

1

## 2 **Supplementary Information for**

### 3 **Monoallelic and skewed expression in mouse CNS-derived clonal cell lines: Balancing** 4 **selection for developmental stage-specific genes**

5 **Sergio Branciamore, Zuzana Valo, Min Li, Jinhui Wang, Arthur D. Riggs, and Judith Singer-Sam**

6 **Sergio Branciamore E-mail: [sbranciamore@coh.org](mailto:sbranciamore@coh.org),**

7 **Arthur D. Riggs E-mail: [ariggs@coh.org](mailto:ariggs@coh.org),**

8 **Judith Singer-Sam E-mail: [jsam@coh.org](mailto:jsam@coh.org)**

#### 9 **This PDF file includes:**

10 Figs. S1 to S2

11 Tables S1 to S6

12 References for SI reference citations

**Table S1. Probability that  $S_{D_3}$  is originated from state  $S_{D_0}$  where S is one of three possible states of expression (monoallelic, bi-allelic, not expressed),  $D_0$  refers to *nsc* before induction of differentiation, and  $D_3$  refers to *asl* after three days of differentiation.**

	$D_0$ mono	$D_0$ bi-allelic	$D_0$ non expressed
$D_3$ mono	53%	3%	44%
$D_3$ bi-allelic	< 1%	79%	21%
$D_3$ not expressed	0.01	0.1	0.89

**Table S2. Ontology analysis of all genes expressed in *nsc***

GO		p-value (Benjamini)
GO:0000279	M phase;	$7.4 \times 10^{-21}$
GO:0022403	cell cycle phase;	$1.7 \times 10^{-18}$
GO:0022402	cell cycle process;	$2.2 \times 10^{-16}$
GO:0000087	M phase of mitotic cell cycle;	$2.9 \times 10^{-15}$
GO:0007067	mitosis;	$2.9 \times 10^{-15}$
GO:0005694	chromosome;	$1.5 \times 10^{-13}$
GO:0044427	chromosomal part;	$3.8 \times 10^{-13}$
GO:0006259	DNA metabolic process;	$4.2 \times 10^{-13}$
GO:0006281	DNA repair;	$3.3 \times 10^{-12}$
GO:0006974	response to DNA damage stimulus;	$1.7 \times 10^{-11}$
GO:0009719	response to endogenous stimulus;	$9.3 \times 10^{-11}$
GO:0000775	chromosome, pericentric region;	$4.2 \times 10^{-10}$
GO:0005634	nucleus;	$2.5 \times 10^{-6}$
GO:0006260	DNA replication;	0.00016
GO:0043228	non-membrane-bound organelle;	0.00024
GO:0043232	intracellular non-membrane-bound organelle;	0.00024
GO:0051327	M phase of meiotic cell cycle;	0.00031
GO:0007126	meiosis;	0.00031
GO:0051321	meiotic cell cycle;	0.00036
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process;	0.00037
GO:0006310	DNA recombination;	0.00058

