

# NHGRI GWAS Catalog: Current uses and future directions

Lucia A. Hindorff, PhD, MPH

Division of Genomic Medicine

NHGRI, NIH

July 18, 2013



National Human Genome  
Research Institute

# Overview

- Goals and logistics of webinar
- Curation and display
- Recent enhancements
- Community adoption & uptake
- Discussion

# Goals for today

- Describe overall goals of the GWAS Catalog
- Review evolution of the GWAS Catalog: history, process, improvements
- Assess community use of Catalog and challenges in its use
- Identify possible future uses and needs
- Obtain advice on useful future directions for the Catalog

<b>Time</b>	<b>Talk</b>	<b>Speaker</b>
09:00	<i>Introduction to the GWAS Catalog</i>	<b>Lucia Hindorff, Ph.D.</b> National Human Genome Research Institute, NIH
09:40	<i>Uncovering hidden genes in intergenic GWAS regions with RNA capture-sequencing</i>	<b>Michael Clark, Ph.D.</b> University of Queensland, Australia
10:20	<i>GWAS and prior knowledge to uncover gene-gene interactions</i>	<b>Marylyn Ritchie, Ph.D.</b> Pennsylvania State University
11:00	<i>Gene regulation and common diseases and traits</i>	<b>John Stamatoyannopoulos, M.D.</b> University of Washington, Seattle
11:40	<i>Regulatory genomics and epigenomics of complex disease genetics</i>	<b>Luke Ward, Ph.D.</b> Massachusetts Institute of Technology
12:20	<i>Using human GWAS data to interrogate complex traits in an outbred mouse population</i>	<b>Karen Svenson, Ph.D.</b> The Jackson Laboratory
12:40	<i>GTEx X NHGRI catalog</i>	<b>Nancy Cox, Ph.D.</b> University of Chicago
13:20	<i>Ontologising the GWAS Catalog</i>	<b>Helen Parkinson, Ph.D.</b> EMBL-EBI, UK
13:40	<i>GWAS as a window to genetic architecture</i>	<b>Benjamin Neale, Ph.D.</b> Broad Institute
14:20	<i>Open Discussion: Feedback and Suggestions</i>	





### Webinar Dial-In Information

Go to: [GWAS Webinar Dial-In Numbers](#)

Attendees can connect via the web or by dial-in (audio only).

### About the Webinar

#### Agenda

**Speakers representing the cutting edge of research in GWAS, EWAS, regulatory genomics and data integration will present their work:**

- Mike Clark, Institute for Molecular Bioscience, University of Queensland
- [Nancy Cox, Ph.D.](#), University of Chicago
- [Lucia Hindorff, Ph.D., M.P.H.](#), National Human Genome Research Institute
- [Luke Ward, Ph.D.](#), Massachusetts Institute of Technology
- [Ben Neale, Ph.D.](#), The Broad Institute
- [Helen Parkinson, Ph.D.](#), European Bioinformatics Institute
- [Marylyn Ritchie, Ph.D.](#), Penn State University
- [John Stamatoyannopoulos, M.D.](#), University of Washington School of Medicine, Seattle
- Karen Svenson, Ph.D., The Jackson Laboratory

#### Archiving

Presentations will be archived and linked from this page when available.



Share Print

GWAS)

, 2013, the  
 cine will hold  
 s of and  
 GWAS  
 nt uses and  
 ture  
 catalog. The  
 .m. Eastern,  
 ime, 3 p.m.  
 nd 11 p.m. in

explore the  
 S catalog by  
 unities and  
 by

biomedical researchers is informing  
 projects such as ENCODE, Mouse

<http://www.genome.gov/27554296>

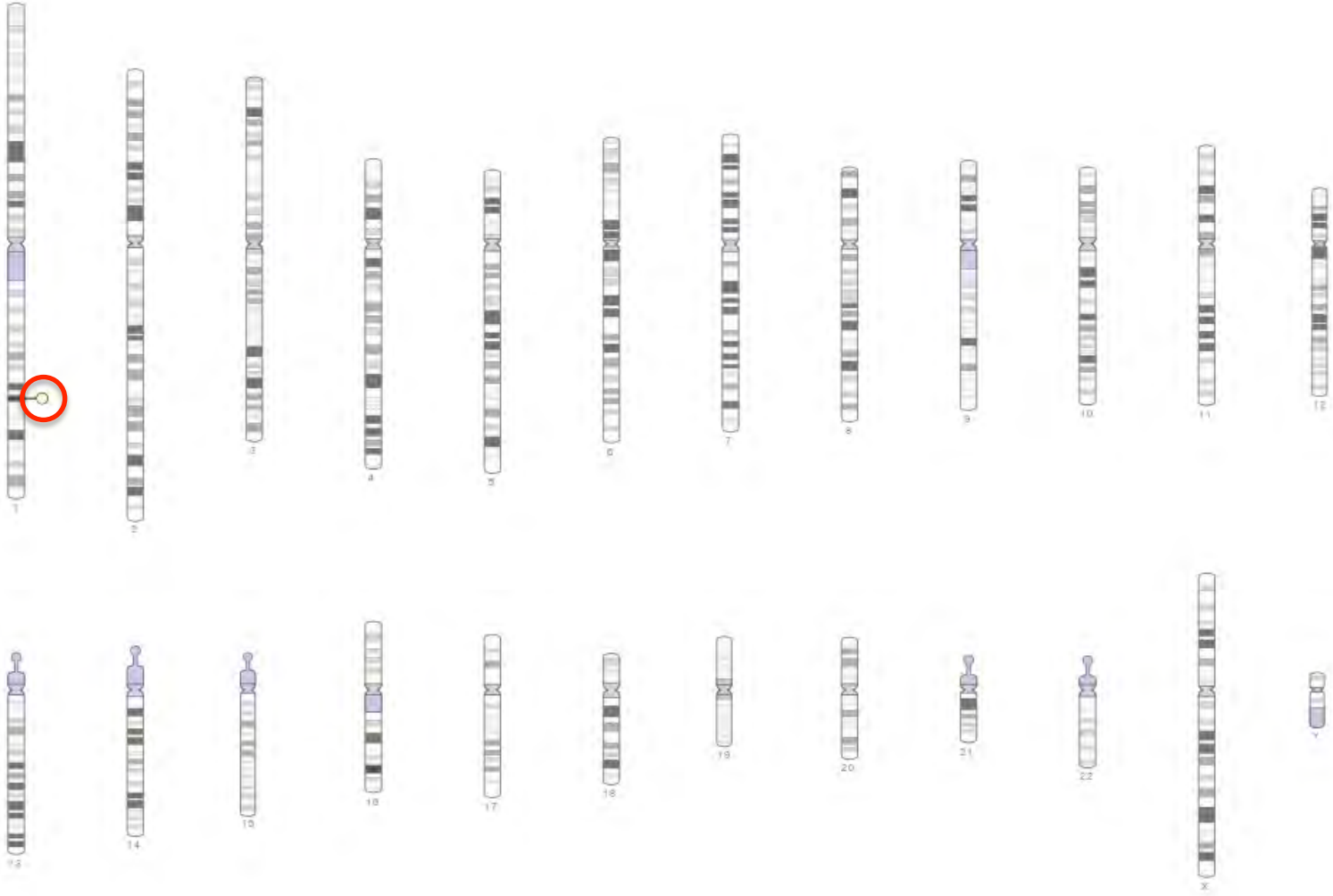
# Logistics

- Presenters joining by webinar / phone
- Adherence to agenda times
- Questions
  - Directly after each talk
  - By phone: \*1 to reach operator
  - Chat feature in WebEx
  - By email: [gwas\\_table@mail.nih.gov](mailto:gwas_table@mail.nih.gov)

