Lineage: Visualizing Multivariate Clinical Data in Genealogy Graphs

Supplementary Material

Carolina Nobre
cnobre@sci.utah.edu
University of Utah

Nils Gehlenborg
nils@hms.harvard.edu
Harvard Medical School

Hilary Coon
hilary.coon@utah.edu
University of Utah

Alexander Lex
alex@sci.utah.edu
University of Utah

1 Scalability Demonstrations

Here we compare Lineage to Progeny, the tool currently used by our collaborators. Progeny is a commercial genealogy drawing tool that closely follows the standard for visualizing genealogies. For details, visit `http://www.progenygenetics.com/`. Figures S1 and S2 show the same dataset visualized with Progeny (Figure S1) and Lineage (Figure S2). Progeny shows three attributes and Lineage shows 15.

2 Additional Information on Case Studies

Supporting evidence for the role of the gene NRXN1

For genes within familial regions with compelling supporting evidence from the literature, one aspect of our analytical pipeline involves an external analysis to identify highly associated genes, then searching all familially shared regions to determine if there is evidence for gene pathways in our results. Using this analysis strategy, two other regions were found with genes closely associated with *NRXN1*. A 320,000 base pair region was shared in family 27251 on chromosome 10q21.3 (p-value for genomic sharing = 1.66E-09). One of two genes in this region is the *leucine rich repeat transmembrane neuronal 3 (LRRTM3)* gene. This gene closely regulates *NRXN1* [1], and has been independently associated with Alzheimer's disease and autism.

In family 68939, a 553,000 base pair region on chromosome 7p21.3 was observed (p-value for sharing = 1.61E-07) implicating the *neurexophilin (NXPH1)* gene, which, like *LRRTM3*, has been shown to interact with *NRXN1*.

Once these supporting regions had been identified, the researcher could immediately turn to Lineage to visualize these families to determine potential overlap in case attributes with the original discovery family. Figure 3 shows these additional families, and the results of demographic and phenotypic exploration in Lineage of the attributes of cases contributing to familial sharing. As Figure S5 shows, multiple cases in both family 27251 and family 68939 again show co-occurring depression. Young age at death was also again apparent. In family 27251, ages ranged from 20 to 30, with one 65-year-old case. In family 68939, ages ranged from 17 to 35. For all 19 cases across the initial NRXN1 family and these two supporting families, the average age at death was 26.68 (standard deviation = 11.28). Even with the small number of cases in these three families, this attribute is significantly different from the mean age in our overall cohort of suicide cases with DNA (t=3.34, p=0.0009). The unexpected co-occurring diagnosis of PD was again present in four of the six cases in family 27251, and in two of the six cases in family 68939.

References

[1] J. W. Um, T.-Y. Choi, H. Kang, Y. S. Cho, G. Choii, P. Uvarov, D. Park, D. Jeong, S. Jeon, D. Lee, H. Kim, S.-H. Lee, Y.-C. Bae, S.-Y. Choi, M. S. Airaksinen, and J. Ko. LRRTM3 Regulates Excitatory Synapse Development through Alternative Splicing and Neurexin Binding. *Cell Reports*, 14(4):808–822, Feb. 2016.

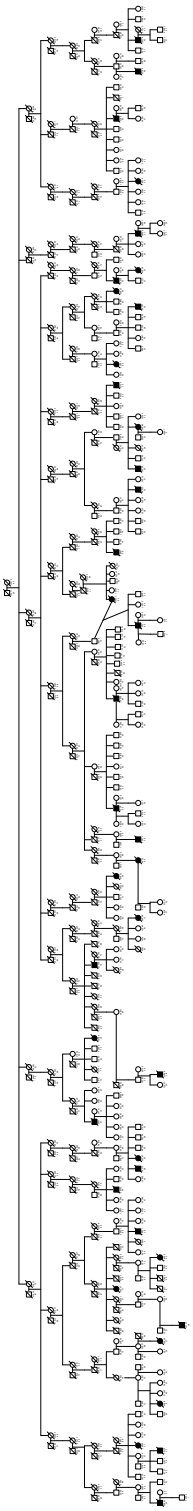

Figure S1: Screenshot of Progeny, the tool previously used by our collaborators. This figure shows a family with 433 members and three attributes.

