Supplementary Figure 1

Screenshot of the MAGI home page and query interface.

(i) Users select a combination of (public or private) datasets to query. (ii) Users can enter up to 25 genes to query at once. (iii) Alternatively, users can view the mutations in a single sample (Supplementary Fig. 6).
Supplementary Figure 2

Schematic of the software technologies used in MAGI.

MAGI uses D3 client-side to load the mutations and annotations from a web server running Nginx. The web server constructs a JSON object payload using the Node.js framework, which performs the query and collates the resulting data from MongoDB.
Supplementary Figure 3

Mutations in \textit{BRAF} in glioblastoma tumors in the TCGA Pan-Cancer data set.

Shown is the screenshot of the transcript plot of \textit{BRAF} mutations in transcript ENST00000288602. Each diamond indicates a missense mutation in an individual sample – five of six are clustered at position 600.
Supplementary Figure 4

Aberrations view of the mutations in the CDKN2A, CDK4 and RB1 genes in the glioblastoma tumors from the TCGA Pan-Cancer dataset.

Full ticks represent SNVs while black stripes represent inactivating SNVs, downticks represent deletions, and upticks represent amplifications. Exclusive mutations – those that occur in only one gene in the gene set – are colored blue, and co-occurring mutations are colored orange.
Supplementary Figure 5

Aberrations view of genes in the SWI/SNF complex from the TCGA Pan-Cancer data set.

The genes are enriched for mutations in kidney cancer (KIRC, red) and endometrial cancer (UCEC, brown). Most of the inactivating SNVs and deletions in these genes occur in the BLCA, KIRC, and UCEC cancer types, which is easily seen in MAGI by toggling off the other cancer types in the sample type legend on the right (not shown).
Supplementary Figure 6

An example of a potential missed target of recurrent CNAs by GISTIC2 in TCGA STAD.

Shown are amplifications assigned by GISTIC2 to the FGF19 gene in the TCGA STAD dataset. (a) View of all the amplifications in FGF19. (b) Zoomed in view of the amplifications, where CCND1 – a well-studied gene frequently amplified in cancer – is visible. CCND1 is only 44kb from FGF19, and thus may also be a target of the CNAs in FGF19.
Supplementary Figure 7

Copy-number aberrations in PDGFRA in the TCGA Pan-Cancer data set.

Amplifications in PDGFRA were identified by GISTIC2 to be significantly recurrent in GBM tumors (green bars), although PDGFRA is also amplified in many other cancers. Vertical bar indicates the genomic coordinates of PDGFRA gene.
<table>
<thead>
<tr>
<th>Supplementary Figure 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of sample linking.</td>
</tr>
</tbody>
</table>

As users scroll over mutations in the transcript view (i), the corresponding samples are highlighted in the aberration matrix (ii), heatmap (iii), and copy number views (iv).
Supplementary Figure 9

Screenshot of MAGI data-upload interface.

(i) Users can choose to upload a single JSON manifest file that includes the URLs for all the data files in a given dataset. (ii) Users select a cancer type for their dataset either from a predefined list of TCGA/ICGC types or by adding their own – ensuring every mutation annotation maps to a particular cancer type. (iii) Users select the data files they want to load into MAGI. MAGI supports six different data types: SNVs, CNAs, other aberrations, heatmap (continuous values), sample annotations, and sample annotation colors. (iv) Users can also upload SNVs in TCGA MAF and CNAs in GISTIC2 format. (v) All data files can be provided as URLs, or uploaded directly to MAGI.
Supplementary Figure 10

Summary of the glioblastoma (GBM) samples from the TCGA Pan-Cancer data set automatically produced by MAGI.

Users can view these summary pages for public datasets or their own uploaded private data. (i) Summary statistics for the mutation data. (ii) Plot of the number of the mutations in each gene in the dataset. The plot is interactive, as users can zoom in and out or scroll.
over points to get additional information. (iii) Users can also choose from 11 mutation categories for the x- and y-axes (number of SNVs and number of CNAs by default) of the mutations plot. (iv) Sortable and searchable table of the genes in the dataset. (v) Sortable and searchable list of pathways with the most mutations in the dataset. Pathways are from the KEGG and PINdb databases.
Supplementary Figure 11

Screenshot of MAGI annotation interface.
Annotations of (i) genes (ii) expression, (iii) interactions, (iv) mutated residues, and (v) copy number aberrations are viewable as tooltips in each view. MAGI displays annotations interactively as users mouse over different views. Users can also upvote or downvote existing annotations in the tool tips. (vi) MAGI lists the annotations in the database for each gene. (vii) Users can click on individual data points (e.g. interactions in the network view) to pre-load the annotation form on the right.
Supplementary Figure 12

The MAGI gene annotation page shows a table listing all the annotations for a given gene.

The PubMed or PubMed Central ID, the source, and (optionally) the associated cancer, mutation class, mutation type, and protein sequence change are shown for each annotation. Logged-in users can upvote or downvote any of the annotations.
Supplementary Figure 13

MAGI’s tumor sample view.

(i) The tumor sample view lists all the aberrations in the given tumor sample in a table. The aberrations are displayed are represented by the gene, mutation class, and locus or protein sequence change, and are ordered by the score of the annotations for that aberration in the MAGI database. The badges next to highlighted genes/classes/loci list the number of annotations for the gene/class/locus. (ii) Users can click on any of the badges to view the annotations in more detail. (iii) The page also includes a table of the attributes (e.g. gender) for the given sample.
a

**Aberrations**

- **ERBB2 (105)**
- **PTEN (29)**
- **EGFR (15)**

**Gender**

- **Survival (days)**
- **ABSOLUTE Purity**
- **Expression subtype**

Legend (mouse over)

Coverage: 18.65% (144/772)

b

**Enrichment statistics**

Datasets used to compute enrichment statistics

- PTEN
- ERBB2
- EGFR

i

**Expression subtype**

ii

**Contingency table**

Removed 266 samples with missing annotation data.

