Navigating the Epidemiology of the Human Genome

CCGTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACGT

Marta Gwinn, MD, MPH
National Office of Public Health Genomics, CDC

Personal Genomics Workshop
Bethesda, MD – December 17, 2008

Navigating the Epidemiology of the Human Genome

CCGTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACGT

Size up the domain

- population-based research studies

Map the topography

- population prevalence, associations, interactions, genetic tests

Identify major features

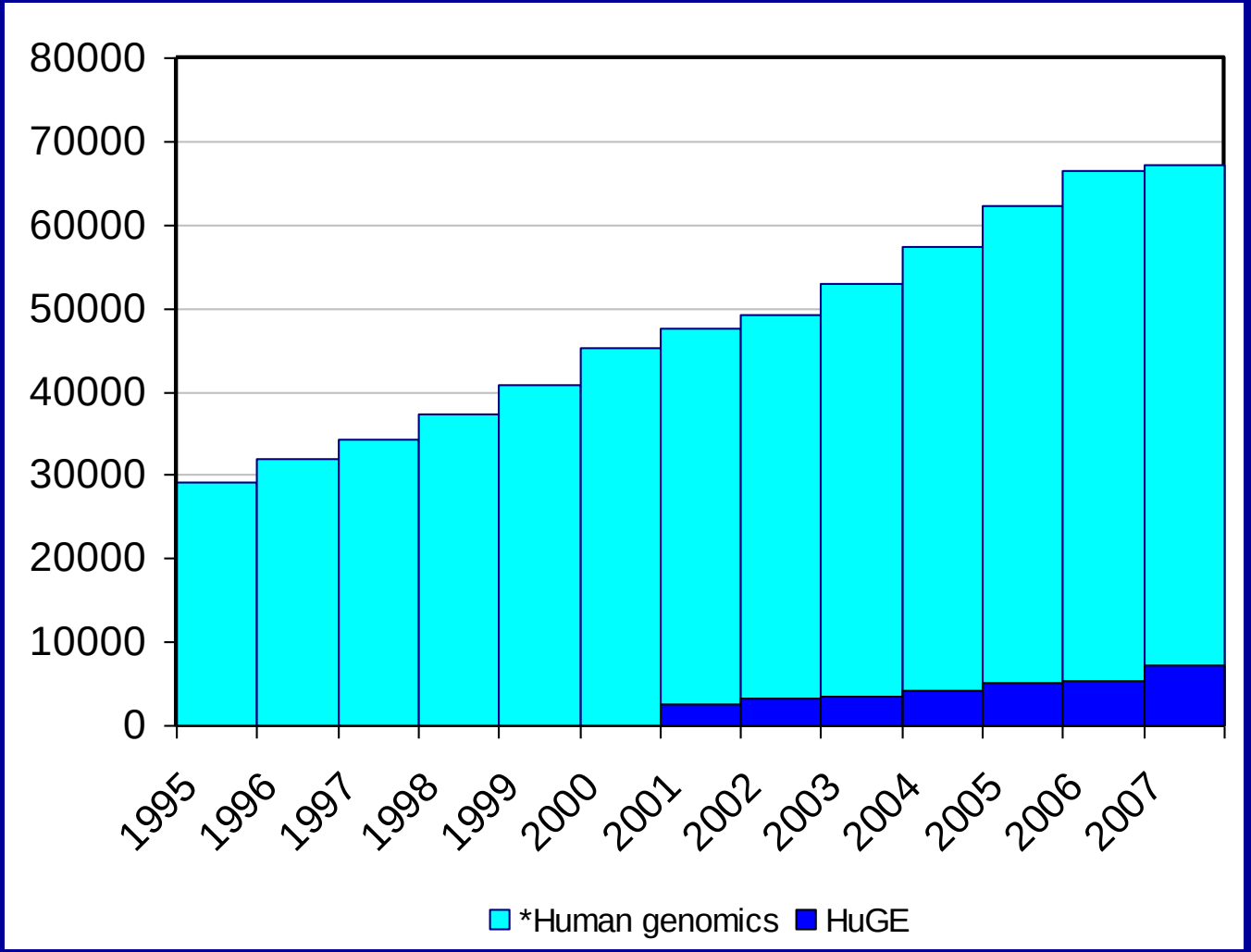
- consistent findings, patterns, and gaps

Develop navigational tools for further exploration

- reliable knowledge base with applications

www.hugenavigator.net

PubMed citations on human genetics/genomics and human genome epidemiology (HuGE)



*PubMed query: “gene or genetic or genome or genomic” / limits: human, year and HuGE Navigator, Oct 2008

