A tool for identification of genes expressed in patterns of interest using the Allen Brain Atlas: Supplementary Information

Fred P. Davis and Sean R. Eddy

HHMI Janelia Farm Research Campus

19700 Helix Dr, Ashburn, VA 20147

davisf@janelia.hhmi.org

http://research.janelia.org/allenminer

January 23, 2009

References

[1] Kurrasch, D. M. *et al.* (2007). The neonatal ventromedial hypothalamus transcriptome reveals novel markers with spatially distinct patterning. *J Neurosci*, **27**(50), 13624–13634.

[2] Lein, E. S. *et al.* (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, **445**(7124), 168–176.

[3] Zirlinger, M. *et al.* (2001). Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid subnuclei. *Proc Natl Acad Sci U S A*, **98**(9), 5270–5275.

List of Figures

(a) Comparison to ABA manually curated lists

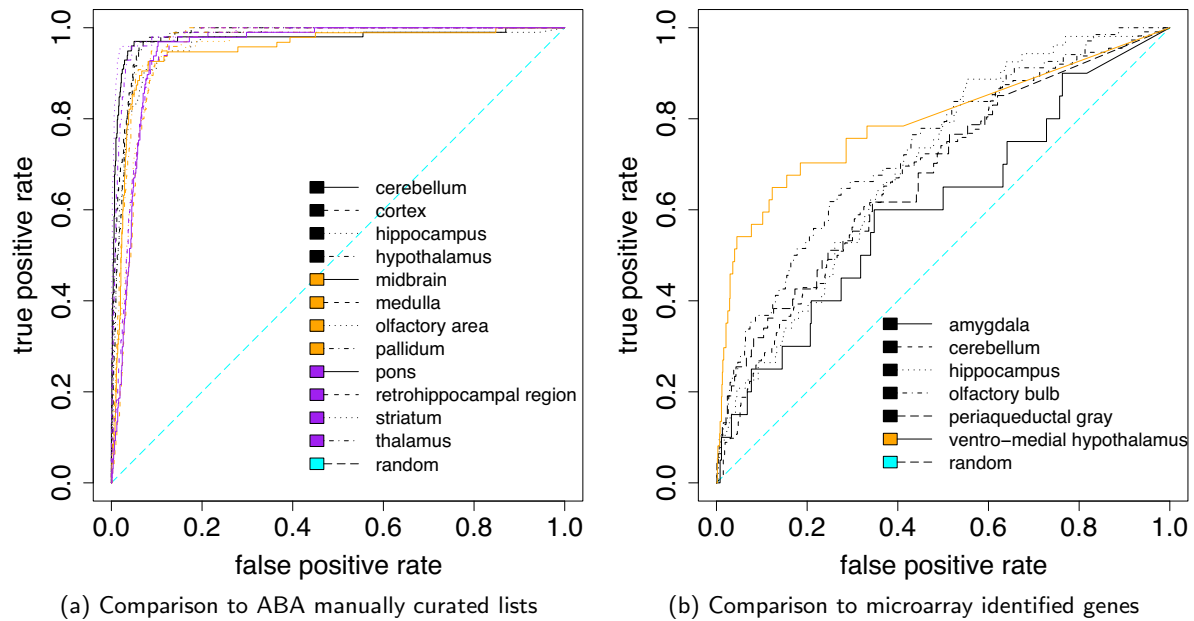(b) Comparison to microarray identified genes

Figure 1: Receiver-operator characteristic of ALLENMINER enrichment score. ALLENMINER results were compared to (a) manually curated lists of genes enriched in 12 ABA brain regions [2] and (b) microarray-identified genes enriched in the cerebellum, amygdala, hippocampus, olfactory bulb, peri-aqueductal gray [3] and fetal ventromedial hypothalamus [1]. Details of the ROC analysis are presented in Table 1 of the manuscript.
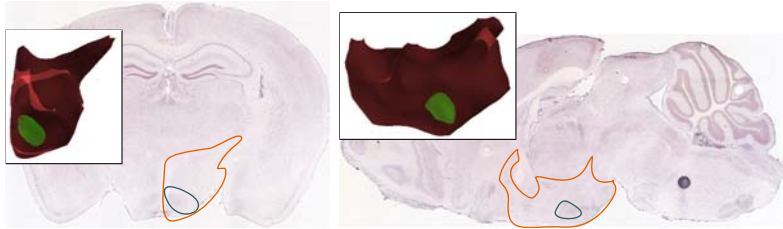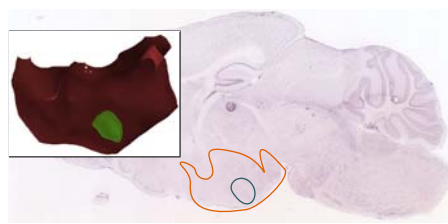
(a) G protein-activated inward rectifier potassium channel 4 (Kcnj5; K: 2.22, A:0; ABA images 72040834, 69890146)
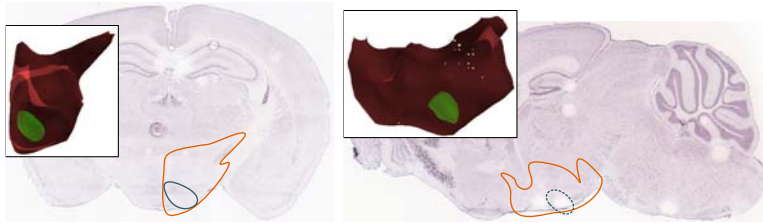
(b) Neuromedin-B receptor (Nmbr; K: 3.30, A: 0; ABA images 77328246, 69148481)
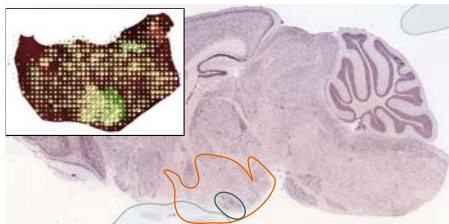
(c) Ataxin-2-binding protein 1 (A2bp1; K:2.4, A:0.95; ABA image 71138095)

