TARGET Conference 2013
Probing Big Data for answers
3 – 5 April, 2013
Groningen, Netherlands

Flexible datamodels for life sciences

K. Joeri van der Velde, Martijn Dijkstra, Morris Swertz, members of the Genomics Coordination Center

umcg

university of groningen / genomics coordination center

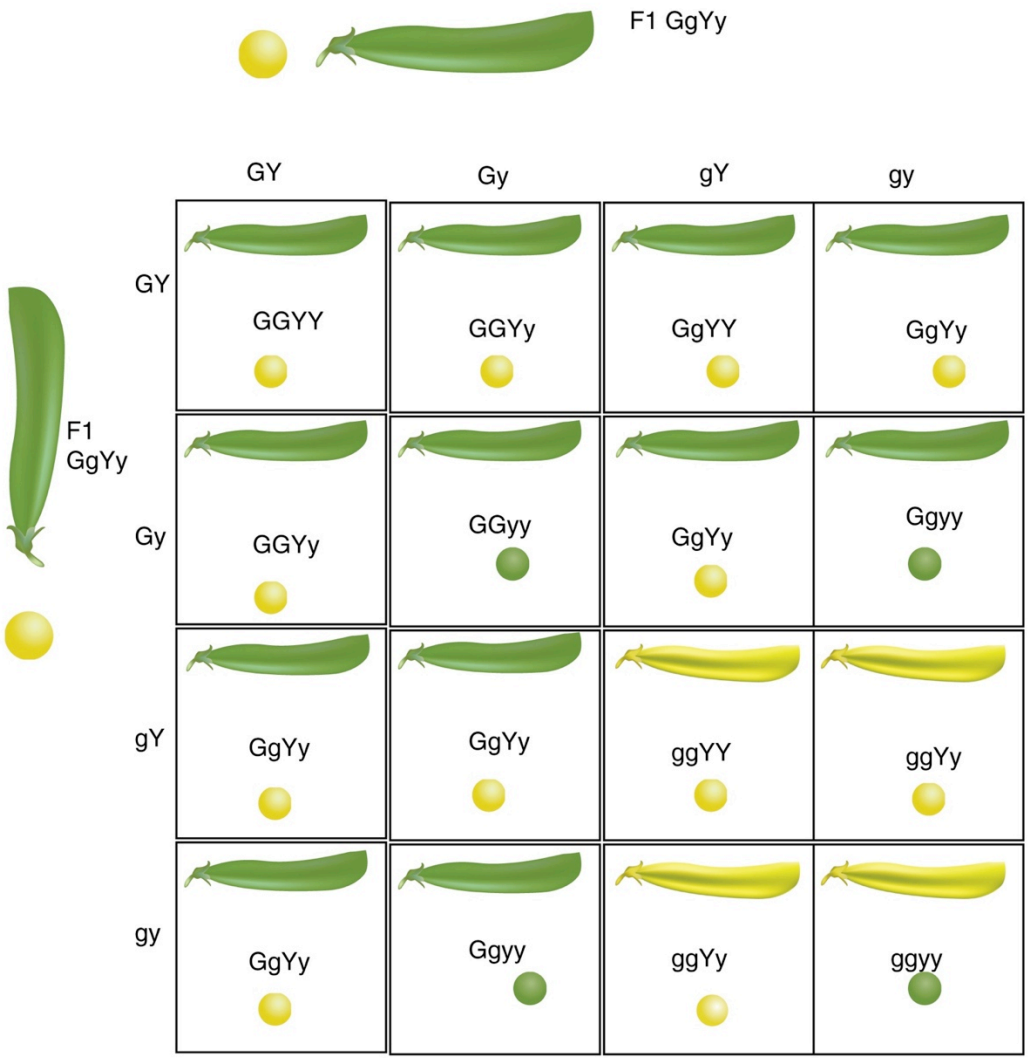
Outline

- Introduction
  - Motivation
  - Dealing with variation in research needs
  - Design-time vs runtime configuration
- Results
  - XGAP model for homogeneous data ('molecules')
  - Observ-OM model for heterogenous data ('phenotypes')
- Showcase
  - EB Registry
  - WormQTL
- Current work

