# SNPedia

## An Applied BioWiki

Mike Cariaso
cariaso@SNPedia.com

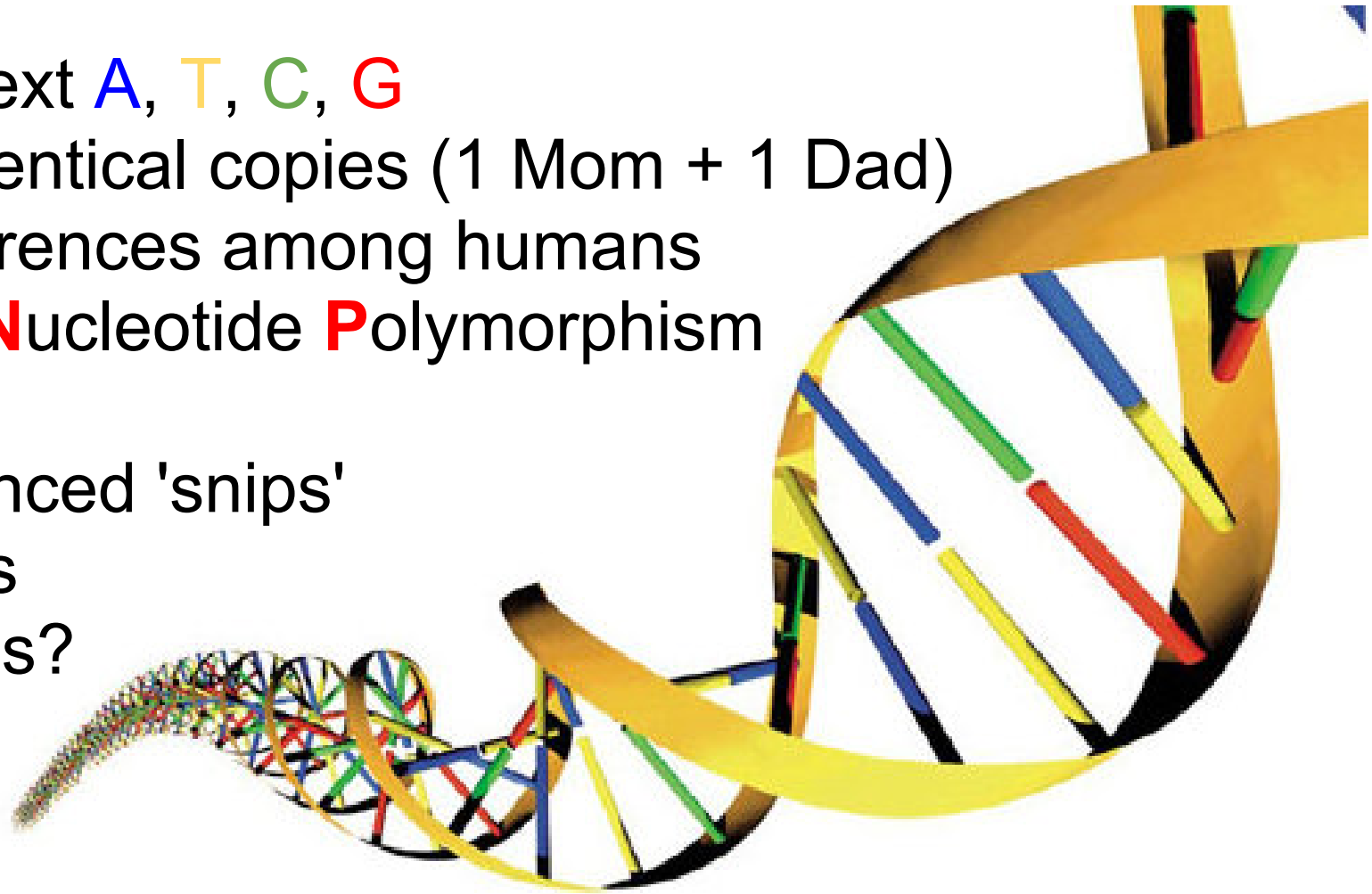
# Bioinformatics Résumé

- Computer programming applied to Biology
- 1997 Human Genome
  - LLNL.gov, Gene Logic, Celera
- 2004 Bacteria - Washington DC Anthrax
  - FBI - SAIC
  - US Navy - BioTeam.net
  - CSHL  454+GA2 Sequencing
- 2008 Plants - Netherlands

# DNA = Genome

## The source code for your body

- ~3 GB of text A, T, C, G
- 2 nearly identical copies (1 Mom + 1 Dad)
- ~20M differences among humans
  - **S**ingle **N**ucleotide **P**olymorphism
  - SNP
  - Pronounced 'snips'
- Differences
  - Features?
  - Bugs?



# Chromosome 22  Position 40,856,638

...TCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGT
CCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTA
GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCC
TGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGC
TGCACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTG
GGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACT
GCTTCGACCAGGTGAGGGAGGAGGTCCTGGAGGGCGGCAG
AGGTCCTGAGGATGCCCCACCAC...

# Usually C but ...

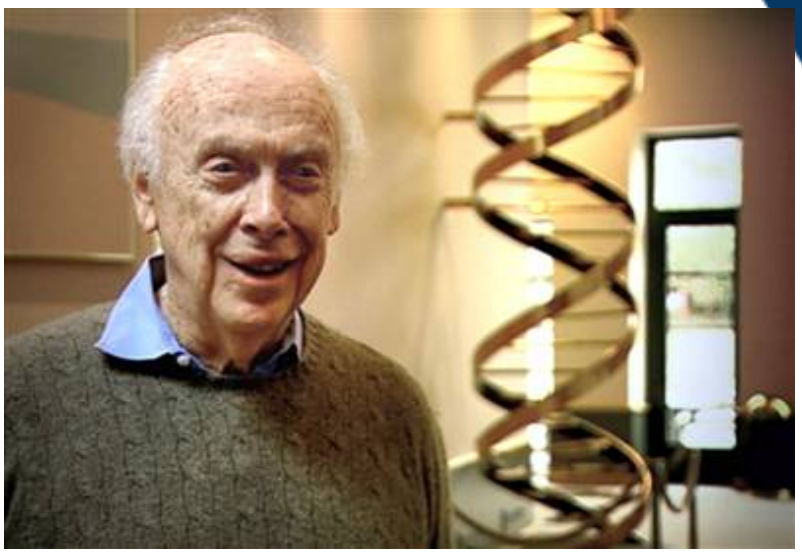
```
...TCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGT
CCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTA
GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCC
TGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGC
TGCACGCTACCCACCAGGCCCCCTGCCACTGCCCGGGCTG
GGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACT
GCTTCGACCAGGTG   GAGGAGGTCCTGGAGGGCGGCAG
AGGTCCTGAGGATGC   CCAC...
```

