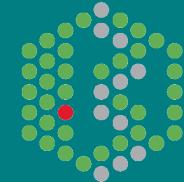


A picture paints a thousand traits

Tony Burdett, Functional Genomics Production Team
Bio-Ontologies SIG, ISMB, 13th July 2012

EMBL-EBI

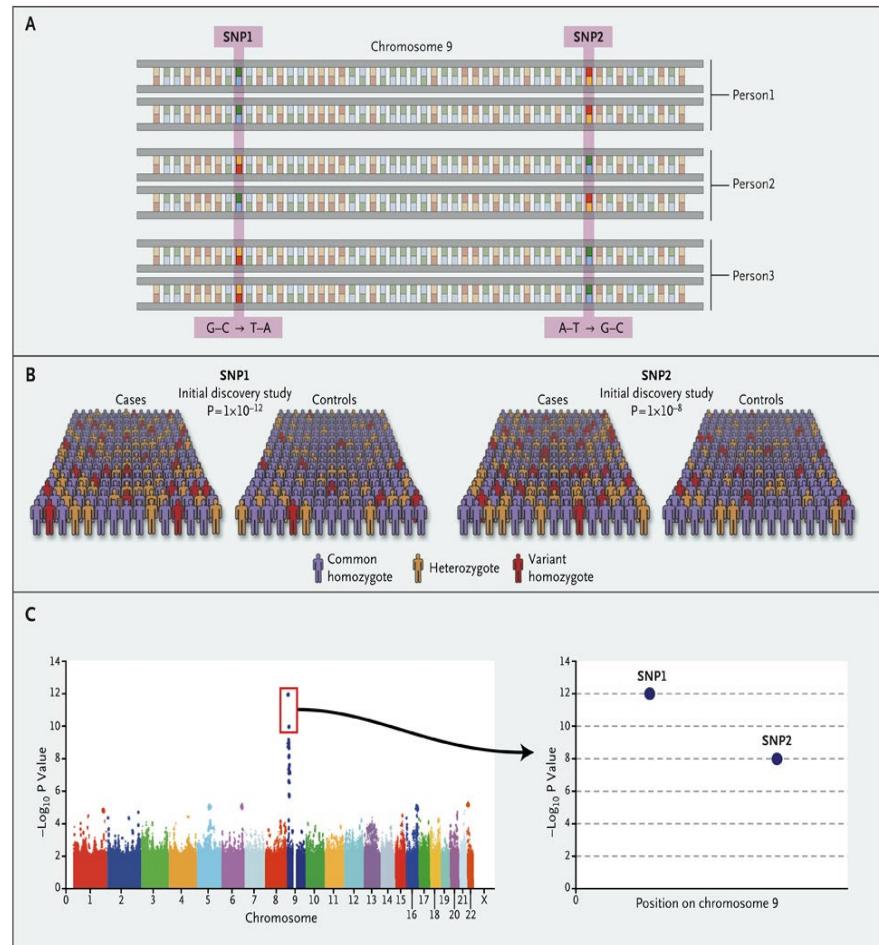


Overview

- Introduction to the GWAS catalogue
- EBI/NHGRI project aims
- Ontology and GWAS KnowledgeBase
- GWAS diagram
- Outlook

Genome-Wide Association Studies

- Hundreds of thousands of SNPs tested for association with a trait or disease in large population samples
- Insights include
 - New potential risk loci in or near genes not previously associated with a disease
 - New associations between traits or diseases not previously thought to have any common causes or share pathways
 - Additional information on chromosomal regions annotated as gene poor



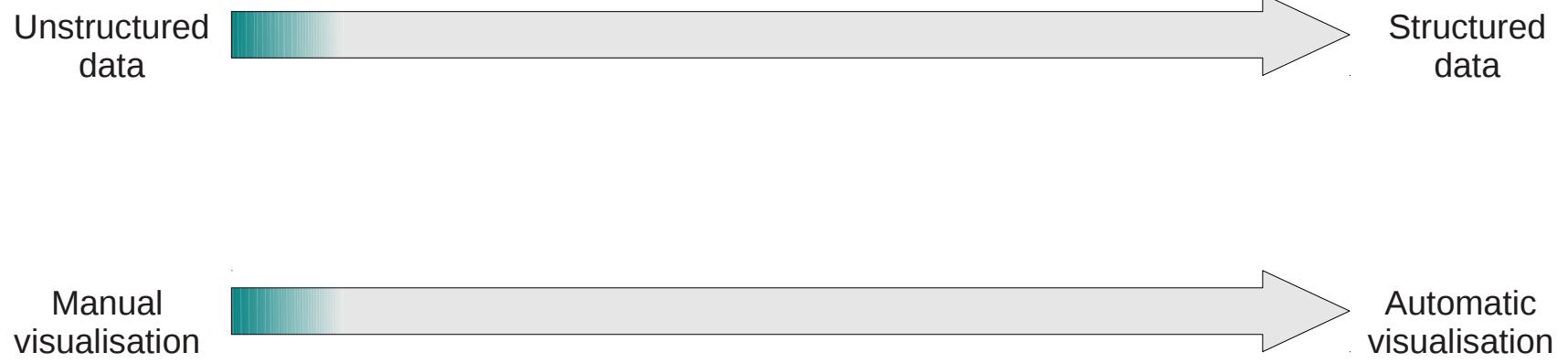
GWAS catalog

- As of July 2012, the catalog includes 1293 publications and 6551 SNPs, with 655 distinct traits and nearly 8000 SNP-trait associations

Date Added to Catalog (since 11/25/08)	First Author/Date/Journal/Study	Disease/Trait	Initial Sample Size	Replication Sample Size	Region	Reported Gene(s)	Mapped Gene(s)	Strongest SNP-Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform [SNPs passing QC]	CNV
11/09/11	Khor CC October 16, 2011 <i>Nat Genet</i> Genome-wide association study identifies susceptibility loci for dengue shock syndrome at <i>MICB</i> and <i>PLCE1</i> .	Dengue shock syndrome	2,008 Vietnamese pediatric cases, 2,018 Vietnamese controls	1,737 Vietnamese cases, 2,934 Vietnamese controls	6p21.33 10q23.33 16p13.3 8q11.23	<i>MICB</i> <i>PLCE1</i> <i>NR</i> <i>NR</i>	<i>MICB</i> <i>PLCE1</i> <i>RBFOX1</i> <i>SEC11B</i> - <i>RP1</i> <i>rs10104997-?</i>	<i>rs3132468-?</i> <i>rs3765524-?</i> <i>rs6500818-?</i> <i>rs10104997-?</i>	intron missense intron intergenic	0.13 0.70 NR NR	4×10^{-11} 3×10^{-10} 2×10^{-7} 9×10^{-7}	1.34 [1.23-1.46] 1.25 [1.16-1.33] 1.31 [NR] 1.2 [NR]	Illumina [481,342]	N

- The data itself is highly curated by expert biologists
- But the structure of the data is loose and the semantics not explicitly defined





GWAS traits

- GWAS catalogue traits previously only available as an unstructured list

Dental caries

Depression and alcohol dependence

Depression--quantitative trait

Diabetes (gestational)

Diabetes (incident)

