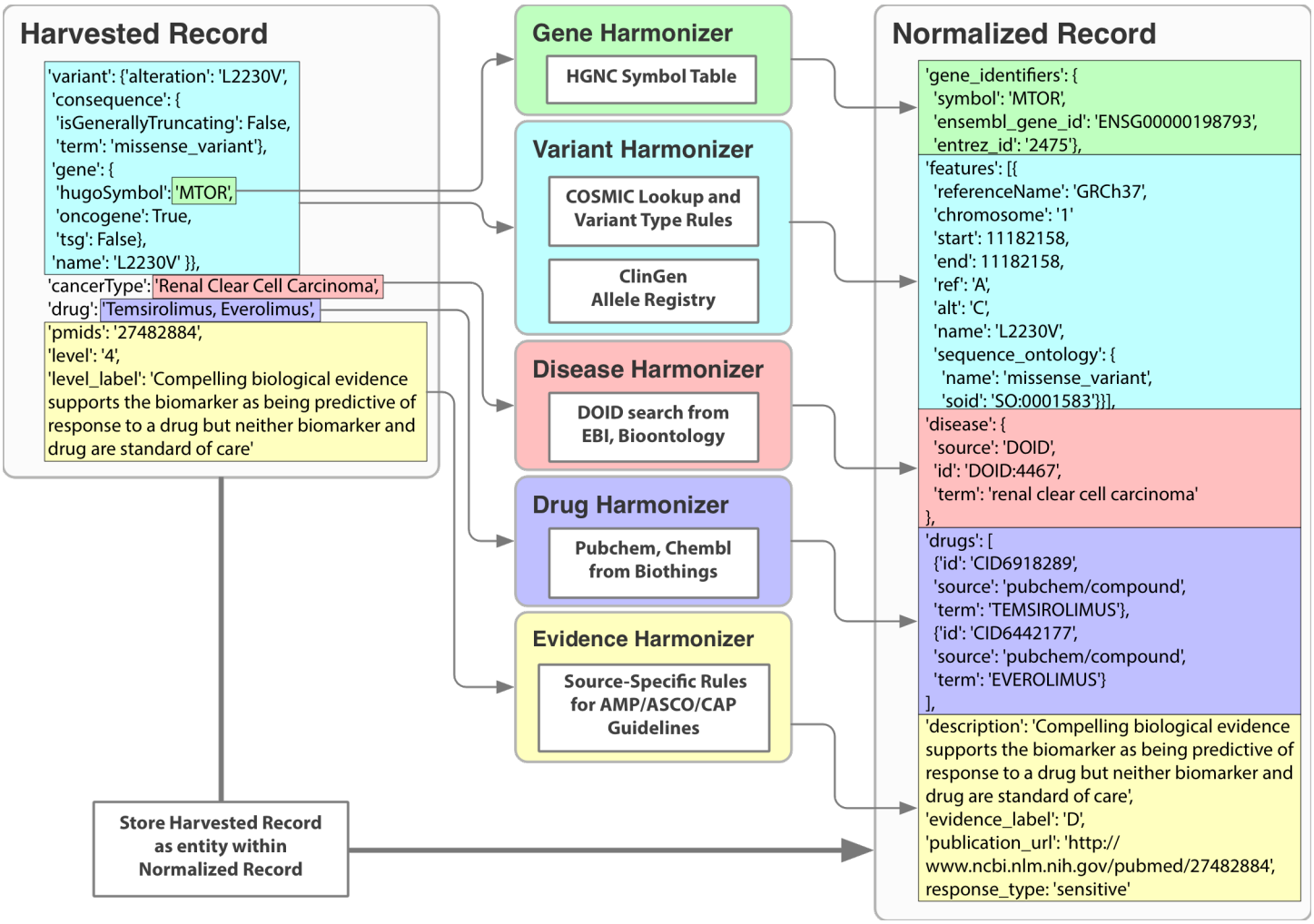

Figure S1 - Harvesting and harmonizing records

Harvested interpretation records (left column) from each knowledgebase vary in structure, a consequence of how they are represented and exported by their parent knowledgebase. Knowledgebase-specific rules are written to select data from harvested records for harmonization across a suite of element-specific harmonizers (center column). Colors represent different elements of an interpretation, which are each harmonized independently: genes (green), variants (cyan), diseases (red), drugs (purple), and evidence (yellow). Outputs from these harmonizers are assembled into normalized records (right column).

