Broad GDAC Pipeline Status
2011_05_25 Run

June 2, 2011
Michael S. Noble
mnoble@broadinstitute.org

May 25, 2011 Run Status

- Two weeks since last run uploaded to DCC

- Still executing: duplicate samples, manual MAF tweaks, internal wiring

- ALSO: bigger data == longer runtimes AND debug cycles

- INTEGRATION workspace: for ironing out kinks, pre-production

  - Exists, but not fully operational

- TCGA-wide buy-in established for version-stamped normalized data

- DATA workspace: for versioned/normed data

  - Also exists, but not fully operational

Summary of TCGA Tumor Data
Ingested into Broad GDAC Pipeline
2011_05_25 Run

TumorType	Biospecimen	Any_Level_1	Clinical	CNA	Methylation	mRNA	miR	MAF
BLCA	35	12	11	9	0	0	0	0
BRCA	704	524	358	507	186	434	0	0
CESC	40	8	5	8	0	0	0	0
COAD	245	202	208	186	167	155	0	102
COADREAD	338	276	287	257	236	224	0	158
GBM	547	511	465	498	288	499	415	199
HNSC	97	59	0	57	0	0	0	0
KIRC	460	453	241	448	219	72	0	0
KIRP	75	16	17	16	36	41	0	0
LAML	202	0	0	0	188	0	178	135
LGG	58	30	19	30	0	0	0	0
LIHC	45	38	0	37	0	0	0	0
LUAD	158	59	47	58	128	33	0	122
LUSC	184	184	72	142	133	134	0	150
OV	592	570	528	519	425	570	566	383
PRAD	65	65	0	64	0	0	0	0
READ	93	74	79	71	69	69	0	56
STAD	111	35	0	81	82	0	0	0
THCA	39	25	0	24	0	0	0	0
UCEC	325	220	127	215	70	0	0	0
Totals	4075	3085	2177	2970	1991	2007	1159	1147
	+222	+738	+58	+486	+0	-7	+0	+291

Website Evolution

This is the online home of the Broad Institute's Genome Data Analysis Center (GDAC). A collaborative effort between cancer researchers, computational biologists and software engineers, under the auspices of The Cancer Genome Atlas (TCGA), we operate a scientific pipeline designed to coordinate the flow of massive, terabyte-scale genomic datasets through scores of quantitative algorithms, in the hope of accelerating discovery of cancer mechanisms and the development of therapies. The current set of data loaded into Firehose is given below, and here is the status of the analyses performed upon these data.

- Data Ingestion
- Documents
- Pipeline Status

Email Archive
Contact Us

Summary of TCGA Tumor Data
Ingested into Broad GDAC Pipeline
04/21/2011 Run

TumorType	Biospecimen	Any_Level_1	Clinical	CNA	Methylation	mRNA	miR	MAF
BLCA	26	12	9	9	0	0	0	0
BRCA	647	390	353	375	186	434	0	0
CESC	23	8	5	8	0	0	0	0
COAD	245	151	207	182	167	155	0	88
COADREAD	338	203	285	253	236	224	0	139
GBM	508	476	465	466	288	506	415	199
HNSC	59	59	0	57	0	0	0	0
KIRC	460	347	192	345	219	72	0	0
KIRP	75	16	17	16	36	41	0	0
LAML	202	0	0	0	188	0	178	135
LGG	30	0	19	0	0	0	0	0
LIHC	38	0	0	0	0	0	0	0
LUAD	158	21	47	56	128	33	0	0
LUSC	184	161	72	142	133	134	0	0
OV	592	570	528	519	425	570	566	383
PRAD	65	0	0	0	0	0	0	0
READ	93	52	78	71	69	69	0	51
STAD	111	35	0	81	82	0	0	0
THCA	39	25	0	24	0	0	0	0
UCEC	298	24	127	133	70	0	0	0
Totals	3853	2347	2119	2484	1991	2014	1159	856

gdac.broadinstitute.org

FIREHOSE
Broad GDAC

Mail

| Pages | Blog | Labels | Attachments | Bookmarks | **Mail** | Forums | Advanced |

