

**Deep Phenotyping of Alzheimer's Disease Leveraging Electronic
Medical Records Identifies Sex-Specific Clinical Associations**

Supplementary Information

**Alice S. Tang^{1,2,3*}, Tomiko Oskotsky^{1,4}, Shreyas Havaldar⁵, William G. Mantyh⁶,
Mesude Bicak^{5,7}, Caroline Warly Solsberg^{8,9,10}, Sarah Woldemariam¹, Billy Zeng³,
Zicheng Hu¹, Boris Oskotsky¹, Dena Dubal⁹, Isabel E. Allen¹⁰, Benjamin S.
Glicksberg^{5,7}, Marina Sirota^{1,4*}**

Supplementary Tables

	Count	Age	Death Status	Race	
Alzheimer's Cohort	6,612	86.4 90 (83-91)	Alive: 6714 (76.3%) Deceased: 2090 (23.7%)	White/Caucasian: 5462 (64.0%) Asian: 879 (10.3%) Black/African American: 586 (6.9%) Hawaiian/Pacific Islander: 452 (5.3%)	American Native: 9 (.1%) Other: 743 (8.7%) Unknown/Declined: 802 (4.7%)
Males	2,382 (36.0%)	85.7 90 (84-91)	Alive: 1778 (74.6%) Deceased: 604 (25.4%)	White/Caucasian: 1570 (66.8%) Asian: 254 (10.8%) Black/African American: 128 (5.4%) Hawaiian/Pacific Islander: 117 (5.0%)	Other: 209 (8.9%) Unknown/Declined: 72 (3.1%)
Females	4,223 (63.9%)	86.8 90 (82-91)	Alive: 3084 (73%) Deceased: 1139 (27%)	White/Caucasian: 2525 (60.5%) Asian: 497 (11.9%) Black/African American: 393 (9.4%) Hawaiian/Pacific Islander: 217 (5.2%)	American Native: 8 (0.2%) Other: 404 (9.7%) Unknown/Declined: 130 (3.1%)
Other or Unknown	7 (0.10%)	90.7 91 (90.5-91)	Alive: 7 (100%)	White/Caucasian: 6 (85.7%) Unknown/Declined: 1 (14.3%)	
Control Cohort	13,224	86.2 90 (83 – 91)	Alive: 13432 (76.3%) Deceased: 4176 (23.7%)	White/Caucasian: 10924 (64.0%) Asian: 1759 (10.3%) Black/African American: 1172 (6.9%) Hawaiian/Pacific Islander: 904 (5.3%)	American Native: 18 (.1%) Other: 1487 (8.7%) Unknown/Declined: 802 (4.7%)
Males	4,674 (35.3%)	85.8 90 (82-91)	Alive: 3248 (69.5%) Deceased: 1426 (30.5%)	White/Caucasian: 3076 (66.8%) Asian: 490 (10.7%) Black/African American: 277 (6.0%) Hawaiian/Pacific Islander: 222 (4.8%)	American Native: 5 (.1%) Other: 384 (8.3%) Unknown/Declined: 247 (3.2%)
Females	8,539 (64.6%)	86.5 90 (84-91)	Alive: 6253 (73.2%) Deceased: 2286 (26.8%)	White/Caucasian: 5225 (61.9%) Asian: 1024 (12.1%) Black/African American: 768 (9.1%) Hawaiian/Pacific Islander: 387 (4.6%)	American Native: 11 (.1%) Other: 783 (9.3%) Unknown/Declined: 246 (2.9%)
Other or Unknown	11 (0.10%)	90.2 90 (90-91)	Alive: 10 (90.9%) Deceased: 1 (9.1%)	White/Caucasian: 1 (11.1%) Other: 3 (33.3%)	Unknown/Declined: 5 (55.6%)

Supplementary Table 1. Patient Demographics with Encounter Thresholds and Controlling.

Distribution of sex, estimated age, death status, and first race among Alzheimer's and control cohorts. These cohorts are thresholded on more than 10 encounters, and over a year representation in the EMR. Patients are matched at a 1:2 Alzheimer to control ratio with the demographics shown in the table. Estimated age shows mean and median (25%ile - 75%ile).