Size up the domain

(((((((((((((((((((((((genetic[All Fields] AND (((((("disease"[MeSH Terms] OR ("disease susceptibility"[MeSH Terms] OR predisposition[Text Word])) OR disease[Text Word]) OR defect[Text Word]) OR susceptibility[Text Word]) OR ("counseling"[MeSH Terms] OR counseling[Text Word]))) OR (("disease susceptibility"[MeSH Terms] OR susceptibility[Text Word]) AND (("genes"[MeSH Terms] OR gene[Text Word]) OR ("genes"[MeSH Terms] OR genes[Text Word])))) OR ((((("mutation"[MeSH Terms] OR mutation[Text Word]) OR (("genes"[MeSH Terms] OR gene[Text Word]) AND ("mutation"[MeSH Terms] OR mutation[Text Word]))) OR (("mutation"[MeSH Terms] OR mutations[Text Word]) AND ("genes"[MeSH Terms] OR gene[Text Word]))) OR (("mutation"[MeSH Terms] OR mutations[Text Word] OR mutation[Text Word]) AND ("genes"[MeSH Terms] OR gene[Text Word])))) OR ("hereditary diseases"[MeSH Terms] OR genetic disorder[Text Word])) OR (genetic[All Fields] AND ((("TEST"[Substance Name] OR ("TEST"[Substance Name] OR test[Text Word])) OR ("research design"[MeSH Terms] OR testing[Text Word])) OR study[All Fields]))) OR ("genetic screening"[MeSH Terms] OR genetic screening[Text Word])) OR (genetic[All Fields] AND ("risk"[MeSH Terms] OR risk[Text Word]))) OR ("polymorphism (genetics)"[MeSH Terms] OR ("polymorphism (genetics)"[MeSH Terms] OR polymorphism[Text Word]))) OR ((((("genotype"[MeSH Terms] OR ("genotype"[MeSH Terms] OR genotype[Text Word])) OR genotyping[All Fields]) OR ("haplotypes"[MeSH Terms] OR haplotype[Text Word])) OR ("haplotypes"[MeSH Terms] OR haplotypes[Text Word]))) OR ((("genome"[MeSH Terms] OR genome[Text Word]) OR genomic[All Fields]) OR ("Genomics"[MeSH Terms] OR genomics[Text Word]))) OR (((gene-environment) OR (gene AND environment)) AND interaction[Text Word])) OR (((genetic[Text Word] OR gene[Text Word]) OR allelic[All Fields]) AND ((variant[All Fields] OR variants[All Fields]) OR (("epidemiology"[MeSH Subheading] OR "epidemiology"[MeSH Terms]) OR frequency[Text Word])))) OR (("alleles"[MeSH Terms] OR allele[Text Word]) OR ("alleles"[MeSH Terms] OR alleles[Text Word]))) OR ("heterozygote detection"[MeSH Terms] OR Heterozygote Detection[Text Word])) OR ((Neonatal[All Fields] OR ("infant, newborn"[MeSH Terms] OR newborn[Text Word])) AND (("diagnosis"[MeSH Subheading] OR "mass screening"[MeSH Terms]) OR Screening[Text Word]))) OR germline[All Fields]) OR somatic[All Fields]) OR ("human genome project"[MeSH Terms] OR human genome project[Text Word])) AND ((((((((((((((((((((("epidemiology"[Subheading] OR "epidemiology"[MeSH Terms]) OR epidemiology[Text Word]) OR ("public health"[MeSH Terms] OR public health[Text Word])) OR ((("alleles"[MeSH Terms] OR allele[Text Word]) OR allelic[All Fields]) AND ((("epidemiology"[MeSH Subheading] OR "epidemiology"[MeSH Terms]) OR frequency[Text Word]) OR frequencies[All Fields]))) OR ("public policy"[MeSH Terms] OR policy[Text Word])) OR (("education"[Subheading] OR "education"[MeSH Terms]) OR education[Text Word])) OR "prevalence"[MeSH Terms]) OR prevalence[Text Word]) OR ("prevention and control"[Subheading] OR prevention[Text Word])) OR ("risk"[MeSH Terms] OR risk[Text Word])) OR (((((((population[Text Word] OR (a number of) OR genetic[All Fields]) OR comparative[All Fields]) OR prospective[All Fields]) OR cohort[All Fields]) OR cross-section[All Fields]) OR cross-sectional[All Fields]) OR case-control[All Fields]) AND (studies OR study[All Fields]))) OR (clinical trial[All Fields] OR randomized controlled trial[All Fields])) OR (("drug interactions"[MeSH Terms] OR interactions[Text Word]) OR (("interpersonal relations"[MeSH Terms] OR "drug interactions"[MeSH Terms]) OR interaction[Text Word]))) OR ("questionnaires"[MeSH Terms] OR questionnaire[Text Word])) OR (("sensitivity and specificity"[MeSH Terms] OR sensitivity[Text Word]) OR ("sensitivity and specificity"[MeSH Terms] OR specificity[Text Word]))) OR ((((case[All Fields] OR cases[All Fields]) OR ("patient[MeSH Terms] OR patients[Text Word])) OR (study[All Fields] AND group[All Fields])) OR ((((((("prevention and control"[MeSH Subheading] OR control[All Fields]) OR controls[All Fields]) OR (healthy[All Fields] AND subjects[All Fields])) OR ("child"[MeSH Terms] OR children[Text Word])) OR ("adult"[MeSH Terms] OR adults[Text Word])) OR individuals[All Fields]))) OR ((((("association"[MeSH Terms] OR association[Text Word]) OR ("association"[MeSH Terms] OR associations[Text Word])) OR ("disease"[MeSH Terms] OR disease[Text Word])) AND (("genes"[MeSH Terms] OR gene[Text Word]) OR ("genes"[MeSH Terms] OR genes[Text Word]))))) OR oversight[All Fields]) OR ((("genotype"[MeSH Terms] OR genotype[All Fields]) OR allelic[All Fields]) AND distribution[Text Word])) OR ((((("genotype"[MeSH Terms] OR genotype[Text Word]) AND ("phenotype"[MeSH Terms] OR phenotype[Text Word])) OR genotype[All Fields]) AND correlation[All Fields])) OR ((positive OR negative) AND predictive value)) OR (odds ratio)) OR (("ethics"[MeSH Terms] OR ethics[Text Word]) OR ethical[All Fields]))))) AND "2004/7/7 8.00"[MHDA]:"2004/7/14 8.00"[MHDA])

HuGE search query 2004

Knowledge base for Human Genome Epidemiology

GTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACG

- prevalence, associations, interactions, genetic tests
- updated weekly from PubMed since 2001
- combination of human and machine curation processes


SVM screens PubMed

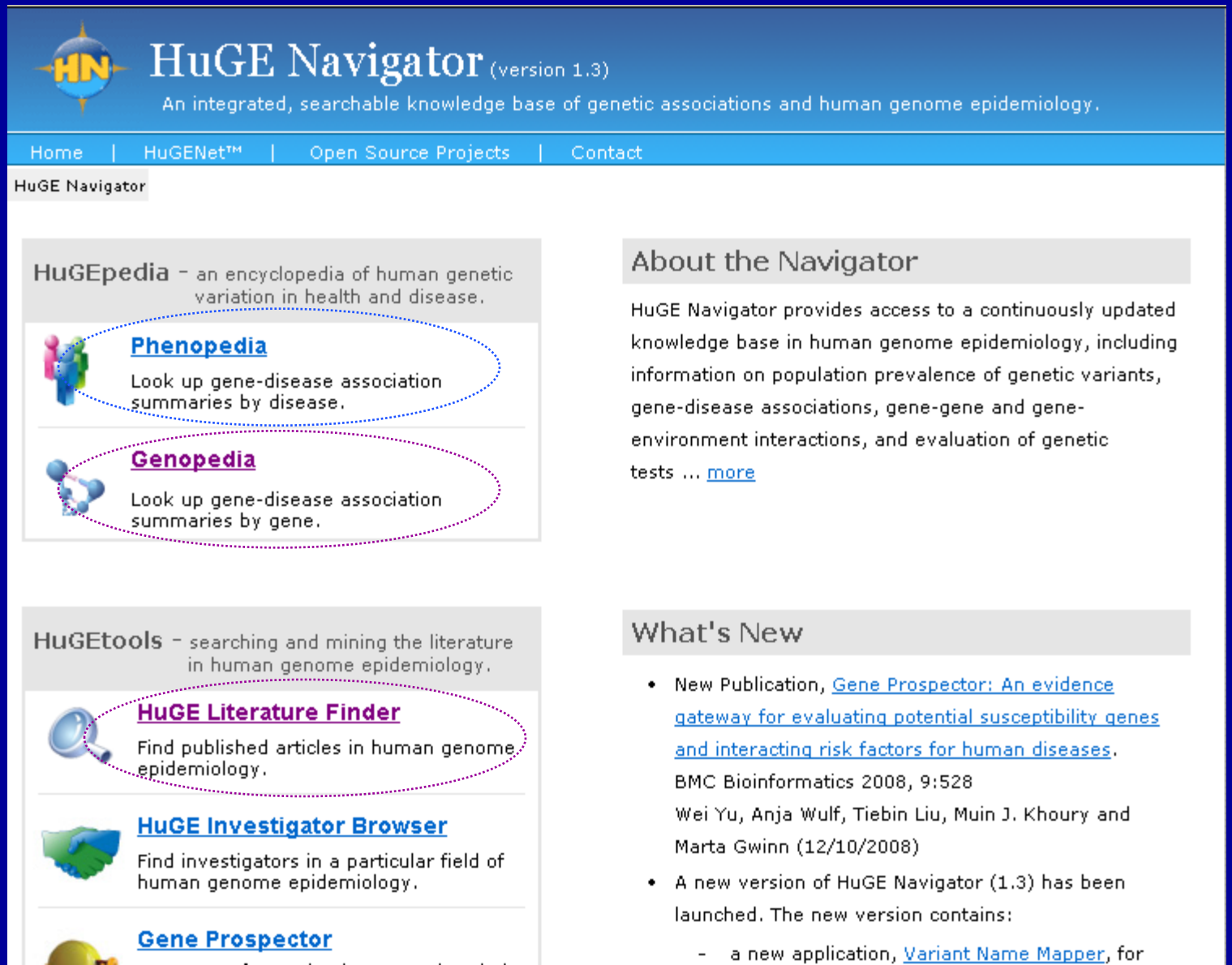

Curator selects, indexes


Auto-indexing (Entrez Gene, MeSH, UMLS)


