

# Synapse R Client Demo

Matt Furia

November 1, 2011

## 1 Sage Bionetworks Synapse project

The recent exponential growth of biological “omics” data has occurred concurrently with a decline in the number of NMEs approved by the FDA. Sage Bionetworks believes that a fundamental reason biological research productivity does not scale with biological data generation is that the analysis and interpretation of genomic data remains largely an isolated, individual activity. Sage Bionetworks’ mission is to catalyze a cultural transition from the traditional single lab, single-company, and single-therapy R&D paradigm to a model with broad precompetitive collaboration on the analysis of large scale data in medical sciences. For this to happen it is critical that: 1) human health data become accessible and reusable by people other than the original data generators allowing multiple parallel approaches to data interpretation; 2) analytical methodologies become fully reproducible and transparent so that results can be vetted and existing analysis techniques quickly applied to new application areas, and; 3) models of biological systems and networks be opened to a variety of users such that theoretical predictions can be rapidly validated experimentally and improve standards of care for patients. Sage Bionetworks is actively engaged with academic and pharmaceutical collaborators in developing technical, policy and strategic solutions to these issues. Part of Sage’s solution is Synapse, a platform for open, reproducible data-driven science, which will provide support for Sage’s research initiatives and serve as a resource for the broader scientific community.

Synapse will support the reusability of information facilitated by ontology-based services and applications directed at scientific researchers and data curators. Sage Bionetworks is actively pursuing the acquisition, curation, statistical quality control, and hosting of human and mouse global coherent datasets for use by Sage Bionetworks researchers, collaborators, and the broader research community. Global coherent datasets integrate both clinical phenotype and genomic data along with an intermediate molecular layer such as gene expression or proteomic data. Typically studies contain genome-wide genetic variation data and/or expression profiling data. We expect the release of these sorts of unique, integrative, high value datasets into the public domain will seed a variety of analytical approaches to drive new treatments based on better understanding of disease states and the biological effects of existing drugs.

Reproducible analysis and disease model reuse require a supporting informatics infrastructure. In the Sage Bionetworks system, users interact with resources via a number of mechanisms depending upon their interests and expertise. The Synapse web portal will be an environment for end user scientists to interact and share data, models, and analysis methods, both in the context of specific research projects, and broadly across otherwise disparate projects. Many other specialized scientific tools can be extended to load data and save results to the Sage Bionetworks platform, or to perform analysis by calling methods executed on a remote service. The Sage Bionetworks Platform is aggressively leveraging and optimizing its architecture to take full advantage of the rapidly maturing cloud computing technologies which will put on-demand supercomputing power in the hands of the average researcher. These more specialized analytical clients would support use cases in data curation and quality control as well as scientific analysis.

## 2 Set up your Synapse work environment

Go to <https://synapse-alpha.sagebase.org/> to register for a new account and then log into Synapse.