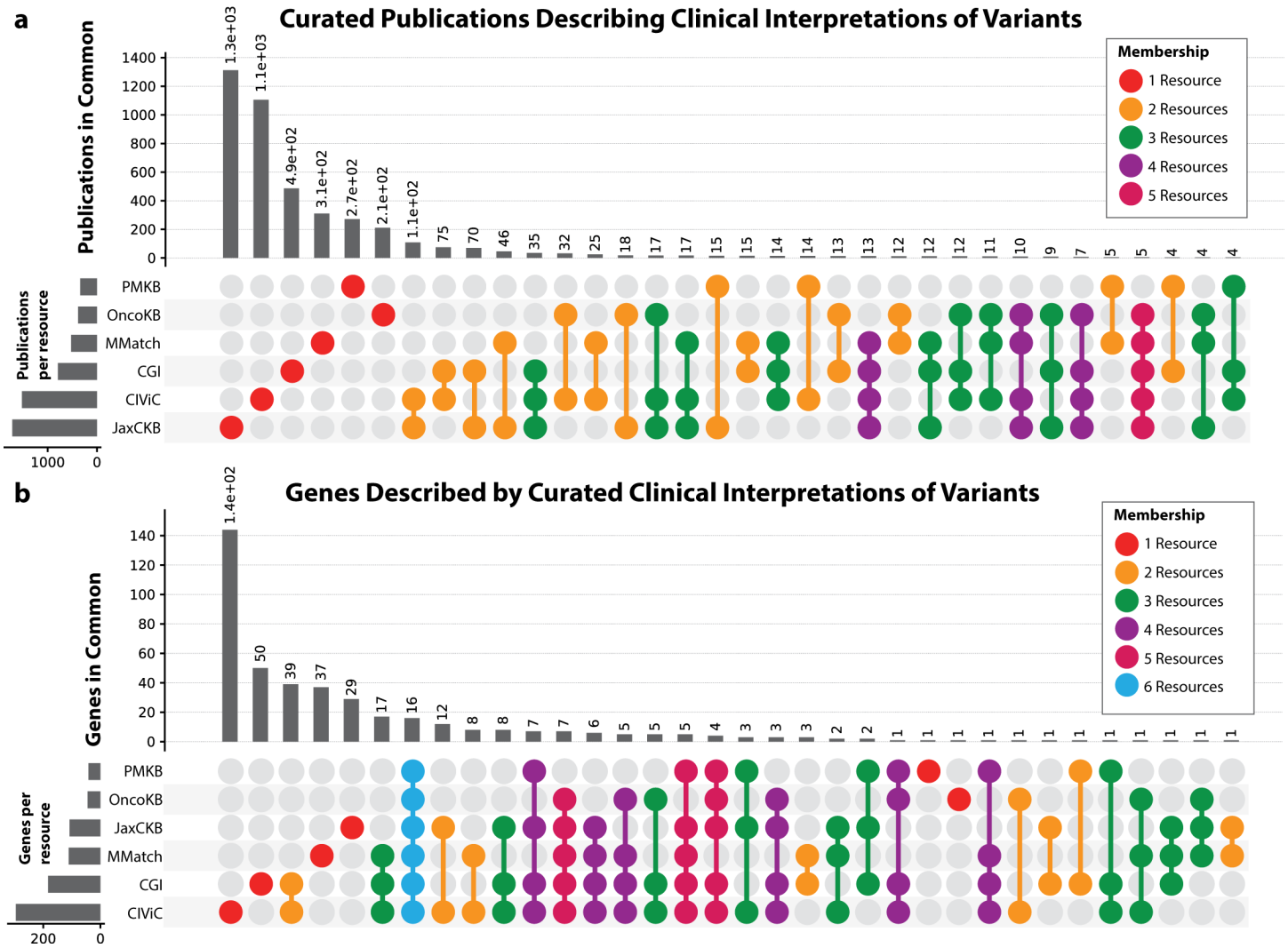

Figure S2 - Knowledgebase overlap

(a) Upset plot of publications supporting clinical interpretations of variants. the overwhelming majority of publications are observed in only 1 of 6 resources. **(b)** Upset plot of genes described by clinical interpretations of variants. Compared to other interpretation elements, genes are much more commonly shared between resources.

