Supplemental Figures

Figure S1 - Comparison of different manifold learning methods for TCX brain region.

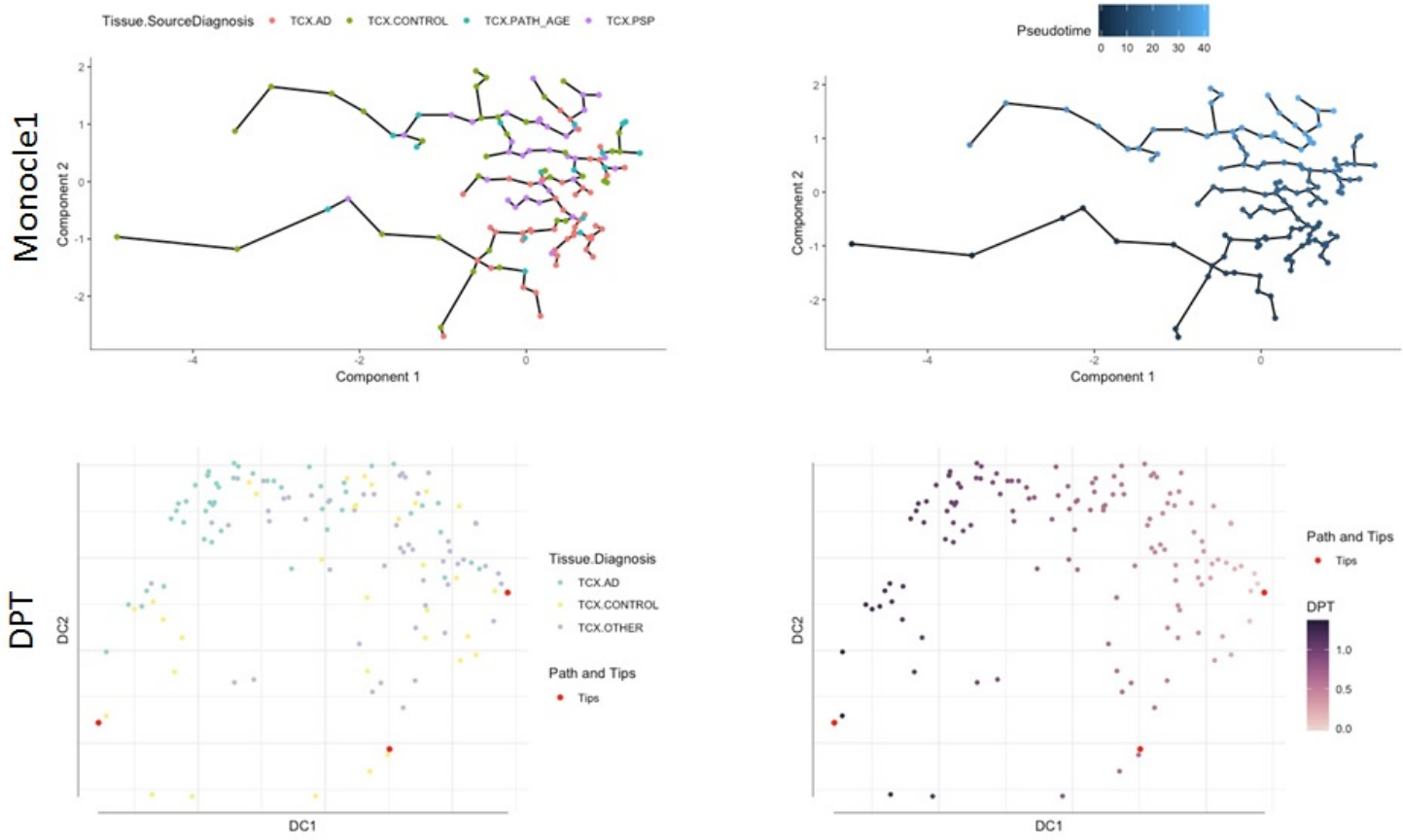


Figure S2 - Comparison of different manifold learning methods for DLPFC brain region.

