

API

ALLEN BRAIN ATLAS API

- ISH Surveys
 - Experimental Overview and Metadata
 - ISH Expression Quantification
 - ISH Expression Search Service
- RNA-Seq Surveys
 - Experimental Overview and Metadata
 - RNA-Seq Expression Download Service
 - RNA-Seq Expression Search Service



The **Ivy Glioblastoma Atlas Project (Ivy GAP)** is a foundational resource for exploring the anatomic and genetic basis of glioblastoma at the cellular and molecular levels.

Six studies were designed to identify the molecular signatures and measure heterogeneity. **In Situ Hybridization (ISH)** was used to screen for gene expression enriched in particular structures and cell clusters, and laser microdissection followed by **RNA sequencing** were used to generate the transcriptomes and identify the genetic markers.

Overall, the dataset spans 42 individual tumors. Each tumor was sub-divided into sub-blocks for processing in one of the six studies. See [whitepapers](#) for more details about the tumor specimens, tissue and informatics processing.

From the API, you can:



Download images



Download quantified ISH expression values by tumor feature



Query the ISH expression differential search services



Download RNA-Seq expression values



Query the RNA-Seq correlative and differential search services

This document provides a brief overview of the data, database organization and example queries. API database object names are in camel case. See the main [API Documentation](#) for more information on data models and query syntax.

ISH Surveys

Experimental Overview and Metadata

Experimental data from the four ISH surveys is associated with the "Glioblastoma" Product.

Multiple genes were assayed using each sub-block Specimen. The Specimen is cryosectioned into 20µm thick sections. ISH sections are interleaved with H&E sections such that every ISH is adjacent to an H&E section, yielding a single image SectionDataSet for each gene and one H&E SectionDataSet with 11-15 images.

Each sub-block Specimen is associated with a SpecimenType identifying the study to which this sub-block belongs.

Each H&E section is processed through a semi-automated annotation application that labels anatomic features using a statistical machine learning algorithm. The algorithm associates each 45x45 pixel neighborhood to a tumor feature label. Additionally, a separate algorithm counts the number of nuclei and the nuclei fraction coverage within each neighborhood.

Tumor feature statistics for each sub-block Specimen are generated by summing up the 45x45 pixel neighborhoods over all H&E images. This process produces area, normalized area, nuclei count, and nuclei fraction coverage for each sub-block Specimen and feature of interest.

Anatomic tumor features (also referred to as structures) are hierarchically organized into a tree in which a child structure is a “part of” its parent structure. Structures are assigned colors to visually emphasize their relationships in the hierarchy. See the [structure ontology page](#) for more information.

Examples:

- All donors in the "Glioblastoma" Product

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::Donor,
rma::criteria,products[name$eq'Glioblastoma']
```

- All sub-block Specimens which are part of the "Anatomic Structures ISH Survey"

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::Specimen
,rma::criteria,specimen_types[name$eq'Anatomic Structures ISH Survey']
,rma::options[num_rows$eqall]
```

- All TumorFeatures in sub-block Specimen "W32-1-1-K.01"

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::TumorFeature
,rma::criteria,data_set(specimen[external_specimen_name$eq'W32-1-1-K.
01'])
,rma::include,structure
```

- All ISH SectionDataSets in sub-block Specimen "W32-1-1-K.01"

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::SectionDataSet
,rma::criteria,specimen[external_specimen_name$eq'W32-1-1-K.01'],trea
tments[name$eq'ISH']
,rma::include,genes,sub_images
```

From the above query, specimen "W32-1-1-K.01" has 21 ISH SectionDataSet (id=278269453 is one of them). In the web application, images from a specimen are displayed in a [specimen page](#). All displayed information, images and structural expression values are also available through the API.



See the [image download](#) page to learn how to download images at different resolutions and regions of interest.

