



- GAP Informatics

Platforms for Human Genetics: From Samples to Sequences

George Grant
Oct. 7, 2010

GAP Informatics

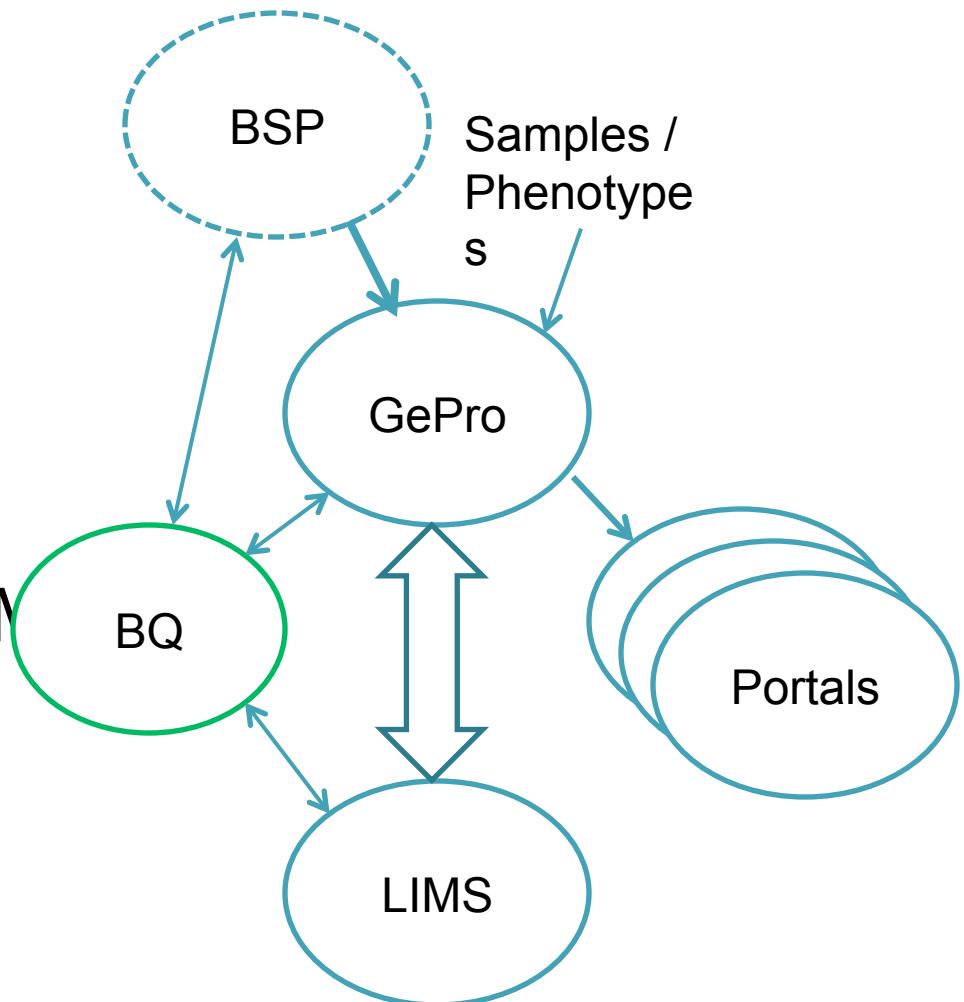
Enable the generation of high-quality Genotyping and Expression data on the following platforms:

- Genotyping
 - Sequenom
 - Affymetrix
 - Illumina Infinium
 - Illumina Golden Gate
 - Illumina BeadXpress
- Expression
 - Affymetrix
 - Illumina
 - Luminex
 - *Nanostring*

Make the data generated by these platforms easily accessible

GAP Applications / Data Flow

- Broad Quotes
- Project Management (GePro)
- Process Dashboard (LIM)
- Portals
 - Analysis Portal
 - GTEx Portal
 - ...