Introduction

Motivation

Biological variation first explained by genetics



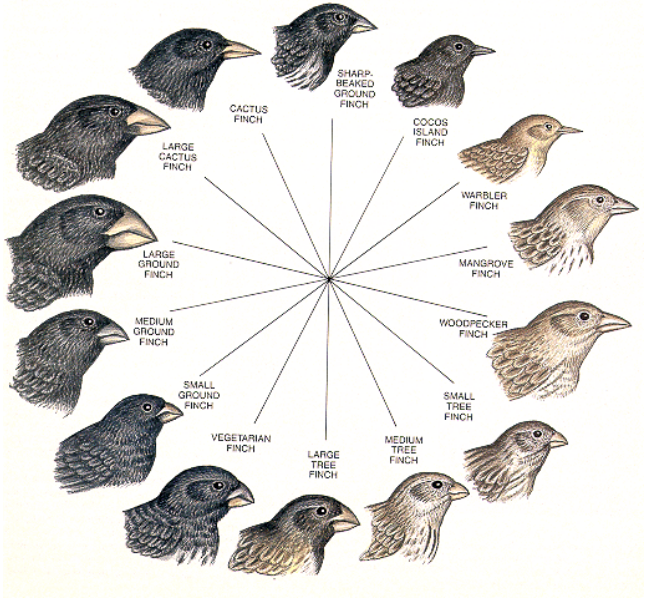
Mendel got lucky.. single-locus traits

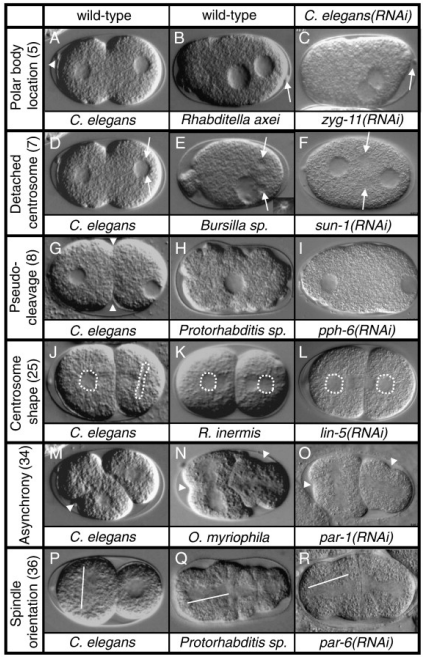
molgenis

umcg

university of groningen / genomics coordination center

Motivation: Understanding complex variation



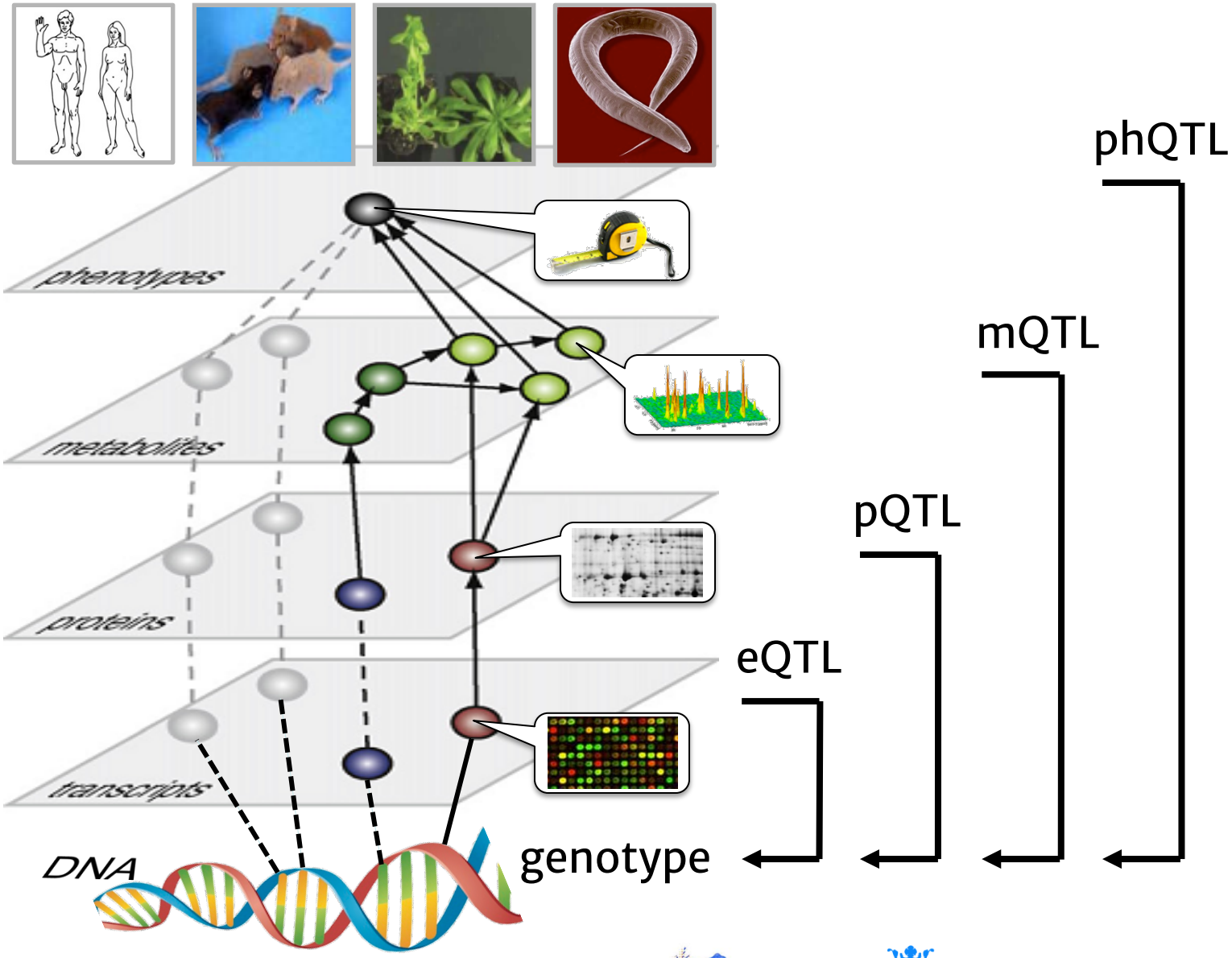


molgenis

umcg

university of groningen / genomics coordination center

How? Understanding geno-to-pheno

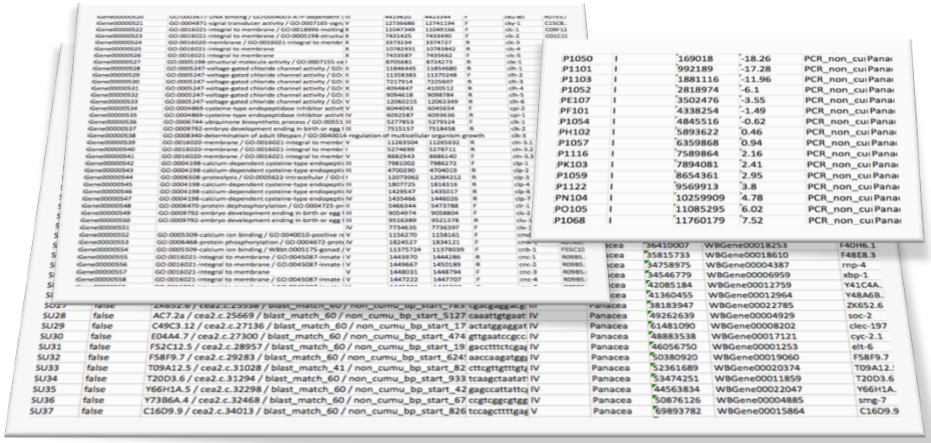


Introduction

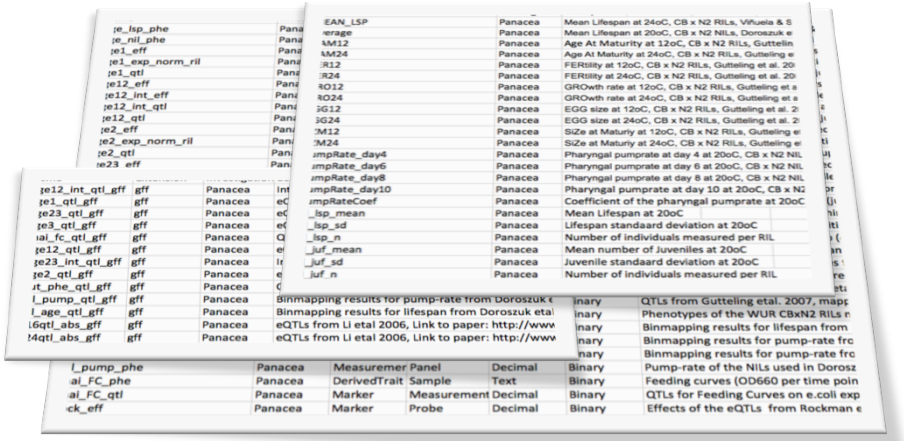
Dealing with variation in research needs

Situation: Many types and flavours of data

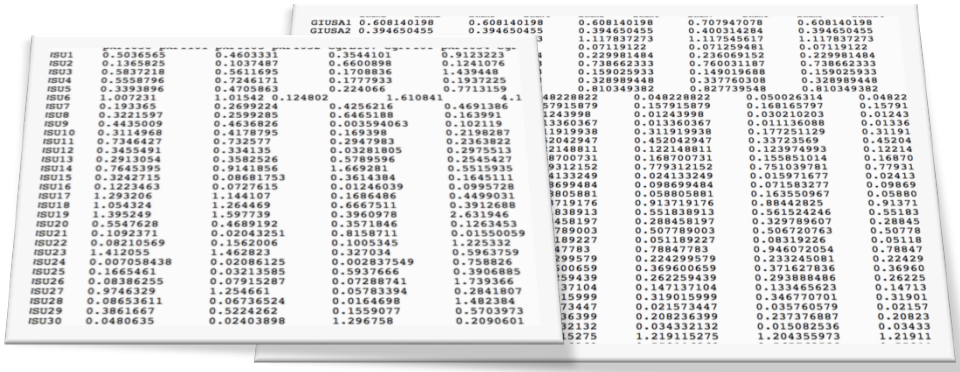
Genomic features, individuals, ontologies ..



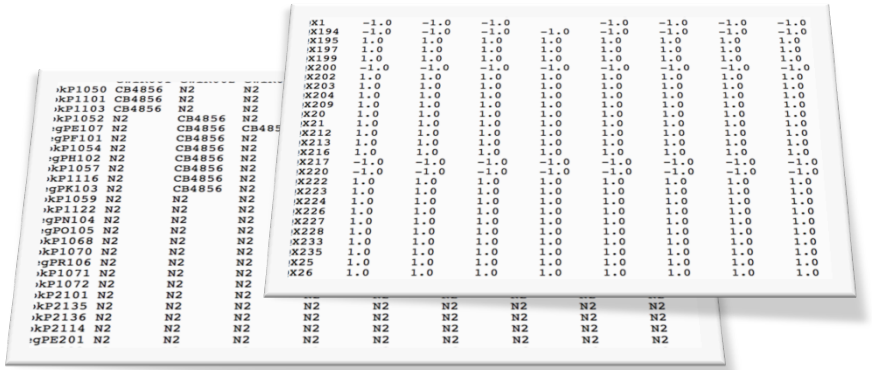
Metadata for phenotypes, datasets, samples, panels ...



Biomolecular measurements, association results ...



Genotypes, conditions ...



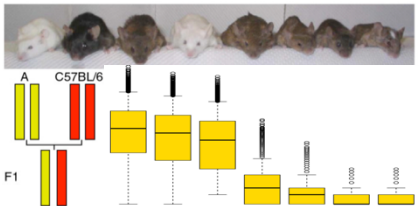
molgenis

umcg

university of groningen

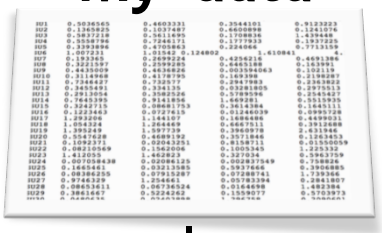
genomics coordination center

What needs to be done?



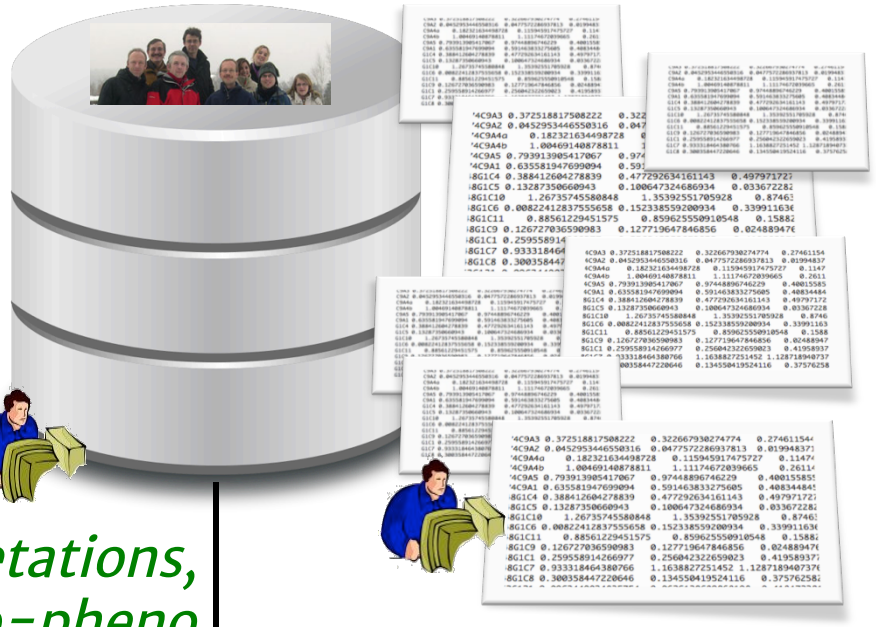
Project that generates data

‘My’ data

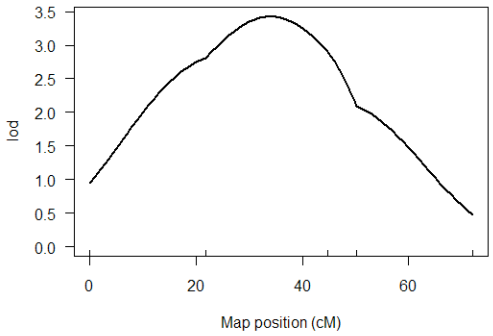


A little extra effort

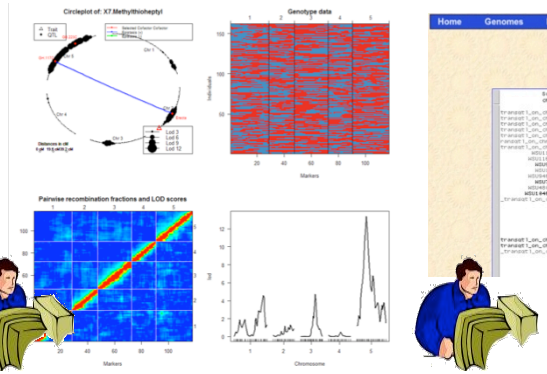
‘Team’ data

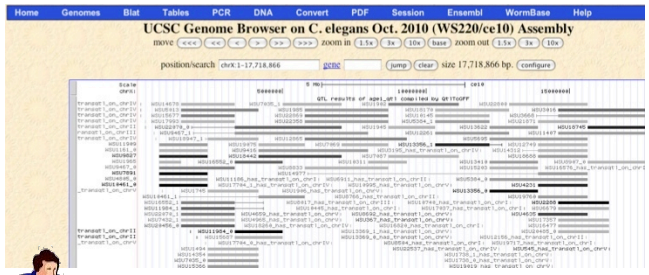


Direct interpretation



Rich interpretations, geno-to-pheno





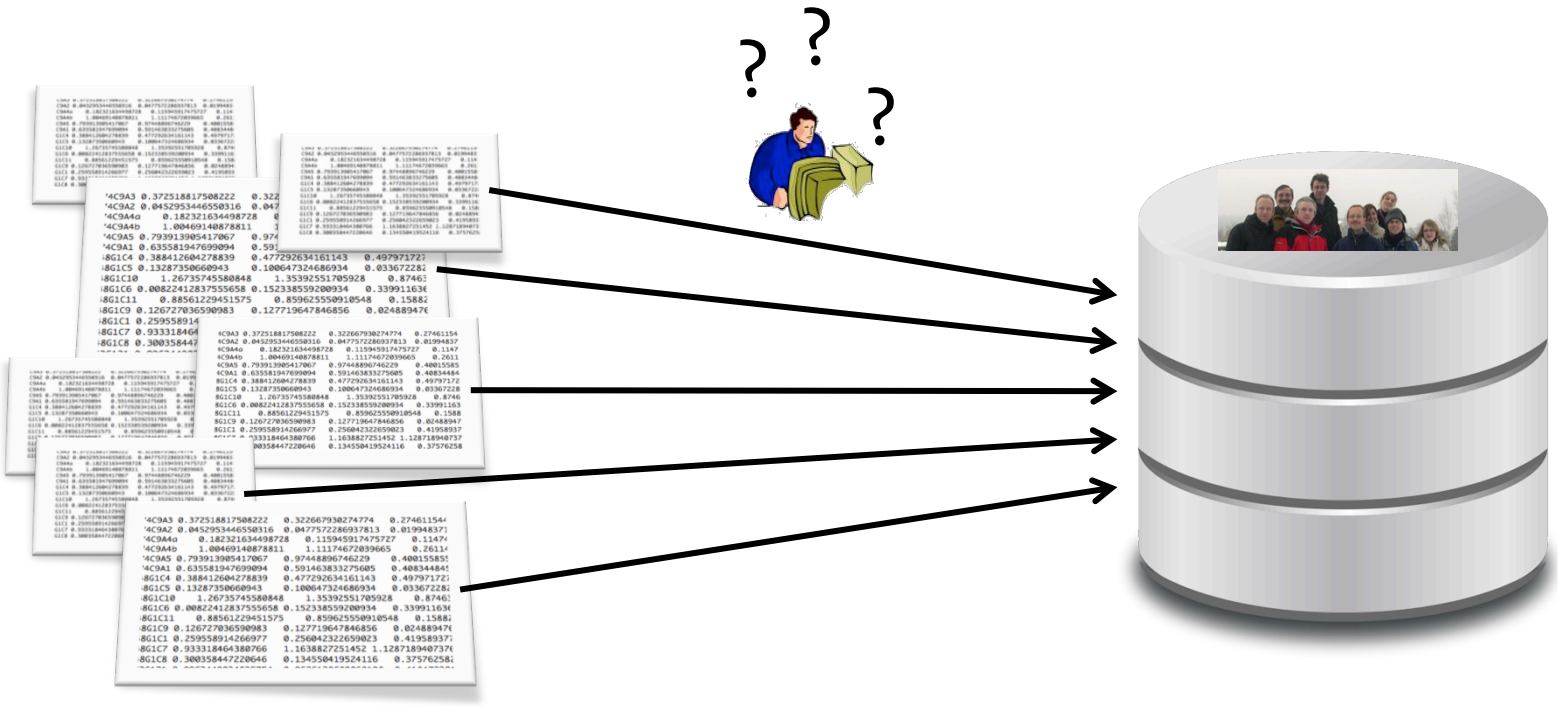
molgenis

umcg

university of groningen

genomics coordination center

Challenge: Building a ‘team’ database



Introduction

Design-time vs runtime configuration

Autogenerate the software

Model in DSL

NextGenSeq

Mutation

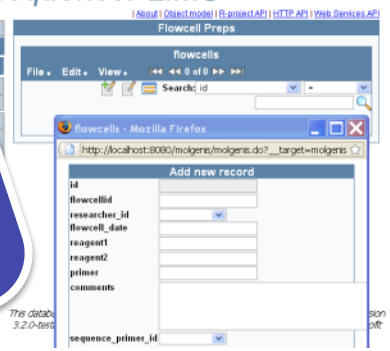
Model organisms

What models to use?

