

BOSC 2013
14th Annual Bioinformatics Open Source Conference
Saturday July 20th, 2013
Berlin, Germany

OmicsConnect: flexible multi-omics data capture and integration tools for high-throughput biology

**K. Joeri van der Velde, Morris Swertz,
members of the Genomics Coordination Center,
& many others**



Genomics Coordination Center, UMCG, Groningen

(NGS) diagnostics & clinic

(Inter)national bioinformatics hub of ~15 on a mission to speed up rare and complex disease research and improve patient care via databases, integration tools and analysis pipelines



Biobanking

Large scale multi-omics

Looking for colleague(s) 😊

Human and model organism research: genotype 2 phenotype

MOLGENIS: Motivation – building many apps

Researcher needs



NextGenSeq data

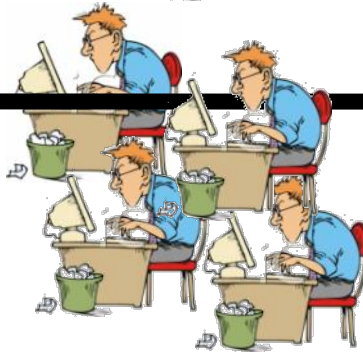
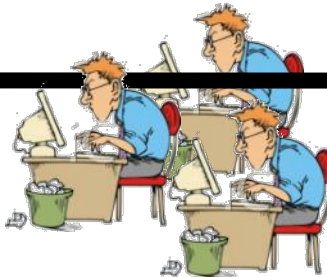
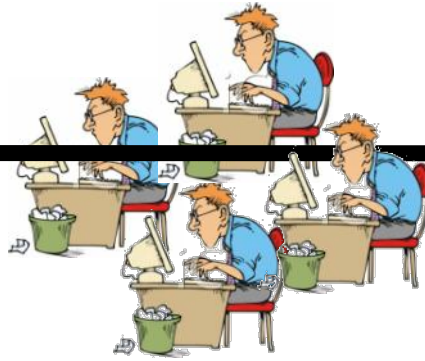


Mutation data



Model organisms data

Work very hard



Use

Solexa Sequencer LIMS

Sampling
Flowcell Preps
flowcells
recipes
sequence_primers
Genome Analyzer Runs
Pipeline Runs
Admin

id	flowcellid	researcher_id	flowcell_date	reagent1	reagent2	primer	comments	sequence_primer_id
1								

database of COL7A1 mutations

SearchDatabase | UploadData | Contact | References | Package | Help | Login

SearchDatabase

Find a specific mutation/variation

Variation: or

Nucleotide No:

Amino Acid No:

Find mutations/variations

Exon/Intron: or

Select mutation type:

Select protein domain:

Search term:

Animal Observatory

Animals Present

id	name	status
1	mouse	alive

Selected animal ID:

Copy event to selected animals:

Details:

Species:

Background:

Mutation:

Genotype:

Sex:

Age:

Source:

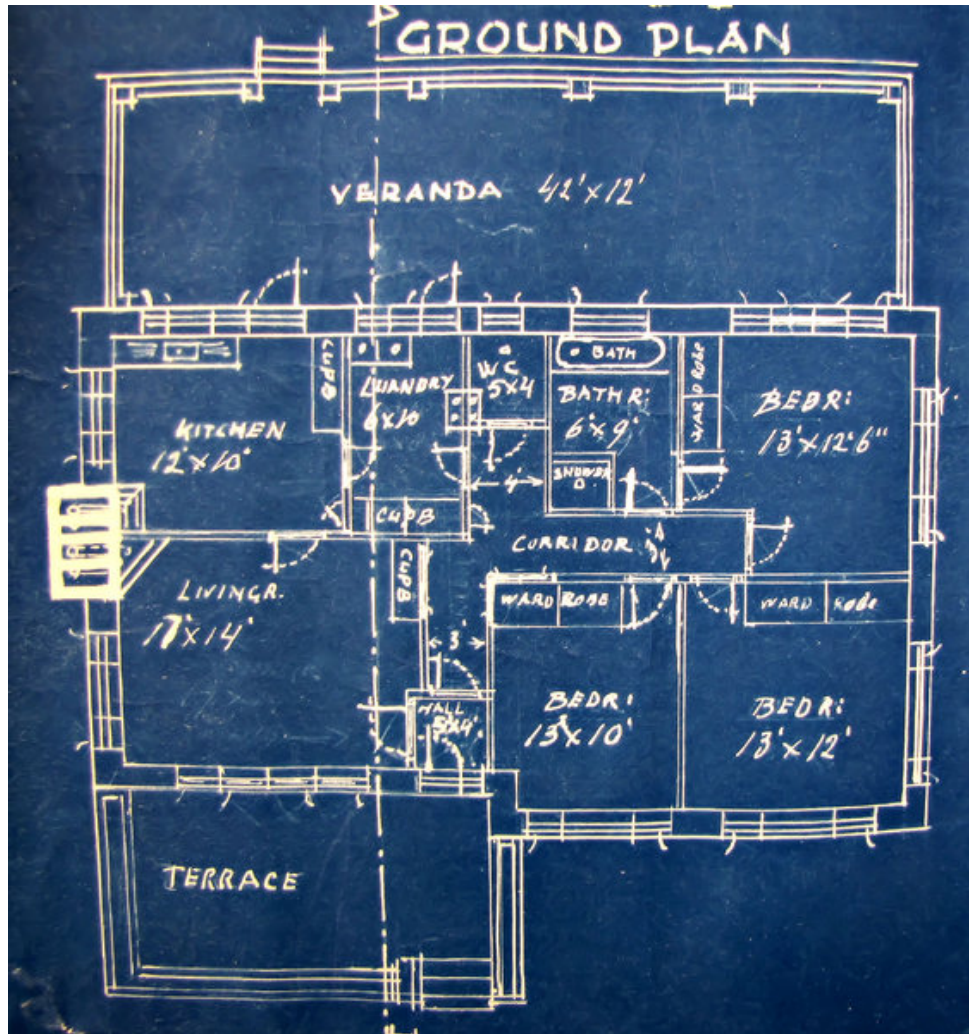
Location:

Participating Group:

Participating Group:

Participating Group:

Instead: design blueprint of data model & GUI



What comes where?

What dimensions?

E.g.

- Phenotype reports?
- Genotype reports?
- Mutation reports?
- News?
- Submissions?
- Literature?
- Clinician contacts?
- ...

MOLGENIS: Solution – generate the software

Model in DSL



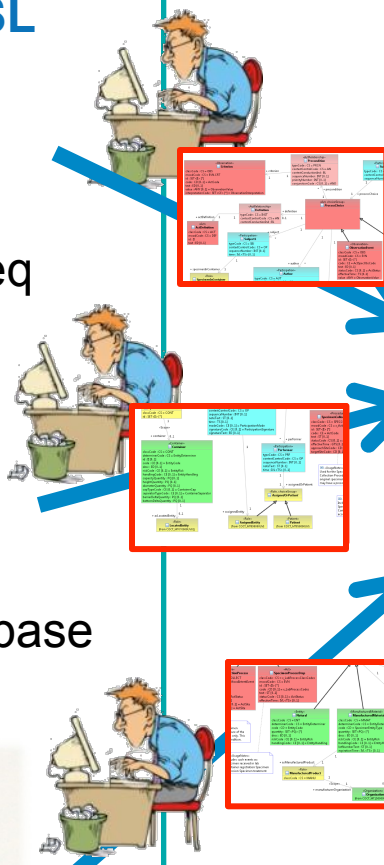
NextGenSeq



Mutation database



Model organisms



Use generated software

Solexa Sequencer LIMS

database of COL7A1 mutations

Animal Observatory

<http://www.molgenis.org>

Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

Swertz et al (2004) *Bioinformatics* 20(13), 2075-83



university of
 groningen

genomics coordination
 center

Example of a basic application

BBMRI-NL Biobank Registry



[Exchange format](#) | [R-project API](#) | [HTTP API](#) | [Web Services API](#)

Blobanks

Blobankers

Institutes

Batch Upload

Admin

Institutes

File Edit View

1 - 10 of 28

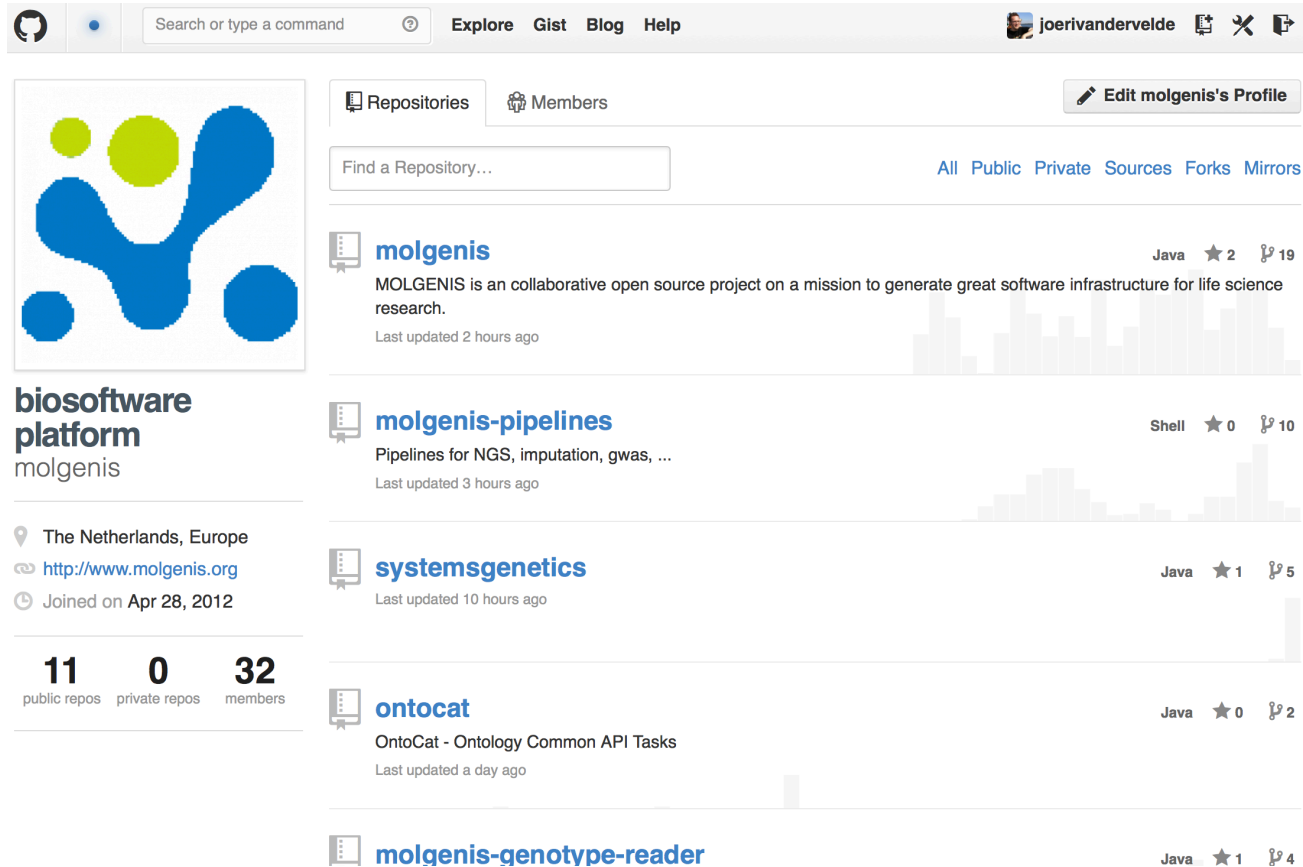
Search: id =

		id ▼	name	Address	Phone	Email	Fax	tollFreePhone
1.		<input type="checkbox"/>	1	All UMCs				
2.		<input type="checkbox"/>	2	AMC				
3.		<input type="checkbox"/>	3	BOOG				
4.		<input type="checkbox"/>	4	DDHK				
5.		<input type="checkbox"/>	5	DCCG				
6.		<input type="checkbox"/>	6	ErasmusMC				
7.		<input type="checkbox"/>	7	EUR				
8.		<input type="checkbox"/>	8	GGD				
9.		<input type="checkbox"/>	9	HIV Monitoring Foundation				
10.		<input type="checkbox"/>	10	ICC consortium (Parelsnoer)				

* = this record is readonly.

This database was generated using the open source [MOLGENIS database generator](#) version 3.3.3.
Please cite [Swertz et al \(2004\)](#) or [Swertz & Jansen \(2007\)](#) on use.

Migrated from SVN to GitHub



The screenshot shows the GitHub profile page for 'molgenis'. At the top, there's a navigation bar with 'Explore', 'Gist', 'Blog', and 'Help'. The user's name 'joerivandervelde' is visible. The profile header includes 'Repositories' and 'Members' tabs, and a search bar. Below this, there are several repository cards, each with a repository icon, name, description, language, star count, and fork count. The repositories listed are: 'molgenis' (Java, 2 stars, 19 forks), 'molgenis-pipelines' (Shell, 0 stars, 10 forks), 'systemsgenetics' (Java, 1 star, 5 forks), 'ontocat' (Java, 0 stars, 2 forks), and 'molgenis-genotype-reader' (Java, 1 star, 4 forks). On the left side of the profile, there's a bio section for 'biosoftware platform molgenis' with location 'The Netherlands, Europe', website 'http://www.molgenis.org', and join date 'Apr 28, 2012'. At the bottom of the bio section, there are statistics: 11 public repos, 0 private repos, and 32 members.