SNP

rs1065852

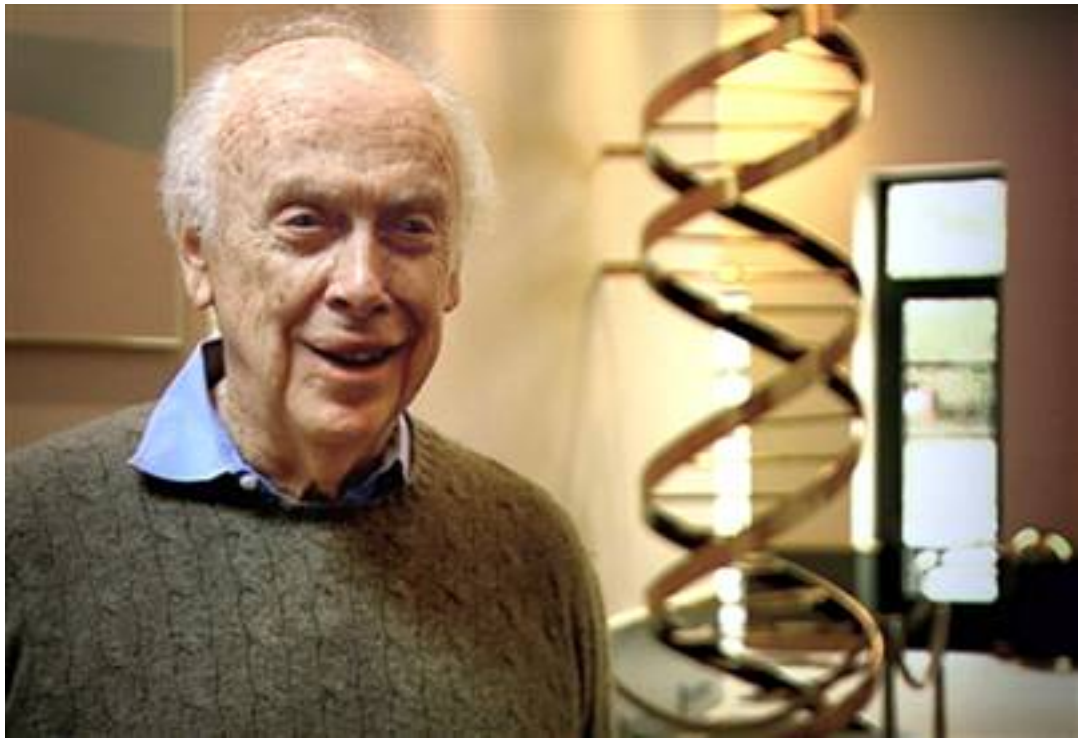
# James Watson has T

```
...TCCTCTGTGCCTGGTGGGGTGGGGGTGCCAGGTGTGT
CCAGAGGAGCCCATTTGGTAGTGAGGCAGGTATGGGGCTA
GAAGCACTGGTGCCCCTGGCCGTGATAGTGGCCATCTTCC
TGCTCCTGGTGGACCTGATGCACCGGCGCCAACGCTGGGC
TGCACGCTACTCACCAGGCCCCCTGCCACTGCCCGGGCTG
GGCAACCTGCTGCATGTGGACTTCCAGAACACACCATACT
GCTTCGACCAGGTG      GAGGAGGTCCTGGAGGGCGGCAG
AGGTCCTGAGGATGC     CCAC...
```



SNP
rs1065852

# James Watson



"I metabolize β-blockers much more slowly than most other Caucasians. Before I had this knowledge, my use of β-blockers to control my blood pressure caused me to constantly fall asleep at inappropriate moments.

Instead of a daily pill, I now take one every week."

- source

SNPedia

navigation

- SNPedia
- FAQ
- Blog
- Recent changes
- Random page

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Browse properties

# Rs1065852

The wild type (normal) allele at this SNP is (C). The (T) variant indicates the presence of a non-wild type CYP2D6 variant, but it appears in many different variants so it can not be used to determine the presence of any particular variant. The most common variants it appears in are CYP2D6*10 and CYP2D6*4, but these are not the only ones. While this is the defining mutation for CYP2D6*10, it is not possible by itself for this mutation to determine a particular variant. In addition, the CYP2D6*4 variant includes this mutation, but the defining mutation 1846G>A SNP is not available in the common direct to consumer testing services.

While it is not possible for this mutation to identify one particular variant, all the variants in which it appears have reduced or no CYP2D6 activity.

If two copies of this (or similar) changes are inherited, poor metabolism ('PM') of debrisoquine [PMID 2211621] is observed.

Other drugs metabolized by CYP2D6 include dextromorphan, sparteine, nortriptyline, and codeine.

Nakamura et al [PMID 12051754] suggest that thermal instabilities and reduced intrinsic clearance by the protein encoded by the rs1065852(T) allele are the main reasons Orientals show lower metabolic activities than Caucasians for drugs metabolized mainly by CYP2D6, since this (T) allele occurs in higher frequency in Orientals.