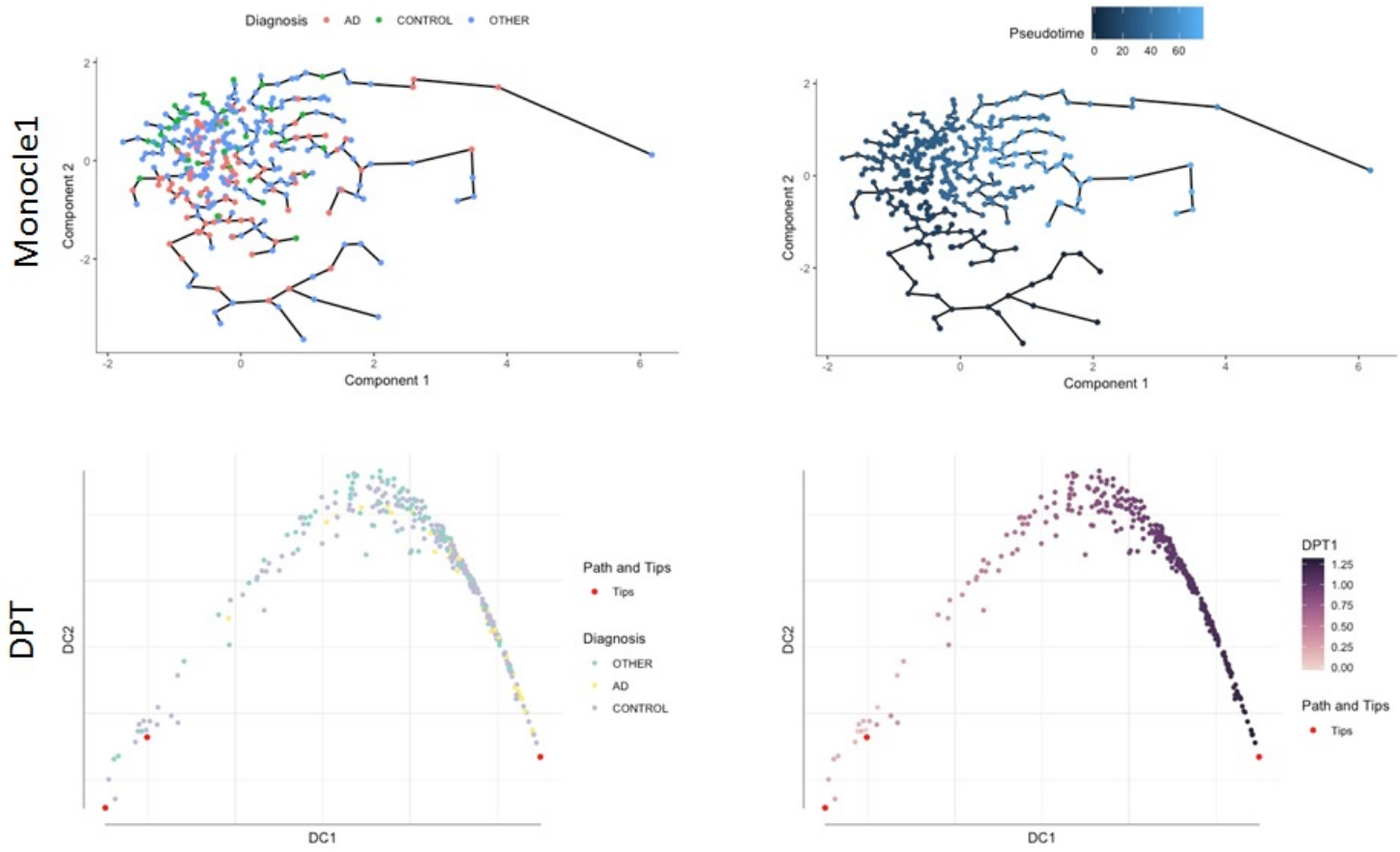


Figure S3 - Correlation between pseudotimes estimated by different manifold learning approaches on both TCX and DLPFC brain region.

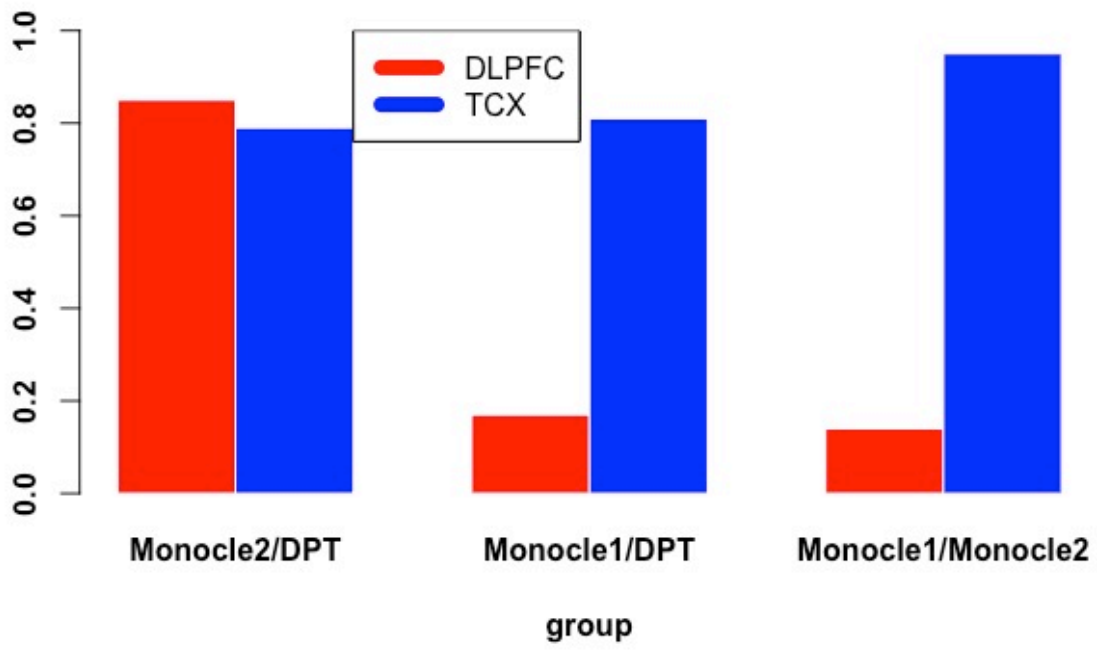


Figure S4 - Comparison between trajectories inferred using different gene sub-set selection methods: i) Differential Expression with an FDR cut-off of 0.1, ii) High variance gene selection.