Diabetes related insulin traits

Diabetic nephropathy

Diabetic retinopathy

Dialysis-related mortality

Diastolic blood pressure

Digit length ratio

Dilated cardiomyopathy

Fatty acids-blood pressure (FABP)

Tuberculosis

Two-hour glucose challenge

Type 1 diabetes

Type 1 diabetes autoantibodies

Type 2 diabetes

Type 2 diabetes and 6 quantitative traits

Type 2 diabetes and gout

Type 2 diabetes and other traits

Irritative colitis

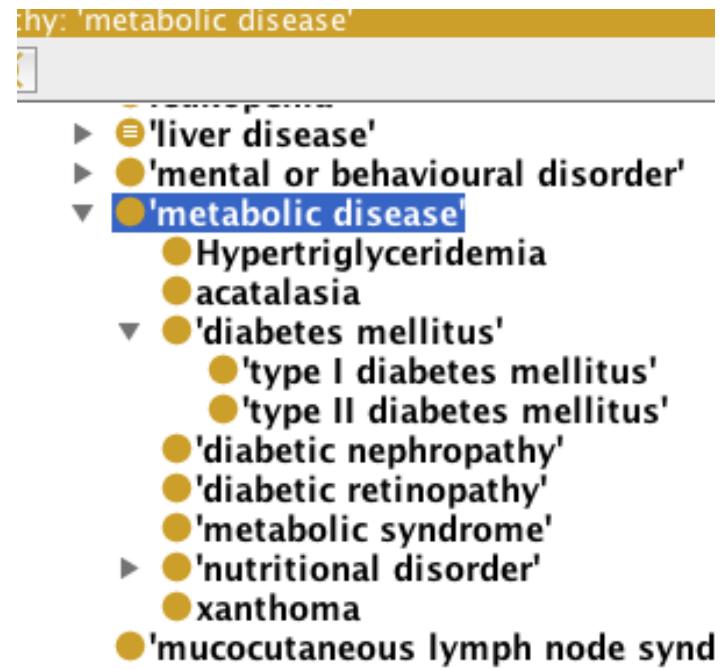
- Traits are highly diverse
 - Phenotypes, e.g. hair colour
 - Treatment responses, e.g. response to anti-neoplastic agents
 - Diseases, e.g. type 2 diabetes
 - Chemical/drug names e.g. C-reactive protein
- Traits are often compound and/or context-dependent, e.g. “Type 2 diabetes and gout” or “Parkinson’s disease (interaction with caffeine)”

Ontology

- Integration of GWAS traits into an ontology
- Additional semantically meaningful links between traits allows much more complex and extensive querying

“Show me all SNPs associated with type 2 diabetes and metabolic syndrome”

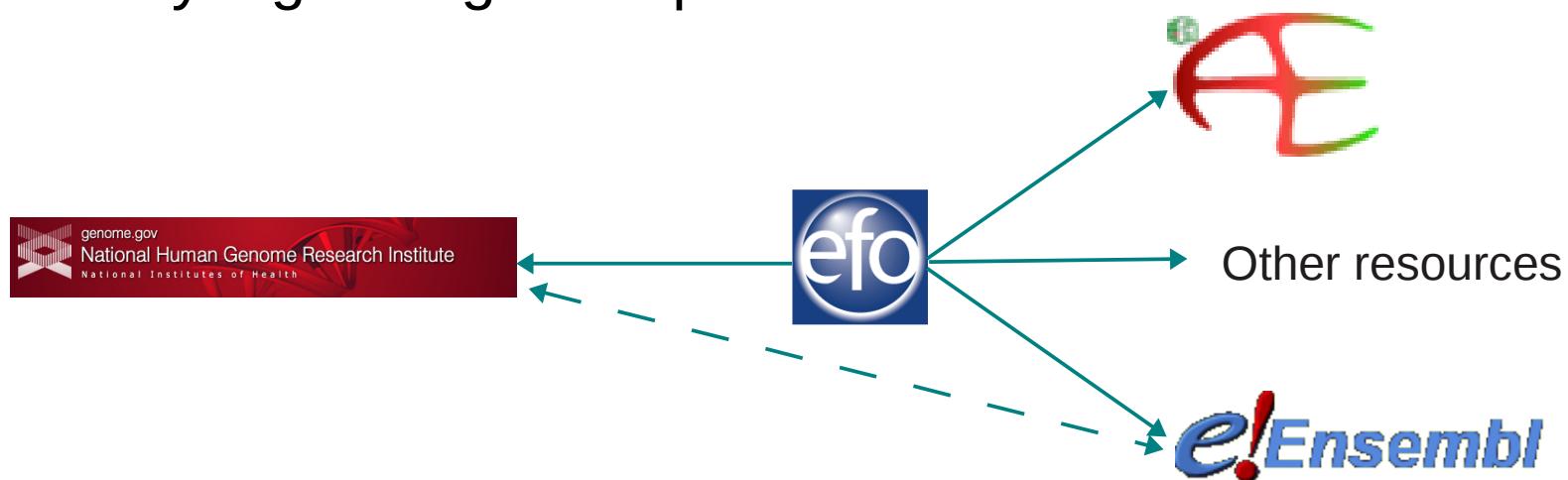
- Two options for ontology integration
 - Create new “GWAS ontology”
 - Integrate with an existing ontology



Integration with “Experimental Factor Ontology”



- EFO is actively developed
- Well-suited to covering diversity of GWAS traits
- 20% of GWAS traits already found in EFO prior to integration process
- 300 new terms added over 5 releases = 80% coverage
- Very high integration potential



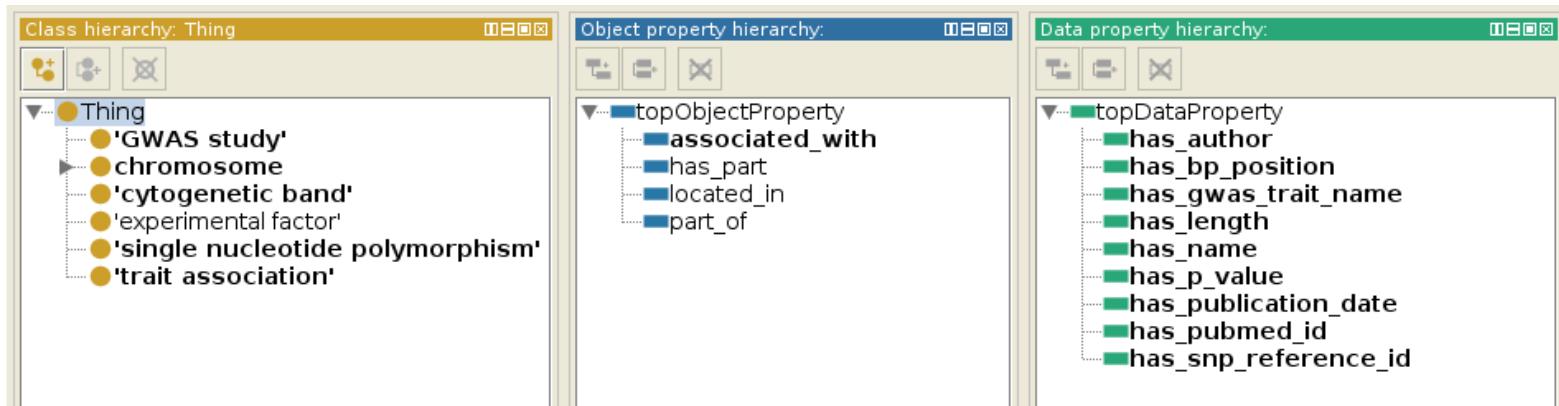
<http://www.ebi.ac.uk/efo/>

Wider GWAS Integration

- Much GWAS data are published with measurement traits with implicit links to disease
 - e.g. 90% queries from GWAS web logs are about diseases
- Connecting the cross species anatomy, phenotype, disease, measurements is essential
- Cross species phenotypic queries implemented in existing resources e.g. Ensembl become feasible
- Leveraging the existing human curated data is essential