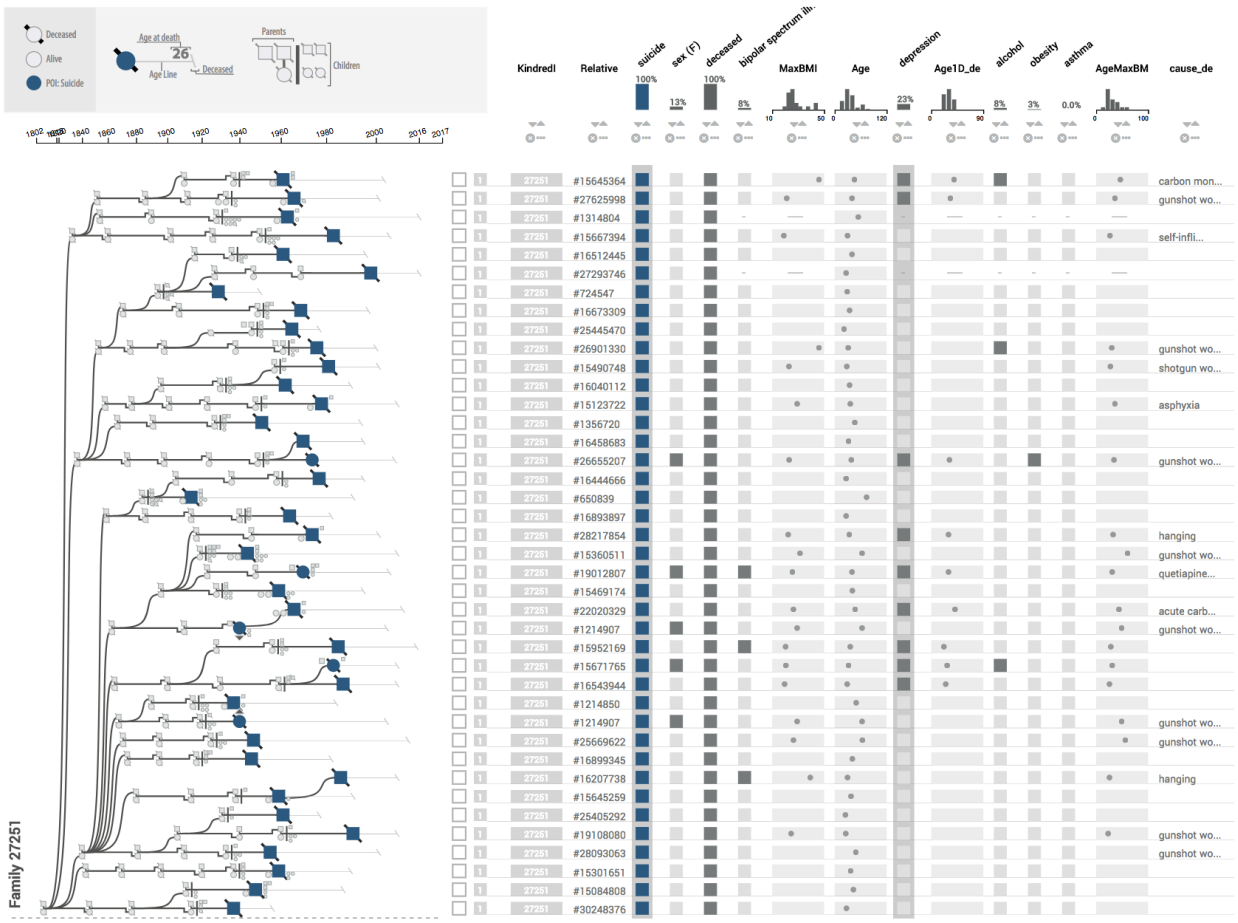

Figure S2: The same family as shown in S1 in Lineage. The 433 family members conveniently fit on a single screen on a 2560x1600 pixel display, and 15 different attributes are shown in the table. In this example, suicide is set as the POI and depression is starred.

