

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



university of
groningen

genomics coordination
center

Genomics Coordination Center, UMCG, Groningen

Biobanking

(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



Looking for colleague(s) ☺

Large scale multi-omics

Human and model organism research: genotype 2 phenotype

MOLGENIS: Motivation – building many apps

Researcher needs



NextGenSeq data

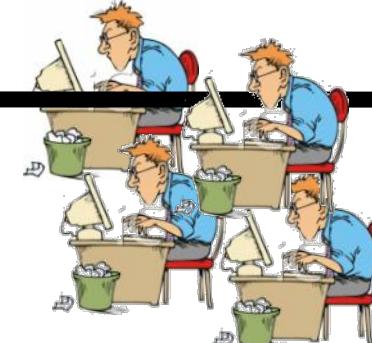
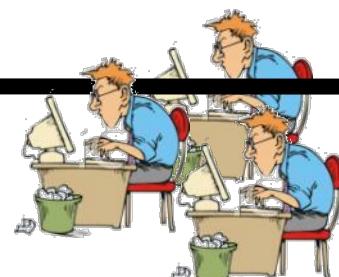
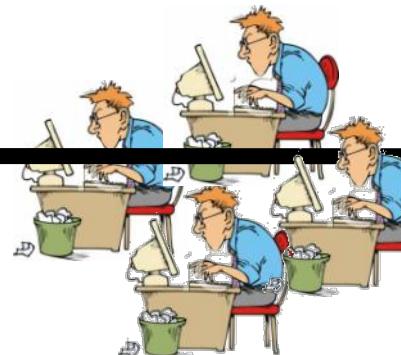


Mutation data



Model organisms data

Work very hard



Use

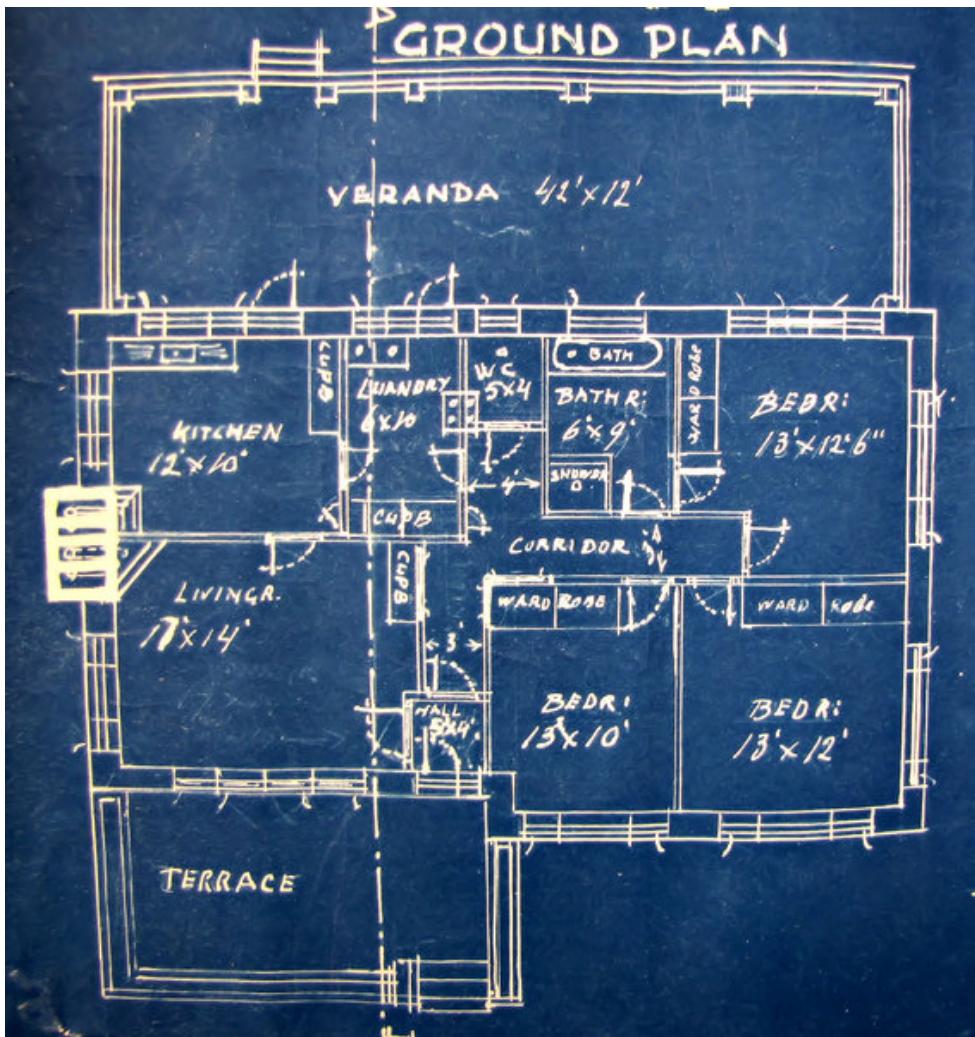
Solexa Sequencer LIMS

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

database of COL7A1 mutations

Animal Observatory

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



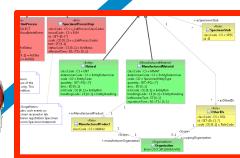
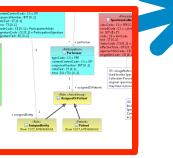
NextGenSeq



Mutation database



Model organisms



Use generated software

Solexa Sequencer LIMS

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

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

B B M R I • N L

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)					
	<input type="checkbox"/>							

* = 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

Search or type a command Explore Gist Blog Help joerivandervelde Edit molgenis's Profile

Repositories Members Find a Repository... All Public Private Sources Forks Mirrors

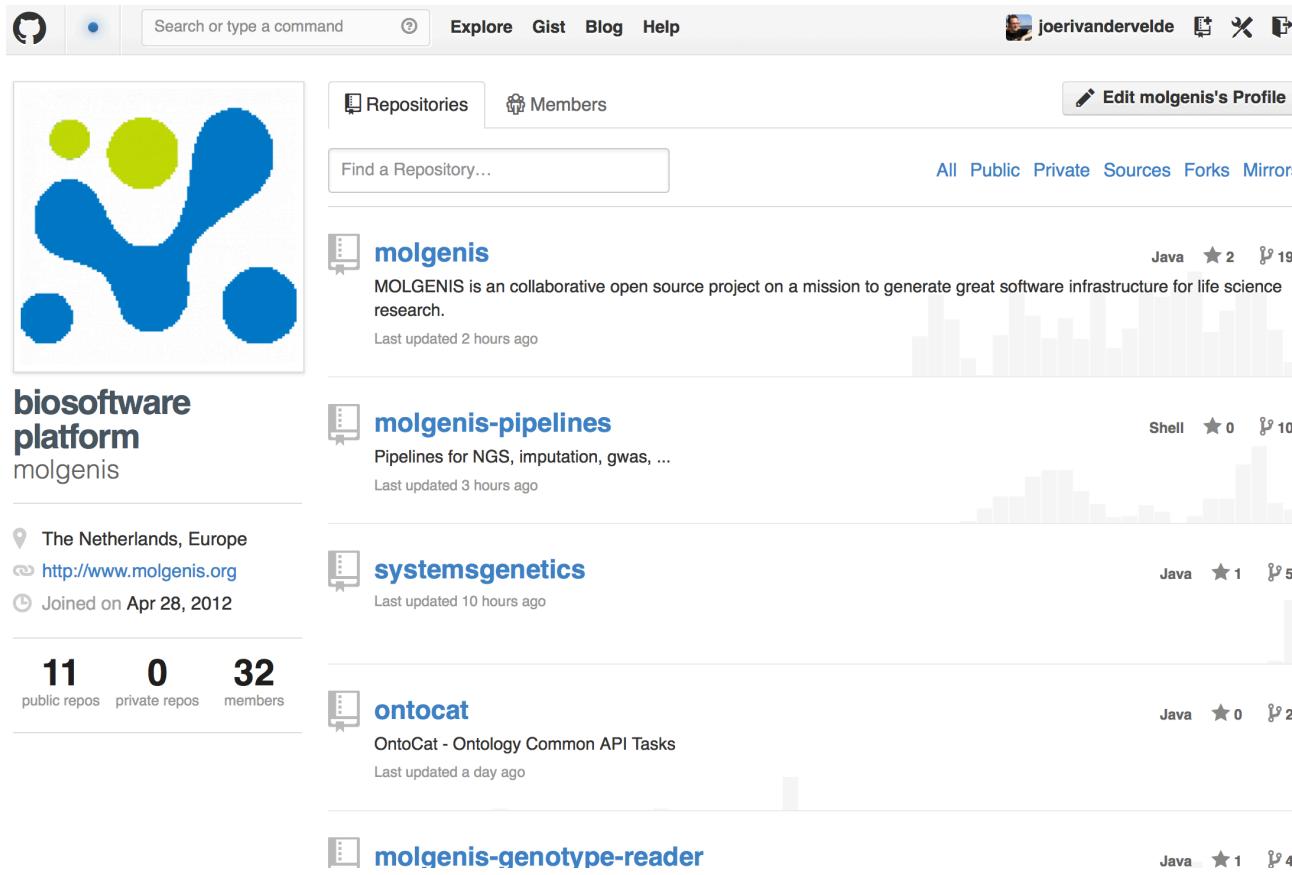
molgenis Java ★ 2 ⚡ 19
MOLGENIS is an collaborative open source project on a mission to generate great software infrastructure for life science research.
Last updated 2 hours ago

molgenis-pipelines Shell ★ 0 ⚡ 10
Pipelines for NGS, imputation, gwas, ...
Last updated 3 hours ago

systemsgenetics Java ★ 1 ⚡ 5
Last updated 10 hours ago

ontocat Java ★ 0 ⚡ 2
OntoCat - Ontology Common API Tasks
Last updated a day ago

molgenis-genotype-reader Java ★ 1 ⚡ 4



~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

Java



TM



JPA / MySQL



elasticsearch.