GWAS KnowledgeBase

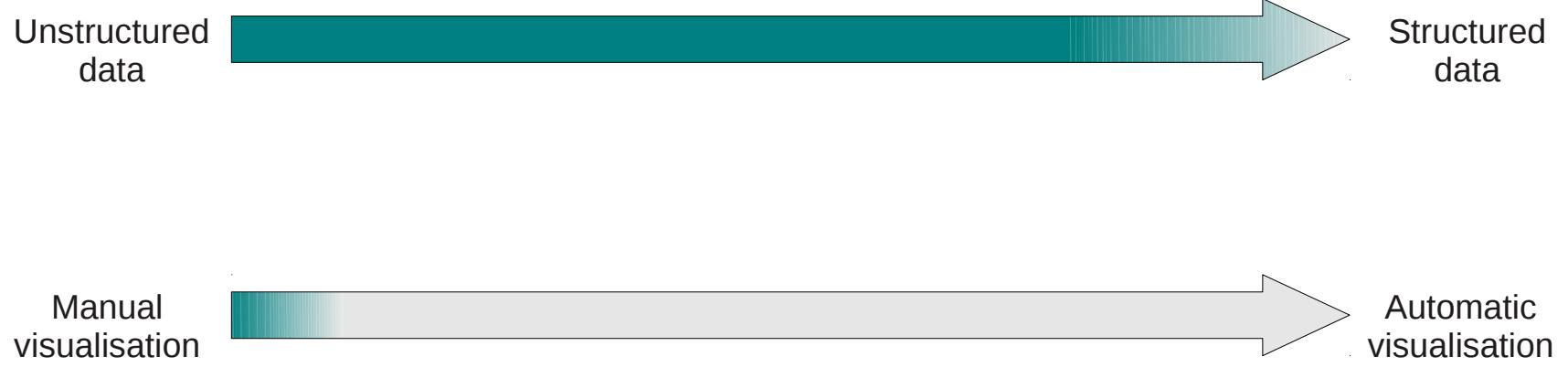
- Created a small schema ontology modelling the concepts displayed in the GWAS diagram



- Built a Java application to convert relevant data from the GWAS database into OWL individuals
- Reasoned over this ontology using Hermit

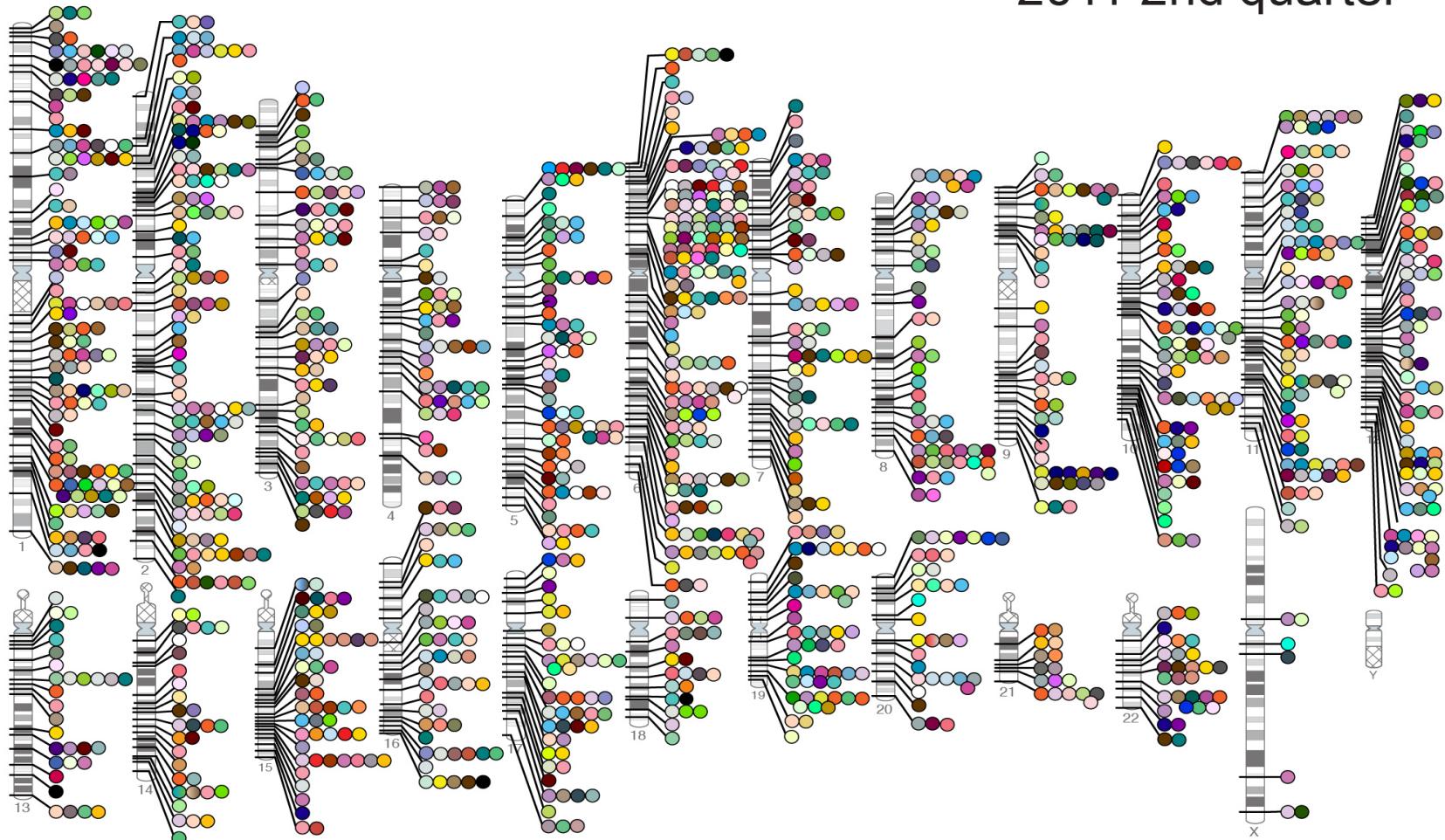
GWAS KnowledgeBase

- Contains 12,716 individuals and 43,395 axioms
- SHOIF(D) expressivity
- Reasoned in ~10hrs on a desktop machine (dual core Intel processor, 4GB memory, 2GB Java heap space) using Hermit 1.3.4
- Performs acceptably well when run from a tomcat server as part of a web application doing up-front reasoning
- We haven't needed to migrate to RDF triplestore backend (yet!)