It is also suggested that poor metabolizers of debrisoquine will be poor metabolizers of metoprolol, diltiazem (brand name cardizem), and propafenone. [PMID 3437726]

- related to DEBRISOQUINE, POOR METABOLISM OF according to omim 124030.0005

CYP2D6 drug metabolism

| | |
|---|---|
| is a | snp |
| is | mentioned by |
| dbSNP | rs1065852 |
| hapmap | rs1065852 |
| hgdp | rs1065852 |
| ensembl | rs1065852 |
| gopubmed | rs1065852 |
| scholar | rs1065852 |
| google | rs1065852 |
| pharmgkb | rs1065852 |
| hgvbaseg2p | rs1065852 |
| medrefsnp | rs1065852 |
| 23andMe | rs1065852 |
| SNP Nexus | rs1065852 |
| Gene | CYP2D6 |
| Chromosome | 22 |
| Orientation | minus |
| Position | 42526694 |
| Reference | GRCh37 37.1/131 |

| Genotype | Effect |
|---|---|
| rs1065852(C;C) | normal |
| rs1065852(C;T) | Carrier of one CYP2D6 decreased or non-functioning variant. |
| rs1065852(T;T) | Homozygous for CYP2D6 decreased or non-functioning variants. |

1: J Biol Chem. 1990 Oct 5;265(28):17209-14.



# Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. S significance of individual mutations by expression of chimeric genes.

Kagimoto M, Heim M, Kagimoto K, Zeugin T, Meyer UA.

Department of Pharmacology, University of Basel, Switzerland.

The debrisoquine/sparteine-type polymorphism is a clinically important inherited variation of drug metabolism characterized by two phen and the poor metabolizer (PM). Five to 10 percent of individuals in Caucasian populations are of the PM phenotype and have deficient m 25 other drugs. Our previous studies have revealed absence of cytochrome P450IID6 protein and aberrant splicing of IID6 premRNA in l mutant alleles of the P450IID6 gene locus (CYP2D6) were identified by restriction fragment length analysis to be associated with the PM mutations of the CYP2D6 gene causing absent P450IID6 protein have not been defined. Here we report the cloning and sequencing of tw CYP2D6 isolated from genomic libraries of three PM individuals. One allele (29-A) was characterized by a single nucleotide deletion in t frameshift and was observed in one individual only. The other type of mutant allele (29-B) was present in all three PM individuals and its mutations, notably four base changes causing amino acid changes in exons 1, 2 and 9, and a point mutation at the consensus sequence of understand the significance of the individual mutations, chimeric genes were constructed between the wild-type IID6 gene and the mutant mutations were introduced into the IID6-cDNA and these DNA constructs were transiently expressed in COS-1 cells. The mutations in e

# Origins of SNPedia

Greg Lennon

PhD Genetics

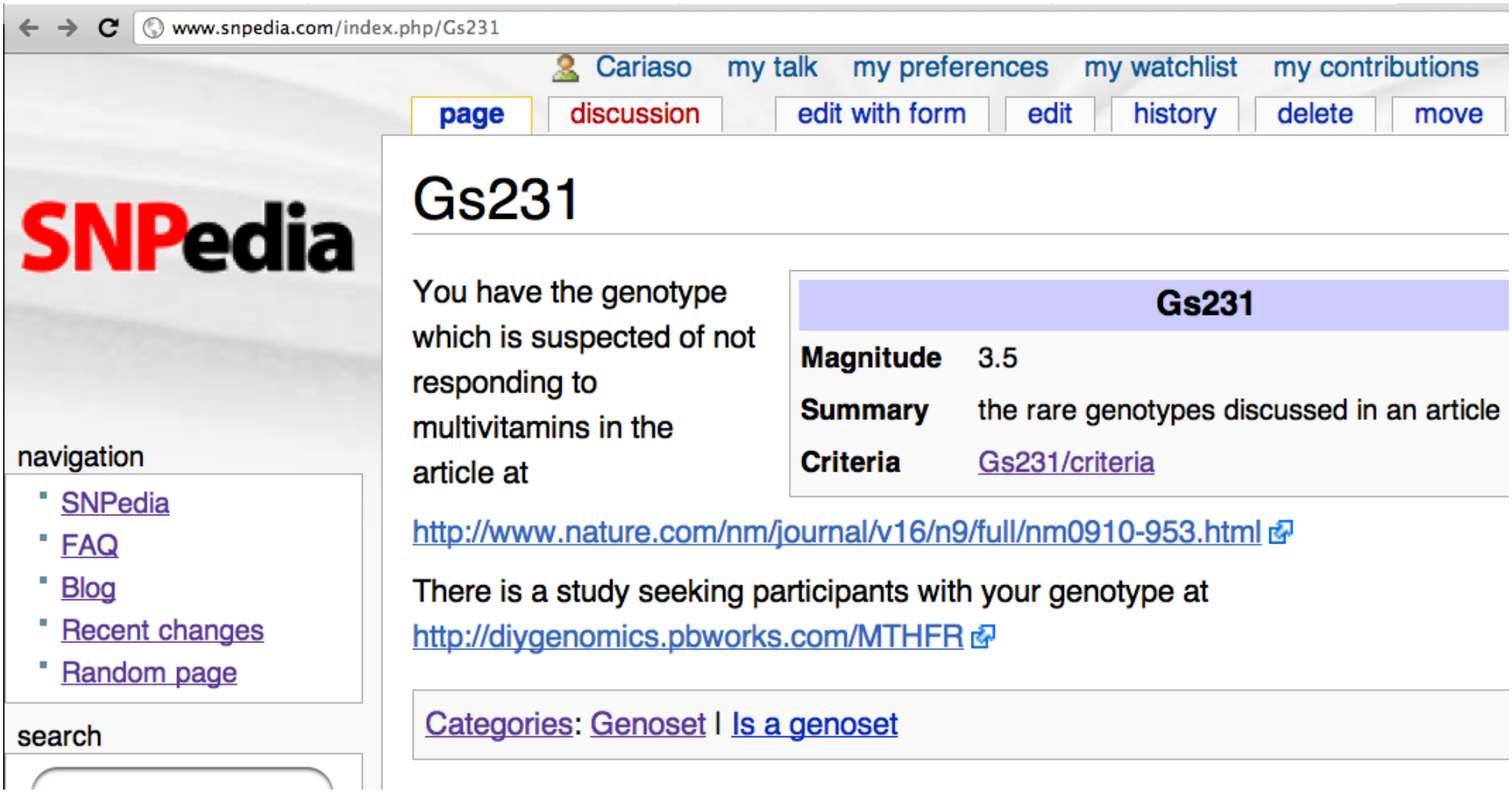
- Read our DNA in May 2007
- How to interpret?
  - Need a program
  - Need a database
- Promethease
  - is the program
- SNPedia
  - is the database