# Overview

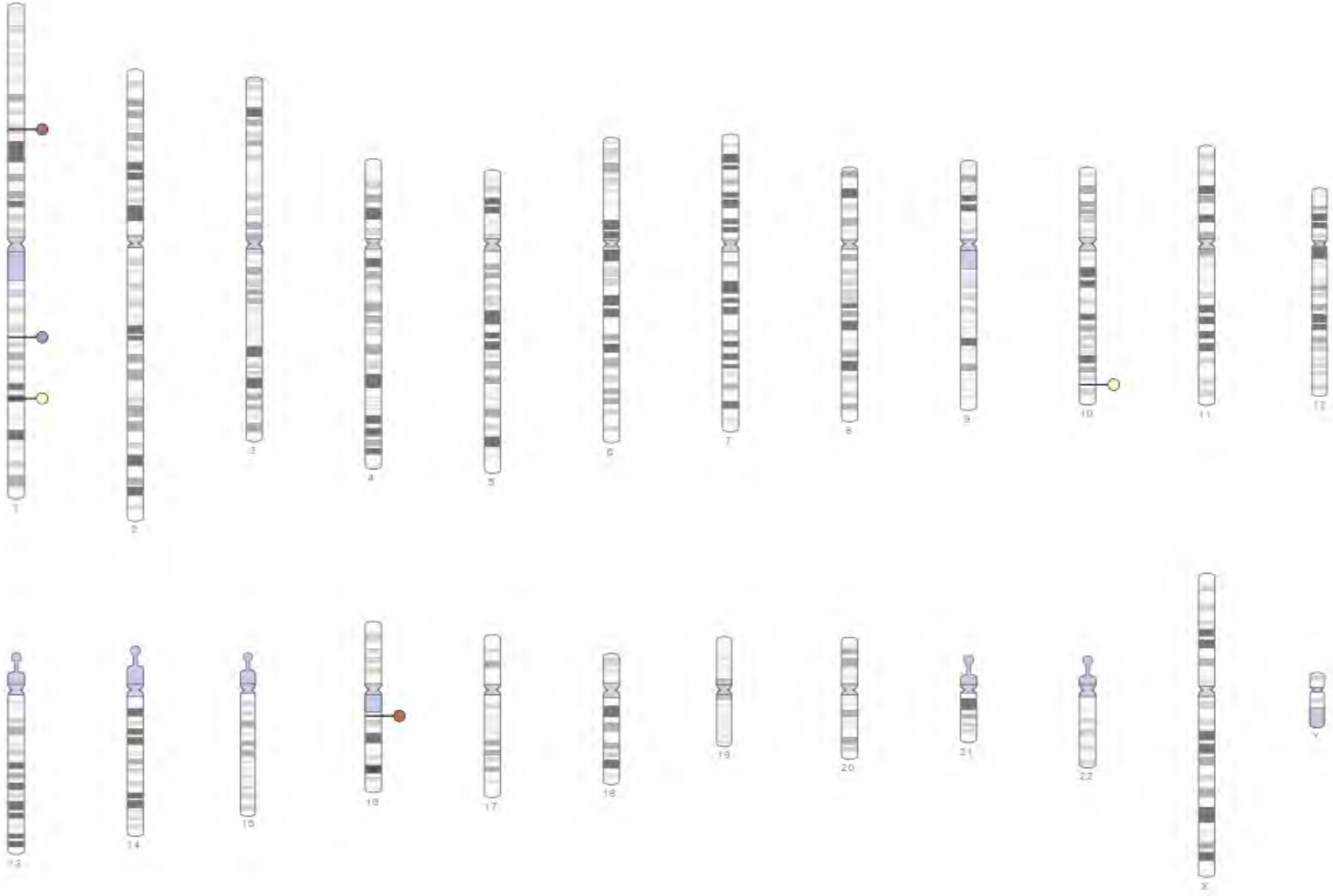
- Goals and logistics of webinar
- Curation and display
- Recent enhancements
- Community adoption & uptake
- Discussion