AD (Alzheimer's disease) (HCC)
 Alzheimer disease (HCC)
 Alzheimer disease type 3 (HCC)
 Alzheimer's dementia (HCC)
 Alzheimer's dementia with behavioral disturbance (HCC)
 Alzheimer's dementia with behavioral disturbance, unspecified timing of dementia onset (HCC)
 Alzheimer's dementia without behavioral disturbance (HCC)
 Alzheimer's dementia without behavioral disturbance, unspecified timing of dementia onset (HCC)
 Alzheimer's dementia, late onset (HCC)
 Alzheimer's dementia, late onset, with behavioral disturbance (HCC)
 Alzheimer's disease (HCC)
 Alzheimer's disease of other onset
 Alzheimer's disease of other onset with behavioral disturbance (HCC)
 Alzheimer's disease of other onset without behavioral disturbance (HCC)
 Alzheimer's disease with delirium (HCC)
 Alzheimer's disease with early onset (CODE) (HCC)
 Alzheimer's disease with early onset (HCC)
 Alzheimer's disease with late onset (CODE) (HCC)
 Alzheimer's disease with late onset (HCC)
 Alzheimer's disease with presenile onset (HCC)
 Alzheimer's disease, early onset (HCC)
 Alzheimer's disease, familial (HCC)
 Alzheimer's disease, focal onset (HCC)
 Alzheimer's disease, unspecified (CODE) (HCC)
 Alzheimer's disease, unspecified (HCC)
 Alzheimer's type dementia (HCC)
 Alzheimer's type dementia with late onset with behavioral disturbance (HCC)
 Alzheimer's type dementia with late onset without behavioral disturbance (HCC)
 Alzheimers disease (HCC)
 DAT (dementia Alzheimer type)
 DAT (dementia of Alzheimer type) (HCC)
 Dementia due to Alzheimer's disease (HCC)
 Dementia in Alzheimer's disease (HCC)
 Dementia in Alzheimer's disease with delusions (HCC)
 Dementia in Alzheimer's disease with depression (HCC)
 Dementia in Alzheimer's disease with early onset (HCC)
 Dementia in Alzheimer's disease with early onset with behavioral disturbance (HCC)
 Dementia in Alzheimer's disease with early onset without behavioral disturbance (HCC)
 Dementia in Alzheimer's disease with early onset, with behavioral disturbance
 Dementia in Alzheimer's disease with early onset, without behavioral disturbance
 Dementia in Alzheimer's disease with late onset
 Dementia of Alzheimer's type with behavioral disturbance (HCC)
 Dementia of Alzheimer's type, with early onset, with depressed mood (HCC)
 Dementia of the Alzheimer's type (HCC)
 Dementia of the Alzheimer's type with early onset with behavioral disturbance (HCC)
 Dementia of the Alzheimer's type with late onset without behavioral disturbance (HCC)
 Dementia of the Alzheimer's type without behavioral disturbance (HCC)
 Dementia of the Alzheimer's type, with late onset, uncomplicated (HCC)
 Dementia of the Alzheimer's type, with late onset, with delirium (HCC)
 Dementia of the Alzheimer's type, with late onset, with delusions (HCC)
 Dementia of the Alzheimer's type, with late onset, with depressed mood (HCC)