# Notable SNPs

- rs1815739 Fast vs slow twitch muscle formation
- rs333 HIV resistance
- rs601338 Norovirus resistance
- rs6265 Learning

# SNP Combinations

- gs121 50% chance of Type-1 Diabetes
- gs152 FDA Warning: Plavix ineffective for you
- gs116 Unable to taste bitterness
- gs212 Maple Syrup Urine Disease
- gs162 CYP2C9 poor metabolizers of Ibuprofen, Viagra, Tamoxifen, + 20% of all medicines


www.snpedia.com/index.php/Gs231
Cariaso
my talk
my preferences
my watchlist
my contributions
page
discussion
edit with form
edit
history
delete
move
SNPedia
navigation
SNPedia
FAQ
Blog
Recent changes
Random page
search
Gs231
You have the genotype which is suspected of not responding to multivitamins in the article at
Gs231
Magnitude 3.5
Summary the rare genotypes discussed in an article
Criteria Gs231/criteria
http://www.nature.com/nm/journal/v16/n9/full/nm0910-953.html
There is a study seeking participants with your genotype at
http://diygenomics.pbworks.com/MTHFR
Categories: Genoset | Is a genoset


www.snpedia.com/index.php/Gs231/criteria
Cariaso
my talk
my preferences
my watchlist
my contril
page
discussion
edit
hist
SNPedia
Gs231/criteria
< Gs231
and(rs1801131(C;C),
rs1801133(C;C))
navigation
SNPedia
FAQ
Blog
Recent changes

# SNPs? ... InDels, CNVs, Epigenetics!

- rs332 is a deletion of three nucleotides with 2 different insertion texts.
- dbSNP handles small indels
  - so does SNPedia
  - dbVAR is new, for larger structural variants

# SNPedia needs named features

- not positions

We also handle

- rs15793179 animal snps
- SL10030_490 plant snps
- i3003626 23andMe's internal identifiers

# FAQ 2.1

```
1.  use MediaWiki::Bot;
2.  my $bot = MediaWiki::Bot->new({
3.        host        => 'www.snpedia.com',
4.        login_data  => { username => "MyBot",
5.                         password => "password" },
6.                       });
7.
8.  my @pagenames = $bot->get_pages_in_category('Category:Is_a_snp');
9.  foreach my $pagename (@pagenames) {
10.       my $text = $bot->get_text($pagename);
11.       print "\n\n$pagename\n$text\n";
12.       my $newtext = $text . "{{Thing}}\n";
13.       $bot->edit($pagename, $newtext);
14. }
```