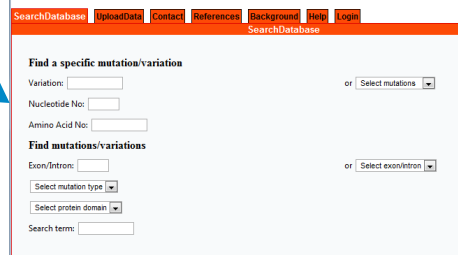
Can we have a model that rules them all?

Use generated software

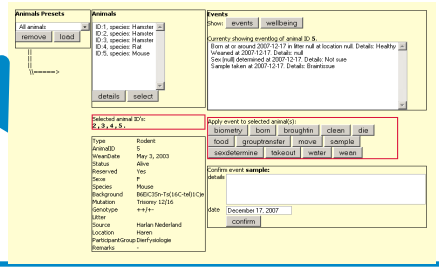
Solexa Sequencer LIMS



database of COL7A1 mutations



Animal Observatory



http://www.molgenis.org

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

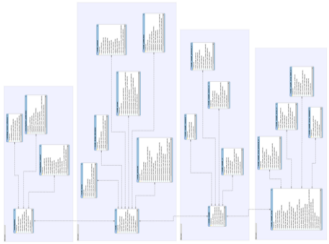
Swertz et al (2004) *Bioinformatics* 20(13), 2075-83

genomics coordination center

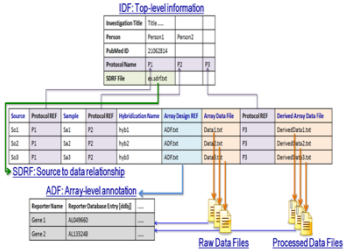
Examples of models.. How to choose?

Stable
'design-time'

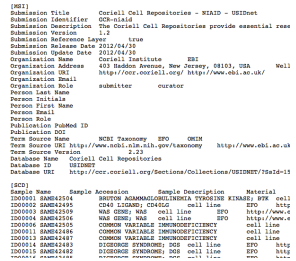
ISA-TAB
Investigation, study, sample



MAGE-TAB
Microarray provenance

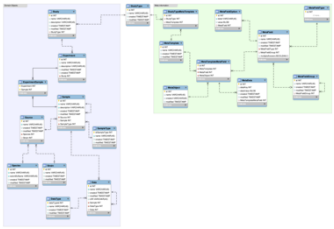


SAMPLE-TAB
Biobank sample tracking

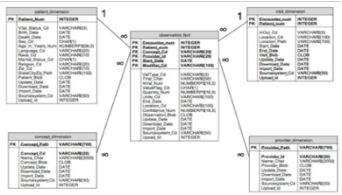


Dynamic
'runtime'

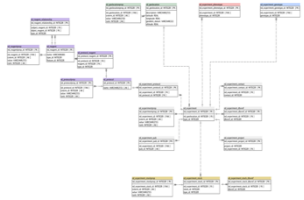
GSCF
Studies, flexible properties



i2b2
Observations on patients during doctor visit

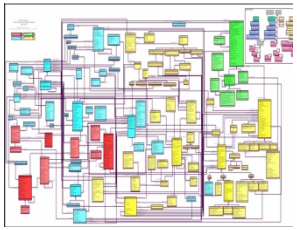


Chado NDM
Boilerplate for biological experiments and protocols

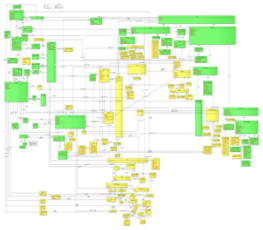


Exhaustive

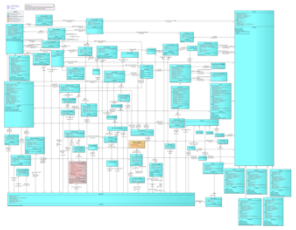
HL7
excerpt



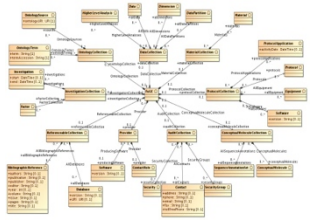
LS-DAM
excerpt



CDISC BRIDG
excerpt



FuGE
excerpt



molgenis

umcg university of groningen / genomics coordination center

What are we dealing with?

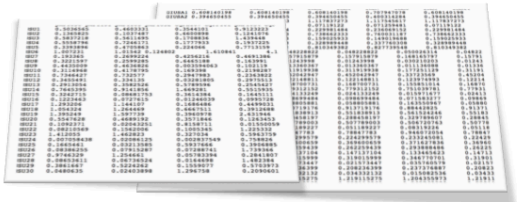
Genomic features, individuals, ontologies ..

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

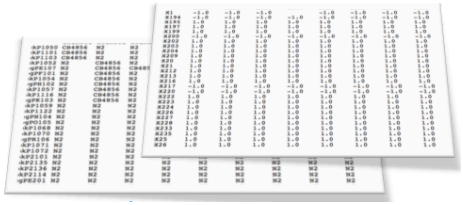
Annotations of concepts used in data sets, mostly static content

Stable?

Biomolecular measurements, association results ...



Genotypes, conditions ...



Experimental data sets, usually flexible and volatile content

Dynamic?

Apparently we need something stable AND dynamic
..without becoming exhaustive

molgenis

umcg

university of groningen / genomics coordination center

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*

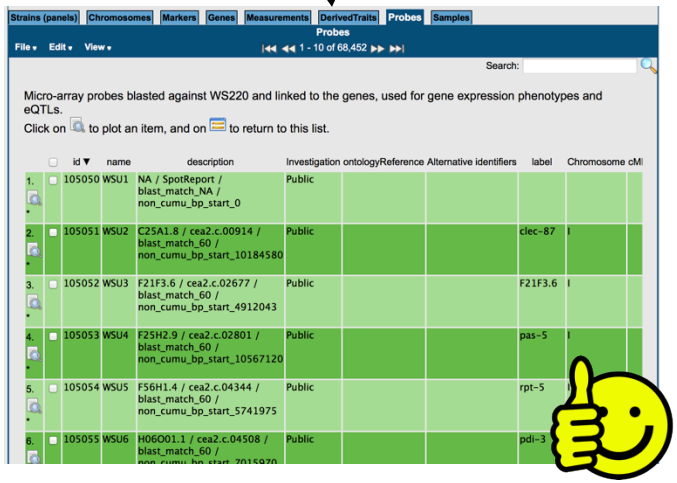
3. import

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="t
                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

molgenis



Challenge: Modeling 'dynamic' data sets

Can we use 'entity-attribute-value' modeling?

	BF221L	FD85C	C6L9	T7M24	BF151L
RIL1	A	A	A	B	B
RIL2	B	A	B	B	A
RIL3	B	B	A	B	A
RIL4	A	B	B	A	B
RIL5	A	B	A	A	B

Not stable columns!
Different per data set...

↓ 1. model

Values all comparable:
Columns all of the same type

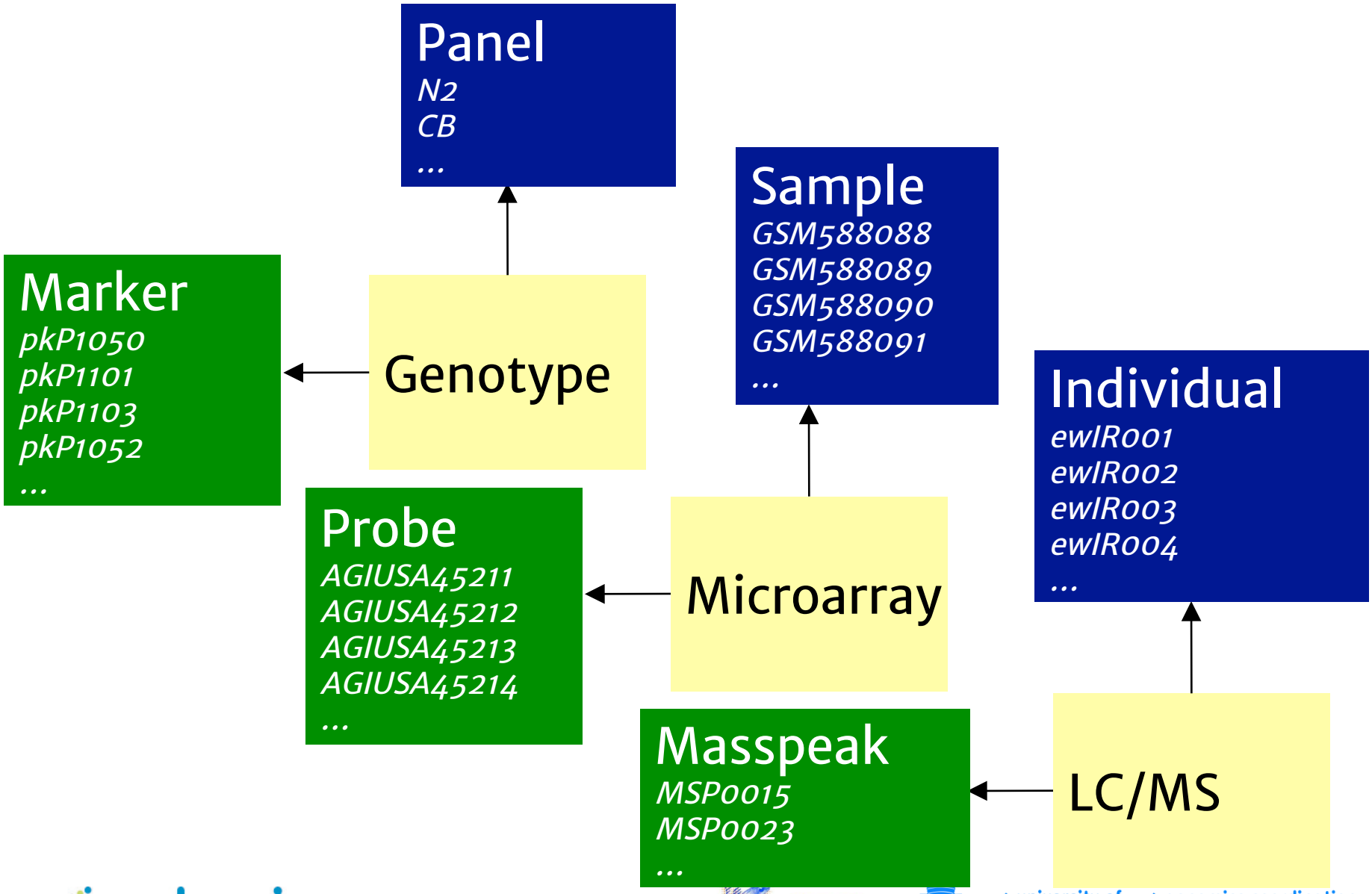
```
<entity name="Locus" abstract="true">
        <description> position. Typical examples of such traits are genes,
                probes and markers. Common structure for entities that have a
                genomic</description>
        <field name="Chromosome" label="Chromosome" type="xref"
                xref_entity="Chromosome" xref_field="id" xref_label="name" nillable="t
                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>
```

. . . ?

