scientific data

Supplementary Materials for

Genomic data resources of the Brain Somatic Mosaicism Network for neuropsychiatric diseases

McKinzie A. Garrison et al.

Corresponding author: Alexej Abyzov, <u>abyzov.alexej@mayo.edu</u>

Pertaining to data hosted by the NIMH Data Archive: Study 967 (DOI: 10.15154/1519293) & Study 814 (DOI: 10.15154/1506068)

The PDF file includes:

Supplementary Figures S1 to S3.

Other Supplementary Materials for this manuscript include the following:

Supplementary Tables S1 to S3.



Stacked barchart of coverage per subject for each cohort with WGS data, including data for cohorts from the Mount Sinai School of Medicine (MSSM), University of California San Diego (UCSD), Yale University, Harvard University, and the Lieber Institute for Brain Development (LIBD). The UCSD section solely describes samples from a single subject, UCSD-19-110. The barplot is colored for tissue of origin, and phenotype is colored by a rugchart at the bottom for subjects affected by autism spectrum disorder (ASD), global geriatric decline, schizophrenia (SCZ), Tourette syndrome (TS), and control subjects.

Supplementary Figure 1.



Supplementary Figure 3. Variant allele frequency (VAF) distributions of the heterozygous SNPs of each sample in the Harvard (NDA ID 2962), MSSM (NDA ID 2965), LIBD (NDA ID 2967), Yale (NDA ID 2961), and UCSD (NDA ID 2968) datasets. Most distributions have a median VAF of 50 ± 2%. Samples with a median outside this short range were determined to be likely contaminated. Potentially contaminated samples are marked with red titles.

50% 50% 300,000 AN02255 50% AN05983 AN06365 50% AN09412 50% AN12434 AN13287 50% AN13654 50% 200,000 100,000 0 300,000 AN14067 50% BEAR12 50% AN19923 50% BEAR10 50% 50% BEAR3 BEAR5 50% BEAR8 50% 200,000 100,000 0 300,000 LBR5221 LBR5313 MSSM001 50% LBR5311 50% 50% LBR5586 50% M3663M 50% 50% MSSM002 50% 200,000 100,000 0 300,000 NEBB004 50% 50% UK20119 UK25363 UK28768 50% UK45353 50% UMB1024 50% NEBB006 50% 50% 200,000 100,000 0 300,000 UMB1174 50% UMB1182 50% UMB1349 50% UMB1445 50% UMB1465 50% UMB1474 50% UMB1499 49% 200,000 100,000 0 100,000 UMB1638 UMB4548 50% 50% 50% UMB1712 50% UMB4231 50% UMB4334 49% UMB4638 50% UMB4643 0 300,000 UMB4671 UMB4672 50% UMB4721 UMB4842 50% 50% 50% UMB4849 50% UMB4899 50% UMB4999 49% 200,000 100,000 0 300,000 UMB5027 49% UMB5115 51% UMB5144 50% UMB5161 50% UMB5176 49% UMB5238 50% UMB5278 50% 200,000 100,000 0 300,000 UMB5294 UMB5297 49% UMB5302 49% UMB5303 UMB5308 49% UMB5340 UMB5391 50% 50% 50% 50% 200,000 100,000 0 300,000 UMB5403 50% UMB5419 50% UMB5565 UMB5574 50% UMB5771 50% UMB5841 50% UMB5864 50% 50% 200,000 100,000 0 300,000 UMB5878 UMB797 50% UMB5939 49% 50% UMB818 50% UMB914 50% 200,000 100,000 0 25% 50% 75% 100% 0% 25% 50% 75% 100% 25% 5Ó% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% VAF

Supp. Fig 3. (a) Harvard (NDA ID 2962) dataset VAF distributions of heterozygous SNPs.