Continuous integration & automated deploy

Hudson

Hudson

search log in
ENABLE AUTO REFRESH

People

Build History

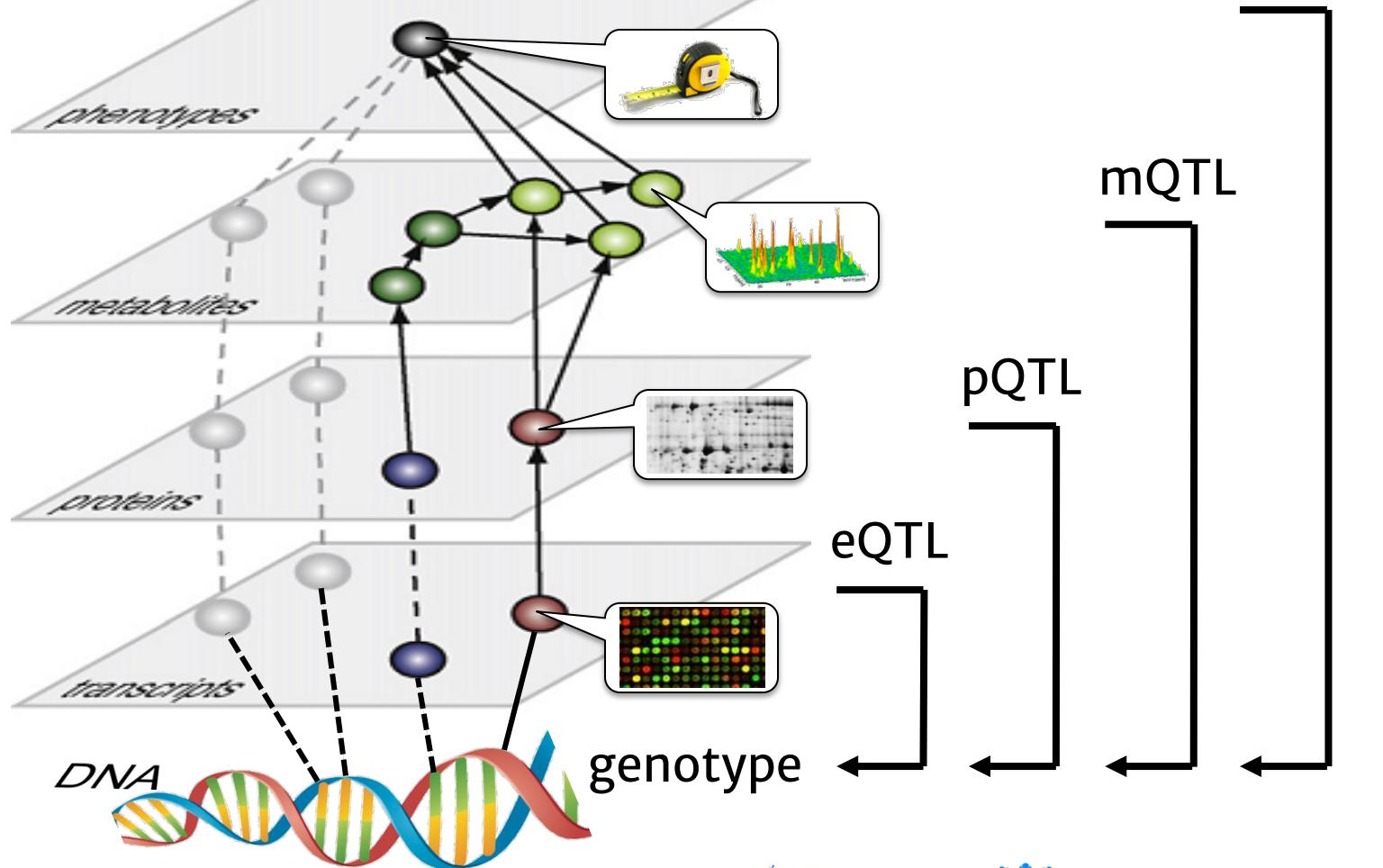
Build Queue
No builds in the queue.

Build Executor Status
Status 1/4
Building molgenis_designgg
#7

Jobs Status

All	Deploy acceptance	Deploy production	Deploy test	Molgenis	Molgenis (Legacy)	Monitor	dash	systemsgenetics
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		
		molaenis-search		6 hr 31 min (#27)	N/A	52 sec		

Motivation: Understanding geno-to-pheno



Situation: Many types and flavours of data

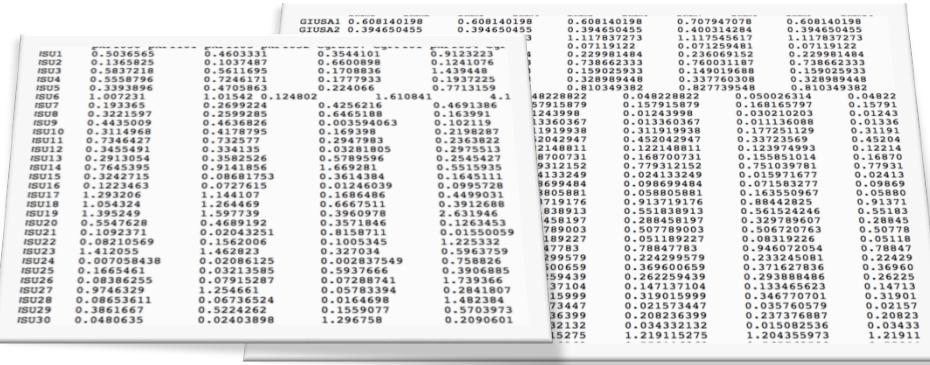
Genomic features, individuals, ontologies ..

SampleID	Sample_label	Description	File
GEO000000322	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000324	GO_0016023	integral to membrane / GO:0016023	1
GEO000000325	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000327	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000328	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000329	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000330	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000331	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000332	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000333	GO_0016023	integral membrane protein / GO:0016023	1
GEO000000337	GO_000792	embryo development ending in birth or egg I	1
GEO000000339	GO_000792	embryo development ending in birth or egg I	1
GEO000000340	GO_000792	embryo development ending in birth or egg I	1
GEO000000341	GO_000792	embryo development ending in birth or egg I	1
GEO000000342	GO_000792	embryo development ending in birth or egg I	1
GEO000000343	GO_000792	embryo development ending in birth or egg I	1
GEO000000344	GO_000792	embryo development ending in birth or egg I	1
GEO000000345	GO_000792	embryo development ending in birth or egg I	1
GEO000000346	GO_000792	embryo development ending in birth or egg I	1
GEO000000347	GO_000792	embryo development ending in birth or egg I	1
GEO000000348	GO_000792	embryo development ending in birth or egg I	1
GEO000000349	GO_000792	embryo development ending in birth or egg I	1
GEO000000350	GO_000792	embryo development ending in birth or egg I	1
GEO000000351	GO_000792	embryo development ending in birth or egg I	1
GEO000000352	GO_000792	embryo development ending in birth or egg I	1
GEO000000353	GO_0006468	calcium ion binding / GO:0006468	1
GEO000000354	GO_0006468	calcium ion binding / GO:0006468	1
GEO000000355	GO_0006468	calcium ion binding / GO:0006468	1
GEO000000356	GO_0006468	calcium ion binding / GO:0006468	1
GEO000000357	GO_0006468	calcium ion binding / GO:0006468	1
GEO000000358	GO_0006468	calcium ion binding / GO:0006468	1
SU27	false	ZANB02.C.2/ce2.c.23289//blast_match/60/_non_cum_bp_start_5127_caatgtatgat	
SU28	false	AC7.2a / ce2.c.25669 / blast_match / 60 / _non_cum_bp_start_5127_caatgtatgat	
SU29	false	C49C3.12 / ce2.c.27136 / blast_match / 60 / _non_cum_bp_start_17_actatggatgt	
SU30	false	EDB2.1 / ce2.c.29766 / blast_match / 60 / _non_cum_bp_start_15_gtttgtttag	
SU31	false	F52C12.5 / ce2.c.28957 / blast_match / 60 / _non_cum_bp_start_19_gacctttcgat	
SU32	false	F5B87.9 / ce2.c.29283 / blast_match / 60 / _non_cum_bp_start_62_aaccaagatgg	
SU33	false	T09A23.2 / ce2.c.31036 / blast_match / 60 / _non_cum_bp_start_82_cttctgttgt	
SU34	false	T09D15.6 / ce2.c.29094 / blast_match / 60 / _non_cum_bp_start_86_cttttgttgg	
SU35	false	Y66H1A.5 / ce2.c.32298 / blast_match / 60 / _non_cum_bp_start_42_gaccattatcc	
SU36	false	Y73B6A.4 / ce2.c.32468 / blast_match / 60 / _non_cum_bp_start_67_cgttgtggatg	
SU37	false	C1609.9 / ce2.c.34013 / blast_match / 60 / _non_cum_bp_start_86_tccacttttgat	

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

SampleID	Sample_label	Description	File
ie_bp_phe	Panacea	Mean Lifespan at 24oC, CB x N2 RILs, Villafuerte & S	
ie_nl_phe	Panacea	Mean Lifespan at 20oC, CB x N2 RILs, Donosuk et al	
ie1_eff	Panacea	Age At Maturity at 20oC, CB x N2 RILs, Gutteling et al	
ie1_norm_ril	Panacea	Fertility at 24oC, CB x N2 RILs, Gutteling et al. 20	
ie1_oef	Panacea	Fertility at 24oC, CB x N2 RILs, Gutteling et al. 20	
ie12_int_eff	Panacea	Pharngyl pumprate at 12oC, CB x N2 RILs, Guttingel et s	
ie12_int_qtl	Panacea	Pharngyl pumprate at 12oC, CB x N2 RILs, Guttingel et s	
ie12_qtl	Panacea	Egg size at 12oC, CB x N2 RILs, Gutteling et al. 2	
pe2_eff	Panacea	EGG size at 24oC, CB x N2 RILs, Gutteling et al. 2	
pe2_norm_ril	Panacea	Size at Maturity at 12oC, CB x N2 RILs, Gutteling et al.	
pe2_qtl	Panacea	Size at Maturity at 24oC, CB x N2 RILs, Gutteling et al.	
impRate_day4	Panacea	Pharngyl pumprate at day 4 at 20oC, CB x N2 RILs	
impRate_day6	Panacea	Pharngyl pumprate at day 6 at 20oC, CB x N2 RILs	
impRate_day8	Panacea	Pharngyl pumprate at day 8 at 20oC, CB x N2 RILs	
impRate_day10	Panacea	Pharngyl pumprate at day 10 at 20oC, CB x N2 RILs	
m12	Panacea	Coefficient of the pharyngeal pumprate at 20oC	
meanLifespan	Panacea	Mean Lifespan at 20oC	
sdLifespan	Panacea	SD of Lifespan distribution at 20oC	
nRILs	Panacea	Number of individuals measured per RIL	
imRL	Panacea	Mean number of juveniles at 20oC	
juStDev	Panacea	Juvenile standard deviation at 20oC	
juRIL	Panacea	Number of individuals measured per RIL	
i1_pump_qtl_gff	Panacea	QTLS from Gutteling et al. 2007, mapped	
je1_qtl_gff	Panacea	Blinmapping results for inferred from Doroszuk et al	
je23_qtl_gff	Panacea	Blinmapping results for lifespans from eQTLs from Li et al 2006, Link to paper: http://www.ncbi.nlm.nih.gov/pubmed/16572891	
je23_qtl_gff_eff	Panacea	eQTLs from Li et al 2006, Link to paper: http://www.ncbi.nlm.nih.gov/pubmed/16572891	
je23_qtl_gff_sd	Panacea	Blinmapping results for pump-rate fro	
je23_qtl_gff_sd_eff	Panacea	edQTLs from Li et al 2006, Link to paper: http://www.ncbi.nlm.nih.gov/pubmed/16572891	
l_pump_qtl_gff	Panacea	Blinmapping results for pump-rate fro	
l_qage_qtl_gff	Panacea	Blinmapping results for inferred from Doroszuk et al	
l_qeff_abs_gff	Panacea	Blinmapping results for lifespans from eQTLs from Li et al 2006, Link to paper: http://www.ncbi.nlm.nih.gov/pubmed/16572891	
l_qeff_abs_gff_eff	Panacea	Blinmapping results for pump-rate fro	
l_pump_phe	Panacea	Feeding curves (OD660 per time point	
ai_FC_phe	Panacea	QTLS for Feeding Curves on e.coli exp	
ai_FC_qtl	Panacea	QTLS for Feeding Curves on e.coli exp	
ck_eff	Panacea	Effects of the eQTLs from Rockman et	