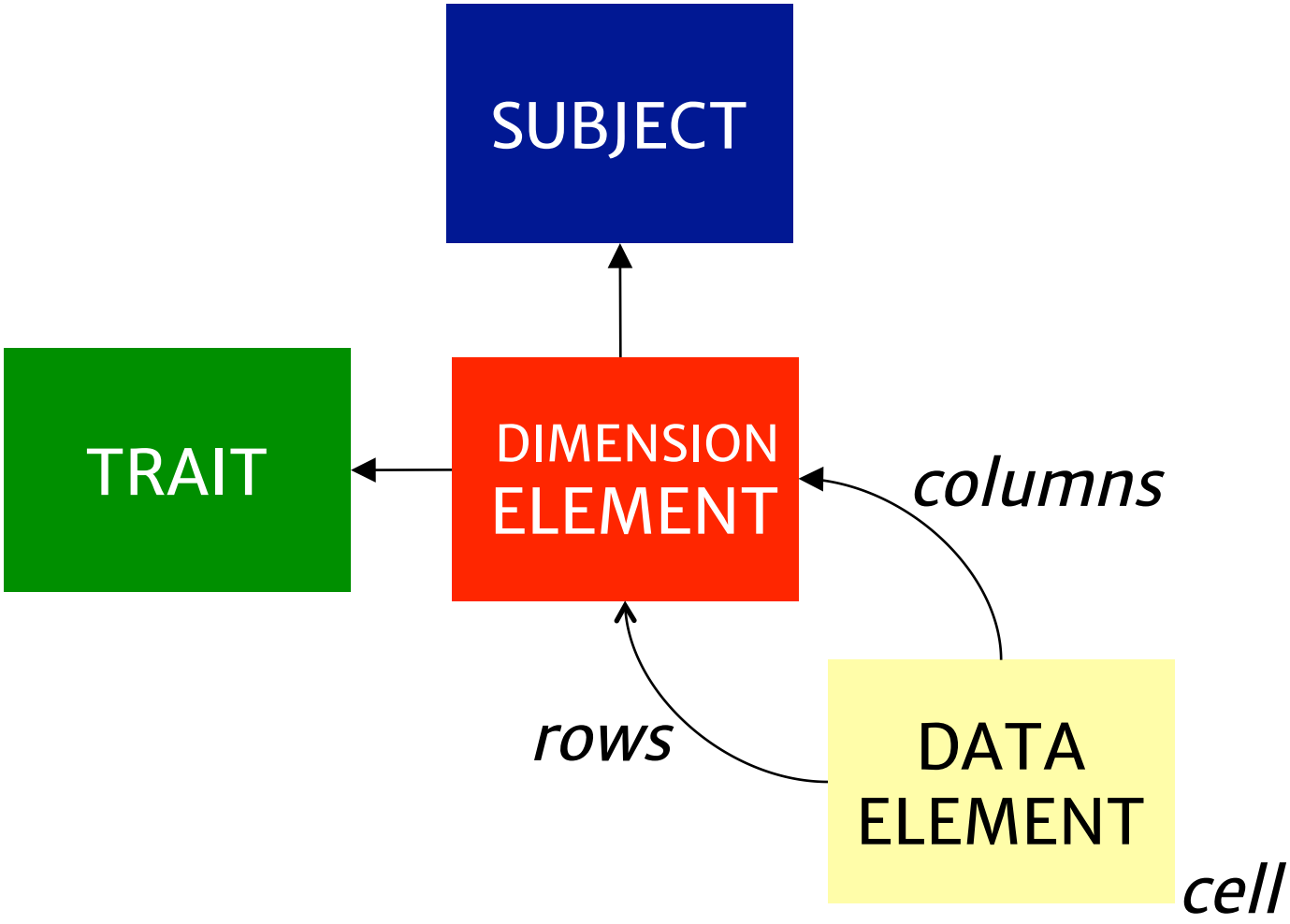
Results

XGAP model

Challenge: Data sets can be variable combinations

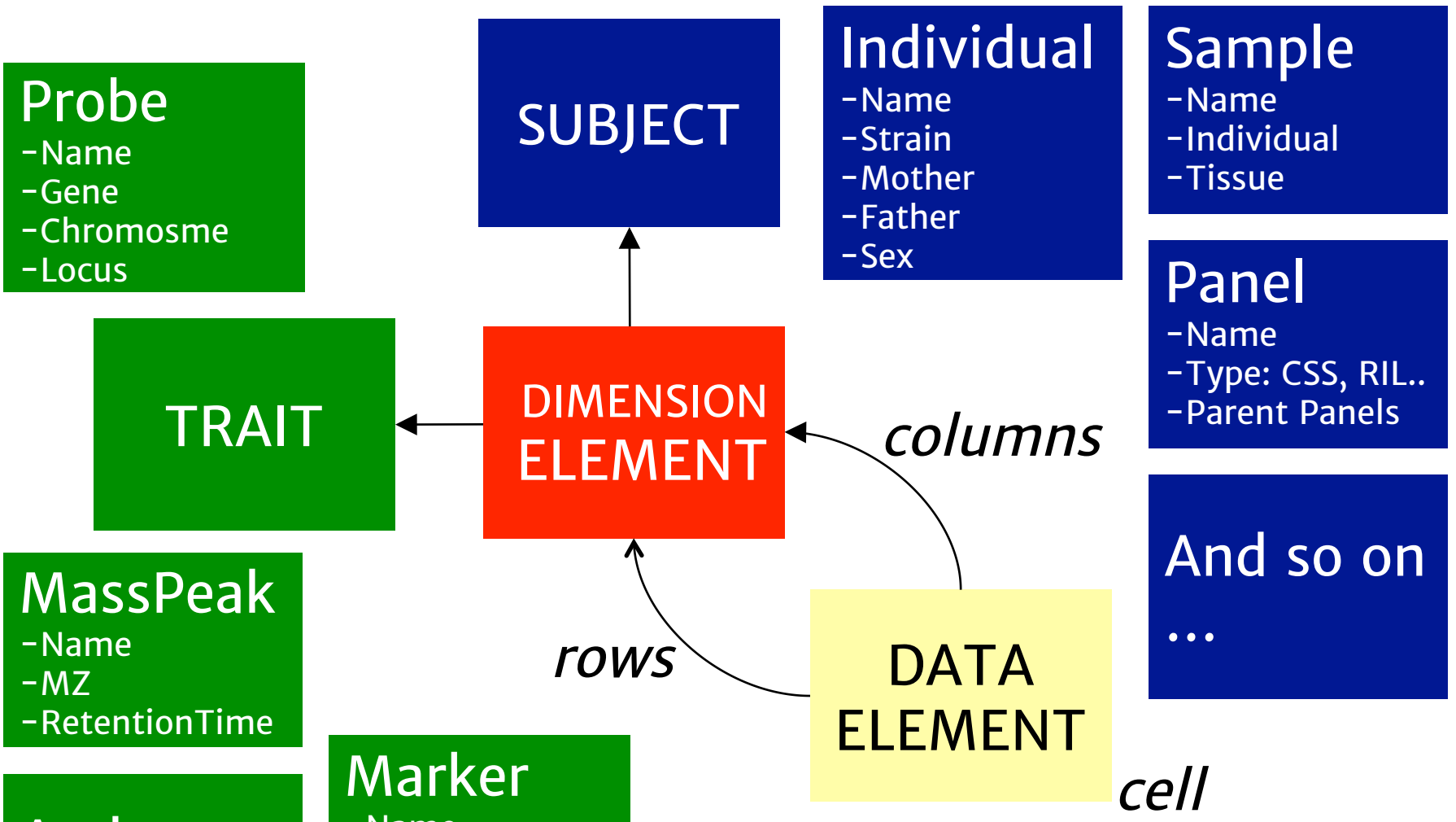


XGAP model: <any trait> X <any subject>



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

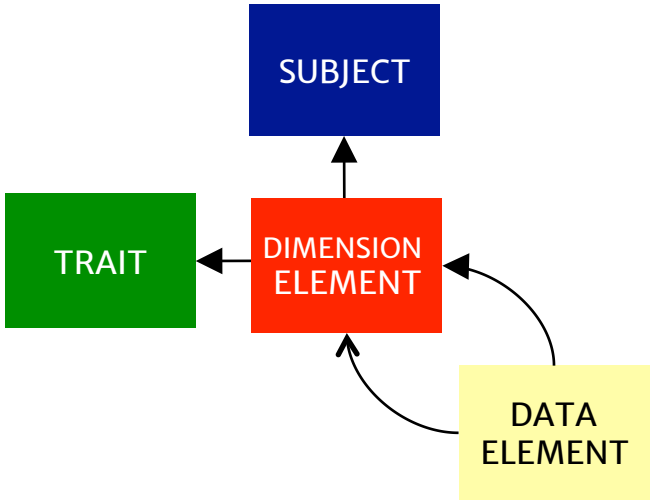
Extensible core model



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

umcg university of groningen genomics coordination center

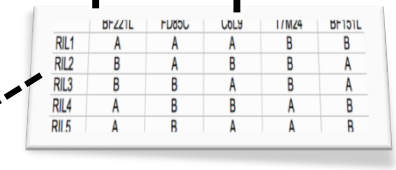
Using the XGAP model



1. model

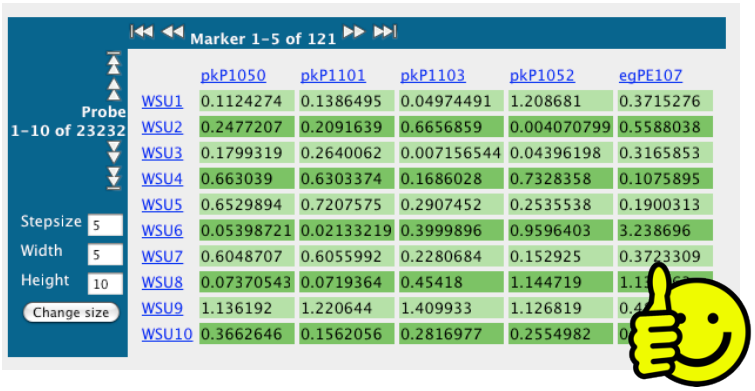
```
        <field name="Feature" type="xref" xref_entity="ObservationElement"
                xref_field="id" xref_label="name"
                description="References the ObservableFeature that this observation was ma
        <description>FIXME: change to ObservationTarget?</description>
        <field name="Target" type="xref" xref_entity="ObservationElement"
                xref_field="id" xref_label="name"
                description="References the ObservationTarget that this feature was made o
</entity>
<entity name="ObservedValue" implements="Observation">
        <description>
                Generic storage of values, relationships and optional ontology
                mapping of the value/relation. Values can be atomatic observations,
                e.g., length (feature) of individual 1 (target) = 179cm (value).
                Values can also be relationship values, e.g., extract (feature) of
                sample 1 (target) = derived sample (relation).
                <br />
                Discussion: how to model sample pooling in this model?
                <br />
                More Discussion: do we want to have type specific subclasses? No,
                because you can solve this by casting during querying?
        </description>
        <field name="ontologyReference" type="xref" xref_entity="OntologyTerm"
                nillable="true"
                description="(Optional) Reference to the
                ontology definition or 'code' for this value (recommended for non-numeric
                values such as codes)" />
```