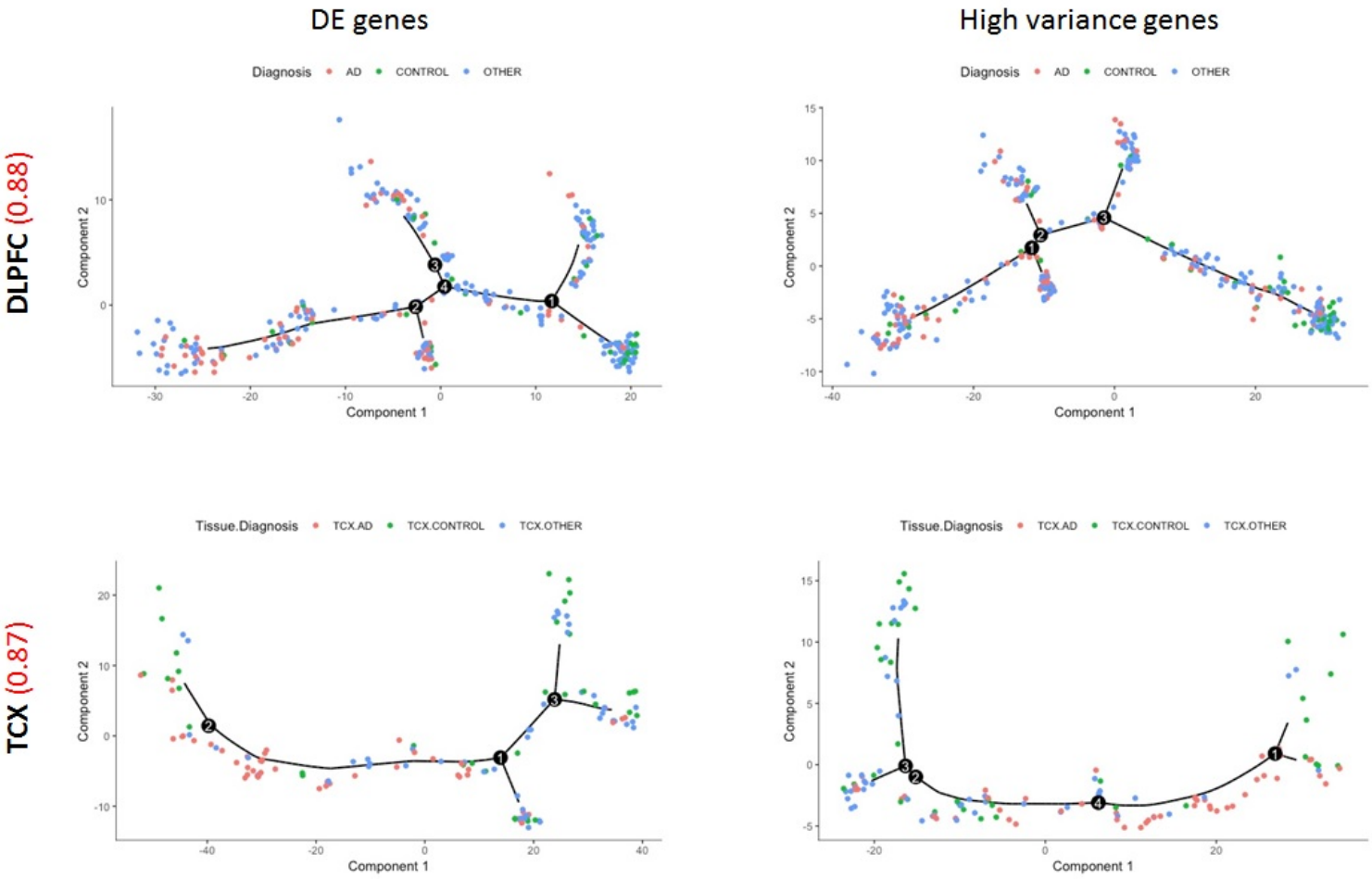


Figure S5 - Average expression of marker genes from neurons, astrocytes, microglia and oligodendrocytes as a function of inferred stage for both brain regions. Inferred cell populations align with known cell type specific effects of various neuropathological outcomes.