Biomolecular measurements, association results ...

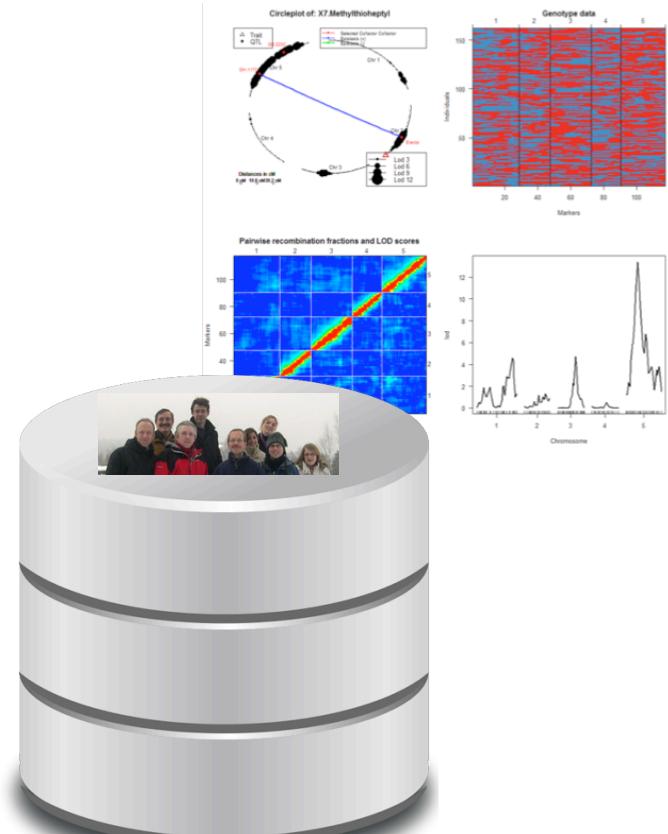
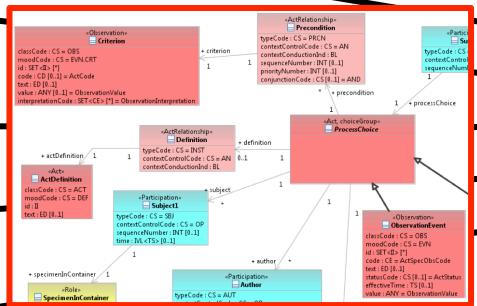


Challenge: Building a ‘team’ database

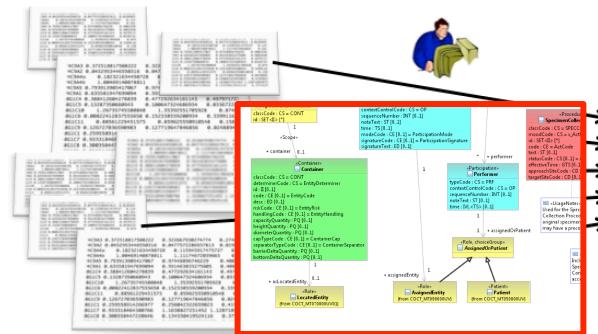
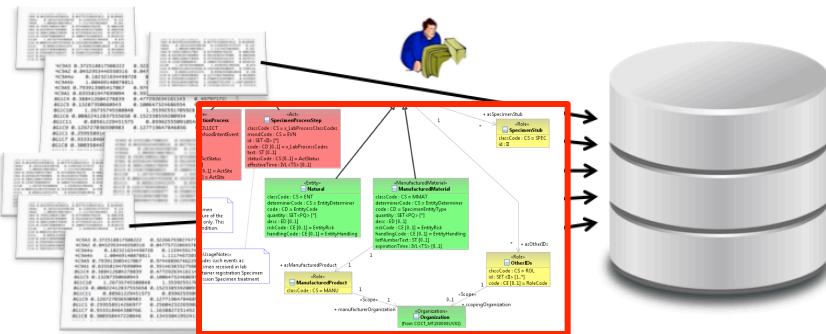
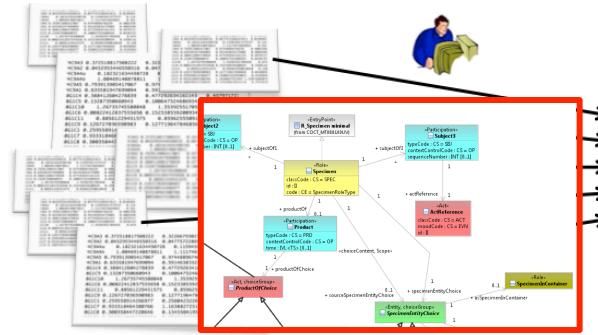
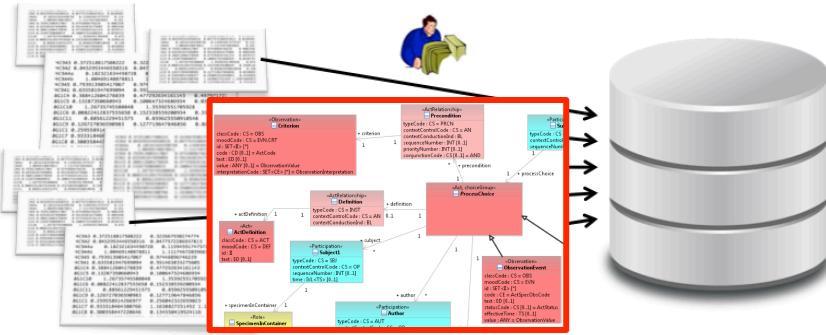
LINX R. 0.12721481179822	0.322
4C9A2 0.0452953446503116	0.04757228637813
4C9A4d 0.18232163449728	0.1584
4C9A4d 1.00869140878811	1
4C9A5 0.791380138417067	0.974
4C9A5 0.853939426314669	0.974
4G1C0 0.388412604277889	0.47729263416143
4G1C5 0.13287350660943	0.100647324686934
4G1C6 1.26735747461861	0.00010233855686
4G1C6 0.152338559208934	0.1584
4G1C9 0.12672793659986	0.127719647846856
4G1C9 0.259558914266977	0.259558914266977
4G1C9 0.300358442	0.00010233855686
4G1C9 0.37521817598822	0.32266793874774
4G1C9 0.4045295344650316	0.04757228637813
4G1C9 0.499461116363	0.019948371
4G1C9 0.584461116363	0.019948371
4G1C9 0.67082241283755658	0.152338559208934
4G1C9 0.859625559910548	0.1584
4G1C9 0.98561229451575	0.98561229451575
4G1C9 0.12672793659986	0.127719647846856
4G1C9 0.127719647846856	0.127719647846856
4G1C9 0.13455813524116	0.375762582
LINX R. 0.12721481179822	0.322
4C9A2 0.0452953446503116	0.04757228637813
4C9A4d 0.18232163449728	0.1584
4C9A4d 1.00869140878811	1
4C9A5 0.791380138417067	0.974
4C9A5 0.853939426314669	0.974
4G1C0 0.388412604277889	0.47729263416143
4G1C5 0.13287350660943	0.100647324686934
4G1C6 1.26735747461861	0.00010233855686
4G1C6 0.152338559208934	0.1584
4G1C9 0.12672793659986	0.127719647846856
4G1C9 0.259558914266977	0.259558914266977
4G1C9 0.300358442	0.00010233855686
4G1C9 0.37521817598822	0.32266793874774
4G1C9 0.4045295344650316	0.04757228637813
4G1C9 0.499461116363	0.019948371
4G1C9 0.584461116363	0.019948371
4G1C9 0.67082241283755658	0.152338559208934
4G1C9 0.859625559910548	0.1584
4G1C9 0.98561229451575	0.98561229451575
4G1C9 0.12672793659986	0.127719647846856
4G1C9 0.127719647846856	0.127719647846856
4G1C9 0.13455813524116	0.375762582



Data model



Bigger challenge: Building many databases



MOLGENIS: Many unique apps is not optimal

Model in DSL



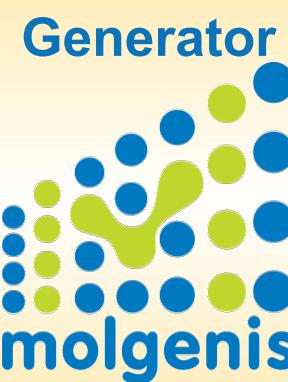
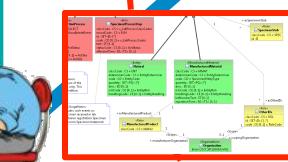
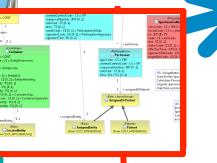
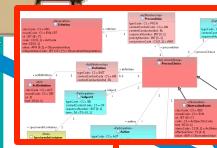
NextGenSeq



Mutation database



Model organisms



Use generated software

Solexa Sequencer LIMS

database of COL7A1 mutations

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

<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 ..