<table>
<thead>
<tr>
<th></th>
<th>Basal-like</th>
<th>HER2-enriched</th>
<th>Luminal-A</th>
<th>Luminal-B</th>
<th>Normal-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutated</td>
<td>5</td>
<td>38</td>
<td>12</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>Not Mutated</td>
<td>88</td>
<td>19</td>
<td>212</td>
<td>110</td>
<td>8</td>
</tr>
</tbody>
</table>

iii

**Statistical tests**

Fisher’s exact test, Mutated vs Not Mutated on HER2-enriched vs Not HER2-enriched membership

P-value = 2.64445e-24

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Mutated, Not Mutated</td>
</tr>
<tr>
<td>Y</td>
<td>HER2-enriched, Not HER2-enriched</td>
</tr>
<tr>
<td>p</td>
<td>0</td>
</tr>
</tbody>
</table>
Supplementary Figure 14

MAGI interface for computing statistical tests of association between mutation status and sample annotations.

(a) Users view enrichments for their query by following the “Enrichment statistics” link on the view page. (b) The MAGI enrichment statistics page, shown here for the genes PTEN, ERBB2, and EGFR in BRCA samples annotated by the gene expression subtypes from [50]. i) Users choose the sample annotation they want to test from a dropdown of all discrete sample annotation categories. ii) MAGI then cross-classifies samples into a contingency table. iii) Users view enrichments by choosing from a dropdown of statistical tests.
Supplementary Figure 15

Transcript plots for mutations in SMAD2 (top) and SMAD4 (bottom) in the TCGA gastric (STAD; dark green) and Pan-Cancer data sets.

(top) Three of the four mutations in SMAD2 in the TCGA STAD dataset are inactivating nonsense mutations, while the fourth mutation is a missense mutation that occurs in the same location in the MH2 binding domain as two missense mutations in the TCGA colorectal cancer (COADREAD; light green) dataset. (bottom) All of the mutations in SMAD4 in the TCGA gastric dataset occur in the MH2 binding domain. In addition, 12 of the 49 mutations in SMAD4 are mutations at position 361. (Note that only 7 of these mutations are...
visible in the screenshot.)
Supplementary Methods

MAGI: Overview of Features

MAGI (http://magi.brown.edu or installed locally) is a web-based platform for creating custom interactive visualizations of cancer mutation data and enabling the collaborative annotation of this data. The visualizations integrate mutation data from one to thousands of tumors with publicly available annotations of genes and proteins including protein domains and protein-protein interactions. MAGI has several key features.

1. MAGI allows the querying of mutation data from TCGA Pan-Cancer project\(^2\) for any combination of genes (Supplementary Fig. 1), detailed in Section Integration of mutation and annotation data.

2. MAGI generates multiple visualizations of genomics data across thousands of samples including an aberrations view, heatmap view, transcript view, network view, and copy number aberration view. These views illustrate relationships between mutations across multiple biological scales from genome and protein sequence through protein domains and protein-protein interactions. The visualizations are dynamic and interactive, allowing users to explore their data by filtering, sorting, and rescaling the visualizations. Any displayed view can be exported in vector graphics format for subsequent presentation/publication. Each view is described further in Section Visualization components.

3. MAGI includes a straightforward system to upload mutation data from additional samples into a private database. This private data can be queried and analyzed in combination with public data. Users can also upload additional aberration data for TCGA samples (e.g. methylation data), and query this data alongside TCGA and/or private mutation data. Users can share links to views generated with public and/or private data to colleagues. This is further detailed in Section Uploading private mutation datasets.

4. MAGI enables collaborative annotation of aberrations in individual genes or transcripts as well as protein interactions using an interactive, context-aware system that makes it easy to add publication support (PubMed IDs) or free-text annotations. This is further detailed in Section Collaborative annotation.

5. MAGI provides a sample view that shows all sample features and annotations of aberrations in a given (public or private) tumor sample. An annotation score prioritizes the display of aberrations with annotations in the MAGI database. This leverages information across many samples of “N=1” analyses. Moreover, the view reveals unusual samples with few known aberrations, encouraging users to further investigate and annotate these samples. This is further detailed in Section Tumor Sample View.

6. MAGI provides interactive computation of statistical tests of association between mutations and sample annotations across a combination of public and user-uploaded datasets, further detailed in Section Statistical tests of enrichment.

Supplementary Table 1 provides a comparison of MAGI and other cancer genomics browsers and web applications.
MAGI is publicly available and accessible through any modern web browser with Javascript enabled. No local installation of software or browser plugins is necessary. Thus, MAGI places no computational burden on its users beyond the ability to use a web browser. MAGI generates visualizations directly in the browser using the Data Driven Documents (D3)\textsuperscript{7} Javascript framework in scalable vector graphics (SVG) format. The MAGI web server is written in Node.js using a MongoDB database, with web services provided by Nginx (See Supplementary Table 2 for additional details on the implementation). We show a schematic of how MAGI uses these technologies to respond to queries in Supplementary Fig. 2. Using these technologies, MAGI can display mutation data for thousands of samples for dozens of genes in under a second. The source code for MAGI is publicly available on GitHub (http://github.com/raphael-group/magi; See Supplementary Note: Software Availability). We also provide an Amazon machine image containing a pre-configured webserver for easy deployment of private MAGI installations (Details in Supplementary Note: Software Availability).