2005

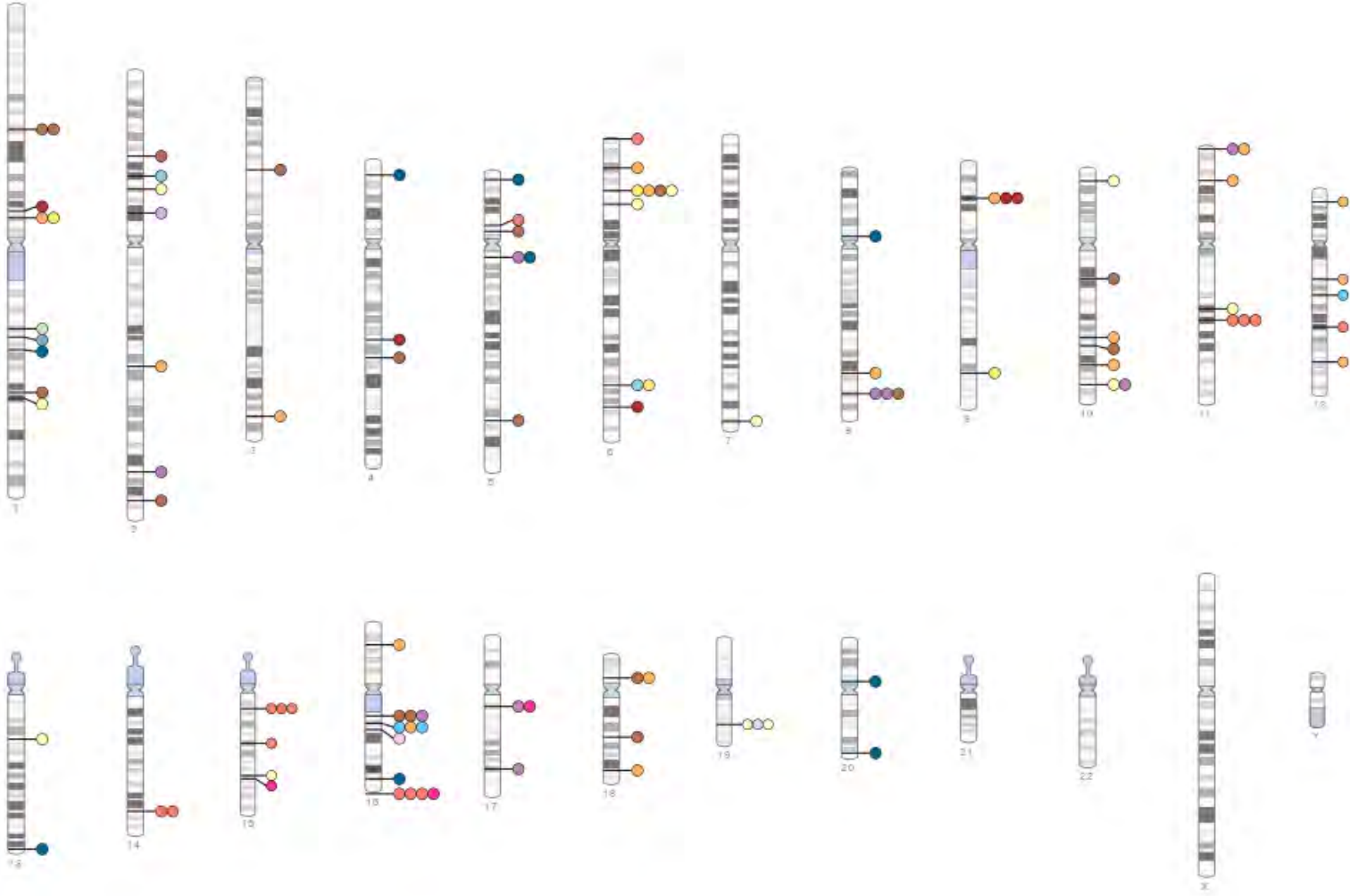




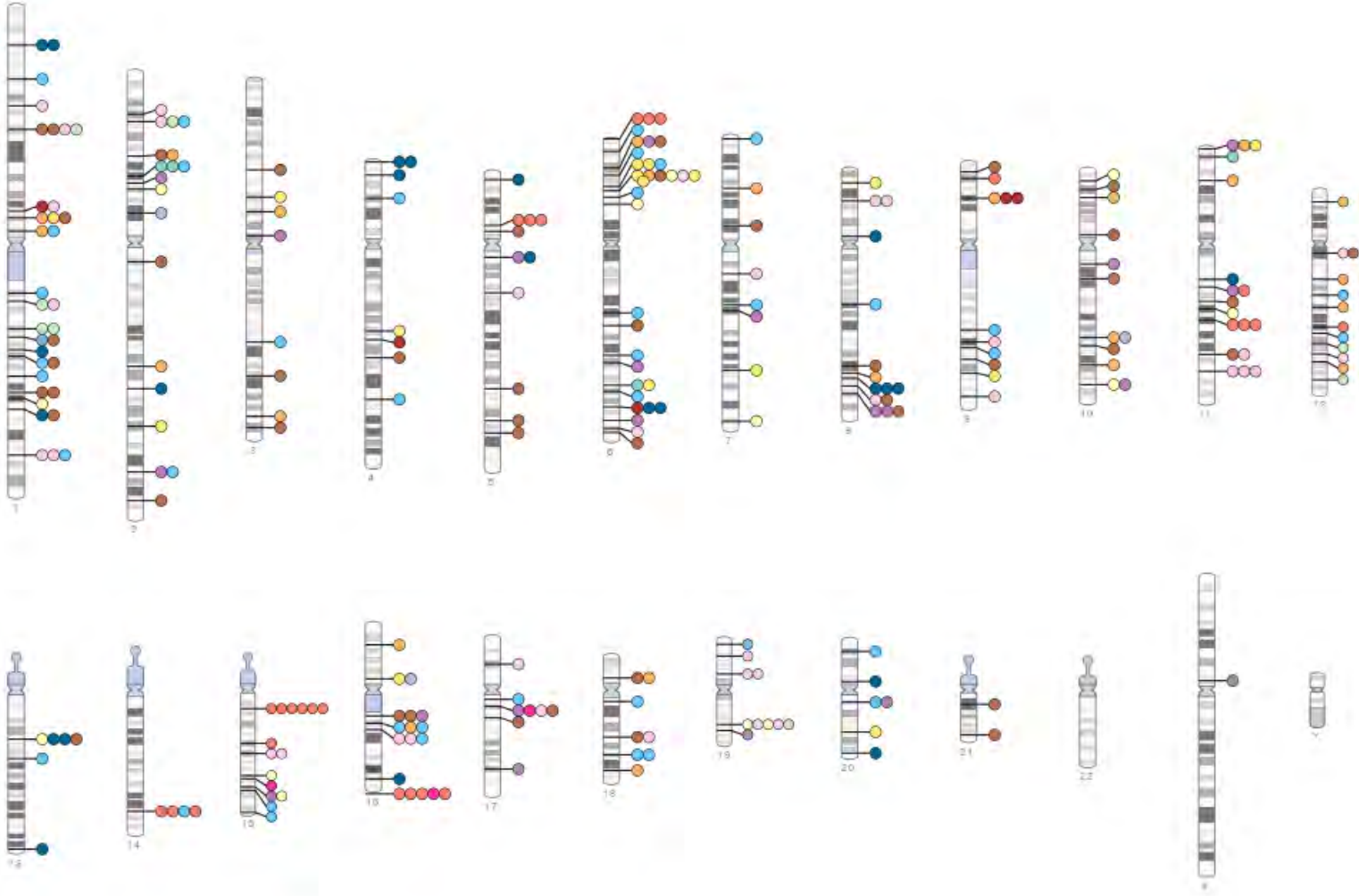
2006



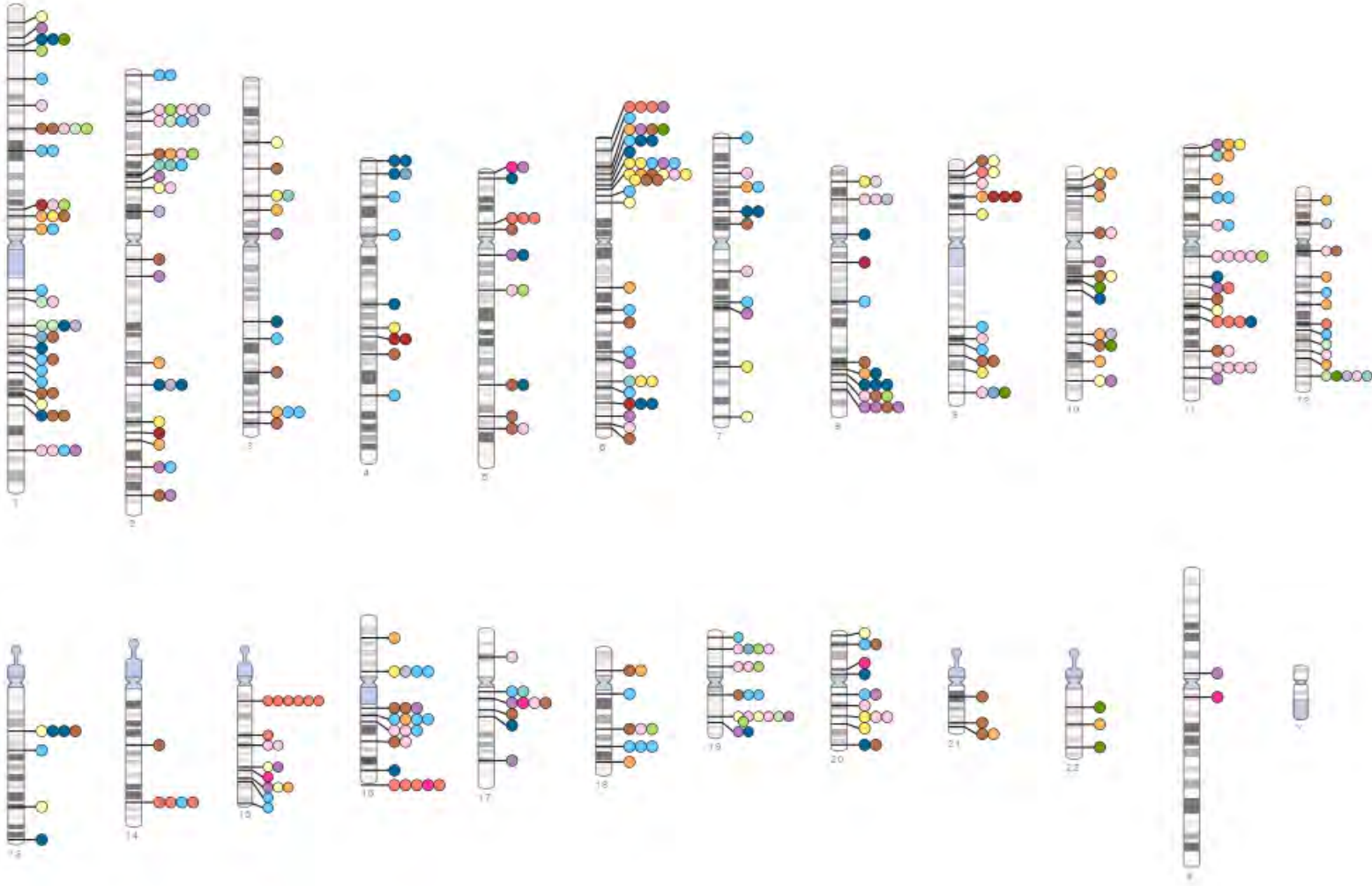
2007



# 2008 2<sup>nd</sup> quarter



2008 4<sup>th</sup> quarter





# 2009 2<sup>nd</sup> quarter



2009 4<sup>th</sup> quarter



2010 2<sup>nd</sup> quarter



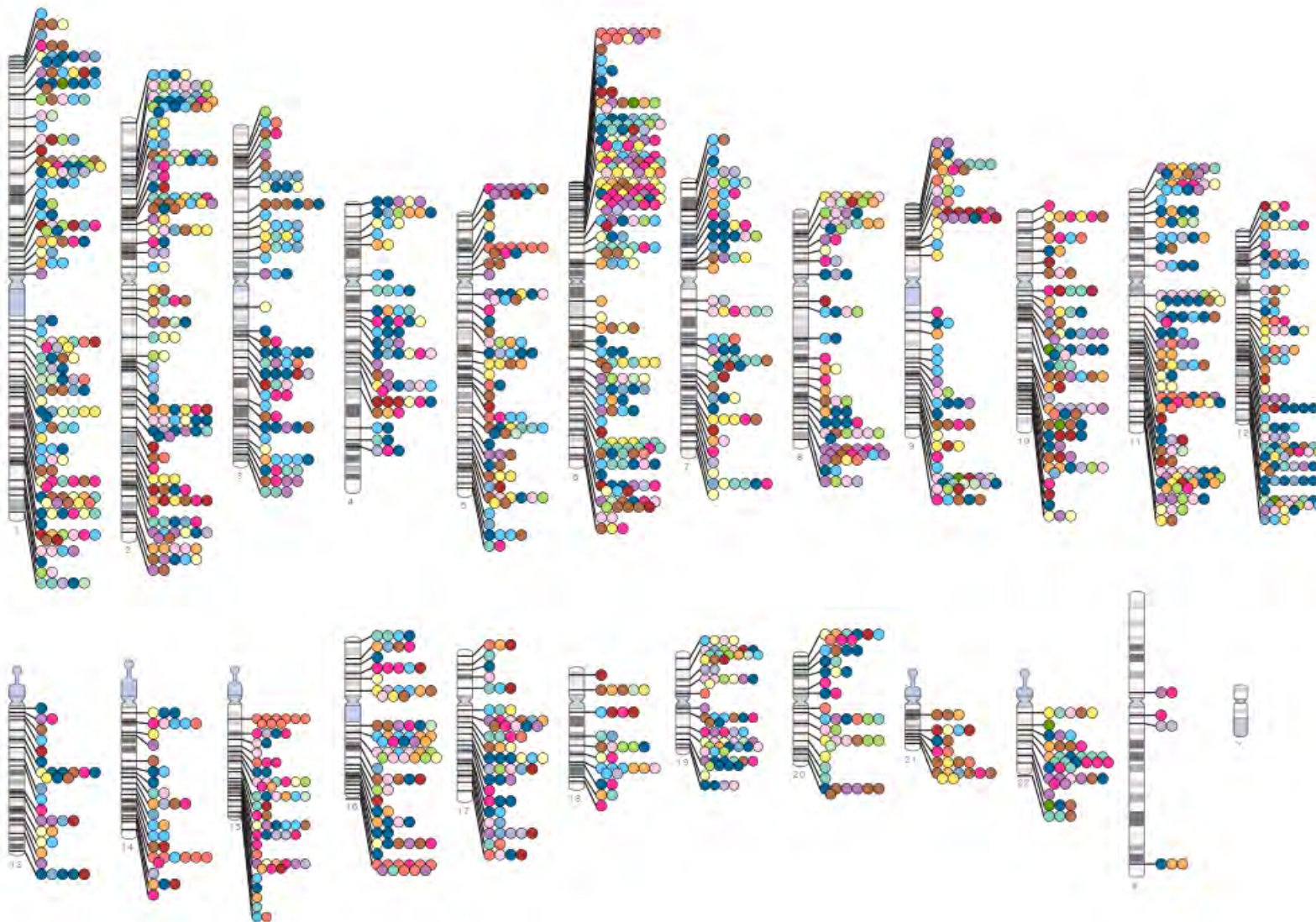


2010 4<sup>th</sup> quarter

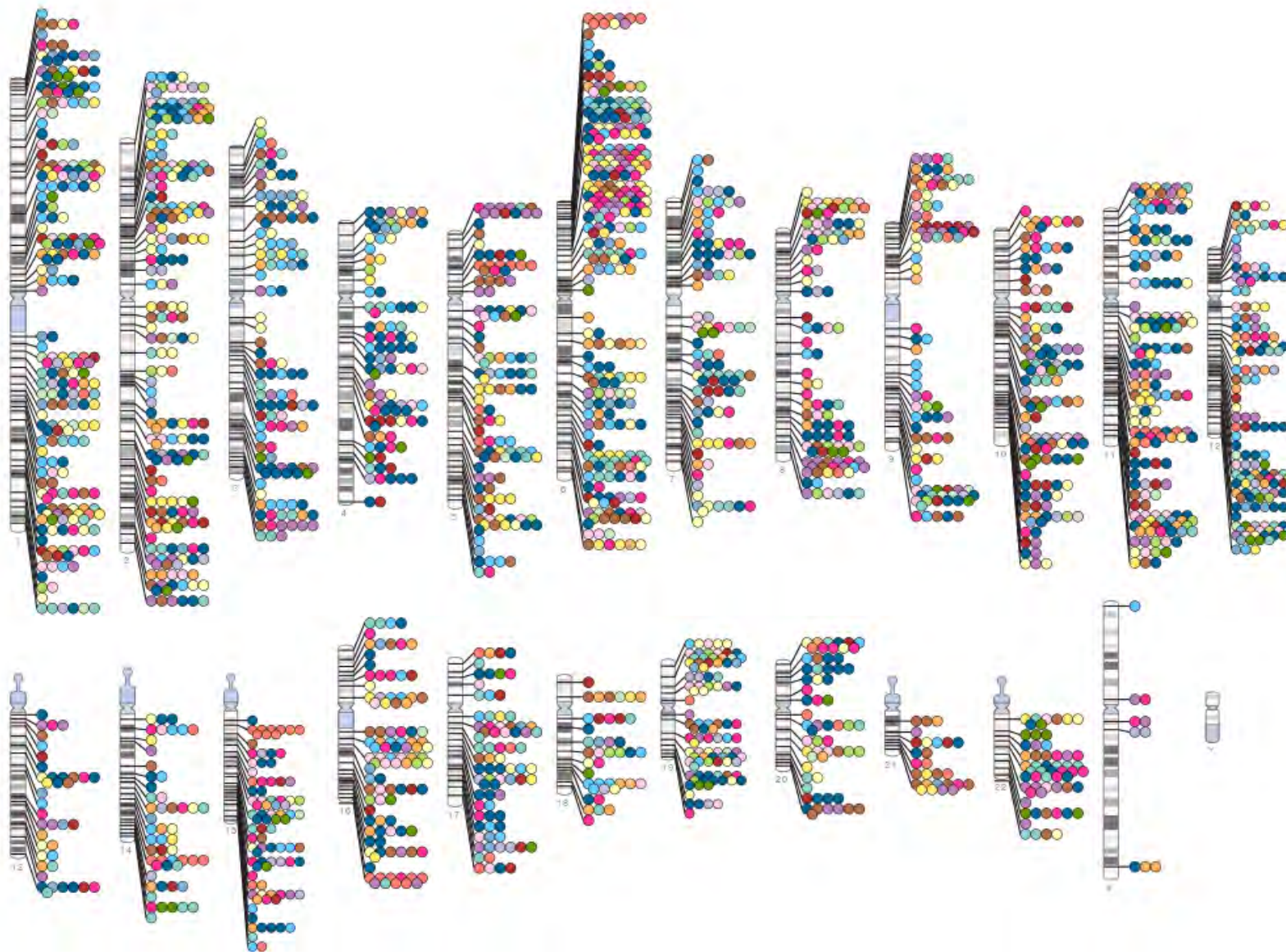




2011 2<sup>nd</sup> quarter



2011 4<sup>th</sup> quarter





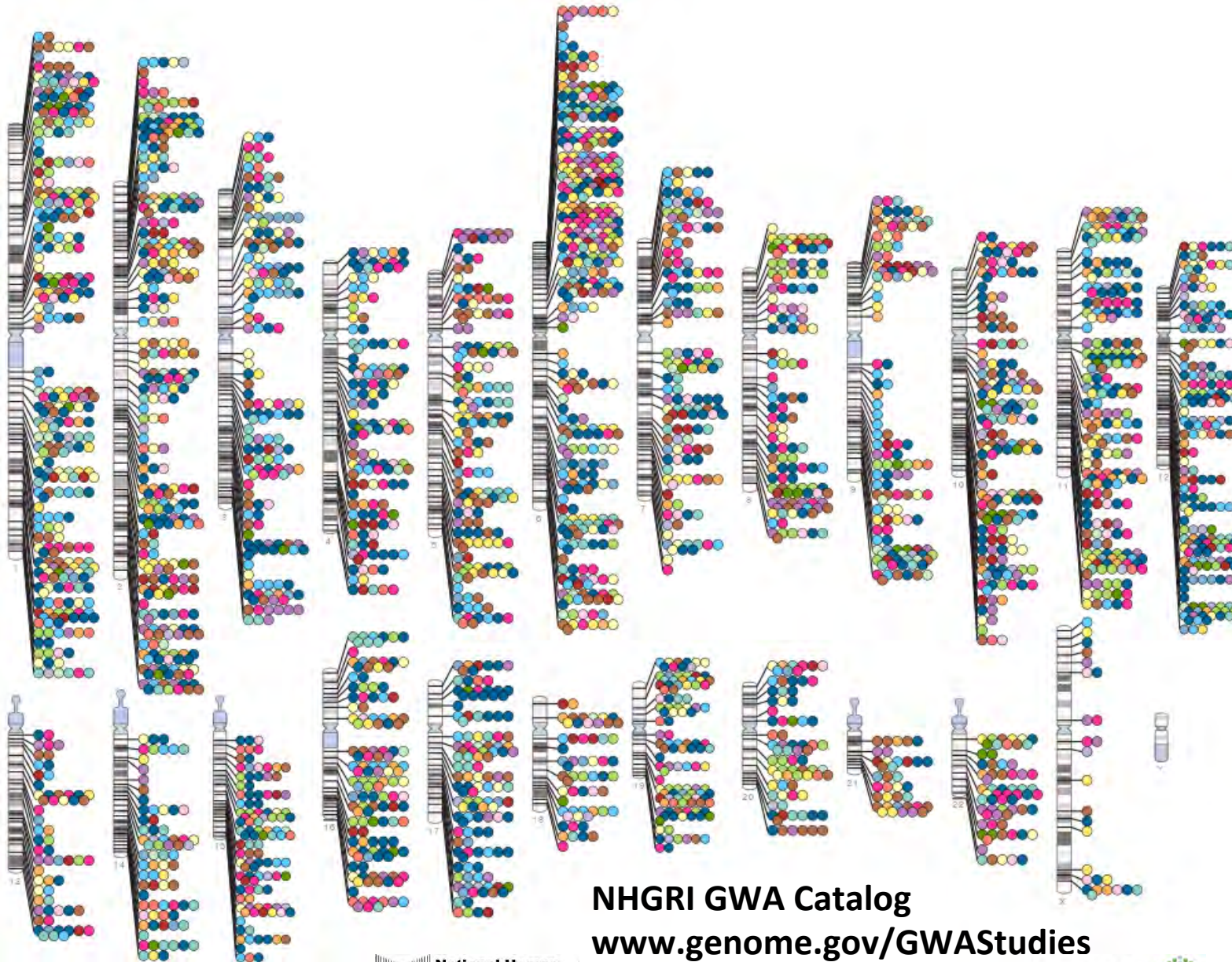
# 2012 2<sup>nd</sup> quarter





# Published Genome-Wide Associations through 12/2012

## Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories



- Digestive system disease
- Cardiovascular disease
- Metabolic disease
- Immune system disease
- Nervous system disease
- Liver enzyme measurement
- Lipid or lipoprotein measurement
- Inflammatory marker measurement
- Hematological measurement
- Body measurement
- Cardiovascular measurement
- Other measurement
- Response to drug
- Biological process
- Cancer
- Other disease
- Other trait

NHGRI GWA Catalog

[www.genome.gov/GWASudies](http://www.genome.gov/GWASudies)

[www.ebi.ac.uk/fgpt/gwas/](http://www.ebi.ac.uk/fgpt/gwas/) EMBL-EBI







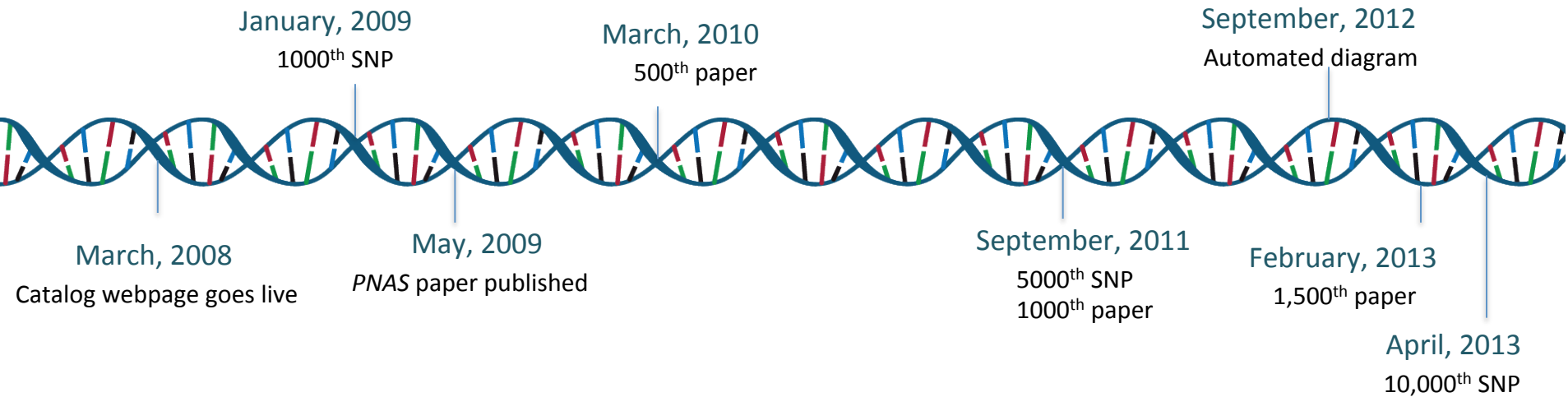
**Table 9**  
GWA studies in various traits