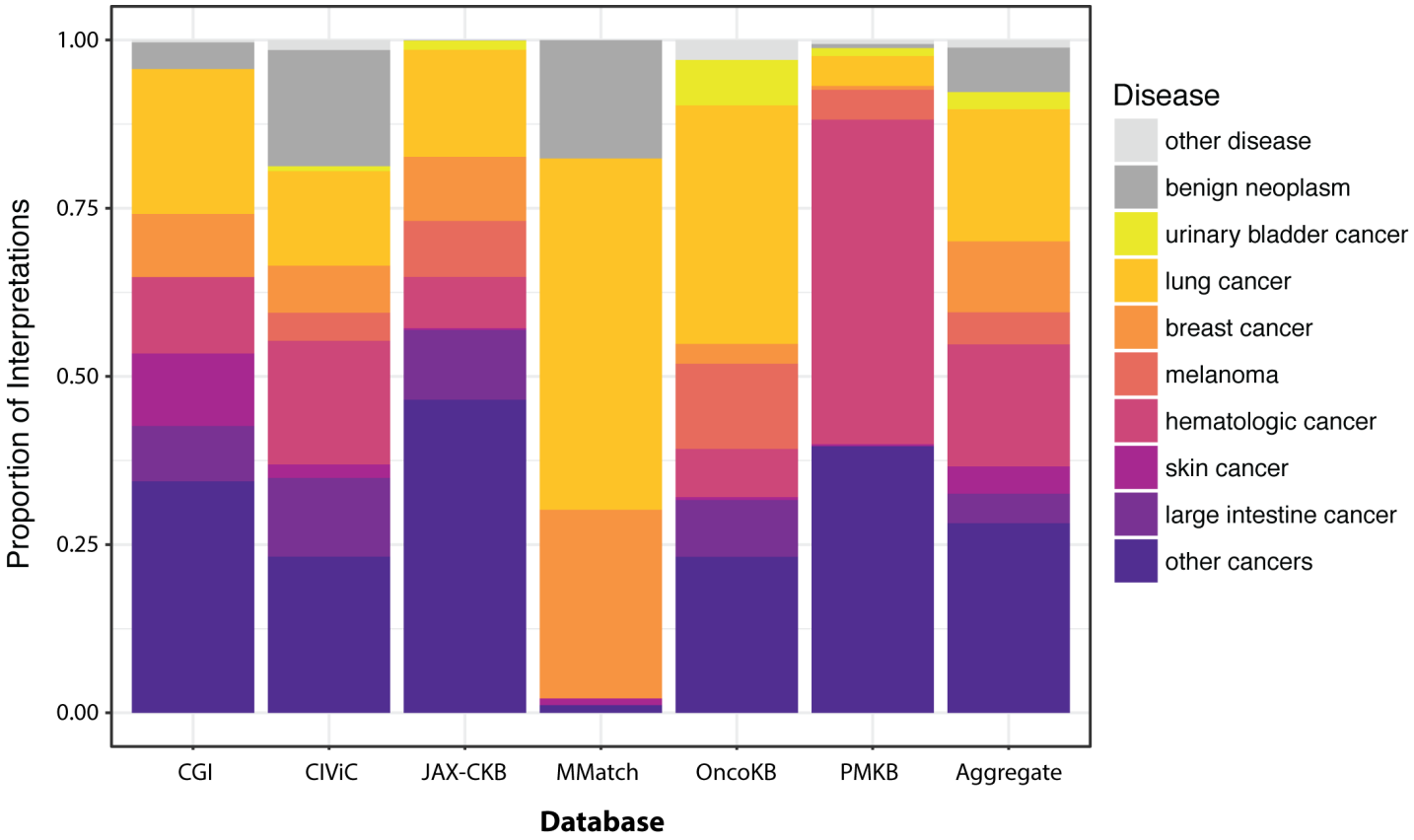

Figure S3 - Knowledgebase disease enrichment

Relative distribution of interpretations describing diseases across the VICC resources. Several resources are strongly enriched for one or more diseases compared to the entire dataset (see related **Table S8**).