Integration of mutation and annotation data
MAGI integrates mutation data, including non-synonymous single nucleotide variants (SNVs), small indels, copy number aberrations (CNAs) and predicted fusion genes, and gene expression data with publicly available annotations of the genome, transcriptome, and proteome. MAGI is initialized with two mutation datasets:

1. **TCGA Pan-Cancer project\textsuperscript{2}:** 18,526 mutated genes in 3110 tumor samples\textsuperscript{†} from twelve different cancer types. Copy number aberrations were extracted from GISTIC2\textsuperscript{2} output via FireHose\textsuperscript{‡}, using recurrently aberrant regions from the GISTIC2 maxpeaks. We restricted aberrations to those within 50kb of regions with significantly recurrent aberrations in one or more cancer types, or across cancer types. MAGI also includes expression data from the TCGA Pan-Cancer project downloaded from syn1715753. Samples from the TCGA Pan-Cancer dataset are annotated with clinical data (survival time and gender) downloaded from Firehose (http://gdac.broadinstitute.org/runs/stddata__2014_02_15/data/PANCAN12/20140215), and purity estimates from the ABSOLUTE algorithm\textsuperscript{9} downloaded from syn1710466. TCGA breast cancer (BRCA) samples are annotated with gene expression subtypes from\textsuperscript{10}, downloaded from http://tcga-data.nci.nih.gov/docs/publications/brca_2012/BRCA.547.PAM50.SigClust.Subtypes.txt.

2. **TCGA Gastric Cancer Project\textsuperscript{3}:** 9,316 mutated genes in 215 tumor samples. Mutations and copy number aberrations were downloaded from syn1725886. CNAs were extracted from GISTIC2 output. We excluded 74 tumor samples classified as hypermutators in\textsuperscript{3}. We also downloaded clinical data (survival time and gender) and tumor purity estimates from the ABSOLUTE algorithm\textsuperscript{9} from syn1725886.

\textsuperscript{†}We removed 71 ultramutator samples from the Pan-Cancer dataset as described in syn1729383, and an additional 95 hypermutator samples with >400 SNVs and CNAs.  
\textsuperscript{‡}doi:10.7908/C1736NW2

Nature Methods: doi:10.1038/nmeth.3412
We provide these datasets online on the MAGI website (see below).

MAGI maps copy number aberrations to gene locations on the hg19 reference genome. Mutations in each gene were mapped to a single ENSEMBL transcript using the vcf2maf software package (https://github.com/ckandoth/vcf2maf). MAGI includes 44 cancer type abbreviations from TCGA (https://tcga-data.nci.nih.gov/datareports/codeTablesReport.htm) and ICGC (https://dcc.icgc.org/projects), such that samples of the same cancer type are colored consistently across views. Because MAGI uses standard file types (MAF files and GISTIC output) and cancer types used in TCGA, it is easy to update the TCGA data in MAGI directly from FireHose output.

MAGI is pre-loaded with ~40,000 annotations of known cancer variants. Each of these annotations maps a PubMed or PubMed Central (PMC) ID to a variant (protein sequence change). MAGI includes nearly 4,000 annotations from the Database of Curated Mutations (DoCM; http://docm.genome.wustl.edu/). We generated the remaining annotations by performing text searches using the PMC API for each of the protein sequence changes in the TCGA Pan-Cancer and TCGA STAD datasets.

MAGI also includes annotations – protein-protein interaction networks and protein domain datasets – that describe features, properties, or interactions between genes. MAGI is initialized with interactions from four different protein-protein interaction networks – HINT11, HPRD12, iRefIndex13, and Multinet14 – which comprise 157,026 interactions among 16,448 proteins. MAGI also includes 174,362 protein domains in 64,186 transcripts from the Conserved Domain15, PFAM16, and SMART17,18 protein domain databases. We describe the format of the data used by MAGI in Supplementary Table 3.

Visualization components
Given a query gene set G and a collection of mutation datasets, MAGI generates five types of visualizations, or views. The views are generated using our publicly available GD3 library (https://github.com/raphael-group/gd3), a library of genomic visualization elements written using the D3.js library7. These views are each displayed as part of a single-page web application (Fig. 1). The views are interactive and linked (described further in Section Interactive Linking of Data Views). We describe the views in detail below.

- **Aberrations view.** The aberrations view is a matrix indexed by genes (rows) and samples (column) (Fig. 1). The matrix displays binary aberration data; in each sample, a gene is marked as containing an aberration or not. By default, the aberrations view shows the SNVs and CNAs in the query gene set. For each mutation in each gene, the corresponding cell in the aberrations matrix is marked with a colored shape that indicates the type of mutation (SNV, inactivating SNV, amplification, deletion, or fusion). Cells may be colored by the sample type; the standard TCGA Pan-Cancer color scheme2 is used for these cancer types and
users may define colors for their own samples. Aberration matrices that include only a single mutation dataset color the shape based on whether or not the sample is mutated in one gene (mutually exclusive) or more than gene (co-occurring) in \( G \) (Supplementary Fig. 4). The samples (columns) of the aberrations view are dynamically sortable based on attributes of the genes, the samples, and the mutations themselves. The aberrations view is interactive, and supports zooming and panning such that users can view the pattern of mutations in a gene set across thousands of tumors, or restrict their view to the mutations in individual tumors. Users can also dynamically filter the aberrations by toggling on and off different mutation types.

The aberrations view offers researchers the opportunity to view mutations in any combination of genes to facilitate understanding of how the genes are mutated within individuals and/or across cancer types. We demonstrate this view using the mutations in eight genes of the SWI/SNF complex in the TCGA Pan-Cancer dataset (See Supplementary Fig. 5). The aberrations view generated by MAGI demonstrates visually that this complex is enriched for mutations in kidney cancer (KIRC); in fact, 153/592 (25\%) samples with mutations in this complex occur in KIRC (Supplementary Fig. 5). Furthermore, MAGI demonstrates that most of the inactivating SNVs or deletions occur in KIRC, bladder, or endometrial cancer (207/301 samples), which is easily accomplished by toggling off the other cancer types.