Web applications search, sort, filter, display data

www.hugenavigator.net

HuGE Navigator: www.hugenavigator.net



HuGE Published Literature

GTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACG

As of December 11, 2008, the knowledge base contained:

34,208 genetic association studies

863 meta-analyses

243 GWAS

3,888 genes

1,958 MeSH disease terms

5,645 common variant names matched to rs numbers

…but numbers alone mean little…

about

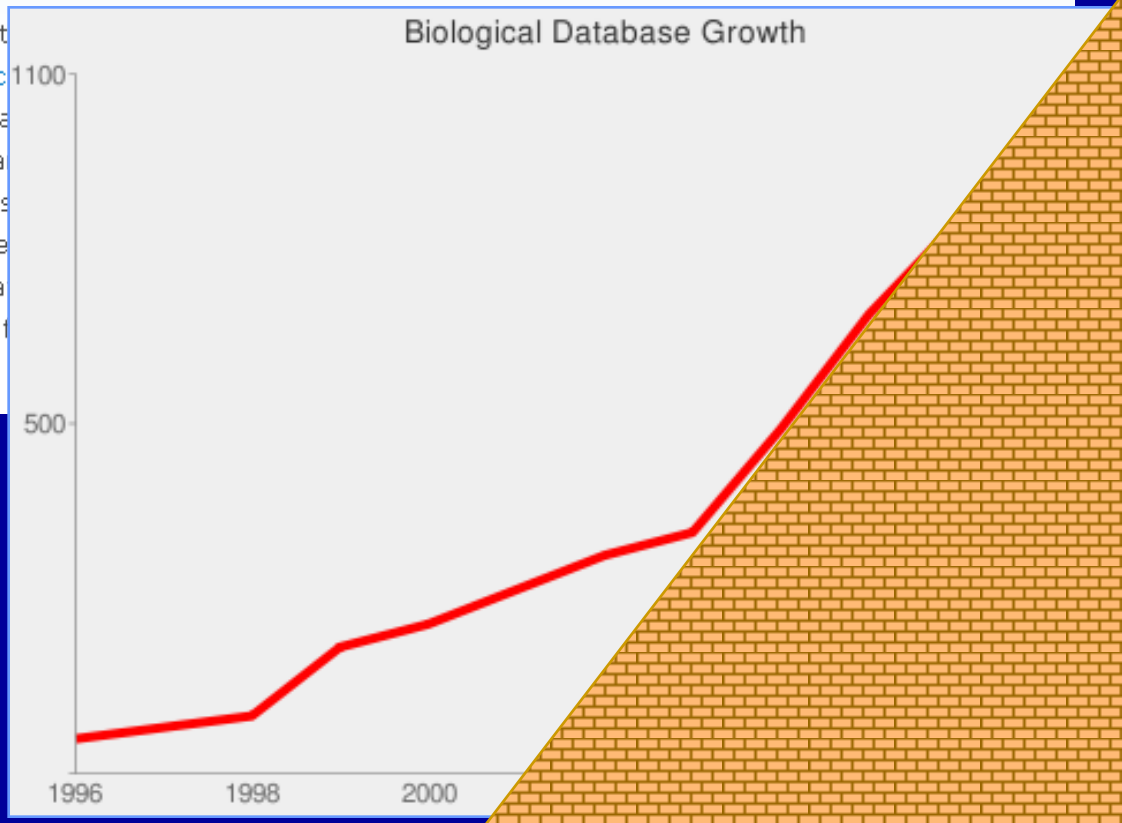
nodalpoint.org

Home › Blogs › Duncan's blog

One Thousand Databases High (and rising)

Fri, 2008-01-18 21:08 — Duncan

Well it's that
stamp collec
also known a
published ea
databases lis
previous one
thousand da
proportion of
be used?