Disease/trait	Sample size		Region	Gene	Strongest SNP-risk allele	Risk allele frequency in controls	P	OR per copy or for heterozygote (95% CI)	Platform manufacturer and SNPs <sup>a</sup>
	Initial	Replication							
Body mass index (93)	10,657 adults	19,424 adults 10,172 children	16q12.2	<i>FTO</i>	rs9939609-A	0.39	$2 \times 10^{-20}$	0.36 (NR) <sup>a</sup> -0.4 (NR) <sup>c</sup>	Affymetrix: 490,032
Height (132)	4,921 studied	29,098 studied <sup>d</sup>	12q14.3	<i>HMG2</i>	rs1042725-C	0.51	$6 \times 10^{-16}$	0.4 (NR) <sup>b</sup>	Affymetrix: 364,301
Height (133)	6,669 studied	28,801 studied	20q11.22	<i>BFZB</i>	rs6060369-C	0.44	$2 \times 10^{-16}$	0.44 (NR) <sup>b</sup>	Illumina and Affymetrix <sup>d</sup>
Skin pigmentation by reflectance spectroscopy (134)	363 maxL* <56 <sup>ii</sup> 374 maxL* >63 <sup>ii</sup>	116 low maxL* 115 high maxL*	15q21.1	<i>SLC24A5</i>	rs1834640-G	0.30	$1 \times 10^{-30}$	12.5 (8.33-20.0)	Perlegen: 1,502,205 <sup>i</sup>
			11q14.3	<i>TYR</i>	rs1042602-C	0.90	$4 \times 10^{-10}$	4.36 (2.64-7.20)	
			5p13.3	<i>SLC45A2</i>	rs16891982-C	0.97	$3 \times 10^{-11}$	4.86 (2.88-8.21)	
Freckles (135)	2,986 studied	3,932 studied	6p25.3	between <i>SEC5L1</i> and <i>IRF4</i>	rs1540711-A	0.50	$2 \times 10^{-9}$	1.40 (1.26-1.57)	Illumina: 317,511
Blond vs. brown hair (135)	2,986 studied	3,932 studied	12q21.33	<i>KITLG</i>	rs12821256-C	0.15	$2 \times 10^{-11}$	2.32 (1.86-2.92)	Illumina: 317,511
			14q32.12	<i>SLC24A4</i>	rs4904868-C+ rs2402130-A	0.60	$9 \times 10^{-24}$	2.56 (2.12-3.09)	
Blue vs. green eyes (135)	2,986 studied	3,932 studied	14q32.12	<i>SLC24A4</i>	rs4904868-C+ rs2402130-A	0.60	$2 \times 10^{-18}$	2.06 (1.76-2.42)	Illumina: 317,511
F cell distribution (136)	179 adults <sup>j</sup>	90 adults	2p16.1	<i>BCL11A</i>	rs1427407-?	0.14	$5 \times 10^{-21}$	13.1% (NR) <sup>k</sup>	Illumina: 308,015
			6q23.3	Intergenic	rs9399137-?	0.23	$3 \times 10^{-28}$	15.8% (NR) <sup>k</sup>	
			11p15.5	<i>XmnI</i> - $\beta\gamma$	NR	0.33	$2 \times 10^{-28}$	10.2% (NR) <sup>k</sup>	
Serum uric acid levels (137)	4,305 Sardinian	1,301 Tuscan	4p16.1	<i>GLUT9</i>	rs6855911-A	0.74	$2 \times 10^{-18}$	0.32 (NR) <sup>b</sup>	Affymetrix: 362,129
Serum urate (117)	1,955 hypertensive individ	2,033 individ in 519 families; 1,461 twins <sup>ll</sup>	4p16.1	<i>SLC2A9</i>	rs7442295-A	0.79	$2 \times 10^{-10}$	0.024 (0.018-0.030) <sup>k</sup>	Affymetrix: 400,496
Recombination	1,887 men	1,248 men	4p16.3	<i>RNF212</i>	rs3796619-T	0.33 <sup>p</sup>	$3 \times 10^{-24}$	70.7 (84.3-57.1) <sup>o</sup>	Illumina:





patterns genetic the ini- common gesting ntifying


# GWAS Catalog Timeline 2008 - present



## A Catalog of Published Genome-Wide Association Studies

[Division Staff](#) : [Funding Opportunities](#) : [Genomic Medicine Activities](#) : [GWAS Catalog](#) : [Meetings & Workshops](#) : [Potential Sample Collections for Sequencing](#) : [Programs](#) : [Publications](#) : [Trans-NIH Sequencing Inventory](#)

Additional information has been added to the HTML catalog columns below. For a description of column headings for the HTML catalog, go to: [Catalog Heading Descriptions](#)  

[Potential etiologic and functional implications of genome-wide association loci for human diseases and traits](#) 

Click here to read our recent *Proceedings of the Academy of Sciences (PNAS)* article on catalog methods and analysis.


[View the Interactive Diagram](#)  [View the Full Catalog](#) [Download the Catalog](#) [Search the Catalog](#)



*Published Genome-Wide Associations*

*Credit: Darryl Leja and Teri Manolio, NHGRI; Tony Burdett, Dani Welter, and Helen Parkinson, EBI*

[View as PDF](#) 

[View PowerPoint slide](#) 

The genome-wide association study (GWAS) publications listed here include only those attempting to assay at least 100,000 single nucleotide polymorphisms (SNPs) in the initial stage. Publications are organized from most to least recent date of publication, indexing from online publication if available. Studies focusing only on candidate genes are excluded from this catalog. Studies are identified through weekly PubMed literature searches, daily NIH-distributed compilations of news and media reports, and occasional comparisons with an existing database of GWAS literature ([HuGE Navigator](#)).

SNP-trait associations listed here are limited to those with p-values  $< 1.0 \times 10^{-5}$  (see full methods for additional details). Multipliers of powers of 10 in p-values are rounded to the nearest single digit; odds ratios and allele frequencies are rounded to two decimals. Standard errors are converted to 95 percent confidence intervals where applicable. Allele frequencies, p-values, and odds ratios derived from the largest sample size, typically a combined analysis (initial plus replication studies), are recorded below if reported; otherwise statistics from the initial study sample are recorded. For quantitative traits, information on % variance explained, SD increment, or unit difference is reported where available. Odds ratios  $< 1$  in the original paper are converted to  $OR > 1$  for the alternate



## Search By:

**Journal:**

**First Author:**  
(last name)

**Disease/Trait:**  
(string search)

Tip: Expand your search by using the OR operator (returns results with either term),  
or narrow your search using the AND operator (returns results with both terms).

**or**

Tip: Hold Ctrl-key to select multiple entries.

**Chromosomal Region:**  
(e.g., "13q21.31")

**Gene:**  
(e.g., "LRP5")

**SNP:**  
(e.g., "rs20755555")

*The SNP data in the catalog has been mapped to dbSNP Build 137 and Genome Assembly, GRCh37/hg19.*

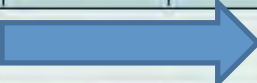
**OR greater than:**

**p-Value threshold:**  
Enter the exponent. For example,

enter "5" for  $p < 10^{-5}$



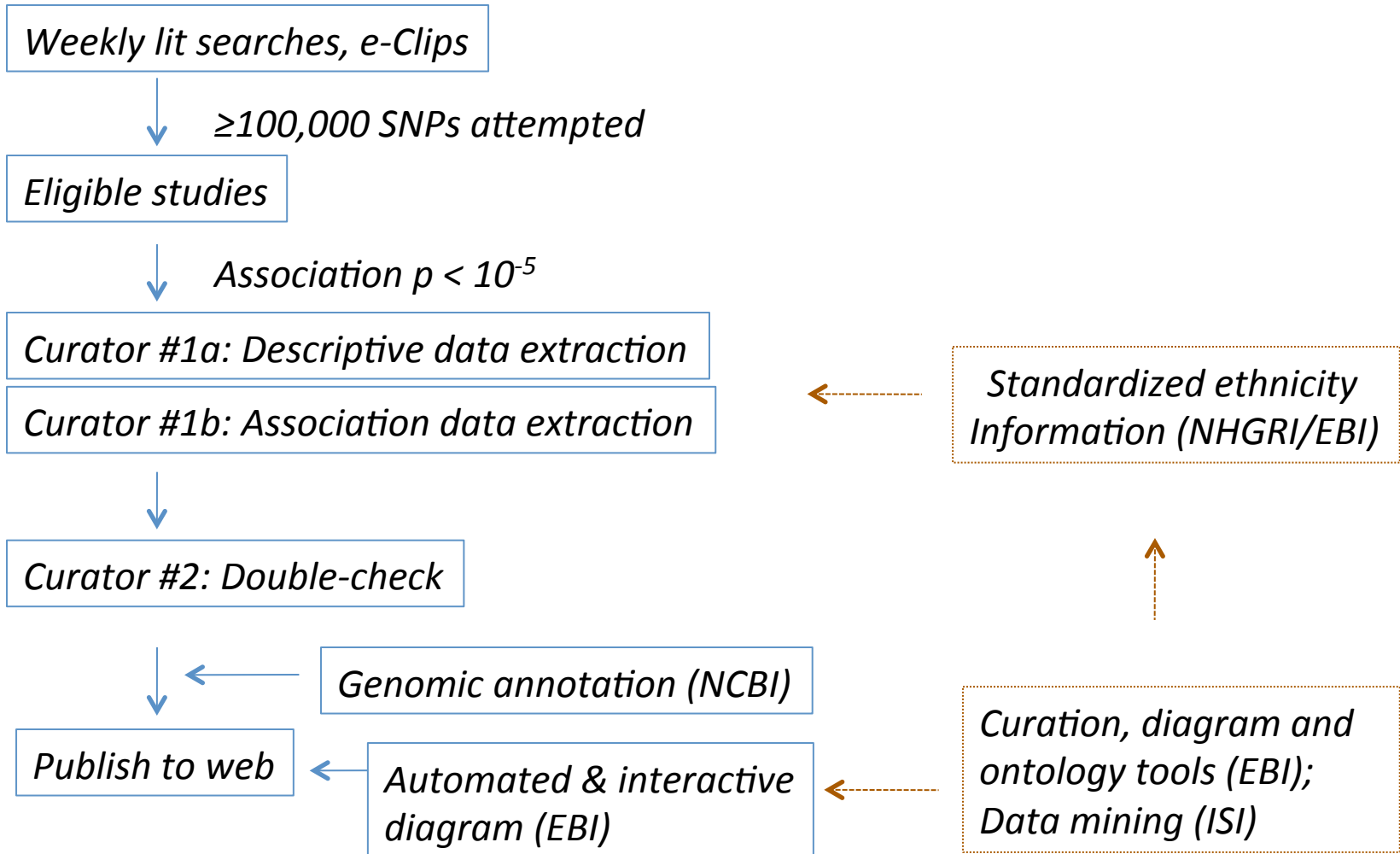
Date Added to Catalog (since 11/25/08)	First Author/Date/Journal/Study	Disease/Trait	Initial Sample Size	Replication Sample Size
--	---------------------------------	---------------	---------------------	-------------------------