2. generate



	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
RIL1	A	A	A	B	B
RIL2	B	A	B	B	A
RIL3	B	B	A	B	A
RIL4	A	B	B	A	B
RIL5	A	B	A	A	B

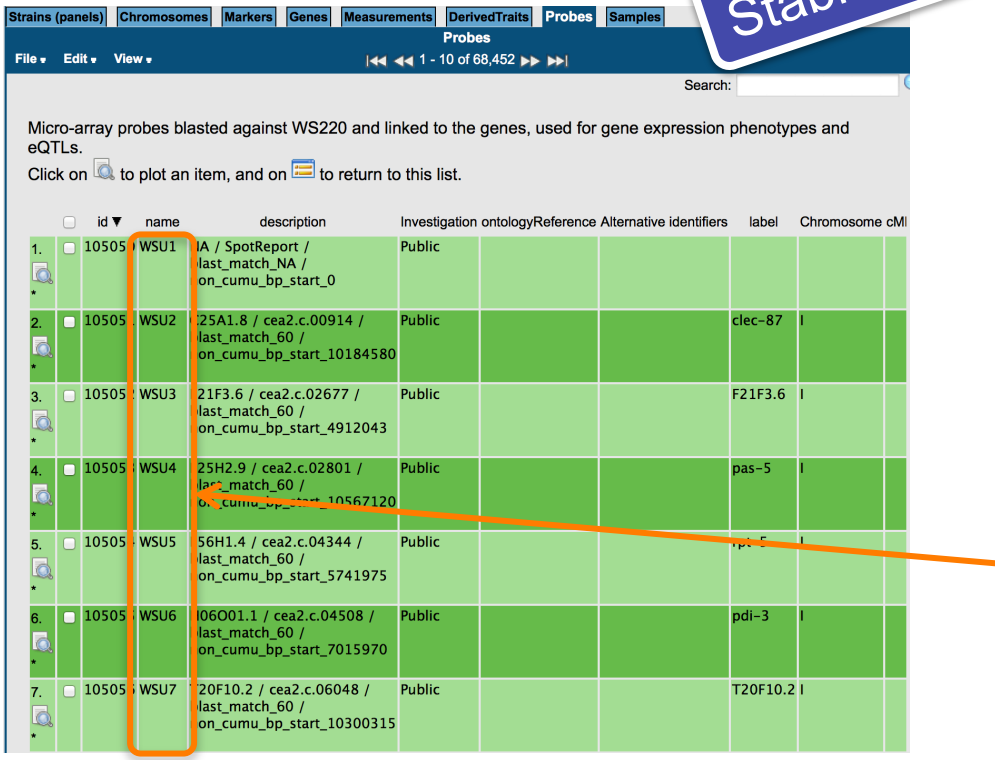
3. program viewer plugin

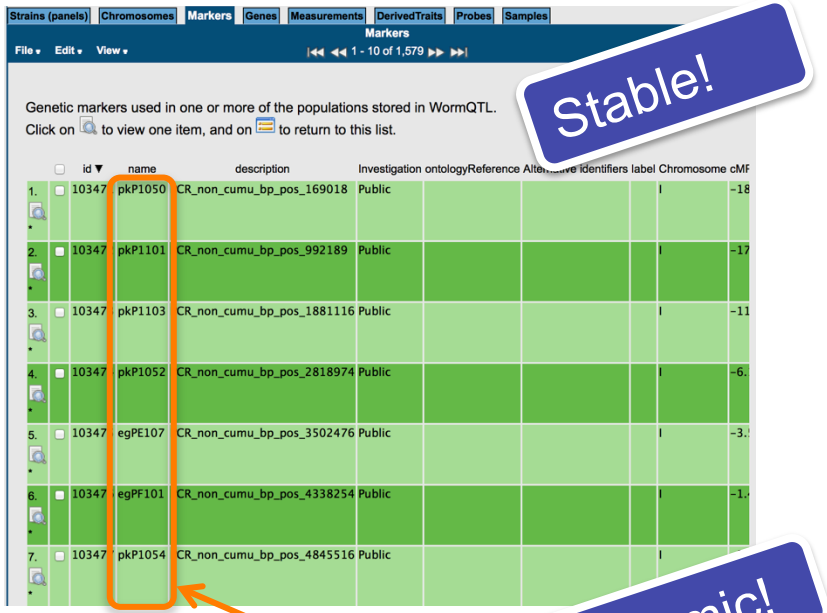


	pkP1050	pkP1101	pkP1103	pkP1052	egPE107
WSU1	0.1124274	0.1386495	0.04974491	1.208681	0.3715276
WSU2	0.2477207	0.2091639	0.6656859	0.004070799	0.5588038
WSU3	0.1799319	0.2640062	0.007156544	0.04396198	0.3165853
WSU4	0.663039	0.6303374	0.1686028	0.7328358	0.1075895
WSU5	0.6529894	0.7207575	0.2907452	0.2535538	0.1900313
WSU6	0.05398721	0.02133219	0.3999896	0.9596403	3.238696
WSU7	0.6048707	0.6055992	0.2280684	0.152925	0.3723309
WSU8	0.07370543	0.0719364	0.45418	1.144719	1.13
WSU9	1.136192	1.220644	1.409933	1.126819	0.4
WSU10	0.3662646	0.1562056	0.2816977	0.2554982	0

molgenis

Outcome: working application





molgenis

umcg

university of groningen

genomics coordination center

Results

Observ-OM model

So far..

- MOLGENIS toolbox
  - Good at rapidly generating databases for **stable** data
  - Values comparable within columns

- XGAP model
  - Good at **flexible homogeneous** data sets
  - All values in data set are semantically equivalent

"Molecules"

- *But how about..*
  - Capture **flexible heterogenous** data
  - Detailed meta-data (value type, unit, ontologies..)
  - Data lineage by protocol & application thereof

"Phenotypes"

Repeat

Can we use 'entity-attribute-value' modeling?

Individual	Gender	Age	Height	Body Weight	Diastolic_Blood_Pressure	Systolic_Blood_Pressure	Cervical_Carcinoma
Individual_1	female	56	180	65	90	140	yes
Individual_2	female	45	178	75	87	130	no
Individual_3	male	65	168	100	78	125	no
Individual_4	male	35	178	45	100	150	no
Individual_5	male	34	190	55	68	134	yes
Individual_6	female	24	185	67	76	134	yes
Individual_7	female	20	179	80	102	145	yes

The columns should be independent (more flexible)

Values NOT all comparable: Columns are NOT of the same type

1. model

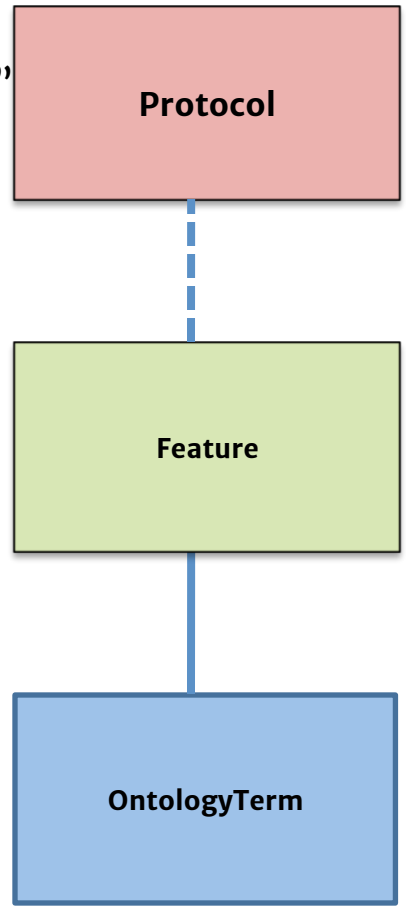
```
<entity name="Locus" abstract="true">
        <description> position. Typical examples of such traits are genes,
                probes and markers. Common structure for entities that have a
                genomic</description>
        <field name="Chromosome" label="Chromosome" type="xref"
                xref_entity="Chromosome" xref_field="id" xref_label="name" nillable="t
                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>
```

. . . ?

Features become columns

- Feature ~= columns, protocols ~= tables

Protocol
Measuring "Gender", "Age", "Height", "Blood pressure"
in some medically standardized way (ISO-54532)