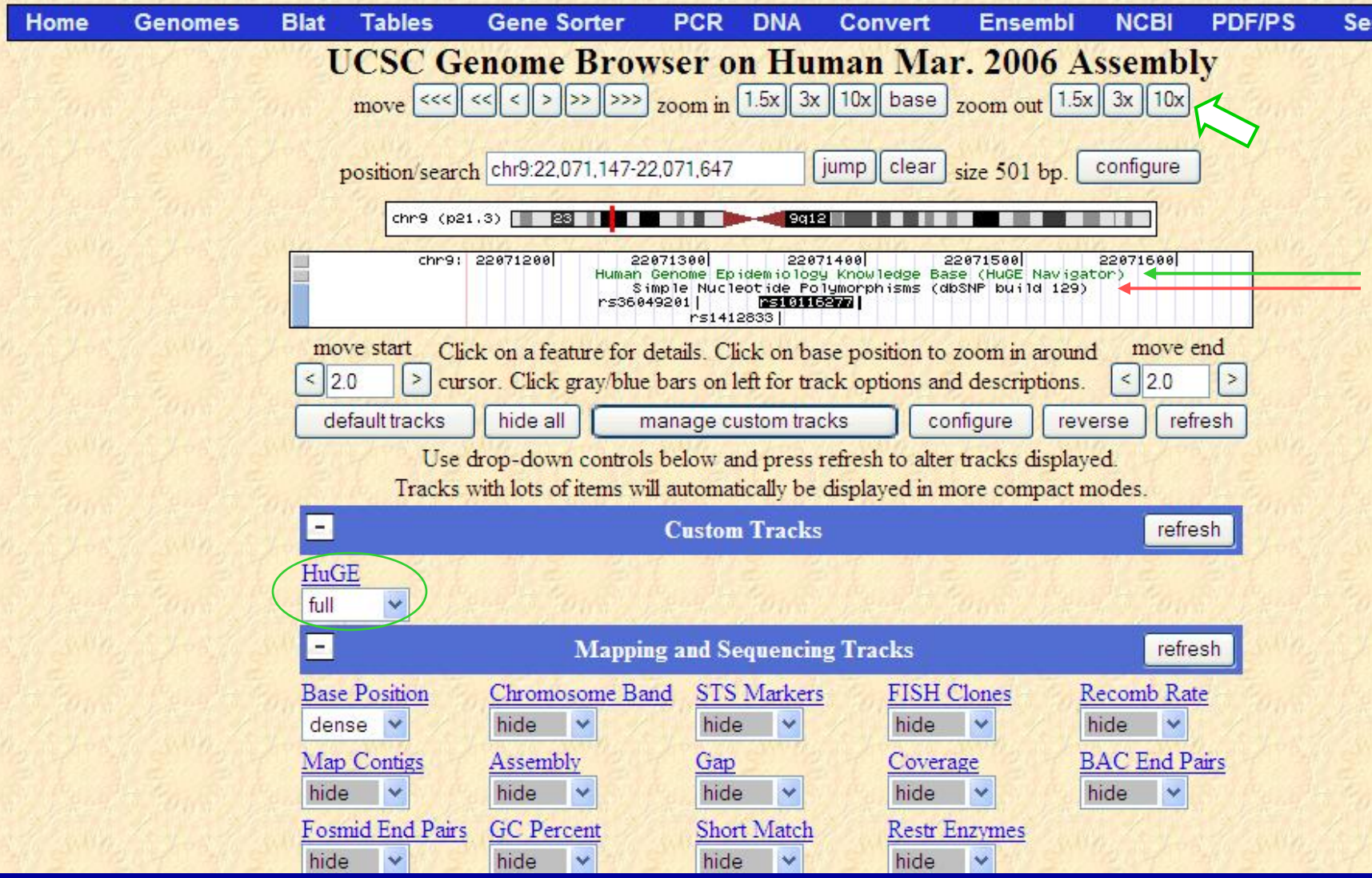
data tombs

Nucleic Acids Research, 2008

Duncan Hull, http://www.nodalpoint.org/2008/01/18/one_thousand_databases_high_and_rising

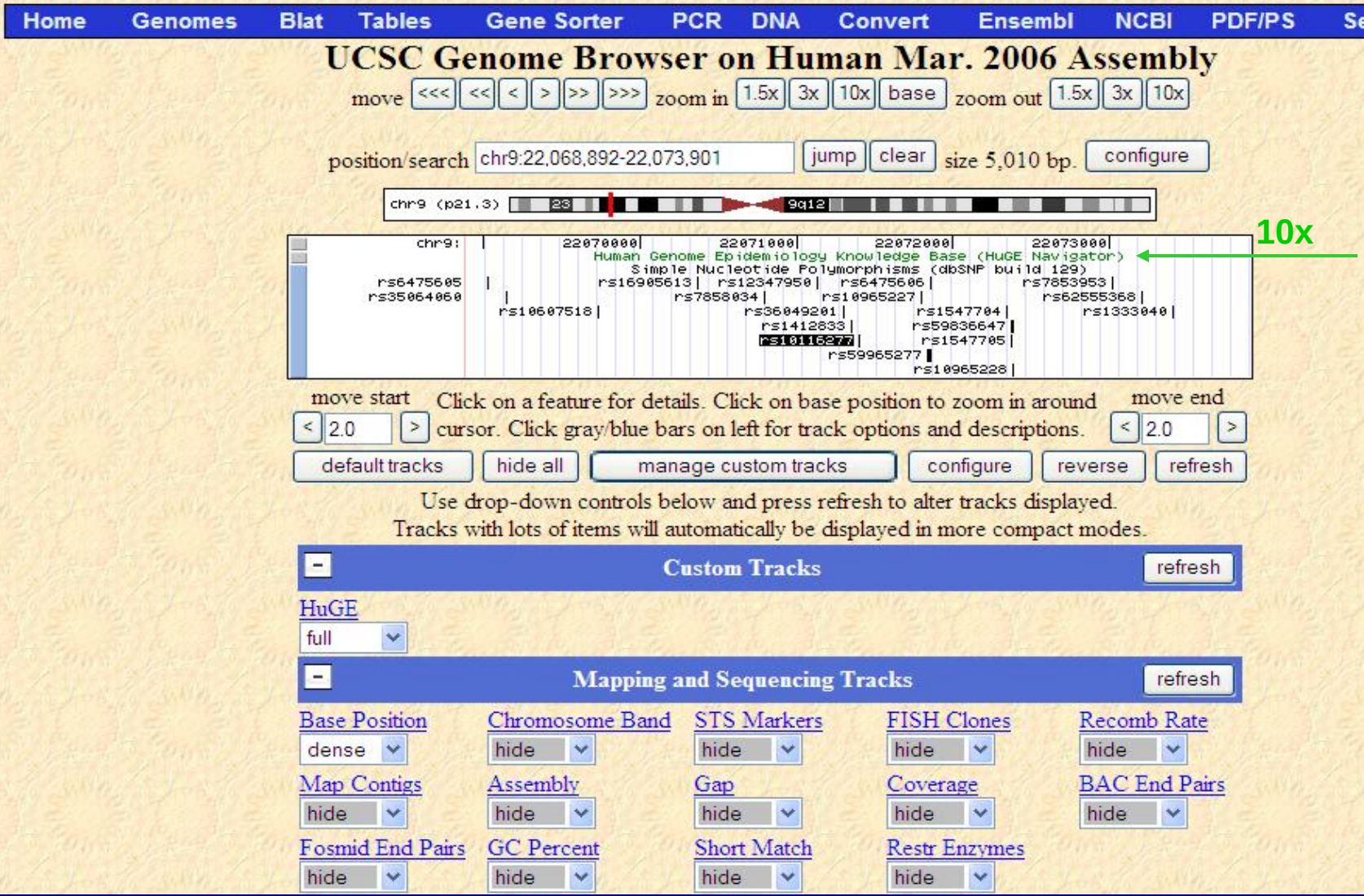
Flying Over the Columbia Hills of Mars
Animated Illustration Credit: Doug Ellison, Randolph Kirk (USGS), MSSS, MER, NASA

