

Broad GDAC Pipeline Run Status March 2011

April 7, 2011
Michael S. Noble
mnoles@broadinstitute.org

**Summary of TCGA Tumor Data
Ingested into Broad GDAC Pipeline
03/27/2011 Run**

TumorType	Biospecimen	Any_Level_1	Clinical	CNA	Methylation	mRNA	miR	MAF
BLCA	12	0	9	0	0	0	0	0
BRCA	534	390	353	375	186	434	0	0
CECSC	8	0	4	0	0	0	0	0
COAD	203	151	187	182	167	155	0	64
COADREAD	282	203	265	257	236	224	0	77
GBM	508	448	465	466	288	506	415	199
HNSC	59	39	0	37	0	0	0	0
KIRC	460	39	66	254	219	72	0	0
KIRP	75	39	15	16	36	41	0	0
LAML	202	0	0	0	188	0	178	135
LGG	30	0	0	0	0	0	0	0
LIHC	38	0	0	0	0	0	0	0
LUAD	132	21	47	56	128	33	0	0
LUSC	161	116	72	142	133	134	0	0
OV	592	570	528	519	425	570	566	384
PRAD	65	0	0	0	0	0	0	0
READ	79	52	78	71	69	69	0	13
STAD	82	35	0	81	82	0	0	0
THCA	25	0	0	0	0	0	0	0
UCEC	269	24	121	133	70	0	0	0
FULL	3534 +465	1924 +243	1945 +289	2332 +256	1991	2014 +205	1159 +178	795

Summary Data Available Online

- gdac.broadinstitute.org/runs/latest/ingested_data.png
- gdac.broadinstitute.org/runs/latest/ingested_data.tsv
- gdac.broadinstitute.org/runs/2011_03_27/...
- gdac.broadinstitute.org/runs/2011_02_17/...

gdac.broadintitute.org



Edit Add Tools

1 Added by [Aaron Ball](#), last edited by [Michael Noble](#) on Apr 07, 2011 ([view change](#))

This is 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.

Summary of TCGA Tumor Data
Ingested into Broad GDAC Pipeline
03/27/2011 Run

TumorType	Biospecimen	Any_Level_1	Clinical	CNA	Methylation	mRNA	miR	MAF
BLCA	12	0	9	0	0	0	0	0
BRCA	534	390	353	375	186	434	0	0
CESC	8	0	4	0	0	0	0	0
COAD	203	151	187	182	167	155	0	64
COADREAD	282	203	265	257	236	224	0	77
GBM	508	448	465	466	288	506	415	199
HNSC	59	39	0	37	0	0	0	0
KIRC	460	39	66	254	219	72	0	0
KIRP	75	39	15	16	36	41	0	0
LAML	202	0	0	0	188	0	178	135
LGG	30	0	0	0	0	0	0	0
LIHC	38	0	0	0	0	0	0	0
LUAD	132	21	47	56	128	33	0	0
LUSC	161	116	72	142	133	134	0	0
OV	592	570	528	519	425	570	566	384
PRAD	65	0	0	0	0	0	0	0
READ	79	52	78	71	69	69	0	13
STAD	82	35	0	81	82	0	0	0
THCA	25	0	0	0	0	0	0	0
UCEC	269	24	121	133	70	0	0	0
FULL	3534	1924	1945	2332	1991	2014	1159	795

Added by [Michael Noble](#), last edited by [Michael Noble](#) on Apr 07, 2011 ([view change](#))

Select a date below to view the status of older Broad GDAC analysis pipeline runs. Input data and analysis results for these runs are available at the [TCGA Data Coordination Center \(DCC\)](#) (PI credentials may be required).

Broad GDAC Analysis Summary 2011_03_27 Run

Tumor Type	# Completed	Percentage
OV	26	<u>100%</u>
GBM	26	<u>100%</u>
FULL	23	<u>88%</u>
READ	16	<u>62%</u>
COAD	16	<u>62%</u>
COADREAD	16	<u>62%</u>
LUSC	15	<u>58%</u>
LUAD	14	<u>54%</u>
KIRC	14	<u>54%</u>
BRCA	14	<u>54%</u>
KIRP	8	<u>31%</u>
UCEC	4	<u>15%</u>
STAD	3	<u>12%</u>
HNSC	3	<u>12%</u>
LAML	2	<u>8%</u>
CESC	1	<u>4%</u>
BLCA	1	<u>4%</u>
THCA	0	<u>0%</u>
PRAD	0	<u>0%</u>
LIHC	0	<u>0%</u>
LGG	0	<u>0%</u>



[February 17, 2011](#)