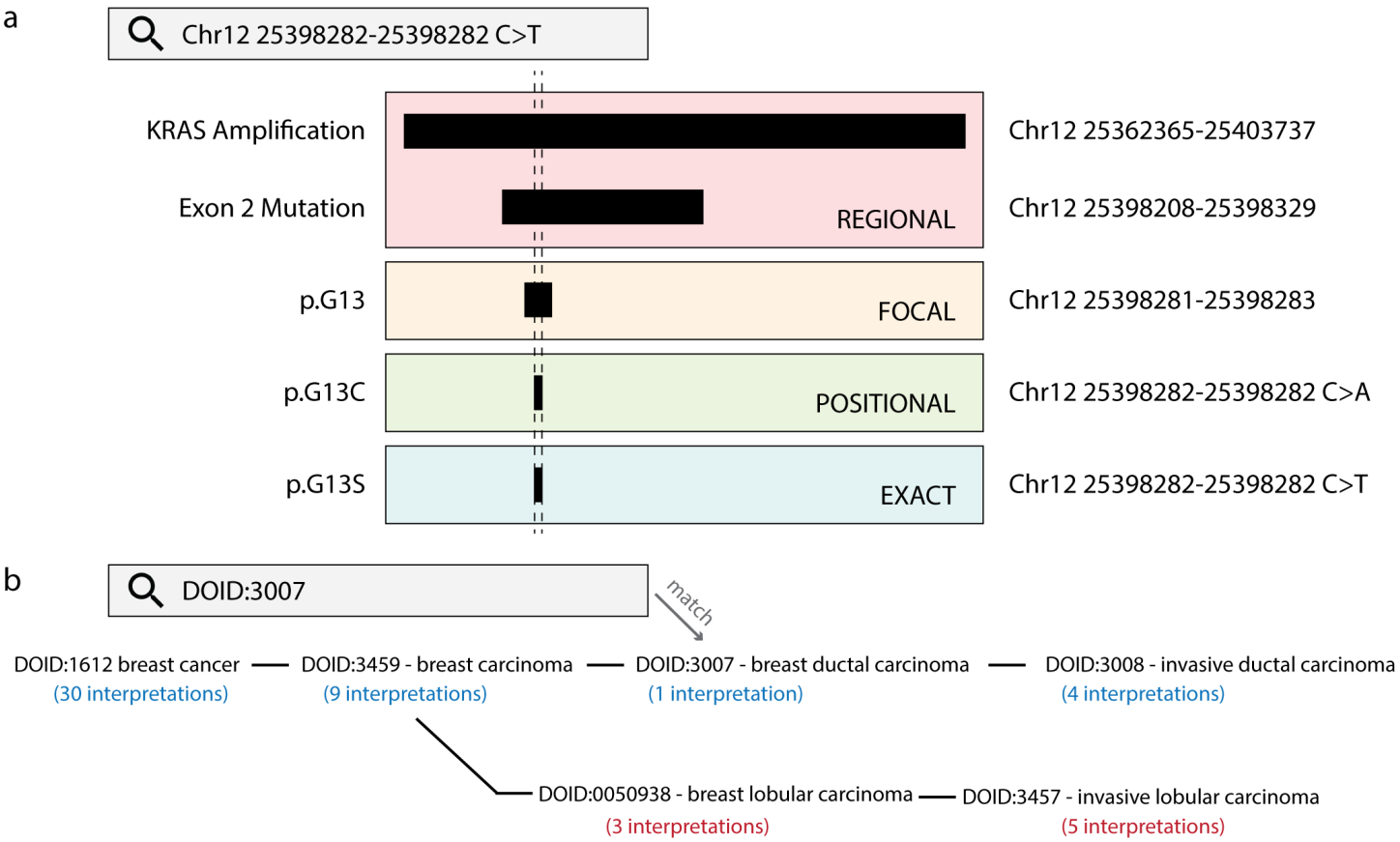

Figure S4 - Search strategies

(a) A variant intersection search strategy. Variants that match at position and allele are referred to as “exact” (blue box), variants matching at position only as “positional” (green box), variants that largely (but not completely) intersect are considered “focal” (orange box), and variants that overlap only a small amount are considered “regional” (red box). The left column shows matched results for a query (red text, top), based on the intersection of coordinates in the right column. **(b)** TopNode disease search strategy. Shown are a subset of disease nodes that all map to the parent TopNode *DOID:1612 - Breast Cancer*. A query for *DOID:3007* would return 44 interpretations (blue) from the queried term, its direct ancestors (*DOID:3459 - Breast Carcinoma* and *DOID:1612 - Breast Cancer*) and descendants (*DOID:3008 - invasive ductal carcinoma*), but no interpretations (red) from indirectly related terms (*DOID:0050938 - breast lobular carcinoma* and *DOID:3457 - invasive lobular carcinoma*).