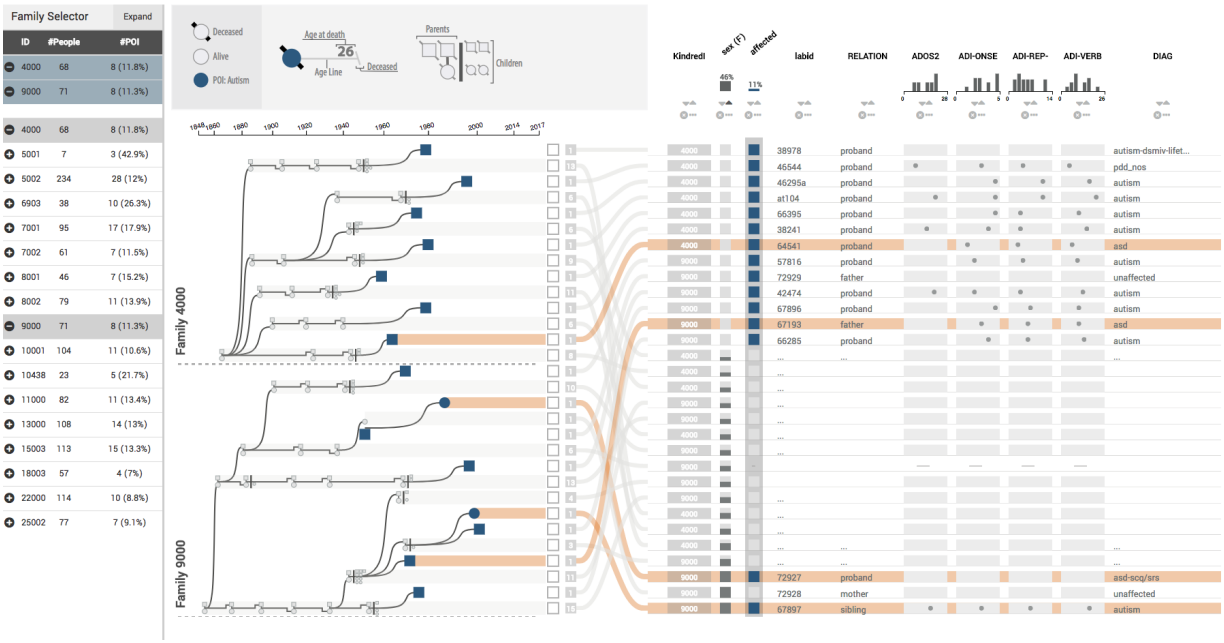

Figure S3: Two families in the Autism Study. The table is sorted by gender to show the prevalence of male cases (there are only two female cases).