Supp. Fig 3. (b) MSSM (NDA ID 2965) dataset VAF distributions of heterozygous SNPs. Some samples have a bumpy, stepwise appearance. This bumpy appearance is caused by shallow, uneven coverage and lower data quality compared to samples in the same cohort, resulting in variant allele frequencies that are stepwise rather than continuous. These samples still have a median VAF around 50% and are not likely contaminated.

400,000 300,000 200,000 100,000	MSSM_027_NeuN_pl 50%	MSSM_033_NeuN_pl 50%	MSSM_055_NeuN_pl 50%	MSSM_056_NeuN_pl 50%	MSSM_063_NeuN_pl 50%	MSSM_065_NeuN_pl 50%	MSSM_069_NeuN_pl 50%	MSSM_097_NeuN_pl 50%	MSSM_099_NeuN_pl 50%
400,000 300,000 200,000 100,000	MSSM_100_NeuN_pl 50%	MSSM_116_NeuN_pl 49%	MSSM_130_NeuN_pl 50%	MSSM_142_NeuN_pl 50%	MSSM_158_NeuN_pl 48%	MSSM_161_NeuN_pl 49%	MSSM_162_NeuN_pl 50%	MSSM_164_NeuN_pl 16%	MSSM_168_NeuN_pl 49%
400,000 300,000 200,000 100,000	MSSM_178_NeuN_pl 50%	MSSM_180_NeuN_pl 50%	MSSM_192_NeuN_pl 50%	MSSM_193_NeuN_pl 49%	MSSM_199_NeuN_pl 50%	MSSM_201_NeuN_pl 50%	MSSM_211_NeuN_pl 51%	MSSM_213_NeuN_pl 50%	MSSM_222_NeuN_pl 50%
400,000 300,000 200,000 100,000	MSSM_224_NeuN_pl 49%	MSSM_227_NeuN_pl 50%	MSSM_254_NeuN_pl 50%	MSSM_265_NeuN_pl 50%	MSSM_266_NeuN_pl 49%	MSSM_269_NeuN_pl 49%	MSSM_273_NeuN_pl 49%	MSSM_287_NeuN_pl 50%	MSSM_291_NeuN_pl 49%
400,000 300,000 200,000 100,000	MSSM_293_NeuN_pl 48%	MSSM_295_NeuN_pl 50%	MSSM_297_NeuN_pl 50%	MSSM_299_NeuN_pl 50%	MSSM_304_NeuN_pl 49%	MSSM_305_NeuN_pl 50%	MSSM_308_NeuN_pl 49%	MSSM_309_NeuN_pl 50%	MSSM_310_NeuN_pl 50%
300,000 300,000 200,000 100,000	MSSM_321_NeuN_pl 50%	MSSM_327_NeuN_pl 50%	MSSM_331_NeuN_pl 49%	MSSM_338_NeuN_pl 50%	MSSM_339_NeuN_pl 49%	MSSM_340_NeuN_pl 50%	MSSM_343_NeuN_pl 50%	MSSM_346_NeuN_pl 50%	MSSM_348_NeuN_pl 49%
400,000 300,000 200,000 100,000	MSSM_352_NeuN_pl 49%	MSSM_362_NeuN_pl 49%	MSSM_363_NeuN_pl 50%	MSSM_364_NeuN_pl 50%	MSSM_366_NeuN_pl 50%	MSSM_370_NeuN_pl 49%	MSSM_372_NeuN_pl 49%	MSSM_379_NeuN_pl	MSSM_405_NeuN_pl 50%
400,000 300,000 200,000 100,000	MSSM_406_NeuN_pl 50%	MSSM_415_NeuN_pl 50%	PITT_004_NeuN_pl 50%	PITT_007_NeuN_pl 50%	PITT_017_NeuN_pl	PITT_020_NeuN_pl	PITT_024_NeuN_pl 49%	PITT_036_NeuN_pl 49%	PITT_050_NeuN_pl
400,000 300,000 200,000 100,000	PITT_052_NeuN_pl 49% 0% 25% 50% 75% 100%	PITT_060_NeuN_pl 50% 0% 25% 50% 75% 100%	PITT_071_NeuN_pl 50% 25% 50% 75% 100%	PITT_072_NeuN_pl 49% 0% 25% 50% 75% 100%	PITT_082_NeuN_pl 50% 0% 25% 50% 75% 100%	PITT_098_NeuN_pl 50% 0% 25% 50% 75% 100%	PITT_101_NeuN_pl 50% 0% 25% 50% 75% 100%	PITT_113_NeuN_pl 50% 0% 25% 50% 75% 100%	PITT_117_NeuN_pl 49% 0% 25% 50% 75% 100%

