



1.

the policy  
environment.



# Office of Science and Technology Policy

[About OSTP](#)[Pressroom](#)[OSTP Blog](#)[Divisions](#)[Initiatives](#)[R&D Budgets](#)[Resource Lib](#)

## Expanding Public Access to the Results of Federally Funded Research [Subscribe](#)

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST

[E-Mail](#)[Tweet](#)[Share](#)

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand of whom recently signed a *We the People* petition asking for expanded



[Home](#) • [Briefing Room](#) • [Presidential Actions](#) • [Executive Orders](#)

The White House

Office of the Press Secretary



E-Mail



Tweet



Share



For Immediate Release

May 09, 2013

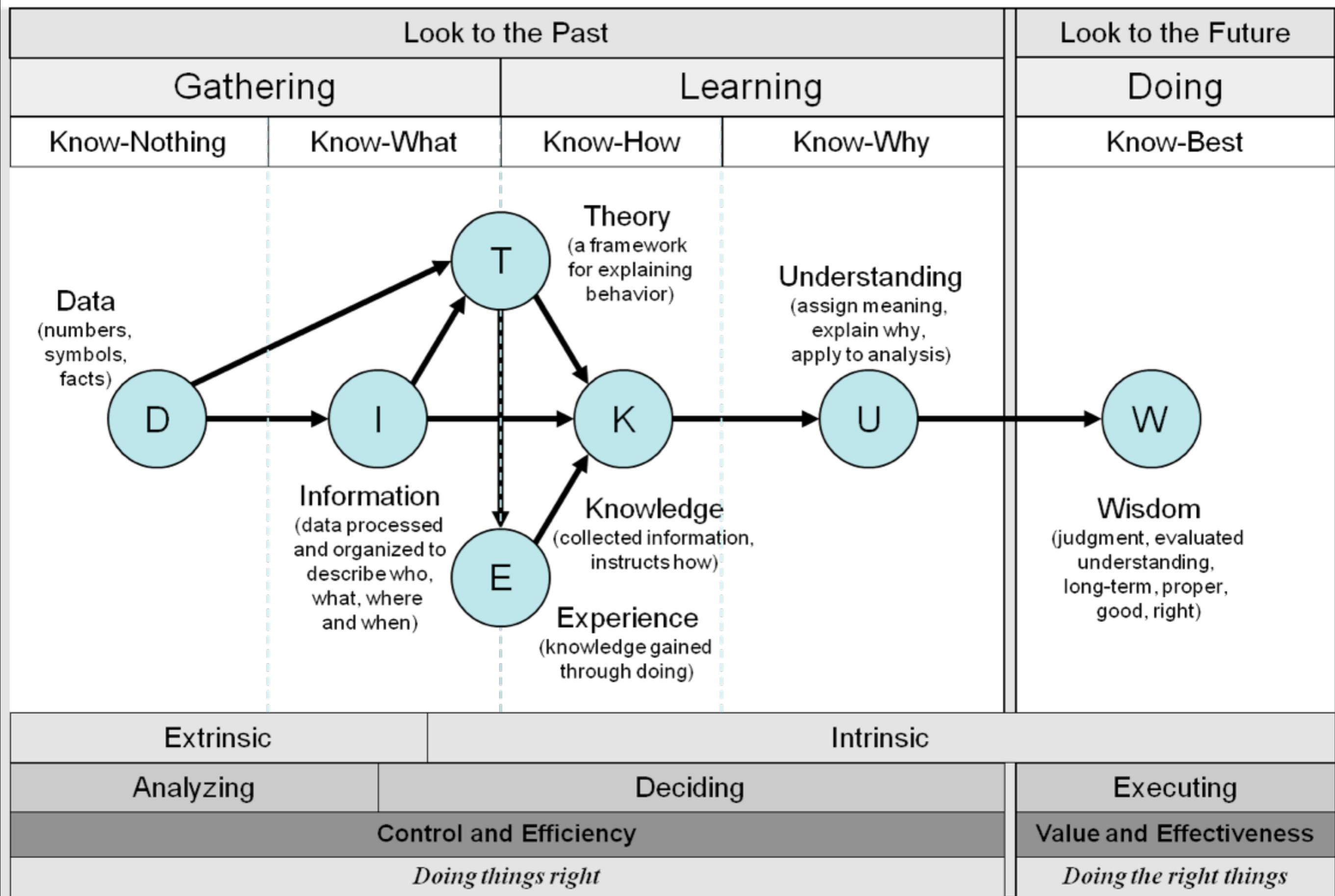
## **Executive Order -- Making Open and Machine Readable the New Default for Government Information**

EXECUTIVE ORDER

-----

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT  
FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:





# Certify your open data

Show that it's easy to  
**find, use and share**

Create a certificate

Browse all certificates

# Untitled

**What's this data called?** This is required for all levels

Data Title

**Where is it described?**

Documentation URL

**Who publishes this data?** This is required for all levels

Data Publisher

**What website is the data published on?**

## Legal Information

Rights, licensing and privacy

## Practical Information

Findability, accuracy, quality and guarantees

## Technical Information

Locations, formats and trust

## Social Information

Documentation, support and services

## **Rights** your right to share this data with people

**Do you have the rights to publish this data as open data?** This is required for all levels

- yes, you have the rights to publish this data as open data
  - no, you don't have the rights to publish this data as open data
  - you're not sure if you have the rights to publish this data as open data
  - the rights in this data are complicated or unclear
- 

## **Licensing** how you give people permission to use this data

**Where have you published the rights statement for this dataset?**

## **Findability** how you help people find your data

Is there a link to your data from your main website?

- no
- yes

Is your data listed within a collection?

- no
- yes

Is this data referenced from your own publications?

## **Locations** how people can access your data

---

## **Formats** how people can work with your data

Is this data machine-readable?

- no
- yes

Is this data in a standard open format?

- no
- yes

## Documentation how you help people understand the context and content of your data

Where is the technical documentation for the data?

Do the data formats use vocabularies or schemas? This is required for all levels

- no
- yes

Are there any codes used in this data? This is required for all levels

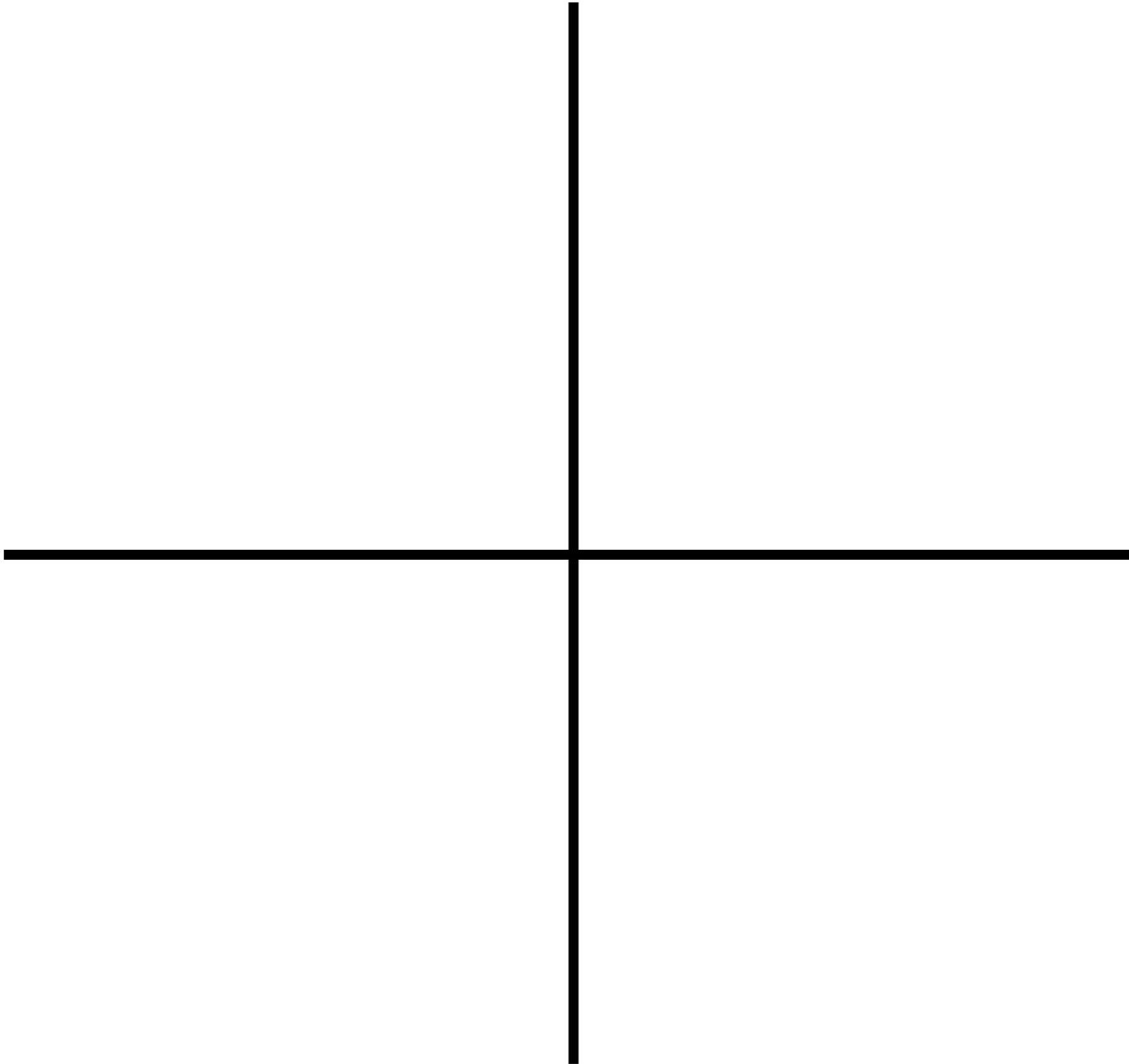
“is it open?” is  
perhaps not the  
right frame.

accessibility

adaptability

ease  
of mastery

leverage



## Western Wind Resources Dataset

This interface gives access to the 30,000+ sites that were modeled by 3TIER as part of the [Western Wind and Solar Integration Study](#).

The data can be accessed in two ways:

- Use interactive map to zoom in and click on a turbine
- Choosing a Station ID from the [metadata file](#) and typing it into the form below

Enter Station Id:

go



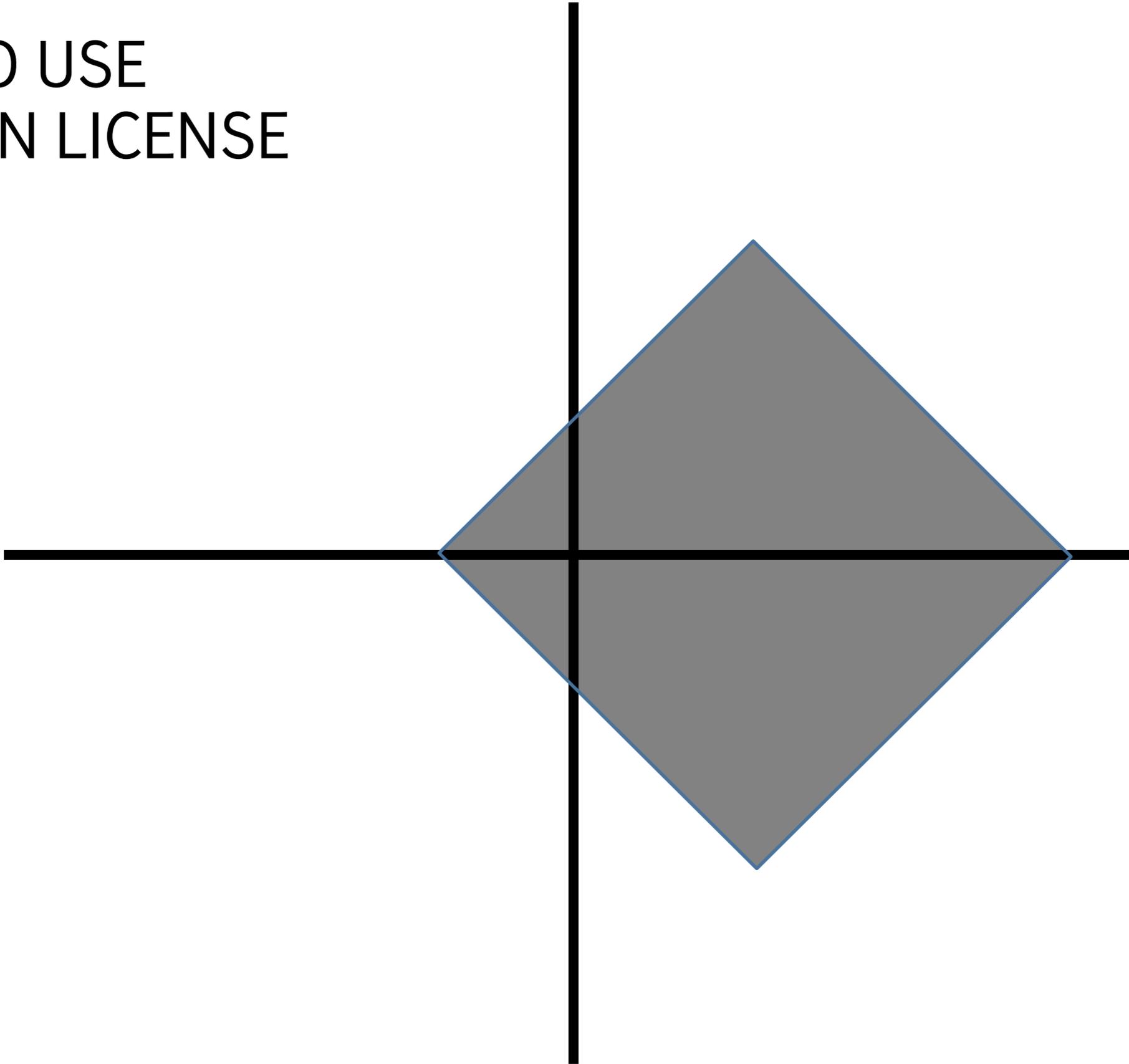
EASY TO USE  
NO OPEN LICENSE

accessibility

adaptability

ease  
of mastery

leverage



	A	B	C	D	E	F	G	H
	SiteID	Latitude	Longitude	Power Dens	SCORE-lite C	Wind Speed	State Code	Model Elevation [
2	1	31.192	-102.242	413.925	30.823	7.741	TX	849.5
3	2	31.192	-102.225	419.46	31.227	7.798	TX	859.4
4	3	31.192	-102.208	429.712	32.042	7.911	TX	873.6
5	4	31.208	-102.258	385.712	29.985	7.668	TX	902.7
5	5	31.208	-102.242	375.495	29.305	7.609	TX	908.1
7	6	31.208	-102.225	378.675	29.539	7.655	TX	915.9
8	7	31.208	-102.208	383.684	30.01	7.723	TX	922.4
9	8	31.225	-102.242	447.048	34.466	8.145	TX	947.9
0	9	31.225	-102.225	441.857	34.324	8.146	TX	953.1
1	10	31.225	-102.208	437.837	34.109	8.131	TX	948.7
2	11	31.275	-104.542	425.601	30.468	7.738	TX	1344.4
3	12	31.275	-104.525	453.123	31.794	7.907	TX	1363.2
4	13	31.292	-104.525	440.34	30.684	7.797	TX	1361.8
5	14	31.292	-104.458	562.009	33.463	8.075	TX	1284.2
6	15	31.292	-104.442	540.765	32.127	7.926	TX	1277.7
7	16	31.308	-104.508	477.308	31.735	7.911	TX	1344.1
8	17	31.308	-104.492	524.97	33.154	8.06	TX	1320.3
9	18	31.308	-104.475	516.722	32.106	7.947	TX	1309.2
0	19	31.325	-104.525	455.937	30.635	7.843	TX	1370.5
1	20	31.325	-104.458	523.132	31.632	7.966	TX	1320.3
2	21	31.342	-104.542	423.342	28.901	7.676	TX	1398.2
3	22	31.342	-104.458	486.605	30.157	7.812	TX	1341
4	23	31.342	-104.375	455.77	28.751	7.574	TX	1330.3
5	24	31.342	-104.358	454.683	28.451	7.485	TX	1298.3
6	25	31.358	-104.508	482.053	30.973	7.905	TX	1390.7
7	26	31.358	-104.442	502.848	30.961	7.913	TX	1347.8
8	27	31.358	-104.425	479.688	30.067	7.805	TX	1353.6
9	28	31.358	-104.408	487.624	30.527	7.854	TX	1358.1
0	29	31.358	-104.392	459.859	29.323	7.674	TX	1349
1	30	31.358	-104.375	467.813	29.509	7.662	TX	1330
2	31	31.375	-104.508	478.267	31.022	7.917	TX	1412.8

## **About the Weather Research & Forecasting Model**

The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. It features multiple dynamical cores, a 3-dimensional variational (3DVAR) data assimilation system, and a software architecture allowing for computational parallelism and system extensibility. WRF is suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers.