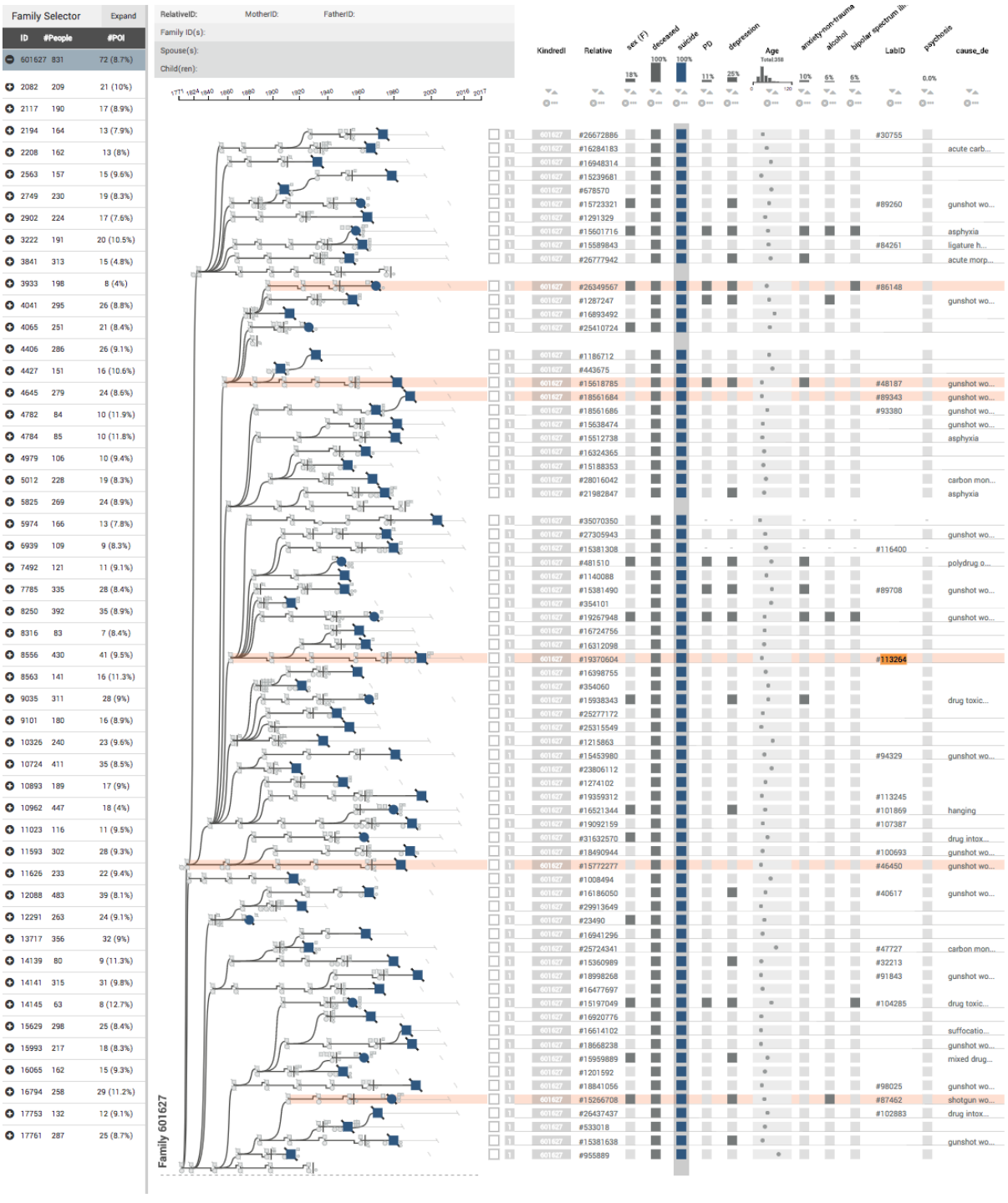

Figure S4: Suicide cases in Family 601627 that share a genetic variant in a region associated with a gene (NRXN1) implicated in psychiatric conditions are highlighted in orange. These cases also share young age at death, and an expected association with depression, but also also an unexpected association with rare personality disorders (column PD).