# Bots do the dirty work

# Gbrowse

GFF

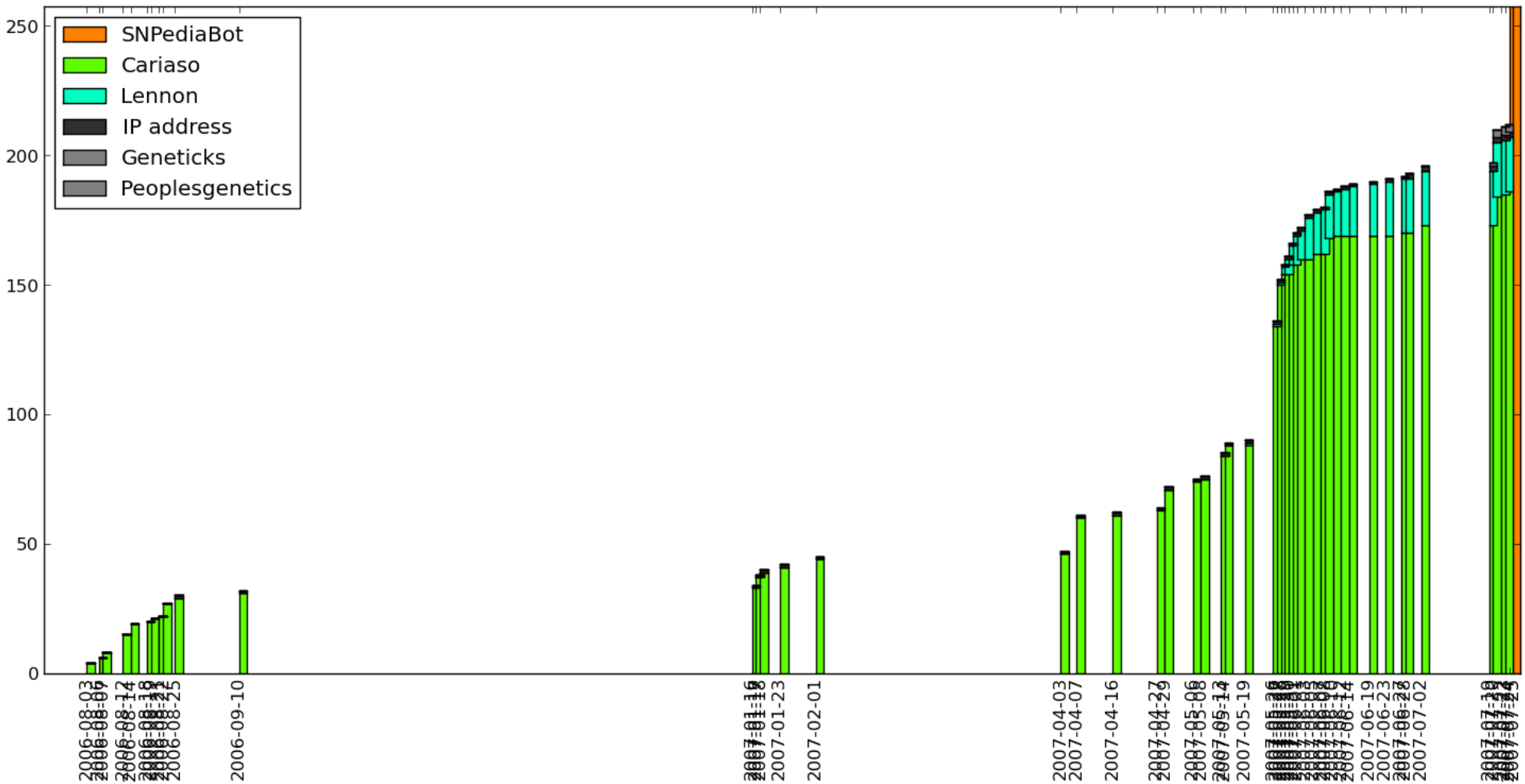

http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/
Overview
chr8
0M 10M 20M 30M 40M 50M 60M 70M 80M 90M 100M 110M 120M 130M 140M
Ideogram
OMIM disease associations
GWA studies (NHGRI Catalog)
Region
chr8
128M 129M 130M 131M
Details
128M 129M 130M 131M
OMIM disease associations
//www.snpedia.com/files/gbrowse/snpedia.gff
snp (illumina)
rs10086908
rs10086908 is a SNP in the 8q24 chromosomal region, which has been linked in several studies to pros...
rs1016342
rs7837328
rs1016343
cancer-genetics these snps influence genetic risk for prostate cancer *the haplotype rs6983267 rs101...
rs1378897
rs6470517
rs891835
[GWAS:Glioma]
rs10092658
[GWAS:Protein quant
rs10808556
[omim:COLORECTAL CANCER, SUSCEPTIBILITY TO, 2; CRCS2]
rs4295627
spittoon rs4295627 G 1.36 Gl
rs2648875
[GWAS:End-stage renal disease]
rs921146
rs1447295
rs1447295 is a SNP on chromosome 8q24, associated with increased risk for prostate cancer in sever
rs7017300
rs7017300 is one of 4 tightly linked SNPs in the 'locus 1' region of 8q24 chromosomal region, wh:

# The Growth of SNPedia

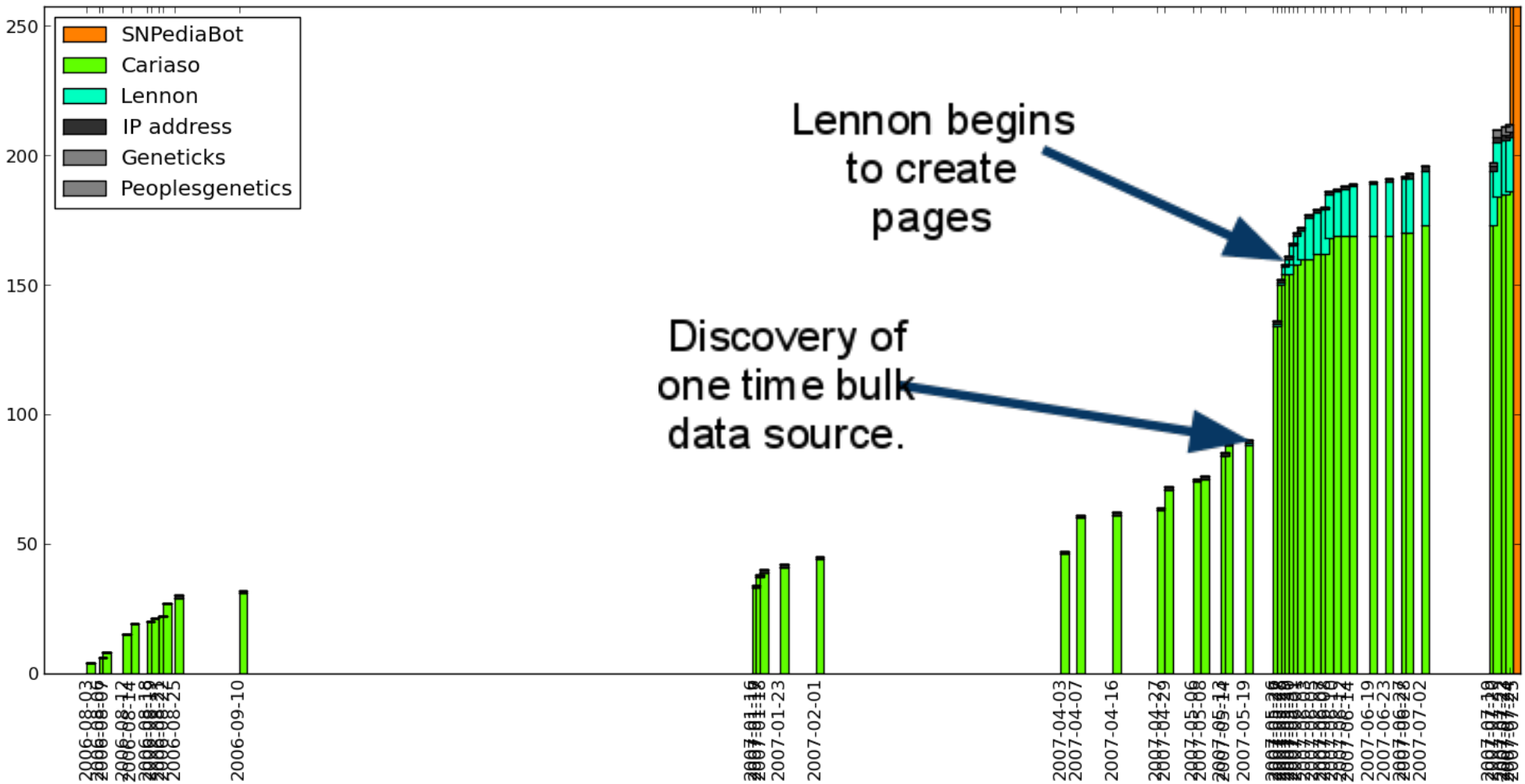
These graphs show who made the FIRST edit on each page.