~15 active devs

~25 projects

github.com/molgenis



Migrated from Ant builds to Maven modules

- MOLGENIS collection of repositories
 - **molgenis**: rich application toolbox
 - **sdk**: bare bones development
 - *...others*

molgenis-app-compute-db	added extra parameter to specify backend credentials; make it run fro...	10 days ago
molgenis-app-lifelines	add .gitignore files	14 days ago
molgenis-app-ngs	update spring 3.2.2 --> 3.2.3	15 days ago
molgenis-app-omx	update spring 3.2.2 --> 3.2.3	15 days ago
molgenis-compute-core	fixed build; removed wrong testing	2 hours ago
molgenis-core-ui	fixed the bug that RestApi throws an error when the value of expanded...	14 days ago
molgenis-core	improve entity importer performance	14 days ago
molgenis-omx-auth	update spring 3.2.2 --> 3.2.3	15 days ago
molgenis-omx-core	update spring 3.2.2 --> 3.2.3	15 days ago
molgenis-omx-dataexplorer	add .gitignore files	14 days ago
molgenis-omx-filemanager	removed jqgrid + removed dataviewer	22 days ago
molgenis-omx-importer	improve dataset importer performance	14 days ago
molgenis-omx-protocolmanager	add .gitignore files	14 days ago
molgenis-omx-protocolviewer	fixed a major in catalogue that protocol might be duplicated twice wi...	21 days ago
molgenis-search-elasticsearch	add .gitignore files	14 days ago
molgenis-search	add .gitignore files	14 days ago

maven

spring
mvc



Bootstrap

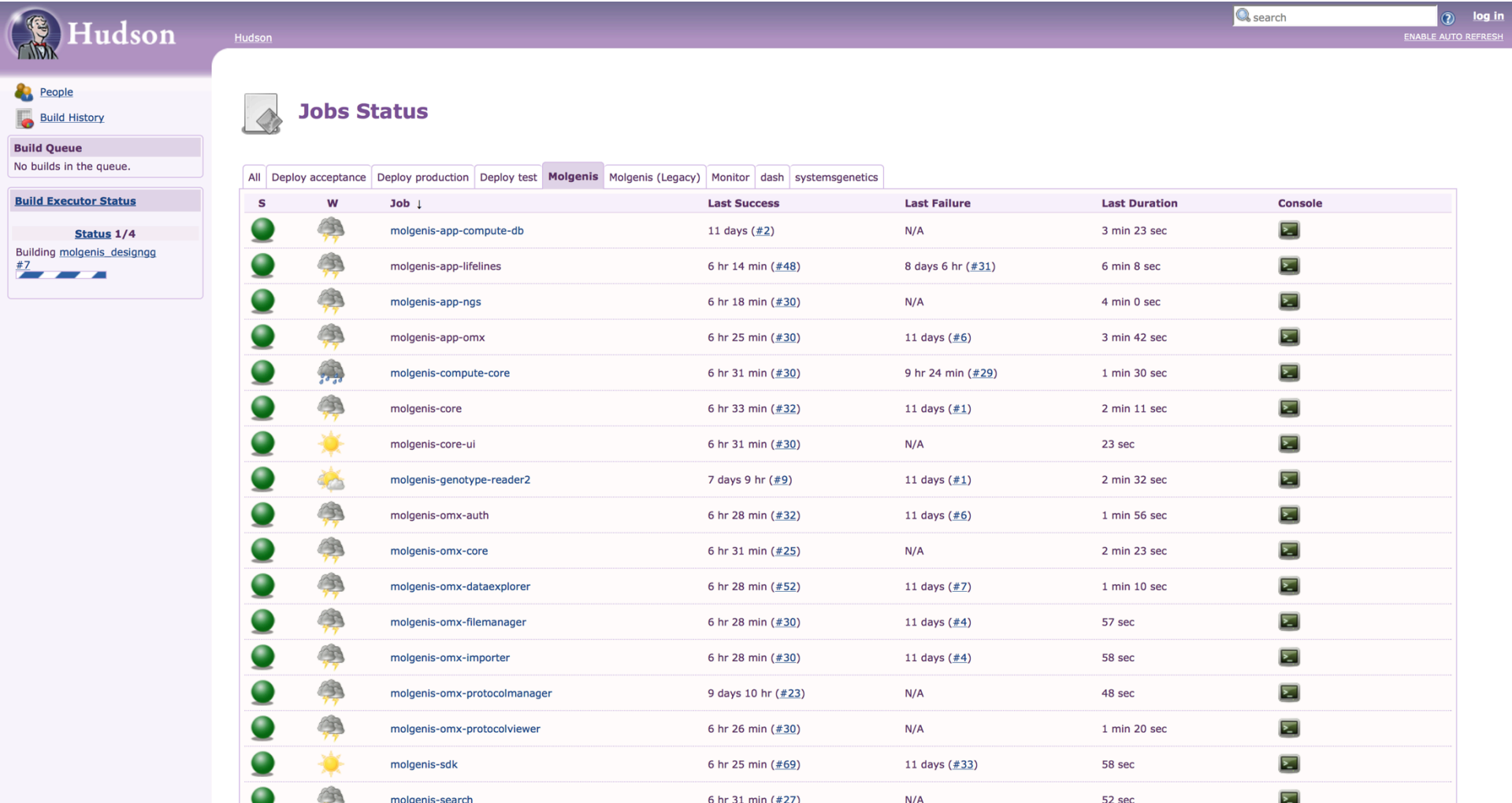


JPA / MySQL

elasticsearch.



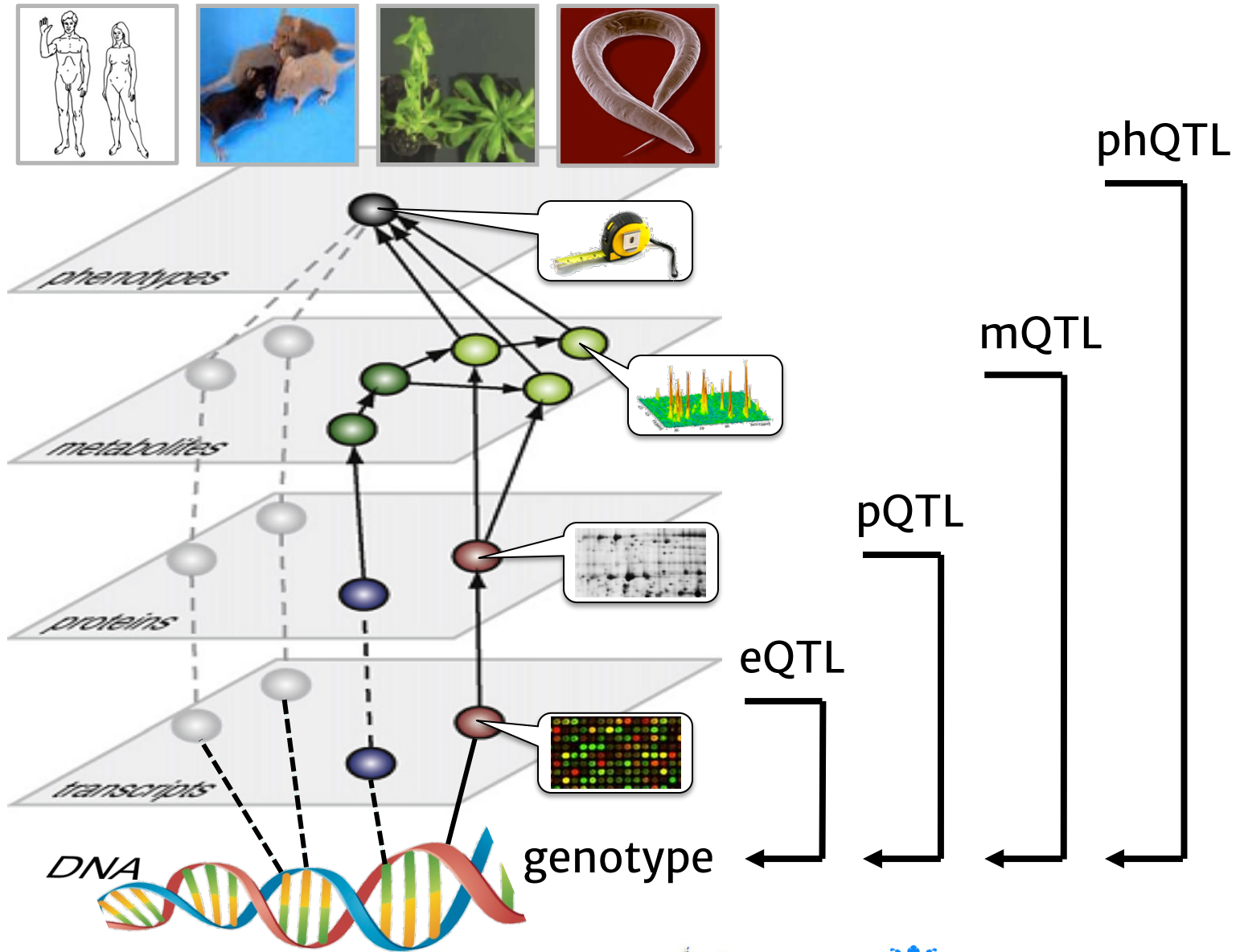
Continuous integration & automated deploy



The screenshot shows the Hudson web interface. At the top left is the Hudson logo. On the right, there is a search bar and a 'log in' link. Below the search bar, the text 'ENABLE AUTO REFRESH' is visible. The main content area is titled 'Jobs Status' and features a navigation bar with tabs for 'All', 'Deploy acceptance', 'Deploy production', 'Deploy test', 'Molgenis' (selected), 'Molgenis (Legacy)', 'Monitor', 'dash', and 'systemsgenetics'. Below the tabs is a table listing various jobs. The table has columns for 'S' (Success), 'W' (Warning), 'Job', 'Last Success', 'Last Failure', 'Last Duration', and 'Console'. The jobs listed include 'molgenis-app-compute-db', 'molgenis-app-lifelines', 'molgenis-app-ngs', 'molgenis-app-omx', 'molgenis-compute-core', 'molgenis-core', 'molgenis-core-ui', 'molgenis-genotype-reader2', 'molgenis-omx-auth', 'molgenis-omx-core', 'molgenis-omx-dataexplorer', 'molgenis-omx-filemanager', 'molgenis-omx-importer', 'molgenis-omx-protocolmanager', 'molgenis-omx-protocolviewer', 'molgenis-sdk', and 'molgenis-search'. On the left side of the dashboard, there are sections for 'Build Queue' (No builds in the queue), 'Build History', and 'Build Executor Status' (Status 1/4, Building molgenis_designgg #7).

S	W	Job ↓	Last Success	Last Failure	Last Duration	Console
		molgenis-app-compute-db	11 days (#2)	N/A	3 min 23 sec	
		molgenis-app-lifelines	6 hr 14 min (#48)	8 days 6 hr (#31)	6 min 8 sec	
		molgenis-app-ngs	6 hr 18 min (#30)	N/A	4 min 0 sec	
		molgenis-app-omx	6 hr 25 min (#30)	11 days (#6)	3 min 42 sec	
		molgenis-compute-core	6 hr 31 min (#30)	9 hr 24 min (#29)	1 min 30 sec	
		molgenis-core	6 hr 33 min (#32)	11 days (#1)	2 min 11 sec	
		molgenis-core-ui	6 hr 31 min (#30)	N/A	23 sec	
		molgenis-genotype-reader2	7 days 9 hr (#9)	11 days (#1)	2 min 32 sec	
		molgenis-omx-auth	6 hr 28 min (#32)	11 days (#6)	1 min 56 sec	
		molgenis-omx-core	6 hr 31 min (#25)	N/A	2 min 23 sec	
		molgenis-omx-dataexplorer	6 hr 28 min (#52)	11 days (#7)	1 min 10 sec	
		molgenis-omx-filemanager	6 hr 28 min (#30)	11 days (#4)	57 sec	
		molgenis-omx-importer	6 hr 28 min (#30)	11 days (#4)	58 sec	
		molgenis-omx-protocolmanager	9 days 10 hr (#23)	N/A	48 sec	
		molgenis-omx-protocolviewer	6 hr 26 min (#30)	N/A	1 min 20 sec	
		molgenis-sdk	6 hr 25 min (#69)	11 days (#33)	58 sec	
		molgenis-search	6 hr 31 min (#27)	N/A	52 sec	

Motivation: Understanding geno-to-pheno



Situation: Many types and flavours of data

Genomic features, individuals, ontologies ..

Metadata for phenotypes, datasets, samples, panels ...