The effort to develop WRF has been a collaborative partnership, principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (the National Centers for Environmental Prediction (NCEP) and the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA). WRF allows researchers the ability to conduct simulations reflecting either real data or idealized configurations. WRF provides operational forecasting a model that is flexible and efficient computationally, while offering the advances in physics, numerics, and data assimilation contributed by the research community.

# This is modeled (not measured) wind data

## Western US:

- 3TIER generates historical wind data by running a Numerical Weather Prediction Model using physical conservation equations that 'recreate the weather' for 2004-6.
- They sampled the weather at a 1 arc-minute (~2km) spatial and 10 minute temporal resolution, and at 5 hub heights (10, 20, 50, 100, 200m).
- Based on a limited number of actual tower measurements for that time period, they did a sophisticated adjustment of (MOS-corrected) the model so that the data more accurately reflects actual wind speed measurements.
- Validation reports comparing the model wind speed results to actual data will be produced by 3TIER when their modeling is complete.
- 3TIER used SCORE-lite to convert wind speed to power output, assuming a Vestas V90 3 MW turbine at 100m hub height. Each grid point can hold 10 turbines or a 30 MW wind plant. The SCORE-lite process applies a probability distribution function to the manufacturer's power curve to replicate actual wind farm output.
- 3TIER produced hourly forecasts for day-ahead wind power output for the 30,000 selected sites.

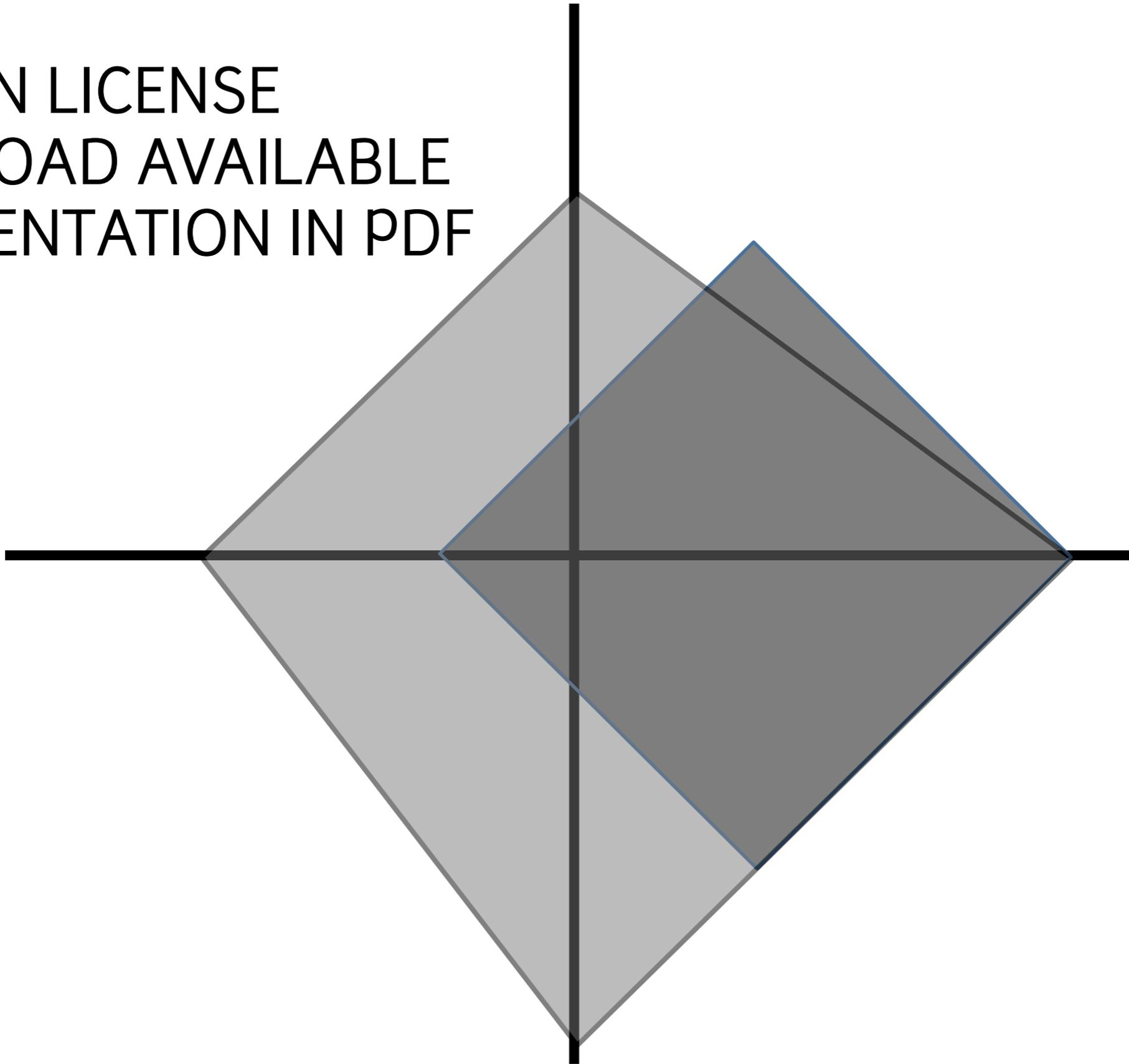
NO OPEN LICENSE  
DOWNLOAD AVAILABLE  
DOCUMENTATION IN PDF

adaptability

accessibility

ease  
of mastery

leverage

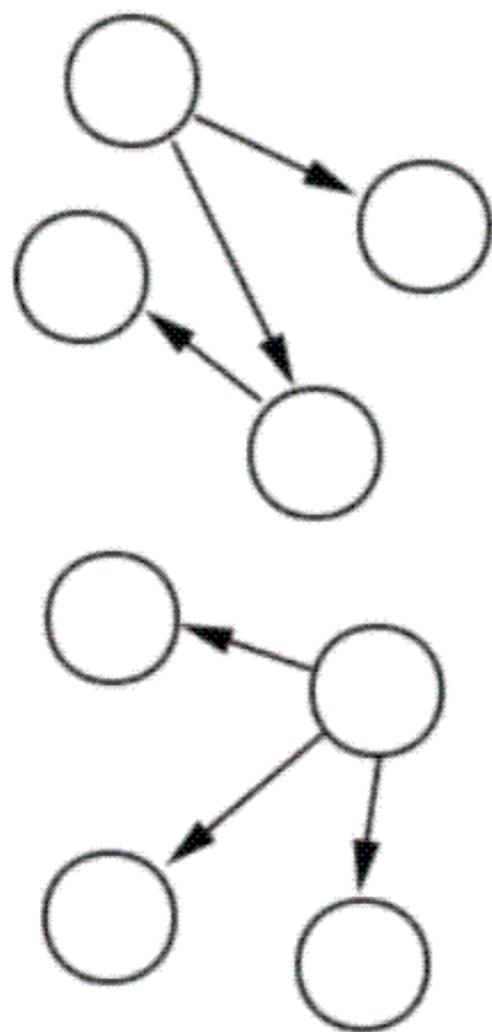


2.

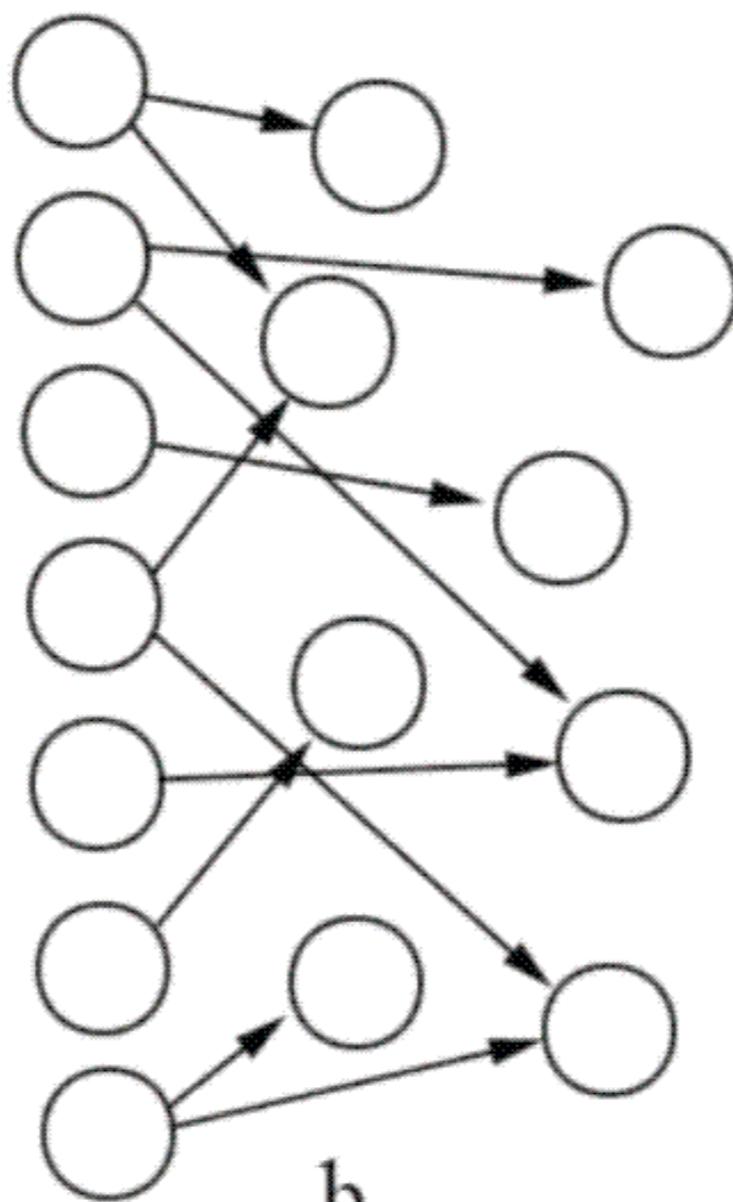
doing research in the  
open: early returns.

“how accurately can we predict if a female breast cancer survivor will develop a second tumor?”