**Table S3. Ontology analysis of all genes expressed in *asl***

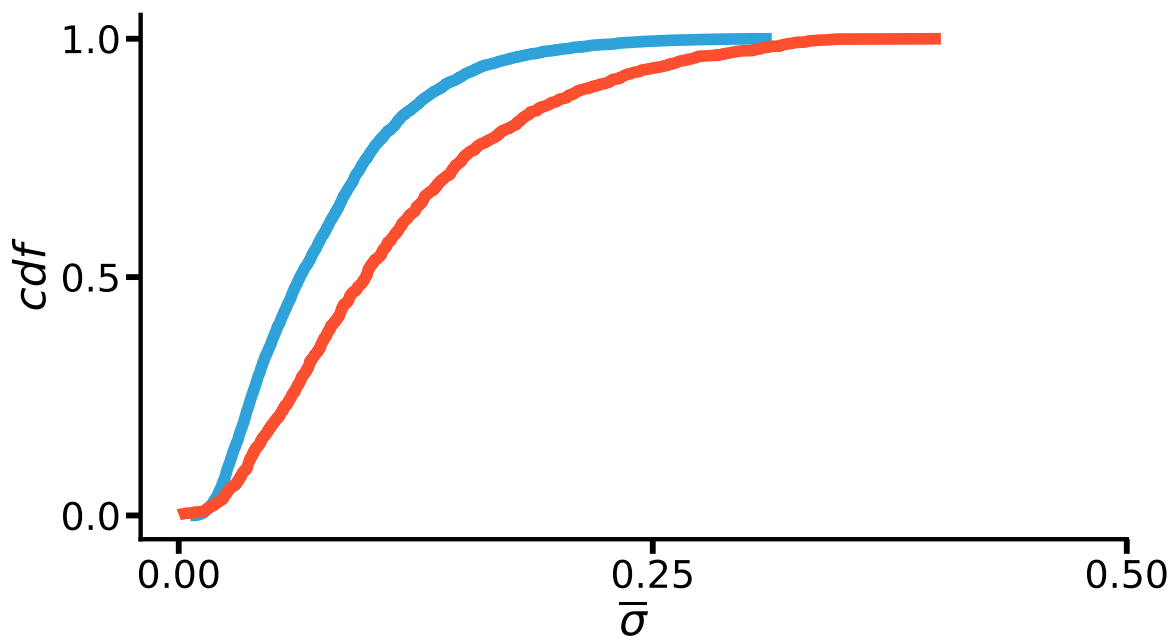
GO		p-value (Benjamini)
GO:0005886	plasma membrane;	$1.5 \times 10^{-9}$
GO:0005578	proteinaceous extracellular matrix;	$4.1 \times 10^{-9}$
GO:0007166	cell surface receptor linked signal transduction;	$2.9 \times 10^{-7}$
GO:0044459	plasma membrane part;	$7.5 \times 10^{-7}$
GO:0022803	passive transmembrane transporter activity;	$1.9 \times 10^{-6}$
GO:0015267	channel activity;	$1.9 \times 10^{-6}$
GO:0004888	transmembrane receptor activity;	$2.5 \times 10^{-6}$
GO:0031226	intrinsic to plasma membrane;	$4.3 \times 10^{-6}$
GO:0022838	substrate specific channel activity;	$6 \times 10^{-6}$
GO:0006811	ion transport;	$7.9 \times 10^{-6}$
GO:0005887	integral to plasma membrane;	$9.4 \times 10^{-6}$
GO:0005216	ion channel activity;	$8.8 \times 10^{-5}$
GO:0048731	system development;	0.00015
GO:0015674	di-, tri-valent inorganic cation transport;	0.00015
GO:0005509	calcium ion binding;	0.00017
GO:0030001	metal ion transport;	0.0002
GO:0048513	organ development;	0.00024
GO:0006816	calcium ion transport;	0.00027
GO:0005261	cation channel activity;	0.00043
GO:0007186	G-protein coupled receptor protein signaling pathway;	0.00057
GO:0022843	voltage-gated cation channel activity;	0.00066
GO:0000267	cell fraction;	0.00068
GO:0031224	intrinsic to membrane;	0.0007
GO:0022836	gated channel activity;	0.00086

**Table S4. Ontology analysis of all genes expressed ( $n_{sc} \cup a_{sl}$ ), ranked by their value of  $\bar{\sigma}$** 