Gene	Accession	Description	Chr	Start	End	Strand	Feature
GO:004291	GO:004291	integral to membrane	chr1	1000000	1000100	+	MEMBRANE
GO:004291	GO:004291	integral to membrane	chr1	1000100	1000200	+	MEMBRANE
GO:004291	GO:004291	integral to membrane	chr1	1000200	1000300	+	MEMBRANE
GO:004291	GO:004291	integral to membrane	chr1	1000300	1000400	+	MEMBRANE
GO:004291	GO:004291	integral to membrane	chr1	1000400	1000500	+	MEMBRANE

Sample	Phenotype	Value	Unit	Method
EA_N_LSP	Mean Lifespan at 240C	12.0	days	Survival
EA_N_LSP	Mean Lifespan at 200C	18.0	days	Survival
EA_N_LSP	Age At Maturity at 240C	4.0	days	Survival
EA_N_LSP	Age At Maturity at 200C	8.0	days	Survival
EA_N_LSP	Fertility at 240C	1.0	progeny	Reproduction

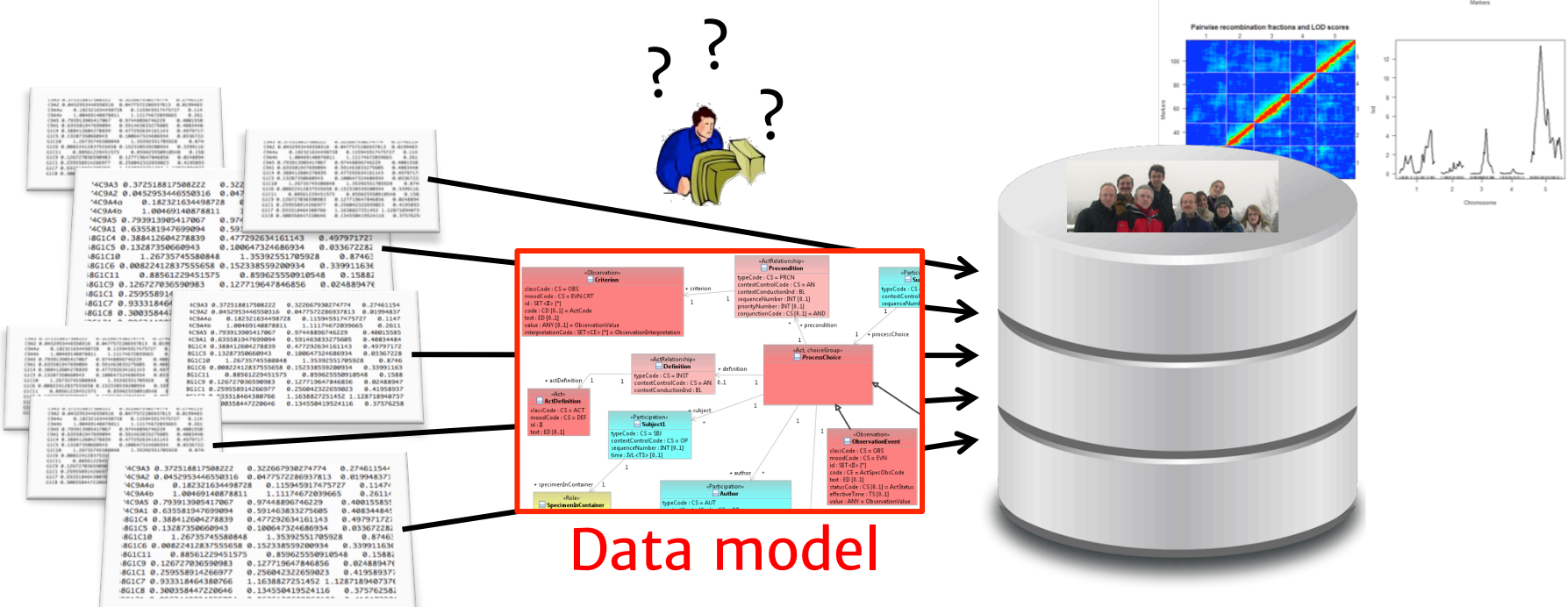
Biomolecular measurements, association results ...

Genotypes, conditions ...

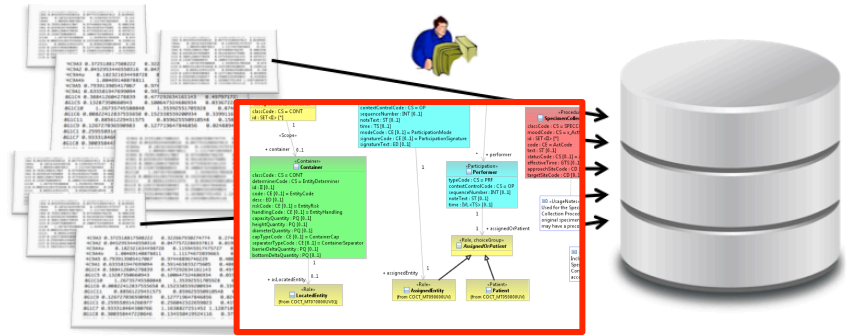
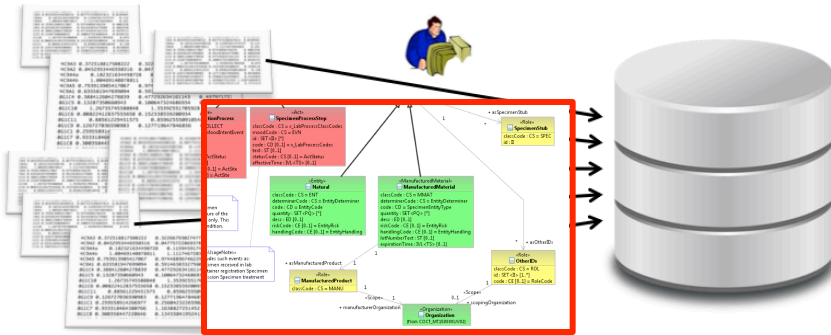
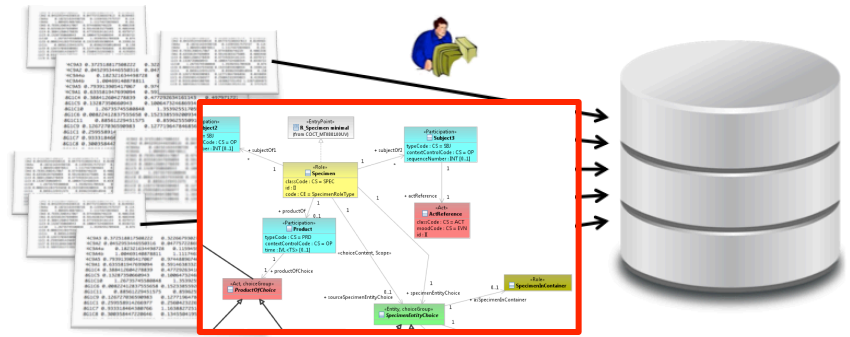
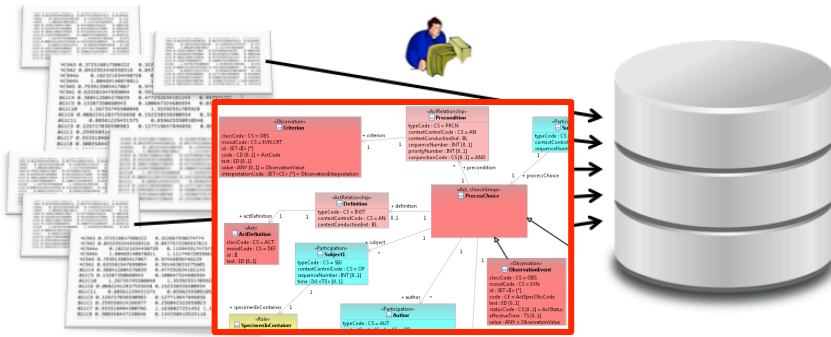
Gene	Accession	Value	Unit	Method
GO:004291	GO:004291	1.234567	log2	RNA-seq
GO:004291	GO:004291	2.345678	log2	RNA-seq
GO:004291	GO:004291	3.456789	log2	RNA-seq
GO:004291	GO:004291	4.567890	log2	RNA-seq

Sample	Genotype	Value	Unit	Method
K1194	CB4856	N2	Genotype	Genotyping
K1195	CB4856	N2	Genotype	Genotyping
K1196	CB4856	N2	Genotype	Genotyping
K1197	CB4856	N2	Genotype	Genotyping
K1198	CB4856	N2	Genotype	Genotyping

Challenge: Building a 'team' database



Bigger challenge: Building many databases



MOLGENIS: Many unique apps is not optimal

Model in DSL



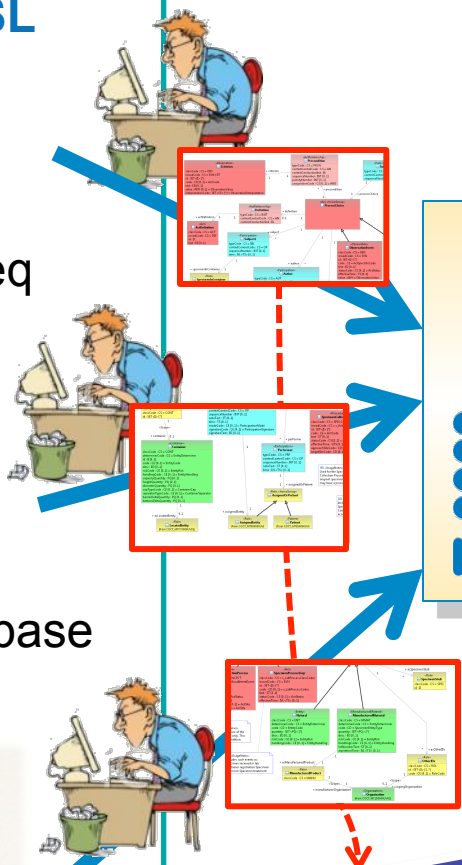
NextGenSeq



Mutation database

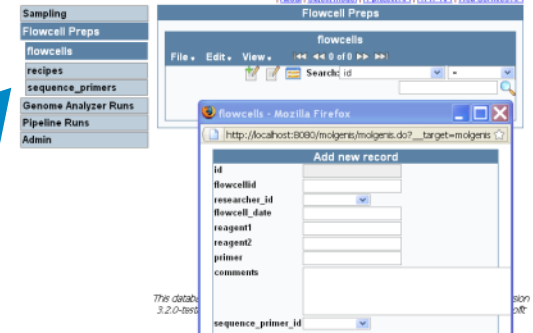


Model organisms



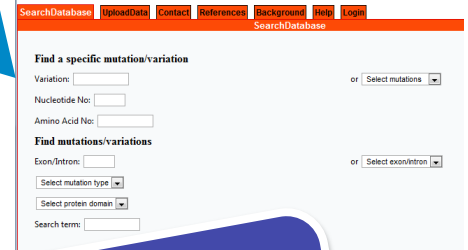
Use generated software

Solexa Sequencer LIMS



What data models to use?
Can we have a model that rules them all?
One application > Many applications

database of COL7A1 mutations



<http://www.molgenis.org>

Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

Swertz et al (2004) *Bioinformatics* 20(13), 2075-83

What are we dealing with?

Genomic features,
individuals, ontologies ...

ID	Name	Value	Unit
P1000	100000	1.0	kg
P1001	100001	1.0	kg
P1002	100002	1.0	kg
P1003	100003	1.0	kg
P1004	100004	1.0	kg
P1005	100005	1.0	kg
P1006	100006	1.0	kg
P1007	100007	1.0	kg
P1008	100008	1.0	kg
P1009	100009	1.0	kg
P1010	100010	1.0	kg

Metadata for phenotypes,
datasets, samples, panels ...

ID	Name	Value	Unit
P1000	100000	1.0	kg
P1001	100001	1.0	kg
P1002	100002	1.0	kg
P1003	100003	1.0	kg
P1004	100004	1.0	kg
P1005	100005	1.0	kg
P1006	100006	1.0	kg
P1007	100007	1.0	kg
P1008	100008	1.0	kg
P1009	100009	1.0	kg
P1010	100010	1.0	kg

Biomolecular measurements,
association results ...

Gene	SNP	Value	Unit
Gene1	SNP1	1.0	kg
Gene1	SNP2	1.0	kg
Gene1	SNP3	1.0	kg
Gene1	SNP4	1.0	kg
Gene1	SNP5	1.0	kg
Gene1	SNP6	1.0	kg
Gene1	SNP7	1.0	kg
Gene1	SNP8	1.0	kg
Gene1	SNP9	1.0	kg
Gene1	SNP10	1.0	kg

Genotypes,
conditions ...

ID	Name	Value	Unit
P1000	100000	1.0	kg
P1001	100001	1.0	kg
P1002	100002	1.0	kg
P1003	100003	1.0	kg
P1004	100004	1.0	kg
P1005	100005	1.0	kg
P1006	100006	1.0	kg
P1007	100007	1.0	kg
P1008	100008	1.0	kg
P1009	100009	1.0	kg
P1010	100010	1.0	kg

Stable?

Annotations of concepts
used in data sets,
mostly static content

Dynamic?