(d) Transcription cofactor vestigial-like protein 2 (Vgll2; K: 8.42, A: 0; ABA image 71279859)

Figure 2: Examples of ALLENMINER false negative errors. Comparison of VMH-enriched genes identified by ALLENMINER to those identified by microarray [1] suggests that ALLENMINER false negatives are caused by: (a) sub-threshold expression, (b) tissue damage, and (c) slice/registration artifacts. Some of the apparent negatives are likely to be true negatives, for example Vgll2 (d) which has been independently shown not to express in the adult VMH [1]. The positions of the hypothalamus (orange outline) and VMH (blue outline) are depicted on the *in situ* images. The 3D registered expression (inset, yellow circles) in the hypothalamus (red) is shown along with the highlighted VMH nuclei (green). The fold enrichment observed by the microarray study (K:, [1]) and the ALLENMINER enrichment score (A:) are listed.
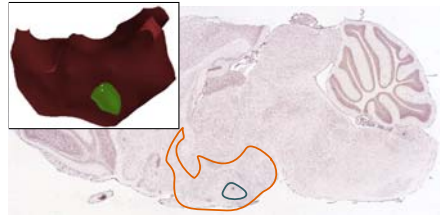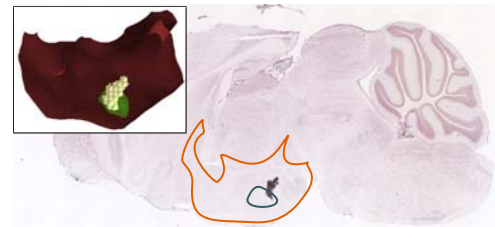
(a) F-box and leucine-rich repeat protein 7 (Fbxl7; A: 37.04; ABA image 73429295)

(b) Guanine nucleotide-binding protein subunit alpha 15 (Gna15; A: 17.31; ABA image 67846082)
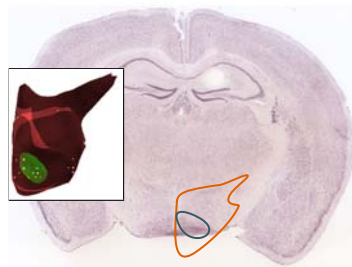
(c) Guanine nucleotide-binding protein G(olf) subunit alpha (Gnal; A: 11.88; ABA image 71514991)

(d) Tctex1 domain containing 1 (Tctex1d1; A: 16.09; ABA image 71774899)

(e) Cylindromatosis (Cyld; A: 27.38; ABA image 73968005)