http://apod.nasa.gov/apod/ap080519m.html



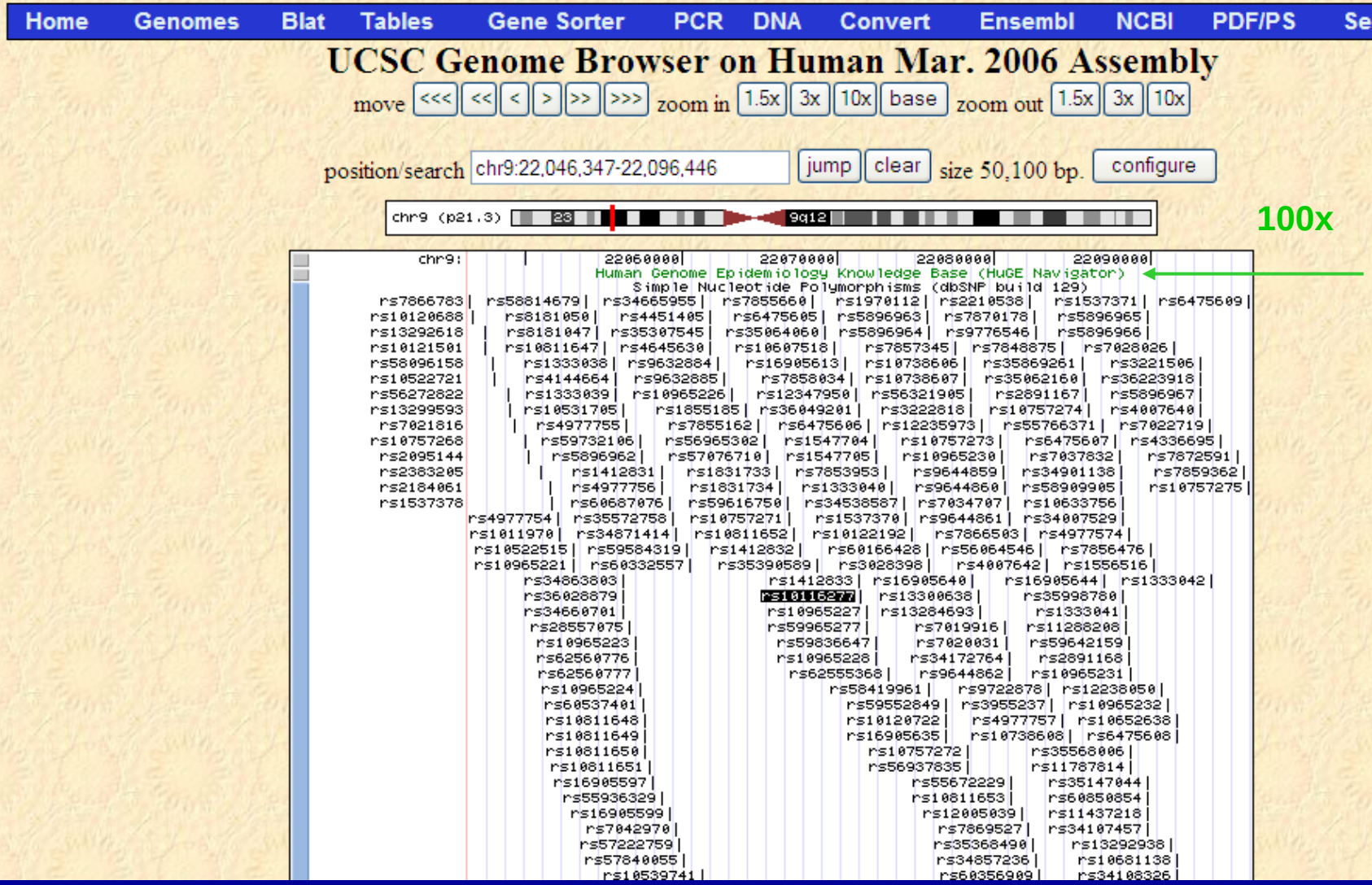
view for **rs10116277** on UCSC Genome Browser

http://genome.ucsc.edu/



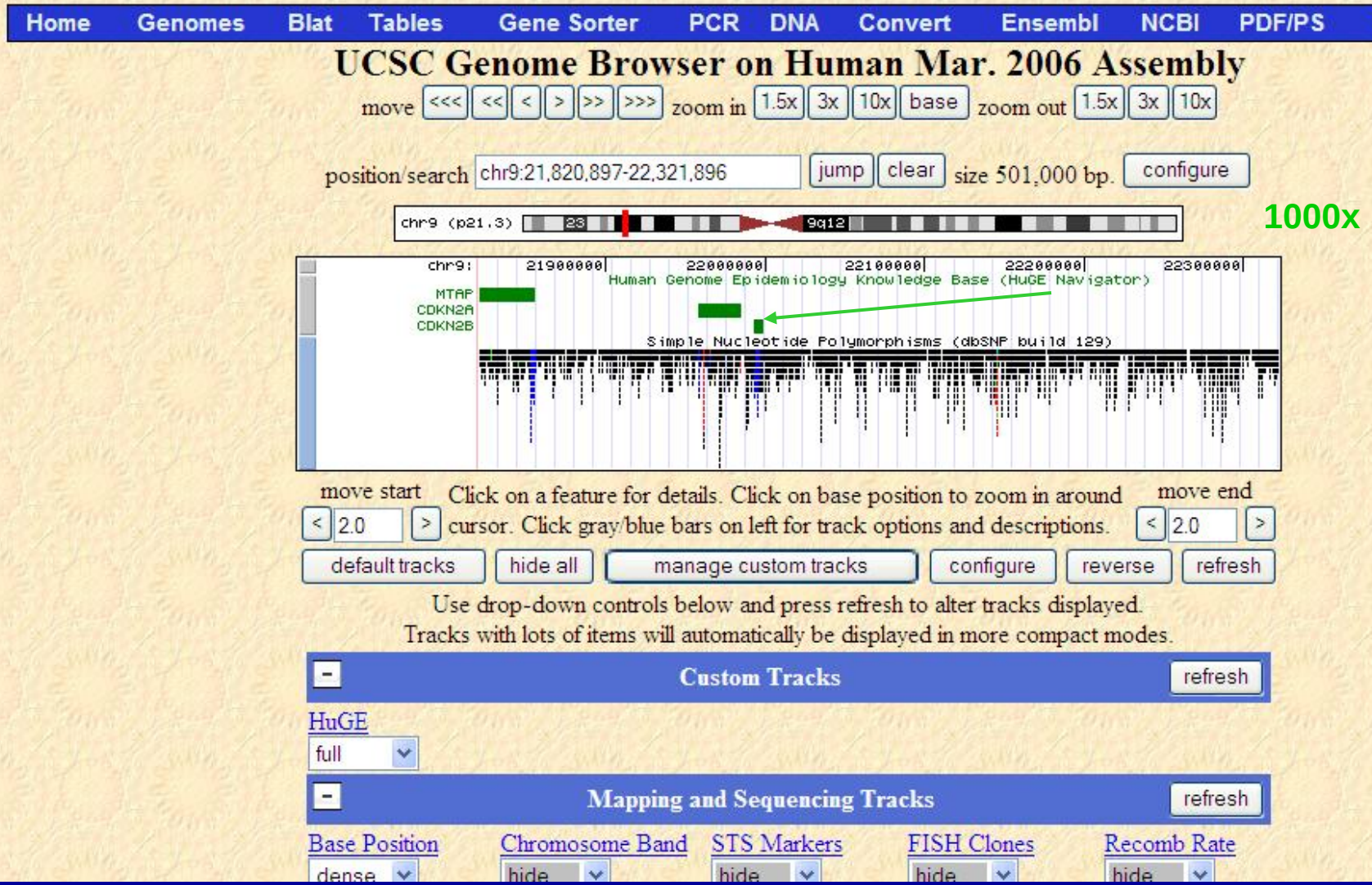
10x view for **rs10116277** on UCSC Genome Browser

http://genome.ucsc.edu/



100x view for **rs10116277** on UCSC Genome Browser

http://genome.ucsc.edu/



1000x view for **rs10116277** on UCSC Genome Browser

http://genome.ucsc.edu/

HuGE Navigator (version 1.3)

An integrated, searchable knowledge base of genetic associations and human genome epidemiology.

Last data upload: 03 Dec 2008. (Total 4315 genes)

Genopedia

Data collected since 2001

CDKN2B SNP

Related Disease Genes

Summary

- Total Publications: 45
- Gene Prevalence: 2
- Meta-Analyses: 5
- MA Summary: NA
- Disease: 23
- Investigator: 82(F/L)/ 534(All)
- Trend

22 disease terms (MeSH) have been reported with CDKN2B gene.