An additional use case for the aberrations view is in examining the pattern of mutations in a gene set across a cohort of tumors (Supplementary Fig. 4). The aberrations view makes it simple to determine visually if a set of mutations are mutually exclusive or co-occurring. Mutually exclusive mutations may indicate that a set of genes is in the same functional pathway\(^{19,20}\). For example, in Supplementary Fig. 4, mutations across 159 glioblastoma samples in three genes (\( CDKN2A \), \( CDK4 \), \( RB1 \)) in the Rb signaling pathway are shown. Only 3/159 samples have mutations in more than one gene, and indeed, these genes have been previously shown to have significant mutually exclusive mutations in glioblastoma\(^{21}\). An alternative explanation for mutual exclusivity is that some genes are targeted more often in different cancer (sub)types. This is easy to observe using MAGI as mutations are color-coded by (sub)type, such as in Supplementary Fig. 5, where \( PBRM1 \) is mutated predominantly in kidney cancers (131/180 mutations).

- **Heatmap view.** MAGI generates a heatmap view for visualizing continuous-valued data in the query gene set across a cohort of samples (Fig. 1). By default, the heatmap shows gene expression data in the samples in which the genes are mutated. The heatmap is not restricted to displaying gene expression data; users can upload any continuous-valued data to MAGI as a matrix (e.g. DNA methylation data).
- **Transcript view.** MAGI generates transcript views for each transcript in each gene in the query set \( G \) (Fig. 1). The transcript plot shows the locations of mutations for a given gene in the protein sequence and constituent protein domains. The protein sequence and domains are shown in the middle of the plot, with in-frame and missense mutations shown above the sequence and inactivating mutations (nonsense, nonstop, splice-site, and frameshift insertions/deletions) shown below. Mutations are represented by different colored symbols, where each symbol represents a different mutation type (e.g. missense or nonsense) and is colored by the dataset in which the mutation occurred. The transcript plot is interactive: users can dynamically switch the protein domain database used to annotate the protein sequence, toggle on and off different mutation types, and zoom and pan along the protein sequence to view mutations at the individual base-pair level, or across hundreds or thousands of base-pairs.

The transcript view is especially useful for addressing a common challenge in cancer genomics: determining the functional impact of mutations. Methods to determine the functional impact of non-synonymous SNVs fall into three categories, and involve determining (1) whether the mutations are inactivating; (2) if there is a cluster of these mutations at a particular position, which suggests that these mutations are activating\(^2\); or (3) if the mutation affected a known protein domain. This latter method is especially useful when examining sets of genes with the transcript view, as it is possible to determine if mutations are targeting an (multiple) interaction domain(s) between pairs of proteins.

To demonstrate the transcript plot, we examined the mutations in \( \text{BRAF} \) in glioblastoma tumors (GBM) from the TCGA Pan-Cancer dataset (Supplementary Fig. 3). \( \text{BRAF} \) is a well-known oncogene and the V600E mutation, a valine to glutamic acid substitution at residue 600, is a driver mutation in multiple cancer types, including \(~60\%\) of melanomas\(^2\) as well as in colon cancer\(^2\). In GBM, we find that \( \text{BRAF} \) mutations are rare with only 7/290 (2.4\%) samples having a \( \text{BRAF} \) mutation. However, 5/7 (71\%) of the mutations in these samples are V600E mutations. This demonstrates the benefits of interactive, exploratory analysis of TCGA data.

- **Network view.** The network view (Fig. 1) shows the interactions among the protein products of the genes in the query gene set \( G \). The subnetwork plot is a multigraph, where each gene (node) is connected by one or more edges, each edge indicating a known protein-protein interaction from one of the four interaction networks\(^11-14\) included in MAGI. Genes in the graph are colored by the number of samples in which they are mutated. The networks used in the plot can be toggled on or off by the user. The positions of nodes are initially determined by the force-directed layout from D3\(^7\) (layout based off of\(^2\)), but nodes can be moved and/or fixed to create custom static layouts. Each edge is annotated with PubMed IDs (PMIDs) of publications that support the underlying interaction.
• **Copy number aberration (CNA) view.** MAGI generates a copy number aberration view (Fig. 1) for each gene in G. The CNA view shows the amplified or deleted segments adjacent to a particular gene in each sample. The middle of the view shows all genes within 500kb of the query gene. The top (respectively bottom) of the view shows the amplified (respectively deleted) segments in each sample with at least one CNA in the region. Each segment is a rectangle that spans the genomic region that is amplified or deleted, and is colored by the dataset of the sample. The CNA view allows for zooming and panning, such that users can determine if the amplified or deleted regions are focused around the particular gene, or span many genes.

Since somatic copy number aberrations often include multiple genes and typically vary greatly in size and position across different samples, it is difficult to determine which, if any, of the genes in a copy number aberration is the target of the aberrations. Methods such as GISTIC2\(^8\) identify recurrently amplified or deleted segments of the genome, and in some cases predict the genes that are the targets of these aberrations. The MAGI CNA view can be useful in identifying likely targets of recurrent aberrations and rare aberrations missed by these computational approaches:

a. **Missed targets of recurrent copy number aberrations.** The MAGI CNA view allows the user to see the entire genomic region in which a set of CNAs occur, including the genomic neighbors of the target identified by a computational approach. This can help users quickly identify likely mistakes in the computational approaches method for assigning targets to recurrent CNAs. For example, we viewed the aberrations the *FGF19* gene from the TCGA STAD dataset (Supplementary Fig. 6a). GISTIC2 identified *FGF19* as the target of amplifications in 30/215 (14%) of STAD samples, potentially a novel discovery, as amplifications in *FGF19* have not been reported in gastric cancer before. However, *FGF19* is located just 44kb from *CCND1* on chromosome 11 (Supplementary Fig. 6b). *CCND1* is a well-studied cancer gene and is frequently amplified in cancer\(^26\), and thus seems likely to be a target of these amplifications. Furthermore, all 12 amplifications that span *FGF19* also span *CCND1*. MAGI makes it easy for users to quickly identify potential missed targets of CNAs by computational approaches.