Feature "Gender" Categr. M/F/O

Feature "Age" Years Integer

Feature "Height" Centimeter Decimal

Feature "Blood pressure" mmHg Decimal

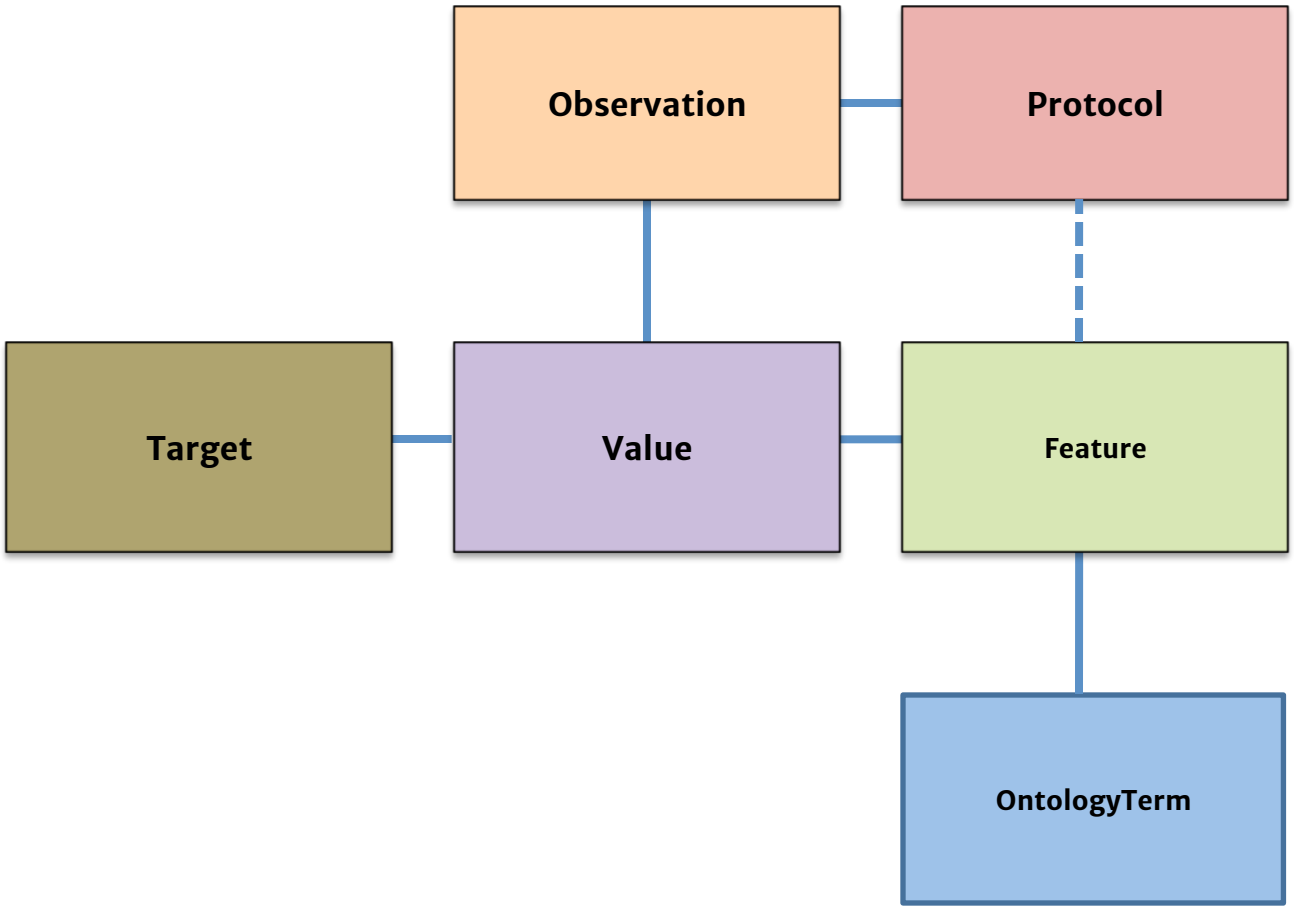
Ontology
MeSH term
"Gender Identity"

Ontology
MeSH term
"Blood pressure"

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

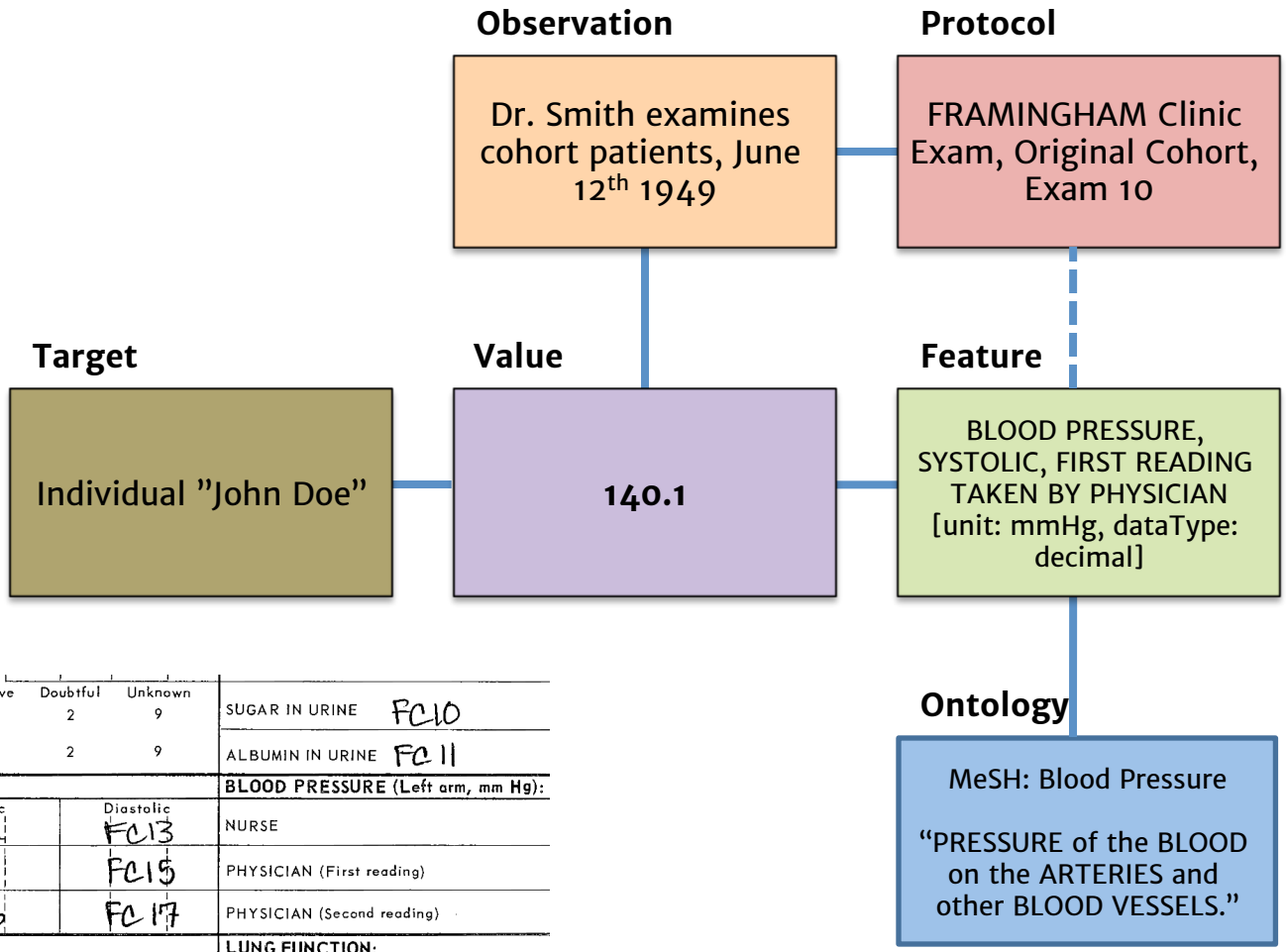
Observations become table rows

- Targets are patients, samples, groups, etc



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

Data example



31	Negative 0	Positive 1	Doubtful 2	Unknown 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
44	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)

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

molgenis

umcg university of groningen genomics coordination center

Individual	Gender	Age	Height	Body Weight	Diastolic_Blood_Pressure	Systolic_Blood_Pressure	Cervical_Carcinoma	Breast_Carcinoma
Individual_1	female	56	180	65	90	140	yes	no
Individual_2	female	45	178	75	87	130	no	yes
Individual_3	male	65	168	100	78	125	no	yes
Individual_4	male	35	178	45	100	150	no	yes
Individual_5	male	34	190	55	68	134	yes	yes
Individual_6	female	24	185	67	76	134	yes	no
Individual_7	female	20	179	80	102	145	yes	no
Individual_8	female	34	175	56	76	134	yes	yes
Individual_9	female	45	181	67	90	144	yes	yes
Individual_10	male	34	160	45	86	132	yes	yes
Individual_11	male	35	194	75	70	123	yes	yes
Individual_12	female	54	182	55	69	120	yes	yes
Individual_13	male	33	170	66	65	121	yes	yes
Individual_14	female	24	180	77	66	122	yes	yes
Individual_15	female	27	170	56	64	125	yes	yes
Individual_16	female	36	167	54	63	127	yes	yes
Individual_17	male	35	186	76	70	131	yes	yes
Individual_18	male	43	190	88	59	119	yes	yes
Individual_19	male	44	168	65	60	120	yes	yes
Individual_20	female	36	175	54	63	121	yes	yes

Ontologies

MeSH: Blood Pressure

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

Protocol

Observation

Features

Targets

Values

Individual	Gender	Age	Height	Body Weight	Diastolic_Blood_Pressure	Systolic_Blood_Pressure	Cervical_Carcinoma	Breast_Carcinoma
Individual_1	female	56	180	65	90	140	yes	no
Individual_2	female	45	178	75	87	130	no	yes
Individual_3	male	65	168	100	78	125	no	yes
Individual_4	male	35	178	45	100	150	no	yes
Individual_5	male	34	190	55	68	134	yes	yes
Individual_6	female	24	185	67	76	134	yes	no
Individual_7	female	20	179	80	102	145	yes	no
Individual_8	female	34	175	56	76	134	yes	yes
Individual_9	female	45	181	67	90	144	yes	yes
Individual_10	male	34	160	45	86	132	yes	yes
Individual_11	male	35	194	75	70	123	yes	yes
Individual_12	female	54	182	55	69	120	yes	yes
Individual_13	male	33	170	66	65	121	yes	yes
Individual_14	female	24	180	77	66	122	yes	yes
Individual_15	female	27	170	56	64	125	yes	yes
Individual_16	female	36	167	54	63	127	yes	yes
Individual_17	male	35	186	76	70	131	yes	yes
Individual_18	male	43	190	88	59	119	yes	yes
Individual_19	male	44	168	65	60	120	yes	yes
Individual_20	female	36	175	54	63	121	yes	yes

Building G2P applications

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

dystrophic eb

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

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

PANACEA xQTL workbench

WormQTL

CropQTL: Learning From Nature
xQTL workbench

Parel IBD xQTL workbench

LifeLines Research Portal

TI FOOD NUTRITION research portal

EURATRANS xQTL workbench

molgenis

umcg

university of groningen

genomics coordination center

Showcase

EB Registry

EB Registry: Dystr. EB mutation database

International dystrophic eb Patient Registry

International registry of patients with dystrophic epidermolysis bullosa and database of associated COL7A1 mutations

Search | Submit data | Contact | References | Background | News | Login

Search

Welcome to the international registry of dystrophic epidermolysis bullosa (DEB) patients and associated COL7A1 mutations.

The International Dystrophic Epidermolysis Bullosa Patient Registry contains anonymised data on both published and unpublished DEB patients, as well as their associated COL7A1 mutations and genotypes, and clinical and molecular phenotypes.

The database currently contains 590 DEB patients, of which 71 unpublished, and 395 COL7A1 mutations. Search or browse below.

Search registry

Search by typing any search term in the search field, like cDNA (e.g. "3G>T") or protein (e.g. "Arg525Ter") notations of mutations, mode of inheritance (e.g. "dominant") or specific phenotypes (e.g. "severe generalized"). Search results are shown at bottom of page.

Enter search term: Advanced Search

(•) Show mutations
( ) Show patients

Search | Display all mutations | Display all patients

Browse the COL7A1 gene

Click anywhere on this schematic representation of the COL7A1 gene to graphically browse the gene. With every click you will zoom in deeper on the COL7A1 gene. Mutated nucleotides are depicted in red. If the cursor is placed over the mutated nucleotide(s), the corresponding mutation is shown.