GWAS diagram

2011 2nd quarter

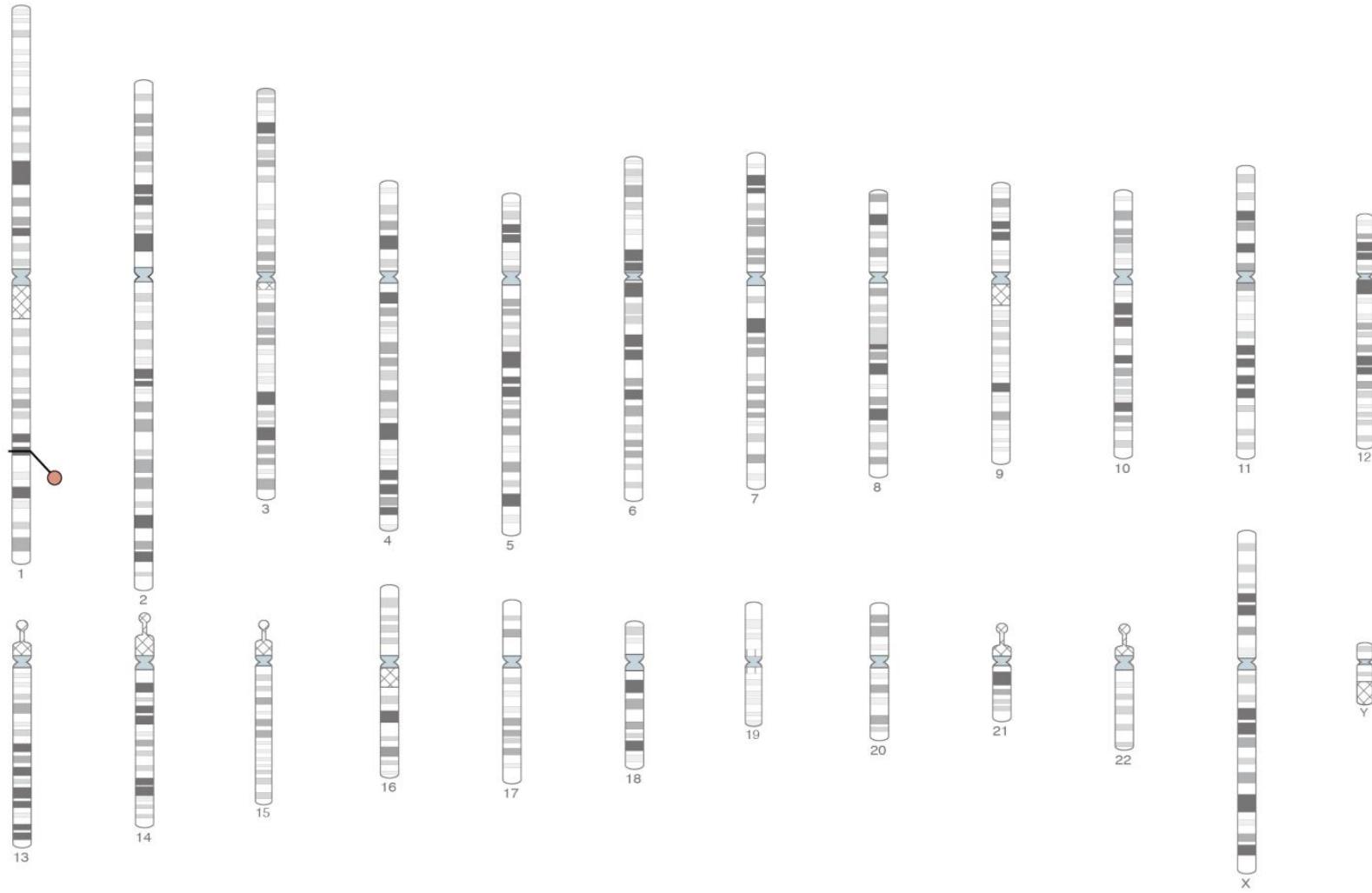


GWAS diagram

- Visualisation of all SNP-trait associations with p-value < 5×10^{-8}
- Generated quarterly by a graphic artist following extensive manual curation of the data
- Static image in PDF or Powerpoint format
- Too many traits and colours to reliably identify any individual feature
- Great way of visualising the evolution of the catalogue over time