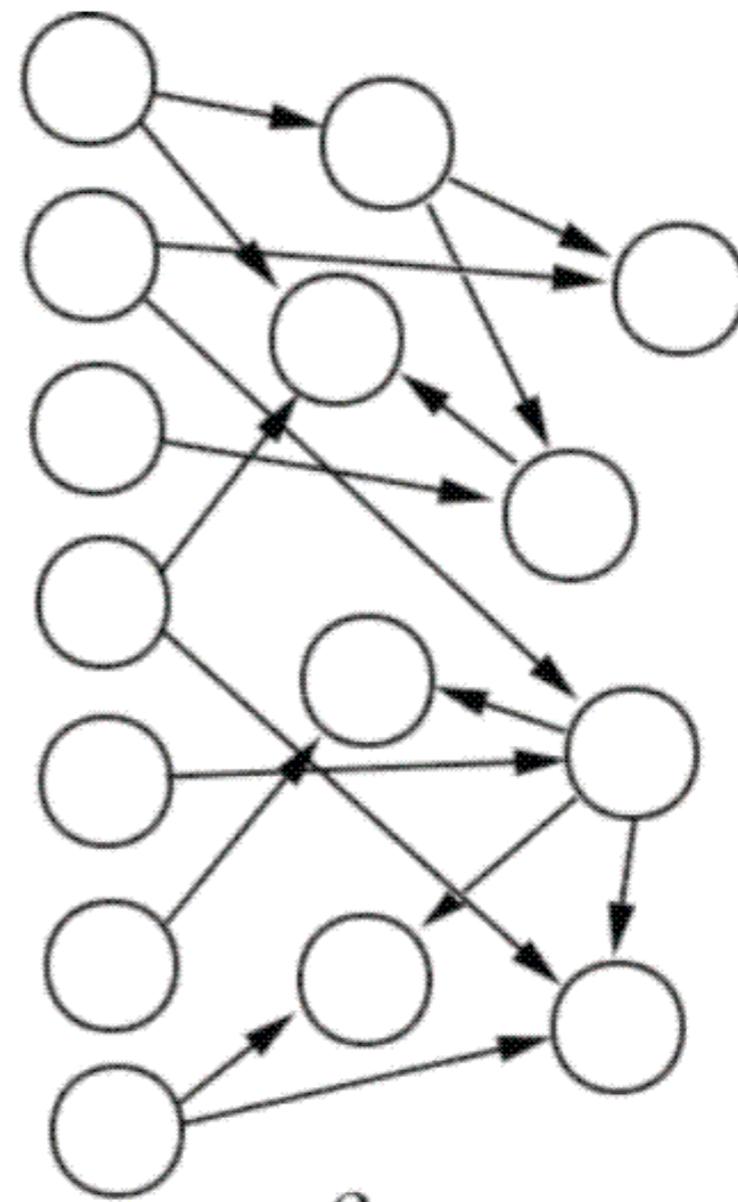
may the best (statistical) model win



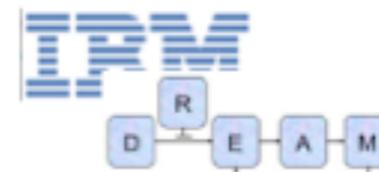
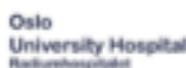
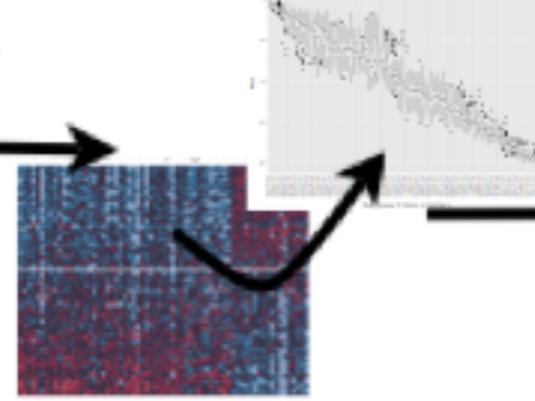
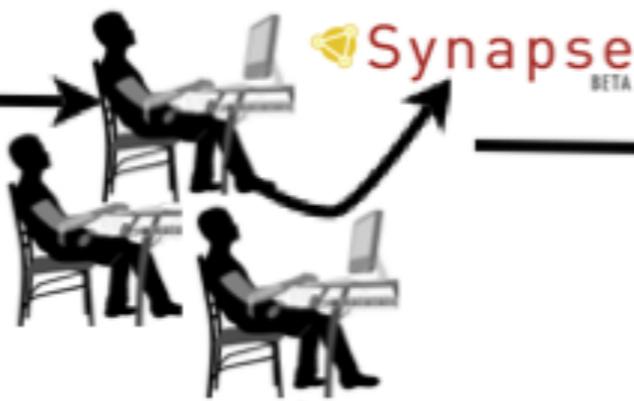
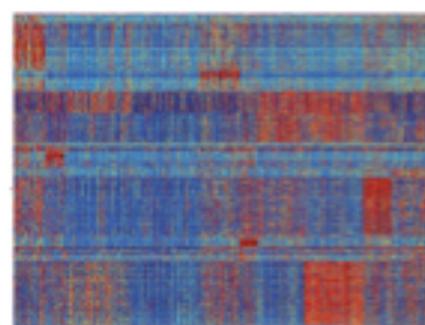
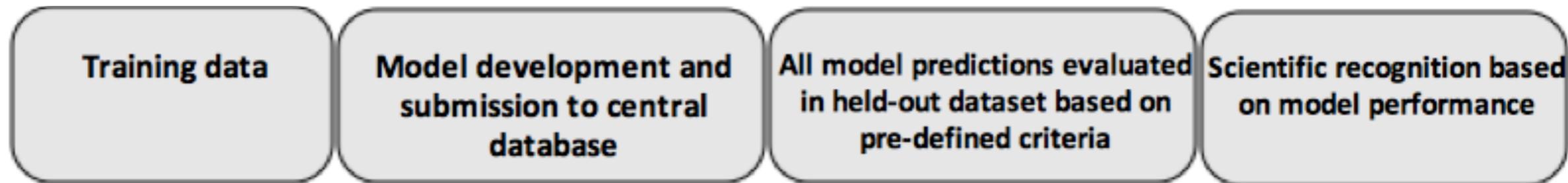
a.



b.



c.



code sharing a prerequisite.

# Public Sage/DREAM Modeling Competition Leaderboard

8/28: We have re-run the evaluation due to a bug found in our exactConcordanceIndex calculation. We apologize for any inconvenience.

ID	Name	Submitter	Status	Received	Train Score	Test Score	Detail
<a href="#">syn1123807</a>	Attractor Metagenes Model 090607	317809	complete	2012-09-06 20:45:22	0.7802159106	0.7449735640	<a href="#">Output</a>
<a href="#">syn1123222</a>	SC_TWEEK_SEP4_3	362391	complete	2012-09-04 23:49:26	0.7758041311	0.7443261519	<a href="#">Output</a>
<a href="#">syn1237990</a>	BH0913S7	962237	complete	2012-09-14 05:50:28	0.7793607196	0.7433010826	<a href="#">Output</a>
<a href="#">syn1335448</a>	re-tweaked SC_TWEEK_SEP4_3	362375	deleted	2012-09-18 00:53:12	0.7774088719	0.7427795562	
<a href="#">syn1125638</a>	BH 0911 BL2	962237	complete	2012-09-11 22:48:49	0.7805026511	0.7424018991	<a href="#">Output</a>
<a href="#">syn1123481</a>	Attractor Metagenes Model 090602	317809	complete	2012-09-06 17:49:20	0.7860865453	0.7420422257	<a href="#">Output</a>
<a href="#">syn1199235</a>	BH0913S3	962237	complete	2012-09-13 20:06:22	0.7789783989	0.7416645686	<a href="#">Output</a>
<a href="#">syn1335400</a>	BH0917S7	962237	complete	2012-09-17 22:48:38	0.7825953538	0.7416106176	<a href="#">Output</a>
<a href="#">syn1158570</a>	BH0912S10	962237	complete	2012-09-13 04:43:14	0.7802209411	0.7414307809	<a href="#">Output</a>
<a href="#">syn1123773</a>	Attractor Metagenes Model 090606	317809	complete	2012-09-06 19:42:20	0.7822432163	0.7414127972	<a href="#">Output</a>
<a href="#">syn1123477</a>	Attractor Metagenes Model 090601	317809	complete	2012-09-06 17:42:54	0.7860865453	0.7407114340	<a href="#">Output</a>
<a href="#">syn1189560</a>	BH0913S1	962237	complete	2012-09-13 17:30:21	0.7781131468	0.7381217854	<a href="#">Output</a>
<a href="#">syn1125753</a>	HybridDriverV15	342024	complete	2012-09-12 01:30:15	0.7827261477	0.7379779161	<a href="#">Output</a>
<a href="#">syn1124564</a>	HybridDriverV6	342024	complete	2012-09-08 02:50:45	0.7820872697	0.7363593857	<a href="#">Output</a>
<a href="#">syn1335353</a>	BH0917S5	962237	complete	2012-09-17 20:53:21	0.8030696327	0.7363414020	<a href="#">Output</a>
<a href="#">syn1123270</a>	WarwickSystemsBiology (Wed Sep 5 13.40.37 2012)	362302	complete	2012-09-05 14:11:27	0.8280864850	0.7363234183	<a href="#">Output</a>
<a href="#">syn1123295</a>	WarwickSystemsBiology (Wed Sep 5 15.42.23 2012)	362302	complete	2012-09-05 16:00:58	0.8378004487	0.7358198756	<a href="#">Output</a>
<a href="#">syn1123283</a>	WarwickSystemsBiology (Wed Sep 5 15.11.15 2012)	362302	complete	2012-09-05 15:26:28	0.8321360659	0.7347048880	<a href="#">Output</a>
<a href="#">syn1123250</a>	WarwickSystemsBiology (Wed Sep 5 13.06.47 2012)	362302	complete	2012-09-05 13:06:47	0.7374460000	0.7344171400	<a href="#">Output</a>

accuracy of model jumped three orders of magnitude in nine days.

Search the Journal

Vol#

Pg#  [GO >](#)

Enter Keyword

Full text

[GO >](#)

[Advanced >](#)

[Contributing to Sci TM](#)

[► Submit a Manuscript](#)

[About Sci TM](#)

[► Mission](#)

[► Advisory Board](#)

[► Associate Advisors](#)

[► Staff](#)

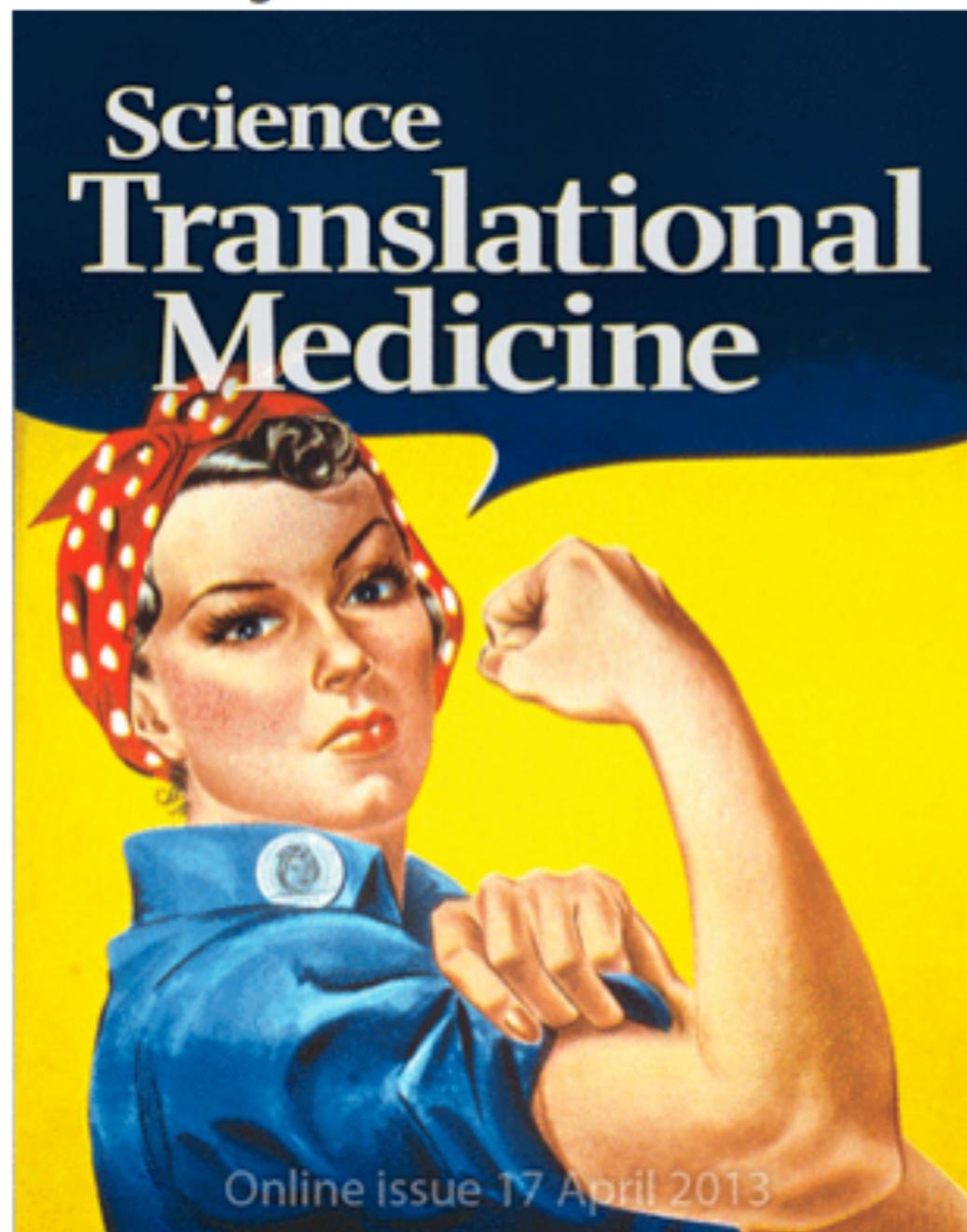
[► Editorial Policies](#)

[► Contact Us](#)

[Information for](#)

[► Authors](#)

### Cover image



76% accurate.

**ONLINE COVER** "Doin' It for Themselves"—and Future Generations. A U.S. cultural icon, Rosie the Riveter represents millions of American women who entered the workforce during World War II. Rosies worked in factories of all kinds, including some that manufactured munitions and other supplies needed

Search the Journal

Vol#

Pg#  [GO >](#)

Enter Keyword

Full text

[GO >](#)

[Advanced >](#)

[Contributing to Sci TM](#)

[► Submit a Manuscript](#)

[About Sci TM](#)

[► Mission](#)

[► Advisory Board](#)

[► Associate Advisors](#)

[► Staff](#)

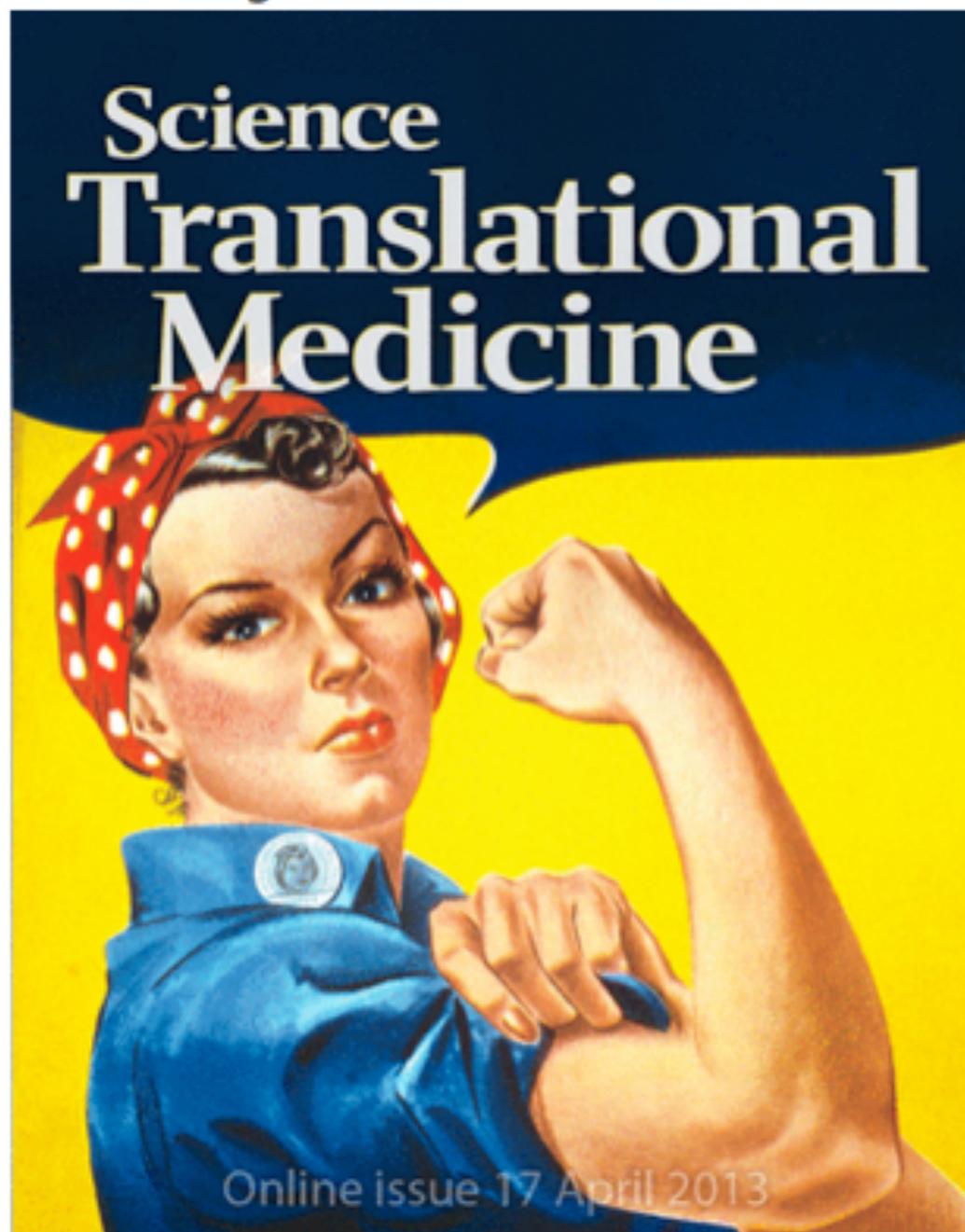
[► Editorial Policies](#)

[► Contact Us](#)

[Information for](#)

[► Authors](#)

### Cover image



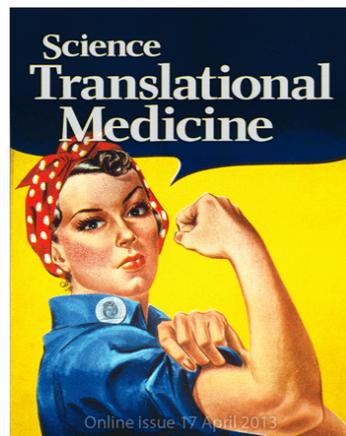
(not a biologist)

**ONLINE COVER** "Doin' It for Themselves"—and Future Generations. A U.S. cultural icon, Rosie the Riveter represents millions of American women who entered the workforce during World War II. Rosies worked in factories of all kinds, including some that manufactured munitions and other supplies needed

21 february 2013

## Biomolecular Events in Cancer Revealed by Attractor Metagenes

Wei-Yi Cheng, Tai-Hsien Ou Yang, Dimitris Anastassiou 



**RESEARCH ARTICLE**

COMPUTATIONAL MODELING

**Development of a Prognostic Model for Breast Cancer Survival in an Open Challenge Environment**

Wei-Yi Cheng, Tai-Hsien Ou Yang and Dimitris Anastassiou\*

17 april 2013

## **Multi-cancer molecular signatures and their interrelationships**

Wei-Yi Cheng<sup>1</sup>, Tai-Hsien Ou Yang<sup>1</sup>, Hui Shen<sup>2</sup>, Peter W. Laird<sup>2</sup>, Dimitris Anastassiou<sup>1</sup> and the Cancer Genome Atlas Research Network

ongoing...