Tumor Subdivision

Tumor: W32-1-1 [Show Patient, Tumor, MRI Summary] [Detailed Clinical Data]

Tissue Blocks from the Tumor: W32-1-1-A, W32-1-1-B, W32-1-1-C, W32-1-1-E, W32-1-1-G, W32-1-1-I, W32-1-1-J, W32-1-1-K, W32-1-1-L, W32-1-1-M

Tissue Sub-Blocks from the Tissue Block: W32-1-1-K.01, W32-1-1-K.02

Tumor Features in this Sub-Block

- Cellular Tumor (CT)
- Perinecrotic zone (CTpnz)
- Pseudopalisading cells but no visible necrosis (CTpnn)
- Pseudopalisading cells around necrosis (CTpan)
- Hyperplastic blood vessels in cellular tumor (CThbv)
- Microvascular proliferation (CTmpv)
- Necrosis (CTne)

Section Information

Gene: TGFBR2
 Experiment: 278271484
 Section Number: 3
 Treatments: ISH
 Study: Cancer Stem Cells ISH Survey

Genes Surveyed in this Sub-block

Select genes below to limit the images shown.

Gene	Selected
MET	<input checked="" type="checkbox"/>
MYC	<input checked="" type="checkbox"/>
NO32	<input checked="" type="checkbox"/>
OLIG2	<input checked="" type="checkbox"/>
PDGFRA	<input checked="" type="checkbox"/>
PDPN	<input checked="" type="checkbox"/>
PI3	<input checked="" type="checkbox"/>
POSTN	<input checked="" type="checkbox"/>
PROM1	<input checked="" type="checkbox"/>
TGFBR2	<input checked="" type="checkbox"/>
TNFAIP3	<input checked="" type="checkbox"/>

ISH Image

Symbol: TGFBR2 ISH [Sync]

Name: transforming growth factor, beta re...

image 2 of 21 zoom:2.56%

1932 microns

Nearest H&E Image

TumorFeatureBoundary

image 1 of 10 zoom:2.56%

1932 microns

Add the 'TGFBR2' experiment from this sub-block to your selections

View Selections (0) Clear Selections

Figure: Specimen page of specimen "W32-1-1-K.01" showing all 21 ISH SectionDataSet on the left image panel and 10 H&E images on the right image panel. The color codes for the anatomic tumor features outlined in the H&E images are displayed on the left.

ISH Expression Quantification

For every ISH image, pixels with gene expression are detected and a grayscale mask generated. The detection algorithm is based on adaptive thresholding and mathematical morphology. Then the ISH image is registered to its closet H&E image using a multi-resolution elastic registration algorithm. Finally, feature label and nuclei count information is transferred from 45x45 pixel neighborhood in the H&E image onto the expression data.



Expression statistics for each tumor feature are computed by combining values from all the neighborhoods with the same label. This process produces expression density, intensity and energy measurements for each experiment and anatomical feature. For an anatomical feature, expression energy is defined as sum of expressing pixel divided by sum product of pixels in each neighborhood label for that feature and its nuclei fraction coverage.

Examples:

- Download tumor feature expression values for POSTN SectionDataSet (id=265857641) in sub-block "W8-1-1-E.1.03"

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::StructureUnionize
,rma::criteria,section_data_set[id$eq265857641]
,rma::include,structure
```

ISH Expression Search Service

The expression search service allows users to instantly search over the ~18000 SectionDataSets to find genes with specific expression patterns:

- The **Expression Search** function allows users to find genes which expression in one structure (or set of structures)
- The **Differential Search** function allows users to find genes which have higher expression in one structure (or set of structures) compared to another structure (or set of structures)



The expression search functionality is available through the Web application and the API.

To perform an **Expression Search**, a user specifies a set of target structures. For each StructureDataSet, the average expression energy is computed for the target structures. The returned results are sorted in descending order of average expression energy.

See the [connected service](#) page for definitions of **service::gbm_ish_expression** parameters

Examples:

- Find genes (SectionDataSets) with expression in pseudopalisading cells around necrosis (CTpan)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
service::gbm_ish_expression[structures$eqCTpan][threshold$eq0,100000]
```