Biomolecular measurements,
association results ...

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

Genotypes,
conditions ...



Stable?

Annotations of concepts
used in data sets,
mostly static content

Dynamic?

Experimental data sets,
usually flexible and
volatile content

Example: eQTL data

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_

Stable?

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

Stable?

	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	1	169018	-18.26	PCR_non_cu
pkP1101	1	992189	-17.28	PCR_non_cu
pkP1103	1	1881116	-11.96	PCR_non_cu
pkP1052	1	2818974	-6.1	PCR_non_cu
egPE107	1	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" nullable="true"
    description="Reference to the chromosome this position belongs to." />
  <field name="cm" label="cMPosition" type="decimal" nullable="true">
    description="genetic map position in centi morgan (cM)." />
  <field name="bpStart" label="Start (5')" type="long" nullable="true">
    description="numeric basepair postion (5') on the chromosome" />
  <field name="bpEnd" label="End" type="long" nullable="true">
    description="numeric basepair postion (3') on the chromosome" />
  <field name="Seq" type="text" nullable="true">
    description="The FASTA text representation of the sequence." />
  <field name="Symbol" type="varchar" nullable="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" nullable="true" description="Length of the chromosome" />
  <field name="Species" label="Species" type="xref" xref_entity="Species">
    xref_field="id" xref_label="name" nullable="true"
    description="Reference to the species this chromosome belongs to." />
</entity>
```

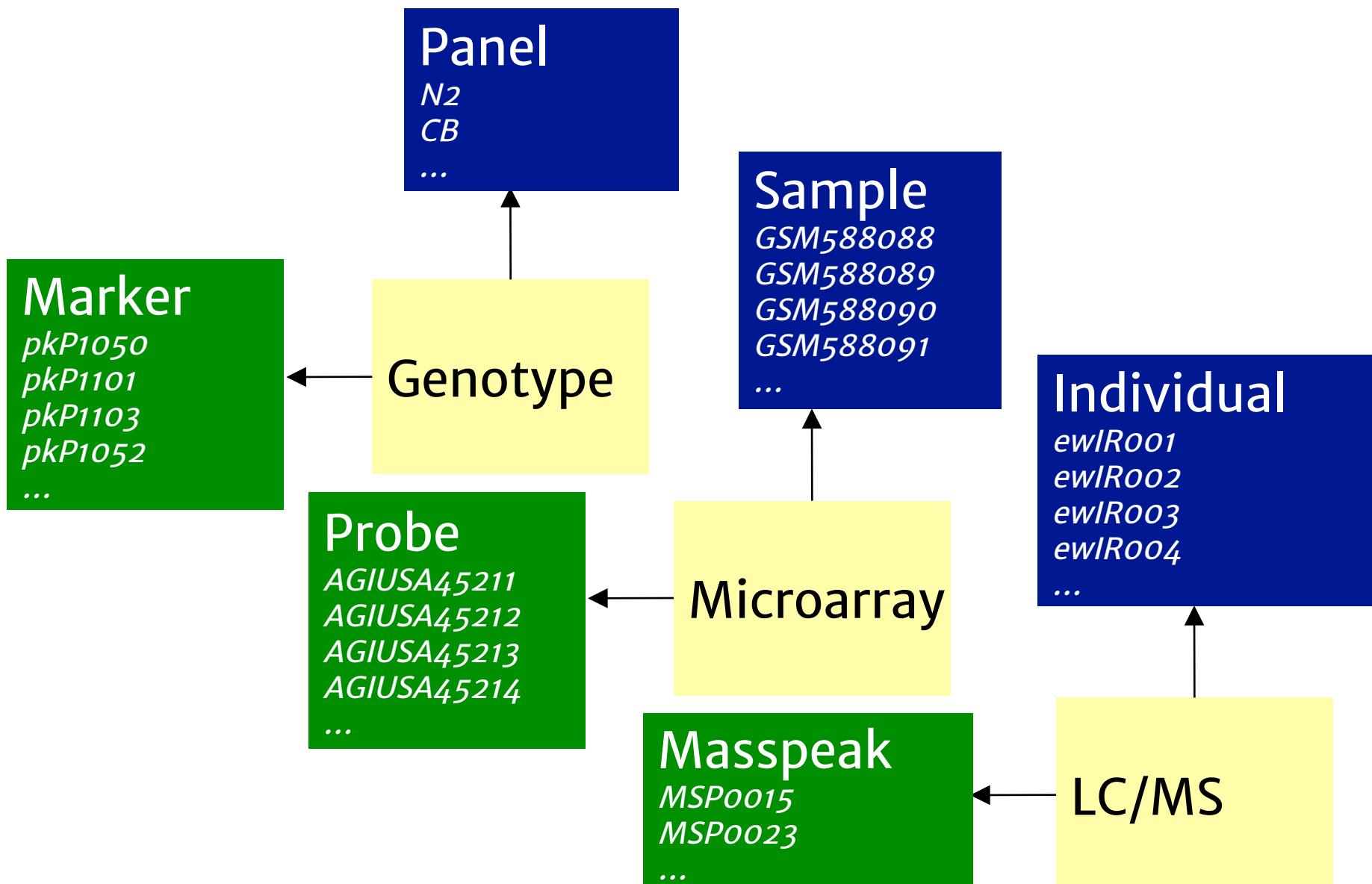
2. generate



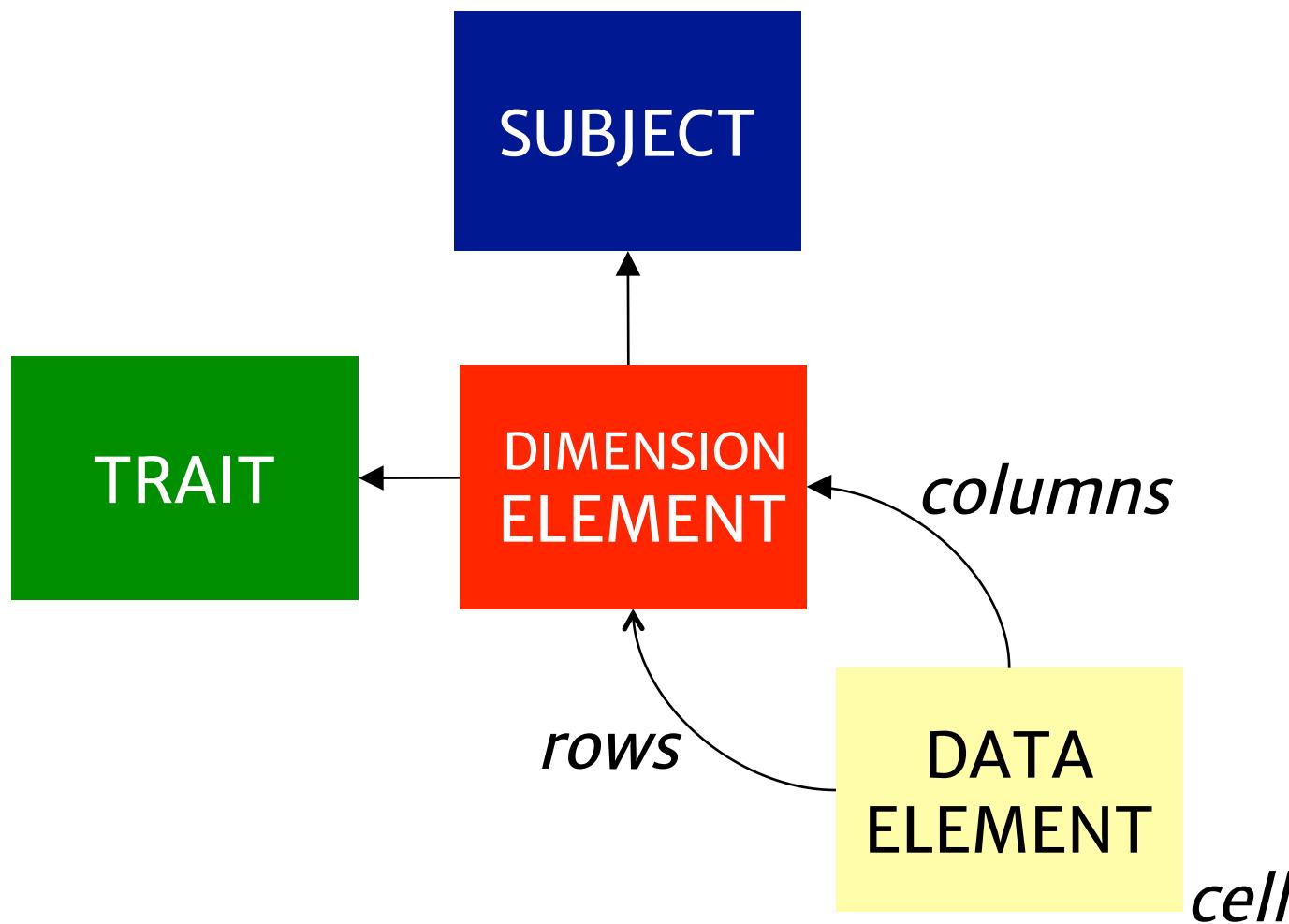
Strains (panels)							Chromosomes	Markers	Genes	Measurements	DerivedTraits	Probes	Samples	
							File	Edit	View	1 - 10 of 68,452				
Micro-array probes blasted against WS220 and linked to the genes, used for gene expression phenotypes and eQTLs.														
1.	105050	WSU1	NA / blast_match_NA / non_cumu_bp_start_0											
2.	105051	WSU2	C25A1.8 / cea2.c.00914 / blast_match_50 / non_cumu_bp_start_10184580											
3.	105052	WSU3	F21F3.6 / cea2.c.02677 / blast_match_60 / non_cumu_bp_start_4912043											
4.	105053	WSU4	F25H2.9 / cea2.c.02801 / blast_match_60 / non_cumu_bp_start_10567120											
5.	105054	WSU5	F56H1.4 / cea2.c.04344 / blast_match_60 / non_cumu_bp_start_5741975											
6.	105055	WSU6	H06O01.1 / cea2.c.04508 / blast_match_60 / non_cumu_bp_start_2015820											



Challenge: Data sets can be variable combinations

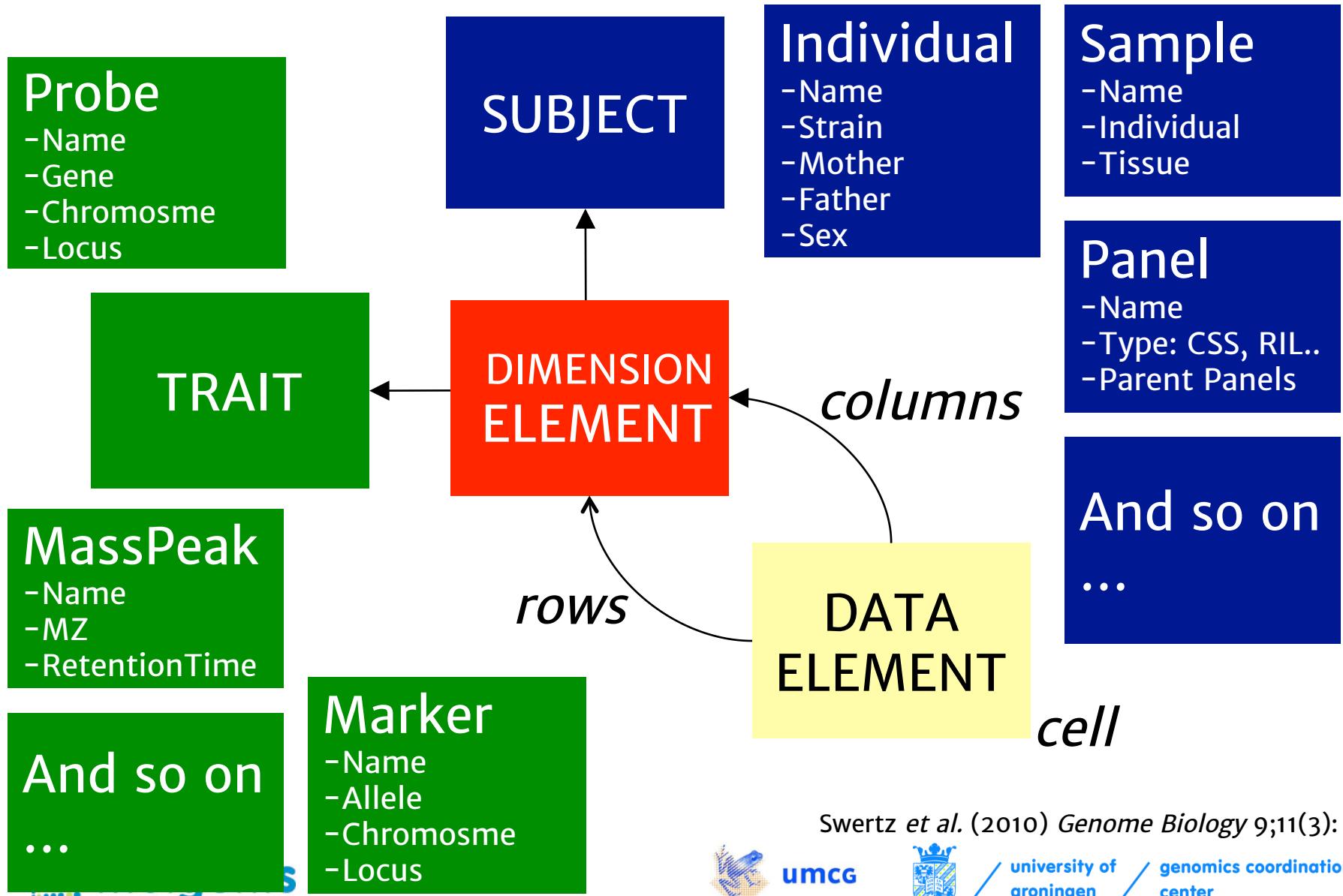


XGAP model: <any trait> X <any subject>



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.



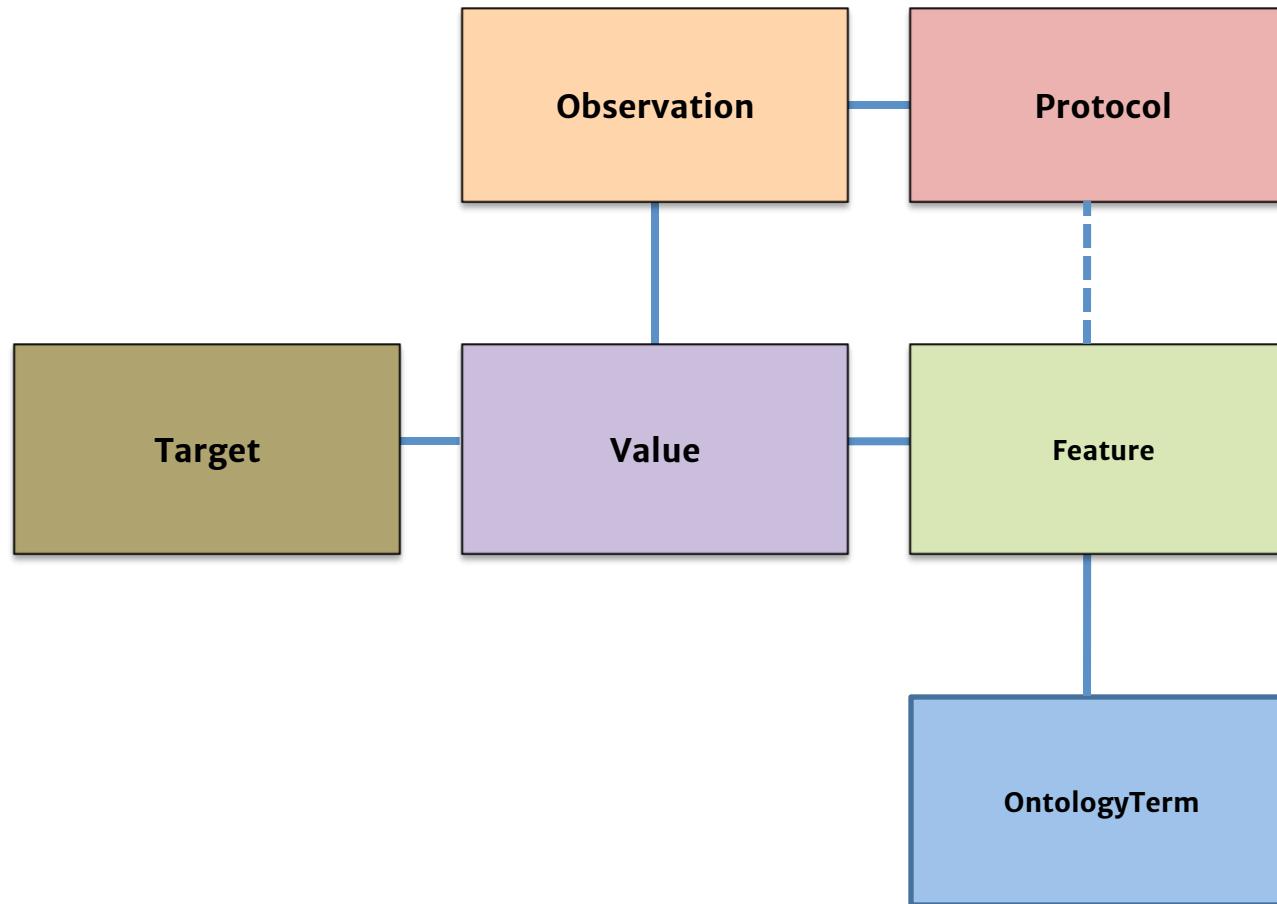
umcg



university of
groningen

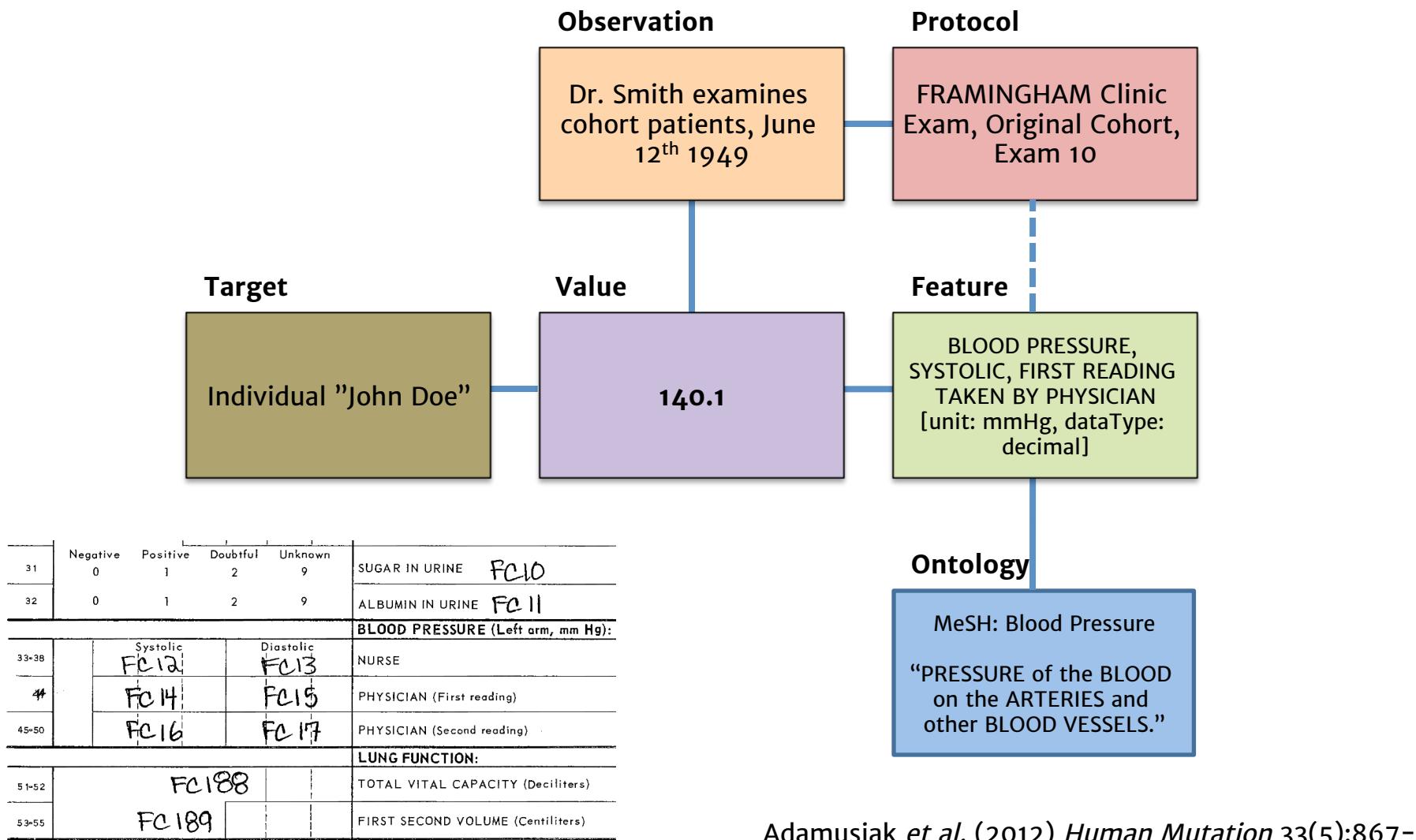
genomics coordination
center

Observ-OM model for flexible columns/prov.



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

Observ-OM data example



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

Evolution

MOLGENIS software

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

<http://www.molgenis.org>



Observ-OM model

Adamusiak *et al*,
Human Mutation (2012)

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



EB Registry

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

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



XGAP model

Swertz *et al*,
Genome Biology (2010)

<http://www.xgap.org>



xQTL workbench

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

<http://www.xqtl.org>



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>



WormQTL



CropQTL: Learning From Nature
xQTL workbench



university of
groningen

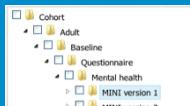
genomics coordination
center

..more

One to rule all? Observ-OMX

Catalogue

Find data item and sample collections



Data

Filter individual data sets and download to Excel & SPSS

	pP1059	aP1101	pP1101	pP1052	aP1107
95U2	-0.1892	-0.1892	0.231	-0.8179	-0.1186
95U2	-0.0192	-0.0192	0.1026	-0.2283	-0.0022
95U2	0.0637	0.0637	0.2159	-0.1184	-0.1068
95U2	0.0546	0.0546	0.2046	-0.1088	-0.0989
95U2	0.054	0.054	0.1649	-0.1768	-0.1621
95U2	0.0873	0.0873	0.1452	-0.0902	-0.0322
95U2	-0.0529	-0.0529	-0.0248	0.0354	0.0405
95U2	0.0608	0.0608	0.1486	0.0506	0.0103

Compute

Run analysis workflows on big data compute infrastructure



GWAS Central

Explore summary level GWAS data



Protocol

CRFs, Questionnaires, Lab protocols, and assays