Dementia of the Alzheimer's type, with late onset, with depressive mood (HCC)
 Dementia, Alzheimer's, with behavior disturbance (HCC)
 Early onset Alzheimer disease
 Early onset Alzheimer's dementia without behavioral disturbance (HCC)
 Early onset Alzheimer's disease with behavioral disturbance (HCC)
 Family history of Alzheimer's disease
 Focal Alzheimer's disease (HCC) 'Late onset Alzheimer disease (HCC)
 Late onset Alzheimer's disease with behavioral disturbance (HCC)
 Late onset Alzheimer's disease without behavioral disturbance (HCC)
 Major neurocognitive disorder due to Alzheimer's disease (HCC)
 Major neurocognitive disorder due to Alzheimer's disease, possible (HCC)
 Major neurocognitive disorder due to Alzheimer's disease, probable, with behavioral disturbance (HCC)
 Major neurocognitive disorder due to Alzheimer's disease, probable, without behavioral disturbance (HCC)
 Major neurocognitive disorder due to Alzheimer's disease, with behavioral disturbance (HCC)
 Major neurocognitive disorder due to possible Alzheimer's disease (HCC)
 Major neurocognitive disorder, due to Alzheimer's disease, with behavioral disturbance, mild (HCC)
 Major neurocognitive disorder, due to Alzheimer's disease, without behavioral disturbance, mild (HCC)
 Major neurocognitive disorder, due to Alzheimer's disease, without behavioral disturbance, moderate (HCC)
 Major neurocognitive disorder, due to Alzheimer's disease, without behavioral disturbance, severe (HCC)
 Mild major neurocognitive disorder due to Alzheimer's disease with behavioral disturbance (HCC)