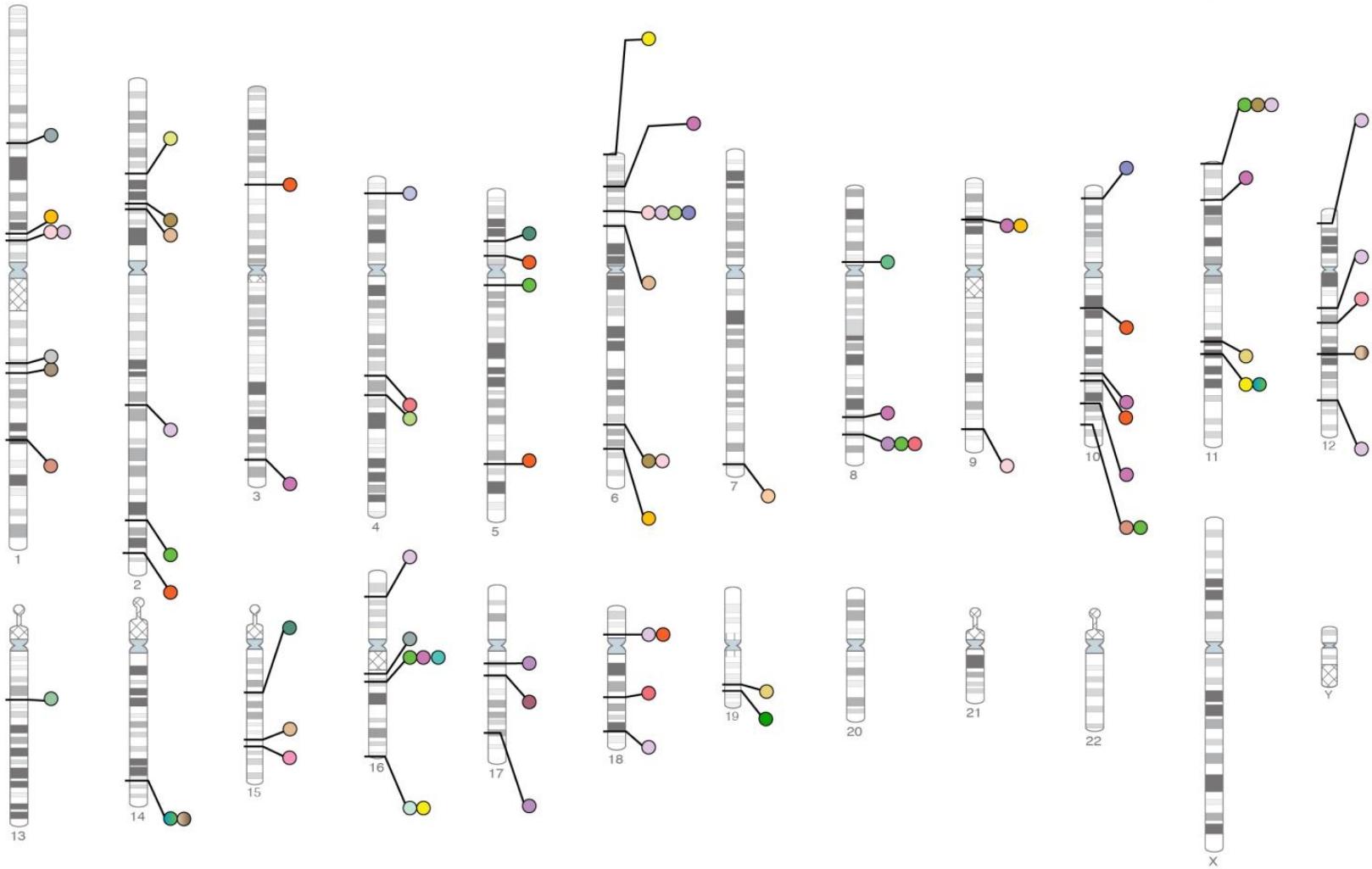
2005



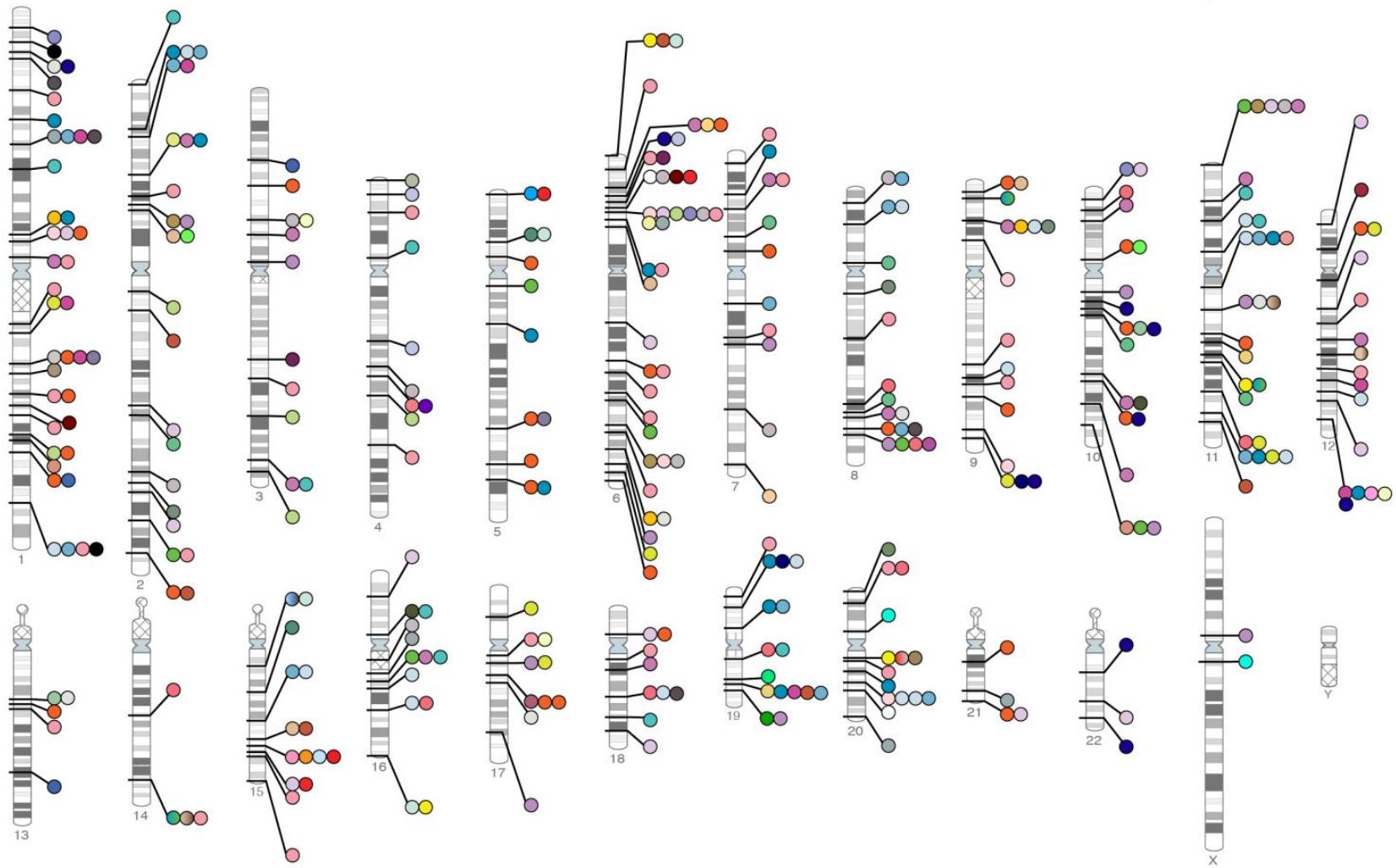
2006



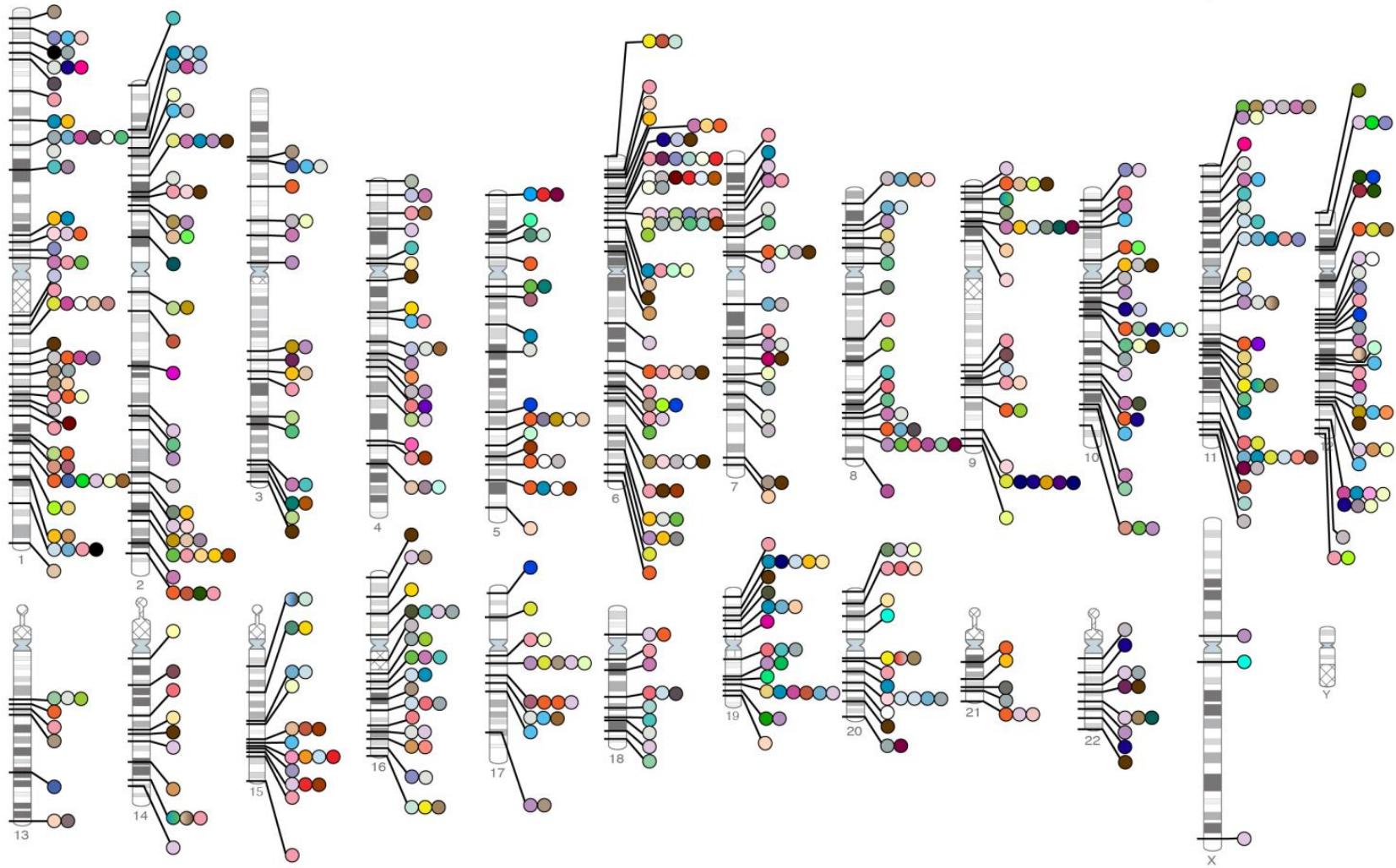
2007 4th quarter



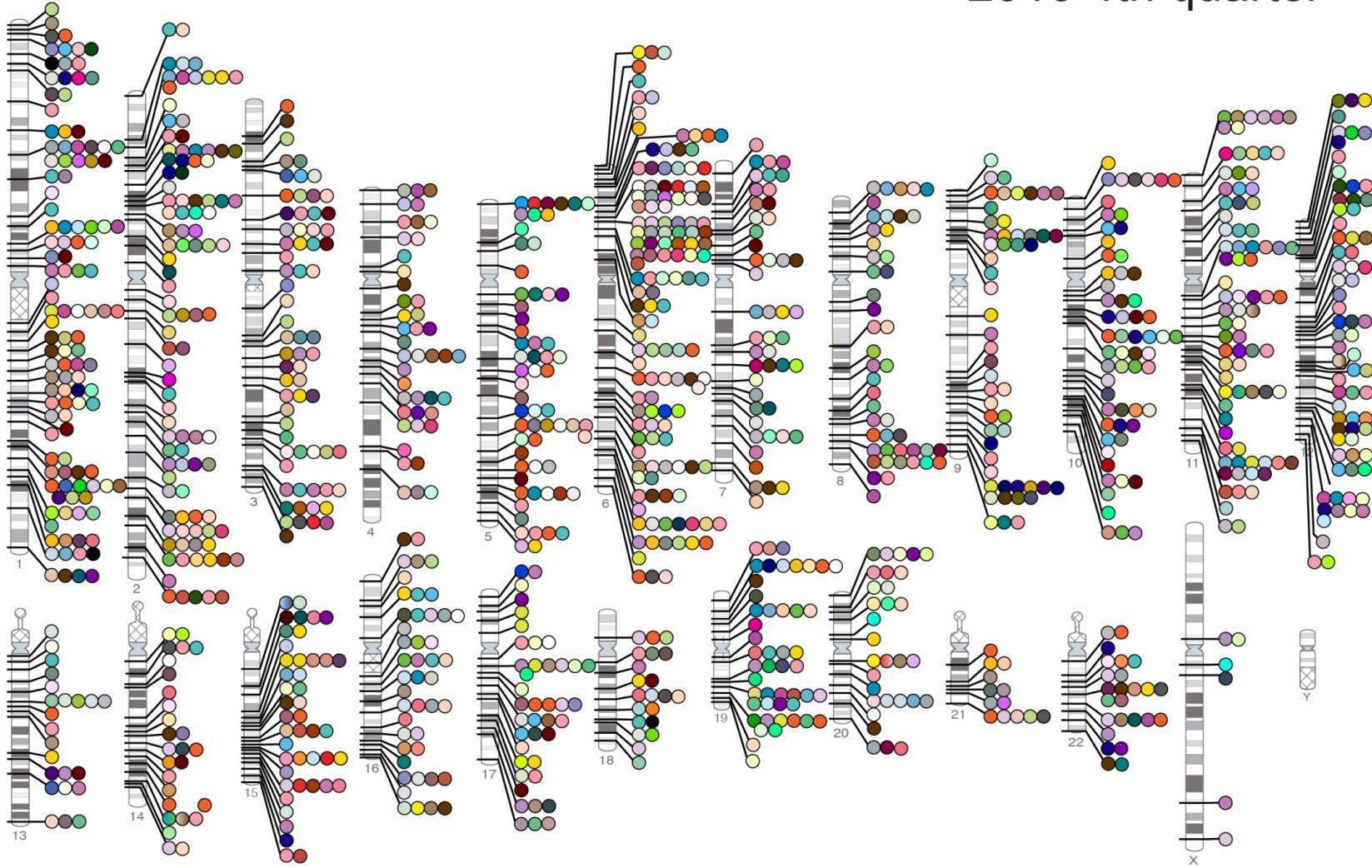
2008 4th quarter



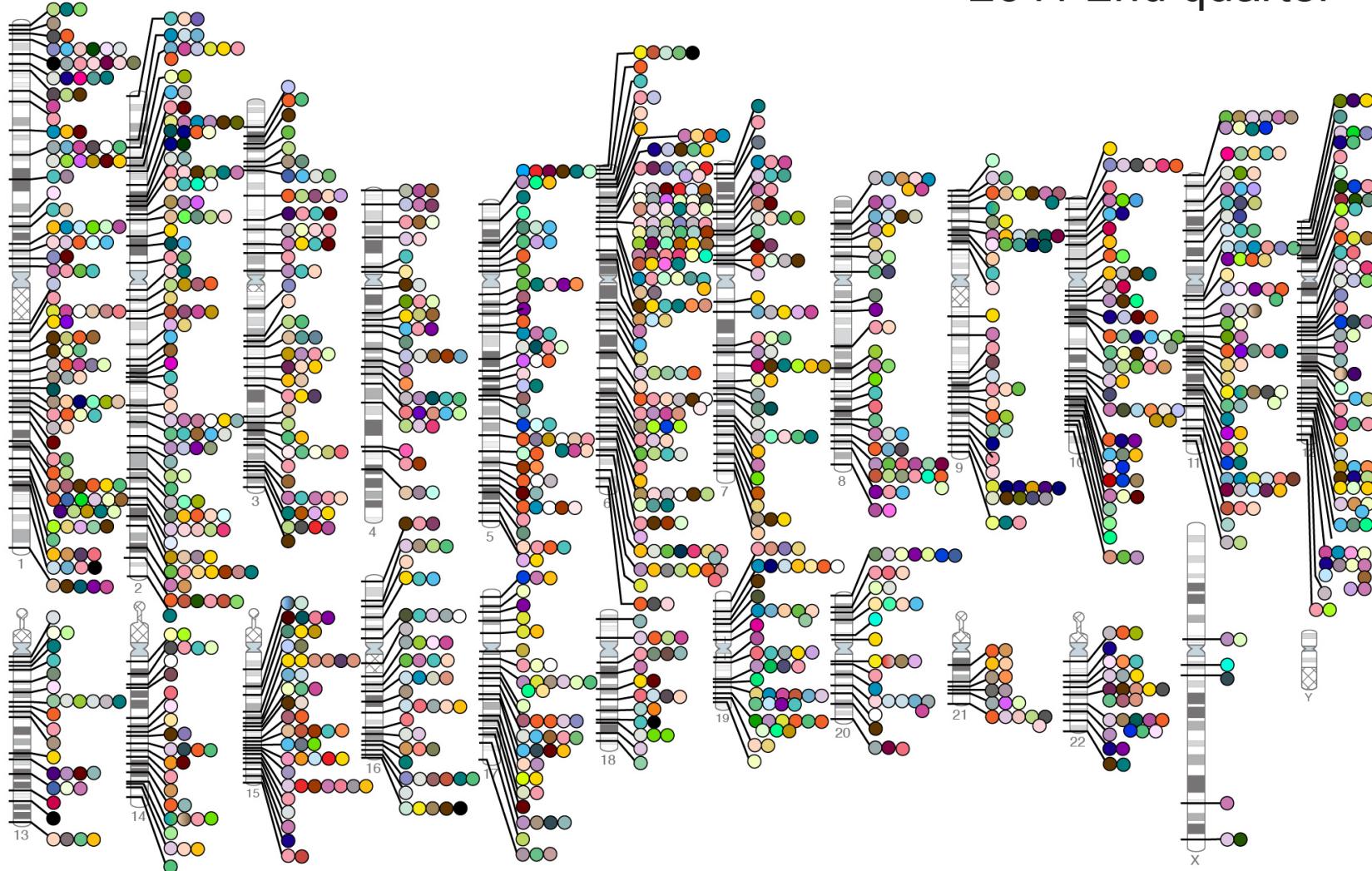
2009 4th quarter



2010 4th quarter