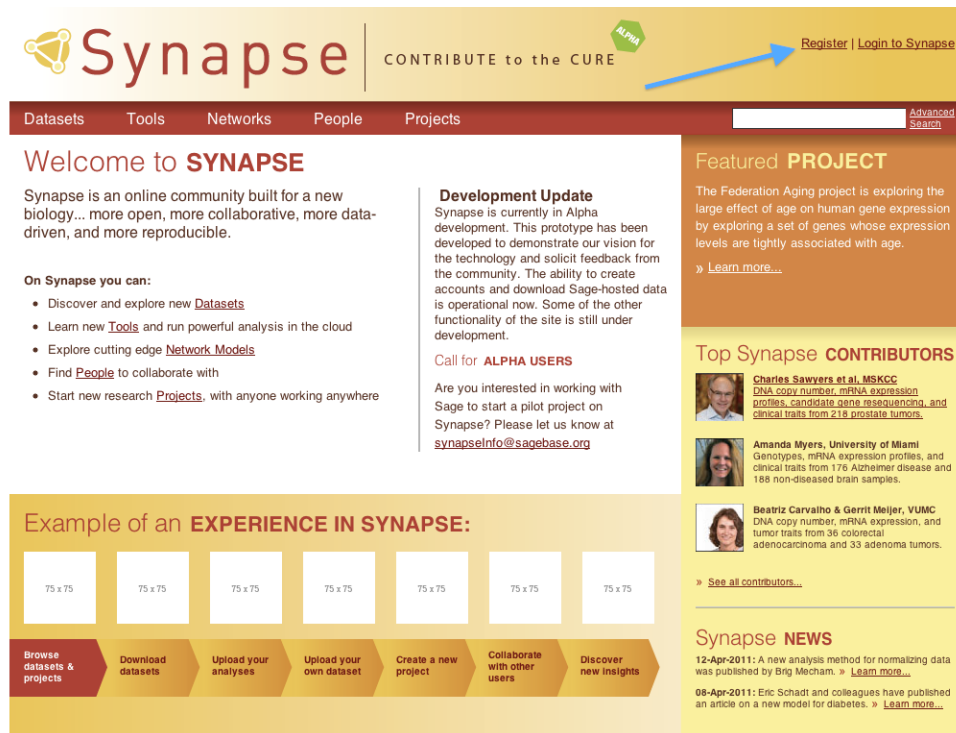


Figure 1: Register for a Synapse account and log in

Use the following R code to setup your Synapse work environment.

Load the synapse client and log in. Be sure to type in your Synapse username and password when prompted from R.

```
> library(affy)
> library(simpleaffy)

> library(synapseClient)
> synapseLogin()
```

Create a Synapse project to hold your analyses results.

Create a project object using it's constructor. The list contains name-value pairs of properties that should be added to the project. See help documentation for details on the properties that can be set. For projects, only name is required.

```
> # Change the values of this variable
> myName <- 'your name'
> projName <- sprintf("%ss Curation Project %s",
+                     myName,
+                     gsub(':', '_', date()))
> myProj <- Project(list(name=projName))
```

Show the project. Note that entity id is missing.

```
> myProj
```

An object of class "Project"

Synapse Entity Name : your names Curation Project Tue Nov 1 15\_22\_49 2011

Synapse Entity Id :

For complete list of annotations, please use the `annotations()` function.

To view this Entity on the Synapse website use the '`onWeb()`' function

or paste this url into your browser:

Create the project in Synapse using `createEntity`. make sure to catch the return value.

```
> myProj <- createEntity(myProj)
```

Now the Synapse Entity Id is populated.

```
> myProj
```

An object of class "Project"

Synapse Entity Name : your names Curation Project Tue Nov 1 15\_22\_49 2011

Synapse Entity Id : 24034

Parent Id : 4489

For complete list of annotations, please use the `annotations()` function.

To view this Entity on the Synapse website use the '`onWeb()`' function

or paste this url into your browser: <http://synapse.sagebase.org/#Lookup:24034>

Go back to <https://synapse-alpha.sagebase.org/> using the `onWeb` function and find your newly created project.

```
> onWeb(myProj)
```

Click on “share” and share your project with the group `AUTHENTICATED_USERS`.