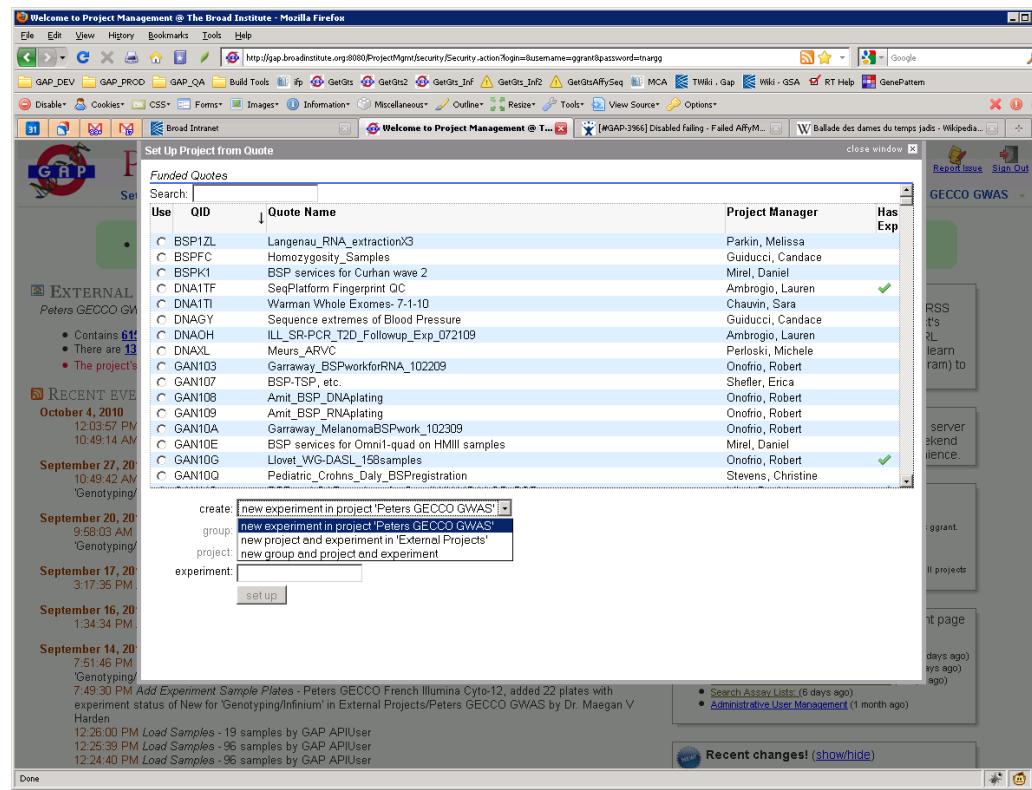
Data Flow: It starts with a Quote

The screenshot displays three sequential pages from the BioMart system:

- Import Price List:** A form for importing price lists. It shows the date as Monday, October 27, 2008 8:33:09 AM EDT, and the file 'SSP.PriceList' is selected. Buttons include 'Choose File', 'Submit', and 'Cancel'.
- Price List:** A table of items. The columns are Platform, Category, Price Item, Price, Effective Date, Source, Unit, Description, and Expires. Items listed include various DNA extraction kits and sample kits, such as 'Compressed Frozen Blood Additive' and 'Sample Kit - 48 Samples'. The table has 13 rows.
- Quote Details:** A detailed quote for a project named 'Pan_SC_mouse ILMN WG-6'. The quote includes:
 - Project Manager:** Jennifer Moran
 - Investigator:** Jen Pan
 - Finance Contact:** Lina Hristova
 - Description:** Pan ILMN WG-6
 - Identifier:** GX94T
 - Note:** 12 samples to be profiled on Illumina WG-6 beadchips.
 - Status:** Terminated
 - Platform:** Gene Expression
 - Total:** \$4,680.00The quote also notes that the cost of the project scope is \$4,055.00 total. It includes a note about the breakdown of costs and a request for a signature on the attached spreadsheet.