Supp. Fig 3. (b continued) MSSM (NDA ID 2965) dataset VAF distributions of heterozygous SNPs.









Supp. Fig 3. (c) LIBD (NDA ID 2967) dataset VAF distributions of heterozygous SNPs.

250,000 200,000 150,000 100,000 50,000	LIBD01	50%	LIBD02	50%	LIBD03	50%	LIBD04	50%	LIBD05	50%	LIBD06	50%	LIBD07	50%
250,000 200,000 150,000 100,000 50,000	LIBD08	50%	LIBD09	50%	LIBD10	50%	LIBD100	50%	LIBD100_DLPFC	50%	LIBD100_Hippocampus	50%	LIBD100_MBPpos	50%
250,000 200,000 150,000 100,000 50,000	LIBD100_SLC17A7pos	50%	LIBD100_SNAP25pos	50%	LIBD100_tripleneg	50%	LIBD101	50%	LIBD101_DLPFC	50%	LIBD101_Hippocampus	49%	LIBD101_MBPpos	50%
250,000 200,000 150,000 100,000 50,000	LIBD101_SLC17A7pos	50%	LIBD101_SNAP25pos	50%	LIBD101_tripleneg	50%	LIBD102	50%	LIBD103	50%	LIBD104	37%	LIBD104_DLPFC	50%
250,000 200,000 150,000 100,000 50,000	LIBD104_Hippocampus	49%	LIBD104_MBPpos	50%	LIBD104_SLC17A7pos	50%	LIBD104_SNAP25pos	16%	LIBD104_tripleneg	50%	LIBD105	50%	LIBD105_DLPFC	49%
250,000 200,000 150,000 100,000 50,000 0	LIBD105_Hippocampus	50%	LIBD105_MBPpos	50%	LIBD105_SLC17A7pos	50%	LIBD105_SNAP25pos	50%	LIBD105_tripleneg	50%	LIBD106	50%	LIBD106_DLPFC	50%
X 250,000 200,000 150,000 100,000 50,000	LIBD106_Hippocampus	49%	LIBD106_MBPpos	50%	LIBD106_SLC17A7pos	50%	LIBD106_SNAP25pos	50%	LIBD106_tripleneg	50%	LIBD107	50%	LIBD107_DLPFC	50%
250,000 200,000 150,000 100,000 50,000	LIBD107_Hippocampus	50%	LIBD107_MBPpos	50%	LIBD107_SLC17A7pos	50%	LIBD107_SNAP25pos	50%	LIBD107_tripleneg	50%	LIBD108	50%	LIBD109	50%
250,000 200,000 150,000 100,000 50,000	LIBD11	50%	LIBD110	50%	LIBD110_DLPFC	49%	LIBD110_Hippocampu	s 49%	LIBD110_MBPpos	50%	LIBD110_SLC17A7pos	50%	LIBD110_SNAP25pos	50%
250,000 200,000 150,000 100,000 50,000	LIBD110_tripleneg	50%	LIBD111	50%	LIBD112	50%	LIBD113	50%	LIBD113_DLPFC	50%	LIBD113_Hippocampus	51%	LIBD113_MBPpos	50%
250,000 200,000 150,000 100,000 50,000	LIBD113_SLC17A7pos	50%	LIBD113_SNAP25pos	50%	LIBD113_tripleneg	50%	LIBD114	50%	LIBD115	52%	LIBD116	50%	LIBD117	50%
ς (0% 25% 50% 75%	% 100%	0% 25% 50% 75%	% 100%	0% 25% 50% 75%	6 100%	0% 25% 50% 75% VAF	6 100%	0% 25% 50% 75	% 100%	0% 25% 50% 75%	100%	0% 25% 50% 75%	% 100%