## Multi-cancer molecular signatures and their interrelationships

Wei-Yi Cheng<sup>1</sup>, Tai-Hsien Ou Yang<sup>1</sup>, Hui Shen<sup>2</sup>, Peter W. Laird<sup>2</sup>, Dimitris Anastassiou<sup>1</sup> and the Cancer Genome Atlas Research Network

- Multi-cancer molecular signatures and their interrelationships -- clearScience supplement (Current Page)
  - Figures
    - Figure 1: Scatter plots of top three genes in each attractor in twelve cancer types
    - Figure 2: Scatter plots connecting the LYM, M+ and M- signatures in 12 cancer types
    - Figure S1: Scatter plots of the top three features of the 15 pan-cancer attractors
    - Figure S2: Association between the MES and END signatures
  - Tables
    - Table S1: Attractor clusters in pancan12 data sets
    - Table S2: The consensus rankings of features in each attractor
    - Table S3: Genomically-localized mRNA attractor clusters in pancan12 data sets

## "Multi-cancer molecular signatures and their interrelationships"

**AUTHORS:** Wei-Yi Cheng, Tai-Hsien Ou Yang, Dimitris Anastassiou

### ABSTRACT

Several molecular signatures, called attractor metagenes, present in multiple cancer types were recently identified and proved prognostic for survival. These signatures were found using an iterative data mining algorithm converging to the core of gene coexpression. We report the identification of additional multi-cancer molecular signatures resulting from analysis of data sets from twelve cancer types, including methylation and protein signatures and an angiogenesis-related signature.



## Multi-cancer molecular signatures and their interrelationships

Wei-Yi Cheng<sup>1</sup>, Tai-Hsien Ou Yang<sup>1</sup>, Hui Shen<sup>2</sup>, Peter W. Laird<sup>2</sup>, Dimitris Anastassiou<sup>1</sup> and the Cancer Genome Atlas Research Network

### Multi-cancer molecular signatures and their interrelationships -- clearScience supplement (Current Page)

#### Figures

Figure 1: Scatter plots of top three genes in each attractor in twelve cancer types

Figure 2: Scatter plots connecting the LYM, M+ and M- signatures in 12 cancer types

Figure S1: Scatter plots of the top three features of the 15 pan-cancer attractors

Figure S2: Association between the MES and END signatures

#### Tables

Table S1: Attractor clusters in pancan12 data sets

Table S2: The consensus rankings of features in each attractor

Table S3: Genomically-localized mRNA attractor clusters in pancan12 data sets

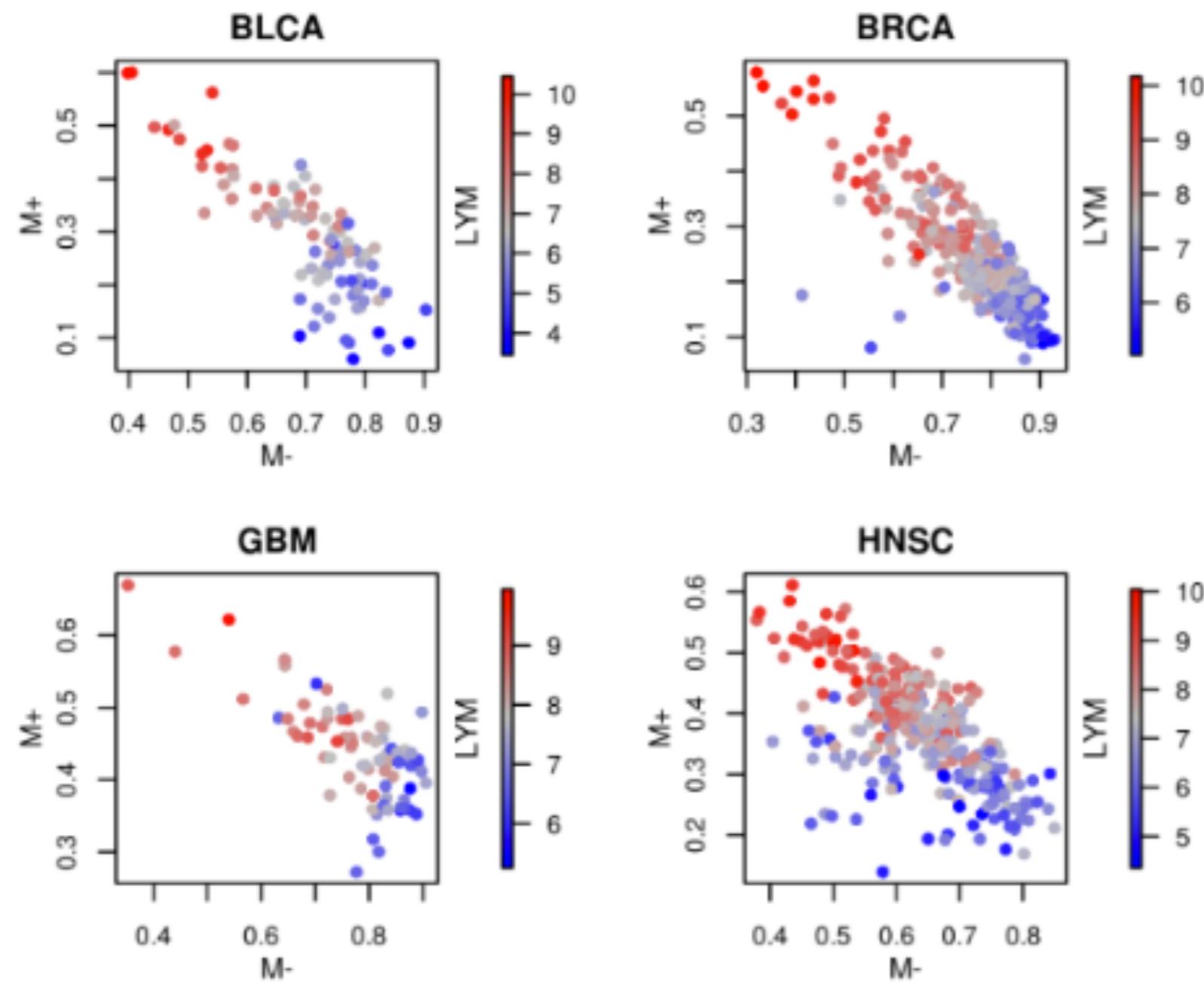
## "Multi-cancer molecular signatures and their interrelationships"

**AUTHORS:** Wei-Yi Cheng, Tai-Hsien Ou Yang, Dimitris Anastassiou

### ABSTRACT

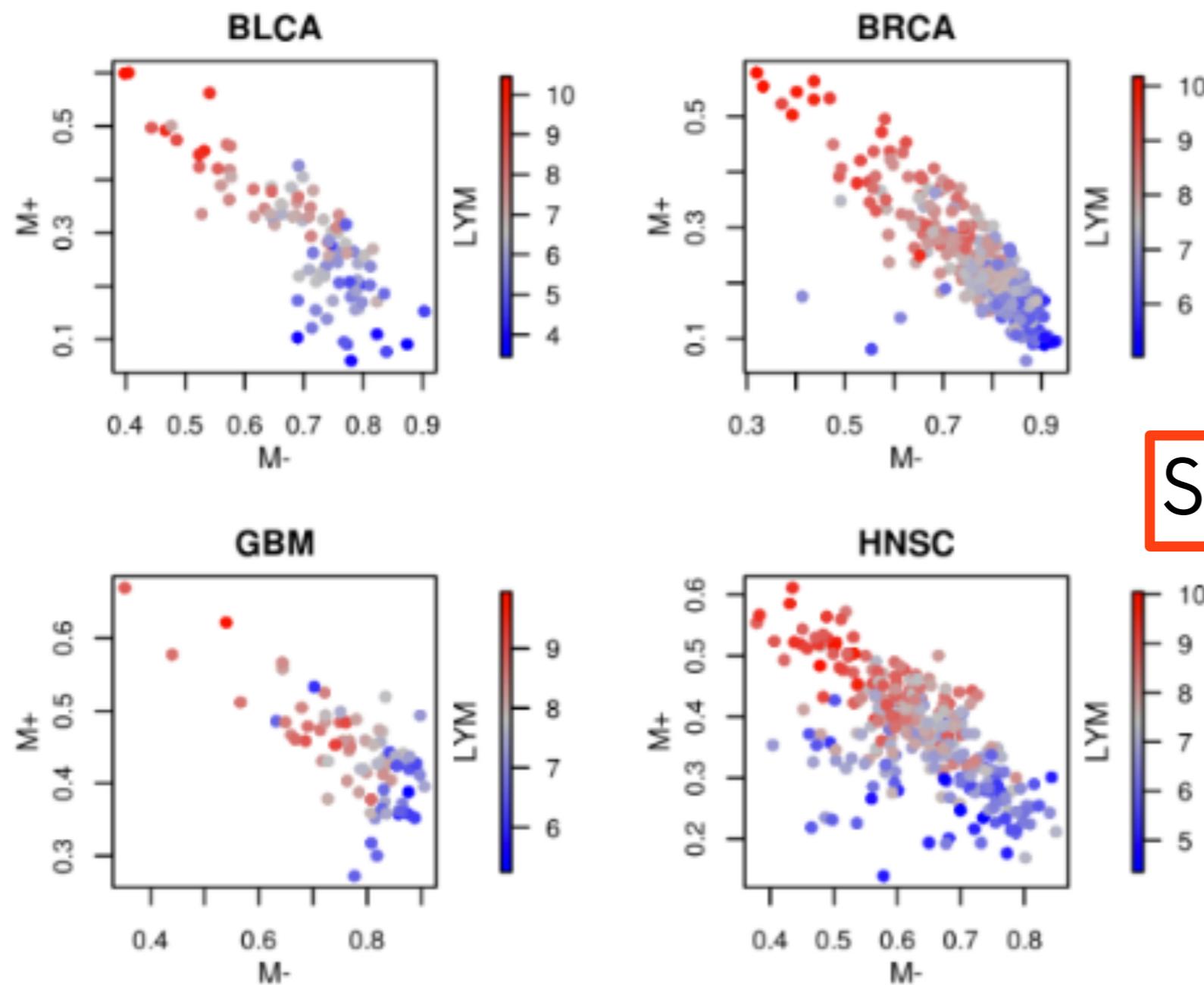
Several molecular signatures, called attractor metagenes, present in multiple cancer types were recently identified and proved prognostic for survival. These signatures were found using an iterative data mining algorithm converging to the core of gene coexpression. We report the identification of additional multi-cancer molecular signatures resulting from analysis of data sets from twelve cancer types, including methylation and protein signatures and an angiogenesis-related signature.

We had previously found all three LYM, M+ and M- signatures from their association with the expression of *miR-142*. We have now confirmed this in the [pancan12 data sets](#), and we found that *miR-150* and *miR-155* are also strongly associated with the LYM signature. We had also previously used the LYM signature as an attractor metagene<sup>3</sup>, and used it in the winning model of the Sage Bionetworks Breast Cancer Prognosis Challenge. The LYM signature is strongly associated with improved prognosis in ER-negative breast cancers, and this fact also provides an explanation for the relative prognosis compared with other types of high-grade breast cancers.



**Figure 2:** Scatter plots connecting the LYM, M+ and M- meta-features in 12 cancer types. Each dot represents a cancer sample. The horizontal axis shows the methylation values of the two methylation signatures, M- and M+, while the value of the expression of the LYM metagene is color-coded.

We had previously found all three LYM, M+ and M- signatures from their association with the expression of *miR-142*. We have now confirmed this in the **pancan12 data sets**, and we found that *miR-150* and *miR-155* are also strongly associated with the LYM signature. We had also previously used the LYM signature as an attractor metagene<sup>3</sup>, and used it in the winning model of the Sage Bionetworks Breast Cancer Prognosis Challenge. The LYM signature is strongly associated with improved prognosis in ER-negative breast cancers, and this fact also provides an explanation for the relationship between LYM and prognosis compared with other types of high-grade breast cancers.



SHOW ME THE CODE!

**Figure 2:** Scatter plots connecting the LYM, M+ and M- meta-features in 12 cancer types. Each dot represents a cancer sample. The horizontal axis shows the methylation values of the two methylation signatures, M- and M+, while the value of the expression of the LYM metagene is color-coded.