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
 <a href="#">View full set of 17 SNPs</a>								Illumina [1,232,008] (imputed)	N
10q25.2	<i>TCF7L2</i>	<a href="#">TCF7L2</a>	<a href="#">rs7903146-T</a>	intron	0.5	$9 \times 10^{-75}$ (South Asian, East Asian, Europeans)	1.19 [1.17 - 1.21]		
10q25.2	<i>TCF7L2</i>	<a href="#">TCF7L2</a>	<a href="#">rs7903146-T</a>	intron	0.3	$2 \times 10^{-38}$ (South Asians, East Asians)	1.15 [1.12 - 1.17]		
10q25.2	<i>TCF7L2</i>	<a href="#">TCF7L2</a>	<a href="#">rs7903146-T</a>	intron	0.3	$3 \times 10^{-35}$ (South Asians)	1.15 [1.13 - 1.18]		
10q25.2	<i>TCF7L2</i>	<a href="#">TCF7L2</a>	<a href="#">rs7903146-T</a>	intron	0.31	$6 \times 10^{-22}$ (All Punjabi)	1.3 [1.23 - 1.37]		
3q27.2	<i>IGF2BP2</i>	<a href="#">IGF2BP2</a>	<a href="#">rs1470579-C</a>	intron	0.5	$2 \times 10^{-19}$ (South Asian, East Asian, Europeans)	1.08 [1.05-1.09]		
10q25.2	<i>TCF7L2</i>	<a href="#">TCF7L2</a>	<a href="#">rs7903146-T</a>	intron	0.31	$3 \times 10^{-19}$ (Punjabi Sikhs)	1.44 [1.33 - 1.56]		
3q27.2	<i>IGF2BP2</i>	<a href="#">IGF2BP2</a>	<a href="#">rs1470579-C</a>	intron	0.45	$2 \times 10^{-13}$ (South Asians, East Asians)	1.06 [1.04-1.08]		
3q27.2	<i>IGF2BP2</i>	<a href="#">IGF2BP2</a>	<a href="#">rs1470579-C</a>	intron	0.45	$4 \times 10^{-9}$ (South Asians)	1.06 [1.04-1.09]		
13q12.12	<i>SGCG, SACS</i>	<a href="#">SGCG</a>	<a href="#">rs9552911-G</a>	intron	0.93	$2 \times 10^{-8}$ (Punjabi Sikhs)	1.49 [1.3-1.72]		
3q27.2	<i>IGF2BP2</i>	<a href="#">IGF2BP2</a>	<a href="#">rs1470579-C</a>	intron	0.41	$4 \times 10^{-7}$ (All Punjabi)	.88 [0.83 - 0.92]		

# GWAS catalog workflow

## Current process

## Ongoing/future improvements



# Data entry

## Add a GWA Study

### Add a GWA Study SNP

Instructions:

- To go back to the Input for GWAS Studies Table without changing or adding data, click on "View GWA Studies".
- To go back to the Edit GWA Studies Table without changing or adding data, click on "Edit GWA Studies".
- Once you have entered all the data for the SNP, click the "Submit" button at the bottom of the form. You will be taken back to the Edit GWA Studies page.

[View GWA Studies](#)

[Edit GWA Studies](#)

Region:

Gene:

Strongest SNP-Risk Allele:

SNP:

Risk Allele Frequency in Controls:

P-value:

0 x 10 0 Enter mantissa and exponent (e.g.; for "7 x 10<sup>-13</sup>", enter "7" and "-13")

P-value (Text):

OR/HR/RR per copy (Num):

(enter reciprocal:

OR/HR/RR-type?:

OR/HR/RR per copy (Range):

OR/HR/RR per copy (Unit Description):



# Overview

- Goals and logistics of webinar
- Curation and display
- Recent enhancements
- Community adoption & uptake
- Discussion

# GWAS diagram improvements

- Automated generation of diagram
- Interactive display
- Consolidation of traits into higher level categories = fewer colors
- Dynamic time-series display
- Interactive filtering on trait (selected browsers)

# GWAS Diagram Browser

Exploring Genome-wide Association Studies

Query by trait

Clear

To show only one trait, e.g. "breast cancer" or "schizophrenia", type

the trait into the box on the left and hit "Query by trait"

Interactive GWAS Diagram

Trait-specific Views

Time Series Views

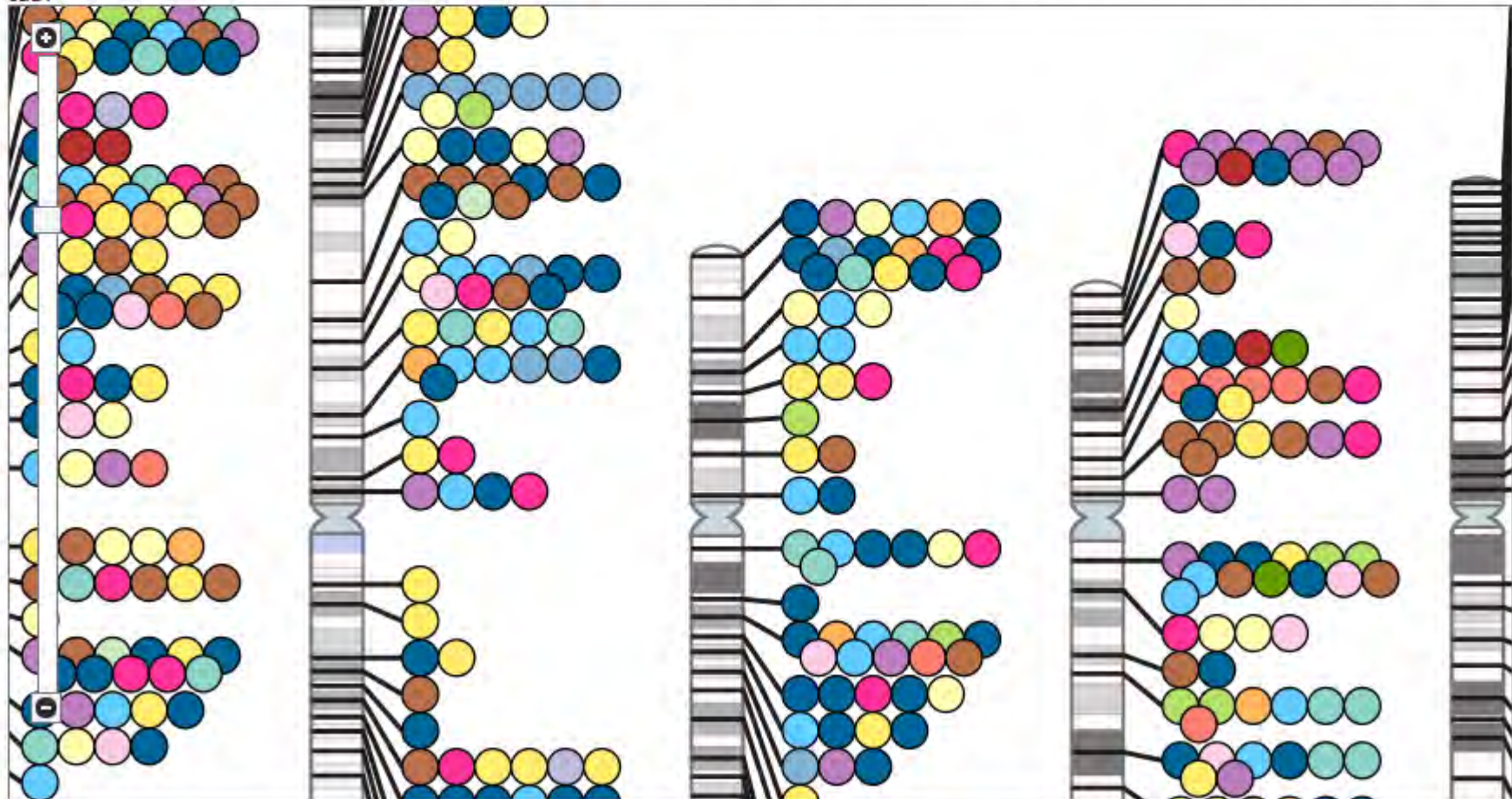
Downloads

Help

About

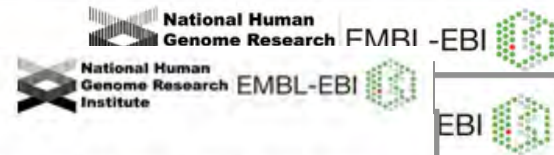
Show Legend

This diagram shows all SNP-trait associations with  $p\text{-value} \leq 5.0 \times 10^{-8}$ , published in the GWAS catalogue (<http://www.genome.gov/qwastudies>) up to the end of December 2012. For information on how to navigate the diagram, see the help tab.