[Click to re-sort the table]

Disease Term (MeSH)	Total	Meta	GWAS	Gene-Env	Trend
Diabetes Mellitus, Type 2	17	4	3	1	
Melanoma	3	0	0	0	
Breast Neoplasms	2	0	0	0	
Insulin Resistance	2	1	1	0	
Coronary Arteriosclerosis	2	0	2	0	
Coronary Disease	2	0	1	0	
Ovarian Neoplasms	2	0	0	0	
Skin Neoplasms	1	0	0	0	
Thyroid Neoplasms	1	0	0	0	
Uveal Neoplasms	1	0	0	0	
Diabetes Mellitus	1	0	0	0	
Diabetes Mellitus, Type 1	1	1	0	0	
Glucose Intolerance	1	0	0	0	
Hyperglycemia	1	0	0	0	
Hyperlipoproteinemia Type II	1	0	0	0	

Links:

- Entrez Gene
- GeneCard
- PharmGKB

HuGE Navigator (version 1.3)

An integrated, searchable knowledge base of genetic associations and human genome epidemiology.

Home | HuGENet™ | Open Source Projects | Contact

HuGE Navigator > HuGE Literature Finder

Last data upload: 11 Dec 2008. (Total 38607 articles)

HuGE Literature Finder

Data collected since 2001

Home | About | Search Instructions | FAQs

Search Literature **for** Coronary Arteriosclerosis and gwas and CDKN2B Go Clear

Search Criteria: Coronary Arteriosclerosis and gwas and CDKN2B[Text+MeSH] **[Query Detail]** Download

Filtered By Disease Gene Category StudyType Year Author Journal Country

Articles 1 - 2 of 2

Export **Display** 25 on Page 1 of 1

1. A common allele on chromosome 9 associated with coronary heart disease. **[Detail]**
GWAS
Science (New York, N.Y.) 2007 Jun 316 (5830): 1488-91.
McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC

2. A common variant on chromosome 9p21 affects the risk of myocardial infarction. **[Detail]**
GWAS
Science (New York, N.Y.) 2007 Jun 316 (5830): 1491-3.
Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K

Export **Display** 25 on page 1 of 1

Articles 1 - 2 of 2

Download

CDKN2B SNP

- Related Disease Genes -

Summary

- Total Publications: 45
- Gene Prevalence: 2
- Meta-Analyses: 5
- MA Summary: NA
- Disease: 23
- Investigator: 82(F/L)/ 534(All)
- Trend

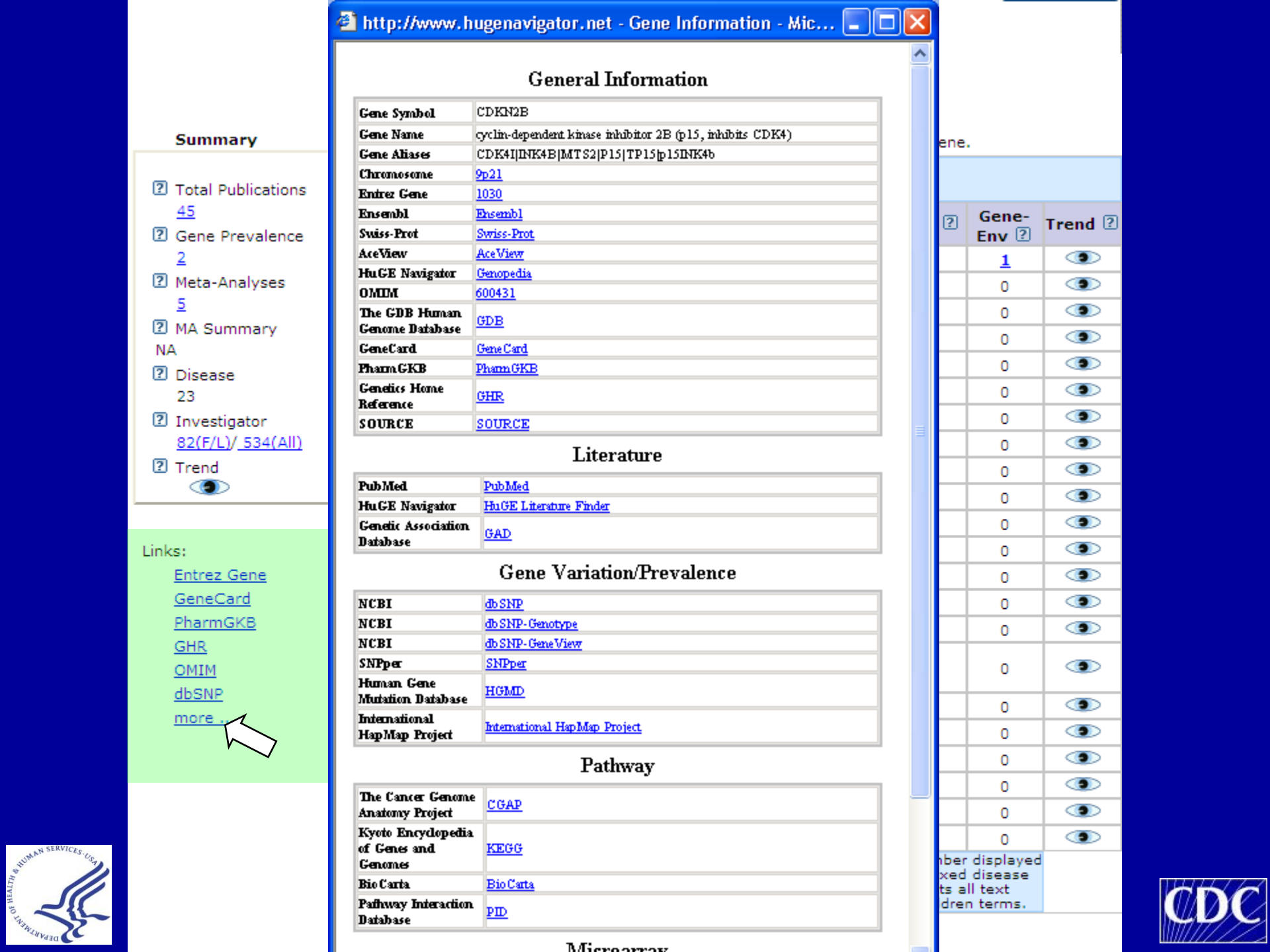
Links:
- Entrez Gene
- GeneCard
- PharmGKB
- GHR
- OMIM
- dbSNP
- more ...

22 disease terms (MeSH) have been reported with CDKN2B gene.