News

Patient and Mutation Update
14 Australasian patients added to Registry.
Jul 20, 2011
More

Article about DEB Registry online in Human Mutation.
Go to Pubmed to view article
Jun 16, 2011
More

Feature Update
Select type of search result tables.
Jun 9, 2011
More

Name change: International Dystrophic Epidermolysis Bullosa Patient Registry
Important update.
May 25, 2011
More

Patient and Mutation Update
22 Tunisian RDEB patients entered to database.
May 24, 2011
More

molgenis umcg university of groningen genomics coordination center

EB Registry: Dystr. EB mutation database

Phenotypic details for patient 'P10'

target

Characteristics

value

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

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

Showcase

WormQTL

WormQTL.org

PANACEA xQTL workbench

WormQTL

Home | Contact | About | **Article in NAR**

Home | Find QTLs | Genome browser | Browse data | Help

provide feedback:

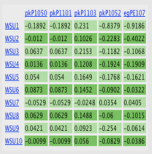
Home

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



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

molgenis

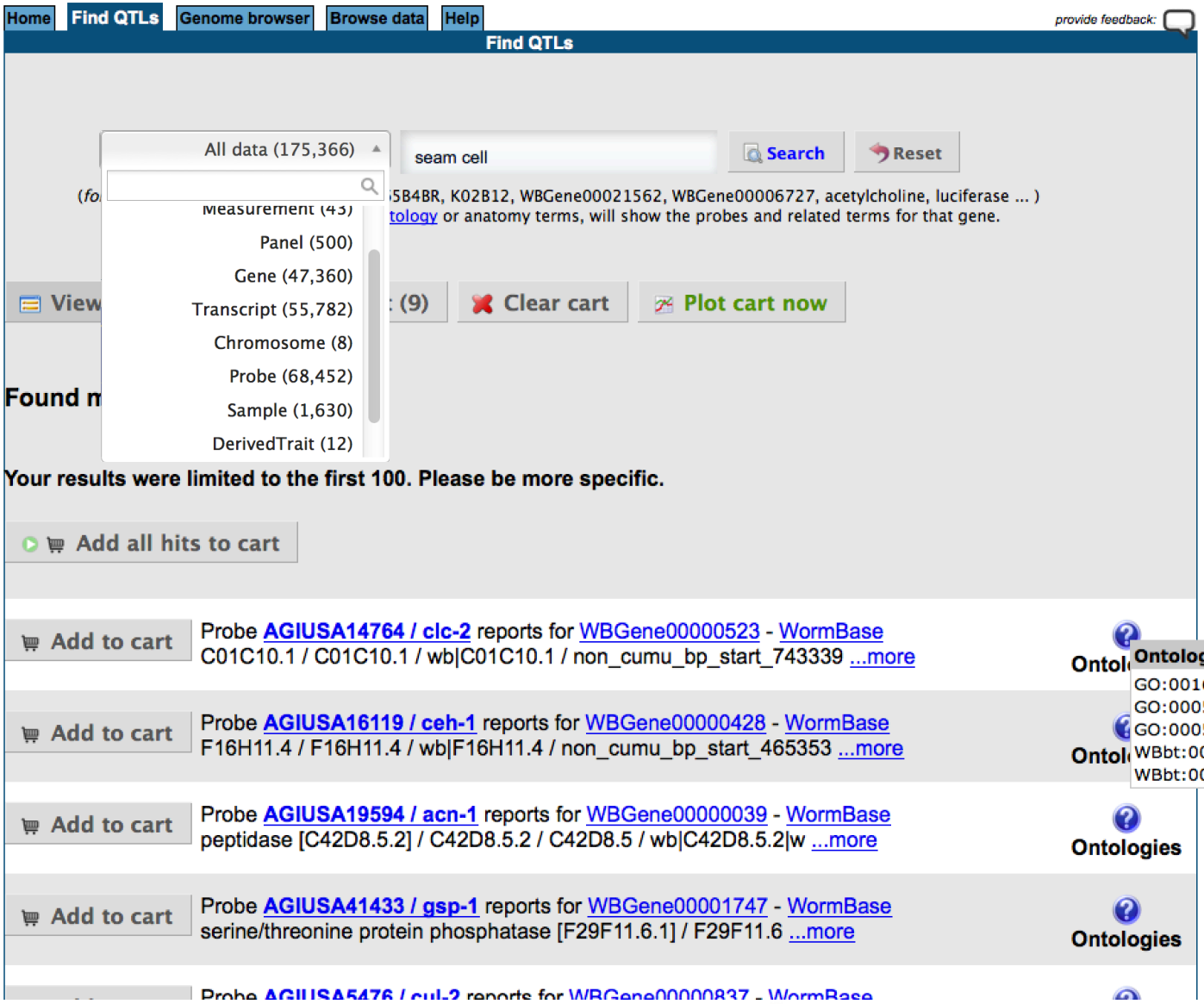
umcg

university of groningen / genomics coordination center

WormQTL.org

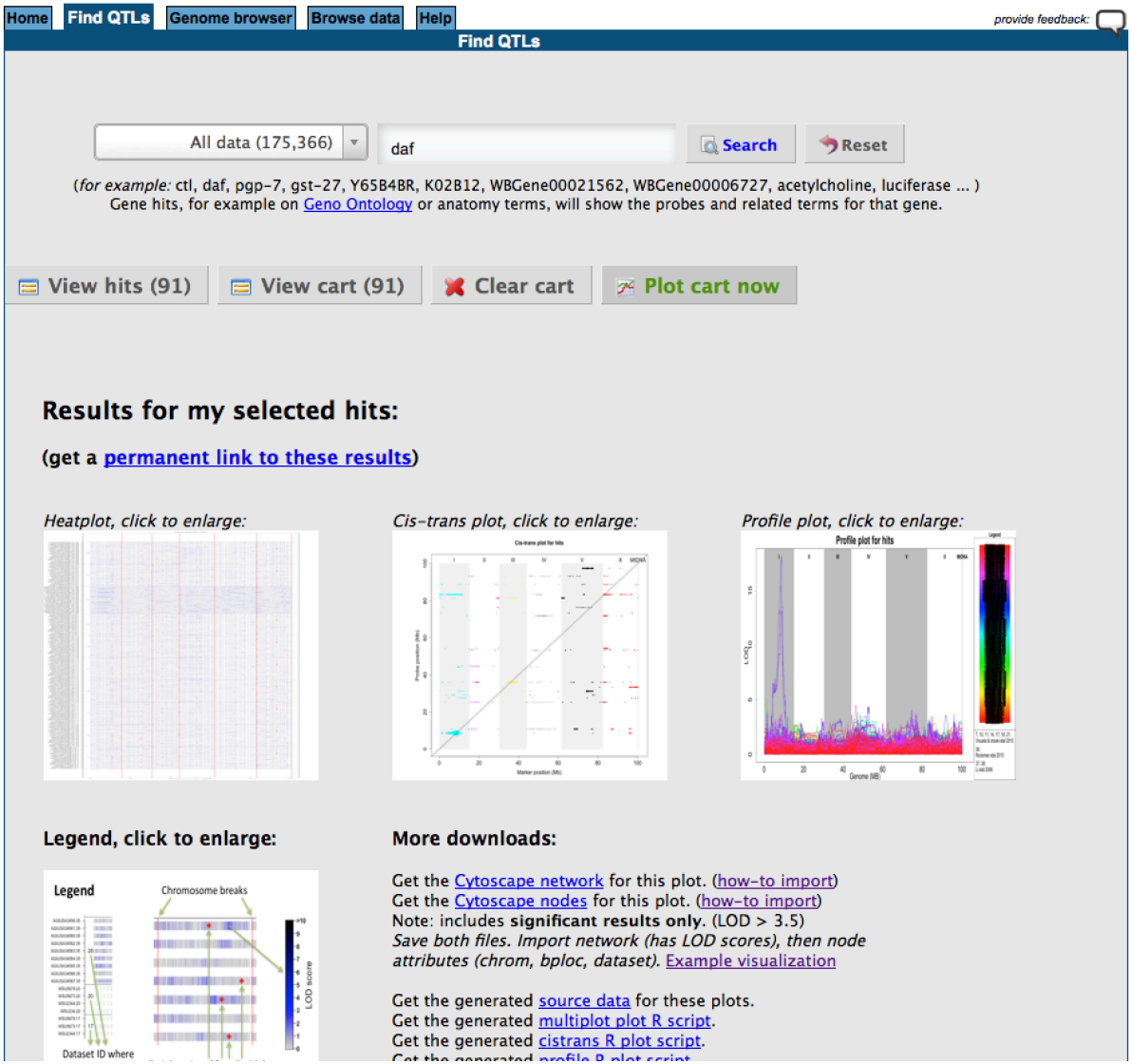
Phenotypes	*Type of array*	*Sample size*	*Parental strains*	*Reference*	*Pubmed link*	*Growing temperature*	*Stage*	*Food*	*Medium*	*Dataset IDs*
Gene expression	Washington State University	2x40 RILs	CB4856; N2	**Li et al. 2006;** Mapping determinants of gene expression plasticity by genetical genomics in C. elegans.	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 Caenorhabditis elegans.	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 C. elegans.	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 RIAILs	CB4856; N2	**Rockman et al. 2010;** Selection at linked sites shapes heritable phenotypic variation in C. elegans.	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 C. elegans 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 Caenorhabditis elegans.	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 Caenorhabditis elegans.	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 C. elegans 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 Caenorhabditis elegans.	22670229	24oC	(40h, 96h and 214h) L4, Adult, Old	OP50	NGM Plate	41, 42, 43