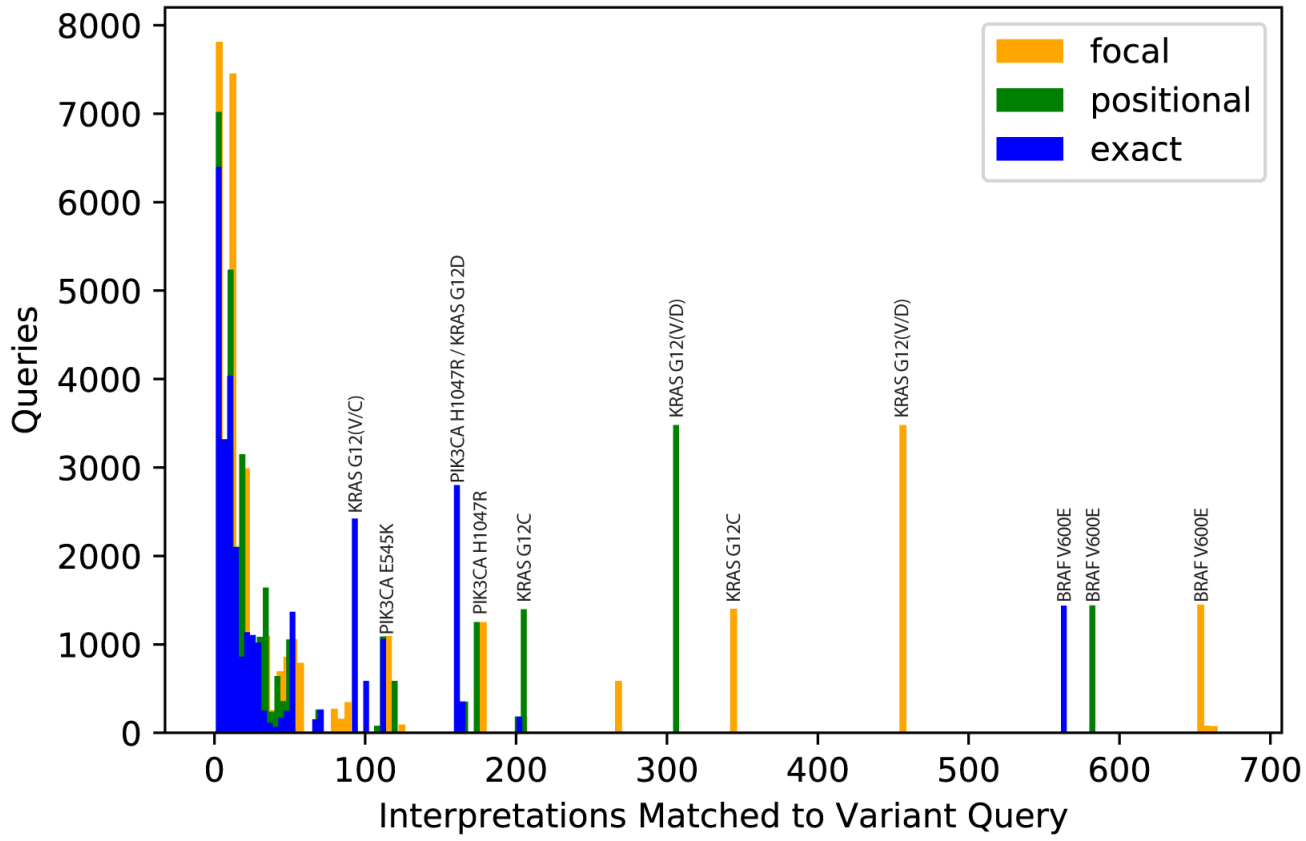

Figure S5 - Commonality of observed mutations and their interpretations

Interpretation count (x-axis) by number of queries (y-axis). Focal (yellow) and positional (green) searches provide a benefit to interpretability over exact matching. Notably, several high interpretation spikes are observed, due to variants that have both a large number of interpretations and are often observed in the GENIE cohort. These include KRAS G12 mutations, BRAF V600E, and several mutations in PIK3CA.