Experimental data sets,
usually flexible and
volatile content

Example: eQTL data

Stable?

Stable?

Probe (annotation)

name	mismatch	description
WSU1	true	NA / SpotReport / blast_match_NA / n
WSU2	false	C25A1.8 / cea2.c.00914 / blast_match
WSU3	false	F21F3.6 / cea2.c.02677 / blast_match
WSU4	false	F25H2.9 / cea2.c.02801 / blast_match
WSU5	false	F56H1.4 / cea2.c.04344 / blast_match

Marker (annotation)

name	chromosome	bpstart	cm	description
pkP1050	I	169018	-18.26	PCR_non_cu
pkP1101	I	992189	-17.28	PCR_non_cu
pkP1103	I	1881116	-11.96	PCR_non_cu
pkP1052	I	2818974	-6.1	PCR_non_cu
egPE107	I	3502476	-3.55	PCR_non_cu

	pkP1050	pkP1101	pkP1103	pkP1052	egPE107
WSU1	0.5036565	0.4603331	0.3544101	0.9123223	0.4157701
WSU2	0.1365825	0.1037487	0.6600898	0.1241076	0.1672705
WSU3	0.5837218	0.5611695	0.1708836	1.439448	1.94431
WSU4	0.5558796	0.7246171	0.1777933	0.1937225	0.4413371
WSU5	0.3393896	0.4705863	0.224066	0.7713159	0.01334126

Dynamic?

eQTL profiles (data set)

Stable = good for code generation

Annotations: Column-oriented data

name	chromosome	bpstart	cm	description
pkP1050	I	169018	-18.26	PCR_non_cu
pkP1101	I	992189	-17.28	PCR_non_cu
pkP1103	I	1881116	-11.96	PCR_non_cu
pkP1052	I	2818974	-6.1	PCR_non_cu
egPE107	I	3502476	-3.55	PCR_non_cu

← Attributes

3. import

↓ 1. model

```
<entity name="Locus" abstract="true">
  <description> position. Typical examples of such traits are genes,
  probes and markers. Common structure for entities that have a
  genomic</description>
  <field name="Chromosome" label="Chromosome" type="xref"
    xref_entity="Chromosome" xref_field="id" xref_label="name" nillable="1"
    description="Reference to the chromosome this
    position belongs to." />
  <field name="cm" label="cmPosition" type="decimal" nillable="true"
    description="genetic map position in centi_morgan (cM)." />
  <field name="bpStart" label="Start (5')" type="long" nillable="true"
    description="numeric basepair postion (5') on the chromosome" />
  <field name="bpEnd" label="End" type="long" nillable="true"
    description="numeric basepair postion (3') on the chromosome" />
  <field name="Seq" type="text" nillable="true"
    description="The FASTA text representation of the sequence." />
  <field name="Symbol" type="varchar" nillable="true"
    description="todo" />
</entity>
<entity name="Chromosome" extends="ObservableFeature">
  <field name="orderNr" type="int" />
  <field name="isAutosomal" type="bool" description="Is 'yes' when number of chr
  <field name="bpLength" type="int" nillable="true" description="Lenght of the c
  <field name="Species" label="Species" type="xref" xref_entity="Species"
    xref_field="id" xref_label="name" nillable="true"
    description="Reference to the species this
    chromosome belongs to." />
</entity>
```

2. generate

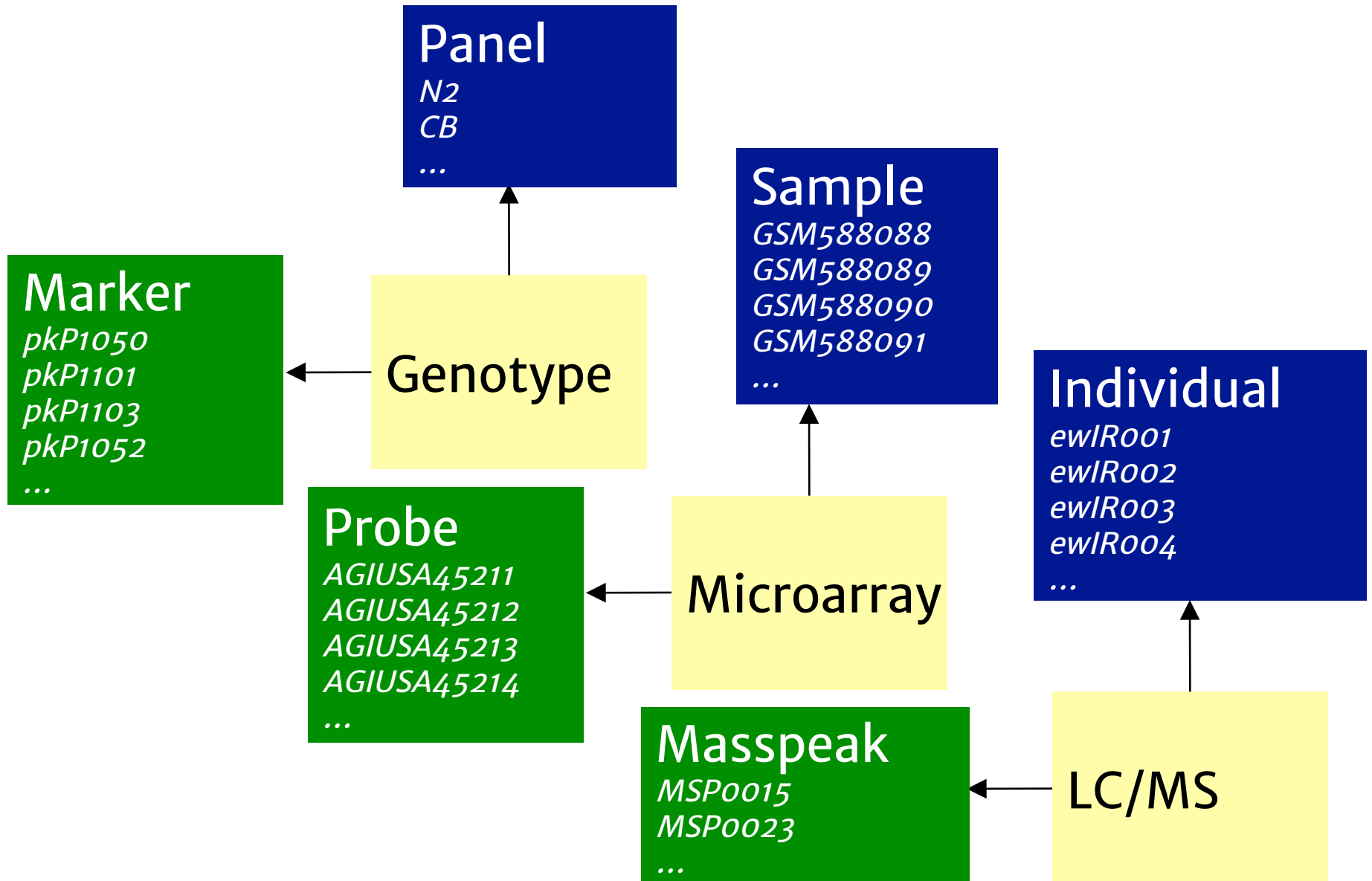


Micro-array probes blasted against WS220 and linked to the genes, used for gene expression phenotypes and eQTLs. Click on to plot an item, and on to return to this list.

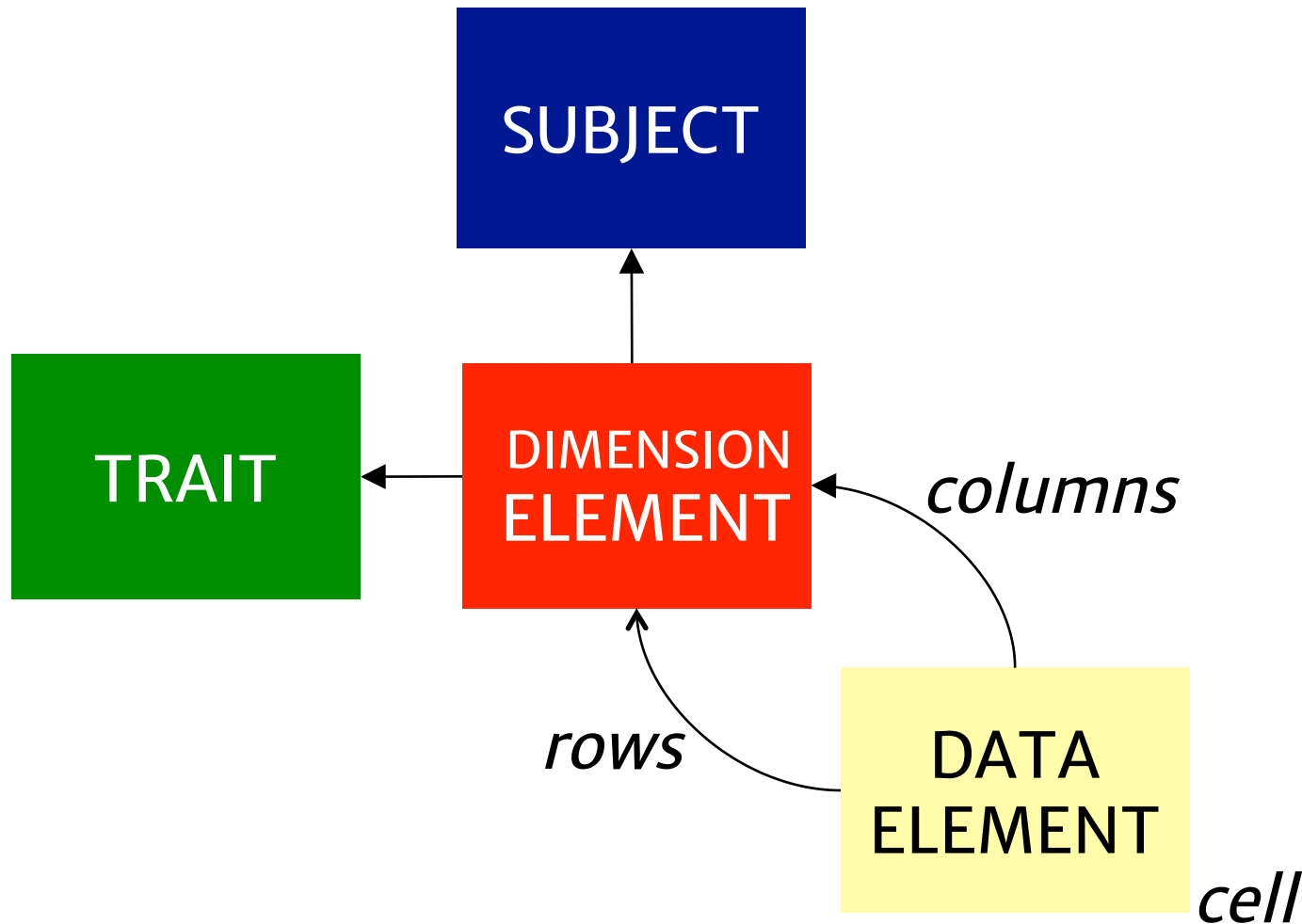
id	name	description	Investigation ontology	Reference	Alternative identifiers	label	Chromosome	cM
105050	WSU1	NA / SpotReport / blast_match_NA / non_cumcu_bp_start_0	Public					
105051	WSU2	C25A1.8 / cea2.c.00914 / blast_match_60 / non_cumcu_bp_start_10184580	Public			dec-87	I	
105052	WSU3	F21F3.6 / cea2.c.02677 / blast_match_60 / non_cumcu_bp_start_4912043	Public			F21F3.6	I	
105053	WSU4	F25H2.9 / cea2.c.02801 / blast_match_60 / non_cumcu_bp_start_10567120	Public			pas-5	I	
105054	WSU5	F56H1.4 / cea2.c.04344 / blast_match_60 / non_cumcu_bp_start_5741975	Public			rpt-5	I	
105055	WSU6	H06001.1 / cea2.c.04508 / blast_match_60 / non_cumcu_bp_start_7011070	Public			pdl-3	I	

A yellow thumbs-up icon with a smiling face, indicating a positive or successful outcome.