## **Multi-cancer molecular signatures and their interrelationships -- clearScience supplement** ★

Synapse ID: [syn1898119](#)

Sharing:  Public ([modify](#))

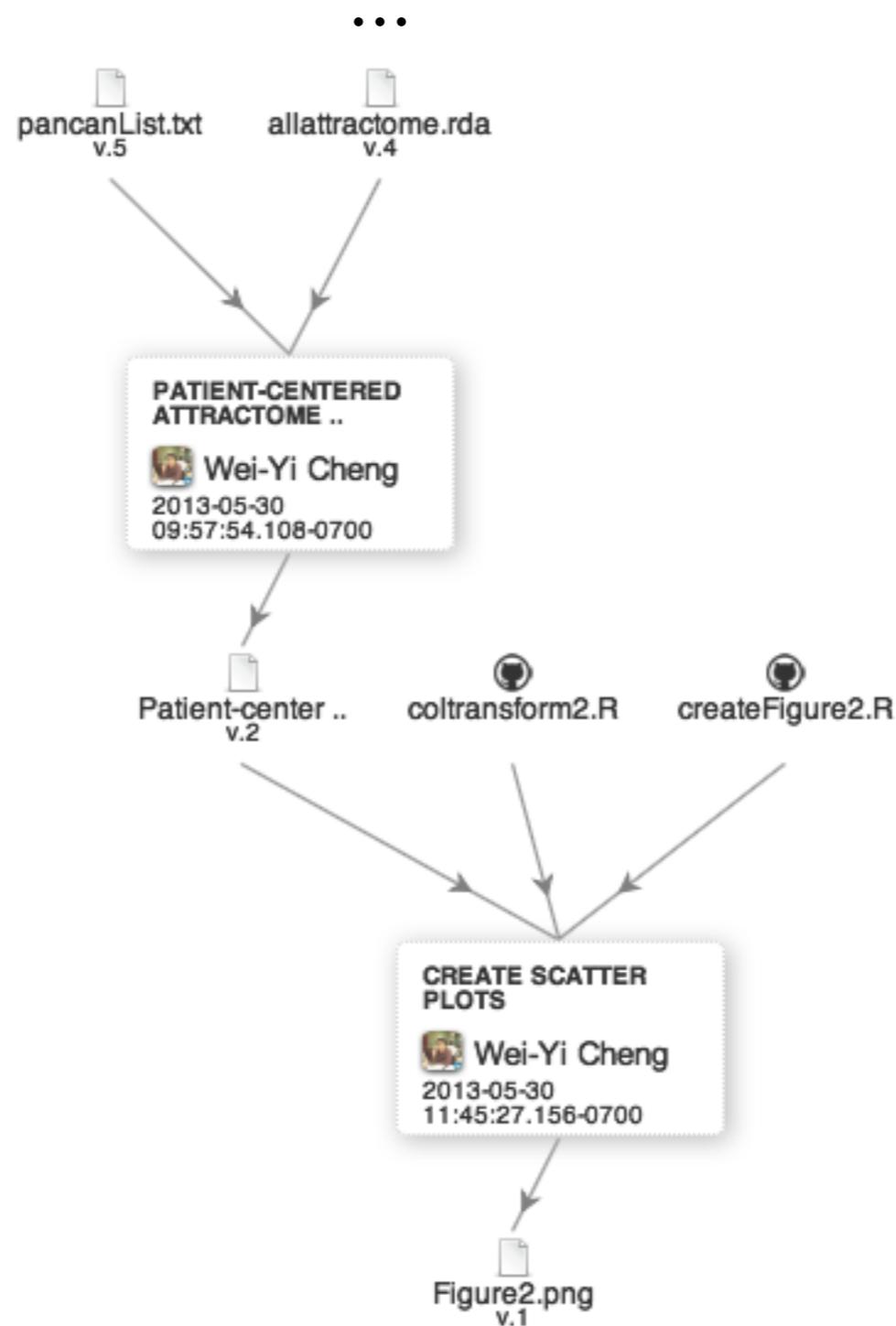
DOI: ([create DOI](#))

## Multi-cancer molecular signatures and their interrelationships -- clearScience supplement

Synapse ID: syn1898119

Sharing:  Public [\(modify\)](#)

DOI: [\(create DOI\)](#)

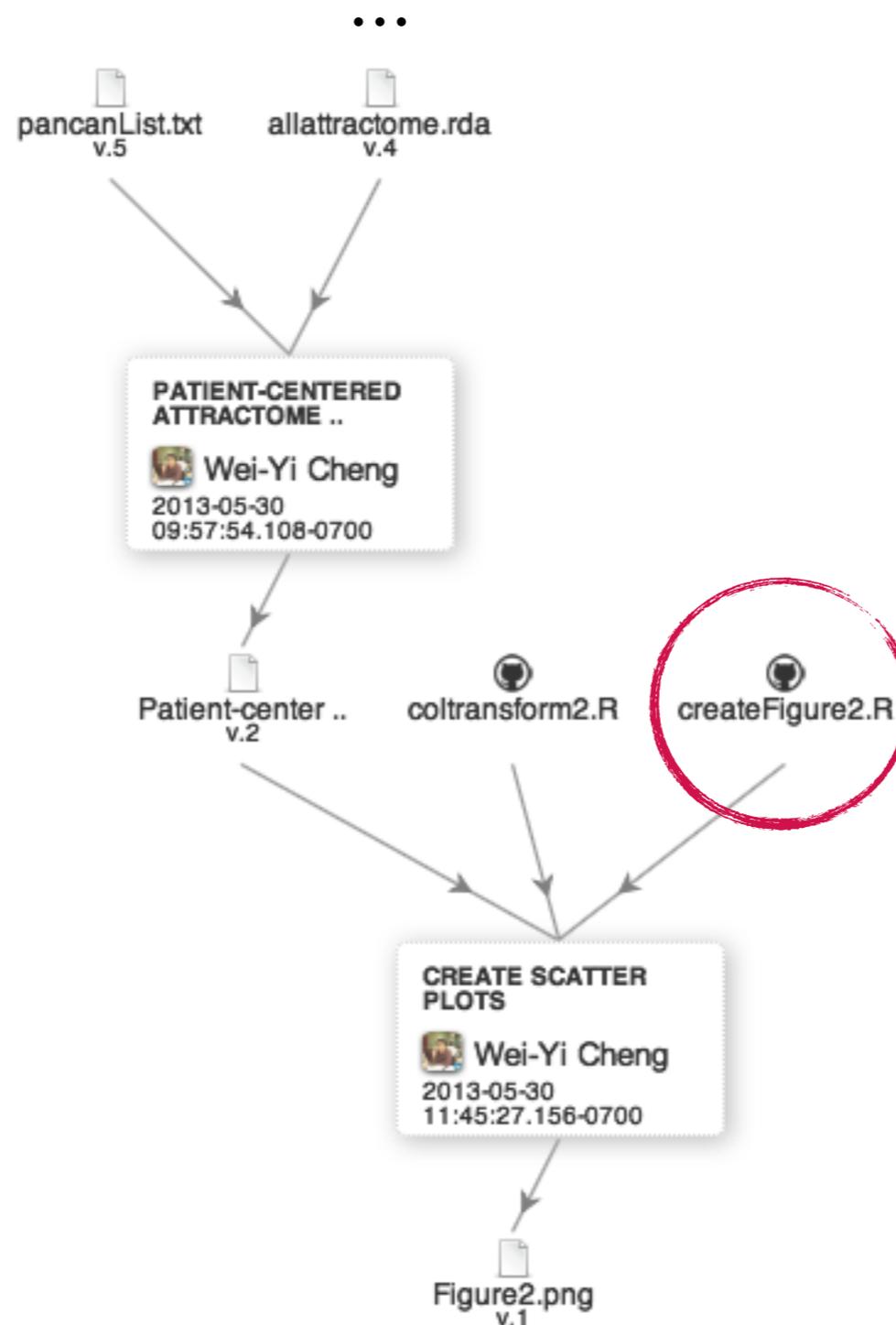


## Multi-cancer molecular signatures and their interrelationships -- clearScience supplement

Synapse ID: syn1898119

Sharing:  Public [\(modify\)](#)

DOI: [\(create DOI\)](#)





tree: 314a0881e6

synapseTCGAPancan / sourceScripts / createFigure2.R



weiyi\_gisl 4 months ago figure 2

0 contributors



file | 55 lines (44 sloc) | 1.527 kb

Edit

Raw

Blame

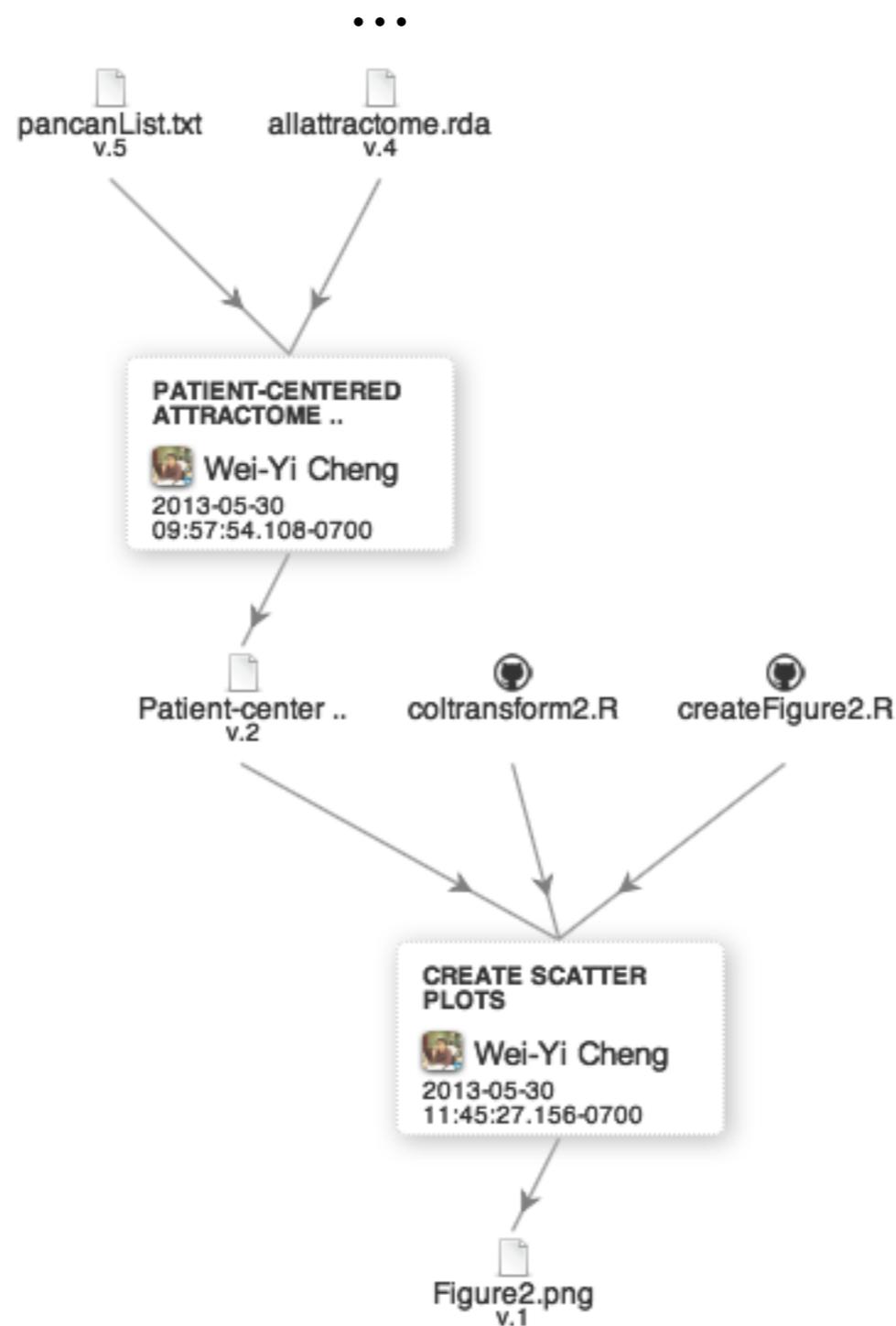
```
1 createFigure2 <- function(meta.pancan, x, y, z){
2
3     fileName <- paste("scatter.", x, "x", y, "x", z, ".png", sep="")
4
5     png(fileName, width=7.3, height=8, units="in", res=300, pointsize=12)
6     par(mar = c(4,4,2,5),          #plot margin
7         mfrow = c(4, 3),
8         oma=c(0, 0, 0, 0),
9         mgp=c(2, 1, 0)
10        )
11
12     # find the features
13     temp <- meta.pancan[[1]]
14     idxx <- NULL
15     idxy <- NULL
16     idxz <- NULL
17     for(d in names(temp)){
18         if(x %in% rownames(temp[[d]])) idxx <- d
19         if(y %in% rownames(temp[[d]])) idxy <- d
```

## Multi-cancer molecular signatures and their interrelationships -- clearScience supplement

Synapse ID: syn1898119

Sharing:  Public [\(modify\)](#)

DOI: [\(create DOI\)](#)

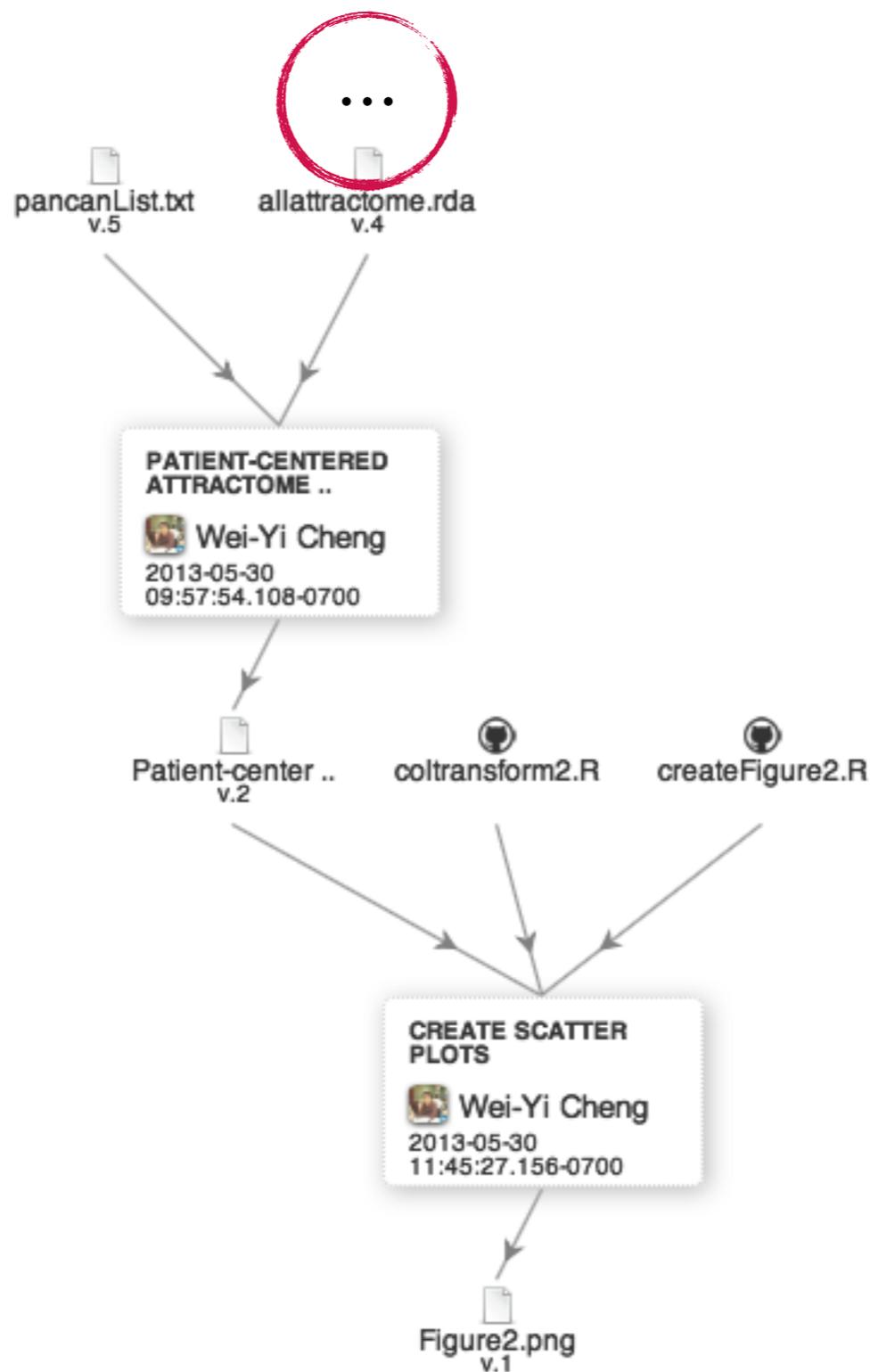


## Multi-cancer molecular signatures and their interrelationships -- clearScience supplement

Synapse ID: syn1898119

Sharing:  Public (modify)