 Mild major neurocognitive disorder due to Alzheimer's disease without behavioral disturbance (HCC)
 Mild neurocognitive disorder due to Alzheimer's disease (HCC)
 Mild possible major neurocognitive disorder due to Alzheimer's disease (HCC)
 Mixed Alzheimer's and vascular dementia (HCC)
 Mixed Alzheimer's and vascular dementia with behavior disturbances (HCC)
 Moderate major neurocognitive disorder due to Alzheimer's disease without behavioral disturbance (HCC)
 Moderate probable major neurocognitive disorder due to Alzheimer's disease with behavioral disturbance
 Other Alzheimer's disease (HCC)
 Possible major neurocognitive disorder due to Alzheimer's disease
 Primary degenerative dementia of Alzheimer type (HCC)
 Primary degenerative dementia of the Alzheimer type, senile onset (HCC)
 Primary degenerative dementia of the Alzheimer type, senile onset, uncomplicated (HCC)
 Primary degenerative dementia of the Alzheimer type, senile onset, with depression (HCC)
 Probable major neurocognitive disorder due to Alzheimer's disease with behavioral disturbance
 Probable major neurocognitive disorder due to Alzheimer's disease without behavioral disturbance
 Progressive aphasia in Alzheimer's disease (HCC)
 SDAT (senile dementia of Alzheimer's type) (HCC)
 Senile dementia of Alzheimer's type (HCC)
 Sporadic Alzheimer's disease (HCC)

Supplementary Table 2. UMAP Exclusion Terms

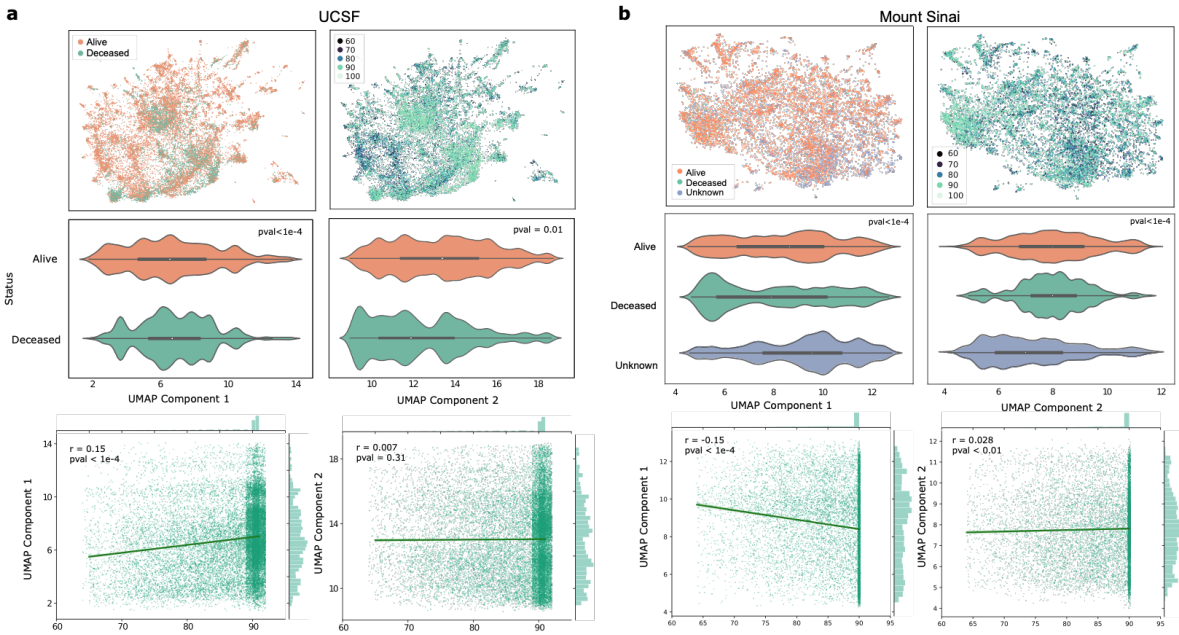
Table of diagnosis excluded in UMAP embedding. These terms contain the word 'Alzheimer'.