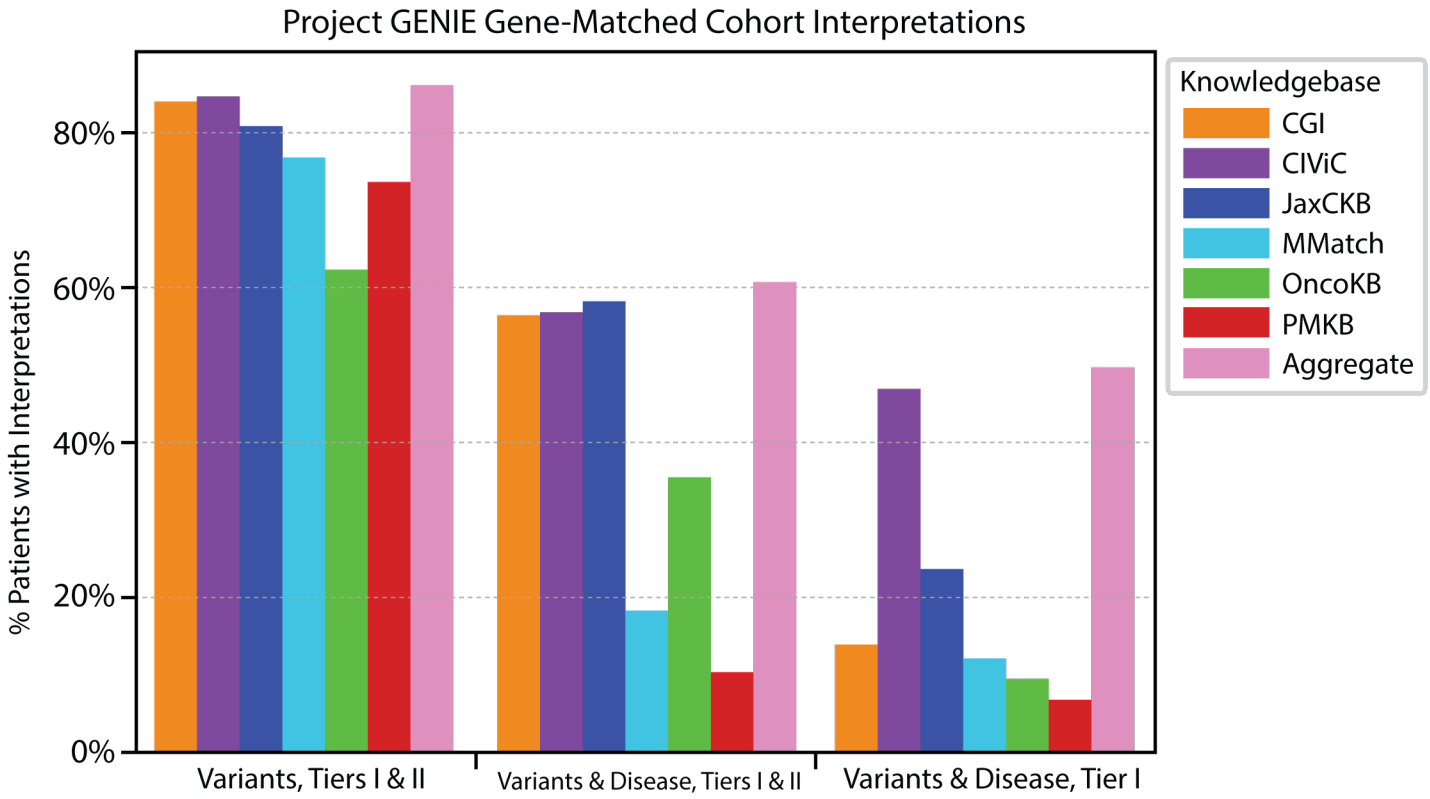

Figure S6 - Gene intersection search

Percentage of Project GENIE cohort with at least one interpretation from the indicated knowledgebase that matches patient variant genes (left group), patient variant genes and disease (center group), or patient variant genes, disease, and a Tier I evidence level (right group). This very broad match strategy is incompatible with the ASCO/AMP/CAP evidence guidelines.