DOI: (create DOI)



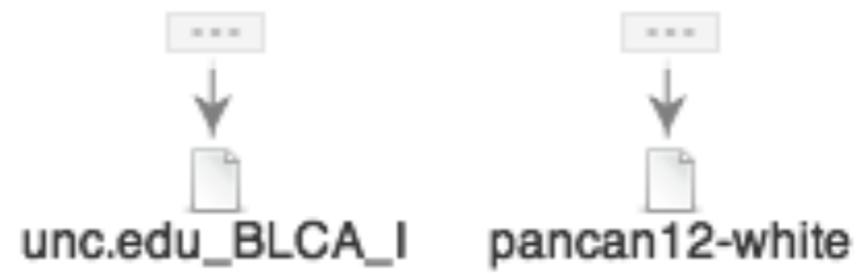
## **Multi-cancer molecular signatures and their interrelationships -- clearScience supplement** ★

Synapse ID: [syn1898119](#)

Sharing:  Public ([modify](#))

DOI: ([create DOI](#))





**WHITELISTING**



**Kyle Ellrott**  
2013-04-05  
16:04:25.664-0700



v:1

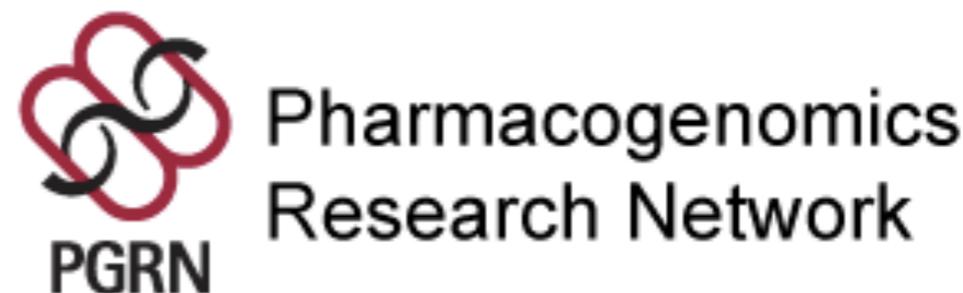
**PARALLEL ATTRACTOR  
SCANNING ..**



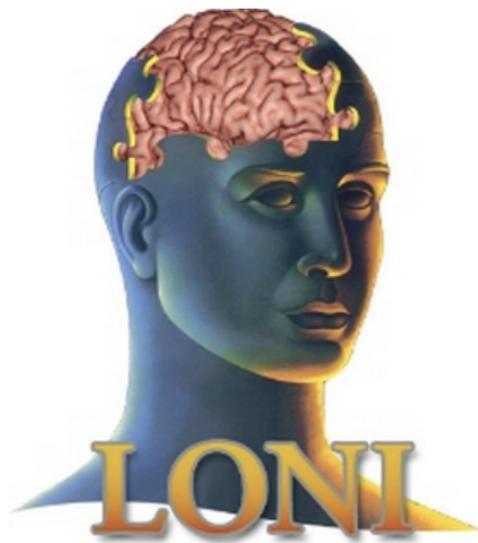
**Wei-Yi Cheng**  
2013-05-21  
10:06:49.950-0700

if we don't have the article in  
machinable form with rights  
to transform? doesn't happen.

can we predict clinical utility from  
genetics of arthritis?



# can we predict scores on alzheimers cognitive tests from existing data?



alzheimer's  association™

the compassion to care, the leadership to conquer

## Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data

---

**Notice Number:** NOT-HG-10-006

### Key Dates

Release Date: October 19, 2009

### Issued by

National Institutes of Health (NIH), (<http://www.nih.gov>)

### Purpose

The purpose of this Notice is to inform the research community of plans by the National Institutes of Health (NIH) to:

1. Update data sharing policies for NIH supported research, including extramural and intramural projects, involving sequence and related genomic data obtained with advanced sequencing technology (*e.g.*, medical resequencing data, sequence data from non-human species, including microorganisms, transcriptomic and epigenomic data, as well as data needed for interpretation, including associated clinical, other phenotype and metadata, such as supporting study documents and methodologies);
2. Encourage investigators and IRBs to consider the potential for broad sharing of sequence and related genomic data in developing informed consent processes and documents for such studies involving human sequence data; and,
3. Communicate the agency's intent and current underlying considerations related to developing a policy pertaining to the deposition of these large datasets into centralized databases, such as the GenBank Short Read Archive (SRA) or the Database of Genotypes and Phenotypes (dbGaP), so that they are available as broadly and rapidly as possible to a wide range of scientific investigators.

### Need for Broad Data Sharing Policies

Rapid advances in applying genomic approaches to developing an understanding of the patterns of genetic and epigenetic variation, gene expression, and chromosomal organization are being made possible by maturing, more-effective methods and technologies for generating very large sequence data sets. These data sets are not only valuable for addressing the questions that the experiments were designed to ask, but also have



Home

## TCGA Data Portal Overview

We provide 3 ways to download data: The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high-throughput sequencing analysis of the tumor genomes.

The TCGA Data Portal does not host lower levels of sequence data. NCI's [Cancer Genomics Hub \(CGHub\)](#) is the new secure repository for storing, cataloging, and accessing BAM files and metadata for sequencing data. New users must still apply for authorized access through NCBI's [Database of Genotypes and Phenotypes \(dbGaP\)](#).

[Download Data](#) >

Choose from three ways to download data

Available Cancer Types	# Cases Shipped by BCR	# Cases with Data*	Date Last Updated (mm/dd/yy)
<a href="#">Acute Myeloid Leukemia [LAML]</a>	200	200	07/16/13
<a href="#">Adrenocortical carcinoma [ACC]</a>	80	80	07/22/13

### Announcements

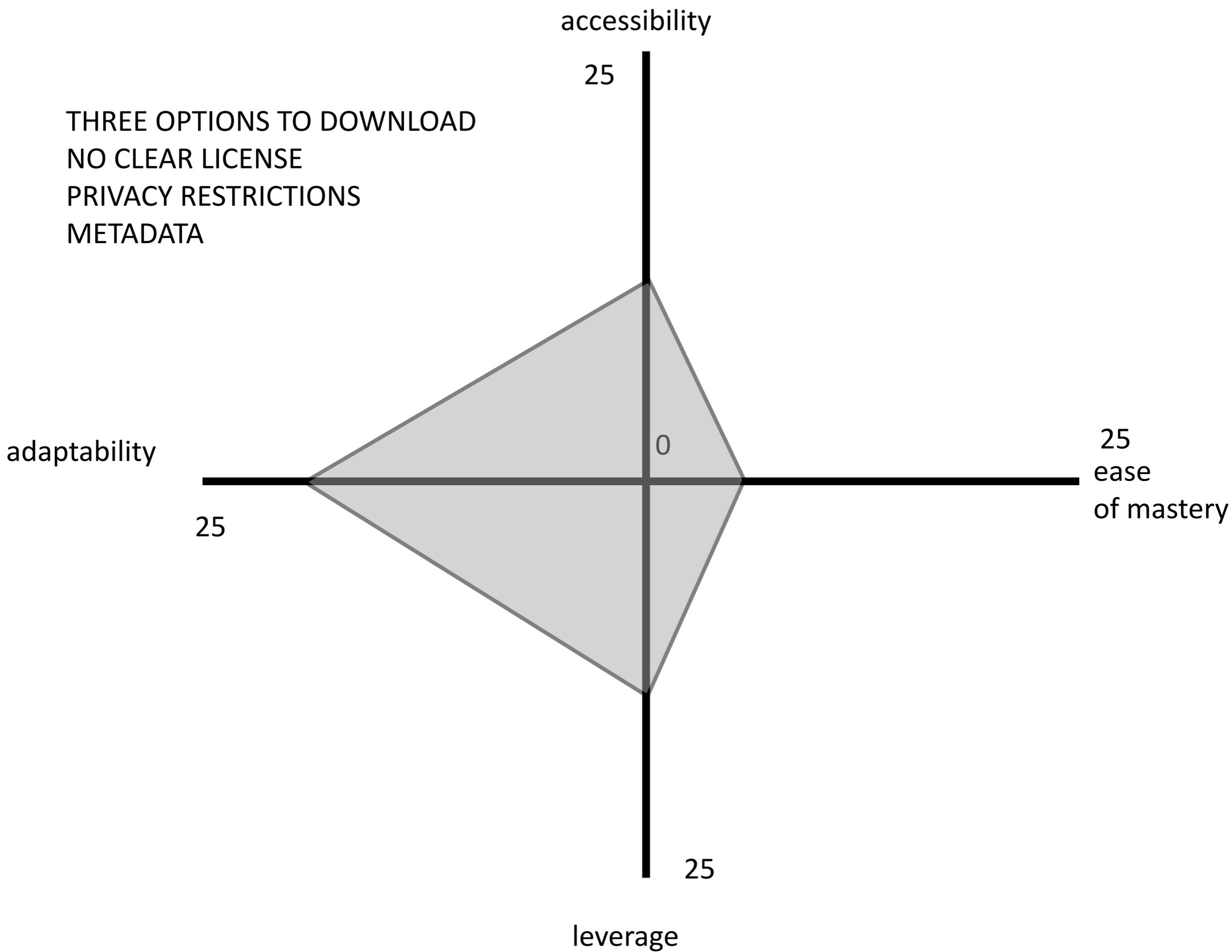
#### 06/13/2013 - DCC Software Released

The software release scheduled for today is complete and the TCGA Data Portal has been returned to normal operation. As part of this release, a new version of the TCGA Archive Validator has been provided and we strongly suggest that data submitting centers download and use this new version. The TCGA Archive Validator can be found on the TCGA Wiki [here](#).

A complete list of the items addressed in this release can be found on the TCGA Wiki [here](#) and for those with JIRA access the tickets covered in this release can be found on the wiki [here](#).

If you have any questions or concerns about this release, contact [tcga-dcc-binf-l@list.nih.gov](mailto:tcga-dcc-binf-l@list.nih.gov).

#### 05/21/2013 - DCC Software Released



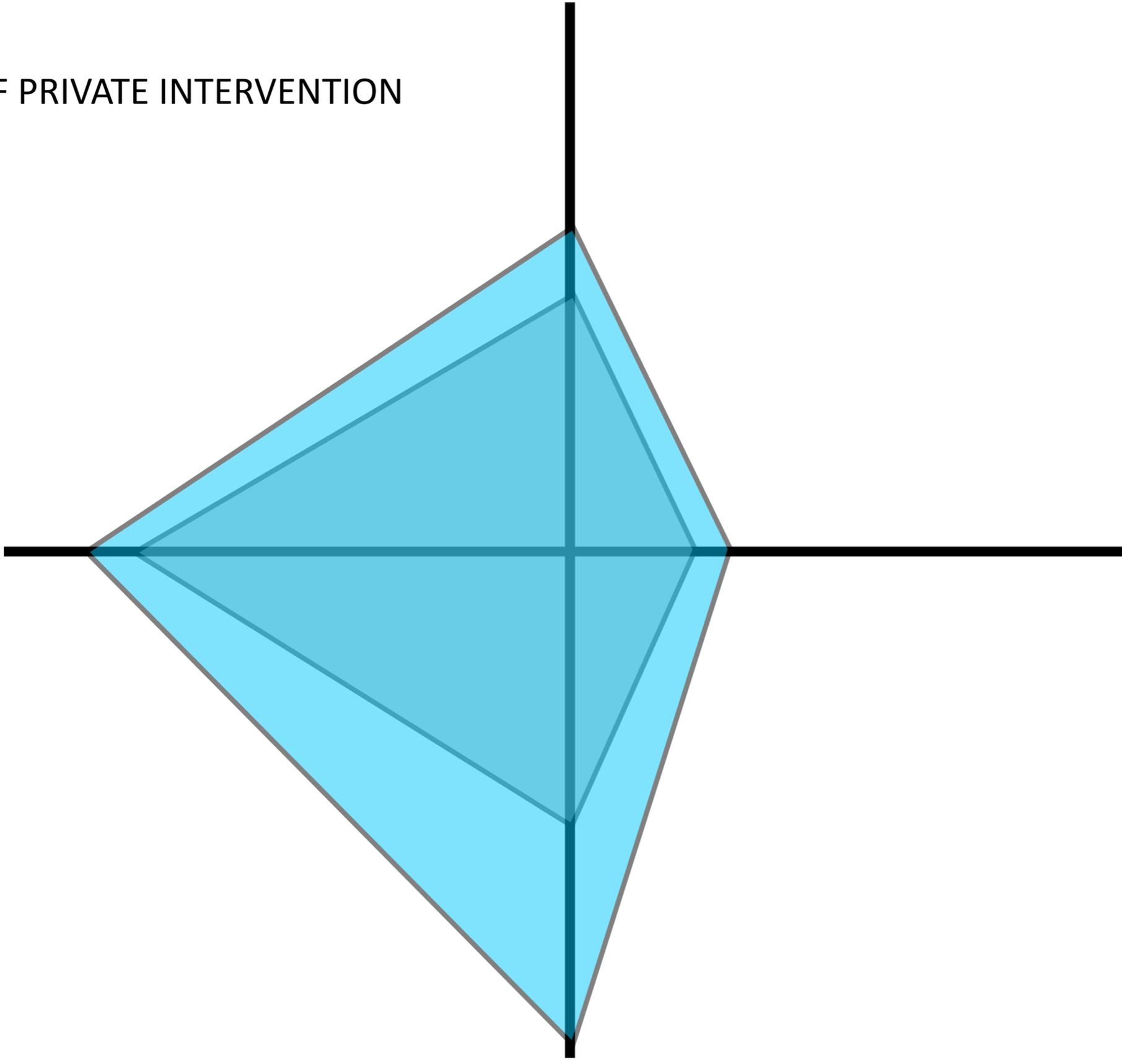
IMPACT OF PRIVATE INTERVENTION

accessibility

adaptability

ease  
of mastery

leverage



68

core projects

248  
researchers

28

institutions

1070

datasets

1723

results

# Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas

Larsson Omberg<sup>1,6</sup>, Kyle Ellrott<sup>2,6</sup>, Yuan Yuan<sup>3,4</sup>, Cyriac Kandoth<sup>5</sup>, Chris Wong<sup>2</sup>, Stephen H Friend<sup>1</sup>, Josh Stuart<sup>2</sup>, Han Liang<sup>3,4</sup> & Adam A Margolin<sup>1</sup>

The Cancer Genome Atlas Pan-Cancer Analysis Working Group collaborated through Synapse, a software platform, to share and evolve data, results and methodologies to perform integrative analysis of molecular profiling data from 12 tumor types. The group's work serves as a pilot case study that provides (i) a template for future large collaborative studies; (ii) a system to support collaborative projects; and (iii) a public resource of highly curated data, results and automated systems for the evaluation of community-developed models.