Supp. Fig 3. (c continued) LIBD (NDA ID 2967) dataset VAF distributions of heterozygous SNPs.

250,000 200,000 150,000 100,000 50,000	LIBD118	50%	LIBD119	50%	LIBD12	50%	LIBD120	50%	LIBD120_DLPFC	50%	LIBD120_Hippocampus	50%	LIBD120_MBPpos	50%
250,000 200,000 150,000 100,000 50,000	LIBD120_SLC17A7pos	50%	LIBD120_SNAP25pos	50%	LIBD120_tripleneg	50%	LIBD121	50%	LIBD122	50%	LIBD122_DLPFC	50%	LIBD122_Hippocampu	ıs 50%
250,000 200,000 150,000 100,000 50,000	LIBD122_MBPpos	50%	LIBD122_SLC17A7pos	50%	LIBD122_SNAP25pos	50%	LIBD122_tripleneg	50%	LIBD123	50%	LIBD123_DLPFC	50%	LIBD123_Hippocampu	ıs 50%
250,000 200,000 150,000 100,000 50,000	LIBD123_MBPpos	50%	LIBD123_SLC17A7pos	50%	LIBD123_SNAP25pos	50%	LIBD123_tripleneg	50%	LIBD124	50%	LIBD125	50%	LIBD13	51%
250,000 200,000 150,000 100,000 50,000 0	LIBD14	50%	LIBD15	50%	LIBD16	50%	LIBD17	50%	LIBD18	50%	LIBD19	50%	LIBD20	50%
250,000 200,000 150,000 50,000 0	LIBD21	50%	LIBD22	50%	LIBD23	50%	LIBD24	50%	LIBD25	50%	LIBD26	50%	LIBD27	50%
250,000 200,000 150,000 100,000 50,000 0	LIBD28	50%	LIBD29	50%	LIBD30	50%	LIBD31	50%	LIBD32	50%	LIBD33	50%	LIBD34	50%
250,000 200,000 150,000 100,000 50,000	LIBD35	50%	LIBD36	50%	LIBD37	50%	LIBD38	50%	LIBD39	50%	LIBD40	50%	LIBD41	50%
250,000 200,000 150,000 100,000 50,000	LIBD42	50%	LIBD43	50%	LIBD44	50%	LIBD45	50%	LIBD46	52%	LIBD64	50%	LIBD65	50%
250,000 200,000 150,000 100,000 50,000	LIBD66	50%	LIBD67	50%	LIBD68	50%	LIBD69	50%	LIBD70	50%	LIBD71	50%	LIBD72	50%
250,000 200,000 150,000 100,000 50,000	LIBD73	50%	LIBD74	50%	LIBD74_DLPFC	50%	LIBD74_Hippocampus	50%	LIBD74_MBPpos	50%	LIBD74_SLC17A7pos	50%	LIBD74_SNAP25pos	50%
	Ο% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100% 0% 25% 50% 75% 100%													

Supp. Fig 3. (c continued) LIBD (NDA ID 2967) dataset VAF distributions of heterozygous SNPs.