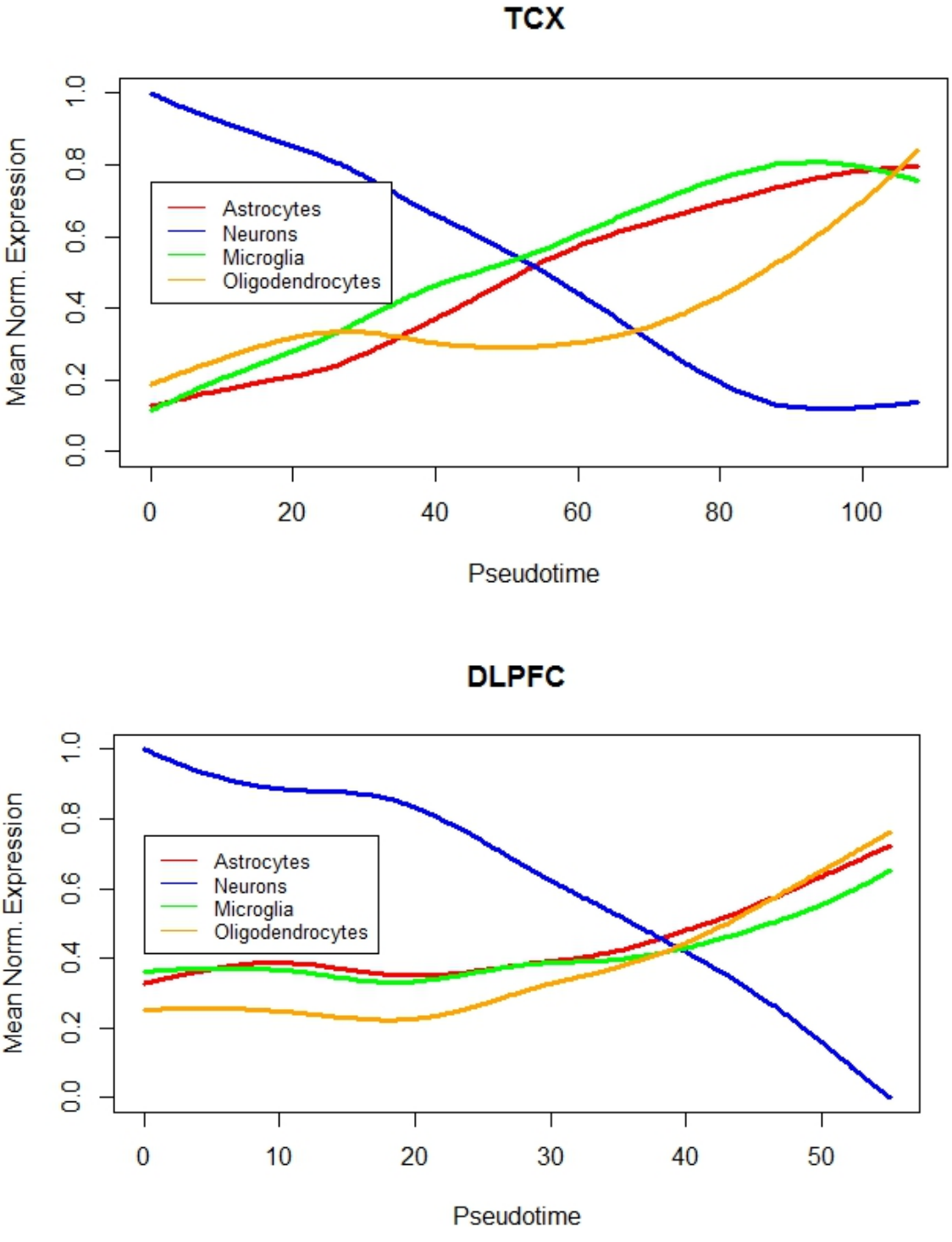


Figure S6 - APOE e4 status of samples overlaid on inferred manifolds for both TCX and DLPFC brain regions.

