrected probeset intensities. For each probeset, we computed its median intensity for each tissue, and then the median of these median values. For each experiment, we calculated the log ratio between the probeset intensity and this median value. For numeric stability, we added a fixed, background-level pseudocount to each observation, which also renders probesets with no expression as constant-valued. The genome browser renders these log ratios as blue–white–red (shown), green–red, or yellow–blue heat maps, with the color selection controlled via the track's details page. Additional details are provided in the Supplementary Material.

Figure 1a shows the Affy Exon Tissues track for TPM2 in mm9. The *constitutive* exons (those included in all transcripts) indicate that TPM2 is expressed most strongly in muscle and embryo, with some expression in ovary. TPM2 has a well-documented pattern of tissue-dependent splicing, with one isoform produced in skeletal muscle tissue and another in non-muscle tissue (Gooding and Smith, 2008). This pattern is apparent in the two mutually-exclusive exons (third and fourth from the left): one is highly-expressed (red) in muscle and embryo (a heterogeneous tissue), while the other is highly-expressed in ovary.

*To whom correspondence should be addressed.

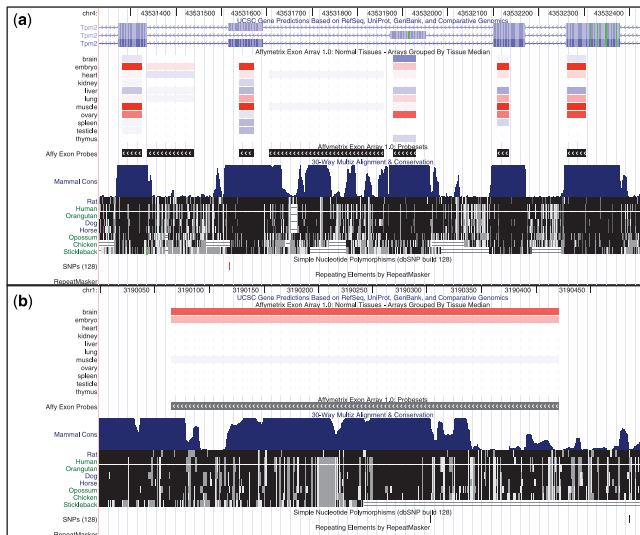


Fig. 1. (a) The Affy Exon Tissues track demonstrates tissue-dependent splicing of two mutually exclusive exons (third and fourth from the left) in the mm9 TPM2 locus. Red indicates probesets that are up-regulated relative to median levels, while blue indicates down-regulated probesets. The *constitutive* exons (those included in all transcripts: first, second and fifth from the left) offer perspectives on overall gene expression, and indicate that the gene is up-regulated (red) in embryo, muscle and ovary. The leftmost mutually-exclusive exon appears upregulated in embryo and muscle, while the second appears upregulated in ovary. The remaining two probesets, which map to no known exons, are rendered mostly white. This indicates no variation in their expression levels, and suggests that they are not expressed in these tissues. (b) A probeset was designed for this unannotated region in mm9 based on more-speculative evidence such as genomic conservation or *ab initio* exon prediction. This region shows evidence of up-regulation (red) in brain, suggesting production of a brain-specific transcript.

For contrast, Figure 1b shows an unannotated conserved region on chromosome 1 in mm9. While it does not overlap with any known gene, its red (up-regulated) log intensity in brain suggests brain-specific expression. This illustrates how this data can offer insights on regions with no annotation but strong conservation.

3 CONCLUSION

The Affy Exon Tissues track displays exon probeset intensities in human, mouse, and rat tissues, including breast, cerebellum, heart, kidney, liver, muscle, pancreas, prostate, spleen, testes, and thyroid. In contrast to traditional microarray tracks such as the GNF Expression Atlas (Su *et al.*, 2004), which provide one measure of overall expression per gene and cannot report any transcript variation, the Affy Exon Tissues track offers the ability to compare intensities of neighboring probesets and observe alternative promoter usage, polyadenylation, and splicing.

Exon probeset intensities are rendered as heat maps to offer rapid visual identification of exons that vary under normal cellular conditions.

Besides the Affy Exon Tissues track, the UCSC Genome Browser currently hosts the hg18 Sestan Brain exon expression track, which contrasts exon probeset intensities between sections of the brain (Johnson *et al.*, 2009). This set of tracks may expand further as additional datasets become available, offering further insights into transcript variation in the mammalian genomes.

ACKNOWLEDGEMENTS

Many people contributed to this track. The authors would like to thank Affymetrix in general and Alan Williams in particular for the track data. This track reflects the work of many individuals in the UCSC Genome Browser team, including Bob Kuhn, Donna Karolchik and Jim Kent. M.S.C. thanks Manuel Ares Jr, for his mentorship and encouragement.

Funding: National Human Genome Research Institute (2 P41 HG002371-06 to UCSC Center for Genomic Science, 3 P41 HG002371-06S1 ENCODE supplement to UCSC Center for Genomic Science); National Cancer Institute (Contract No. N01-CO-12400 for Mammalian Gene Collection); NIH GM-040478 (to M.S.C.)

Conflict of Interest: none declared.

REFERENCES

- Chu,C.Y. and Rana,T.M. (2007) Small RNAs: regulators and guardians of the genome. *J. Cell Physiol.*, **213**, 412–419.
- Clark,T.A. *et al.* (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.*, **8**, R64.
- Gooding,C. and Smith,C.W. (2008) Tropomyosin exons as models for alternative splicing. *Adv. Exp. Med. Biol.*, **644**, 27–42.
- Hartmann,B. and Valcarcel,J. (2009) Decrypting the genome's alternative messages. *Curr. Opin. Cell Biol.*, **21**, 377–386.
- Johnson,M.B. *et al.* (2009) Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*, **62**, 494–509.
- Kuhn,R.M. *et al.* (2009) The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res.*, **37**, D755–D761.
- Kwan,T. *et al.* (2007) Heritability of alternative splicing in the human genome. *Genome Res.*, **17**, 1210–1218.
- Soreq,L. *et al.* (2008) Identifying alternative hyper-splicing signatures in MG-thymoma by exon arrays. *PLoS ONE*, **3**, e2392.
- Su,A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
- Thorsen,K. *et al.* (2008) Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol. Cell Proteomics*, **7**, 1214–1224.
- Wang,E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
- Yeo,G.W. *et al.* (2005) Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA*, **102**, 2850–2855.
- Yeo,G.W. *et al.* (2007) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.*, **3**, e85.
- Zhang,W. *et al.* (2009) Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum. Genet.*, **125**, 81–93.