# colorectal cancer subtyping



## Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines

Andreas Schlicker<sup>1</sup>, Garry Beran<sup>3</sup>, Christine M Chresta<sup>3</sup>, Gael McWalter<sup>4</sup>, Alison Pritchard<sup>3</sup>, Susie Weston<sup>4</sup>, Sarah Runswick<sup>4</sup>, Sara Davenport<sup>3</sup>, Kerry Heathcote<sup>3</sup>, Denis Alferez Castro<sup>3</sup>, George Orphanides<sup>3</sup>, Tim French<sup>4\*</sup> and Lodewyk FA Wessels<sup>1,2,5\*</sup>



A colorectal cancer classification system that associates cellular phenotype and responses to therapy

Anguraj Sadanandam, Costas A Lyssiotis, Krisztian Homicsko, Eric A Collisson, William J Gibb, Stephan Wullschleger, Liliane C Gonzalez Ostos, William A Lannon, Carsten Grotzinger, Maguy Del Rio, Benoit Lhermitte, Adam B Olshen, Bertram Wiedenmann, Lewis C Cantley, Joe W Gray & Douglas Hanahan



## Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value

Laetitia Marisa, Aurélien de Reyniès, Alex Duval, Janick Selves, Marie Pierre Gaub, Laure Vescovo, Marie-Christine Etienne-Grimaldi, Renaud Schiappa, Dominique Guenot, Mira Ayadi, Sylvain Kirzin, Maurice Chazal, Jean-François Fléjou, Daniel Benchimol, Anne Berger, Arnaud Lagarde, Erwan Pencreach, Françoise Piard, Dominique Elias, Yann Parc, Sylviane Olschwang, Gérard Milano, Pierre Laurent-Puig , Valérie Boige [ [view less](#) ]



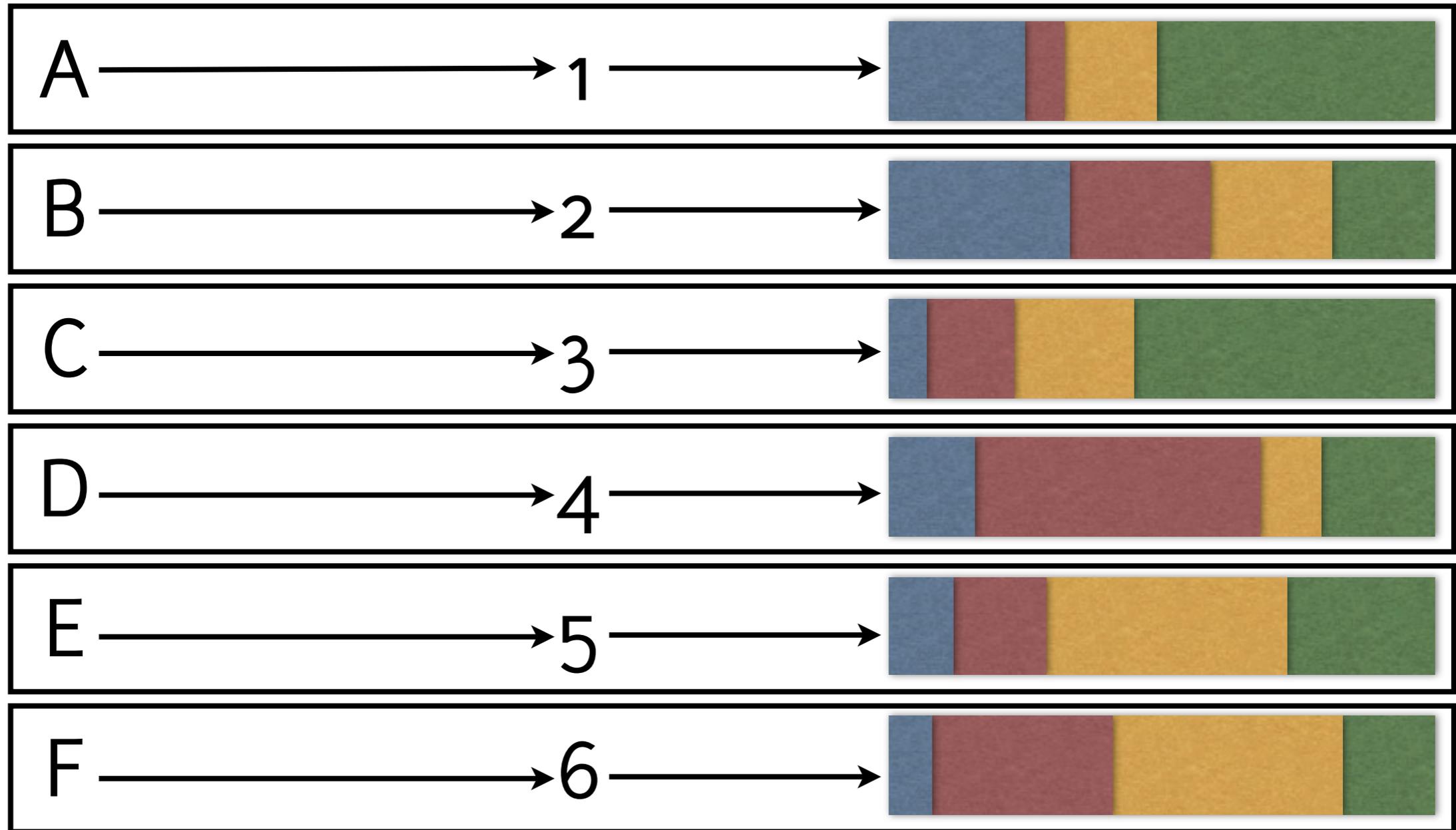
Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer

Eva Budinska,<sup>1,2\*</sup> Vlad Popovici,<sup>1,2</sup> Sabine Tejpar,<sup>3</sup> Giovanni D'Ario,<sup>1</sup> Nicolas Lapique,<sup>1</sup> Katarzyna Otylia Sikora,<sup>1</sup> Antonio Fabio Di Narzo,<sup>1</sup> Pu Yan,<sup>4</sup> John Graeme Hodgson,<sup>5</sup> Scott Weinrich,<sup>5</sup> Fred Bosman,<sup>5</sup> Arnaud Roth<sup>6,7</sup> and Mauro Delorenzi<sup>1,8</sup>

analysis  
groups

datasets

subtypes



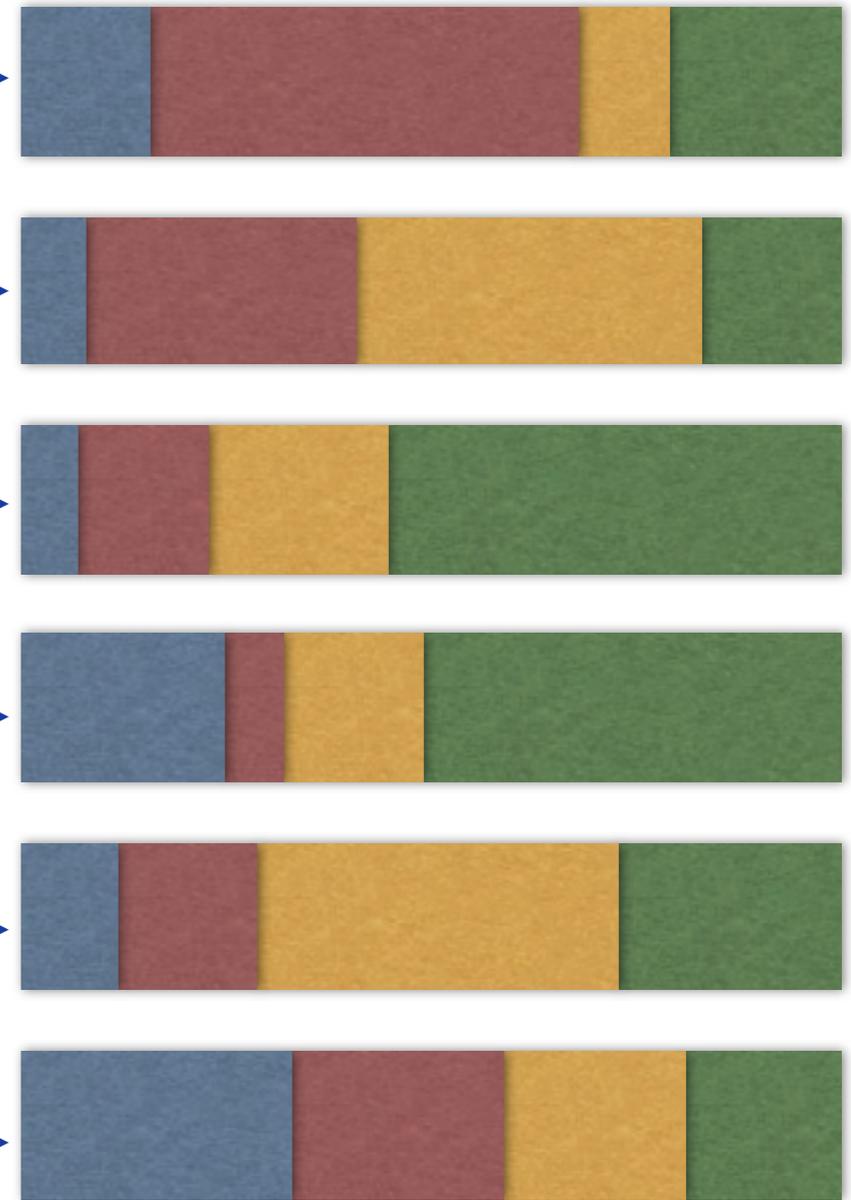
analysis  
groups

A  
B  
C  
D  
E  
F  
G

datasets

1  
2  
3  
4  
5  
6  
...

subtypes

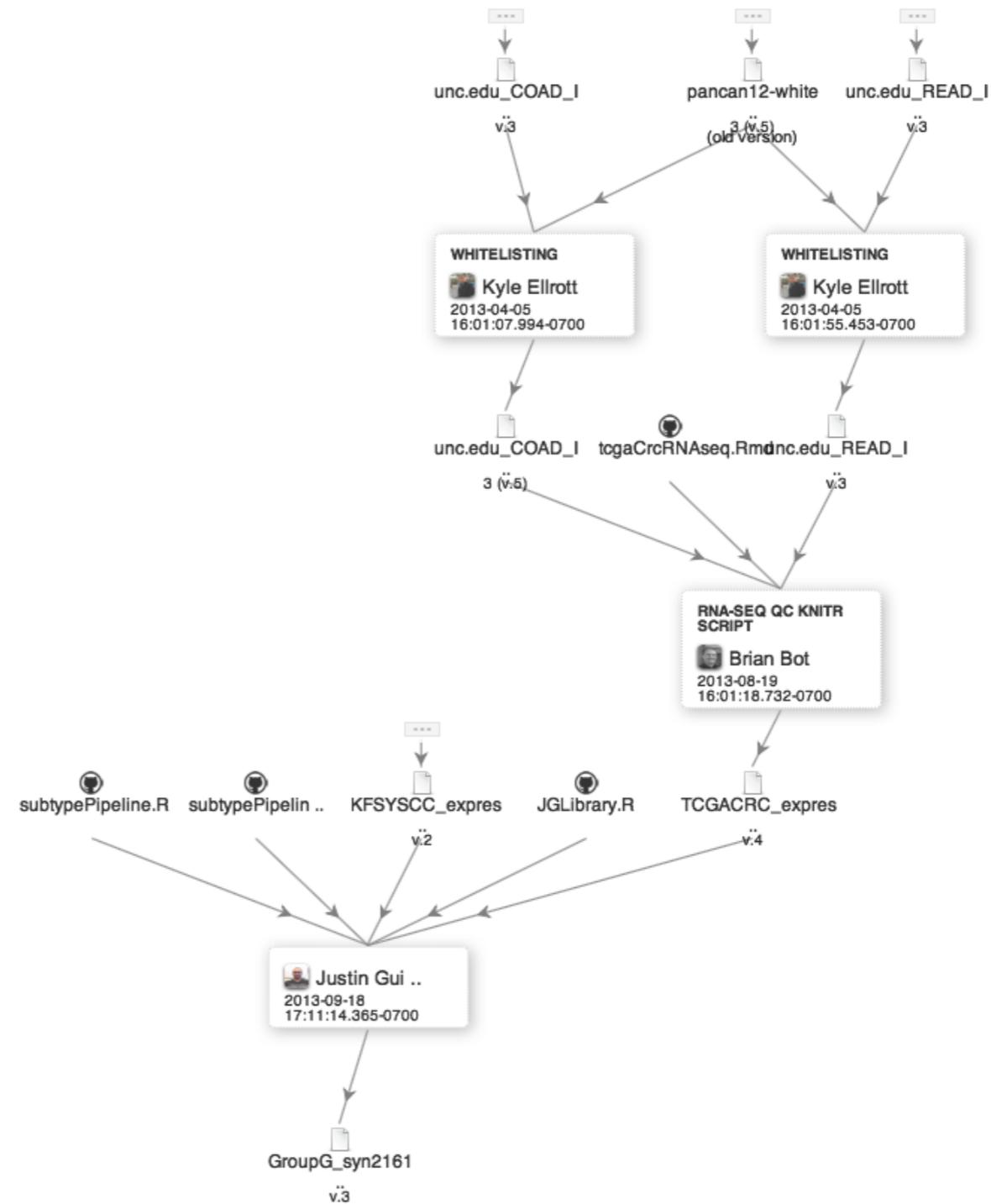


# analysis groups

## G

phasel

- ▶ GroupG\_syn2159423\_amc\_ajccii.tsv
- ▶ GroupG\_syn2161141\_tcga\_rnaseq.tsv
- ▶ GroupG\_syn2169565\_kfsyscc.tsv
- ▶ GroupG\_syn2171434\_french.tsv
- ▶ GroupG\_syn2175581\_petacc3.tsv
- ▶ GroupG\_syn2176657\_nki\_az.tsv
- ▶ GroupG\_syn2177169\_gse2109.tsv
- ▶ GroupG\_syn2177179\_gse4107.tsv
- ▶ GroupG\_syn2177187\_gse4183.tsv
- ▶ GroupG\_syn2177194\_gse10961.tsv
- ▶ GroupG\_syn2177199\_gse15960.tsv
- ▶ GroupG\_syn2177888\_gse13067.tsv
- ▶ GroupG\_syn2177894\_gse13294.tsv
- ▶ GroupG\_syn2177899\_gse20916.tsv
- ▶ GroupG\_syn2177902\_gse23878.tsv
- ▶ GroupG\_syn2178082\_gse37892.tsv
- ▶ GroupG\_syn2178128\_gse17537.tsv
- ▶ GroupG\_syn2181079\_gse14333.tsv
- ▶ GroupG\_syn2181088\_gse8671.tsv
- ▶ GroupG\_syn2192792\_agendia\_gse42284.tsv
- ▶ GroupG\_syn2192796\_agendia\_ico208.tsv
- ▶ GroupG\_syn2192799\_agendia\_vhb70.tsv



analysis  
groups

A  
B  
C  
D  
E  
F  
G

datasets

1  
2  
3  
4  
5  
6  
...

subtypes



3.

research and culture are  
on a collision course,  
driven by data.

tension between  
anonymity and utility.