☐ Fetch new mail

1 2   Next >>

Sender	Subject	Date
Michael S. Noble	✉ Re: [Gdac-users] Latest Run	May 27, 2011
Michael S. Noble	✉ Re: [Gdac-users] We'd like to add a module to Firehose	May 26, 2011
Thomas Robinson	✉ Re: [Gdac-users] In re, Firehose analysis of DCC data	May 26, 2011
mark.jensen at NIH.GOV (Jensen, Mark (NIH/NCI) [C])	✉ [Gdac-users] Patched 2011.04.21 Broad pipeline data available	May 24, 2011
mnoble at broadinstitute.org (Michael S. Noble)	✉ [Gdac-users] firehose outputs	May 24, 2011
Sheila.Reynolds at systemsbiology.org (Sheila Reynolds)	✉ [Gdac-users] firehose outputs	May 24, 2011
gsaksena at broadinstitute.org (Gordon Saksena)	✉ [Gdac-users] In re, Firehose analysis of DCC data	May 24, 2011
gsaksena at broadinstitute.org (Gordon Saksena)	✉ [Gdac-users] DCC annotations?	May 23, 2011
mnoble at broadinstitute.org (Michael S. Noble)	✉ [Gdac-users] DCC annotations?	May 23, 2011
nils at hms.harvard.edu (Gehlenborg, Nils)	✉ [Gdac-users] DCC annotations?	May 23, 2011
gsaksena at broadinstitute.org (Gordon Saksena)	✉ [Gdac-users] DCC annotations?	May 23, 2011
nils at hms.harvard.edu (Gehlenborg, Nils)	✉ [Gdac-users] DCC annotations?	May 23, 2011
RYao at mdanderson.org (Yao,Rong)	✉ [Gdac-users] question regarding Ovarian Copy Nimber data from Broad GDAC run	May 23, 2011
mnoble at broadinstitute.org (Michael S. Noble)	✉ [Gdac-users] question regarding Ovarian Copy Nimber data from Broad GDAC run	May 23, 2011
RYao at mdanderson.org (Yao,Rong)	✉ [Gdac-users] question regarding Ovarian Copy Nimber data from Broad GDAC run	May 23, 2011
Jianhua_Zhang at dfci.harvard.edu (Jianhua Zhang)	✉ [Gdac-users] We'd like to add a module to Firehose	May 23, 2011
schum at broadinstitute.org (Steven Schumacher)	✉ [Gdac-users] question regarding Ovarian Copy Nimber data from Broad GDAC run	May 23, 2011
Adam.Norberg at systemsbiology.org (Adam Norberg)	✉ [Gdac-users] We'd like to add a module to Firehose	May 20, 2011

Documentation repository

Browse ▾   Log In   Search

FIREHOSE
Broad GDAC

Documents

⚙ Tools ▾

📎5 Added by Michael Noble, last edited by Michael Noble on May 27, 2011 (view change)

Firehose_Archive_Name_Convention

Pipeline_Overview_Apr_2011

Pipeline_Overview_Nov_2010

Run_Status_2011_02_17

Run_Status_2011_03_27

FIREHOSE
Broad GDAC

Pipeline Status

Added by Michael Noble, last edited by Michael Noble on May 14, 2011 (view change)

Input data and analysis results for these runs are available at the TCGA Data Coordination Center (DCC), and require appropriate login

Previous Runs

March 27, 2011

February 17, 2011

Broad GDAC Analysis Summary
2011_04_21 Run

Tables of Ingested Data: HTML   PNG   TSV

Tumor Type	# Completed	Percentage
OV	24	100%
GBM	24	100%
READ	17	71%
COAD	17	71%
COADREAD	17	71%
LUSC	12	50%
LUAD	12	50%
BRCA	12	50%
KIRC	10	42%
KIRP	9	38%
UCEC	4	17%
CESC	4	17%
BLCA	4	17%
STAD	3	13%
HNSC	3	13%
THCA	2	8%
LAML	2	8%
LGG	1	4%
PRAD	0	0%
LIHC	0	0%