WormQTL.org



molgenis umcg university of groningen genomics coordination center

WormQTL.org



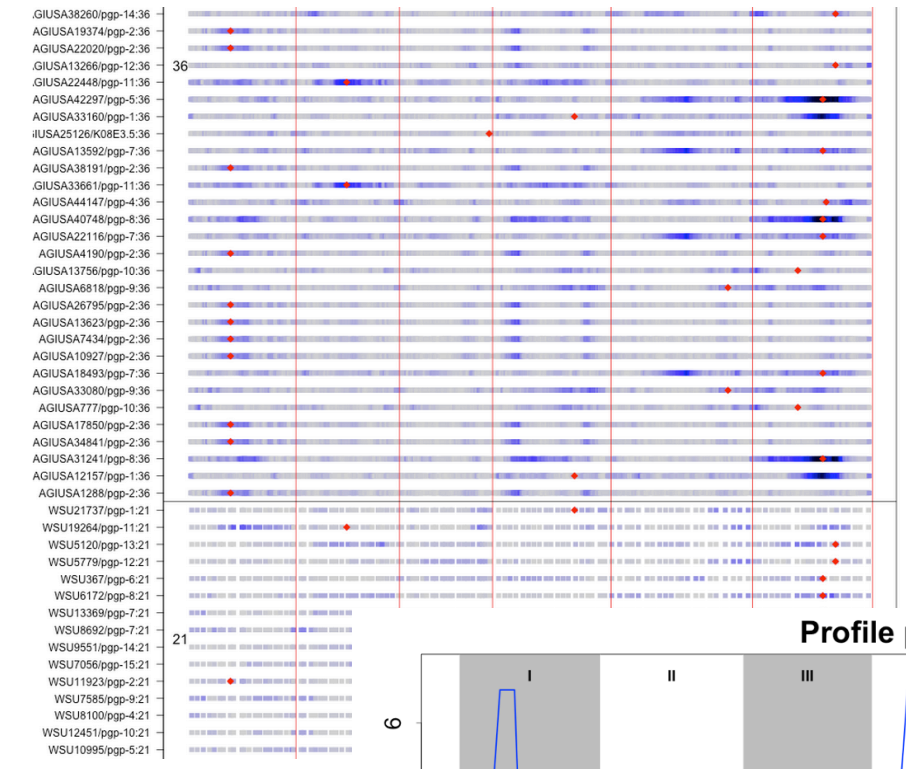
molgenis

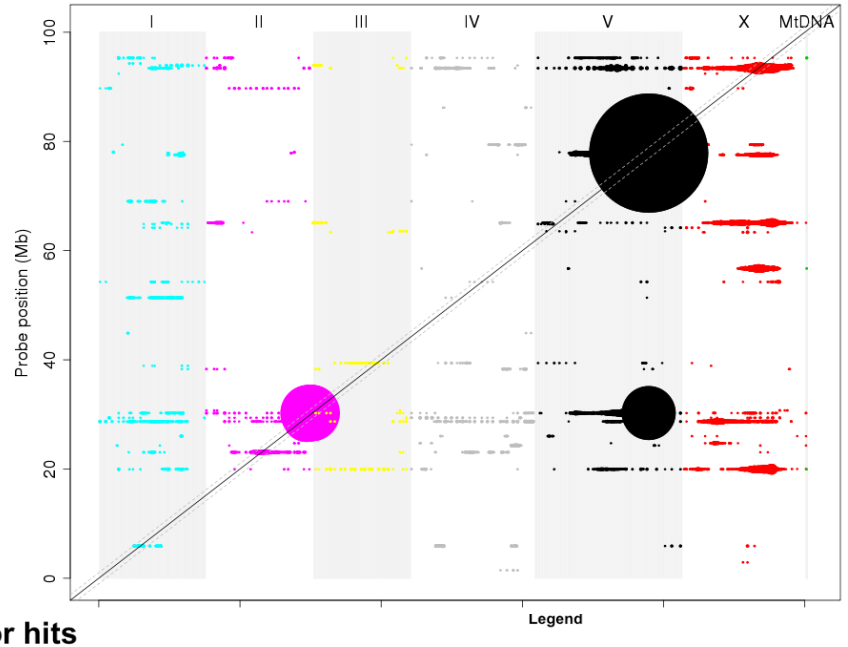
umcg

university of groningen

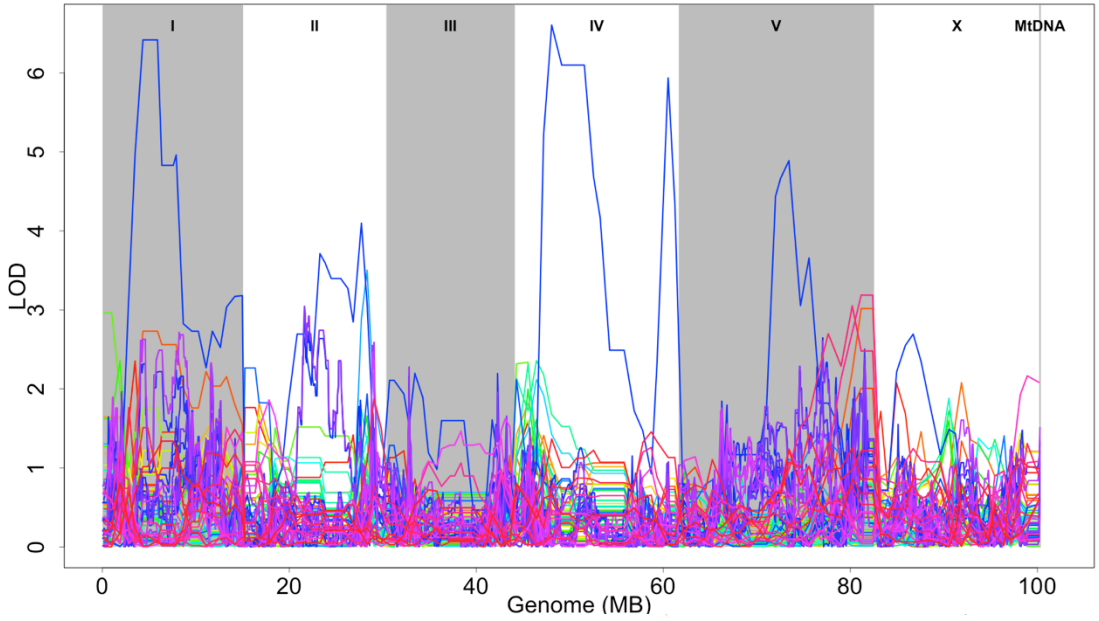
genomics coordination center

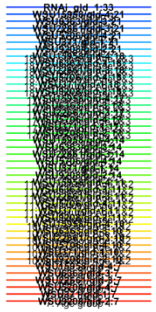
WormQTL.org





Profile plot for hits

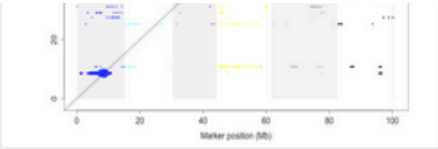


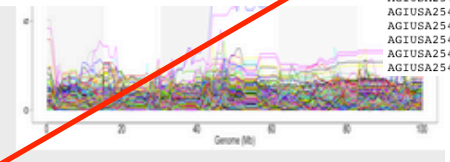


molge

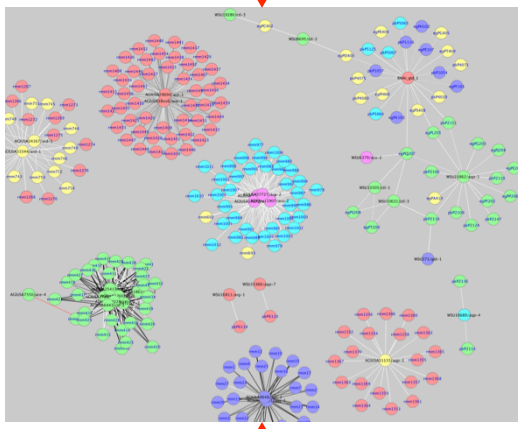
nics coordination

WormQTL.org

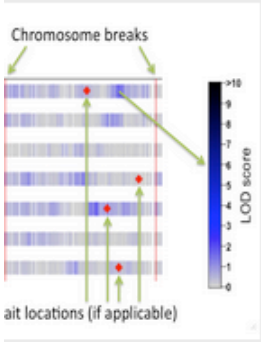




```
WSU6355/daf-3   QTL     pkP1101 4.6026
WSU6355/daf-3   QTL     pkP1103 3.5561
WSU6355/daf-3   QTL     pkP1050 5.0597
WSU6355/daf-3   QTL     pkP1101 5.0597
WSU7870/daf-16  QTL     egPD403 3.9109
WSU6355/daf-3   QTL     pkP1050 3.9762
WSU6355/daf-3   QTL     pkP1101 3.9762
WSU3844/daf-3   QTL     egPK601 3.6122
WSU7870/daf-16  QTL     pkP5071 3.5478
AGIUSA25467/daf-36      QTL     rmm743  3.893693484
AGIUSA25467/daf-36      QTL     rmm744  4.302316576
AGIUSA25467/daf-36      QTL     rmm745  4.302316576
AGIUSA25467/daf-36      QTL     rmm746  4.302316576
AGIUSA25467/daf-36      QTL     rmm747  4.302316576
AGIUSA25467/daf-36      QTL     rmm748  4.302316576
AGIUSA25467/daf-36      QTL     rmm749  4.302316576
AGIUSA25467/daf-36      QTL     rmm750  3.824765495
```



k 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.

```
WSU6355/daf-3   trait   X       83377539        age12_int_qtl
pkP1050 marker  I       169018  age12_int_qtl
pkP1101 marker  I       992189  age12_int_qtl
pkP1103 marker  I       1881116 age12_int_qtl
WSU7870/daf-16  trait   I       10763660        age23_int_qtl
egPD403 marker  IV      47198842        age23_int_qtl
WSU3844/daf-3   trait   X       83371790        age3_qtl
egPK601 marker  X       91884289        age3_qtl
pkP5071 marker  V       76788020        age3_qtl
AGIUSA25467/daf-36      trait   V       71855469        rock_qtl
rmm743  marker  IV      52297746        rock_qtl
rmm744  marker  IV      52364173        rock_qtl
rmm745  marker  IV      52422112        rock_qtl
rmm746  marker  IV      52482040        rock_qtl
rmm747  marker  IV      52569872        rock_qtl
rmm748  marker  IV      52584390        rock_qtl
rmm749  marker  IV      52626947        rock_qtl
rmm750  marker  IV      52705021        rock_qtl
rmm751  marker  IV      52722889        rock_qtl
rmm752  marker  IV      52809758        rock_qtl
```

/ daf-1 [explore deeper] - protein kinase [F29C4.1b] / F29C4.1b / F...
/ daf-3 [explore deeper] - F25E2.5b.3 / F25E2.5 / wb|F25E2.5b.3|wb|...
af-11 [explore deeper] - B0240.3 / cea2 p.107079 / blast_match_60

molgenis

umcg

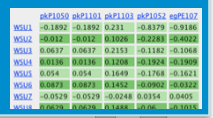
university of groningen / genomics coordination center

Current work

OmicsConnect running on Observ-OMX

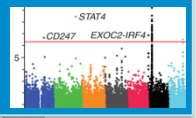
Catalogue
Find data item and sample collections

Data
Filter individual data sets and download to Excel & SPSS



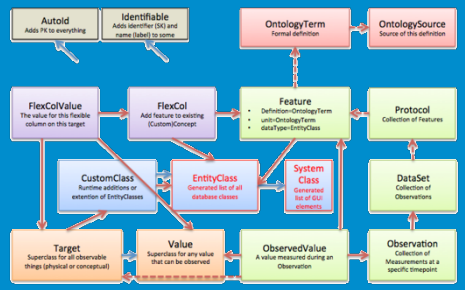
Compute
Run analysis workflows on big data compute infra-structure

GWAS Central
Explore summary level GWAS data

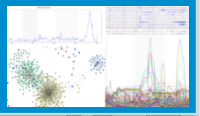


Protocol
CRFs, Questionnaires, Lab protocols, and assays

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 patho-genicity effects

Organization
Institutes, Departments, People, Locations & Containers

File
File storage and drivers for images and data

molgenis

umcg

university of groningen / genomics coordination center

Acknowledgements

Martijn Dijkstra
Despoina Antonakaki
Tomasz Adamuziak
Rob Hastings
Sirisha Gollapudi
Gudmundur Thorisson
Chao Pang
Myles Byrne
David van Enckvort
Linda Mook
Pieter Neerincx
Ger Strikwerda
Danny Arends
Roan Kanninga
Jan Bot
George Byelas
Yang Li
Basten Snoek
Noortje Festen
Konrad Zych

And more ...

Lude Franke
Juha Muilu
Anthony Brookes
Helen Parkinson
Vincent Ferreti
Gert-Jan van Ommen
Jan Jurjen Uitterdijk
Ritsert C. Jansen
Jan Kammenga
Cisca Wijmenga
Paul de Bakker
Irene Nooren
Rob Hooft
Salome Scholtens
Hans Hillege
Ronald Stolk
Morris Swertz

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 consoritum (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

Thank you!
Questions?

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

molgenis.org
Your database at the push of a button

Summary

- Complex variation, geno-to-pheno
- Exploiting the data requires structure
- Best-of flexible and stable parts
- Support homo- and heterogeneous data

Read more

- MOLGENIS: http://www.molgenis.org
- MOLGENIS Compute: http://www.molgenis.org/wiki/ComputeStart
- 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