UCSF: Graph (>1%)	Number		Avg Number		Network Diameter		Network Radius		Characteristic Path Length	Clustering Coefficient	Network density		Network heterogeneity		Network Centralization		Connected Components	
	Nodes	Edges	Neighbors	Neighbors	Diameter	Radius	Path Length	density			heterogeneity	Centralization	Components	Singletons				
ADDiagnosisNameAll	1056	27504	62.15		4	4	2	2.043	0.830	0.070	1.626	0.763	0.763	171	169			
ADDiagnosisNameFemale	962	25459	61.64		4	4	2	2.038	0.832	0.075	1.586	0.761	0.761	137	136			
ADDiagnosisNameMale	924	20102	52.90		4	2	2	2.057	0.823	0.070	1.633	0.739	0.739	164	162			
ADL3NameAll	483	23505	97.33		2	2	1	1.798	0.899	0.202	1.054	0.801	0.801	1	0			
ADL3NameFemale	452	21958	97.16		2	2	1	1.785	0.899	0.215	1.021	0.788	0.788	1	0			
ADL3NameMale	445	20099	90.33		2	2	1	1.797	0.899	0.203	1.046	0.800	0.800	1	0			
ADL2NameAll	165	7960	96.48		2	2	1	1.412	0.896	0.588	0.479	0.417	0.417	1	0			
ADL2NameFemale	160	7531	94.73		2	2	1	1.400	0.897	0.600	0.469	0.406	0.406	2	1			
ADL2NameMale	158	7257	91.86		2	2	1	1.415	0.892	0.585	0.481	0.420	0.420	1	0			
ConDiagnosisNameAll	421	2445	18.04		4	4	2	2.048	0.738	0.067	1.671	0.843	0.843	151	150			
ConDiagnosisNameFemale	167	2109	25.72		3	3	2	1.892	0.848	0.158	1.156	0.797	0.797	4	3			
ConDiagnosisNameMale	321	1417	13.43		4	4	2	2.078	0.681	0.064	1.717	0.815	0.815	111	110			
ConL3NameAll	318	5772	43.89		2	2	1	1.832	0.760	0.168	1.135	0.839	0.839	56	55			
ConL3NameFemale	190	5434	57.20		2	2	1	1.697	0.830	0.303	0.829	0.705	0.705	1	0			
ConL3NameMale	282	4195	37.46		3	3	2	1.837	0.750	0.168	1.125	0.835	0.835	59	58			
ConL2NameAll	150	3990	55.80		2	2	1	1.607	0.854	0.393	0.697	0.616	0.616	8	7			
ConL2NameFemale	122	3760	61.64		2	2	1	1.491	0.866	0.509	0.552	0.499	0.499	1	0			
ConL2NameMale	137	3200	48.48		2	2	1	1.630	0.862	0.370	0.726	0.640	0.640	6	5			
Mount Sinai: Graph (>1%)	Number Nodes	Number Edges	Avg Number Neighbors	Network Diameter	Network Radius	Characteristic Path Length	Clustering Coefficient	Network density	Network heterogeneity	Network Centralization	Connected Components	Singletons						
ADDiagnosisNameAll	483	1788	15.96		4	4	2	2.030	0.782	0.072	1.696	0.756	0.756	260	259			
ADDiagnosisNameFemale	482	1753	15.72		4	4	2	2.035	0.769	0.071	1.700	0.751	0.751	260	259			
ADDiagnosisNameMale	446	1034	12.16		4	2	2	2.084	0.722	0.072	1.674	0.700	0.700	277	276			
ADL3NameAll	348	10434	59.97		2	2	1	1.827	0.910	0.173	1.152	0.832	0.832	1	0			
ADL3NameFemale	352	10145	59.68		2	2	1	1.824	0.909	0.176	1.142	0.829	0.829	13	12			
ADL3NameMale	332	8162	52.32		2	2	1	1.832	0.905	0.168	1.166	0.837	0.837	21	20			
ADL2NameAll	141	4625	65.60		2	2	1	1.531	0.875	0.469	0.608	0.539	0.539	1	0			
ADL2NameFemale	141	4480	64.93		2	2	1	1.526	0.876	0.474	0.602	0.534	0.534	4	3			
ADL2NameMale	139	4037	59.37		2	2	1	1.560	0.878	0.440	0.642	0.569	0.569	4	3			
ConDiagnosisNameAll	461	13	3.25		3	3	2	1.607	0.558	0.464	0.527	0.524	0.524	454	453			
ConDiagnosisNameFemale	461	13	3.25		3	3	2	1.607	0.558	0.464	0.527	0.524	0.524	454	453			
ConDiagnosisNameMale	423	13	3.25		3	3	2	1.607	0.558	0.464	0.527	0.524	0.524	416	415			
ConL3NameAll	347	1038	17.16		2	2	1	1.857	0.788	0.143	1.267	0.871	0.871	227	226			
ConL3NameFemale	351	1038	17.16		2	2	1	1.857	0.788	0.143	1.267	0.871	0.871	231	230			
ConL3NameMale	331	980	16.47		3	3	2	1.863	0.780	0.140	1.280	0.867	0.867	213	212			
ConL2NameAll	141	1323	32.67		3	3	2	1.594	0.877	0.408	0.677	0.568	0.568	61	60			
ConL2NameFemale	141	1323	32.67		3	3	2	1.594	0.877	0.408	0.677	0.568	0.568	61	60			
ConL2NameMale	139	1290	31.85		3	3	2	1.606	0.865	0.398	0.690	0.566	0.566	59	58			