Growth in the **depth** and **quality** of each page is not visible in these graphs.

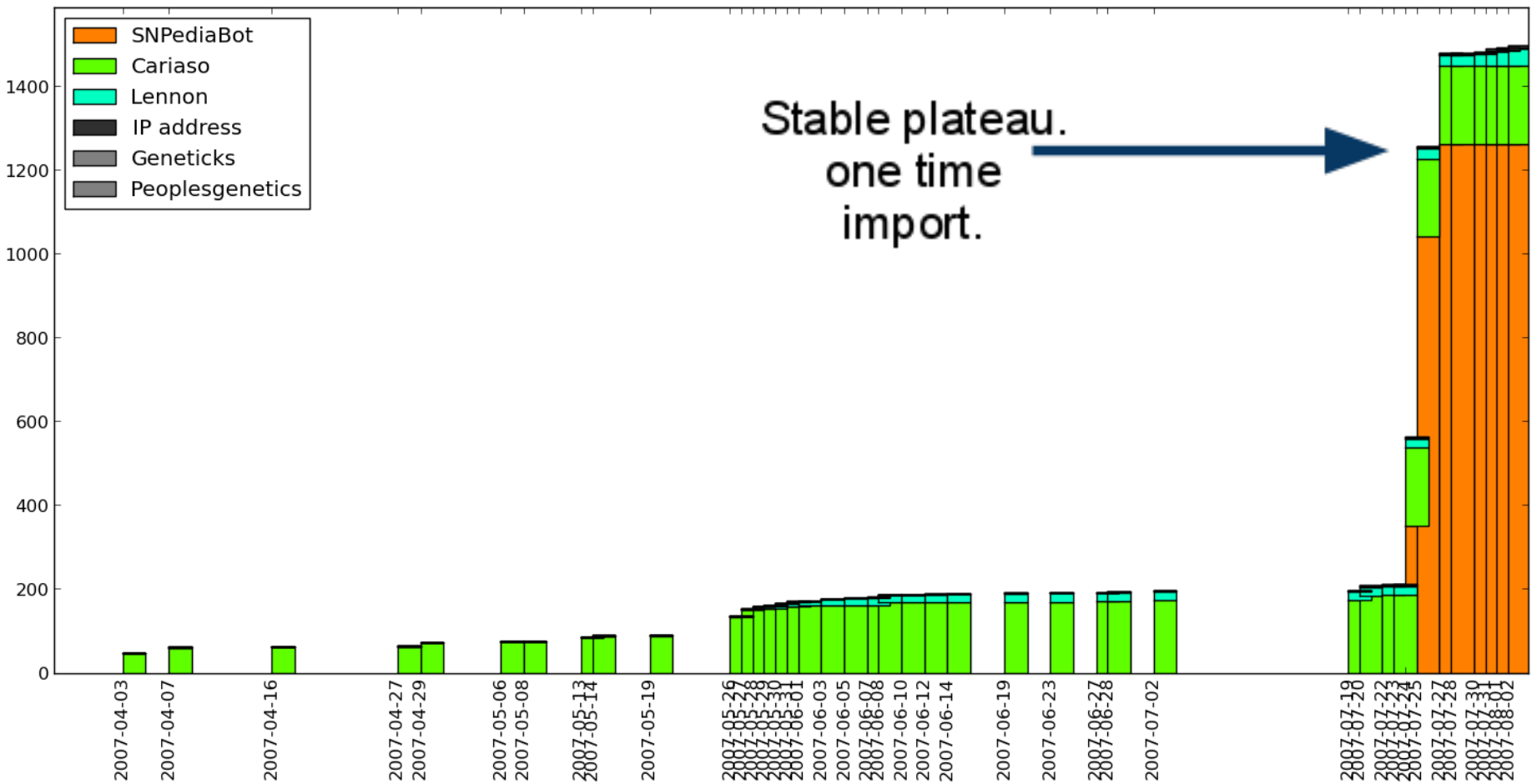
# 1st year, pre-Bot



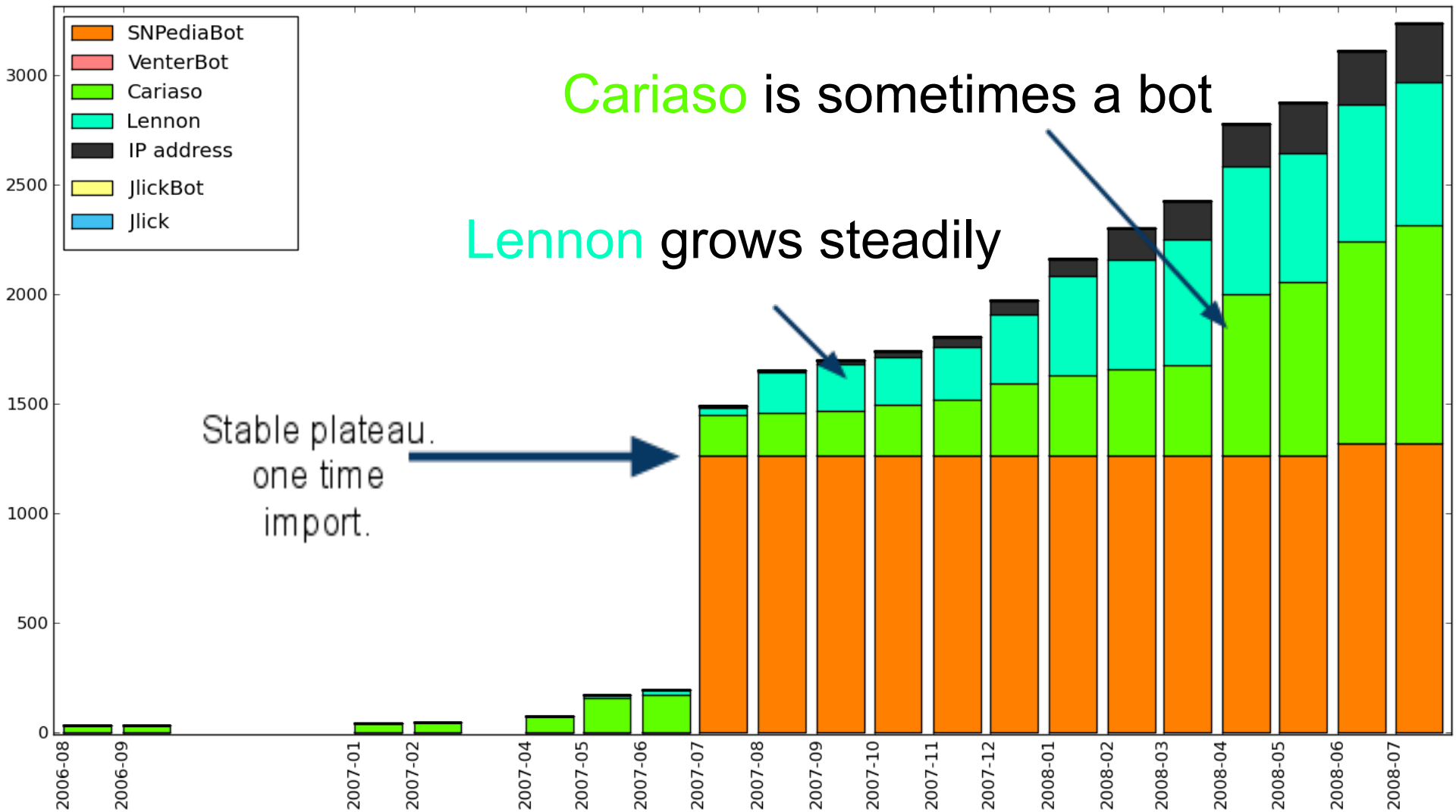
# 1st year, pre-Bot



# Month 13, welcome SNPediaBot



# First 2 years



# First 2 years

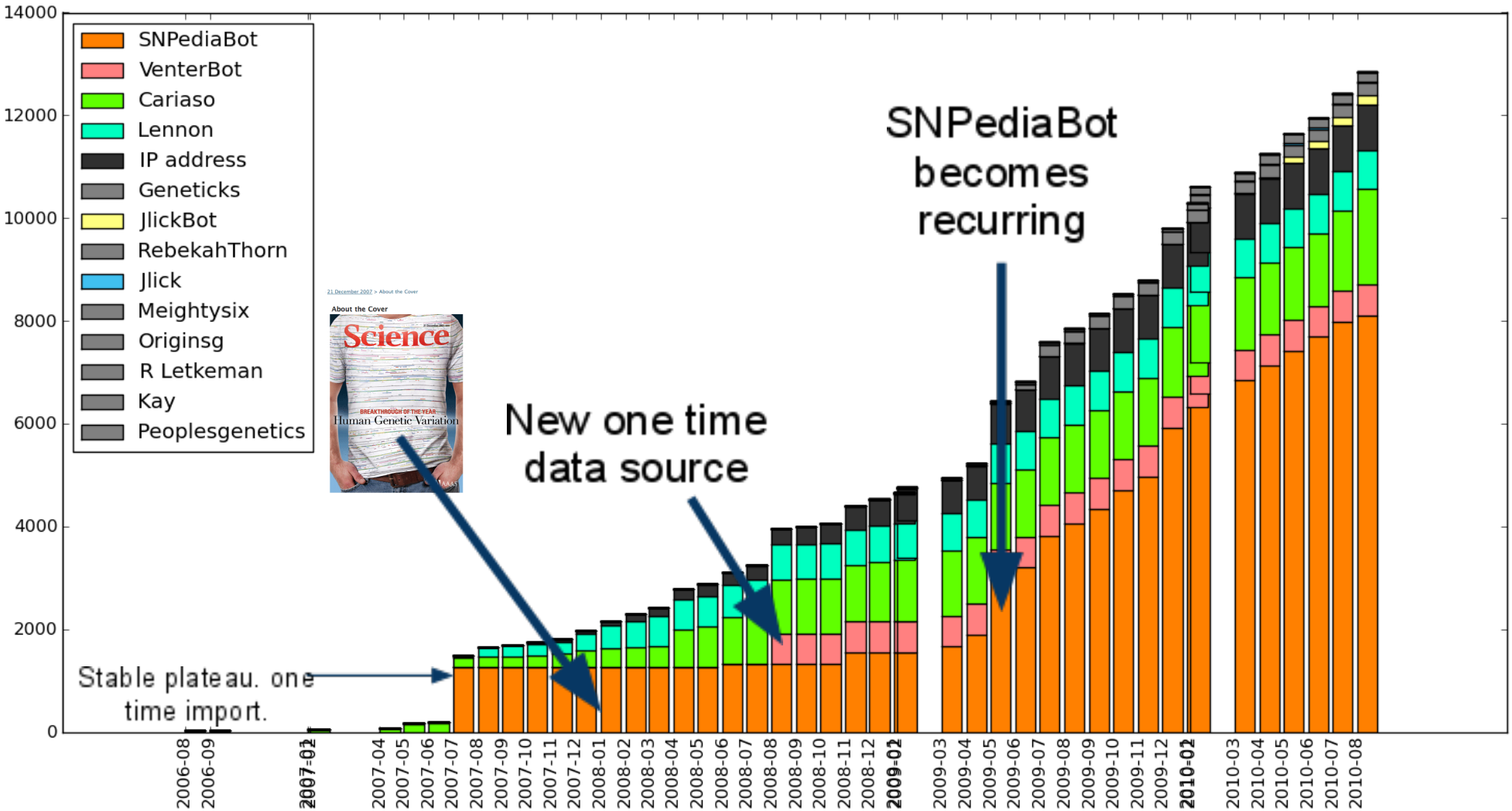
21 December 2007 > About the Cover