- Find genes (SectionDataSets) with expression in hyperplastic blood vessels (LEhbv, IThbv, CThbv)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
service::gbm_ish_expression[structures$eqLEhbv,IThbv,CThbv][threshold
$eq0,100000]
```

- [Visualize expression search in the web application](#)

To perform a **Differential Search**, a user specifies a set of target structures and a set of contrast structures. For each StructureDataSet, the sum expression energy is computed for the target structures and for the contrast structures. The returned results are sorted in descending order of the ratio of the sum expression energy of the target structures over the sum expression energy of the contrast structures.

See the [connected service](#) page for definitions of **service::gbm_ish_differential** parameters

Examples:

- Find genes (SectionDataSets) with higher expression in pseudopalisading cells around necrosis (CTpan) than cellular tumor cells (CT)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=service::gbm_
ish_differential
[structures1$eqCT][structures2$eqCTpan][threshold1$eq0,100000][thresh
old2$eq1,100000]
```

- Find genes (SectionDataSets) with higher expression in infiltrating tumor (IT) than cellular tumor (CT) and leading edge (LE)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=service::gbm_ish_differential
[structures1$eqIT,CT][structures2$eqIT][threshold1$eq0,100000][threshold2$eq1,100000]
```

- Visualize expression search in the web application

It should be noted that the expression quantification values used by the search services are generated by a fully automated processing pipeline. False positive and false negative results can occur due to artifacts on the tissue section or slide and/or algorithmic inaccuracies. Users should confirm results by visually inspecting the ISH images.

RNA-Seq Surveys

Experimental Overview and Metadata

Experimental data from the two RNA-Seq surveys is associated with the "Human Glioblastoma RNASeq" Product.

For the "Anatomic Structures RNA-Seq" study, sampling locations were manually identified after H&E staining. For the "Cancer Stem Cells RNA-Seq", 17 reference gene probes were used to identify 35 types of putative cancer stem cell clusters.

Each sampling site is associated with a Structure with the following naming scheme:

- **XX-reference-histology**: tumor feature XX sampled by reference histology
- **XX-reference-genes**: tumor feature XX sampled by reference gene(s)
- **XX-reference-controls**: tumor feature XX sampled by low expression of reference genes

Structures are organized hierarchically into a tree in which children structures are "parts of" their parent structure. Structures are assigned colors that visually emphasize the hierarchical relationships. See the [structure ontology page](#) for more information.

See [whitepapers](#) for more details about the RNA-Seq data generation and normalization. All gene and sampling site information can be accessed through the API.

Examples:

- All donors in the "Human Glioblastoma RNASeq" Product

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::Donor,
rma::criteria,products[name$eq'Human Glioblastoma RNASeq']
```

- All samples in the "Human Glioblastoma RNASeq" Product

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::Sample
,rma::criteria,tissue_sample_data_set(products[name$eq'Human
Glioblastoma RNASeq'])
,rma::include,structure,tissue_sample_data_set(specimen(donor))
,rma::options[num_rows$eqall]
```

RNA-Seq Expression Download Service



Normalized gene-level expression values can be downloaded in several ways:

- From the web application [Download](#) page
- From the connected data service in the API

Using the connected service, expression values can be obtained by specifying:

- a list of "probes" (genes),
- a list of "donors" (tumor specimens, optional), and
- a list of structures (optional)

See the [connected service](#) page for definitions of `service::gbm_expression` parameters.

Example:

Download expression values for gene ESM1 restricted to samples from tumor 'W1-1-2'

- Find Specimen ID for tumor 'W1-1-2' (id=703393)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::Specimen
,rma::criteria,[external_specimen_name$eq'W1-1-2']
```

- Find Gene ID for human gene ESM1 (id=10924)

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
model::NcbiGene
,rma::criteria,[acronym$eqESM1],organism[name$eq'Homo Sapiens']
```

- Use tumor and gene ID as parameters to `service::gbm_expression`

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
service::gbm_expression[probes$eq10924][donors$eq703393]
```