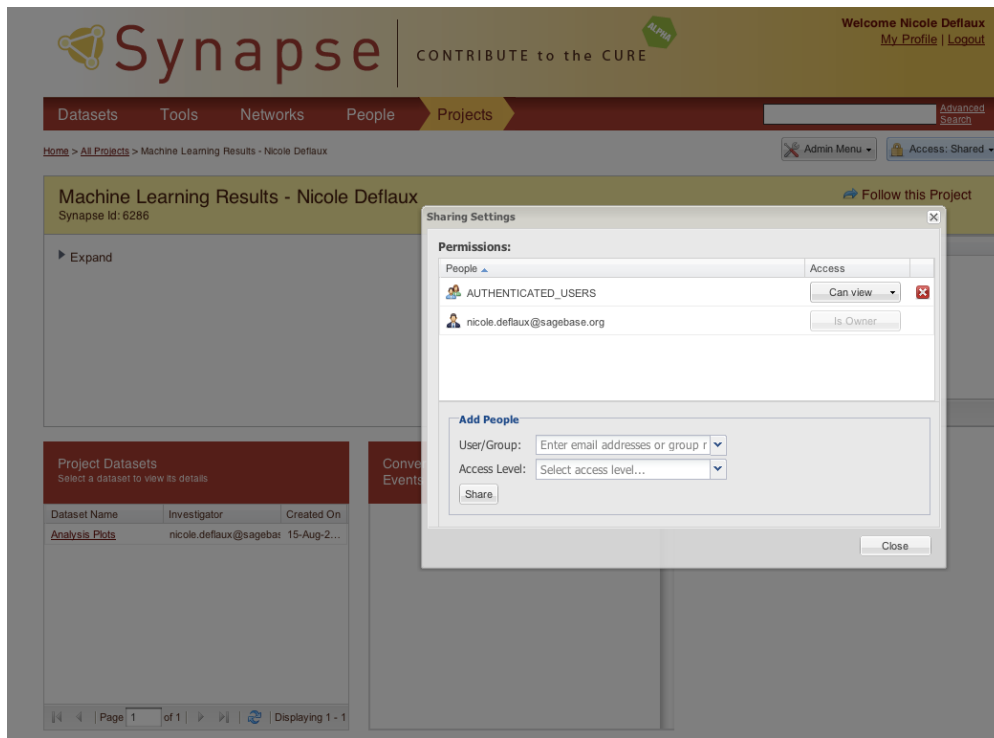


Figure 2: Find your project and share it with AUTHENTICATED\_USERS

### 3 Load data from Synapse

Navigate to the “metaGeo” in Synapse <https://synapse.sagebase.org/#Project:16114>. Click on the two datasets listed there and view their details. Note that you can also browse data available in Synapse via the Synapse R Client. See the help documentation for synapseClient for more detail.

TODO change the screenshot

The screenshot shows the Synapse website interface. At the top, there's a yellow header with the Synapse logo and the text 'CONTRIBUTE to the CURE'. Below this is a navigation bar with tabs for 'Datasets', 'Tools', 'Networks', 'People', and 'Projects'. The 'Projects' tab is selected. The main content area displays the 'Cell Line Project' page. It includes a table with 'Name' and 'Value' columns, showing a single entry with 'uri' as the name and '/repo/v1/project/2657' as the value. Below this, there's a section titled 'Project Datasets' which lists several datasets, including 'CCLE Data' and 'Sanger IC50 Cell Line Drug Response'. A blue arrow points to the 'Project Datasets' section. To the right of the 'Project Datasets' section, there's a 'Conversations and Events for this Project' section.

Figure 3: Find the CCLE and Sanger Dataset Layers in Synapse

Download a metageo expression layer.

```
> geoEntityId <- "23994"
> expr <- loadEntity(geoEntityId)
```

Inspect the contents.

```
> expr
```

```
An object of class "ExpressionLayer"
Synapse Entity Name : GSE10024_rawExpression
Synapse Entity Id   : 23994
Parent Id           : 23993
Type                : E
Version             : 1.0.0
```

```
loaded object(s):
[1] "expression" (list)
```

```
12 File(s) cached in "/Users/deflaux/.synapseCache/pub/geo/DATA/supplementary/series/GSE10024/GSE10024_RA"
```

For complete list of annotations, please use the `annotations()` function.  
To view this Entity on the Synapse website use the `'onWeb()'` function  
or paste this url into your browser: <http://synapse.sagebase.org/#Lookup:23994>

## 4 Create Data in Synapse

Create a dataset.

```
> myDataset <- createEntity(Dataset(list(name="my Data",
+                                       parentId=propertyValue(myProj, "id"))))
```

View the dataset on the web to add a description.

```
> onWeb(myDataset)
```

TODO screen shot here?

Refresh the local copy of myDataset

```
> myDataset <- refreshEntity(myDataset)
```

Write the pm values to a text file.

```
> write.table(pm(expr$objects$expression[["HG-U133A"]]),
+             file="pm.txt",
+             sep="\t",
+             quote=F,
+             row.names=F)
```

Create a new expression layer.

```
> myExpr <- createEntity(Layer(list(name="curated expression",
+                                   type="E",
+                                   parentId = propertyValue(myDataset, "id"),
+                                   status="curated")))
```

Add an annotation specifying the data format.

```
> annotValue(myExpr, "format") <- "sageBioCurated"
```

Add the pm data file to the entity

```
> myExpr <- addFile(myExpr,
+                   "pm.txt",
+                   path="GSE10024/expression/affymetrix")
```