b. **Rare copy number aberrations.** The MAGI CNA view allows the user to see the size and position of copy number aberrations in each sample. This aids in the identification of CNAs that may be significant, particularly in combination with other types of mutation in the same or different genes, shown in the other view. We viewed the CNAs in the *PDGFRA* gene in the TCGA Pan-Cancer dataset (Supplementary Fig. 7). GISTIC2 identified *PDGFRA* as a significantly recurrently amplified gene only in GBM tumors, but Supplementary Fig. 7 shows that *PDGFRA* contains less frequent amplifications in other cancers as well. For example, in the
lungsquamous cancer (LUSC) samples in the TCGA Pan-Cancer data\textsuperscript{2}, 
*PDGFRA* has focal aberrations in 9/178 tumors (5%). While the 
amplifications of *PDGFRA* in LUSC are not significantly recurrent\textsuperscript{27}, they 
may be worthy of additional investigation since similar amplifications in 
*PDGFRA* are significantly recurrent in GBM.

**Interactive linking of data views**
The views shown in MAGI are interactively linked in two ways. First, users can 
dynamically filter the data shown in all views (e.g. by toggling on or off individual 
datasets). Second, when a user interacts with (i.e. moves the cursor over) a sample in 
one view, all the data for that sample is highlighted in each other view (Supplementary 
Fig. 8). This improves user exploration of the data in a number of ways. For example, 
when a user moves the cursor over a sample (column) in the aberrations view, all of the 
mutations in that sample for the transcript shown in the transcript view are highlighted. 
This shows the user immediately if a sample has multiple mutations in a single sample. 
Additional examples of tasks enabled by linked views include the correspondence 
between gene expression and mutation types (does a gene with a deletion or 
inactivating mutation in a sample also have lower gene expression?), or showing 
whether a sample has two hits in a gene, e.g. a deletion and inactivating mutation.

**Uploading private mutation datasets**
With the declining costs of whole-genome/-exome sequencing, many researchers are 
now sequencing their own tumor cohorts. MAGI provides a platform for these 
researchers to analyze and explore their mutation data in concert with larger datasets of 
public mutation data; e.g. from TCGA. By combining public and private mutation data in 
this manner researchers may interactively explore hypotheses about individual 
mutations or genes, and combinations thereof. For example, researchers can quickly 
determine if the mutations in their dataset are present in the larger public datasets. 
MAGI is not intended to replace sophisticated algorithms to predict driver mutations or 
genes in new datasets, but rather complements these approaches by allowing 
researchers to quickly determine the nature of mutations in a gene or set of genes in a 
particular cancer type.

MAGI offers a tool for users to upload their own data using simple plain text tab-
separated value (TSV) formats. Users may upload files directly from their personal 
machine or provide a public URL to the file (e.g. from Synapse, or a TCGA Publication 
Page). Users can upload SNVs using a TSV format, or the standard mutation annotation 
(MAF) format used by TCGA for SNV data. Similarly, users can upload CNAs using a 
TSV format, or a TAR file (http://www.gnu.org/software/tar/) of the output of the GISTIC2 
algorithm\textsuperscript{8}. Users can also upload TSV files of “generic” aberration data that lists the 
mutated genes in a set of samples to MAGI, which can include for example gene 
expression or methylation data. MAGI also offers users the ability to upload continuous-
valued data (e.g. expression or methylation data) which will be drawn as a heatmap. 
Users can annotate the samples in each of the views by uploading a matrix of sample 

Nature Methods: doi:10.1038/nmeth.3412
annotations—such as clinical data or the output of an algorithm analyzing the tumor—
containing continuous (e.g. survival time) or categorical data (e.g. expression subtype).
MAGI supports GZIP compression (http://www.gzip.org/) and will decompress any
provided GZIP files. The file formats and specifications are available online at
http://magi.cs.brown.edu/upload.

Alternatively, users can upload a single manifest file that includes the URLs for all the
data files in a given dataset. As example, we have posted the data files for the TCGA
Pan-Cancer and gastric cancer datasets that are pre-loaded into MAGI (http://compbio-
research.cs.brown.edu/software/magi/data), and created manifest files for each of these
datasets (http://magi.cs.brown.edu/manifests). This feature also makes it easy to pull the
latest TCGA mutation data from Firehose (http://gdac.broadinstitute.org/).

After the data upload is complete, MAGI annotates the mutation data using the
annotations described above, and then makes the data available to be queried by the
user. Thus, users can upload and query their own mutation data on MAGI in seconds.

MAGI automatically generates a dataset summary page that allows users to sort or
search for mutated genes or functionally-related gene sets (from KEGG28 and PINdb29)
based on their mutations in order to aid users in creating their query gene set(s)
(Supplementary Fig. 10). Both the genes and gene sets are displayed in sortable and
searchable tables that list the number of SNVs, CNAs, and mutated samples per gene or
gene set. The rows of these tables include a link to an automatically generated MAGI
query, which allows users to quickly view the mutation data of genes and gene sets with
possible driver mutations in their data. Users can also dynamically create plots of two
attributes of the mutations in each gene in a dataset (by default the number of SNVs and
number of CNAs). Users can zoom and pan these plots to quickly identify outlier genes
that may have interesting patterns of mutations.

No local software installation or browser plugins are required for a user to upload their
own data. The only requirement is that a user must login to MAGI using an Open ID
provider (http://openid.net/), including Google, Facebook, or Twitter. Mutation data
uploaded with MAGI is available exclusively to the user who uploaded, but can be
queried in combination with public mutation data, which facilitates comparing the
mutations in a gene set from two studies (e.g. of the same cancer type). MAGI allows
users to share the results of individual queries with colleagues and collaborators without
making their mutation data publicly available by hashing each query to make a unique,
bookmarkable link, such that only individuals with the link can view the MAGI page
generated from the query.