A Quote is needed in GAP

- GePro manages Groups, Projects, Users and Security
- The Experiment defines a chunk of work to be done in GAP
- Can be initiated from Quote Server:



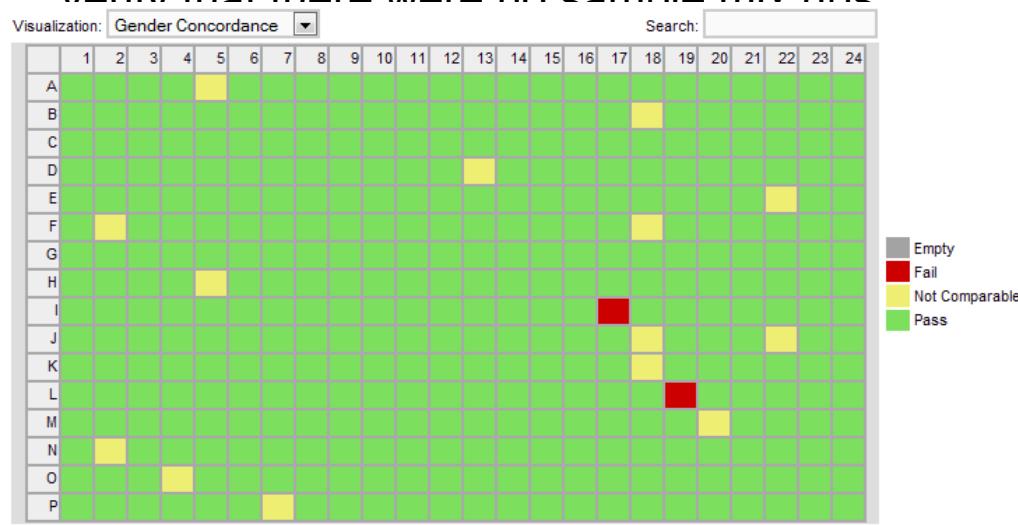
GAP ‘Experiment’

- A GAP Experiment associates:
 - Samples
 - Plates – loaded directly into GAP from BSP
 - Target (GAP) platform
 - Funding (Quote)
 - Assays (SNPs or Genes)
- The Experiment is handed off to the LIMS as a unit of work.
- Consider two Experiments in a typical GAP Project
 1. A Fingerprinting Experiment
 2. A WGA Experiment (Illumina Infinium)

GAP Fingerprinting Experiment

Fingerprint experiments are run on the

- Fingerprint experiments are run on the Sequenom Platform
 - Use one of several special pools designed for overlap with our Whole Genome products or for sequencing
 - i.e. The ILMN_CYTO12_FP pool shares 24 SNPs with the Illumina Cyto12 chip
 - All include gender assay
 - Sample plates are run on Sequenom using pool and curated
 - Upon curation they are flagged as fingerprint genotypes
 - The fingerprint gender is compared with the reported gender to verify that there were no sample mix ups





A GAP Infinium Experiment

- A similar process is used in Ge/Pro to define an Illumina Infinium Experiment
 - Sample plates in the experiment are run through the LIMS workflows
 - Scanned on the Illumina iScan
 - Once scanned, LIMS pulls data files and calls genotypes
 - Default caller is ‘Autocall’ – Illumina-supplied genotype caller
- QC run on the scans:
 - Call Rate, Heterozygosity
 - Concordance
 - Fingerprint Concordance
 - Verify that the SNPs on the Infinium Chip that overlap with the fingerprint panel are concordant
 - HapMap Concordance
 - Typically load one HapMap II Control sample on each plate
 - Gender Concordance
- The Experiment Summary page is one way to determine the success of the experiment

Infinium Experiment Summary

Summary of Experiment '

Illumina Cyto-12'