The output of the service has two top level ordered arrays "probes" and "samples". For example:

```

"probes":[{ "id":10924,
            "name":"endothelial cell-specific molecule 1",
            "gene-id":10924,
            "gene-symbol":"ESM1","gene-name":"endothelial cell-specific molecule
1",

"entrez-id":11082,"chromosome":"5","start-position":"n/a","end-position":"
n/a",

"expression_level":["1.0498","0.3790","0.1323","0.1894","0.2724","0.0268",
"0.0000","1.3249","1.0818","0.9904","1.4085","0.0301","0.0000","0.6687","6
.9482","7.9008","6.8374","8.5132"],

"z-score":["-0.6017","-0.8612","-0.9566","-0.9345","-0.9024","-0.9974","0.
0000","-0.4953","-0.5893","-0.6247","-0.4630","-0.9961","0.0000","-0.7491"
,"1.6797","2.0481","1.6368","2.2850"]}
],"samples":[
{"donor":{"id":703393,"name":"W1-1-2","age":"","color":"646464"},"sample":
{"well":300629386,"polygon":298763268,"mri":[-1,-1,-1]},
"structure":{"id":298764274,"name":"Leading Edge sampled by reference
histology","abbreviation":"LE-reference-histology","color":"218FA5"},
"top_level_structure":{"id":9429,"name":"Glioblastoma","abbreviation":"GBM
","color":"006EC6"}},
{"donor":{"id":703393,"name":"W1-1-2","age":"","color":"646464"},"sample":
{"well":300629325,"polygon":298726093,"mri":[-1,-1,-1]},
"structure":{"id":298764274,"name":"Leading Edge sampled by reference
histology","abbreviation":"LE-reference-histology","color":"218FA5"},
"top_level_structure":{"id":9429,"name":"Glioblastoma","abbreviation":"GBM
","color":"006EC6"}},
{"donor":{"id":703393,"name":"W1-1-2","age":"","color":"646464"},"sample":
{"well":300173638,"polygon":298727595,"mri":[-1,-1,-1]},
"structure":{"id":298764274,"name":"Leading Edge sampled by reference
histology","abbreviation":"LE-reference-histology","color":"218FA5"},
"top_level_structure":{"id":9429,"name":"Glioblastoma","abbreviation":"GBM
","color":"006EC6"}}, ...
],

```

Each **probe** (Gene) contains information about:

- the Gene (id, name)
- the associated NCBI Gene (id, acronym, name, entrez-id), along with
- a vector of normalized expression values and z-scores in the same order as the "samples" array.

Each **sample** contains information about:

- the Tumor (id and name returned in the "donor" field)
- the associated Structure (id, name, acronym and color)

RNA-Seq Expression Search Service

Differential Search

The Differential Search finds genes that show the greatest difference in expression values between two sets (target and contrast) of user-defined structures. For each probe, a 2-sample t-test is performed followed by a Benjamini and Hochberg false discovery rate correction. The null

hypothesis is that the average expression level of samples in the contrast set of structures is greater than or equal to the average expression level of samples in the target set of structures. A statistically significant result (p-value less than user-defined threshold) allows us to reject the null hypothesis and conclude that the average expression level of samples in the target set of structures is greater than the average expression level of samples in the contrast set of structures. Resulting p-values are sorted in ascending order. Search results can also be sorted by fold-change (log ratio of expression) in descending order.



The differential search functionality is available through the Web application and the API.

See the [connected service page](#) for definitions of `service::gbm_differential` parameters.

Example:

- Differential search for genes with higher expression in microvascular proliferation (sampled by reference histology) than in cellular tumor, infiltrating tumor, leading edge and pseudopalisading cells around necrosis.

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
service::gbm_differential
[structures1$eq'CT-reference-histology', 'IT-reference-histology', 'LE-
reference-histology', 'CTpan-reference-histology' ]
[structures2$eq'CTmvp-reference-histology' ][sortBy$eq'fold-change' ]
```

- Visualize differential search results in the web application



Figure: Screenshot of top returns of a differential search for genes with higher expression in microvascular proliferation.

Correlative Search

The Correlative Search finds genes with expression profiles similar to that of a selected seed "probe" over all samples within a user-specified structure and for user-specified tumors/donors. Pearson's correlation coefficients are computed for all probes and the results ranked in descending order.



The correlative search functionality is available through the Web application and the API.

See the [connected service page](#) for definitions of **service::gbm_correlation** parameters.

Example:

- Correlative search for genes with a similar expression to gene VEGFA (id=7379) over all samples

```
http://api.brain-map.org/api/v2/data/query.xml?criteria=
service::gbm_correlation[probes$eq7379][structures$eqGBM]
```