

Fig S1: Data unavailability is a persistent feature of the MAVE literature. We compiled a list of 159 publications that contained at least one new deep mutational scanning or massively parallel reporter assay dataset and manually inspected the publication’s supplementary data and methods to determine whether counts or scores for the assayed variants were present. Refer to <https://github.com/VariantEffect/MaveReferences> for the full table. This figure was generated from release v0.1.1. Of the 159 total publications, 91 (57%) provide scores or counts.

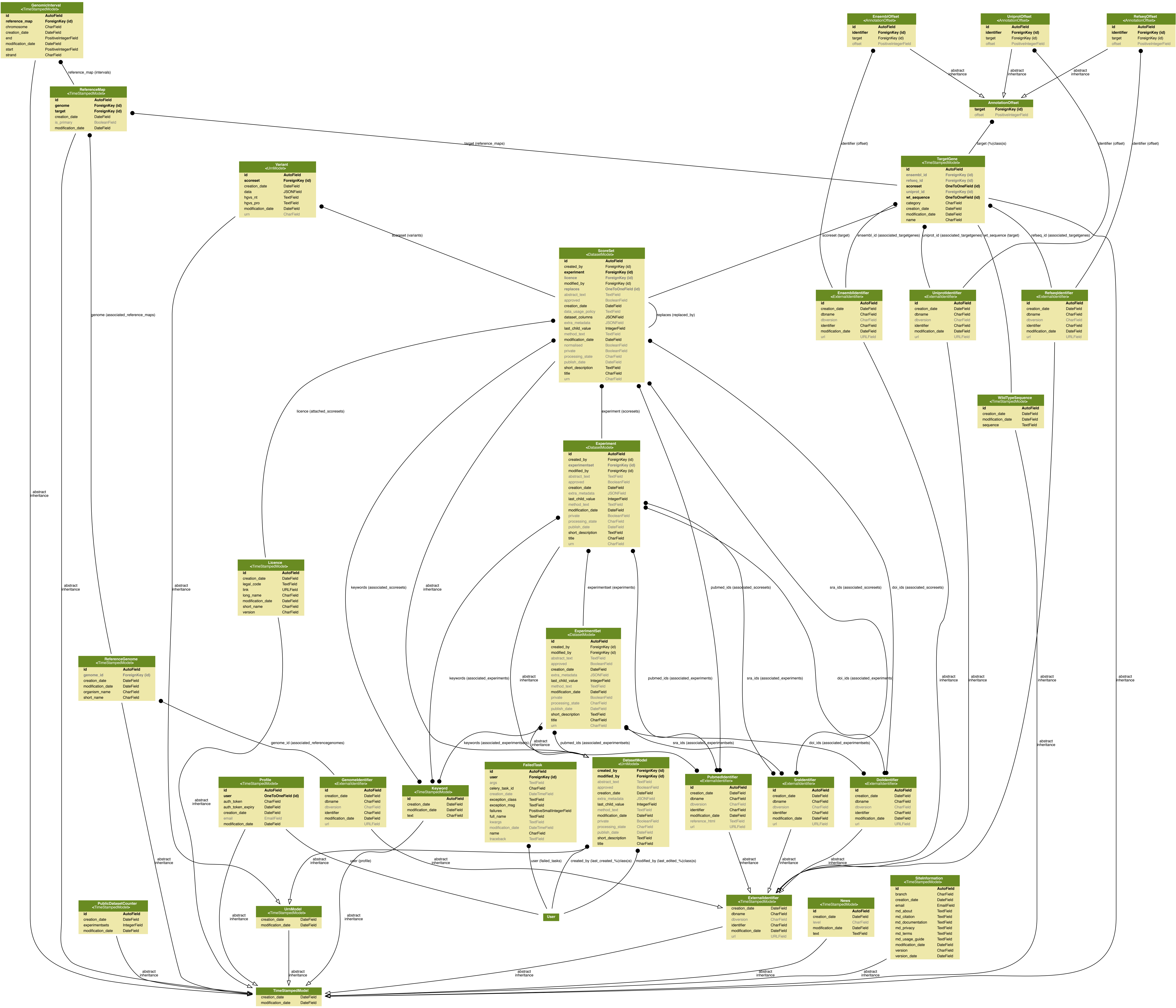


Fig S2: UML (Unified Markup Language) diagram of the complete MaveDB schema in PDF format. The diagram was generated using the Django Extensions package and visualized using Graphviz.