Challenge: Data sets can be variable combinations

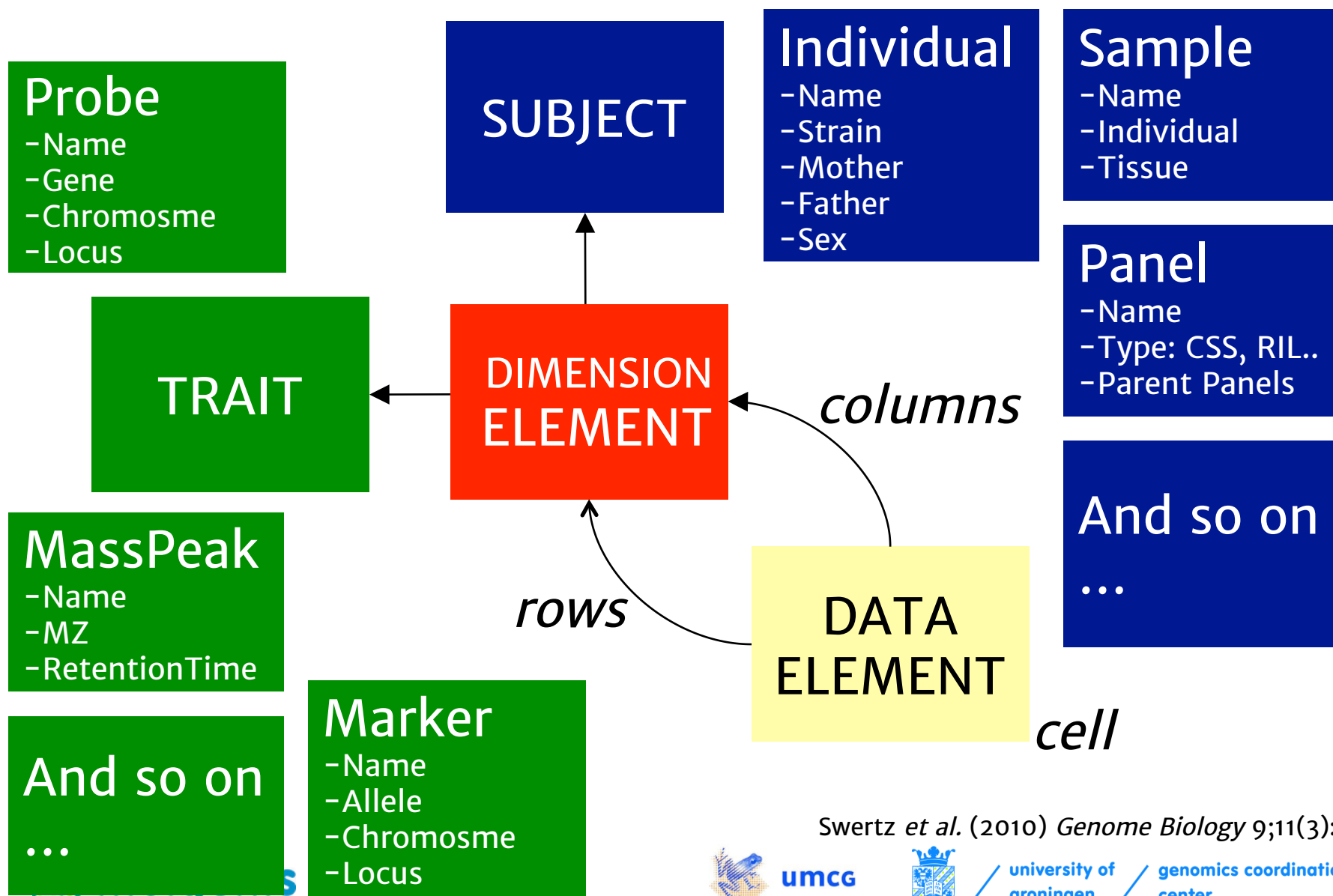


XGAP model: <any trait> X <any subject>



Swertz et al. (2010) *Genome Biology* 9;11(3): R27.

Extensible core model for homogen. datasets



Swertz et al. (2010) *Genome Biology* 9;11(3): R27.



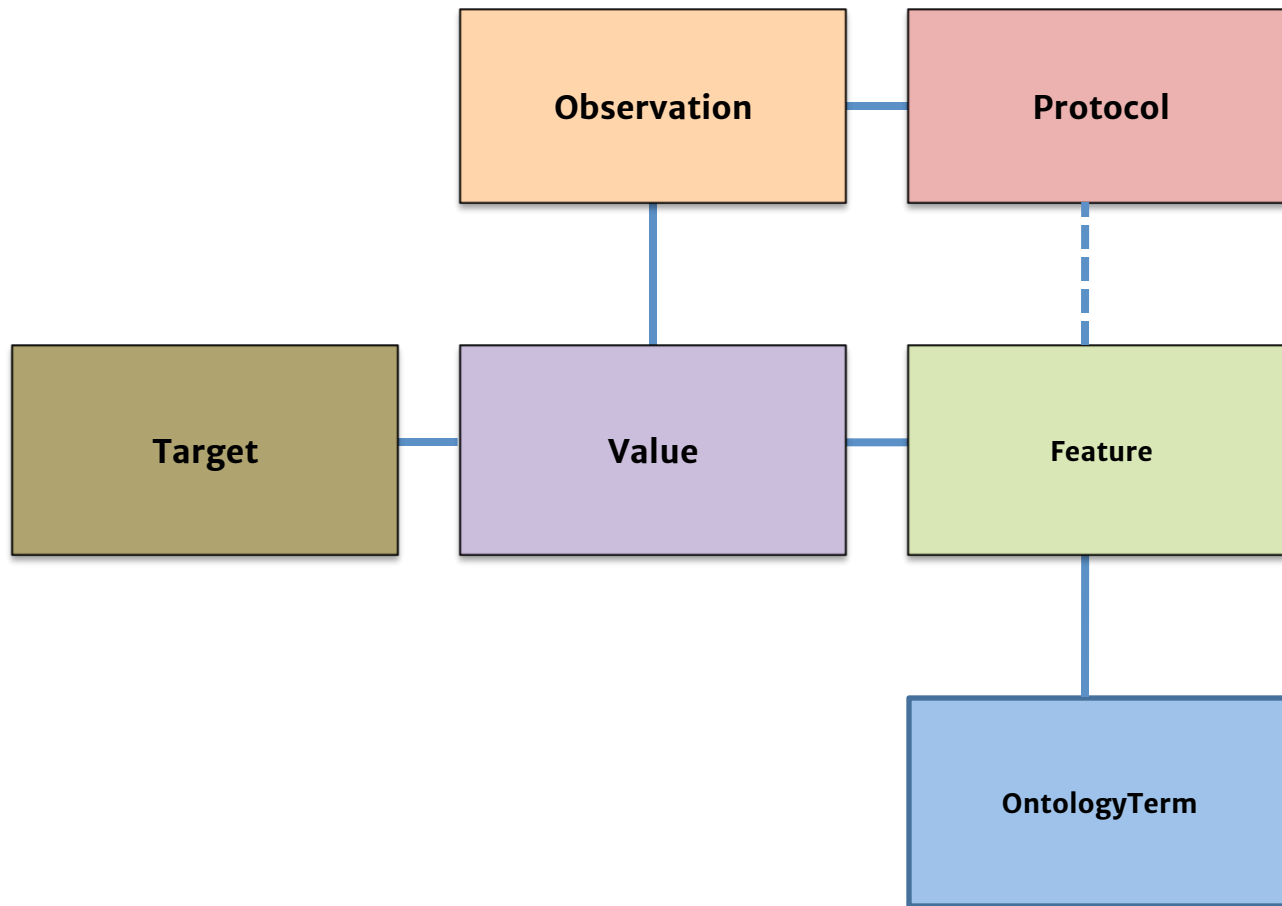
umcg



university of
 groningen

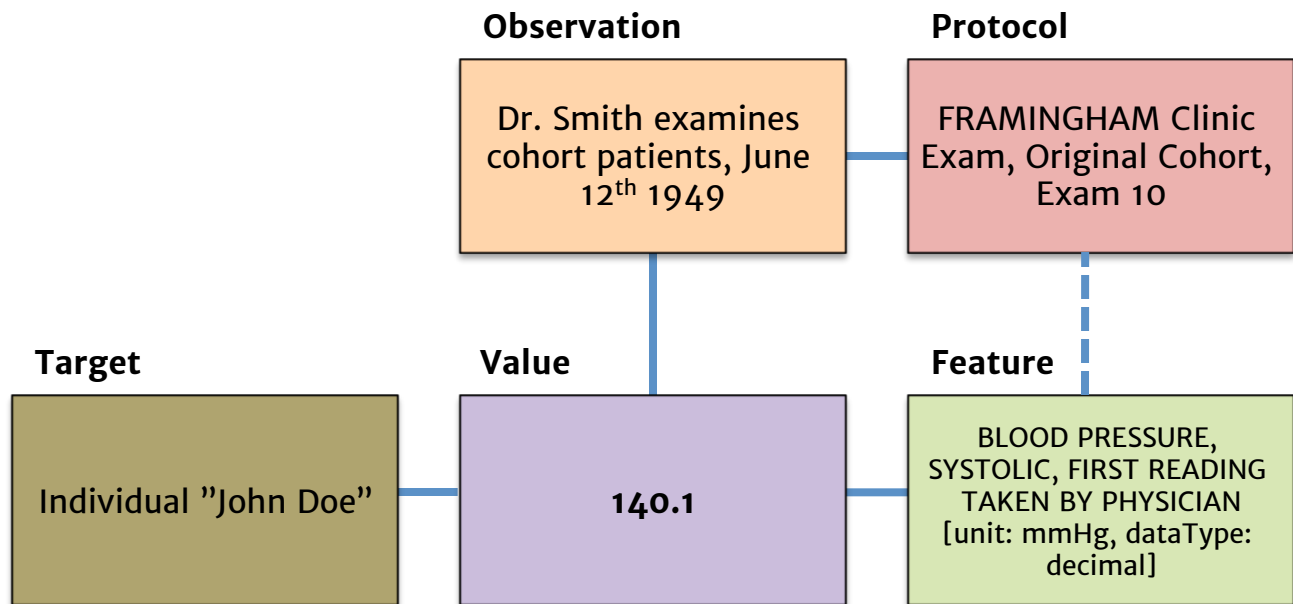
genomics coordination
 center

Observ-OM model for flexible columns/prov.



Adamusiak *et al.* (2012) *Human Mutation* 33(5):867-73

Observ-OM data example



	Negative	Positive	Doubtful	Unknown	
31	0	1	2	9	SUGAR IN URINE FC10
32	0	1	2	9	ALBUMIN IN URINE FC11
BLOOD PRESSURE (Left arm, mm Hg):					
33-38	Systolic FC12		Diastolic FC13		NURSE
4	FC14		FC15		PHYSICIAN (First reading)
45-50	FC16		FC17		PHYSICIAN (Second reading)
LUNG FUNCTION:					
51-52	FC188		TOTAL VITAL CAPACITY (Deciliters)		
53-55	FC189		FIRST SECOND VOLUME (Centiliters)		
GLUCOSE CHALLENGE:					

Ontology

MeSH: Blood Pressure

"PRESSURE of the BLOOD on the ARTERIES and other BLOOD VESSELS."

Adamusiak et al. (2012) *Human Mutation* 33(5):867-73

Evolution

MOLGENIS software

Swertz *et al*,
BMC Bioinf. (2010)

<http://www.molgenis.org>

XGAP model

Swertz *et al*,
Genome Biology (2010)

<http://www.xgap.org>

Observ-OM model

Adamusiak *et al*,
Human Mutation (2012)

<http://www.observe-om.org>

EB Registry

Van den Akker *et al*,
Human Mutation (2011)


<http://www.deb-central.org>

dystrophic eb 

xQTL workbench

Arends & van der Velde *et al*,
Bioinformatics (2012)

<http://www.xqtl.org>

XQTL WORKBENCH 

AnimalDB

Track and trace of animal life
events in research laboratories

<http://www.animaldb.org>

 AnimalDB

WormQTL

- Panacea project, *C. elegans* data
- ~300 million measurements

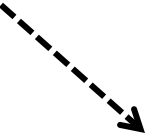
Snoek, van der Velde, Arends & Li *et al*,
Nucl. Acids Res. (2013)

<http://www.wormqtl.org>

CropQTL

Learning From Nature project,
arabidopsis thaliana data

- 1400 plants
- SNP genotypes (~70 million values)
- Classical traits, e.g. flowering time

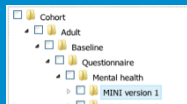
 ..more



One to rule all? **Observ-OMX**

Catalogue

Find data item and sample collections



Data

Filter individual data sets and download to Excel & SPSS

rs11050	rs11051	rs11052	rs11053	rs11054	rs11055
NSLU	-0.1892	-0.1892	0.2131	-0.8379	-0.9180
WSLU	0.0027	0.0027	0.0061	0.0298	0.0028
NSLU	0.0637	0.0637	0.2153	-0.1182	-0.1048
WSLU	0.0116	0.0116	0.1208	-0.1924	-0.1909
NSLU	0.0514	0.0514	0.1649	-0.1768	-0.1621
WSLU	0.0073	0.0073	0.0352	0.0060	0.0028
NSLU	-0.0528	-0.0528	-0.0248	0.0354	0.0405
WSLU	0.0029	0.0029	0.0108	0.0006	0.0002

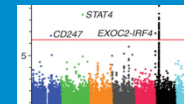
Compute

Run analysis workflows on big data compute infrastructure



GWAS Central

Explore summary level GWAS data

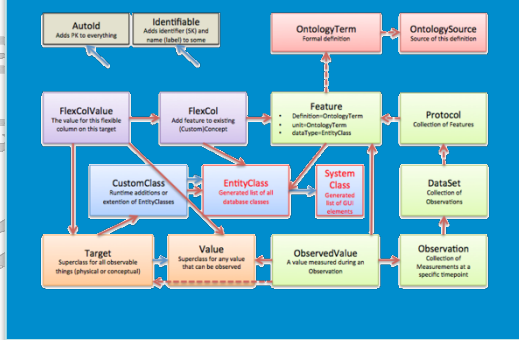


Protocol

CRFs, Questionnaires, Lab protocols, and assays

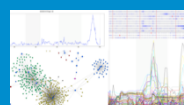


Core model



xQTL

Multi-omics association & visualization tools



NGS

Next-Generation Sequencing



XGAP

Multi-omics genotypes and phenotypes



Share

Friends, Groups and Permission management



Mutation

Explore genetic mutations and pathogenicity effects



Organization

Institutes, Departments, People, Locations & Containers



File

File storage and drivers for images and data

8-18	5P	83	08	47	14	10	00
0-1E	73	D9	02	HGD	36	11	04
4-20	A3	01	76	aPPaU	10	12	1
E-92	D8	AB	F4	bB3	8	11	4
8-9B	DC	99	62	w308	11	6	4
5-67	99	18	EC	o	11	92	0
8-5F	F4	CA	CB	4	94	1e	11
8-91	66	65	E6	.	19	34	11
8-5F	F0	82	EC	RCa	13	8	0

Outcome: working applications (e.g. xQTL)

Strains (panels) Chromosomes Markers Genes Measurements DerivedTraits Probes Samples

Probes

File Edit View 1 - 10 of 68,452

Search:

Micro-array probes blasted against WS220 and linked to the genes, used for gene expression phenotypes and eQTLs.