Collaborative annotation
Users who log into MAGI can record annotations for mutations and protein interactions.
These annotations are available to all users, enabling collaborative annotation of public
datasets. The main MAGI view includes a panel (Supplementary Fig. 11) that lists the
number of annotations in the MAGI database (in the form of PubMed IDs) for the displayed genes. Each gene links to a page that lists all the variant annotations for each gene as well as links to the corresponding PubMed pages (Supplementary Fig. 12).

The interface for adding annotations is integrated with each of the views, thus reducing the effort required to add data. Clicking on a mutation or interaction in any view populates an annotation form (Supplementary Fig. 11) with information about the corresponding mutation or interaction (Supplementary Fig. 11). For example, clicking on an SNV in the transcript view records the position, protein domain, mutation type (e.g. missense), and cancer type of the mutation. Users may edit this default information and then enter one or more PubMed identifiers (PMIDs) and optionally a text comment. Users may also upvote or downvote existing annotations. By linking the information in the view directly with the annotation form, MAGI automates data entry thereby encouraging users to add annotations.

MAGI also links out to public databases to facilitate user addition of new annotations. For each mutation in the transcript view, we provide a link to a PMC search for the same protein sequence change. This simplifies the process of determining whether a variant has been previously reported or functionally validated. For example, examining the mutations in the SMAD4 gene in the TCGA Pan-Cancer dataset, we see four mutations at position 537, three in colorectal cancer and one in lung squamous cancer. The PMC searches reveal two publications for the lung squamous variant (D537E) and two publications for one of the colorectal variants (D537Y). These publications date to 1998, 2000, 2001, and 2004, and report functional studies of these variants. MAGI also links the gene names in the aberrations view to the corresponding NCBI pages to further facilitate the discovery of annotations which the user can then add to MAGI.

For protein-protein interactions, users can annotate existing interactions or add new interactions to a “Community” interaction database. Users may also upvote or downvote (Supplementary Fig. 11) the recorded annotations – including those that are pre-loaded from PPI databases – so that others will know whether the community believes the annotation to be appropriate and useful (See example below). This feature is intended to facilitate the annotation of protein-protein interactions and correct errors that are present in the curated databases30. One example of a possible error is in the iRefIndex 913 protein interaction network. iRefIndex 9 includes interactions of NOTCH1 with PIK3CA, PIK3R1, PIK3CG. These interactions are curated from31, which reports NOTCH1 interacting with the PI3K proteins as part of a NOTCH1-p56<sup>loxp</sup>-PI3K complex, but only p85 (PIK3R1) is specifically listed in one of the co-immunoprecipitation experiments§. MAGI is designed to help remedy such ambiguities through collaborative annotation, and we downvoted each of these interactions in MAGI.

§ There is additional support for the NOTCH1-PIK3CG interaction from the Online Predicted Human Interaction Database32.
By default, all annotations recorded for public data (e.g. TCGA data or protein-protein interaction data), will be available to other users. Mutation annotations are stored for a given gene and mutation type, with optional fields for protein sequence change and cancer type. The votes for each annotation are stored with the user ID of the voter, in order to ensure that the same voter cannot vote more than once on a given annotation. Users who login will see all contributed annotations in the interactive views by clicking on the corresponding mutation (e.g. deletions in \textit{ARID1A}). All users can view expert-sourced interactions as a “community” PPI network.

The annotation features in MAGI facilitate “expert-sourcing” of public datasets, further increasing the value of these datasets by linking them directly to the scientific literature or expert knowledge. To our knowledge, MAGI is the first web platform to offer expert-sourcing of annotations for cancer genomics.

\textit{Tumor sample view}

A major issue in cancer genomics is to interpret the variants in a single sample. MAGI includes a sample view that shows the variants and sample annotation for a single tumor sample (Supplementary Fig. 13). The view lists all the mutations in a given tumor sample, ordering them by the annotation score of the variants in the MAGI database. Thus mutations with more annotations – presumably the drivers – are shown near the top of the list. As more researchers use MAGI and annotate variants, the mutations that are supported by literature will be further separated from likely passenger mutations.

Furthermore, since MAGI allows users to upload mutation data using a simple web form, researchers or clinicians can upload data from an individual tumor and have the mutations prioritized by MAGI using literature annotations.

MAGI shows the tumor sample view in a separate page from the gene set visualizations. For example, Supplementary Fig. 13 displays a bladder urothelial carcinoma (BLCA) tumor sample’s mutations.

\textit{Statistical tests of enrichment}

Users can load sample annotations (e.g. sample clusters or subtypes) into MAGI – either for public or private datasets – and test whether a group of annotated samples is enriched for mutations (Supplementary Fig. 14). MAGI computes the enrichment by cross-classifying samples into an \( r \times c \) contingency table, and then using Fisher’s exact test (when \( r = c = 2 \)) and the \( \chi^2 \) association test. MAGI displays the enrichments for a given query set of genes and datasets on a separate page.