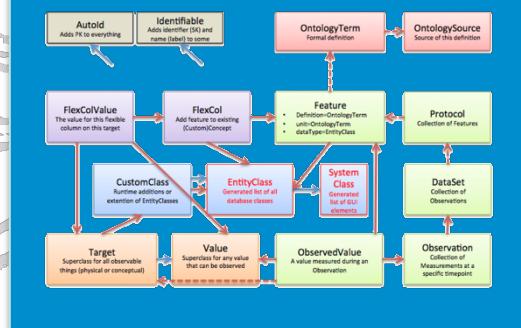


NGS

Next-Generation Sequencing

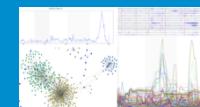


Core model



xQTL

Multi-omics association & visualization tools



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

C	H	Z	D	G	S	T	M	S	P
A-BE	23	D9	R2	HGD3cp shh					
4-2B	A3	B1	26	aPPaU H21					
E-97	DB	AB	F4	b63=H b6A1					
B-9B	DC	99	G2	*w68 H e					
6-62	99	18	EC	o'W'gFOPC					
8-5P	P4	CH	CB	491e1kr l					
8-9	66	65	E6	Ig3 trv f					
8-5F	F9	82	FC	Ko7 hs3-e					

Outcome: working applications (e.g. xQTL)

Strains (panels) | Chromosomes | Markers | Genes | Measurements | DerivedTraits | Probes | Samples

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.

	<input type="checkbox"/> id	<input type="checkbox"/> name	description	Investigation	ontology	Reference	Alternative	identifiers	label	Chromosome	cM
1.	<input type="checkbox"/>	10505 WSU1	NA / SpotReport / last_match_NA / on_cumu_bp_start_0	Public						I	-18
2.	<input type="checkbox"/>	10505 WSU2	25A1.8 / cea2.c.00914 / last_match_60 / on_cumu_bp_start_10184580	Public					clec-87	I	-17
3.	<input type="checkbox"/>	10505 WSU3	21F3.6 / cea2.c.02677 / last_match_60 / on_cumu_bp_start_4912043	Public					F21F3.6	I	-11
4.	<input type="checkbox"/>	10505 WSU4	25H2.9 / cea2.c.02801 / last_match_60 / on_cumu_bp_start_10567120	Public					pas-5	I	-6
5.	<input type="checkbox"/>	10505 WSU5	56H1.4 / cea2.c.04344 / last_match_60 / on_cumu_bp_start_5741975	Public					P56H1.4	I	-3
6.	<input type="checkbox"/>	10505 WSU6	06O01.1 / cea2.c.04508 / last_match_60 / on_cumu_bp_start_7015970	Public					pdi-3	I	-1
7.	<input type="checkbox"/>	10505 WSU7	20F10.2 / cea2.c.06048 / last_match_60 / on_cumu_bp_start_10300315	Public					T20F10.2	I	

Stable!

Strains (panels) | Chromosomes | Markers | Genes | Measurements | DerivedTraits | Probes | Samples

File | Edit | View | 1 - 10 of 1,579 | Click on to view one item, and on to return to this list.

Genetic markers used in one or more of the populations stored in WormQTL.

	<input type="checkbox"/> id	<input type="checkbox"/> name	description	Investigation	ontologyReference	Alternative	identifiers	label	Chromosome	cM
1.	<input type="checkbox"/>	10347 pkP1050	CR_non_cumul_bp_pos_169018	Public					I	-18
2.	<input type="checkbox"/>	10347 pkP1101	CR_non_cumul_bp_pos_992189	Public					I	-17
3.	<input type="checkbox"/>	10347 pkP1103	CR_non_cumul_bp_pos_1881116	Public					I	-11
4.	<input type="checkbox"/>	10347 pkP1052	CR_non_cumul_bp_pos_2818974	Public					I	-6
5.	<input type="checkbox"/>	10347 egPE107	CR_non_cumul_bp_pos_3502476	Public					I	-3
6.	<input type="checkbox"/>	10347 egPF101	CR_non_cumul_bp_pos_4338254	Public					I	-1
7.	<input type="checkbox"/>	10347 pkP1054	CR_non_cumul_bp_pos_4845516	Public					I	

Stable!



Dynamic!

Dystrophic EB Register (COL7A1)

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

Data Explorer

Search data items



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

Choose a dataset:

Patient ID	Phenotype	cDNA change 1	Protein change 1	Exon/Intron 1	Consequence 1	cDNA change 2	Protein change 2	Exon
P37	RDEB, severe generalized	c.3G>T	p.0	Exon 1	pathogenic	c.3G>T	p.0	Exon
P292	RDEB, severe generalized	c.3G>A	p.0	Exon 1	pathogenic	c.353delGinsCCCCCTTGCAA	p.Arg118ProfsX14	Exon
P172	RDEB, unknown	c.3G>T	p.0	Exon 1	pathogenic	c.448G>A	p.Gly150Arg	Exon
P34	RDEB, severe generalized	c.3G>T	p.0	Exon 1	pathogenic	c.4997dupG	p.Pro1668AlafsX4	Exon
P168	RDEB, unknown	c.58C>T	p.Arg20Ter	Exon 1	pathogenic	c.8245_8246insC	p.Gly2749AlafsX41	Exon
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
P367	RDEB, generalized other	c.112G>T	p.Asp38Tyr	Exon 2	pathogenic	c.2157G>A	p.Trp719Ter	Exon
P604	RDEB, pretibial	c.130G>A	p.Asp44Asn	Exon 2	pathogenic	c.8569G>T	p.Glu285Ter	Exon
P213	RDEB, severe generalized	c.143C>T	p.Ser48Phe	Exon 2	pathogenic	c.3625_3635del	p.Ser1209LeufsX6	Exon
P1	RDEB, pruriginosa	c.151C>G	p.Arg51Gly	Exon 2	pathogenic	c.7474C>T	p.Arg2492Ter	Exon
P2	RDEB, severe generalized	c.238G>C	p.Ala80Pro	Exon 2	pathogenic	c.3631C>T	p.Gln1211Ter	Exon
P435	RDEB, severe generalized	c.267-3C>G		IVS2	pathogenic	c.267-3C>G		IVS2
P240	RDEB, pretibial	c.267-1G>C		IVS2	pathogenic	c.5096C>T	p.Pro1699Leu	Exon
P110	RDEB, severe generalized	c.313dupC	p.Arg105ProfsX5	Exon 3	pathogenic	c.5047C>T	p.Arg1683Ter	Exon
P433	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.325_326insCG	p.Glu109AlafsX39	Exon
P434	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.325_326insCG	p.Glu109AlafsX39	Exon