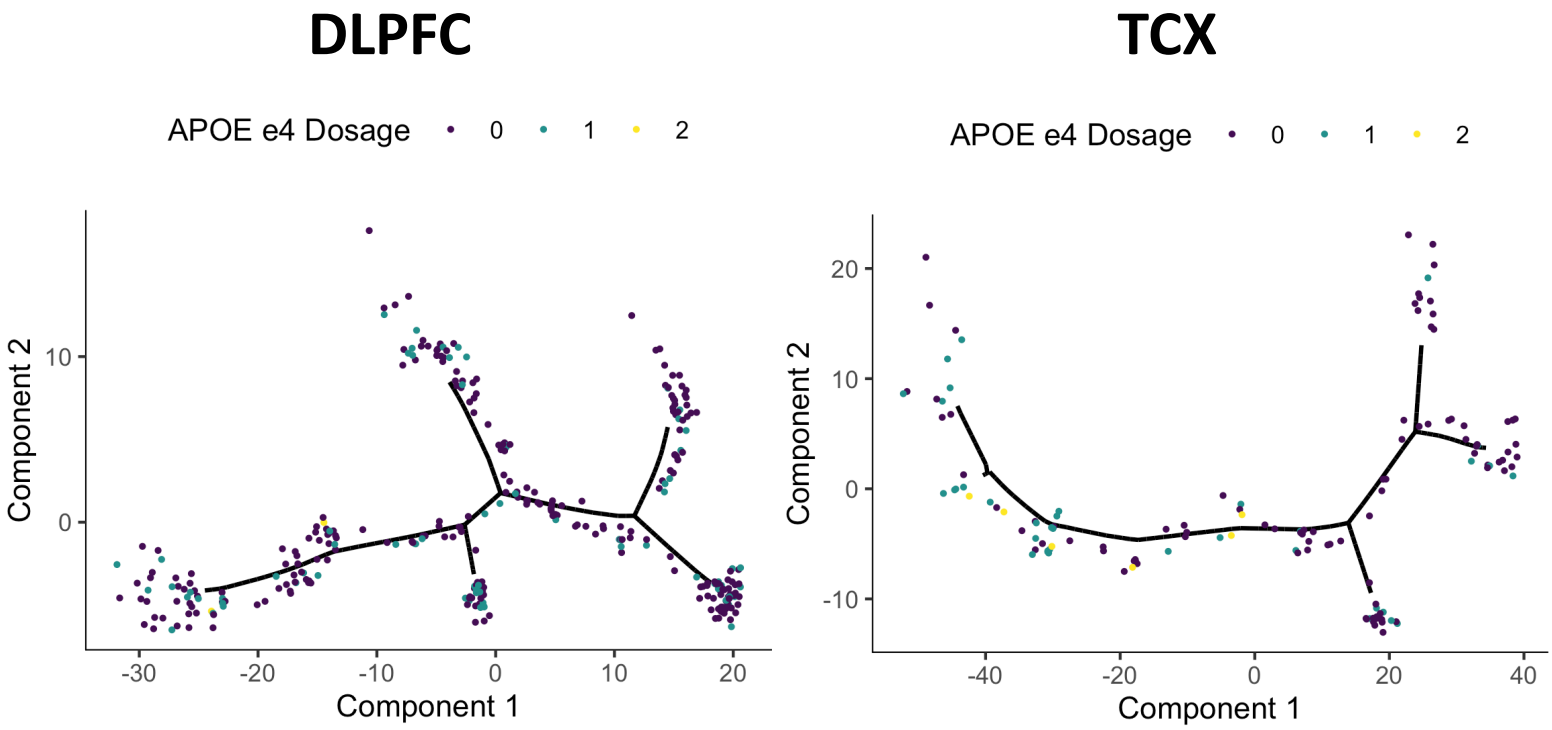


Figure S7 - DLPFC manifolds with samples colored by inferred disease state.

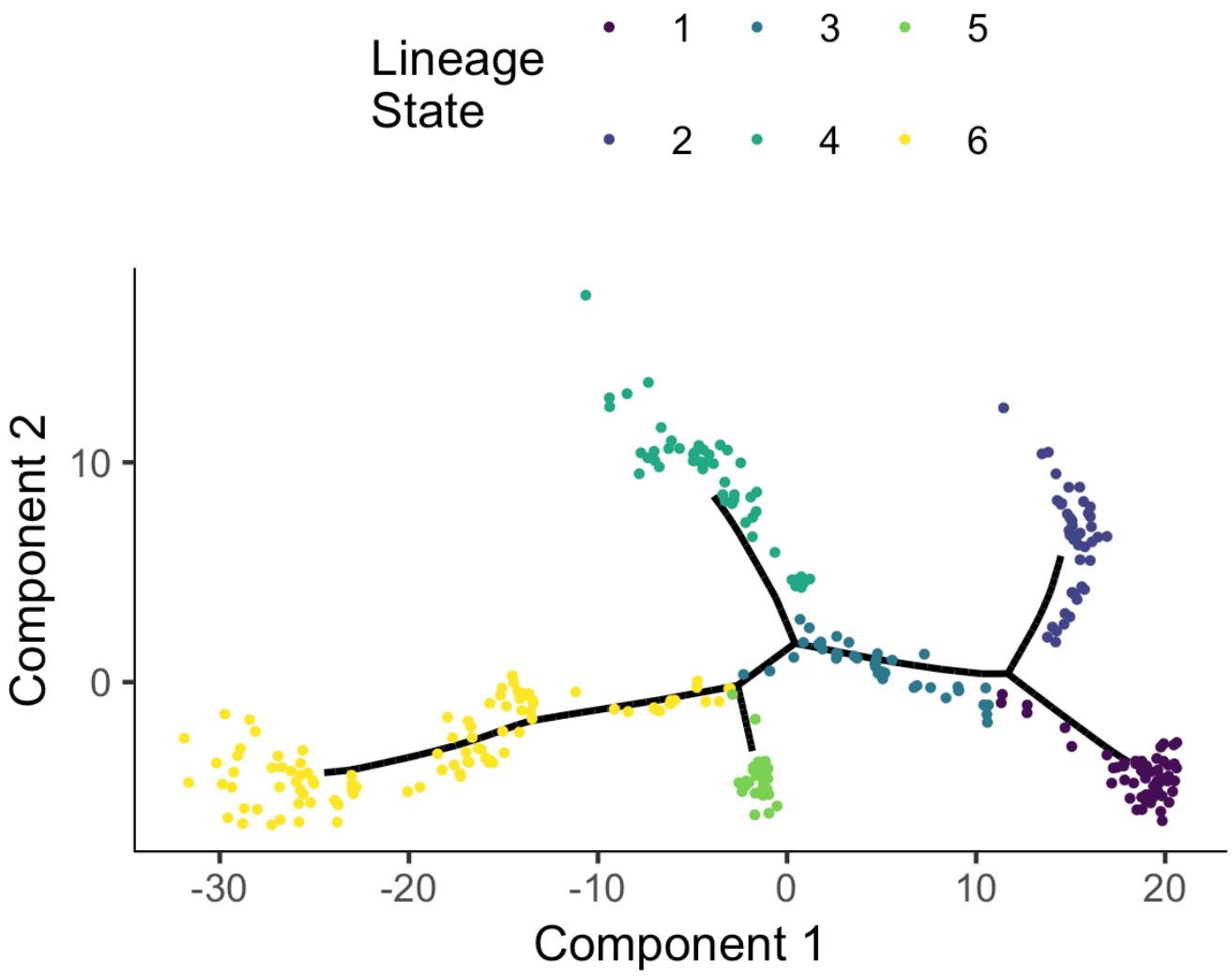


Figure S8 - Quantile-quantile plot for the association with pseudotime in 305 female patients in the ROS/MAP cohort. The graph shows the Q-Q plot for GWAs of pseudotime in the ROS/MAP cohort with a genomic Inflation factor (lambda) of 0.981.