Illumina Infinium Genotyping:																										
Chemistry Plates: 10 (960 wells) - from 10 distinct DNA Plates (total of 10 DNA Plates in experiment)																										
Samples: 940 samples (20 empties) 933 unique clinical samples 933 unique individuals [489 Male, 444 Female, 0 Unknown]																										
Control Samples: 10 control (HapMap) samples 3 unique control clinical samples from 3 unique control individuals																										
Chips: 79																										
Scan Summary: <table border="1"> <thead> <tr> <th>Summary</th> <th>Total</th> <th>Scannable</th> <th>Scanned</th> <th>Pass QC Call Rate ≥ 97.0%</th> </tr> </thead> <tbody> <tr> <td>Samples:</td> <td>940</td> <td>100.0% (940 of 940)</td> <td>100.0% (940 of 940)</td> <td>99.4% (934 of 940)</td> </tr> <tr> <td>Call Rates:</td> <td></td> <td></td> <td>99.21% ± 0.68</td> <td>99.24% ± 0.37</td> </tr> </tbody> </table>												Summary	Total	Scannable	Scanned	Pass QC Call Rate ≥ 97.0%	Samples:	940	100.0% (940 of 940)	100.0% (940 of 940)	99.4% (934 of 940)	Call Rates:			99.21% ± 0.68	99.24% ± 0.37
Summary	Total	Scannable	Scanned	Pass QC Call Rate ≥ 97.0%																						
Samples:	940	100.0% (940 of 940)	100.0% (940 of 940)	99.4% (934 of 940)																						
Call Rates:			99.21% ± 0.68	99.24% ± 0.37																						
Gender Concordance: X Failed (4 of the 887 gender-comparable samples had discordant genders (discordance between infinium, fingerprint, or reported genders). Only 887 of 940 samples were gender-comparable (had infinium, fingerprint and reported genders))																										
Fingerprint Concordance: X Failed QC (3 of 884 fingerprinted samples have > 10.0% discordance with their Sequenom Fingerprints)																										
HapMap Concordance: ✓ Mean HapMap Concordance: 99.78% ± 0.02 Passed QC (10 of 10 HapMap samples found have > 90.0% concordance with their control)																										
Summary By Chemistry Plate:																										
Chemistry Plate	Samples	Num Scanned	Pct Scanned	Call Rate	Num Pass QC	Pct Pass QC	Pass QC Call Rate	Pass Gender Concordance	Pass Fingerprint Concordance	Pass HapMap Concordance	HapMap Concordance															
Cyto-12_01	96	96	100.0%	99.32% ± 0.62	95	99.0%	99.37% ± 0.33	88 / 88	88 / 88	1 / 1	99.8% ± 0.0															
Cyto-12_02	96	96	100.0%	99.27% ± 0.6	94	97.9%	99.34% ± 0.35	92 / 92	90 / 92	1 / 1	99.79% ± 0.0															
Cyto-12_03	96	96	100.0%	99.17% ± 0.39	96	100.0%	99.17% ± 0.39	91 / 91	90 / 90	1 / 1	99.75% ± 0.0															
Cyto-12_04	96	96	100.0%	99.4% ± 0.36	96	100.0%	99.4% ± 0.36	90 / 90	90 / 90	1 / 1	99.79% ± 0.0															
Cyto-12_05	96	96	100.0%	99.06% ± 0.37	96	100.0%	99.06% ± 0.37	85 / 85	85 / 85	1 / 1	99.79% ± 0.0															
Cyto-12_06	96	96	100.0%	99.17% ± 0.52	94	97.9%	99.22% ± 0.39	92 / 93	92 / 92	1 / 1	99.76% ± 0.0															
Cyto-12_07	96	96	100.0%	99.06% ± 0.28	96	100.0%	99.06% ± 0.28	92 / 93	92 / 93	1 / 1	99.78% ± 0.0															
Cyto-12_08	96	96	100.0%	99.16% ± 0.38	96	100.0%	99.16% ± 0.38	89 / 90	90 / 90	1 / 1	99.74% ± 0.0															
Cyto-12_09	96	96	100.0%	99.15% ± 1.67	95	99.0%	99.32% ± 0.29	94 / 95	94 / 94	1 / 1	99.77% ± 0.0															