P267	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.3277-1G>C		IVS2
P268	RDEB, severe generalized	c.325_326insCG	p.Glu109AlafsX39	Exon 3	pathogenic	c.3277-1G>C		IVS2
P49	RDEB, generalized other	c.344dupG	p.Asn116GlnfsX13	Exon 3	pathogenic	c.6082G>A	p.Gly2028Arg	Exon

« 1 2 3 4 5 6 ... 158 »

3156 data items found

Dystrophic EB Register (COL7A1)

Phenotypic details for patient 'P10'

Characteristics

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

target

feature

value

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

Protocol Application

WormQTL.org (C. Elegans)



WormQTL

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

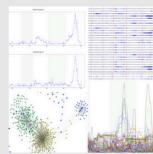
[Home](#) | [Find QTLs](#) | [Genome browser](#) | [Browse data](#) | [Help](#)

[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](#)

sp10150	sp10110	sp10100	sp10102	sp10107
0.13892	-0.13892	0.231	-0.8379	-0.9336
0.60121	0.60121	0.616	-0.2343	-0.602
0.90677	0.90677	0.2133	-0.1132	-0.1668
0.50100	0.50100	0.208	0.0192	-0.0009
0.92557	0.92557	0.208	-0.0192	-0.0009
0.80077	0.80077	0.195	0.0086	-0.0009
0.05358	0.05358	-0.024	0.0154	0.0005
0.62021	0.62021	0.188	0.006	-0.0005
0.04221	0.04221	0.092	-0.234	-0.0034
0.03095	0.03095	0.056	-0.032	-0.0338

[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 e.g. search with [Find QTLs](#)

WormQTL.org (C. Elegans)

<i>Phenotypes</i>	<i>Type of array</i>	<i>Sample size</i>	<i>Parental strains</i>	<i>Reference</i>	<i>Pubmed link</i>	<i>Growing temperature</i>	<i>Stage</i>	<i>Food</i>	<i>Medium</i>	<i>Dataset IDs</i>
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 RILs	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

WormQL.org (C. Elegans)

Home Find QTLs Genome browser Browse data Help

provide feedback: 

Find QTLs

All data (175,366)  seam cell  

(for example: [measurement](#) (45))

measurement (45)  

Panel (500) 

Gene (47,360) 

Transcript (55,782) 

Chromosome (8) 

Probe (68,452) 

Sample (1,630) 

DerivedTrait (12) 

Found n: (9) 

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

 Probe [AGIUSA14764 / clc-2](#) reports for [WBGene00000523 - WormBase](#)
C01C10.1 / C01C10.1 / wb|C01C10.1 / non_cumu_bp_start_743339 [...more](#)

 Probe [AGIUSA16119 / ceh-1](#) reports for [WBGene00000428 - WormBase](#)
F16H11.4 / F16H11.4 / wb|F16H11.4 / non_cumu_bp_start_465353 [...more](#)

 Probe [AGIUSA19594 / acn-1](#) reports for [WBGene00000039 - WormBase](#)
peptidase [C42D8.5.2] / C42D8.5.2 / C42D8.5 / wb|C42D8.5|w [...more](#)

 Probe [AGIUSA41433 / gsp-1](#) reports for [WBGene0001747 - WormBase](#)
serine/threonine protein phosphatase [F29F11.6.1] / F29F11.6 [...more](#)

 Probe [AGIUSA5476 / cul-2](#) reports for [WBGene00000837 - WormBase](#)

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

 **Ontologies**

 **Ontologies**

 **Ontologies**

WormQL.org (C. Elegans)

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

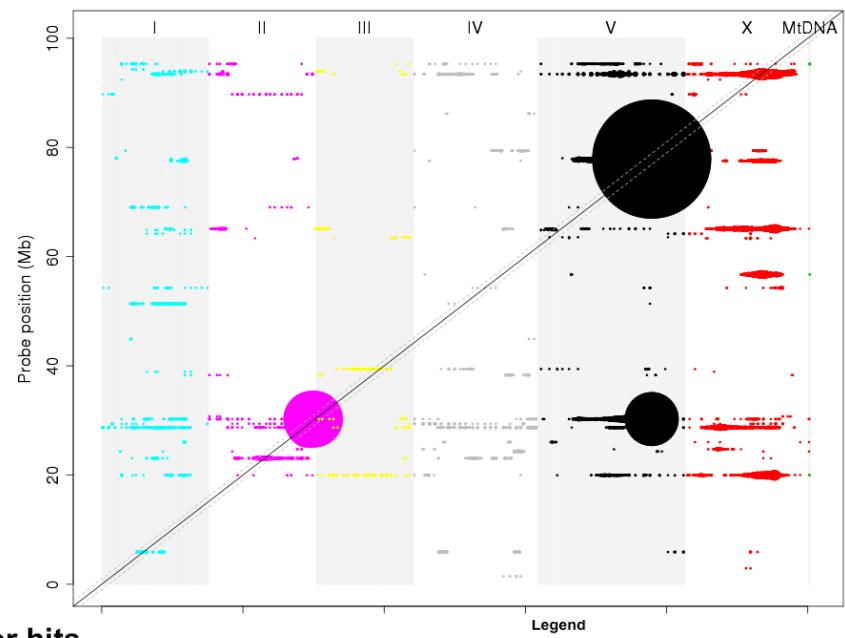
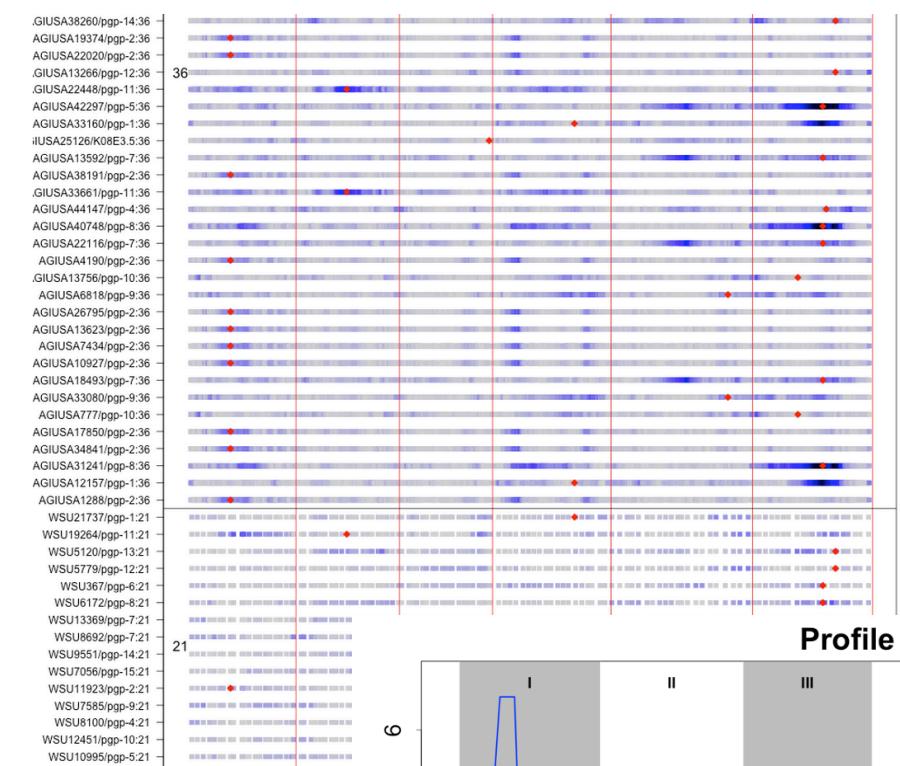
Legend, click to enlarge:

More downloads:

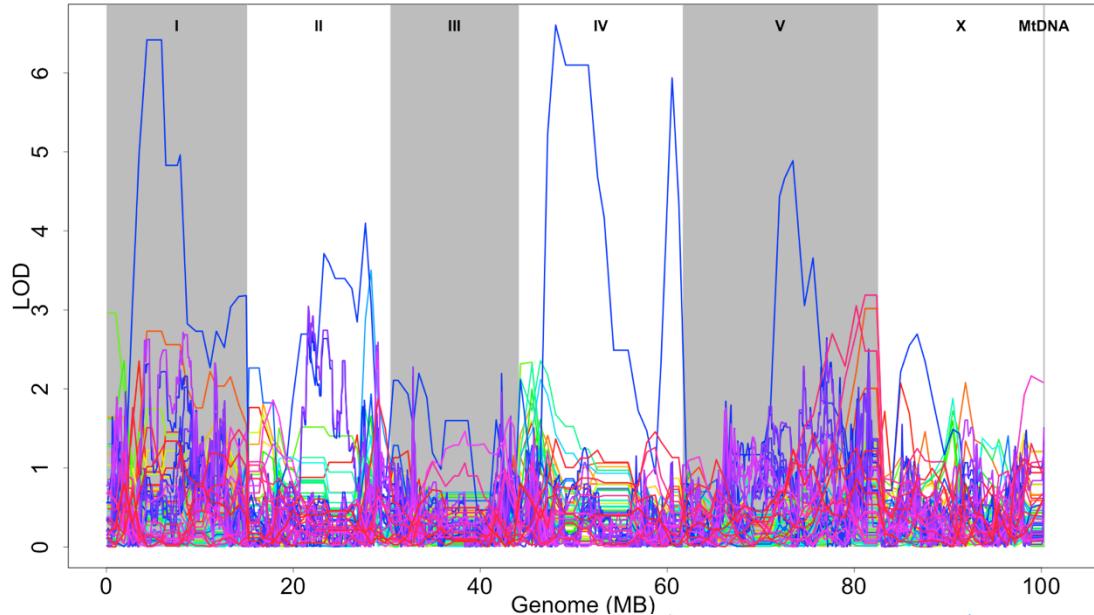
Get the [Cytoscape network](#) for this plot. ([how-to import](#))
Get the [Cytoscape nodes](#) for this plot. ([how-to import](#))
Note: includes **significant results only**. (LOD > 3.5)
Save both files. Import network (has LOD scores), then node attributes (chrom, bploc, dataset). [Example visualization](#)

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](#).

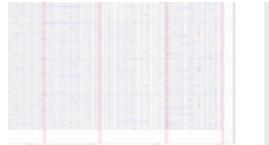
WormQTL.org (C. Elegans)



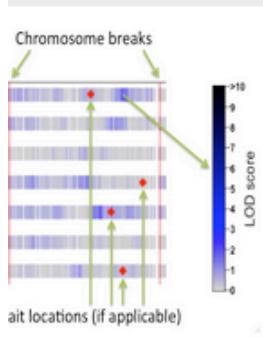
Profile plot for hits



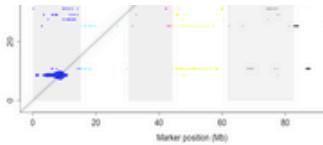
WormQL.org (*C. Elegans*)



Click to enlarge:



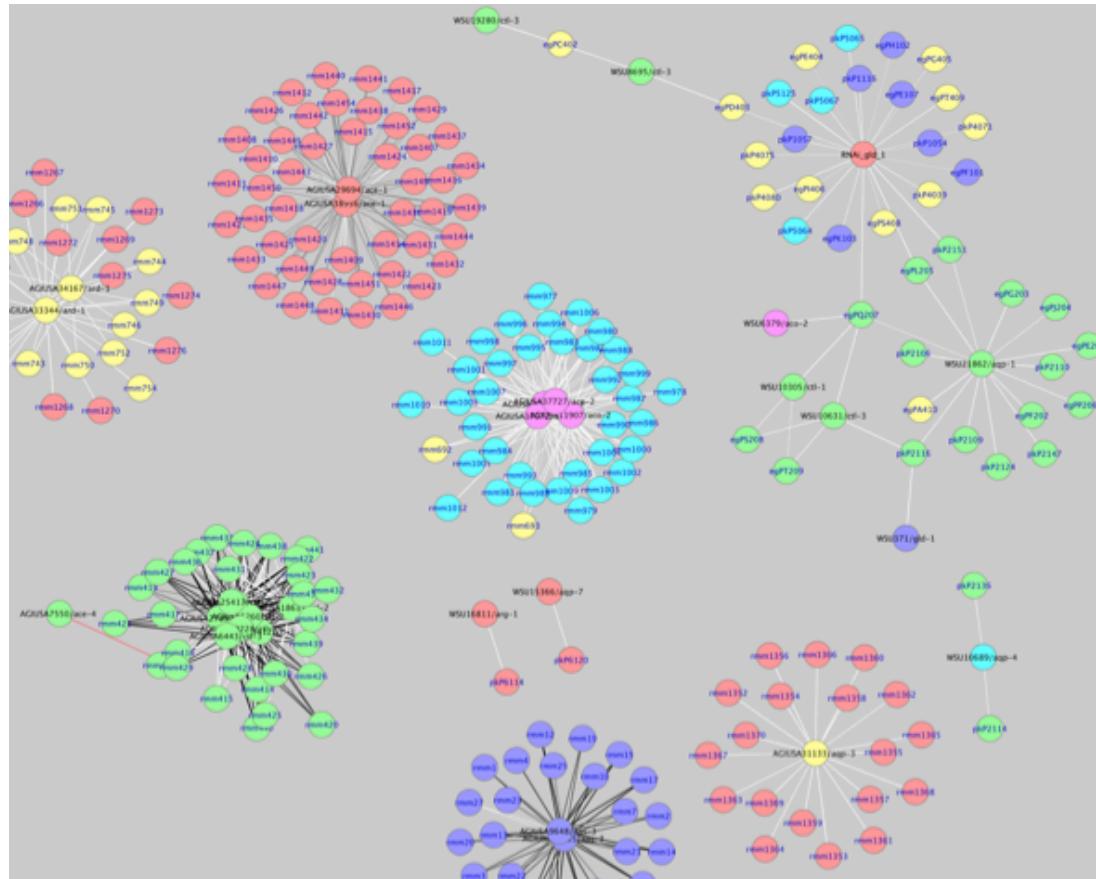
[/daf-1 \[explore deeper\]](#) – protein kinase [F29C4.1b] / F29C4.1c
[/daf-3 \[explore deeper\]](#) – F25E2.5b.3 / F25E2.5 / wb|F25E2.5b.3
[/daf-11 \[explore deeper\]](#) – R0240.3 / cea2 n 107079 / blas



More downloads:

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

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



PlantQTL (Arabidopsis & Brassica, pilot)

PlantQTL
xQTL workbench for plant research

rijksuniversiteit
groningen

WAGENINGEN UR
For quality of life

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

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

report Utilities Admin

Navigate investigations (studies)

File Edit View ◀◀ ◀ 1 of 1 ▶ ▶

Search:

name * Brassica_Nutrigenomics

Show additional fields

Save Refresh Cancel

Experiment overview Investigation overview Datasets Identifiers