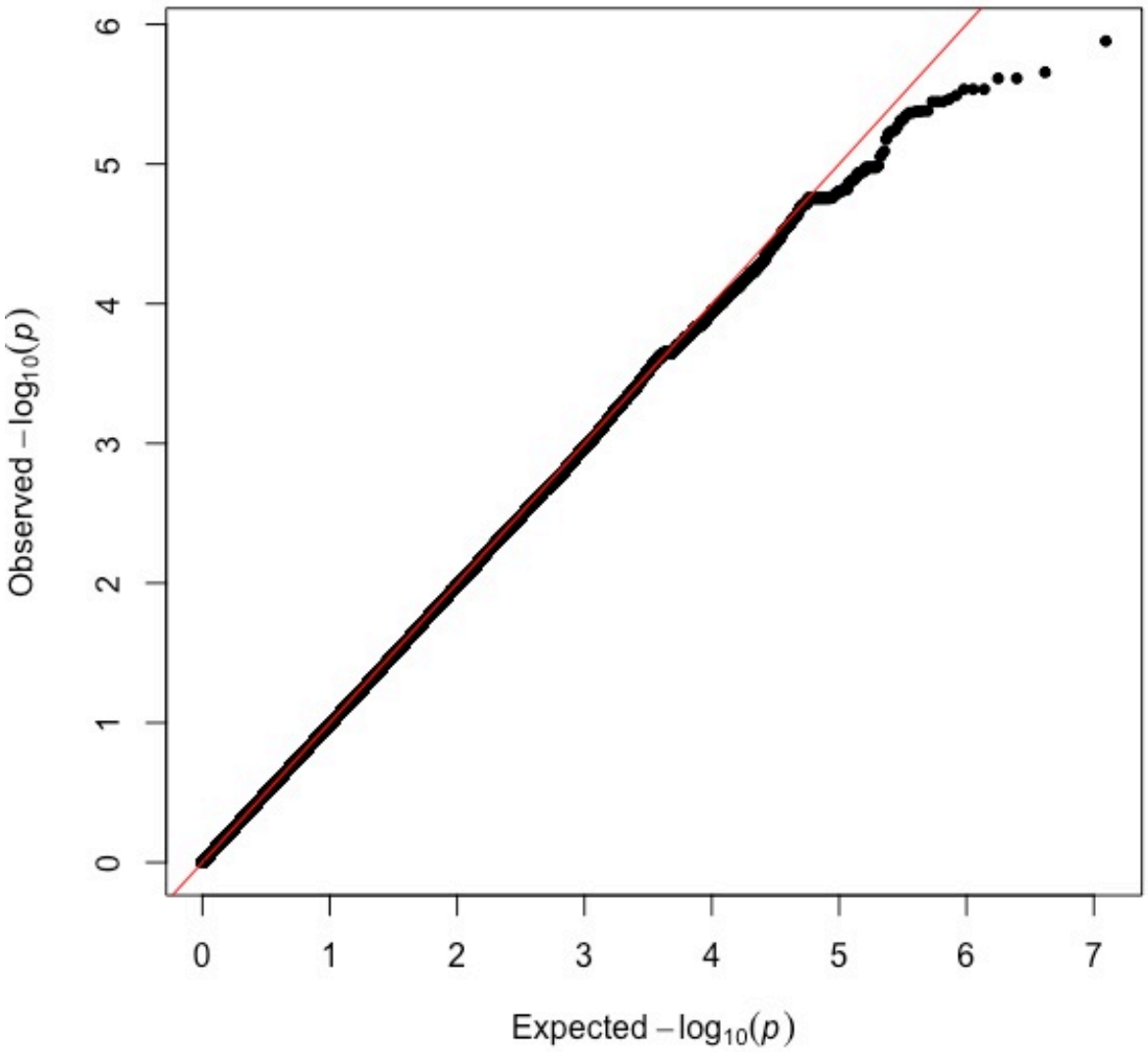


Figure S9 - Quantile-quantile plot for the association with pseudotime in 131 female patients in the Mayo cohort.