July 10, 2011

# Harvard Researchers Accused of Breaching Students' Privacy

Social-network project shows promise and peril of doing social science online



Jason Kaufman  
Society, sa  
profiles are

[Enlarge Image](#)

*By Marc Parry*

In 2006, Harvard sociologists struck a mother lode of social-science data, offering a new way to answer big questions about how race and cultural tastes affect relationships.

The source: some 1,700 Facebook profiles,

“more like plutonium than gold”

the Harvard team began to realize that potential by publicly releasing part of its archive.

But today the data-sharing venture has collapsed. The Facebook archive is more like plutonium than gold—its contents yanked offline, its future release uncertain, its creators scolded by some scholars for downloading the profiles without students' knowledge and for failing to protect their privacy. Those students have been identified as Harvard College's Class of 2009.

The story of that collapse shines a light on emerging ethical challenges faced by scholars researching social networks and other online environments.

tension between  
expectation and reuse.

# JAMA Internal Medicine

Formerly Archives of Internal Medicine

[Home](#) [Current Issue](#) [All Issues](#) [Online First](#) [Topics](#) [CME](#) [Multimedia](#) [For Au](#)

[Online First >](#)

Full Content is available to subscribers  
[Subscribe/Learn More](#)

Original Investigation | August 19, 2013

## Public Preferences About Secondary Uses of Electronic Health Information **ONLINE FIRST**

David Grande, MD, MPA<sup>1,2</sup>; Nandita Mitra, PhD<sup>3</sup>; Anand Shah, MD, MSHP<sup>4</sup>; Fei Wan, MS<sup>3</sup>; David A. Asch, MD, MBA<sup>1,2,5</sup>

[\[+\] Author Affiliations](#)

JAMA Intern Med. Pub

Article

Figures

Ta

68% want their data shared for science

tension between value of  
individual and value of  
aggregate.

# How Much Are You Worth?

When you visit certain Web sites on the Internet, ad requests are sent to advertisers. They compete for a chance to serve ads to you. The bid prices they submitted to auctions are generally based on your information that advertisers possess, for example a profile inferred from your Web history, and your browsing context. The prices reflect how they evaluate your profile. We capture these prices to give you a quantification of your value from advertisers' perspective.

**Results:** We do not have any data associated with you. If you use Firefox or Chrome, please install the plugin and enable cookies. Note that our plugin does NOT work with Ad blocker extensions such as AdblockPlus or any addons of these types. If you use Ad blockers, and still want to know how advertisers estimate your private data, you have three options:

- Deactivate Ad blockers, browse the Internet as usual, and reactivate your Ad blockers whenever enough prices are collected (i.e. our plugin starts showing you your average price and the number of prices is larger than 20-40). This could take a couple of days or just one day, depending on your browsing habits.
- Deactivate your Ad blockers, click on some (10 to 15, with e.g. several refreshes) of the links [example from this list](#), and reactivate your Ad blockers later once you are done. This option is faster, but might somehow affect the results.
- For Firefox create a [new Firefox profile](#) and browse the Web as usual with our plugin installed. Of course do not use Ad Blockers. For Chrome, create a new browser user profile: [see here](#) or alternatively check [this tutorial](#).

We show below some general highlights, if you do not use Firefox or Chrome, or do not wish to install the plugin.

**The average price paid for user's private data (items in Web browsing history) for our users is:**

**\$0.000564**

## 25 Data Broker Giants Hacked by ID Theft Service

SEP 13

An identity theft service that sells Social Security numbers, birth records, credit and background reports on millions of Americans has infiltrated computers at some of America's largest consumer and business data aggregators, according to a seven-month investigation by KrebsOnSecurity.

The Web site **ssndob[dot]ms** (hereafter referred to simply as SSNDOB) has for the

past  
under  
reliab  
custo  
birthd

U.S. resident. Prices range from 50 cents to \$2.50 per record, and from \$5 to \$15 for credit and background checks. Customers pay for their subscriptions using largely unregulated and anonymous virtual currencies, such as Bitcoin and WebMoney.

Until very recently, the source of the data sold by SSNDOB has remained a mystery. That mystery began to unravel in March 2013, when teenage hackers allegedly associated with the hacktivist group **UGNazi** showed just how deeply the service's access went. The young hackers used SSNDOB to collect data for **exposed.su**, a Web site that listed the SSNs,



**\$.50 to \$2.50 for SSN,  
birthdate, etc.**

## 25 Data Broker Giants Hacked by ID Theft Service

SEP 13

An identity theft service that sells Social Security numbers, birth records, credit and background reports on millions of Americans has infiltrated computers at some of America's largest consumer and business data aggregators, according to a seven-month investigation by KrebsOnSecurity.

The Web site `ssndob[dot]ms` (hereafter referred to simply as SSNDOB) has for the

past  
under  
reliab  
custo  
birth

U.S. resident. Prices range from 50 cents to \$2.50 per record, and from \$5 to \$15 for credit and background checks. Customers pay for their subscriptions using largely unregulated and anonymous virtual currencies, such as Bitcoin and WebMoney.

Until very recently, the source of the data sold by SSNDOB has remained a mystery. That mystery began to unravel in March 2013, when teenage hackers allegedly associated with the hacktivist group **UGNazi** showed just how deeply the service's access went. The young hackers used SSNDOB to collect data for **exposed.su**, a Web site that listed the SSNs,



\$5 to \$15 for credit,  
background checks.

HOME

ABOUT

NEWS

MEET THE U.S. ATTORNEY

DIVISIONS

PROGRAMS

Home » News » Press Release

NEWS



FOR IMM  
Tuesday,

~40 records for \$2100

(202) 252-0933

<http://www.justice.gov/usao/dc/index.html>

**Former Howard University Hospital Employee Pleads Guilty  
to Selling Personal Information About Patients  
- Information Was Then Used on Forged Prescriptions for Oxycodone -**

WASHINGTON - Laurie Napper, a former medical technician at Howard University Hospital, pled guilty today to a federal charge stemming from the sale of personal information about patients, along with blank prescription forms, announced U.S. Attorney Ronald C. Machen Jr. and James W. McJunkin,

tension between  
“research” data and  
“consumer” data.



# disease risk

Share my health results with family and friends

My Home

Inbox (5)

### My Health

Disease Risk

Carrier Status

Drug Response

Traits

Health Labs

### My Ancestry

Maternal Line

Paternal Line

Relative Finder

Ancestry Painting

Global Similarity

Ancestry Labs

### Sharing & Community

Compare Genes

Family Inheritance

23andMe Community

Show results for

[See new and recently updated reports »](#)

23andMe Discoveries were made possible by 23andMe members who took surveys.

### Elevated Risk ?

Name	Confidence	Your Risk	Avg. Risk	Compared to Average
Prostate Cancer ♂	★★★★★	31.9%	17.8%	1.79x
Psoriasis	★★★★★	22.4%	11.4%	1.98x
Alzheimer's Disease	★★★★★	14.2%	7.2%	1.98x
Ankylosing Spondylitis	★★★			↑
Asthma	★★★			↑
Bipolar Disorder: Preliminary Research	★★★			↑
Chronic Lymphocytic Leukemia	★★★			↑
Follicular Lymphoma	★★★			↑
High Blood Pressure (Hypertension)	★★★			↑



## BLUE BUTTON

- [Blue Button Home](#)
- [Getting Started](#)
- [License the Blue Button®](#)
- [Blue Button Partners](#)
- [Resources](#)
- [VA Homepage](#)
- [Site Search](#)



## Blue Button Download My Data



### Need to Know: Update for Blue Button Partners

The rollout of the upgraded Blue Button on September 29 and October 9 means enhancements for our Veterans who download their health care data from VA's electronic health records. For our industry partners it means enhancements and improvements for their supporting software's functionality.

Based on Veteran user feedback, the minor enhancements in this rollout will make the output more readable and user friendly. There is also a new set of data for 'VA Immunizations,' enhancements to the online user interface and improved data from the Military Service Information (MSI) section. We have provided these instructions on our [Resources page](#) along with updated sample files.

[One Million Blue Button Users - Veterans Talk to Markle](#)



Manage Your VA Health Care Online

### Blue Button Partners

Industry is building applications to help Veterans use the Blue Button to better manage their health or find employment. See which companies have committed to supporting the Blue Button Initiative on our list of [Blue Button partners](#).

If your organization has created an application for Blue Button and would like to be added to this list, email us at [VABlueButton@va.gov](mailto:VABlueButton@va.gov).

*DISCLAIMER: VA does not endorse these products or companies.*

# The Scientific Services Marketplace

The easiest way to get experiments conducted by researchers in top core facilities and institutions.

**RNA microarray** **\$107.50**  
per Sample

**DNA Sequencing** **\$2.50**  
per Sample

**Real Time qPCR** **\$3.50**  
per Sample

**Mass Spectrometry** **\$10.00**  
per Sample

**Immunohistochemistry** **\$10.00**  
per Sample

**Bioinformatics** **\$50.00**  
per Hour

## Featured RNA microarray Providers

Science Exchange has 63 verified RNA microarray providers including the following featured providers.



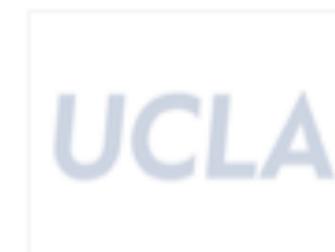
### Virginia Bioinformatics Institute

Virginia Polytechnic Institute and State University | Blacksburg, VA, United States

The Core Laboratory Facility (CLF) at VBI functions as a multi-user resource dedicated to the development and application of various high-throughput technologies to aid in the discovery of biological macromolecules.

**\$284.00 USD** per Sample

[REQUEST ESTIMATE](#)



<https://www.scienceexchange.com>

 [View all providers for this experiment →](#)

 [Request estimates from all providers →](#)







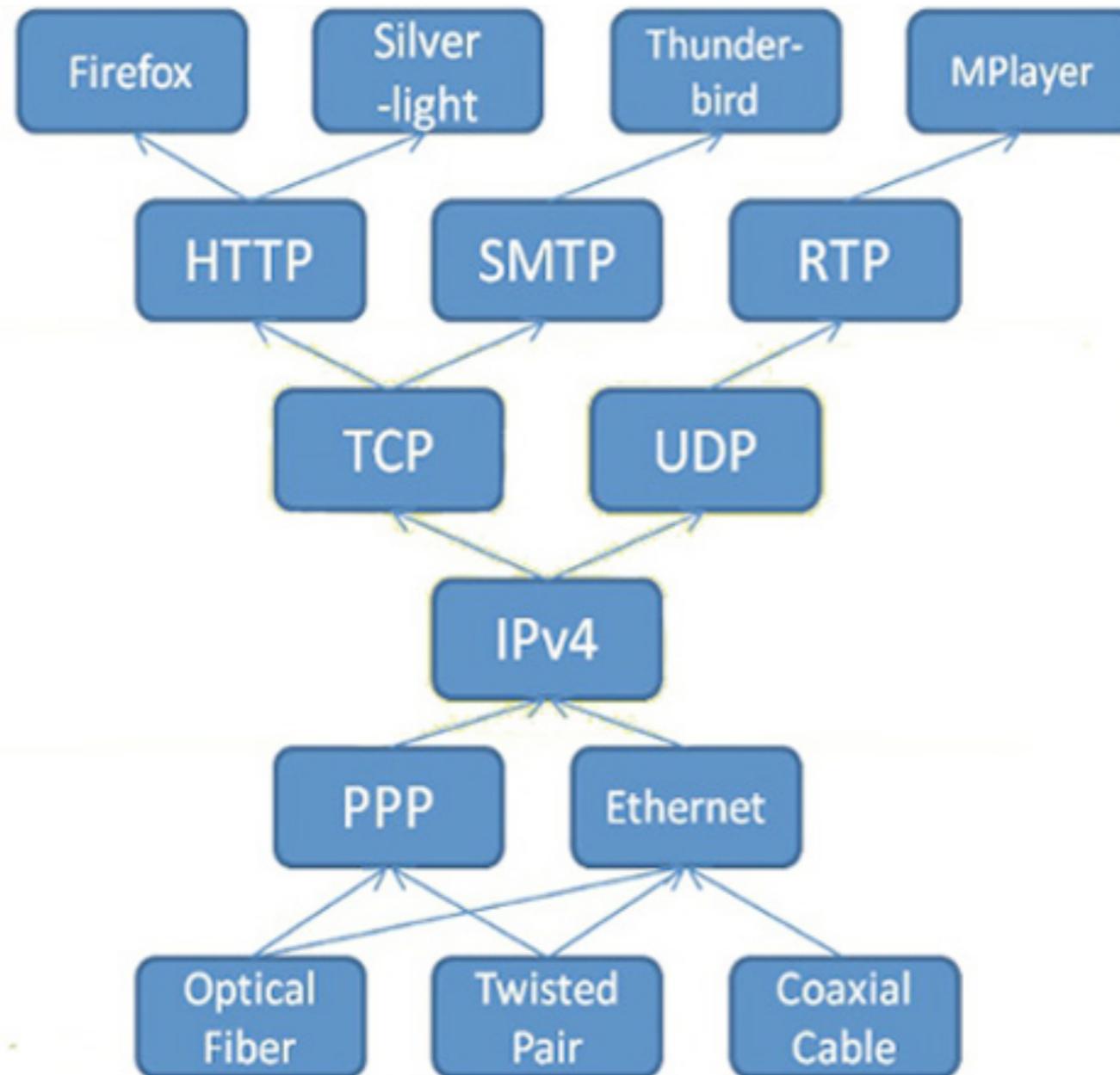


it's likely that we will end  
up with a data network  
effect of some sort.

a. the incremental  
institution.

b. the walled garden.

c. big networks of  
small things.



thank you

@wilbanks  
wilbanks@nitrd.gov