Birdsuite for Illumina

- Running Birdsuite on the Illumina OmniQuad Chip
 - A calling algorithm originally developed for Affymetrix GeneChips
 - Currently Evaluating results
 - Results in more SNPs called with reduced time and effort spent on manual review than is needed with Autocall
 - Equal or greater concordance
 - 99.96% concordant (Birdsuite vs. BeadStudio)
 - 2.5M and Metabochip support soon
 - Run in parallel with Autocall (the default calling algorithm)
 - Both sets of genotypes are stored in the database

Infinium Plate Called with both AutoCall and Birdsuite

Infinium Chemistry Plate '										14'					
Name:	Scan Summary:									Summary:	Total	Scannable	Scanned	Pass QC Call Rate $\geq 97.0\%$	
Group:										Samples:	96	100.0% (96 of 96)	100.0% (96 of 96)	100.0% (96 of 96)	
Project:										Call Rates:			99.87% \pm 0.08	99.87% \pm 0.08	
Experiment:															
Chemistry: HumanOmni1-Quad_v1-0_B (Infinium 1 million Omni - 4 sample chip) 24 Chips Used	Gender Concordance:										Failed (1 of the 90 gender-comparable samples had discordant genders (discordance between infinium, fingerprint and reported genders)) Only 90 of 96 samples were gender-comparable (had infinium, fingerprint and reported genders)				
DNA Plate: (CO-2086533) 96 samples (0 empties) 96 unique clinical samples 96 unique individuals [38 M, 58 F, 0 U]	Fingerprint Concordance:										Failed QC (1 of 90 fingerprinted samples have > 10.0% discordance with their Sequenom Fingerprints)				
	HapMap Concordance:										Mean HapMap Concordance: 99.72% \pm 0.0 Passed QC (1 of 1 HapMap samples found have > 90.0% concordance with their control)				
Plate (Widget View)		Plate (List View)	Amp Plates	Workflow Steps											
Contents of Chemistry Plate:															
DNA Plate Well	Chip Barcode	Chip	Amp2 Plate	Amp2 Plate Well	Call Rate	Het %	Gender (Infinium / FP / Reported)	Birdsuite Call Rate	FP Discordance (Infr->Seq)	FP SNPs Called (Infr/Seq)	HapMap Concordance	Current Step	Date	Comments	Sample Instance Name
A01	5054681099_R01C01	5054681099	WG0031545-MSA1	A01	99.97%	23.749%	M / ! / M	99.0%	0/0	20 / 0	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMI
A02	5054761006_R01C01	5054761006	WG0031545-MSA1	A02	99.96%	23.93%	M / M / M	98.97%	0/19	19 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMJ
A03	5054753009_R01C01	5054753009	WG0031545-MSA1	A03	99.96%	23.866%	M / M / M	98.93%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMK
A04	5054753028_R01C01	5054753028	WG0031545-MSA1	A04	99.87%	24.415%	F / F / F	99.12%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQML
A05	5054745039_R01C01	5054745039	WG0031545-MSA1	A05	99.8%	24.082%	F / F / F	99.17%	1/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMM
A06	5054745048_R01C01	5054745048	WG0031545-MSA1	A06	99.86%	25.197%	F / F / F	99.04%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMN
A07	5054745059_R01C01	5054745059	WG0031545-MSA1	A07	99.95%	22.97%	M / M / M	98.95%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMO
A08	5054745156_R01C01	5054745156	WG0031545-MSA1	A08	99.86%	23.461%	F / F / F	99.17%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMP
A09	5054745074_R01C01	5054745074	WG0031545-MSA1	A09	99.86%	25.516%	F / F / F	99.15%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMQ
A10	5054745079_R01C01	5054745079	WG0031545-MSA1	A10	99.95%	24.264%	M / M / M	99.06%	0/20	20 / 23	n/a		Birdsuite Succeeded	9/15/10	SM-1CQMR