Supplementary Table 3. All Diagnosis Network Metrics

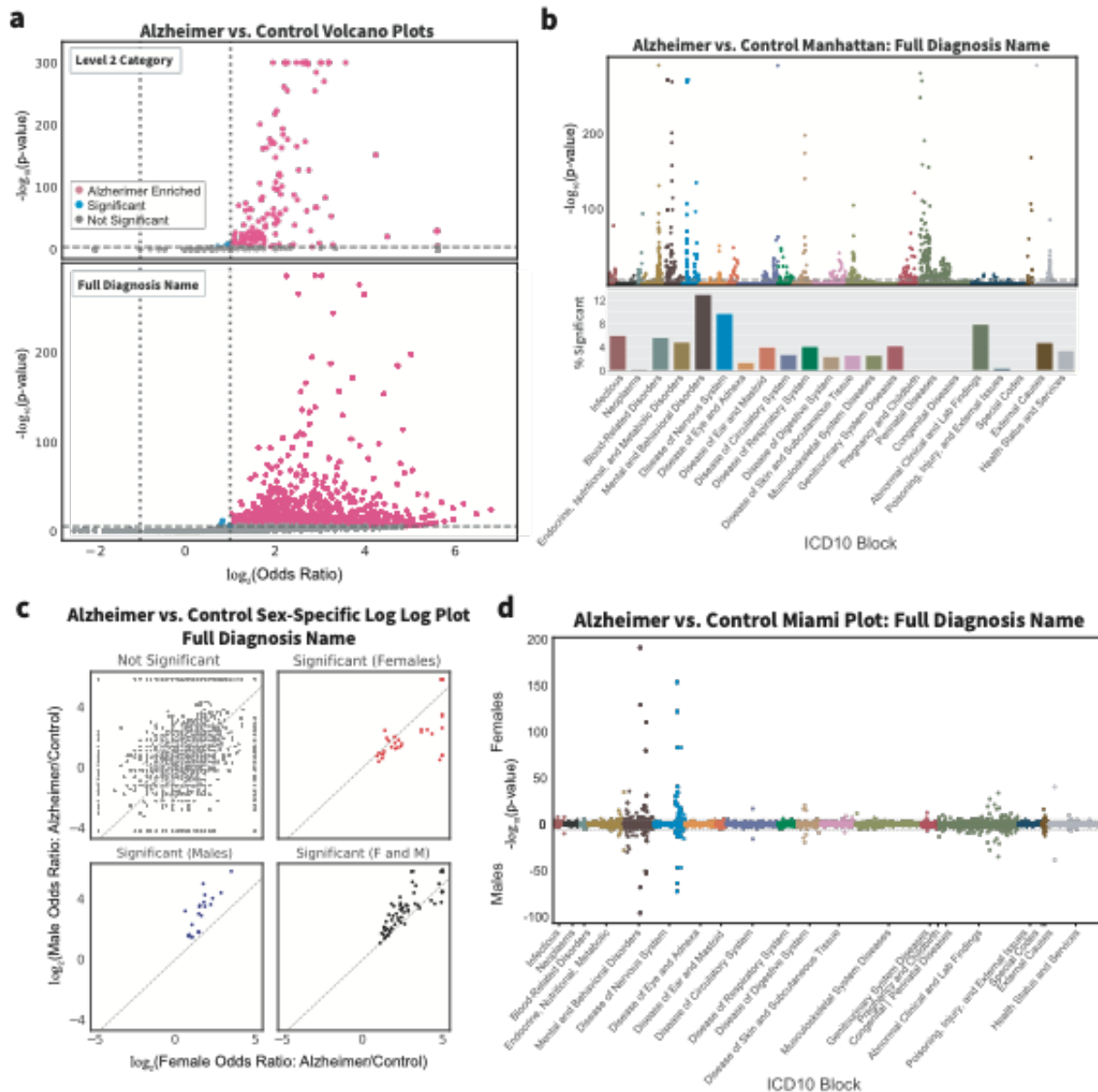
An attached excel sheet with 3 tabs. **Tab 1:** Diagnosis Networks are created with nodes representing a diagnostic category or diagnosis shared among >1% of patients in a group, and edges representing >1% of co-diagnosis in a group. **Tab 2 (UCSF) and 3 (Mount Sinai):** Network metrics are computed for nodes in each network, and the distribution of metrics are compared between networks. Comparisons are performed with and without the removal of singletons (single nodes with no neighbors). A Mann-Whitney U-test is performed to compare the distribution of each network metric, with colors based upon p-value cutoff. The mean difference in metric between comparison groups is also shown.

Supplementary Figures



Supplementary Figure 1. Demographic correlation across UMAP principal components.

- a. The top two graphs show the UMAP of AD and control cohorts at UCSF, colored by deceased status (left) and estimated age (right). The middle graphs show distribution of deceased status among the two UMAP components, which are compared with a two-sided Mann-Whitney U-Test. The bottom graphs show estimated age across the two UMAP components, with marginal distributions shown on the sides. A regression line is plotted, and a Pearson’s R correlation test is performed.
- b. The same UMAP plots are shown as in **a**, but for Mount Sinai.