2011 2nd quarter

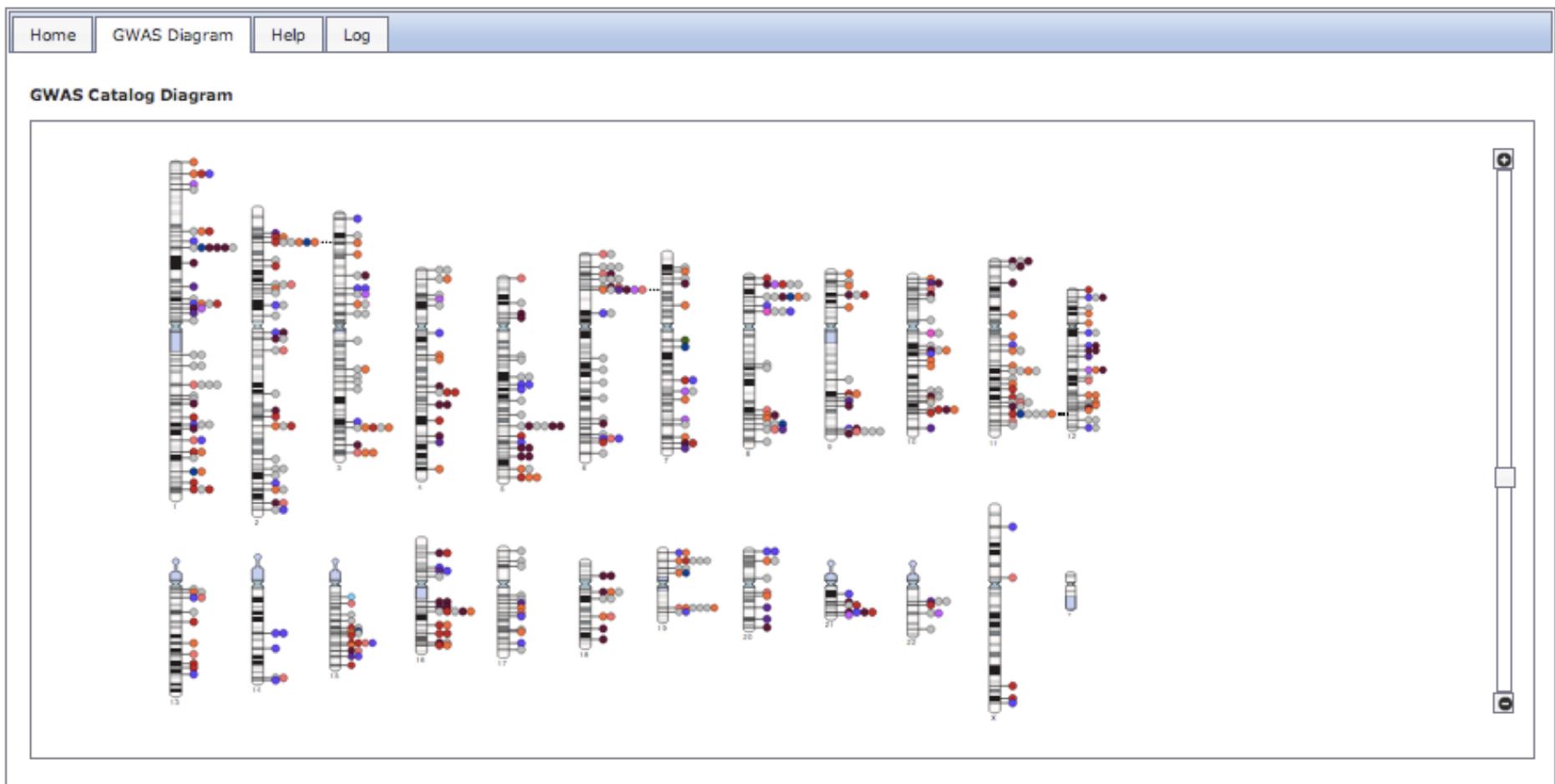




GWAS diagram automation

- Programmatic generation of the GWAS diagram from the GWAS KnowledgeBase
- We utilise “renderlets” (semantically-aware visualisation plugins) to generate components of the diagram directly from the knowledge base
- Renderlets use the strong semantics of OWL individuals and OWL classes to produce representation and layout of glyphs on-screen
 - e.g. colours of dots are based on high-level ontology classes
- SVG, complete with embedded javascript and stylesheet markup, is produced