Accessing your data

- When an experiment completes in the LIMS, the data is made available to GePro and is stored in a readily accessible format
 - Same storage strategy used for genotyping and expression experiments
 - Stored in relational database or datavault (managed file system) depending on scale of data
- Use the ‘Analysis Set’ function in Ge/Pro to access your data in a flexible fashion
 - Filter by samples in project/experiment/plate/sample list...
 - Export to browser, Unix file system, or GAP Analysis Portal
 - Many Export Formats:
 - Plink-ready
 - Ped, Bed/Bim/Fam/Map
 - Native
 - CEL (Affymetrix)
 - CHP (Affymetrix)
 - IDAT (Illumina)
 - CSV (Luminex)
 - Birdsuite-generated files
 - Summary files
 - VCF

Analysis Set

EDIT ANALYSIS SET TEST (INFINIUM - HUMANCYTOSNP-12V2-1_A)

This page allows you to transfer data from GAP into files. The transfer is specified in three steps:

1. creating the set of data you want (inventory)
2. setting up the files that you want (set up transfer)
3. performing the transfer to a specified location (do transfer)

Each step of this process is segregated here into discrete sections. There are many options for manipulating the inventory and for control preferences.

Files are transferred to the GAP Analysis Portal. You are currently set up to use the Analysis Portal with the account: ggrant. [Click here](#)

To start the QC Pipeline, add 'QC Pipeline' to your data transfer items in step 2 below and remove any other items if you have them. The

[Show My Analysis Sets](#)

Create a new analysis set with the data from this analysis set (if successful, you will be taken to the new set):

new analysis set name:

Step One: Filter Inventory

Add Everything in Project

Inventory:

Experiment Plates	Lab	Variants
933 individuals	940 runs	1 assay lists
933 clinical samples		299140 assays
10 DNA plates		299140 polymorphisms
940 sample instances		
0 empties		
10 controls		

Step Two: Set Up Data Transfer

download type:

Step Three: Transfer

Please select Download Type

Step Two: Set Up Data Transfer

download type:

- Gtc File
- Gtc File (as text)
- Idat File (Green)
- Idat File (Red)
- Birdsuite Files

Step Two: Set Up Data Transfer

download type:

genotype:

phenotype:

call algorithm:

strand: always forward

species:

genome build:

gender:

family:

individual:

polymorphisms:

extract assay (auto-complete)
list:

space substitute:



Portals

- One target of export from Analysis Set is the GAP Analysis Portal
 - Visible from outside of Broad
 - Useful for timely dissemination of results
- Other portals deployed by GAP:
 - GTEx
 - CARE
 - ISC
 - Yale U19 (Immune responsiveness signatures)
 - Viral Genomics Initiative
 - Broad Quotes

GAP Analysis Portal

The screenshot displays the GAP Analysis Portal interface. At the top, there is a navigation bar with links for Home, Access Data, QC, and Report Issue. A user status message indicates "grant is logged in | sign out". On the left, a sidebar titled "ACCESS DATA" contains sections for "Study Datasets" (with a link to "Show/hide all 2 studies"), "Batch1DataSet" (description, maintenance links for "Archive the study" and "Permanently delete the study", and a list of authorized users), and "Manage Files" (with a note about file size limits and a document upload form). Below the sidebar, a table lists 37 datasets ready for download, showing columns for File Name, Type, File Size, Author, Date, and Operations (Download, Report Issue, Delete).

File Name	Type	File Size	Author	Date	Operations
Batch1DataSet.2010-10-01_102431.log	text/plain	94.5 KB	hrafal	2010-10-01 13:03:29	Download Report Issue Delete
Batch1DataSet.zip-35.zip	application/zip	380.5 MB	hrafal	2010-10-01 13:01:41	Download Report Issue Delete
Batch1DataSet.zip-34.zip	application/zip	1 GB	hrafal	2010-10-01 12:57:05	Download Report Issue Delete
Batch1DataSet.zip-33.zip	application/zip	1 GB	hrafal	2010-10-01 12:52:38	Download Report Issue Delete
Batch1DataSet.zip-32.zip	application/zip	1 GB	hrafal	2010-10-01 12:48:14	Download Report Issue Delete
Batch1DataSet.zip-21.zip	application/zip	1 GB	hrafal	2010-10-01 12:42:41	Download Report Issue Delete

GAP Portals

The screenshot shows the International Stroke Genetics Consortium (ISGC) website. At the top, there is a logo for ISGC and a banner for 'three sisters'. Below the banner, there is a sidebar with various links and a main content area with text about schizophrenia and stroke.