Navigate datasets

File Edit View ◀◀ ◀ 14 of 15 ▶ ▶

name * gene_quant

description

Show additional fields

Save Refresh Cancel

View Inspect Delete

View

Individual 1-5 of 146

Probe 1-10 of 36831 Stepsize 5

	LerKas144	LerKas135	LerKas155
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

Legend

molgenis39.target.rug.nl:8080/plantqtl/molgenis.do?__target=Investigation

Gene quant

The heatmap displays gene expression levels for the 'gene_quant' dataset. The y-axis lists genes: At2g4676, CHR5_00, CHR4_00, CHR2_01, At5g0522, At4g3342, At5g3537, At3g1414, At4g2453, and At2g3216. The x-axis shows individuals: Kas135, Kas64, Kas144, Kas155, and Kas106. A dendrogram on the left indicates hierarchical clustering of both genes and individuals. The color scale ranges from yellow (low) to red (high), with a white border separating the clustered regions.

GoNL (Genome of the Netherlands) variants



Find GoNL variants

Find GoNL variants

Chromosome	Start bp position	Stop bp position	Find
<input type="text" value="3"/>	<input type="text" value="20000000"/>	<input type="text" value="21000000"/>	<input type="button" value="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

BioSHARE
molgenis
www.bioshare.eu

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.

Existing tools to enable data integration include the MOLGENIS platform, XGAP ‘omics’ database, Observ-OM ‘phenes’ model, xQTL ‘workbench’, and successful adaptations such as WormQLP. However, now biologists want an all-in-one package.

MOLGENIS
The integrated analysis environment for life science researchers. MOLGENIS is a Java-based framework for building and maintaining complex biological databases. It is designed to support the needs of modern genomic and proteomic research. MOLGENIS is a Java-based framework for building and maintaining complex biological databases. It is designed to support the needs of modern genomic and proteomic research.

Observ-OM
Generic Observations, Protocols, Studies

XGAP
model Molecular / classical phenotypes & genotypes

WormQLP
A phenome and genotype database for C. elegans. WormQLP is a web-based application that integrates data from WormBase and the C. elegans Genetic Reference Panel (GPR).

Variant model
Genomic variation

GWAS Central
Genetic association study findings

PAGE-OM
Observation Model for DNA variation

WormQTL.org
The phenome and genotype experiment database for C. elegans. WormQTL.org is a web-based application that integrates data from WormBase and the C. elegans Genetic Reference Panel (GPR).

bioSHARE
The bioSHARE system is a cloud-based platform for sharing and managing biological data. It provides a user-friendly interface for uploading, managing, and sharing data across different platforms and databases.

molgenis
The molgenis platform is a Java-based framework for building and maintaining complex biological databases. It is designed to support the needs of modern genomic and proteomic research.

OmicsConnect toolbox

OmicsConnect
Flexible multi-omics data capture and integration tools for high-throughput biology.

OmicsConnect platform

Here we present the next generation in Observ-OMX, a data model that captures all data modalities and context information. In addition, we have developed a ‘Best of Practices’ for existing domain-specific models and 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-OM model, a software toolbox for labs and consortia to import, curate, analyze, query and analytically interface with large multi-omics data and complex ontologies in biobanks, translational medicine, epidemiology, genetics and model organism studies.

OmicsConnect
Flexible multi-omics data capture and integration tools for high-throughput biology.

OmicsConnect toolbox

The right column on this poster shows a simple example that illustrates in three simple steps how Molgenis Compute enables you to run your big data analyses on clusters, GRID and other computational infrastructures.

GoNL alignment workflow

Run on clusters and GRID, ...

ComputeStart

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.

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

OmicsConnect platform

Here we present the next generation in Observ-OMX, a data model that captures all data modalities and context information. In addition, we have developed a ‘Best of Practices’ for existing domain-specific models and 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-OM model, a software toolbox for labs and consortia to import, curate, analyze, query and analytically interface with large multi-omics data and complex ontologies in biobanks, translational medicine, epidemiology, genetics and model organism studies.