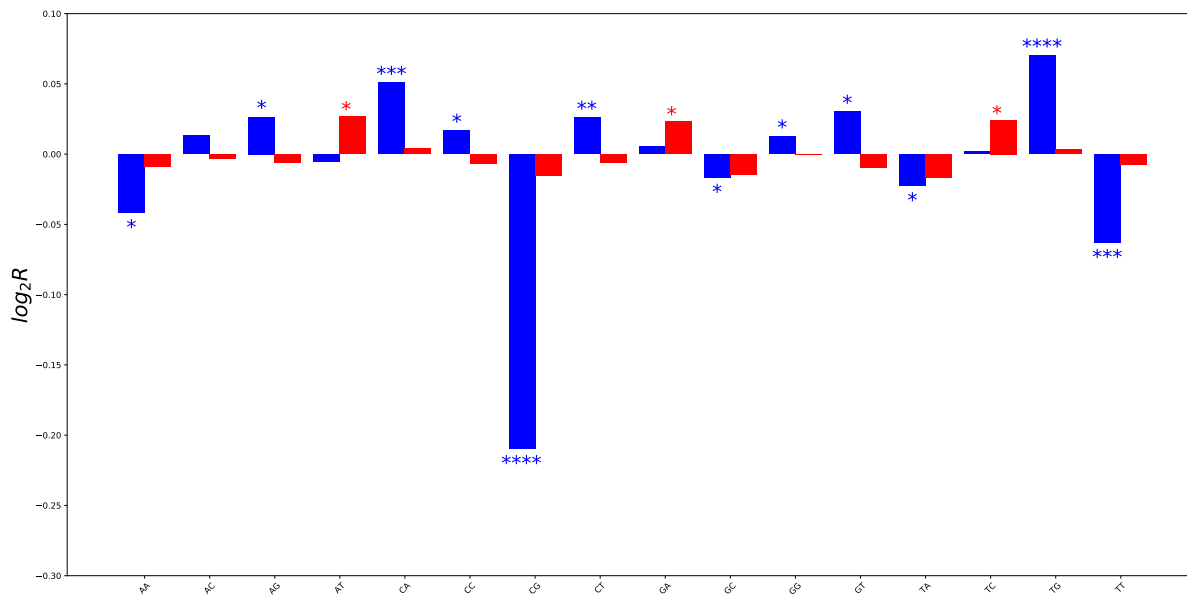
GO		p-value (Benjamini)
GO:0044424	intracellular part	0
GO:0005578	proteinaceous extracellular matrix	1.04e-08
GO:0031224	intrinsic to membrane	2.6e-08
GO:0005509	calcium ion binding	4.24e-08
GO:0016021	integral to membrane	5.24e-08
GO:0031226	intrinsic to plasma membrane	3.78e-05
GO:0004888	transmembrane receptor activity	0.000112
GO:0005887	integral to plasma membrane	0.00014
GO:0044425	membrane part	0.000266
GO:0048503	GPI anchor binding	0.000277
GO:0005886	plasma membrane	0.00131
GO:0009611	response to wounding	0.00346
GO:0022803	passive transmembrane transporter activity	0.00362
GO:0015267	channel activity	0.00362
GO:0005261	cation channel activity	0.0042
GO:0044420	extracellular matrix part	0.00467
GO:0035150	multicellular organismal development regulation of tube size	0.00467
GO:0003018	vascular process in circulatory system	0.00467
GO:0050880	regulation of blood vessel size	0.00467
GO:0008217	blood pressure	0.00575
GO:0022836	gated channel activity	0.0069
GO:0022838	substrate specific channel activity	0.00957
GO:0007186	G-protein coupled receptor protein signaling pathway	0.0114
GO:0006811	ion transport	0.0114
GO:0005216	ion channel activity	0.013
GO:0001584	rhodopsin-like receptor activity	0.0158
GO:0004930	G-protein coupled receptor activity	0.016
GO:0006817	phosphate transport	0.0166
GO:0005830	cytosolic ribosome (sensu Eukaryota)	0.0172
GO:0008015	blood circulation	0.0232

**Table S5. Ontology analysis of all genes expressed following removal of the *MAE* gene set, ranked by their value of  $\bar{\sigma}$** 

GO		p-value (Benjamini)
GO:0044424	intracellular part	0
GO:0031224	intrinsic to membrane	2.65e-05
GO:0016021	integral to membrane	2.7e-05
GO:0005578	proteinaceous extracellular matrix	0.00228
GO:0005261	cation channel activity	0.00335
GO:0005509	calcium ion binding	0.00757
GO:0022803	passive transmembrane transporter activity	0.00757
GO:0015267	channel activity	0.00757
GO:0044425	membrane part	0.0154
GO:0004888	transmembrane receptor activity	0.0154
GO:0005216	ion channel activity	0.022
GO:0022838	substrate specific channel activity	0.0223
GO:0046873	metal ion transmembrane transporter activity	0.0223
GO:0022836	gated channel activity	0.0232
GO:0009611	response to wounding	0.0332
GO:0031226	intrinsic to plasma membrane	0.0332



**Fig. S1.** Developmental genes are associated with skewed allele-specific expression even if the *MAE* gene set is removed. The cumulative density distribution is shown for the variable mean of  $\sigma$  (see eq. 1) with the *MAE* gene set subtracted from genes expressed in both cell types (blue) or with *MAE* subtracted from genes expressed in only one cell type (red).