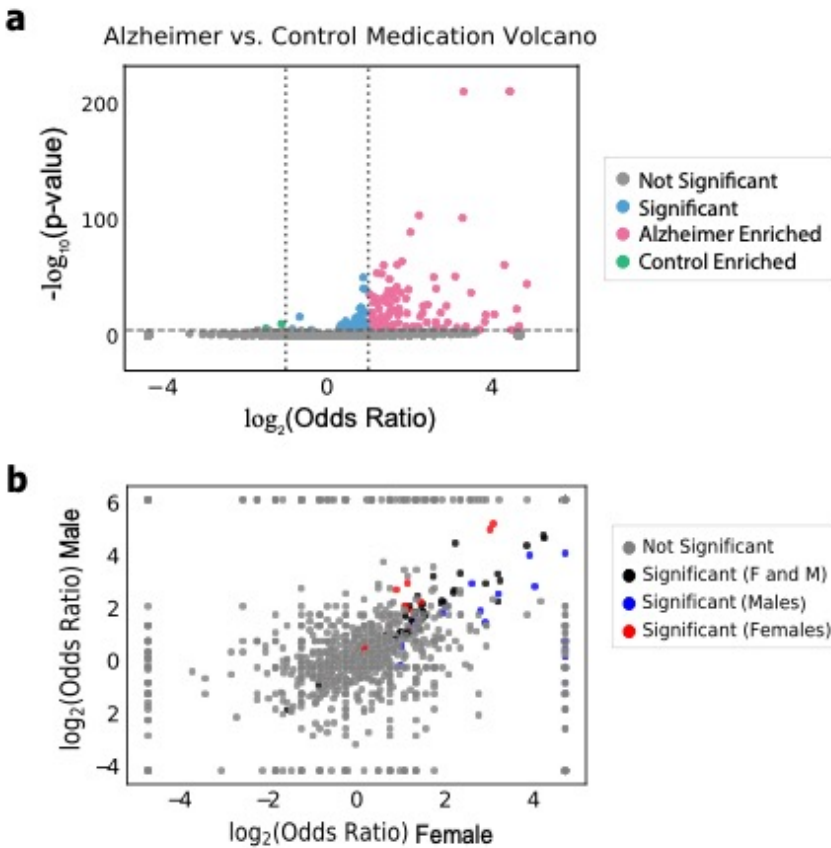


Supplementary Figure 2. Comorbidity Enrichment Analysis identifies enriched diagnosis in Patients with AD vs. Controls and Sex-Specific Enrichments at Mount Sinai.

- Volcano plot for Level 2 categories (top) and full diagnosis names (bottom) compared between AD and control cohorts using two-sided Fisher Exact or Chi-Squared test. P-value cutoff is Bonferroni corrected at 0.05 with log2 odds ratio cutoff at 1 for AD enriched (pink)

and remaining significant diagnoses in blue.

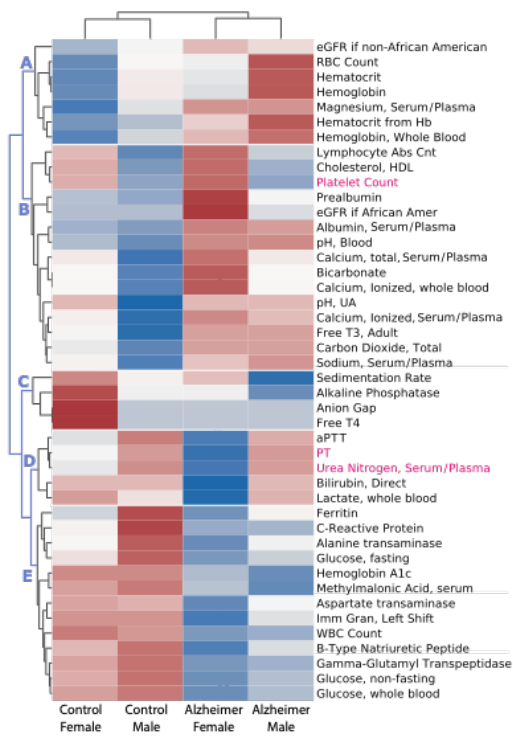
- b. Above, a Manhattan plot with full diagnosis names colored by ICD-10-CM categories with significance determined by two-sided Fisher Exact or Chi Square Test with Bonferroni-corrected p-value cutoff of 0.05. Bottom, percentage of diagnosis in each ICD-10-CM category that is significant.
- c. Full diagnosis names compared between patients with AD and controls within each sex. The log of the odds ratio is plotted on the axis, and points are colored by significance (two-sided Fisher Exact or Chi Square test with Bonferroni-corrected p-value cutoff of 0.05).
- d. Miami plot of the diagnosis names grouped by sex and ICD-10-CM categories.