We demonstrate the statistical tests using breast cancer gene expression subtypes from TCGA\(^{10}\), computing the association between \textit{ERBB2} aberrations and the HER2-enriched subtype (Supplementary Fig. 14). \textit{ERBB2} amplifications largely define the HER2-enriched subtype\(^{10}\), and MAGI indeed finds that the HER2-enriched subtype is enriched for aberrations (\( P < 10^{-15} \)). Identifying associations between aberrations and sample
annotations is a very common task in cancer genomics research. MAGI allows the user to effortlessly perform these calculations, integrating visual exploration of the data with statistical analyses. To the best of our knowledge, MAGI is the only online tool that computes statistical tests of enrichment between mutations and sample annotations interactively across a combination of public and user-uploaded datasets.
Supplementary Notes

Related work

MAGI fulfills different needs from the current data portals, genome browsers and web servers that have been developed to view and analyze genomics data. Portals such as CBio\textsuperscript{33,34} allow users to query and explore cancer genomics data, but do not allow straightforward uploads of private data. Several systems\textsuperscript{33,34} allow direct queries or bulk downloads of TCGA and/or ICGC data. With these systems, joint analysis of public and private datasets typically requires complicated software installation and/or bioinformatics expertise. Recently, the UCSC Genome Browser has added data track hubs\textsuperscript{35} for users to upload browser tracks, but the genome-centric analysis does not facilitate analysis of mutation combinations in pathways/networks. In contrast, genome browsers such as the Integrative Genomics Viewer\textsuperscript{36}, UCSC Cancer Genomics Browser\textsuperscript{37}, or Galaxy\textsuperscript{38} allow researchers to view the genomic locations of aberrations from multiple tumor samples, but are not integrated with public datasets. Different goals are pursued by web servers such as IntOGen-mutations\textsuperscript{39,40} and others\textsuperscript{41,42} that facilitate analysis of the mutations in a cohort of tumors through a simple web interface.

Current cancer data portals and browsers\textsuperscript{33,34,42,43} (See\textsuperscript{1} for a review) have limited features to facilitate a bidirectional interaction between large-scale public and smaller scale/private datasets. In all current systems, information flows in only one direction. There is no ability for researchers to record insights about the public datasets, and the results of follow-up studies remain scattered in the literature and disconnected from the datasets. These computational/software limitations impede the scientific process and lessen the impact of these large cancer datasets.

Case study

To demonstrate the power of using MAGI to analyze private mutation data in conjunction with TCGA data, we uploaded mutation data from the TCGA stomach adenocarcinoma (STAD) study, a cancer type that was not one of the 12 cancer types in the Pan-Cancer dataset\textsuperscript{2}. The TCGA STAD dataset\textsuperscript{3} consists of SNVS and CNAs in 9,326 genes in 215 samples (See Supplementary Note: Integration of mutation and annotation data). We queried the TCGA STAD dataset using a list of 22 known gastric cancer genes compiled from an earlier publication\textsuperscript{44}. This list includes both canonical cancer genes (\textit{TP53, ERBB2, KRAS, MYC, RB1, and PTEN}), as well as genes with roles in cancer more specific to STAD (\textit{ERBB3, EZH2, RUNX3, SMAD1, SMAD2, SMAD3, SMAD4, SMAD7, CDKN1A, CDKN1B, RELA, and TGFBR2}). We found that two of the canonical cancer genes are mutated in >25 samples – \textit{TP53} (106 samples), \textit{MYC} (57), \textit{ERBB2} (56), and \textit{KRAS} (34) – while \textit{PTEN} (10) and \textit{RB1} (4) were less frequently mutated. \textit{PTEN} and \textit{RB1} each have multiple inactivating mutations consistent with their role as tumor suppressors.

\*\* http://www.broadinstitute.org/oncotator/
Of the remaining gastric cancer genes, CDH1 (21), SMAD4 (48) ERBB3 (10), TGFBR2 (6), and SMAD2 (4) are the most mutated. SMAD4, TGFBR2, and SMAD2 each have patterns of mutations that indicate a strong functional impact (Supplementary Fig. 15).

- In SMAD4, all 19 mutations (including three frameshift insertions) are in the MH2 binding domain, including three missense mutations and two in-frame insertions at position 361 where mutations associated with polyposis have been previously reported. We added this publication as an annotation of the SMAD4 mutation, thus making this information available to all users of MAGI. SMAD4 was also deleted in 3 samples.
- In SMAD2, three of the four mutations are inactivating nonsense mutations, and the fourth mutation is a missense mutation in the MH2 binding domain.
- In TGFBR2, 4 of the 5 missense mutations occur in the protein kinase domain, while the other mutation is an inactivating frameshift deletion.

The observed pattern of inactivating mutations in SMAD2, SMAD4, and TGFBR2 is particularly striking as TGFBR2 is a member of the transforming growth factor beta (TGF-β) signal transduction pathway. This pathway activates the SMAD family of proteins and is known to have a key role in many cancers, including gastric cancer.

Combining the STAD mutation data with TCGA Pan-Cancer data in MAGI suggests a strong link between SMAD4 and SMAD2 mutations in gastric cancer and colorectal cancer (COADREAD).

- 12 of the 49 mutations in SMAD4 in the STAD and COADREAD datasets occur at position 361 in the MH2 binding domain. Three missense and two in-frame insertions occur at position 361 in STAD, and six missense mutations occur at position 361 in COADREAD. Furthermore, the MAGI database includes 37 annotations for variants at position 361 in SMAD4, including 21 annotations from DoCM specifically for COADREAD. This provides further support for a mutational hotspot at position 361.
- There are two R321Q missense mutations in SMAD2 in COADREAD, which match the position and residue change of the single SMAD2 missense mutation in STAD. While the single R321Q mutation in STAD would not stand out among the many other mutations measured in the STAD samples, the conservation of this mutation in COADREAD lends support for this mutation having a role in cancer. Moreover, there are five nonsense mutations at position 464 in SMAD2, including two mutations in STAD and three in COADREAD.

These results show that not only are the SMADs mutated in both STAD and COADREAD, but particular mutations in SMAD2 and SMAD4 are common to these two cancer types. In addition, mutations in TGFBR2 and more generally in the TGF-β pathway are well-known in colon cancer, and mutations in the TGF-β pathway were discussed in the TCGA COADREAD publication. However, none of these particular

†† TGFBR2 has been previously reported to be mutated in hypermutator tumors. However, as the TCGA STAD and Pan-Cancer datasets were both filtered for hypermutators, these numbers do not include any mutations from hypermutators.
mutations were discussed in either the COADREAD⁴ or STAD publications³, nor were the similarities between the mutations in these cancer types reported. The interactive visualizations generated by MAGI led immediately to these discoveries.