Broad TCGA GDAC Analyses Status
Tumor Type: GBM
March 27, 2011

	Pipeline	Not Ready	Failed	Succeed
1	Clinical_Aggregate_Tier1	0	0	1
2	Clinical_Pick_Tier1	0	0	1
3	CopyNumber_GeneBySample	0	0	1
4	CopyNumber_Gistic2	0	0	1
5	CopyNumber_Preprocess	0	0	1
6	Correlate_Clinical_vs_miR	0	0	1
7	Correlate_Clinical_vs_miR_Clusters_CNMF	0	0	1
8	Correlate_Clinical_vs_miR_Clusters_Consensus	0	0	1
9	Correlate_Clinical_vs_mRNA	0	0	1
10	Correlate_Clinical_vs_mRNA_Clusters_CNMF	0	0	1
11	Correlate_Clinical_vs_mRNA_Clusters_Consensus	0	0	1
12	Correlate_Clinical_vs_Mutation	0	0	1
13	Correlate_CopyNumber_vs_miR	0	0	1
14	Correlate_CopyNumber_vs_mRNA	0	0	1
15	Correlate_GenomicEvents	0	0	1
16	Correlate_Methylation_vs_mRNA	0	0	1
17	miR_Clustering_CNMF	0	0	1
18	miR_Clustering_Consensus	0	0	1
19	miR_FindDirectTargets	0	0	1
20	mRNA_Clustering_CNMF	0	0	1
21	mRNA_Clustering_Consensus	0	0	1
22	mRNA_Preprocess_Median	0	0	1
23	Mutation_Assessor	0	0	1
24	Mutation_Significance	0	0	1
25	Pathway_FindEnrichedGenes	0	0	1
26	Pathway_Paradigm	0	0	1
	Total	0	0	26

Broad TCGA GDAC Analyses Status

Tumor Type: FULL

March 27, 2011

	Pipeline	Not Ready	Failed	Succeed
1	Clinical_Aggregate_Tier1	0	0	1
2	Clinical_Pick_Tier1	0	0	1
3	CopyNumber_GeneBySample	0	0	1
4	CopyNumber_Gistic2	0	0	1
5	CopyNumber_Preprocess	0	0	1
6	Correlate_Clinical_vs_miR	0	0	1
7	Correlate_Clinical_vs_miR_Clusters_CNMF	0	0	1
8	Correlate_Clinical_vs_miR_Clusters_Consensus	0	0	1
9	Correlate_Clinical_vs_mRNA	0	1	0
10	Correlate_Clinical_vs_mRNA_Clusters_CNMF	1	0	0
11	Correlate_Clinical_vs_mRNA_Clusters_Consensus	0	0	1
12	Correlate_Clinical_vs_Mutation	0	0	1
13	Correlate_CopyNumber_vs_miR	0	0	1
14	Correlate_CopyNumber_vs_mRNA	0	0	1
15	Correlate_GenomicEvents	0	0	1
16	Correlate_Methylation_vs_mRNA	0	0	1
17	miR_Clustering_CNMF	0	0	1
18	miR_Clustering_Consensus	0	0	1
19	miR_FindDirectTargets	0	0	1
20	mRNA_Clustering_CNMF	0	0	0
21	mRNA_Clustering_Consensus	0	0	1
22	mRNA_Preprocess_Median	0	0	1
23	Mutation_Assessor	0	0	1
24	Mutation_Significance	0	0	1
25	Pathway_FindEnrichedGenes	0	0	1
26	Pathway_Paradigm	0	0	1
	Total	1	1	23

News

- New Pipe: **Correlate_GenomicEvents**
- Run Schedule : shifting to twice monthly
- MAFs omitted for GBM, COAD, READ, COADREAD, and FULL
 - Pending contamination assessment & other validation by AWG
- 2 new SWE hires devoted 100% to engineering pipeline

Notes cont ...

- Upload to DCC in pending: passwords, passwords ...
- Multiple versions of many pipe results this month: **use latest**
- Available from **[tinyurl.com/tcga-gdac-broad/\[TUMOR\]/2011021700](http://tinyurl.com/tcga-gdac-broad/[TUMOR]/2011021700)**
- 2 new SWE hires devoted 100% to engineering pipeline

Questions / Comments: gdac@broadinstitute.org

Archive Name Generalization

<Domain>_<TumorType>.<GDAC_Pipeline>.<DataLevel>.<Runcode>.<Revision>.0



<Domain>_<IndivSet>.<GDAC_Pipeline>.<DataLevel>.<Runcode>.<Revision>.0

Accommodates:

- Individual tumor types (OV, GBM, ...)
- Aggregates : COADREAD
- Subsets

Essentially no difference for this month (but UPPERCASE)