# GWAS Diagram Browser



## GWAS Diagram Browser

Exploring Genome-wide Association Studies

*To show only one trait, e.g. "breast cancer" or "schizophrenia", type the trait into the box on the left and hit "Query by trait"*

[Interactive GWAS Diagram](#) [Trait-specific Views](#) [Time Series Views](#) [Downloads](#) [Help](#) [About](#)



### Background

The National Human Genome Research Institute (NHGRI) [Catalog of Published Genome-Wide Association Studies](#) or GWAS catalogue provides a quality controlled, manually curated, literature-derived collection of all published [GWA studies](#) assaying at least 100,000 [SNPs](#) and all SNP-trait associations with p-values  $< 1.0 \times 10^{-5}$  (*Hindorff et al., 2009*). The catalogue can be searched by a number of options including journal, first author, trait, chromosomal region, gene, SNP, odds ratio and p-value threshold. As of 20/08/12, it includes 1355 publications and 7226 SNPs. In addition to the SNP-trait association data, the catalogue also publishes the iconic GWAS diagram of all SNP-trait associations, with p-values  $\leq 5.0 \times 10^{-8}$ , mapped to the SNPs' chromosomal locations. The diagram is released quarterly and the latest version of the diagram is made available on the GWAS catalogue website in PDF format and as a PowerPoint slide.

### The GWAS diagram browser

This page hosts a novel system that generates the GWAS diagram dynamically using semantic web technologies. The diagram produced in this fashion can be filtered and searched at different levels of granularity and by different criteria, currently including trait (main diagram tab) and publication date (time series tab, currently available for a fixed set of time points only). It is possible to zoom in over chromosomes in order to see all SNP-trait associations for a given region. SNP-trait associations are shortly going to be fully interactive. Trait name information is already available on mouse-over and traits being clickable to allow proceeding from an association to the catalogue entry and to the publication is currently under development.

The diagram is generated in SVG (Scalable Vector Graphics) as a web application with a Java back-end using an Apache Tomcat server. The catalogue data is presented as a knowledge base, based on the GWAS Catalog schema ontology, which formalises the relationship between various concepts such as SNP, trait and study. Both the knowledge base and the schema ontology can be

"Enter" or clicking the "Query" button.

# Future diagram improvements

- Interactive links to other genome browsers
- Improved filtering features
  - Autocomplete (with synonyms)
  - PubMed ID
  - Combinatorial queries
- More frequent availability of updated data
- Improved browser compatibility
- (Your suggestion here!)

# Ethnicity

- Disproportionate disease burdens observed in non-European populations
- Frequency of GWAS variants varies up to 40-fold between different populations (Adeyamo, 2010)
- GWAS performed at a ratio of ~10:1 European ancestry vs. all groups combined (Need, 2009)
- Systematic efforts needed to define and extract ancestry information



## Add a GWA Study Ethnicity

### Instructions:

- To go back to the Input for GWAS Studies Table without changing or adding data, click on "View GWA Studies".
- To go back to the Edit GWA Studies Table without changing or adding data, click on "Edit GWA Studies".
- Once you have entered all the data for the Ethnicity, click the "Submit" button at the bottom of the form. You will be taken back to the Edit GWA Studies page.

Number of Individuals:

Inferred Ethnic/Ancestral Group:  
(Press the Ctrl key when selecting multiple groups)

  
 Select a group  
 European  
 Sub-Saharan African  
 African unspecified

Country of Origin:  
(Press the Ctrl key when selecting multiple countries)

  
**Africa**  
 Algeria  
 Angola  
 Benin  
 Botswana  
 Burkina Faso  
 Burundi  
 Cameroon  
 Cape Verde  
 Central African Republic  
 Chad  
 Comoros  
 Congo  
 Côte d'Ivoire  
 Democratic Republic of the Congo

Country of Recruitment:  
(Press the Ctrl key when selecting multiple countries)

  
**Africa**  
 Algeria  
 Angola  
 Benin  
 Botswana  
 Burkina Faso  
 Burundi  
 Cameroon  
 Cape Verde  
 Central African Republic  
 Chad  
 Comoros  
 Congo  
 Côte d'Ivoire  
 Democratic Republic of the Congo

Additional Description (free text):

# Ethnicity analysis

- 575 GWAS papers, 2011-2012
  - 55% EA only
  - 27% non-EA only
  - 13% both
- 74% of study participants are EA
- Countries of recruitment dominated by Europe, United States

eland

Hong Kong SAR

Mauritius

Gambia

Hungary

Argentina

Slovenia

Ethiopia







# Obesity-related traits

# Ontology

- Facilitates broad categorization of traits

200+ manually-defined traits

Abdominal aortic aneurysm	Coffee consumption	Hepatitis B vaccine response	Neuroblastoma	Response to metformin
Acute lymphoblastic leukemia	Cognitive function	Hepatohepatic carcinoma	Neovine dependence	Response to statin therapy
Adhesion molecules	Conduct disorder	Hirschsprung's disease	Obesity	Restless legs syndrome
Adiponectin levels	Colorectal cancer	HIV-1 control	Open angle glaucoma	Retinal vascular caliber
Age-related macular degeneration	Cornel thickness	Hodgkin's lymphoma	Open personality	Retinol levels
AIDS progression	Coronary disease	Homocysteine levels	Optic disc parameters	Rheumatoid arthritis
Alcohol dependence	Cortical thickness	HPV seropositivity	Osteoarthritis	Ribavirin-induced anemia
Alpecia areata	Crouzet-Jakob disease	Hypospadias	Osteoporosis	Schizophrenia
Alzheimer disease	Crohn's disease	Idiopathic pulmonary fibrosis	Otosclerosis	Serum metabolites
Amyloid A levels	Crohn's disease and celiac disease	IFN-related cytopeni	Other metabolic traits	Skin pigmentation
Amyotrophic lateral sclerosis	Culicivirus nevii	IgA levels	Ovarian cancer	Smoking behavior
Angiotensin-converting enzyme activity	Cyclic fibrosis severity	IgG levels	Pancreatic cancer	Speech perception
Ankylosing spondylitis	Dermatitis	Inflammatory bowel disease	Pain	Sphingolipid levels
Arterial stiffness	DHEA-4 levels	Insulin-like growth factors	Paget's disease	Statin-induced myopathy
Asparagins anionemia	Diabetic retinopathy	Intraocular aneurysm	Planic disorder	Stevens-Johnson syndrome
Aspirin	Dilated cardiomyopathy	Iris color	Parkinson's disease	Stroke
Atherosclerosis in HIV	Drug-induced liver injury	Iron status markers	Periodontitis	Sudden cardiac arrest
Atial fibrillation	Drug-induced liver injury	Ischemic stroke	Peripheral arterial disease	Suicide attempts
Atsisa atid hyperactivity disorder	Endometrial cancer	Juvenile idiopathic arthritis	Personality dimensions	Systemic lupus erythematosus
Autism	Endometriosis	Keloid	Phenolphthaleine levels	Systemic sclerosis
Basal cell cancer	Eosinophil count	Kidney stones	Phosphorus levels	T-tau levels
Behcet's disease	Eosinophilic esophagitis	LDL cholesterol	Photic sneeze	Tau AB1-42 levels
Bipolar disorder	Epubiclin-induced leukopenia	Leprosy	Phytosterol levels	Telomere length
Biliary atresia	Erdle-Sykes and prostate cancer treatment	Leplin receptor levels	Platelet count	Testicular germ cell tumor
Bilirubin	Erythrocyte parameters	Liver enzymes	Polycystic ovary syndrome	Thyroid cancer
Bitter taste response	Esophageal cancer	Longevity	Primary biliary cirrhosis	Thyroid volume
Birth weight	Essential tremor	LP (a) levels	Primary sclerosing cholangitis	Tooth development
Bladder cancer	Exfoliation glaucoma	LpPLA2) activity and mass	PR interval	Total cholesterol
Bleomycin sensitivity	Eye color traits	Lung cancer	Progesterone levels	Triglycerides
Blood or brown hair	F cell distribution	Magnesium levels	Progressive macular paly	Tuberculosis
Blood pressure	Fibronogen levels	Major mood disorders	Prostate cancer	Type 1 diabetes
Blue or green eyes	Folate pathway vitamins	Malaria	Protein levels	Type 2 diabetes
BMI, waist circumference	Follicular lymphoma	Male pattern baldness	PSA levels	Liverware coils
Bone density	Fuch's corneal dystrophy	Mammographic density	Psoriasis	Urate
Breast cancer	Freckles and burning	Matrix metalloproteinase levels	Psoriatic arthritis	Urinary albumin excretion
Buylrylcholinesterase levels	Gastones	MCP-1	Pulmonary funct. COPD	Urinary metabolites
C-reactive protein	Gastric cancer	Melanoma	QRS interval	Uremia fibrosis
Calcium levels	Glioma	Menarche & menopause	QT interval	Venous thrombembolism
Cardiac structure/function	Glycemic traits	Meningioma	Quantitative traits	Verticolar conduction
Cardiovascular risk factors	Graves disease	Meiringospl disease	Recombination rate	VEGF levels
Carotid levels	Hair color	Metabolic syndrome	Red vs non-red hair	Vertical cup-disc ratio
Carotenoid/cholesterol levels	Hair morphology	Migrane	Refractive error	Vitamin B12 levels
Carotid atherosclerosis	Hardness in dyslexia	Moyamoya disease	Renal cell carcinoma	Vitamin D insufficiency
Celiac disease	HDL cholesterol	Multiple sclerosis	Renal function	Vitamin E levels
Celiac disease and rheumatoid arthritis	Heart failure	Myeloproliferative neoplasms	Response to antidepressants	Vitigo
Cerebral atrophy measures	Heart rate	Myopia (pathological)	Response to antipsychotic therapy	Warfarin dose
Chronic lymphocytic leukemia	Height	N-glycan levels	Response to carbamazepine	Weight
Chronic myeloid leukemia	Height	Narcolepsy	Response to clopidogrel therapy	White cell count
Cleft lip/palate	Hemostasis parameters	Nasopharyngeal cancer	Response to hepatitis C treat	White matter hyperintensity
	Hepatic steatosis	Nutrenetic peptide levels	Response to interferon beta therapy	YKL-40 levels
	Hepatitis			