250,000 200,000 150,000 100,000 50,000	LIBD74_tripleneg	50%	LIBD75	50%	LIBD75_MBPpos	50%	LIBD75_SLC17A7pos	50%	LIBD75_SNAP25pos	50%	LIBD75_tripleneg	50%	LIBD76	50%
250,000 200,000 150,000 100,000 50,000	LIBD76_MBPpos	50%	LIBD76_SLC17A7pos	50%	LIBD76_SNAP25pos	50%	LIBD76_tripleneg	50%	LIBD77	50%	LIBD77_DLPFC	50%	LIBD77_Hippocampus	50%
250,000 200,000 150,000 100,000 50,000	LIBD77_MBPpos	50%	LIBD77_SLC17A7pos	50%	LIBD77_SNAP25pos	50%	LIBD77_tripleneg	50%	LIBD78	50%	LIBD79	50%	LIBD80	50%
250,000 200,000 150,000 100,000 50,000	LIBD80_DLPFC	50%	LIBD80_Hippocampus	49%	LIBD80_MBPpos	50%	LIBD80_SLC17A7pos	50%	LIBD80_SNAP25pos	50%	LIBD80_tripleneg	50%	LIBD81	50%
250,000 200,000 150,000 100,000 50,000	LIBD82	50%	LIBD82_DLPFC	50%	LIBD82_Hippocampus	50%	LIBD82_MBPpos	50%	LIBD82_SLC17A7pos	50%	LIBD82_SNAP25pos	50%	LIBD82_tripleneg	50%
250,000 200,000 150,000 50,000 0	LIBD83	50%	LIBD83_DLPFC	50%	LIBD83_Hippocampus	49%	LIBD83_SLC17A7pos	50%	LIBD83_SNAP25pos	50%	LIBD83_tripleneg	50%	LIBD84	50%
X 250,000 200,000 150,000 100,000 50,000	LIBD85	50%	LIBD86	52%	LIBD87	50%	LIBD87_DLPFC	50%	LIBD87_Hippocampus	50%	LIBD87_MBPpos	50%	LIBD87_SLC17A7pos	50%
250,000 200,000 150,000 100,000 50,000	LIBD87_SNAP25pos	50%	LIBD87_tripleneg	50%	LIBD88	50%	LIBD89	50%	LIBD90	50%	LIBD91	50%	LIBD92	50%
250,000 200,000 150,000 100,000 50,000	LIBD93	50%	LIBD94	50%	LIBD95	50%	LIBD96	50%	LIBD96_DLPFC	50%	LIBD96_Hippocampus	49%	LIBD96_MBPpos	50%
250,000 200,000 150,000 100,000 50,000	LIBD96_SLC17A7pos	50%	LIBD96_SNAP25pos	50%	LIBD96_tripleneg	50%	LIBD97	50%	LIBD98	50%	LIBD98_DLPFC	50%	LIBD98_Hippocampus	50%
250,000 200,000 150,000 100,000 50,000	LIBD98_MBPpos	50%	LIBD98_SLC17A7pos	50%	LIBD98_SNAP25pos	50%	LIBD98_tripleneg	50%	LIBD99	50%	LIBD99_DLPFC	50%	LIBD99_Hippocampus	49%
0	0% 25% 50% 75	% 100%	0% 25% 50% 75%	6 100%	0% 25% 50% 75%	6 100%	0% 25% 50% 75% VAF	% 100%	0% 25% 50% 75%	% 100%	0% 25% 50% 75%	6 100%	0% 25% 50% 75%	6 100%



Supp. Fig 3. (c continued) LIBD (NDA ID 2967) dataset VAF distributions of heterozygous SNPs.



VAF

Supp. Fig 3. (d continued) Yale (NDA ID 2961) dataset VAF distributions of heterozygous SNPs.



Supp. Fig 3. (d continued) Yale (NDA ID 2961) dataset VAF distributions of heterozygous SNPs.



VAF

Supp. Fig 3. (d continued) Yale (NDA ID 2961) dataset VAF distributions of heterozygous SNPs.



VAF

Supp. Fig 3. (e) UCSD (NDA ID 2968) dataset VAF distributions of heterozygous SNPs.