Store the data

```
> myExpr <- storeEntity(myExpr)
```

## 5 Create a Plot in Synapse

Create a heatmap of some probes and push that to Synapse.

```
> jpeg(file = "heatmap.jpg")
> hmap.eset(expr$objects$expression[[1]],
+           probesets=101:200)
> dev.off()
```

null device

1

```

> plot <- synapseClient::Media(list(name = "heatmap",
+                                   parentId=propertyValue(myDataset,"id")))
> plot <- addFile(plot,"heatmap.jpg")
> plot <- storeEntity(plot)

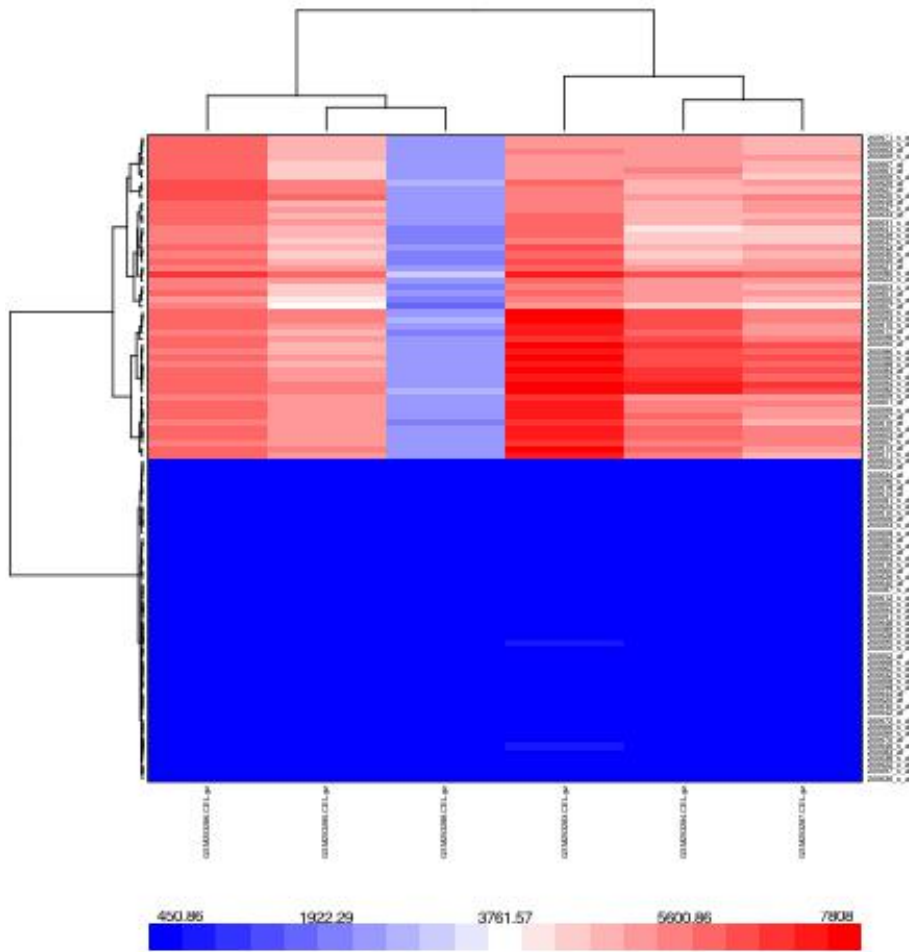
```

Show the plot from R.

```

> plot

```



Show the plot on the web

```

> onWeb(plot)

```

## 6 Next steps

Implement your own customized predictive models and use the cross validation evaluation procedures to test if your method achieves more accurate predictive performance than currently used methods.

## 7 Session Information

```

> toLatex(sessionInfo())

```

- R version 2.13.1 (2011-07-08), x86\_64-apple-darwin9.8.0
- Locale: C/en\_US.UTF-8/C/C/C/C
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: Biobase 2.12.2, affy 1.30.0, gcrma 2.24.1, genefilter 1.34.0, hgu133acdf 2.8.0, simpleaffy 2.28.0, synapseClient 0.8-1
- Loaded via a namespace (and not attached): AnnotationDbi 1.14.1, Biostrings 2.20.4, DBI 0.2-5, IRanges 1.10.6, RCurl 1.6-9, RJSONIO 0.95-0, RSQLite 0.10.0, affyio 1.20.0, annotate 1.30.1, digest 0.5.1, preprocessCore 1.14.0, splines 2.13.1, survival 2.36-9, tools 2.13.1, xtable 1.6-0