Software availability

- **MAGI web server** ([https://github.com/raphael-group/magi](https://github.com/raphael-group/magi)). Git repository including the Node.js web server used to run MAGI.
- **GD3 visualization library** ([https://github.com/raphael-group/gd3](https://github.com/raphael-group/gd3)). Git repository including the Javascript library (GD3) that uses D3 to generate the MAGI visualizations.
- **Amazon machine image** (AMI ID: ami-44e4442c). Preconfigured Amazon machine image named “MAGI Web Server” that includes the MAGI web server and all of its software dependencies. Users can clone the AMI to create their own private version of MAGI.

Future work

We demonstrate MAGI’s functionality using cancer mutation data. However, the software is not limited to the display and annotation of cancer data. The aberrations view in MAGI could equally be used to display germline variants or other transcriptomic, genomic, or proteomic data across samples. Similarly, the network view can display other types of interactions between genes, proteins, protein domains, or individual residues. Any of these can then be annotated interactively, enabling collaborative annotation of a wide variety of datasets. Individual researchers, teams of researchers, or international collaborations may create custom MAGI installations to leverage other public and private datasets and engage different communities of researchers in annotation of these datasets.
Supplementary References


## Supplementary Tables

<table>
<thead>
<tr>
<th>Feature</th>
<th>MAGI</th>
<th>Cancer Regulome</th>
<th>cBioPortal</th>
<th>CGWB</th>
<th>UCSC Cancer Genomics Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allows private data (upload tool)</td>
<td>Yes  (web upload)</td>
<td>No</td>
<td>Yes</td>
<td>Yes (Java Web Tool)</td>
<td>Yes (web upload)</td>
</tr>
<tr>
<td>Combine private and public data</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Collaborative annotation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Interactive (zoom, pan, and filter data)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Aberrations matrix</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Heatmap</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Transcript plot</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PPI networks</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CNA browser</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample annotations (e.g. clinical data)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Statistical tests of association</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (survival)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Interactive linking of visualizations</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Supplementary Table 1: Comparison of features offered by MAGI and similar web tools.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Technology</th>
<th>Framework</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backend</td>
<td>Web server</td>
<td>Node.js</td>
<td>Express</td>
<td>Javascript</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Database</td>
<td>MongoDB</td>
<td>Mongoose</td>
<td>Javascript</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>User authentication</td>
<td>OAuth2</td>
<td>Passport.js</td>
<td>Javascript</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HTML templating</td>
<td>Jade</td>
<td>Express</td>
<td>HTML, CSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontend</td>
<td>Visualizations</td>
<td>D3</td>
<td>N/A</td>
<td>Javascript</td>
</tr>
</tbody>
</table>

Nature Methods: doi:10.1038/nmeth.3412
**Supplementary Table 2**: Technologies used by MAGI.

<table>
<thead>
<tr>
<th>Visualization Component(s)</th>
<th>Data</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aberrations view and transcript view</td>
<td>SNVs and small indels</td>
<td>Gene symbol</td>
<td>ARID1A</td>
</tr>
<tr>
<td></td>
<td>Sample ID</td>
<td>TCGA-A6-2676</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcript ID</td>
<td>ENST00000324856</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcript length</td>
<td>2285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Position</td>
<td>1262</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mutation type</td>
<td>Missense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Original Amino Acid</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>New Amino Acid</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Aberrations view and CNA view</td>
<td>CNAs</td>
<td>Gene symbol</td>
<td>ARID1A</td>
</tr>
<tr>
<td></td>
<td>Sample ID</td>
<td>TCGA-04-1367</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type (amplification or deletion)</td>
<td>DEL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left point</td>
<td>19153764</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right point</td>
<td>33238714</td>
<td></td>
</tr>
<tr>
<td>Transcript view</td>
<td>Protein domain annotations*</td>
<td>Dataset</td>
<td>PFAM</td>
</tr>
<tr>
<td></td>
<td>Transcript ID</td>
<td>ENST00000324856</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Domain name</td>
<td>DUF3518</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Domain start</td>
<td>1975</td>
<td></td>
</tr>
<tr>
<td>Visualization Component</td>
<td>Description</td>
<td>Gene Symbol</td>
<td>Sample</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>CNA view</td>
<td>Genome annotations*</td>
<td>ARID1A</td>
<td>TCGA-A6-2676</td>
</tr>
<tr>
<td></td>
<td>Domain end</td>
<td>2231</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chromosome</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Start</td>
<td>27022521</td>
<td></td>
</tr>
<tr>
<td></td>
<td>End</td>
<td>27108601</td>
<td></td>
</tr>
<tr>
<td>Network view</td>
<td>Protein-protein interactions</td>
<td>ARID1A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Source (gene symbol)</td>
<td>ARID1A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Target (gene symbol)</td>
<td>SMAD2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Network name</td>
<td>iRefIndex</td>
<td></td>
</tr>
<tr>
<td>Heatmap view</td>
<td>Continuous-valued data (e.g. gene expression, DNA methylation)</td>
<td>ARID1A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample</td>
<td>TCGA-A6-2676</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Values</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Aberrations view</td>
<td>Binary data</td>
<td>ARID1A</td>
<td></td>
</tr>
<tr>
<td>Sample annotations</td>
<td>Categorical (e.g. expression subtype) or continuous (e.g. tumor purity) data</td>
<td>TCGA-A6-2676</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample</td>
<td>TCGA-A6-2676</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value (continuous)</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value (categorical)</td>
<td>HER2-Enriched</td>
<td></td>
</tr>
</tbody>
</table>

**Supplementary Table 3:** Data types included in MAGI, organized by the visualization component in which the data is used. Users can upload all data types, except those indicated with an asterisk (*).