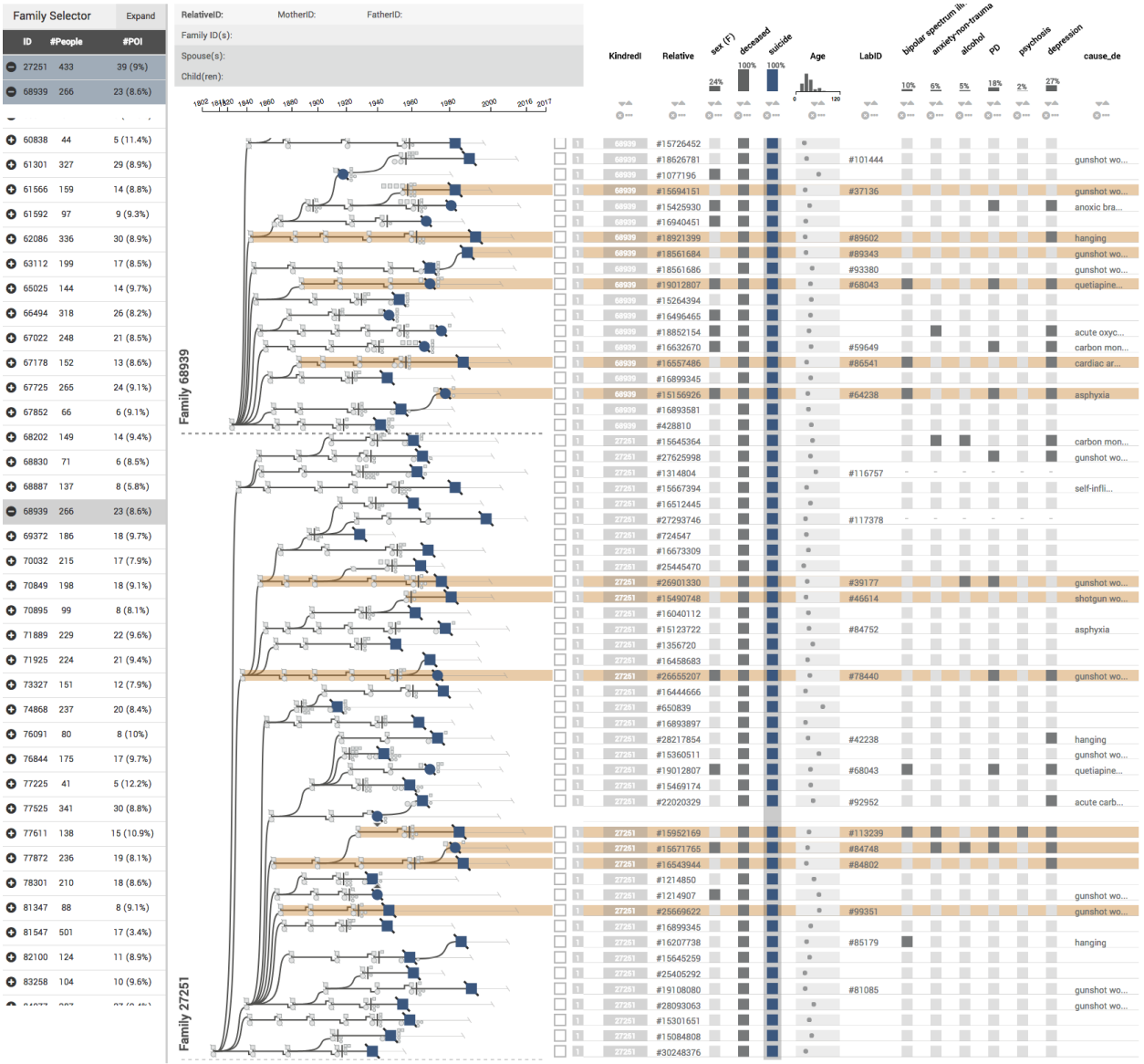

Figure S5: Families 27251 and 68939 identified based on a pathway analysis for a pathway containing the previously identified gene *NRXN1*. The highlighted cases share a genetic region that contains regulatory genes that interact with *NRXN1*. As with family 601627, these cases are disproportionately young, female, and associated with personality disorders (column PD).

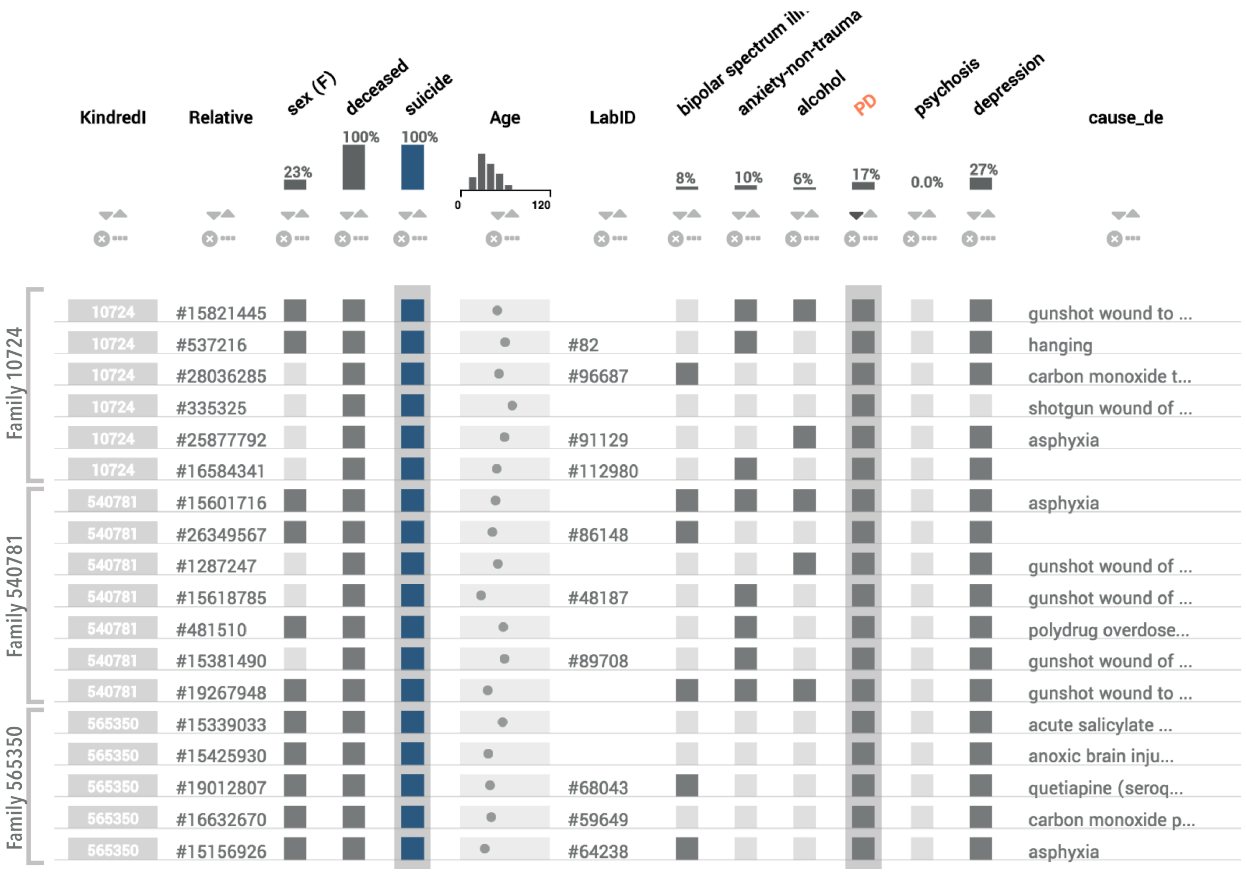

Figure S6: Personality Disorder (PD) Families 10724, 540781 and 565350 identified by browsing families with increased numbers of personality disorders, sorted to show all the PD cases together. These cases are also disproportionately female (10/17).