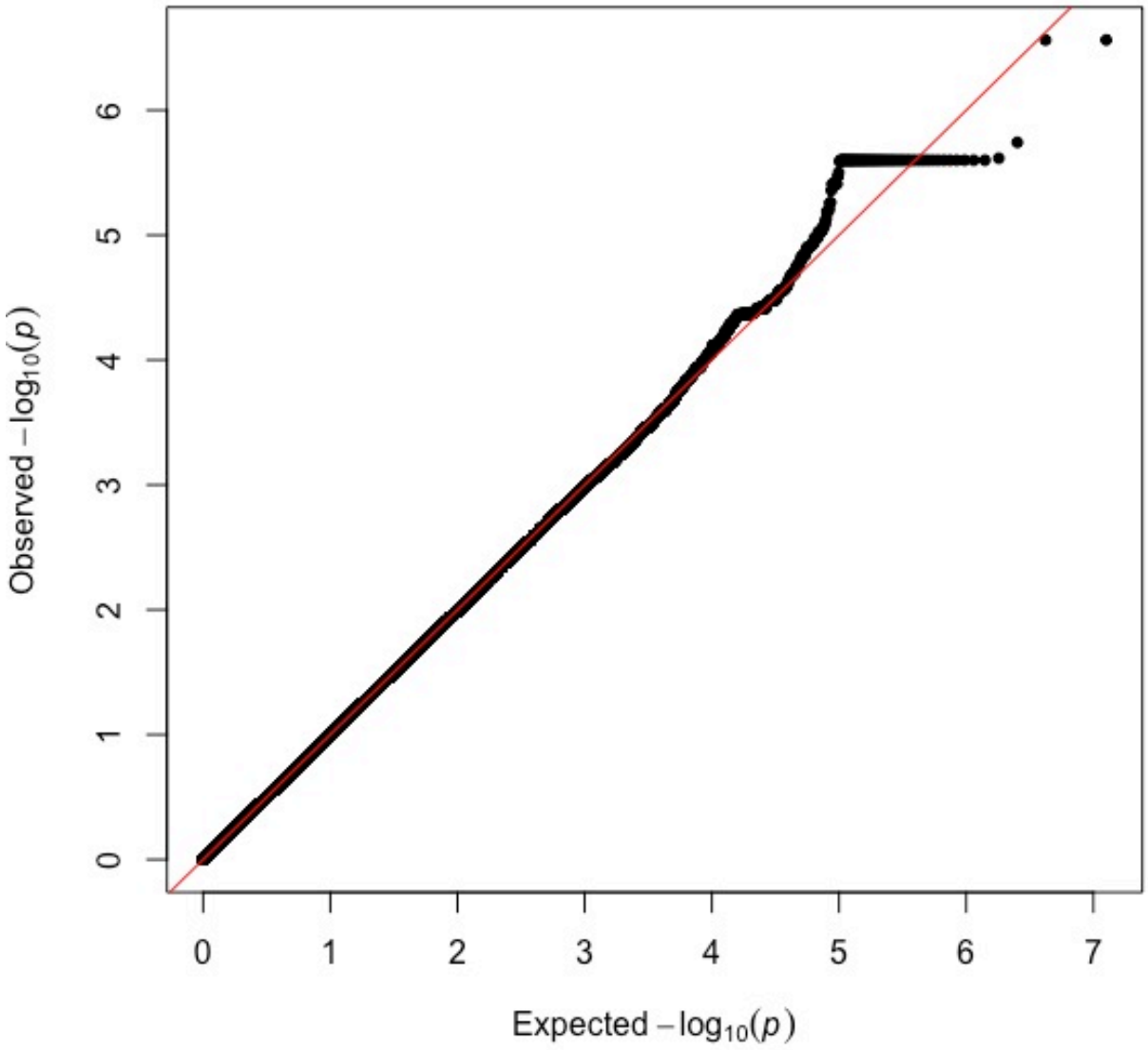
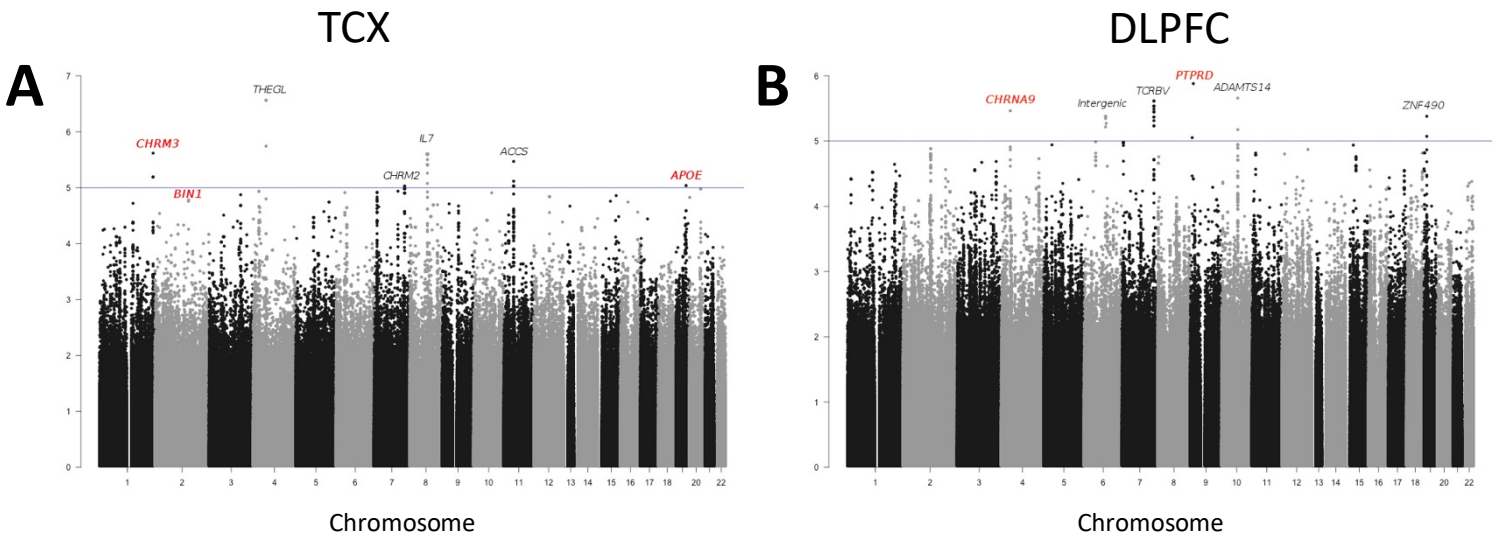


Figure S10 - Manifold learning identified potential genetic factors of stage progression and subtypes of LOAD. A-B) GWA analysis was performed on the Mayo (A) and ROSMAP (B) cohorts using whole genome sequenced data and LOAD pseudotime as the phenotype. Despite the small sample sizes of both analyses (N = 131 in Mayo, N = 306 in ROSMAP), several genomic loci were identified harboring SNPs with a genome wide suggestive p-value ($p < 1x10^{-5}$). These include several loci that were previously associated with LOAD or LOAD related endophenotypes (red labels; see also **Table S5**)



Supplemental Tables

Table S3 - Association between mean expression of cell specific signatures and inferred disease severity (pseudotime).