**Fig. S2.** Comparison (R) of the mean observed to expected dinucleotide distribution for genes with MAE vs. (BAE-hkg) where hkg are housekeeping genes. Results are shown for the 5'-proximal (blue) and 3'-proximal (red) region (see Fig 5 gene neighborhoods). \*,\*\*,\* and \*\*\*\* mark the  $p_{value}$  for Mann-Whitney rank test at significance  $< 10^{-2}$ ,  $10^{-5}$ ,  $10^{-10}$  and  $10^{-15}$ , respectively. The list of housekeeping genes was obtained from (1).

**Table S6. List of primers. For the first six genes, reverse transcription and PCR were carried out as previously described (2), with 1.5 mM MgCl<sub>2</sub>, and an annealing temperature of 60 C. For *Gstm5* and *Ncam1*, see Materials and Methods. For each set, the forward and reverse primers are in the first and second rows, respectively. The reference position is the SNP location relative to the upstream primer. Primers were designed with the aid of Primer3 (<http://frodo.wi.mit.edu/primer3>).**

Gene	NCBI Refseq	SNP (B6/JF1)	primer No.	Ref. pos.	Product size	Substrate	Primer sequence
Aldoc	NM_009657	C/T	453	131	225	RNA	ACAGAGAAGGTCCTGGCTGCT
			454			RNA	AGCCTCCTCTTCACTCTGACC
			465	132	213	DNA	TGCTGTATACAAGGCCCTGAG
			466			DNA	GGAAAACCCCTCAAGTCAGACC
Chi3l1	NM_007695	G/A	455	98	210	RNA	ACCCTCGCTTAAGAGACAAGC
			456			RNA	CTCCATGGAAATCGTAGGTCA
			467	177	289	DNA	TCTGTGTCTGAACTCCAAGCA
			468			DNA	GGAAAAGGTATTGCCTGACAAG
Cpxm1	NM_019696	T/C	457	118	263	RNA	GTCCTGTGACAAGTCCCTCA
			458			RNA	TATGGTAACCCTCAGCACTGG
			469	203	296	DNA	TGCCTAGTCTCAGTGCAACCT
			470			DNA	CTTCTCTCTCTGCCACACAC
Itga7	NM_008398	C/T	459	128	235	RNA	AGATGCATCCACAGTGATTCC
			460			RNA	CTTGAAGTCTGTGCGGTCTTC
			471	148	234	DNA	CCACTCACTCCAGCTTTTGAG
			472			DNA	TGTGTGGGTGAGAGGCTAACT
Osmr	NM_011019	T/C	461	224	290	RNA	GGAAGAACCTTGGTCCCAATA
			462			RNA	CTGCATCTCCTTGGATTTAC
			524	123	192	DNA	TGACCATTTCCACAGCATGAC
			525			DNA	GCACTGCATCTCCTTGGATTT
Tm4sf1	NM_008536	T/C	463	132	263	RNA	TCCTCTATGTGGTCCAAGTGC
			464			RNA	TGCAAAGTCTCACACACAGG
			475	132	295	DNA	TCCTCTATGTGGTCCAAGTGC
			476			DNA	CCCCACAAAGATGAAGTGAGA
Gstm5	NM_010360	C/G	265	128	285	RNA	AGCTACCTGCACAGCTGAAAC
			266			RNA	CTCTGGCTCAGCATAAGCACT
Ncam1	NM_001113204	A/G	507	108	181	RNA	GACGTCCGGTTCATAGTCTCT
			508			RNA	CATTCACGATGCTCTGTCTGG

## 13 References

- 14 1. Bin L, et al. (2017) A comprehensive mouse transcriptomic bodymap across 17 tissues by rna-seq. *Scientific Reports* 7(1).
- 15 2. Wang J, Valo Z, Smith D, Singer-Sam J (2007) Monoallelic expression of multiple genes in the CNS. *PLoS One* 2(12):e1293.