First Beta Version



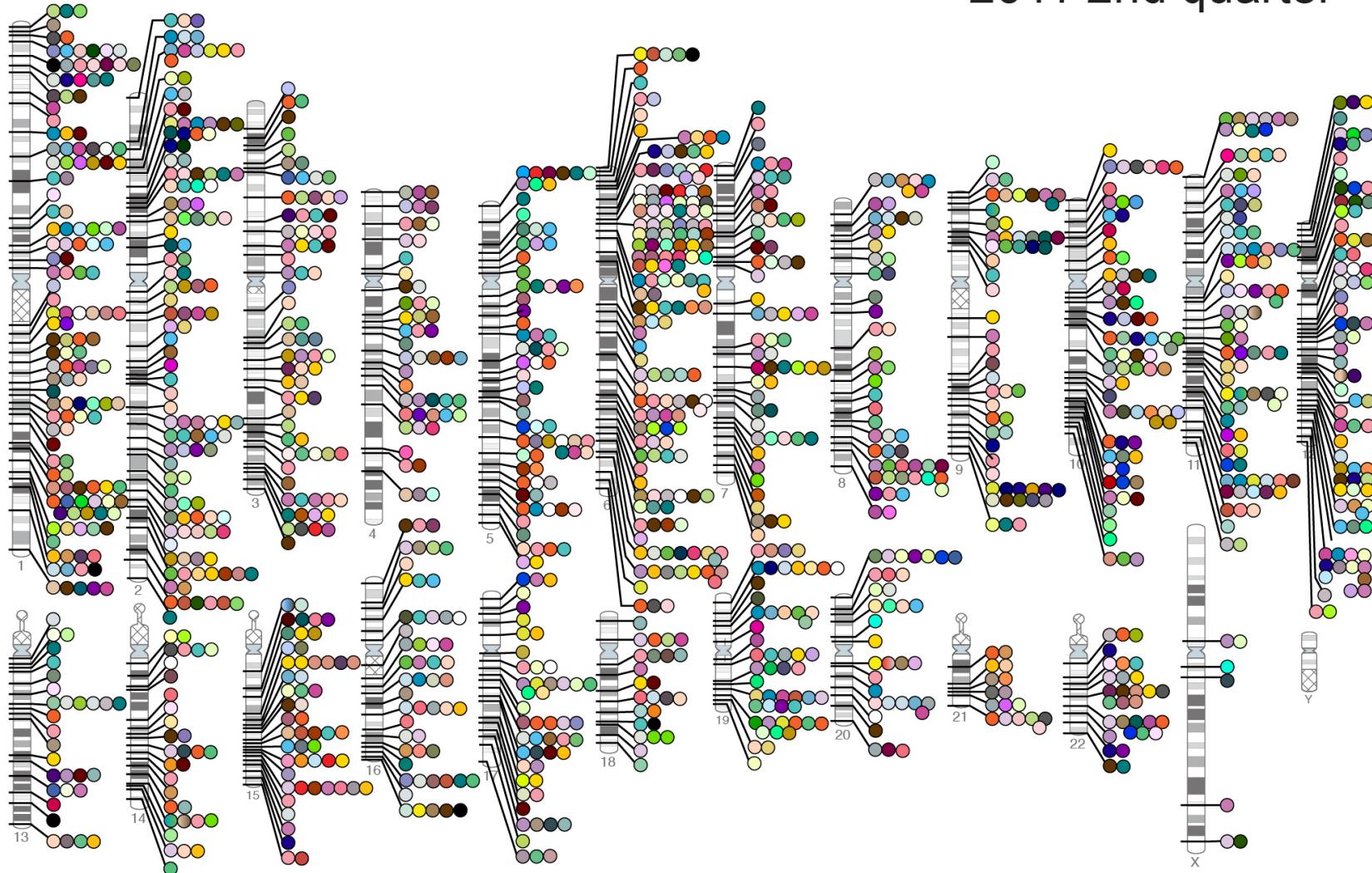
Current Live Version



wwwdev.ebi.ac.uk/fgpt/gwas

EMBL-EBI

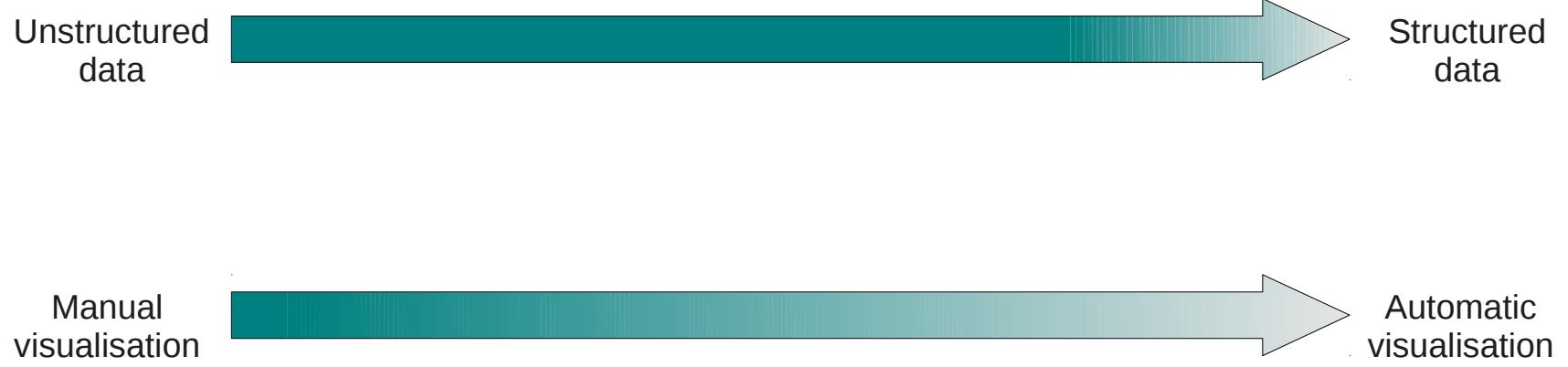
2011 2nd quarter





GWAS Diagram – New Features & Next Steps

- Dynamic, zoomable diagram with mouse-over showing traits
- New colour scheme with fewer colours representing higher-level trait categories, e.g. mental health disorders, cancers, cardio-vascular diseases
- Interactive diagram that can be filtered by a number of criteria, e.g. to show only traits associated with a given disease or from a given time period
- Interactive traits (“dots”) that link directly into the catalogue



Future work

- Short-term goals
 - Finish integration of GWAS traits into EFO
 - Produce a fully interactive diagram browser, including filtering (trait, time & p-value), by September 2012
- Long-term goals
 - Extensive, finely-grained querying
 - Explore different resolution strategies for high-density regions
 - Make GWAS diagram an entry point to the GWAS catalogue
 - Capture, model and show ethnicity information
 - Using mappings and ontology inside Ensembl

Summary

- Restructured GWAS catalogue data to allow querying beyond direct string matching
- Remapped catalogue data for easier integration with other data sources
- Removed manual processing from catalogue visualisation
- Used semantic web technologies for querying and visualisation of catalogue data, with excellent results!

Acknowledgements

- NHGRI
 - Peggy Hall
 - Lucia Hindorff
 - Heather Junkins
 - Kent Klemm
 - Darryl Leja
 - Teri Manolio
- Also thanks to:
 - Rob Davey
(TGAC)
- EBI
 - Tony Burdett
 - Jon Ison
 - Simon Jupp
 - James Malone
 - **Jackie MacArthur**
 - **Joannella Morales**
 - **Dani Welter**
 - Fiona Cunningham
 - Paul Flicek
 - Helen Parkinson