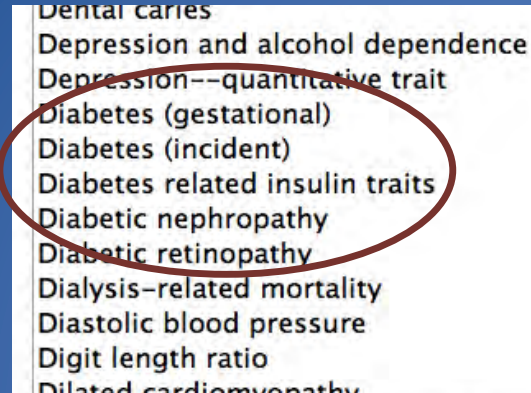
~20 ontology-defined traits

- Biological process
- Cancer
- Other disease
- Other traits
- Chemical compound
- Body measurements
- Lipid or lipoprotein measurement
- Cardiovascular measurements
- Hematological measurement
- Inflammatory marker measurement
- Liver enzyme measurement
- Other measurement
- Cardiovascular disorder
- Neurological disorder
- Immune system disorder
- Digestive system disorder
- Metabolic disorder

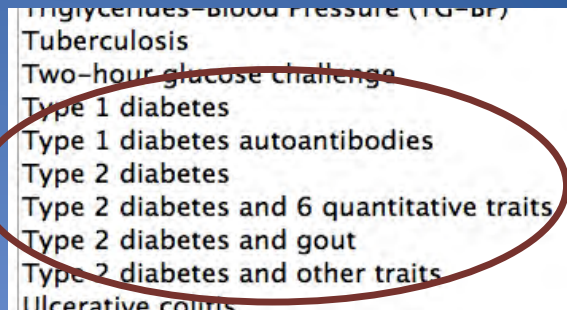


# Ontology

- Relatively unstructured trait list

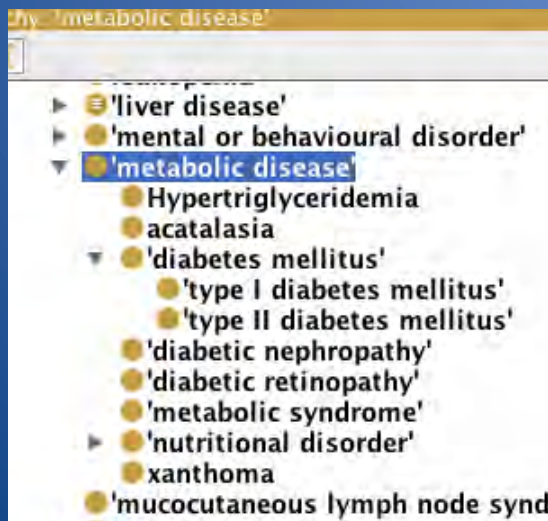


Dental caries  
Depression and alcohol dependence  
Depression--quantitative trait  
Diabetes (gestational)  
Diabetes (incident)  
Diabetes related insulin traits  
Diabetic nephropathy  
Diabetic retinopathy  
Dialysis-related mortality  
Diastolic blood pressure  
Digit length ratio  
Dilated cardiomyopathy



Triglycerides-blood pressure (TG-BP)  
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  
Ulcerative colitis

- Integration of traits into existing ontology (EFO) and facilitates systematic and more powerful searches



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



# Overview

- Goals and logistics of webinar
- Curation and display
- Recent enhancements
- Community adoption & uptake
- Discussion

# “Repurposing” GWAS catalog data



## Practice of Epidemiology

### Research Guidelines in the Era of Large-scale Collaborations: An Analysis of Genome-wide Association Study Consortia

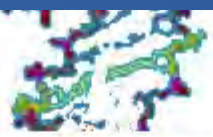
Melissa A. Austin\*, Marilyn S. Hair, and Stephanie M. Fullerton

\* Correspondence to Dr. Melissa A. Austin, Department of Epidemiology, School of Public Health, University of Washington, 1959 NE Pacific Street, Box 357236, Seattle, WA 98195-7236 (e-mail: maustin@u.washington.edu).

Initially submitted September 9, 2011; accepted for publication November 3, 2011.

Scientific research has shifted from studies conducted by single investigators to the creation of large consortia. Genetic epidemiologists, for example, now collaborate extensively for genome-wide association studies (GWAS). The effect has been a stream of confirmed disease-gene associations. However, effects on human subjects oversight, data-sharing, publication and authorship practices, research organization and productivity, and intellectual property remain to be examined. The aim of this analysis was to identify all research consortia that had published the results of a GWAS analysis since 2005, characterize them, determine which have publicly accessible guidelines for research practices, and summarize the policies in these guidelines. A review of the National Human Genome Research Institute's *Catalog of Published Genome-Wide Association Studies* identified 55 GWAS consortia as of April 1, 2011. These consortia were comprised of individual investigators, research centers, studies, or other

CNVs are  
that it v  
at least



Variation displays

- Explore this variation
- Genomic context
  - Genes and regulation (39)
  - Flanking sequence
- Population genetics
- Individual genotypes (2566)
- Linkage disequilibrium
- Phenotype Data (4)
- Phylogenetic Context (6)
- Citations (58)
- External Data
  - SNPedia
  - LOVD

- Configure this page
- Add your data
- Export data
- Bookmark this page
- Share this page
- Download view as CSV

rs7901695 SNP

**Original source** Variants (including SNPs and indels) imported from dbSNP (release 137) | [View in dbSNP](#)

**Alleles** Reference/Alternatives: **T/G/A/C** | Ancestral: C | Ambiguity code: N | MAF: 0.27 (C)

**Location** Chromosome **10:114754088** (forward strand) | [View in location tab](#)

**Co-located** with HGMD-PUBLIC [CS071265](#)

**Evidence status**

**Synonyms** [Archive dbSNP rs56472218](#), [rs56913805](#)

**HGVS names** This variation has **42** HGVS names - click the plus to show

**Genotyping chips** This variation has assays on **10** chips - click the plus to show

Phenotype Data

Significant data

Disease/Trait	Source(s)	Study	Clinical significance	Reported gene(s)	Associated variant(s)	Most associated allele	P value
Fasting proinsulin	<a href="#">[MAGIC]</a>				<a href="#">rs7901695</a>	T	1.009e-16
Type 2 diabetes	<a href="#">[NHGRI_GWAS_catalog]</a>	<a href="#">pubmed:17463249</a>		<a href="#">TCF7L2</a>	<a href="#">rs7901695</a>	rs7901695-C	1e-48

Association Results

<b>Cited Variants</b>	<b>114751086..114757090</b> SNPs from Genome-wide association analyses. They are from NHGRI Catalog and from association results submitted to dbGaP	1
-----------------------	--	---

Genes [Links & Tools](#)  
[GaP Browser](#)



Search All Databases  
  
Search Clear

Welcome

The Ph  
housed  
oriented  
variants  
location  
dynam  
terms a

Search

Search

Phen

Tra

European Journal of Human Genetics (2013), 1-4  
© 2013 Macmillan Publishers Limited. All rights reserved 1018 4813/13  
www.nature.com/ejhg



**SHORT REPORT**

# Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources

Erin M Ramos<sup>1</sup>, Douglas Hoffman<sup>2</sup>, Heather A Junkins<sup>1</sup>, Donna Maglott<sup>2</sup>, Lon Phan<sup>2</sup>, Stephen T Sherry<sup>2</sup>,  
Mike Feolo<sup>\*,2</sup> and Lucia A Hindorf<sup>\*,1</sup>

Rapidly accumulating data from genome-wide association studies (GWASs) and other large-scale studies are most useful when synthesized with existing databases. To address this opportunity, we developed the Phenotype-Genotype Integrator (PheGenI), a user-friendly web interface that integrates various National Center for Biotechnology Information (NCBI) genomic databases with association data from the National Human Genome Research Institute GWAS Catalog and supports downloads of search results.

Chromosome:   
Range (bps):   
(from:to)

SNP Functional Class

# Key challenges and opportunities

- Curation a key bottleneck; automated data mining may facilitate
- Integration with other resources and data types will further biological understanding
- Providing data in a user-friendly way will increase adoption
- Collaborations can enable high priority improvements

# Acknowledgements

## NHGRI

Vivien Bonazzi  
Peggy Hall  
Heather Junkins  
Kent Klemm  
Darryl Leja  
Teri Manolio

## NCBI

Mike Feolo  
Zhen Wang  
Ming Xu



## EBI

Helen Parkinson  
Tony Burdett  
Jon Ison  
Simon Jupp  
James Malone  
Jackie MacArthur  
Joannella Morales  
Dani Welter

Funding:  
NHGRI grant 3U41-  
HG006104-01S1



# Overview

- Goals and logistics of webinar
- Curation and display
- Recent enhancements
- Community adoption & uptake
- Discussion