Study Datasets:

- 5 iter
- FILE
- phen
- Plink
- Affy
- Raw

PI: [redacted]
Sponsor: [redacted]
Grant Number: [redacted]

Required: A

Datasets:

- Click on the yellow button
- We report compared stroke or reversible with ticlopiptoprotein more effe

Required: A

Study Datasets:

- create a co
- genotype 7 phenotypes
- perform ana disorders a

The proposed res consequences. All the phenotypic a community.

PI: [redacted]
Sponsor: [redacted]
Grant Number: [redacted]

Copyright © 2007-2008 Ma

Trait Catalog - Mozilla Firefox

Tuesday, October 28, 2008 1:59:32 PM EDT
dinsmore@broad.mit.edu is logged in | logout

Tuesday, October 28, 2008 1:54:16 PM EDT
signup for an account | log in | password reminder

CARE GAPAnalysis Portal

Home Access Data QC Report Issue

TRAIT CATALOG

ACCESS DATA

Batch1DataSet

Description: Batch1DataSet

Study Maintenance: Archive the study | Permanently delete the study

Authorized Users

Manage Files

37 datasets ready for download

File Name	Type	File Size	Author	Date	Operations
Batch1DataSet.2010-10-01_102431.log	text/plain	94.5 KB	hrafal	2010-10-01 13:03:29	Download Report Issue
Batch1DataSet.zip-35.zip	application/zip	380.5 MB	hrafal	2010-10-01 13:01:41	Download Report Issue
Batch1DataSet.zip-34.zip	application/zip	1 GB	hrafal	2010-10-01 12:57:05	Download Report Issue
Batch1DataSet.zip-33.zip	application/zip	1 GB	hrafal	2010-10-01 12:52:38	Download Report Issue
Batch1DataSet.zip-32.zip	application/zip	1 GB	hrafal	2010-10-01 12:48:14	Download Report Issue
Batch1DataSet.zip-31.zip	application/zip	1 GB	hrafal	2010-10-01 12:42:41	Download Report Issue

Integration with the Broad

- Strongly integrated with the BSP platform
 - Most GAP samples are loaded directly from BSP (plated in a platform-appropriate fashion)
 - GAP Uploads fingerprint genotypes to BSP
 - Many web services developed to share sample information and to handle search details
- Chem Bio
 - loads plates (for Luminex) directly into GAP using GAP-developed API
 - Retrieves Luminex results and assay definitions
- Sequencing access GAP fingerprinting (and genotypes in general) via web services in GAP and BSP
- Broad Quotes integrated with GAP and BSP via APIs.
 - With SAP by shared files
- Aided Genepattern group to add support for Illumina DASL expression to Genepattern
- Host and Support the SNAP service
 - Adding the 1KG pilot 1 data set



Contact us: Suggestions / Complaints

- GAP Informatics Home Page at: <http://iwww.broadinstitute.org/gapinformatics>
 - Links to GAP applications
 - Link to ‘Jira’ – tool for tracking bugs/feature requests
 - can also create a Jira issue by mailing to gapsupport@broadinstitute.org
- Weekly user meetings (Mondays – 2:00) in Chang-Jiang room in 301 Binney
- ggrant@broadinstitute.org



Thank You

- Boris Bugalter
- Jason Carey
- Michael Dinsmore
- Nina Lapchyk
- Pei Lin
- Huy Nguyen
- Howard Rafal
- Greg Rushton
- Dennis Ryan
- Ellen Winchester
- Bob Handsaker
- Jim Nemesh
- Josh Korn
- Jason Flannick
- Kristian Cibulskis
- Tim Fennel
- Doug Voet