[Click to re-sort the table]

Disease Term (MeSH)	Total	Meta	GWAS	Gene-Env	Trend
Diabetes Mellitus, Type 2	17	4	3	1	
Melanoma	3	0	0	0	
Breast Neoplasms	2	0	0	0	
Insulin Resistance	2	1	1	0	
Coronary Arteriosclerosis	2	0	2	0	
Coronary Disease	2	0	1	0	
Ovarian Neoplasms	2	0	0	0	
Skin Neoplasms	1	0	0	0	
Thyroid Neoplasms	1	0	0	0	
Uveal Neoplasms	1	0	0	0	
Diabetes Mellitus	1	0	0	0	
Diabetes Mellitus, Type 1	1	1	0	0	
Glucose Intolerance	1	0	0	0	
Hyperglycemia	1	0	0	0	
Hyperlipoproteinemia Type II	1	0	0	0	
Leukemia, Lymphocytic, Acute, L1	1	0	0	0	
Leukemia, Pre-B-Cell	1	0	0	0	
Lymphoma, Non-Hodgkin	1	0	0	0	
Carcinoma, Medullary	1	0	0	0	
Colorectal Neoplasms	1	0	0	0	
Myocardial Infarction	1	0	1	0	
Obesity	1	0	0	0	

Note: The number of publications displayed in this table will differ from the number displayed in the HuGE Literature Finder as the number in GenoPedia reflects only the indexed disease term without children terms, but the number in the HuGE Literature Finder reflects all text searches of the disease term including the indexed term and corresponding children terms.


http://www.hugenavigator.net - Gene Information - Mic...
General Information
Gene Symbol CDKN2B
Gene Name cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
Gene Aliases CDK4I|INK4B|MTS2|P15|TP15|p15INK4b
Chromosome 9p21
Entrez Gene 1030
Ensembl Ensembl
Swiss-Prot Swiss-Prot
AceView AceView
HuGE Navigator Genopedia
OMIM 600431
The GDB Human Genome Database GDB
GeneCard GeneCard
PharmGKB PharmGKB
Genetics Home Reference GHR
SOURCE SOURCE
Literature
PubMed PubMed
HuGE Navigator HuGE Literature Finder
Genetic Association Database GAD
Gene Variation/Prevalence
NCBI dbSNP
NCBI dbSNP-Genotype
NCBI dbSNP-GeneView
SNPper SNPper
Human Gene Mutation Database HGMD
International HapMap Project International HapMap Project
Pathway
The Cancer Genome Anatomy Project CGAP
Kyoto Encyclopedia of Genes and Genomes KEGG
BioCarta BioCarta
Pathway Interaction Database PID
Summary
Total Publications
45
Gene Prevalence
2
Meta-Analyses
5
MA Summary
NA
Disease
23
Investigator
82(F/L)/ 534(All)
Trend
Links:
Entrez Gene
GeneCard
PharmGKB
GHR
OMIM
dbSNP
more ...
Gene-Env
Trend
1
0
ene.
ber displayed
xed disease
ts all text
dren terms.

HuGE Navigator (version 1.3)

An integrated, searchable knowledge base of genetic associations and human genome epidemiology.

Home | HuGENet™ | Open Source Projects | Contact

HuGE Navigator > Phenopedia (HuGEpedia)

Last data upload: 11 Dec 2008. (Total 1958 disease terms)

Phenopedia

Data collected since 2001

Home | About | Search Instructions | FAQs

Search Phenopedia for Enter a disease term Go Clear All

Download

Myocardial Infarction

Related Diseases

Summary

- Total Publications 1012
- Meta-Analyses 30
- MA Summary
- Genes 318
- GWAS Publications 11
- Investigators 1315(F/L) 4753(All)
- Trend

Last Update: 11 Dec 2008

Field Synopsis

NA

367 genes have been reported with Myocardial Infarction

[Click to re-sort the table]

Group genes by KEGG

Gene	Total	Meta	GWAS	Gene-Env	Trend
ACE SNP	80	4	0	17	
APOE SNP	49	1	0	9	
MTHFR SNP	45	3	0	5	
NOS3 SNP	42	1	0	7	
F5 SNP	41	3	0	4	
SERPINE1 SNP	35	1	0	5	
ITGB3 SNP	34	3	0	4	
AGTR1 SNP	31	1	0	4	
F2 SNP	28	2	0	3	
IL6 SNP	26	0	0	1	
LTA SNP	25	1	1	2	
FGB SNP	25	0	0	2	
CETP SNP	24	0	0	5	
AGT SNP	22	2	0	4	

About Entrez
Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI

Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

1: Clin Chem Lab Med. 2006;44(3):274-81.

Related Articles, Link

FULL TEXT ARTICLE de Gruyter ReferenceGlobal

Associations of apolipoprotein E exon 4 and lipoprotein lipase S447X polymorphisms with acute ischemic stroke and myocardial infarction.

Baum L, Ng HK, Wong KS, Tomlinson B, Rainer TH, Chen X, Cheung WS, Tang J, Tam WW, Goggins W, Tong CS, Chan DK, Thomas GN, Chook P, Woo KS.

Department of Medicine and Therapeutics, Chinese University of Hong Kong, Shatin, Hong Kong. lwbaum@cuhk.edu.hk

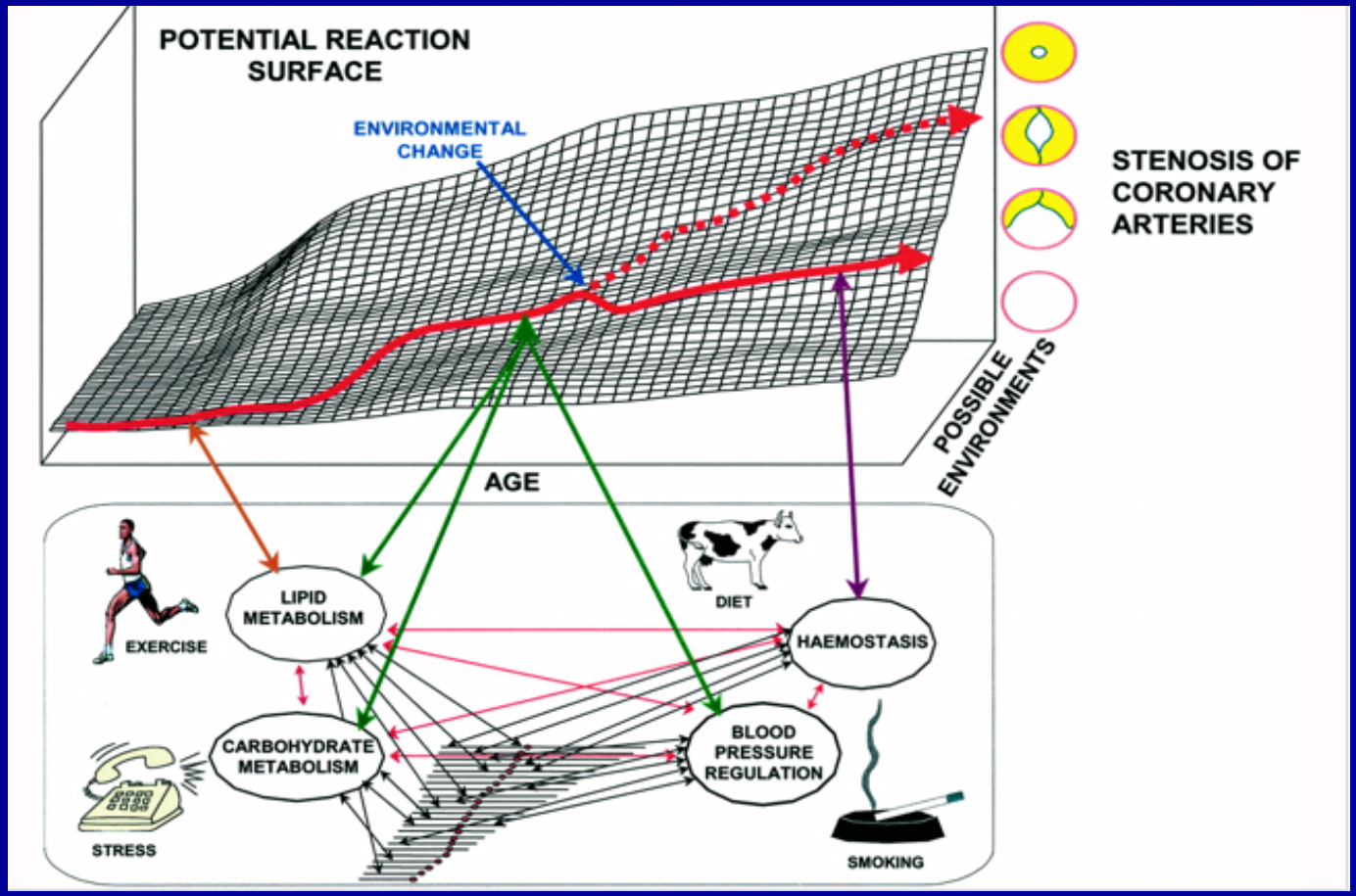
BACK… with each other and with other fa… ion with ischemic vascular… patients, 234 acute myocar… RESULTS:APOE epsilon2… infarction the epsilon4 allele w… 06). For subjects with either A… 01). LPL X alleles displaye… r APOE genotype. Meta-a… LUSION: This is the first rep… es. Therefore, APOE genotyp… on and ischemic stroke