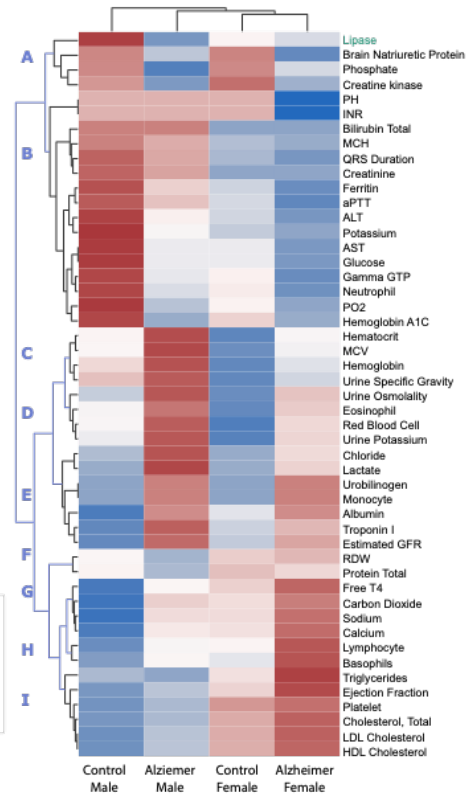
Supplementary Figure 3. Medication Enrichment Analysis identifies Enriched Medications between AD and Control Cohorts.

- a. Volcano plot for generic medication names compared between patients with AD and controls using two-sided Fisher Exact or Chi-Squared Test. P-value cutoff is Bonferroni-corrected at 0.05 with odds ratio cutoff at 2 for AD enriched (pink) or 1/2 for controlled enriched (green). Remaining significant diagnoses are in blue.
- b. Log-log plot of generic medication names compared between patients with AD and controls within each sex. The log of the odds ratio for each sex is plotted on the axis, with points colored by significance (Bonferroni-corrected p-value of 0.05) if female only (red), male only (blue), or both (black).

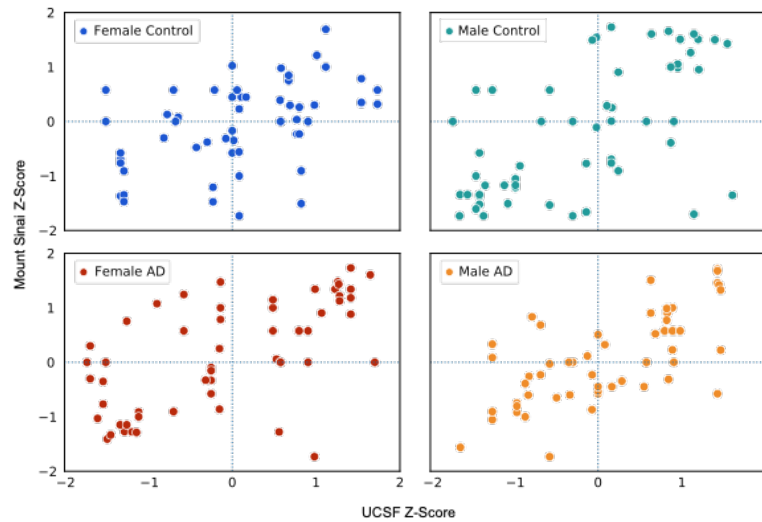
a UCSF Significant Median Lab Values



Mount Sinai Significant Median Labs



b



Supplementary Figure 4. Stratifying by AD status and sex allows identification of lab trends between groups.

a. Heatmap of lab values filtered on significance at UCSF in AD vs control comparison across

sex-specific groups. Labs are clustered with light blue lines representing significant cluster breaks (FWER corrected p-value 0.05). Text color represents significant labs among females only (pink), or significant between patients with AD vs control patients (black). Heatmap colors represent z-score of the average median value across the 4 groups.

- b. Heatmap of lab values filtered on significance at Mount Sinai in AD vs control comparison across sex-specific groups. Labs are clustered with light blue lines representing significant cluster breaks (FWER corrected p-value 0.05). Text color represents significant labs among males only (green), or significant between patients with AD vs control patients (black). Heatmap colors represent z-score of the average median value across the 4 groups.
- c. Comparison of z-scored lab values between UCSF and Mount Sinai showing significant correlations within each AD/sex-stratified groups. Female control: Spearman $\rho = 0.45$, p-value < 0.001 ; Male control: 0.46 , p-value < 0.001 ; Females with AD: 0.59 , p-value $< 1e-5$; Males with AD: 0.64 , p-value $< 1e-5$.