Study (Brain Region)	Cell Signature	P-value	R^2
Mayo RNAseq (TCX)	Neuronal	$3.6x10^{-42}$	0.76
	Microglial	$9.1x10^{-29}$	0.61
	Oligodendroglial	$6.7x10^{-11}$	0.28
	Astrocytic	$6.7x10^{-22}$	0.51
ROS/MAP (DLPFC)	Neuronal	$1.6x10^{-78}$	0.65
	Microglial	$1.5x10^{-31}$	0.33
	Oligodendroglial	$1.4x10^{-44}$	0.44
	Astrocytic	$1.0x10^{-50}$	0.48

Table S4 - Overview of suggestive ($p < 10^{-5}$) results from single variant association with pseudotime

SNP (dbSNP 150)	Location (hg19)	Nearest Gene(s)	region	A1 (Effect Allele)	A2	Allele Freq. (A1)	Beta (Pseudotime)	SE (beta)	P	Cohort	Previous Association
rs4421019	4:40309851	*CHRNA9*	*intergenic*	T	A	0.35	-6.18	1.31	3.44E-06	ROS/MAP	LOAD
rs12216400	6:96292130	*intergenic*	*intergenic*	A	G	0.24	6.86	1.46	4.17E-06	ROS/MAP	/
rs1573618	7:142244415	*TCRBV*	*intronic*	T	C	0.44	-6.22	1.29	2.43E-06	ROS/MAP	/
rs7870388	9:8660693	*PTPRD*	*intronic*	G	C	0.21	-6.40	1.42	1.32E-06	ROS/MAP	Tangle burden
rs4746059	10:72465488	*ADAMTS14*	*intronic*	G	A	0.42	5.85	1.21	2.20E-06	ROS/MAP	/
rs55786848	19:12669655	*ZNF490; ZNF564*	*intergenic*	C	T	0.15	8.01	1.71	4.16E-06	ROS/MAP	/
rs12136200	1:240138130	*CHRM3*	*intergenic*	C	T	0.39	-16.61	3.36	2.42E-06	Mayo	Plaque burden
rs73818121	4:57397157	*THEGL*	*exonic*	G	C	0.07	33.19	6.63	1.81E-06	Mayo	/
rs7809318	7:136419969	*CHRM2*	*intergenic*	C	T	0.07	-34.03	7.37	9.41E-06	Mayo	/
rs3808616	8:79868493	*IL7*	*intergenic*	G	A	0.35	-17.70	3.59	2.51E-06	Mayo	/
rs11037791	11:44022056	*ACCS;ACCSL*	*intergenic*	A	G	0.49	-16.41	3.38	3.39E-06	Mayo	/
rs6857	19:45392254	*PVRL2; TOMM40; APOE*	*intronic*	C	T	0.17	-18.23	3.95	9.18E-06	Mayo	LOAD, Tangle burden, Plaque burden

Table S5 - Associations of known AD variants associated with pseudotime in the IGAP cohort.

Chr.	Position (hg19)	SNP	Minor Allele Frequency	IGAP p-value (Stage1+2)	Pseudotime Cohort	Pseudotime p-value	Gene
2	127887750	rs62158731	0.26	3.41E-13	Mayo	4.68E-05	*BIN1*
3	151018968	rs66927386	0.24	1.40E-04	ROS/MAP	0.0090	*MED12L*
6	32570051	rs9270823	0.25	5.77E-10	ROS/MAP	0.0068	*HLA-DRB1*
7	99809921	rs1727128	0.48	4.43E-06	ROS/MAP	0.0029	*STAG3*
9	129197516	rs887656	0.11	1.40E-04	ROS/MAP	0.0079	*MVB12*
10	72524413	rs2688767	0.36	1.39E-04	ROS/MAP	0.0078	*ADAMTS14*
11	85862728	rs72962020	0.13	8.09E-06	Mayo	0.0075	*PICALM*
16	11199352	rs12929596	0.13	6.43E-05	ROS/MAP	0.0067	*CLEC16A*
19	45392254	rs6857	0.17	1.06E-15	Mayo	9.18E-06	*APOE*
20	55020557	rs16979933	0.09	1.08E-07	Mayo	0.0054	*CASS4*

Table S8 - Number of genes differentially expressed at an FDR of 0.05 between the control branch (Branch 1) and other branches based on an ANOVA test.

Study (Brain Region)	Change in expression	Branch 2	Branch 3	Branch 4	Branch 5	Branch 6
ROSMAP (DLPFC)	Increased	718	468	1121	662	1239
	Decreased	781	611	1017	783	1094
MayoRNAseq (TCX)	Increased	506	2067	2034	2733	1815
	Decreased	699	1912	2441	1966	1494

Supplemental Table Legends

Table S1: AD LOAD GWAS genes[23]. Genes are from Tables 1-3 from previously published work[23].

Table S2: Cell specific gene sets used to compute mean expression of cell signatures across the lineages, as previously described[32].

Table S6: ANOVA summary statistics from differential expression analysis in DLPFC.

Table S7: ANOVA summary statistics from differential expression analysis in TCX.

Table S9: Significant GO pathway enrichments (FDR < 0.05) for DLPFC differential expressed gene sets.

Table S10: Significant GO pathway enrichments (FDR < 0.05) for TCX differential expressed gene sets.

Table S11: Significant GO pathway enrichments from biclustering analysis of mean expression of six branches (states) in TCX with four clusters.