Click on to plot an item, and on to return to this list.

id	name	description	Investigation ontology	Reference	Alternative identifiers	label	Chromosome	cM
10505	WSU1	IA / SpotReport / last_match_NA / on_cumulative_bp_start_0	Public					
10505	WSU2	25A1.8 / cea2.c.02677 / last_match_60 / on_cumulative_bp_start_10184580	Public			clec-87		
10505	WSU3	21F3.6 / cea2.c.02677 / last_match_60 / on_cumulative_bp_start_4912043	Public			F21F3.6		
10505	WSU4	25H2.9 / cea2.c.02801 / last_match_60 / on_cumulative_bp_start_10567120	Public			pas-5		
10505	WSU5	56H1.4 / cea2.c.04344 / last_match_60 / on_cumulative_bp_start_5741975	Public					
10505	WSU6	06O01.1 / cea2.c.04508 / last_match_60 / on_cumulative_bp_start_7015970	Public			pdi-3		
10505	WSU7	20F10.2 / cea2.c.06048 / last_match_60 / on_cumulative_bp_start_10300315	Public			T20F10.2		

Stable!

Strains (panels) Chromosomes Markers Genes Measurements DerivedTraits Probes Samples

Markers

File Edit View 1 - 10 of 1,579

Genetic markers used in one or more of the populations stored in WormQTL. Click on to view one item, and on to return to this list.

id	name	description	Investigation ontology	Reference	Alternative identifiers	label	Chromosome	cM
10347	pkP1050	CR_non_cumulative_bp_pos_169018	Public					-18
10347	pkP1101	CR_non_cumulative_bp_pos_992189	Public					-17
10347	pkP1103	CR_non_cumulative_bp_pos_1881116	Public					-11
10347	pkP1052	CR_non_cumulative_bp_pos_2818974	Public					-6
10347	egPE107	CR_non_cumulative_bp_pos_3502476	Public					-3
10347	egPF101	CR_non_cumulative_bp_pos_4338254	Public					-1
10347	pkP1054	CR_non_cumulative_bp_pos_4845516	Public					

Stable!

Marker 1-5 of 121

Probe 1-10 of 23232

Stepsize 5

Width 5

Height 10

Change size

Probe	pkP1050	pkP1101	pkP1103	pkP1052	egPE107
WSU1	-0.1892	-0.1892	0.231	-0.8379	-0.9186
WSU2	-0.012	-0.012	0.1026	-0.2283	-0.4022
WSU3	0.0637	0.0637	0.2153	-0.1182	-0.1068
WSU4	0.0136	0.0136	0.1208	-0.1924	-0.1909
WSU5	0.054	0.054	0.1649	-0.1768	-0.1621
WSU6	0.0873	0.0873	0.1452	-0.0902	-0.0322
WSU7	-0.0529	-0.0529	-0.0248	0.0354	0.0405
WSU8	0.0629	0.0629	0.1488	-0.06	-0.1015
WSU9	0.0421	0.0421	0.0923	-0.254	-0.0614
WSU10	-0.0099	-0.0099	0.056	-0.0829	-0.0386

Dynamic!

Dystrophic EB Register (COL7A1)

Home Catalogue Dataset Viewer **Data Explorer** Upload Converters Admin My Account

Data Explorer

Search data items

Choose a dataset:

Data item filters

Data item selection

- Patients_protocol
 - Patient ID
 - Phenotype
 - cDNA change 1
 - Protein change 1
 - Exon/Intron 1
 - Consequence 1
 - cDNA change 2
 - Protein change 2
 - Exon/Intron 2
 - Consequence 2
 - PubMed ID

Patient ID	Phenotype	cDNA change 1	Protein change 1	Exon/Intron 1	Consequence 1	cDNA change 2	Protein change 2	Exon/Intron 2	Consequence 2
P37	RDEB, severe generalized	c.3G>T	p.0	Exon 1	pathogenic	c.3G>T	p.0	Exon 1	pathogenic
P292	RDEB, severe generalized	c.3G>A	p.0	Exon 1	pathogenic	c.353delGinsCCCCCTTGCAA	p.Arg118ProfsX14	Exon 1	pathogenic
P172	RDEB, unknown	c.3G>T	p.0	Exon 1	pathogenic	c.448G>A	p.Gly150Arg	Exon 1	pathogenic
P34	RDEB, severe generalized	c.3G>T	p.0	Exon 1	pathogenic	c.4997dupG	p.Pro1668AlafsX4	Exon 1	pathogenic
P168	RDEB, unknown	c.58C>T	p.Arg20Ter	Exon 1	pathogenic	c.8245_8246insC	p.Gly2749AlafsX41	Exon 1	pathogenic
P163	RDEB, unknown	c.62dupT	p.Arg22AlafsX17	Exon 1	pathogenic				
P177	RDEB, severe generalized	c.111_112insA	p.Asp38ArgfsX50	Exon 2	pathogenic	c.497dupA	p.Val168GlyfsX12	Exon 2	pathogenic
P367	RDEB, generalized other	c.112G>T	p.Asp38Tyr	Exon 2	pathogenic	c.2157G>A	p.Trp719Ter	Exon 2	pathogenic
P604	RDEB, pretibial	c.130G>A	p.Asp44Asn	Exon 2	pathogenic	c.8569G>T	p.Glu2857Ter	Exon 2	pathogenic
P213	RDEB, severe generalized	c.143C>T	p.Ser48Phe	Exon 2	pathogenic	c.3625_3635del	p.Ser1209LeufsX6	Exon 2	pathogenic
P1	RDEB, pruriginosa	c.151C>G	p.Arg51Gly	Exon 2	pathogenic	c.7474C>T	p.Arg2492Ter	Exon 2	pathogenic
P2	RDEB, severe generalized	c.238G>C	p.Ala80Pro	Exon 2	pathogenic	c.3631C>T	p.Gln1211Ter	Exon 2	pathogenic
P435	RDEB, severe generalized	c.267-3C>G		IVS2	pathogenic	c.267-3C>G		IVS2	pathogenic
P240	RDEB, pretibial	c.267-1G>C		IVS2	pathogenic	c.5096C>T	p.Pro1699Leu	Exon 2	pathogenic
P110	RDEB, severe generalized	c.313dupC	p.Arg105ProfsX5	Exon 3	pathogenic	c.5047C>T	p.Arg1683Ter	Exon 3	pathogenic
P433	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic
P434	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic
P267	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.3277-1G>C		IVS2	pathogenic
P268	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.3277-1G>C		IVS2	pathogenic
P49	RDEB, generalized other	c.344dupG	p.Asn116GlnfsX13	Exon 3	pathogenic	c.6082G>A	p.Gly2028Arg	Exon 3	pathogenic

« 1 2 3 4 5 6 ... 158 »

1316 data items found

Dystrophic EB Register (COL7A1)

Phenotypic details for patient 'P10'

target

Characteristics

Age	20
Gender	m
Ethnicity	unknown
Deceased	yes
Cause of death	
MMP1 allele 1	
MMP1 allele 2	

value

feature

Protocol Application

Cutaneous

Blistering	yes
Location	generalized
Hands	unknown
Feet	unknown
Arms	unknown
Legs	unknown
Proximal body flexures	unknown
Trunk	unknown
Mucous membranes	yes
Skin atrophy	yes
Milia	unknown
Nail dystrophy	yes

WormQTL.org (C. Elegans)



WormQTL

[Home](#) | [Contact](#) | [About](#) | [Article in NAR](#)

[Home](#) [Find QTLs](#) [Genome browser](#) [Browse data](#) [Help](#)

provide feedback:

Home

WormQTL – Public archive and analysis web portal for natural variation data in *Caenorhabditis* spp.

WormQTL is an online scalable system for QTL exploration to service the worm community. WormQTL provides many publicly available datasets and welcomes submissions from other worm researchers.



[Find QTLs](#)



[Genome browser](#)

id	chr	start	end	score	trait
NS01	1	100000	100000	0.0000	0.0000
NS02	1	100000	100000	0.0000	0.0000
NS03	1	100000	100000	0.0000	0.0000
NS04	1	100000	100000	0.0000	0.0000
NS05	1	100000	100000	0.0000	0.0000
NS06	1	100000	100000	0.0000	0.0000
NS07	1	100000	100000	0.0000	0.0000
NS08	1	100000	100000	0.0000	0.0000
NS09	1	100000	100000	0.0000	0.0000
NS10	1	100000	100000	0.0000	0.0000

[Browse data](#)



[Help](#)

What can you do?

- I want to search (e)QTLs for my trait or gene
 1. Go to [Find QTLs](#)
 2. Type the name or identifier of your trait or gene and press *Search*
 3. Put any relevant hits in the shopping cart
 4. Click *Plot cart* now and explore the results
- I want to know which genes have a QTL on my favourite position
 1. Go to [Genome browser](#)
 2. Add tracks from experiments of interest
 3. Navigate to your favourite location (tip: use *open in new window*)
 4. Collect significant probe identifiers from that region
 5. Use the identifiers to do a search with [Find QTLs](#)

WormQTL.org (C. Elegans)

Phenotypes	Type of array	Sample size	Parental strains	Reference	Pubmed link	Growing temperature	Stage	Food	Medium	Dataset IDs
Gene expression	Washington State University	2x40 RILs	CB4856; N2	Li et al. 2006; Mapping determinants of gene expression plasticity by genetical genomics in <i>C. elegans</i> .	17196041	16oC and 24oC	(72h at 16 and 40h at 24); L4	OP50	NGM Plate	37 , 38
Gene expression	Affymatrix tiling array	60 RILs	CB4856; N2	Li et al. 2010; Global genetic robustness of the alternative splicing machinery in <i>Caenorhabditis elegans</i> .	20610403	24oC	(40h) L4	OP50	NGM Plate	n/a
Gene expression	Washington State University	36x3 RILs	CB4856; N2	Vinuela & Snoek et al. 2010; Genome-wide gene expression regulation as a function of genotype and age in <i>C. elegans</i> .	20488933	24oC	(40h, 96h and 214h) L4, Adult, Old	OP50	NGM Plate	3 , 5 , 6 , 7 , 8 , 9 , 10 , 11 , 12 , 13 , 14 , 15 , 16 , 17 , 18 , 19 , 20 , 21
Gene expression	Agilent 4x44k microarrays	208 RIALs	CB4856; N2	Rockman et al. 2010; Selection at linked sites shapes heritable phenotypic variation in <i>C. elegans</i> .	20947766	20oC	YA	OP50	NGM Plate	22 , 34 , 35 , 36
Feeding curves RNAi exposure	n/a	56 RILs * 12 RNAi	CB4856; N2	Elvin & Snoek et al. 2011; A fitness assay for comparing RNAi effects across multiple <i>C. elegans</i> genotypes.	22004469	20oC	Multi-generational	n/a	Liquid S-medium	24 , 32 , 33
Life-history traits	n/a	80 RILs	CB4856; N2	Gutteling et al. 2007; Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in <i>Caenorhabditis elegans</i> .	16955112	12oC and 24oC	Egg, L4, YA	OP50	NGM Plate	25 , 26 , 27
Lifespan and pharyngeal-pumping	n/a	90 NILs	CB4856; N2	Doroszuk et al. 2009; A genome-wide library of CB4856/N2 introgression lines of <i>Caenorhabditis elegans</i> .	19542186	20oC	All; synchronised	OP50	NGM Plate	4 , 23 , 28 , 29 , 30 , 31
Lifespan, Recovery and reproduction after heat-shock	n/a	58 RILs	CB4856; N2	Rodriguez et al. 2012; Genetic variation for stress-response hormesis in <i>C. elegans</i> lifespan.	22613270	20oC and 35oC heat-shock	L4 and Adult	OP50	NGM Plate	39 , 40
Gene expression	Washington State University	CB4856 and N2	CB4856; N2	Vinuela & Snoek et al. 2012; Aging Uncouples Heritability and Expression-QTL in <i>Caenorhabditis elegans</i> .	22670229	24oC	(40h, 96h and 214h) L4, Adult, Old	OP50	NGM Plate	41 , 42 , 43

WormQTL.org (C. Elegans)

Home Find QTLs Genome browser Browse data Help provide feedback

Find QTLs

All data (175,366) Search Reset

(for measurement (43) 5B4BR, K02B12, WBGene00021562, WBGene00006727, acetylcholine, luciferase ...)
biology or anatomy terms, will show the probes and related terms for that gene.