Only one meta-analysis has examined APOE alleles in association with myocardial infarction, concluding:

"For subjects with either APOE epsilon2 or epsilon4 alleles, LPL X alleles were increased in vascular disease (OR = 2.2, p = 0.01). LPL X alleles displayed opposite tendencies toward association with disease when subjects were divided by sex, smoking, or APOE genotype. Meta-analysis and regression analysis of previous studies supported the sex and smoking dichotomies…."

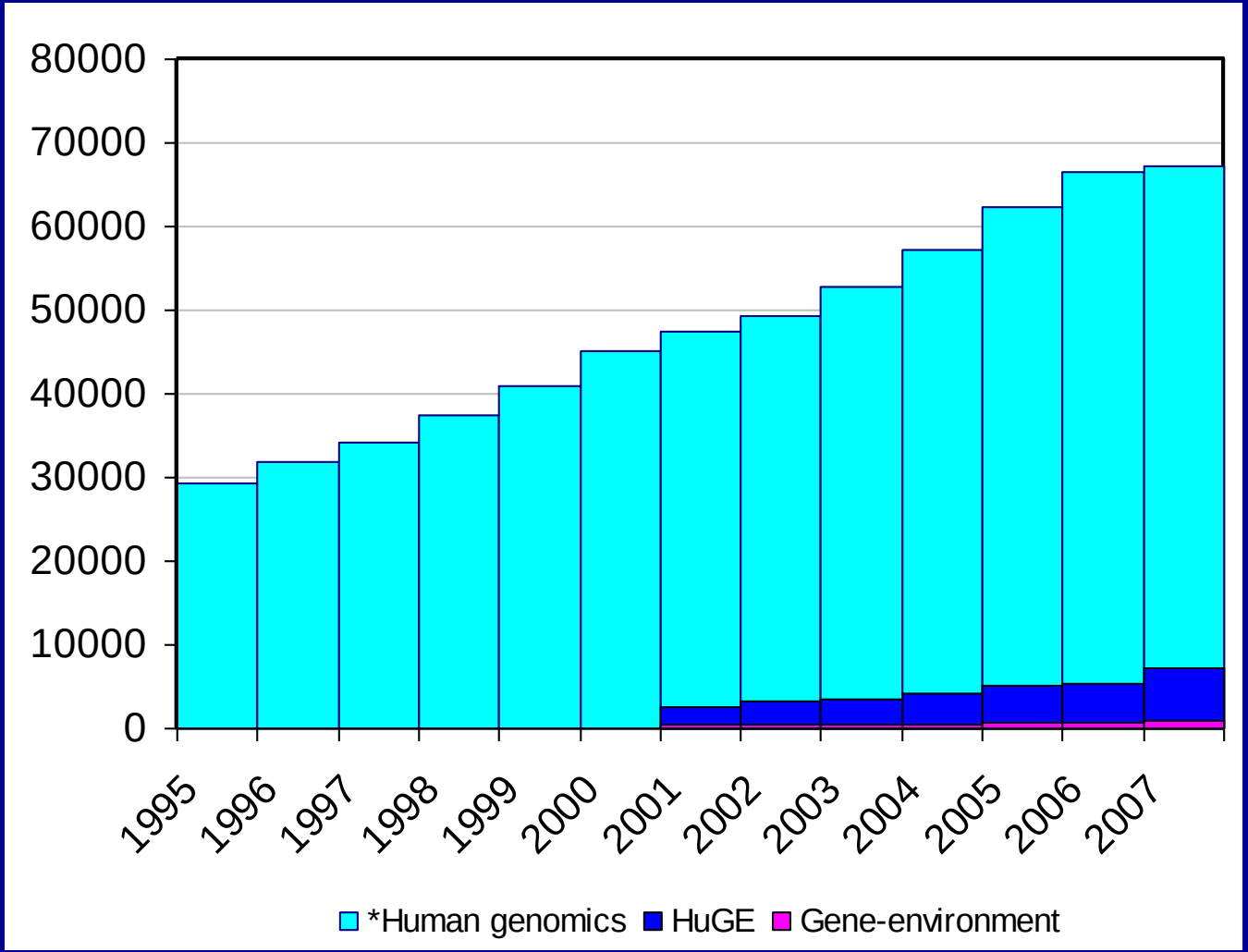
Publication Types:
- Research Support, Non-U.S. Gov't

Gene-environment interaction in atherosclerosis



Sing CF, Stengard JH, Kardia SL. Genes, environment, and cardiovascular disease. *Arterioscler Thromb Vasc Biol 2003;23:1190-6.*

PubMed citations on human genetics/genomics and human genome epidemiology (HuGE)



*PubMed query: “gene or genetic or genome or genomic” / limits: human, year and HuGE Navigator, Oct 2008

Navigating the Epidemiology of the Human Genome

CCGTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACGT

Size up the domain

- population-based research studies
- document data sources, define populations, measure exposures

Map the topography

- population prevalence, associations, interactions, genetic tests
- use data repositories, controlled vocabularies, standard reporting

Identify major features

- consistent findings, patterns, and gaps
- promote knowledge synthesis

Develop navigational tools for further exploration

- reliable knowledge base with applications
- Web 2.0: interconnected, interactive Web-delivered information

Acknowledgements

CCGTCGACTGGAGTGTCTGTGAATTGACTTTTGTTGCCAGTTGGCAGCGGCAGAAGCAGCAAAGCCCGGCCAACAGCAACAAGCTCCTGCCAGATCCCAAAAGCAAACACGT

National Office of Public Health Genomics, CDC

- Muin Khoury, Director
- Wei Yu, HuGE Navigator architect
- Mindy Clyne, curator since 2001

HuGENet Coordinating Centers

- University of Ottawa
- University of Ioannina
- University of Cambridge

Special thanks to NCBI, National Library of Medicine