Figure 3: Examples of ALLENMINER false positive errors. Comparison of genes enriched in the ventromedial hypothalamus identified by ALLENMINER to those identified by microarray [1] suggests that ALLENMINER false positives are caused by (a) particulate aggregates, (b) bubble artifacts, (c) tissue damage such as folded tissue, and (d) debris on the slide. In some cases, enrichment is clearly evident on the *in situ* image and may represent a true positive that was missed by the microarray study or that is differentially expressed in the fetal VMH (e). The positions of the hypothalamus (orange outline) and VMH (blue outline) are depicted on the *in situ* images. The 3D registered expression (inset, yellow circles) in the hypothalamus (red) is shown, along with the highlighted VMH nuclei (green). The ALLENMINER enrichment score is listed for each gene (A:).
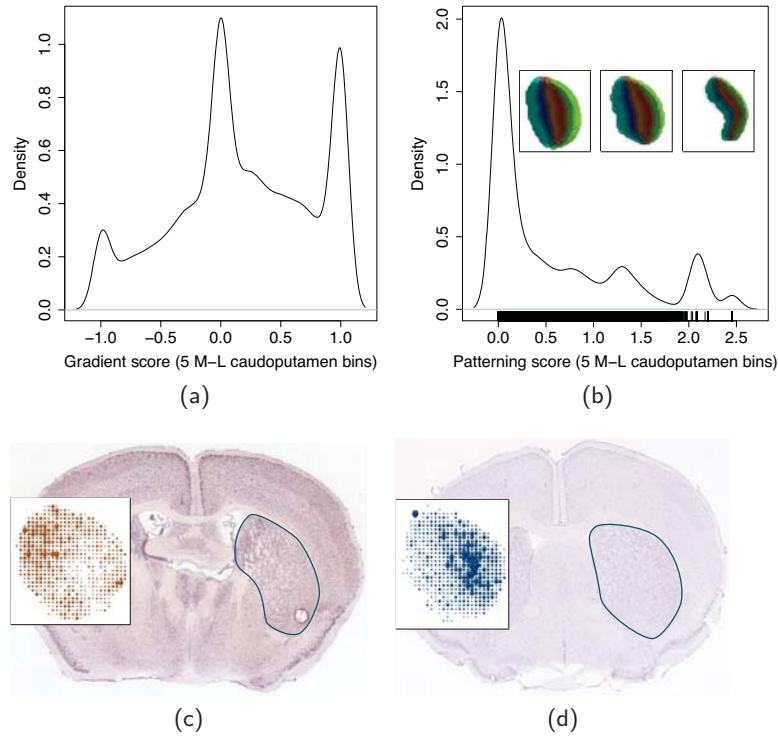
Figure 4: Patterned expression in the caudoputamen. Gradient (a) and patterning scores (b) were computed across five mediolateral caudoputamen bins (rainbow) for all ABA 3D expression profiles. (b) The mediolateral bins adjust to the varying width of the caudoputamen, as shown by three sections from rostral (left) to caudal (right). *In situ* images and caudoputamen voxel calls (inset) are shown for two genes with graded expression across the mediolateral axis: (c) Early growth response 2 (Egr2; gradient = 0.85; ABA image 72034276) and (d) Sodium- and chloride-dependent taurine transporter (Slc6a6; gradient = -0.72; ABA image 15241).

List of Tables

	Gene	Kurrasch enrichment	ALLENMINER enrichment	rank
1	Gpr149	4.43	20.382	43
2	BC034076	2.27	20.223	45
3	Fezf1	4.38	19.158	54
4	Cckbr	2.60	12.532	145
5	Ddn	4.29	12.174	162
6	Gda	2.81	9.854	227
7	Nkx2-2	2.94	9.744	235
8	Gabra5	2.89	9.388	250
9	Fbxw7	2.62	8.728	280
10	Amigo2	3.13	8.636	287
11	Nr5a1	7.75	7.523	338
12	Pdyn	3.16	6.594	428
13	Col6a3	2.53	6.432	438
14	Bdnf	5.37	5.789	522
15	Cdh4	2.58	5.274	595
16	Ldb2	2.65	5.148	626
17	Nptx2	6.85	5.121	629
18	Grhl1	2.24	4.632	735
19	Calb1	2.21	4.311	819
20	Acvr1c	2.44	3.967	936
21	Adcyap1	3.28	2.742	1,578
22	C030019I05Rik	2.20	2.235	2,082
23	Tmem35	2.20	2.024	2,361
24	Tac1	2.36	1.938	2,499
25	Plp1	3.05	1.624	3,156

(a) ALLENMINER enrichment score greater than 1.5

Table 1: Differences between ALLENMINER and microarray identified [1] genes enriched in the ventromedial hypothalamus. (a) Twenty five of the 37 genes identified by microarray received an ALLEN-MINER enrichment score of at least 1.5. (b) Visual inspection of *in situ* images for the remaining 12 genes suggested several potential reasons to explain their enrichment scores of less than 1.5.

	Gene	Kurrasch enrichment	ALLENMINER enrichment	rank	comments
26	Sema3a	2.52	1.412	3,764	sub-threshold: subtle enrichment on one of three coronal data-sets; few registered voxels
27	Satb2	4.68	0.948	5,786	slice/registration effect: subtle VMH enrichment evident on coronal and sagittal images; no VMH voxels called
28	A2bp1	2.40	0.946	5,796	slice/registration effect: enrichment evident on the *in situ* image; sagittal slices only catch edge of VMH
29	Titf1	2.27	0.716	6,723	strong expression throughout hypothalamus
30	B3gnt3	2.25	0.000	11,118	no detectable expression
31	B530002L08	6.13	0.000	11,137	no detectable expression
32	Card14	2.32	0.000	11,689	no detectable expression, elongated brain
33	Kcnj5	2.22	0.000	14,603	sub-threshold: slight signal on coronal and sagittal in situ, no voxels registered
34	Mst1r	3.83	0.000	16,453	sub-threshold, *in situ* artifact: subtle sagittal signal, but close to artifact; no expression called
35	Nmbr	3.30	0.000	16,690	sagittal *in situ* artifact: missing VMH tissue on sagittal section; coronal sub-threshold: very subtle signal in 1-2 slices
36	Ntn2l	2.21	0.000	16,783	sub-threshold: some low level expression, few called pixels, no registered voxels
37	Vgll2	8.42	0.000	19,791	no detectable expression. very low expression on image, few pixels called, no registered voxels

(b) ALLENMINER enrichment score less than 1.5