OmicsConnect
Flexible multi-omics data capture and integration tools for high-throughput biology.

OmicsConnect toolbox

The right column on this poster shows a simple example that illustrates in three simple steps how Molgenis Compute enables you to run your big data analyses on clusters, GRID and other computational infrastructures.

GoNL alignment workflow

Run on clusters and GRID, ...

ComputeStart

Big data analysis made easy Using Molgenis Compute

Motivation

MOLGENIS Compute is a toolbox to configure and execute computational pipelines, achieved by a DNA re-sequence, genome imputations with QTL and GWAS pipelines underway. It was first used in the Genome of the Netherlands (GoNL), a nationwide BBMRI-NL project of 769 whole genome sequence samples (12x). All those needed alignment and variant calling. This required a complex pipeline of 10 shell files, analysis required >30 analysis steps, the complete analysis was >50,000 jobs accumulating 200,000 hours. Top candidate systems Taverna and Galaxy are not optimized for these large numbers and distributed analysis. Moreover, bioinformaticians needed complete influence on the shell scripts executed to optimally run these on big servers, clusters and grids.

Big data analysis made easy

Molgenis Compute is a lightweight, easy to change, pipeline system that can be run from command line or web interface.

It uses three simple text files: (1) a Workflow CSV file (2) your shell scripts, which given a (3) Parameters CSV automatically generate the cluster/grid shell scripts.

We provide automatic mechanisms to submit the generated scripts and for resource, file and tool management.

The right column on this poster shows a simple example that illustrates in three simple steps how Molgenis Compute enables you to run your big data analyses on clusters, GRID and other computational infrastructures.

Step 1: Design

Step 2: Generate

Step 3: Run

Big data analysis made easy Using Molgenis Compute

Motivation

MOLGENIS Compute is a toolbox to configure and execute computational pipelines, achieved by a DNA re-sequence, genome imputations with QTL and GWAS pipelines underway. It was first used in the Genome of the Netherlands (GoNL), a nationwide BBMRI-NL project of 769 whole genome sequence samples (12x). All those needed alignment and variant calling. This required a complex pipeline of 10 shell files, analysis required >30 analysis steps, the complete analysis was >50,000 jobs accumulating 200,000 hours. Top candidate systems Taverna and Galaxy are not optimized for these large numbers and distributed analysis. Moreover, bioinformaticians needed complete influence on the shell scripts executed to optimally run these on big servers, clusters and grids.

Big data analysis made easy

Molgenis Compute is a lightweight, easy to change, pipeline system that can be run from command line or web interface.

It uses three simple text files: (1) a Workflow CSV file (2) your shell scripts, which given a (3) Parameters CSV automatically generate the cluster/grid shell scripts.

We provide automatic mechanisms to submit the generated scripts and for resource, file and tool management.

The right column on this poster shows a simple example that illustrates in three simple steps how Molgenis Compute enables you to run your big data analyses on clusters, GRID and other computational infrastructures.

Step 1: Design

Step 2: Generate

Step 3: Run

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

1. Motivation

Newly developed prediction model need to be tested for their performance with new datasets. Common practice is to validate 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.

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

1. Motivation

Newly developed prediction model need to be tested for their performance with new datasets. Common practice is to validate 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



university of
groningen
genomics coordination
center

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 ...

B B M R I • N L



NBIC/BioAssist consortium (bioinfo)
BBMRI-NL catalogue group(Hs)
CTMM/TraIT consortium (Hs)
EU-GEN2PHEN consortium (Hs)
EU-PANACEA consortium (Ce)
EU-BioSHARE consortium (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...



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



github.com/molgenis