Panel (500)
Gene (47,360)
Transcript (55,782)
Chromosome (8)
Probe (68,452)
Sample (1,630)
DerivedTrait (12)

View (9) Clear cart Plot cart now

Found n

Your results were limited to the first 100. Please be more specific.

Add all hits to cart

Add to cart	Probe AGIUSA14764 / clc-2 reports for WBGene00000523 - WormBase C01C10.1 / C01C10.1 / wb C01C10.1 / non_cumu_bp_start_743339 ...more	Ontologies
Add to cart	Probe AGIUSA16119 / ceh-1 reports for WBGene00000428 - WormBase F16H11.4 / F16H11.4 / wb F16H11.4 / non_cumu_bp_start_465353 ...more	Ontologies
Add to cart	Probe AGIUSA19594 / acn-1 reports for WBGene00000039 - WormBase peptidase [C42D8.5.2] / C42D8.5.2 / C42D8.5 / wb C42D8.5.2 w ...more	Ontologies
Add to cart	Probe AGIUSA41433 / gsp-1 reports for WBGene00001747 - WormBase serine/threonine protein phosphatase [F29F11.6.1] / F29F11.6 ...more	Ontologies
Add to cart	Probe AGIUSA5476 / cul-2 reports for WBGene00000837 - WormBase	Ontologies

Ontologies

Ontological terms

- GO:0016021-integral to membrane
- GO:0005198-structural molecule activity
- GO:0005923-tight junction
- WBbt:0005733-hypodermis
- WBbt:0005753-seam cell

WormQTL.org (C. Elegans)

Home Find QTLs Genome browser Browse data Help provide feedback

Find QTLs

All data (175,366) daf Search Reset

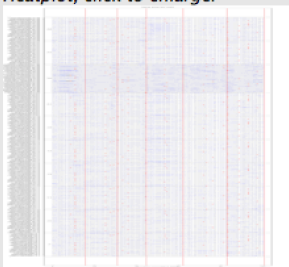
(for example: ctl, daf, pgp-7, gst-27, Y65B4BR, K02B12, WBGene00021562, WBGene00006727, acetylcholine, luciferase ...)
Gene hits, for example on [Geno Ontology](#) or anatomy terms, will show the probes and related terms for that gene.

View hits (91) View cart (91) Clear cart Plot cart now

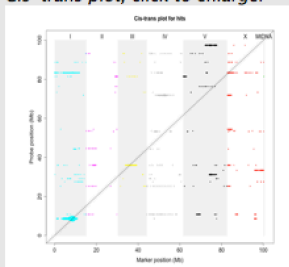
Results for my selected hits:

(get a [permanent link to these results](#))

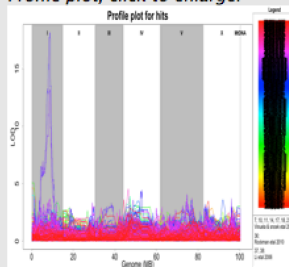
Heatplot, click to enlarge:



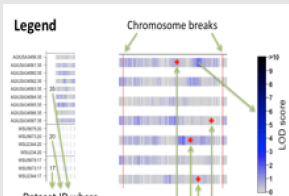
Cis-trans plot, click to enlarge:



Profile plot, click to enlarge:



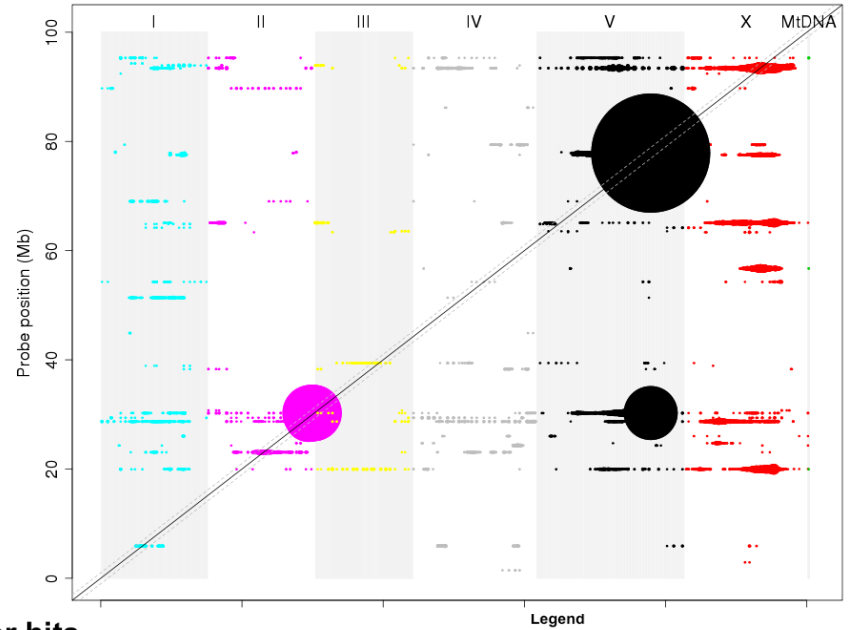
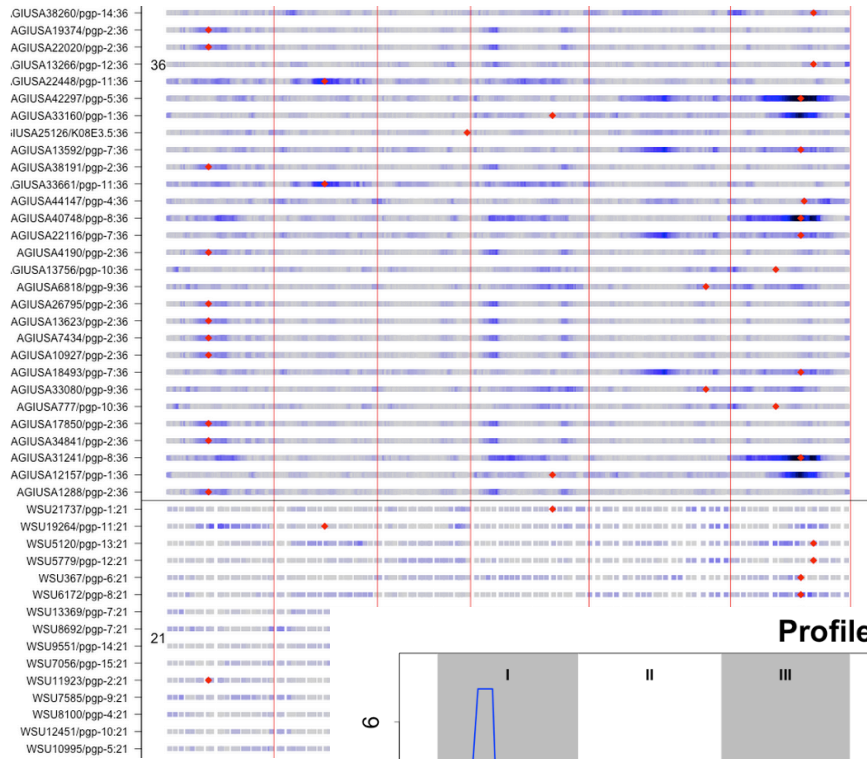
Legend, click to enlarge:



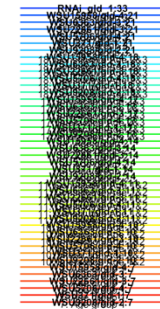
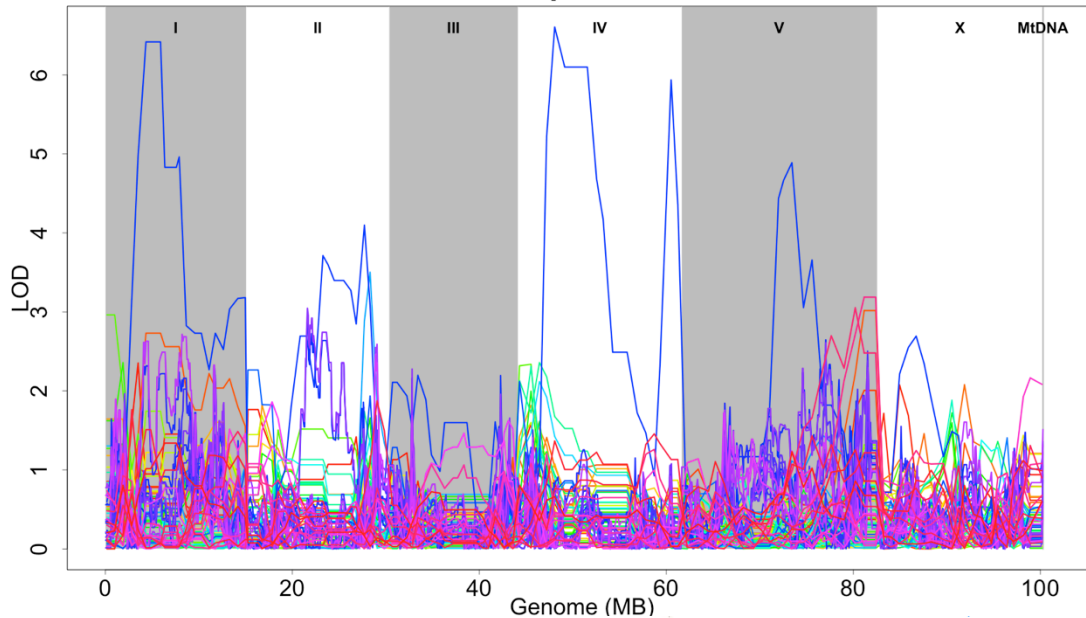
More downloads:

- Get the [Cytoscape network](#) for this plot. ([how-to import](#))
- Get the [Cytoscape nodes](#) for this plot. ([how-to import](#))
- Note: includes **significant results only**. (LOD > 3.5)
- Save both files. Import network (has LOD scores), then node attributes (chrom, bploc, dataset). [Example visualization](#)
- Get the generated [source data](#) for these plots.
- Get the generated [multiplot plot R script](#).
- Get the generated [cistrans R plot script](#).
- Get the generated [profile R plot script](#).

WormQTL.org (C. Elegans)

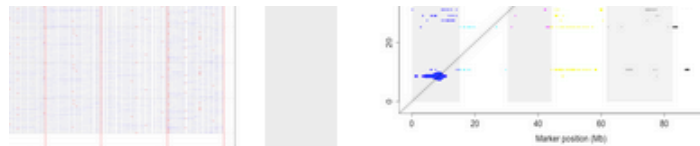


Profile plot for hits



nics coordination
r

WormQTL.org (C. Elegans)

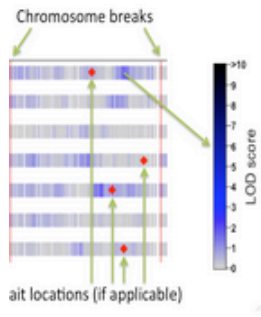


Click to enlarge:

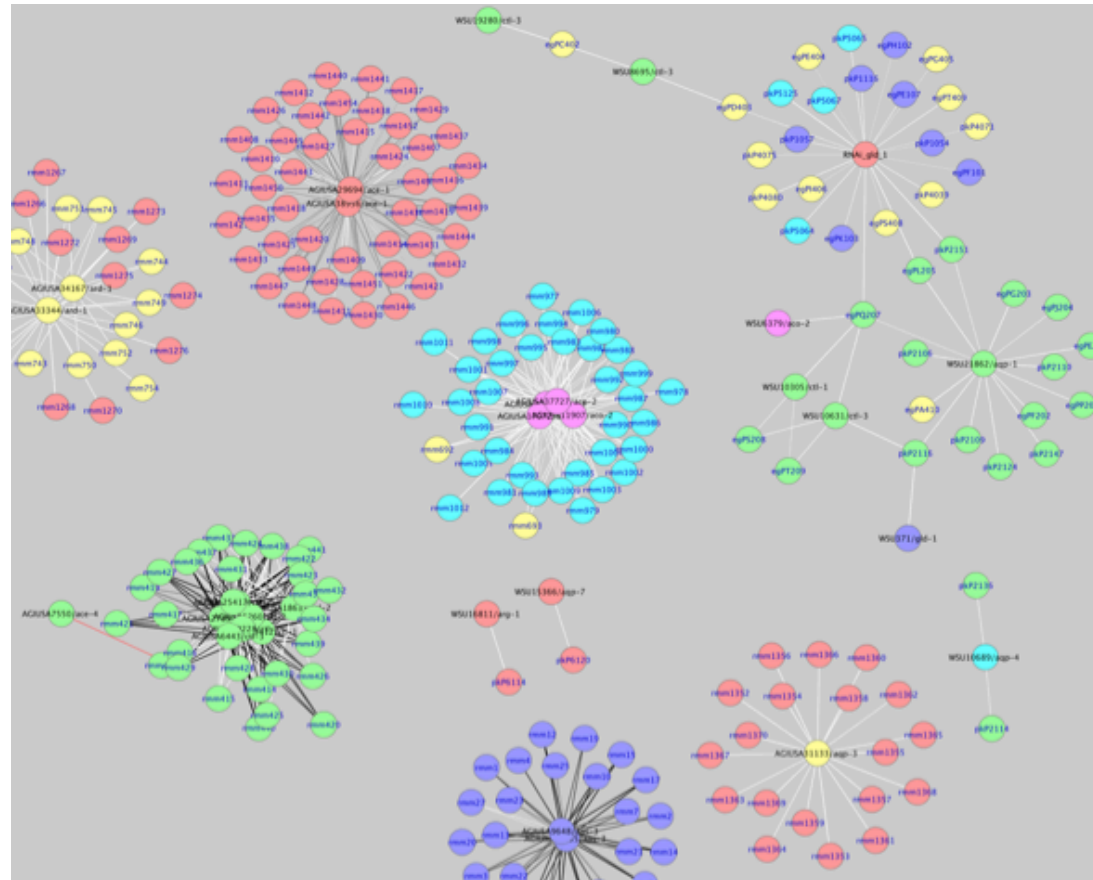
More downloads:

Get the [Cytoscape network](#) for the QTL
 Get the [Cytoscape nodes](#) for the QTL
 Note: includes significant results
 Save both files. Import network into Cytoscape
 attributes (chrom, bploc, data)

Get the generated [source data](#)
 Get the generated [multiplot plot](#)
 Get the generated [cistrans R plot](#)
 Get the generated [profile R plot](#)



[/ daf-1 \[explore deeper\]](#) - protein kinase [F29C4.1b] / F29C4.1
[/ daf-3 \[explore deeper\]](#) - F25E2.5b.3 / F25E2.5 / wb|F25E2.5b.3
[/ af-11 \[explore deeper\]](#) - R0240.3 / cea2 p 107079 / bla|af-11



PlantQTL (Arabidopsis & Brassica, pilot)



Home | Contact | About | Logged in as: admin | Logout

Home Find QTLs Genome browser Browse data Help Upload data Run QTL mapping Configure analysis

report Utilities Admin **Navigate investigations (studies)**

File Edit View <<< 1 of 1 >>>

name * Brassica_Nutrigenomics

Show additional fields

Experiment overview Investigation overview **Datasets** Identifiers

Navigate datasets

File Edit View <<< 14 of 15 >>>

name * gene_quant

description

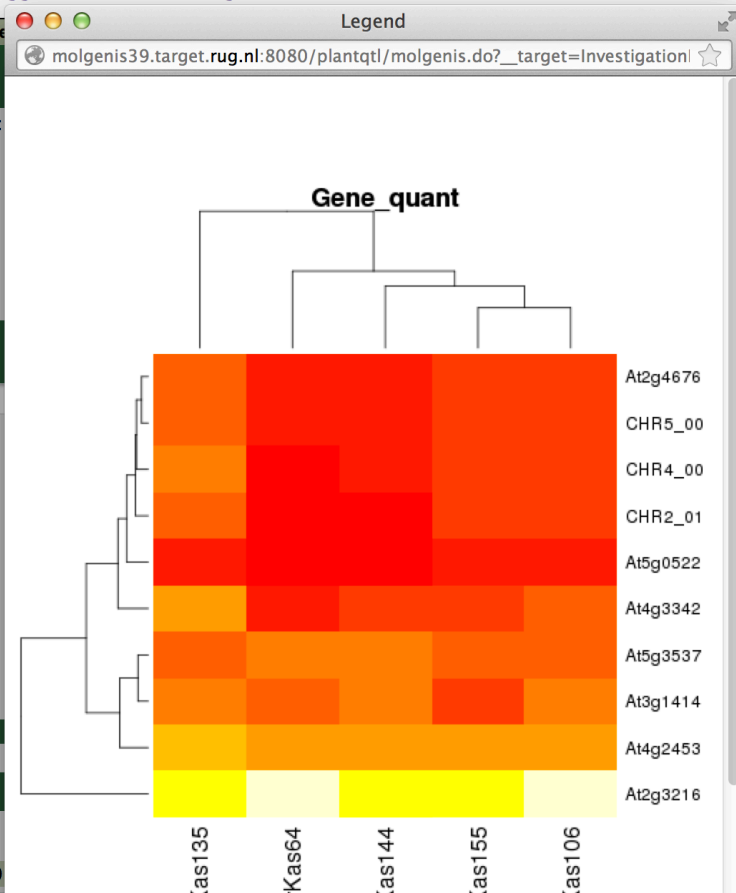
Show additional fields

View Inspect Delete

View

<<< Individual 1-5 of 146 >>>

	LerKas144	LerKas135	LerKas155
Probe			
1-10 of			
36831			
CHR2_019408191to019408250	2.99910328488402	4.44183051724418	3.8157182229
At2g32160	6.87071885817011	7.05370937232414	7.24146941173863
CHR5_007276699to007276758	3.41002880765293	4.49908296230188	3.73788077390656
At4g33420	4.16636737500485	5.51868558280554	3.74922028206199
CHR4_008348949to008349008	3.5669092070861	5.09362408610677	4.0993738832211
At3g14140	4.99742874302464	4.76907448580219	4.08007660092876
Stepsize 5			



GoNL (Genome of the Netherlands) variants



Find GoNL variants

Find GoNL variants

Chromosome

3

Start bp position

20000000

Stop bp position

21000000

Find

Results

Chr	BpPos	RsID	Ref	Alt	Qual	Info
3	20000025	.	G	A	1160.83	AC=4;AN=996;GTC=494,4,0
3	20000196	.	C	T	222.78	AC=1;AN=996;GTC=497,1,0
3	20000228	rs62241288	G	C	77580.26	AC=214;AN=996;GTC=309,164,25
3	20000269	.	G	T	478.83	AC=1;AN=996;GTC=497,1,0
3	20000403	.	C	A	197.56	AC=1;AN=996;GTC=497,1,0
3	20000418	.	G	A	172.12	AC=2;AN=996;GTC=496,2,0
3	20000558	.	T	C	244.53	AC=3;AN=996;GTC=495,3,0
3	20000644	.	T	G	60.28	AC=1;AN=996;GTC=497,1,0
3	20000773	.	A	G	82.27	AC=1;AN=996;GTC=497,1,0
3	20000875	.	A	T	4591.63	AC=29;AN=996;GTC=472,23,3
3	20000944	rs6774670	A	G	138.22	AC=1;AN=996;GTC=497,1,0
3	20000971	.	T	C	2927.61	AC=22;AN=996;GTC=479,16,3
3	20001194	rs6550237	T	A	147218.08	AC=767;AN=994;GTC=34,159,304
3	20001262	.	G	A	378.88	AC=3;AN=996;GTC=495,3,0
3	20001309	rs6550238	C	G	227373.71	AC=979;AN=996;GTC=1,15,482
3	20001318	rs6550239	G	A	206430.74	AC=775;AN=996;GTC=31,159,308

www.nlgenome.nl

Posters @ BOSC Poster area

OmicsConnect: flexible multi-omics data capture and integration tools for high-throughput biology

Challenge

Recent advances in molecular characterization generate many high-throughput multi-omics data for biomedical research. Integrating and analyzing these large multi-dimensional data in an effective and reproducible way across experiments and data types is an ever increasing challenge.

Observ-OMX platform

Here we present the next generation in Observ-OMX, a data model that captures all data modalities and context information in a flexible and powerful way. "Best of practices from existing domain-specific models are harmonized and re-used via the core model. Observ-OMX captures the essential features, protocols, targets and values of any experimental setup without sacrificing details or strong data types.

OmicsConnect toolbox

We created OmicsConnect using the Observ-OMX model, a software toolbox for labs and consortia to import, catalogue, manage, query, and analytically interface with large multi-omics data and complex phenotypes in biobanks, translational medicine, epidemiology, genetics and medical organism studies.

WormQTL.org

www.molgenis.org

k.j.van.der.velde@umcg.nl

Big data analysis made easy

Using Molgenis Compute

Motivation

MOLGENIS Compute is a toolbox to configure and run parallel bioinformatic pipelines, actively used for DNA resequencing, genotype imputations with QTL and GWAS pipelines underway. It was first used in the Genome of the Netherlands (GoNL), a nationwide BBMRI-HL project of 769 whole genome sequence samples (12x). All those needed alignment and variant calling, a daunting challenge: input was 2250/50 TB * fq files; analysis required >30 analysis steps; the complete analysis was >50 000 jobs accumulating 200,000 hours.

Light weight and full control

Step 1: Design

WORKFLOW_CSV

step	script	parameterMapping
step1	assessRisk.sh	sample_name, genome_build
step2	report.sh	name_disease, step1_out

Step 2: Generate

sample	glucose	disease
patient1	5.6	diabetes
patient2	7.8	diabetes
patient3	12.3	diabetes

Step 3: Run

Run on clusters and GRID, ...

Platform independent

resources	Generated script
file management	assessRisk_0.sh
tool management	assessRisk_1.sh
variable list	assessRisk_2.sh
file management	report_0.sh

Local, PBS, GRID, SGE, easy to add new computational infrastructure

<http://molgenis.org/wiki/ComputeStart>

Towards a system for semi-automatic matching of biobank variables using ontology terms

1. Motivation

Newly developed prediction models need to be tested for their performance with new datasets. Common practice is to duplicate one cohort at a time because it is difficult to get data access and to harmonize data variables.

2. Challenge

Manually searching for candidate variables across studies is very time intensive because there are usually thousands of data items described using different terminologies.

3. System

We prototyped a semi-automated system to shortlist candidate data variables from cohorts that could be mapped to prediction model parameters. We implemented three steps:

- Step one: annotate predictors
- Step two: expand predictors
- Step three: map parameters

4. Future work

- This semi-automatic matching system could become a pre-processing DataShaper to speed harmonization.
- The system will be evaluated against manually generated variable matches using current DataShaper mappings as the gold standard.
- We will explore if annotation of parameters (step 1) can be automated since manual work is not desirable when searching long lists of parameters.

OmicsConnect

Compute

BiobankConnect

Acknowledgements

Robert K. Hastings

Charalambos Chrysostomou

Chao Pang

Dennis Hendriksen

Anthony J. Brookes

Martijn Dijkstra

Despoina Antonakaki

Tomasz Adamuziak

Sirisha Gollapudi

Gudmundur Thorisson

Myles Byrne

David van Enckvort

Linda Mook

Ger Strikwerda

Danny Arends

Roan Kanninga

Jan Bot

George Byelas

Erwin Winder

Yang Li

Konrad Zych

And more ...

Pieter Neerincx

Basten Snoek

Noortje Festen

Lude Franke

Juha Muilu

Helen Parkinson

Vincent Ferreti

Gert-Jan van Ommen

Bart Charbon

Jan Jurjen Uitterdijk

Ritsert C. Jansen

Jan Kammenga

Paul de Bakker

Irene Nooren

Rob Hooft

Salome Scholtens

Hans Hillege

Ronald Stolk

Cisca Wijmenga

And more ...

NBIC/BioAssist consortium (bioinfo)

BBMRI-NL catalogue group(Hs)

CTMM/TraIT consortium (Hs)

EU-GEN2PHEN consortium (Hs)

EU-PANACEA consortium (Ce)

EU-BioSHARE consortitum (Hs)

EU-CASIMIR consortium (Mm)

EU-BioMedBridges cosortium (all)

NL Brassica Nutr. consortium (At)

Learning from Nature (At)

LifeLines (Hs)

TIFN (Hs)

BigGrid (info)

Target + CIT (info)

And more...



B B M R I • N L

BiG Grid
the dutch e-science grid



TIFOOD
NUTRITION



Wrap-up

Summary

- MOLGENIS software generator
- Exploiting bio data requires structure
- *Best-of* flexible and stable components
- OmicsConnect as modular platform of apps

Read more

- MOLGENIS: <http://www.molgenis.org>
- xQTL: <http://www.xqtl.org>
- Adamusiak *et al* (2011) *BMC Bioinformatics*
- Akker *et al* (2011) *Human Mutation*
- Arends *et al* (2010) *Bioinformatics* 26: 2990-2992
- Brandsma *et al*, *Norsk Epidemiologi* 2012
- Snoeks *et al* (2013) *Nucleic Acids Res*
- Swertz *et al* (2010) *Genome Biology* 9;11(3): R27.
- Smedley *et al* (2008) *Briefings in bioinformatics* 9(6):532-44.
- Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

Thank you!
Questions?

k.j.van.der.velde@umcg.nl

molgenis
Your database at the push of a button

github.com/molgenis