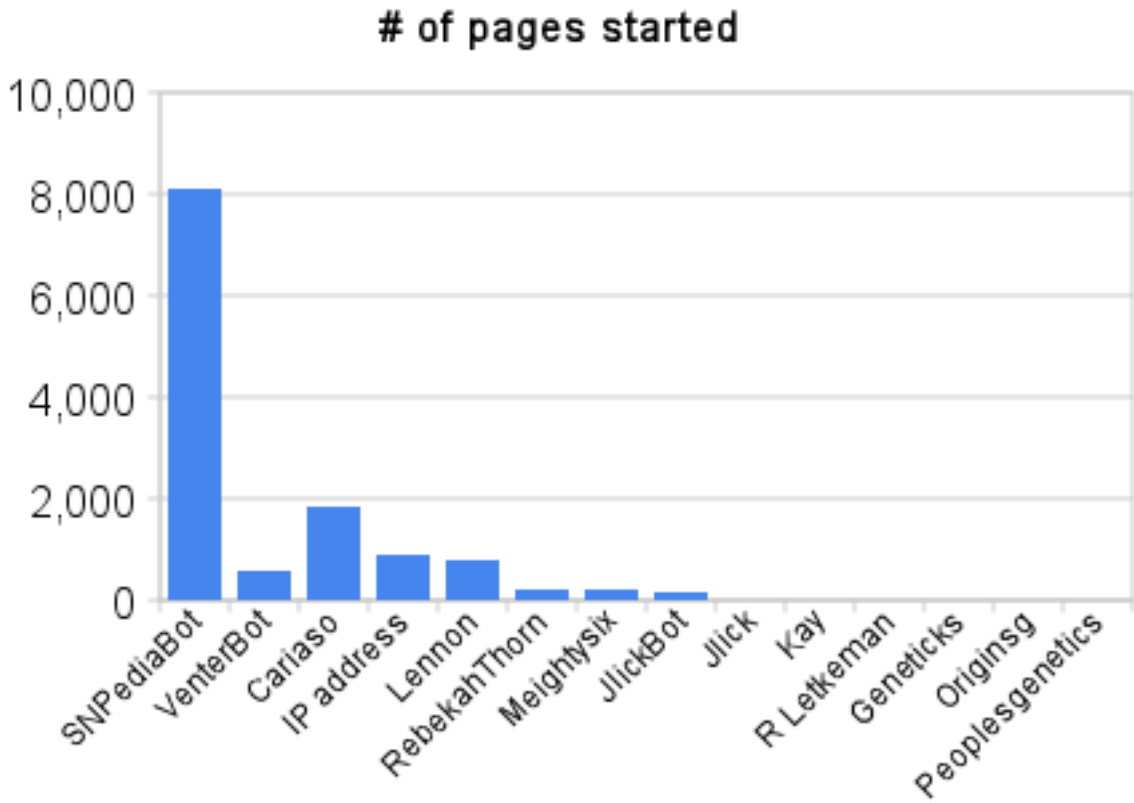
**About the Cover**

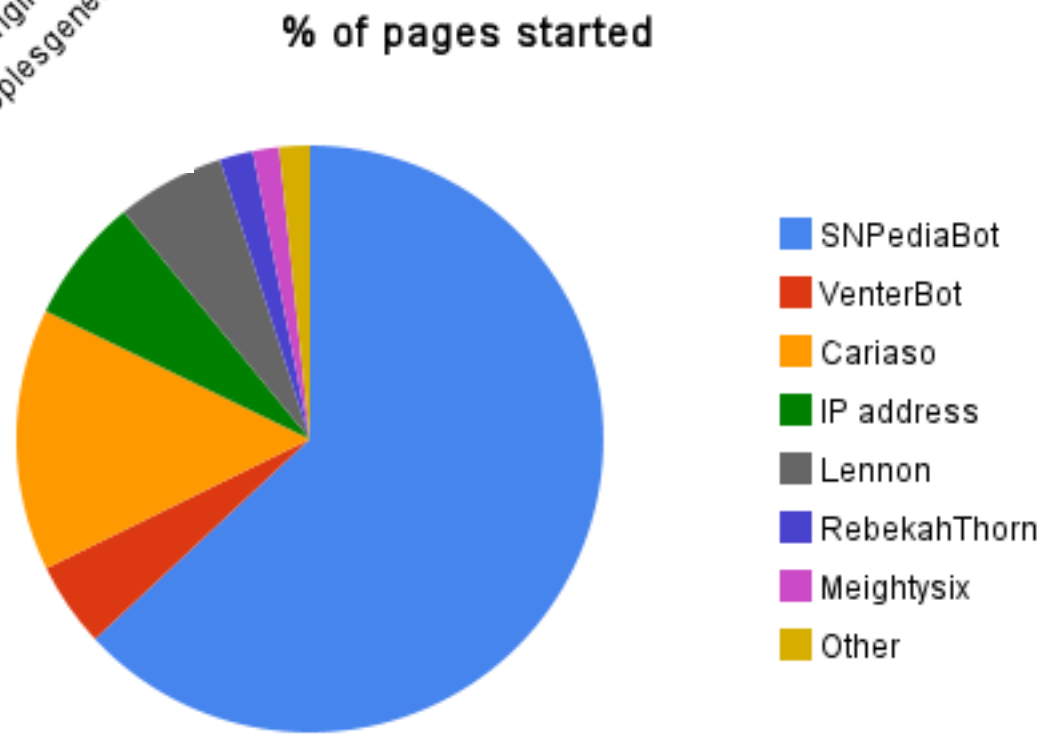

Cariaso is sometimes a bot

ion grows steadily

# All time, by month




# of pages started
10,000
8,000
6,000
4,000
2,000
0
SNPediaBot
VenterBot
Cariaso
IP address
Lennon
RebekahThorn
Meightysix
JlickBot
Jlick
Kay
R Letkeman
Geneticks
Originsg
Peoplesgenetics


% of pages started
SNPediaBot
VenterBot
Cariaso
IP address
Lennon
RebekahThorn
Meightysix
Other

SNPedia

# Active users list

This is a list of users who had some kind of activity within the last 91 days.

Active users list

Display users starting at:

Hide bots Hide administrators

Go

navigation

- SNPedia
- FAQ
- Blog
- Recent changes
- Random page

search

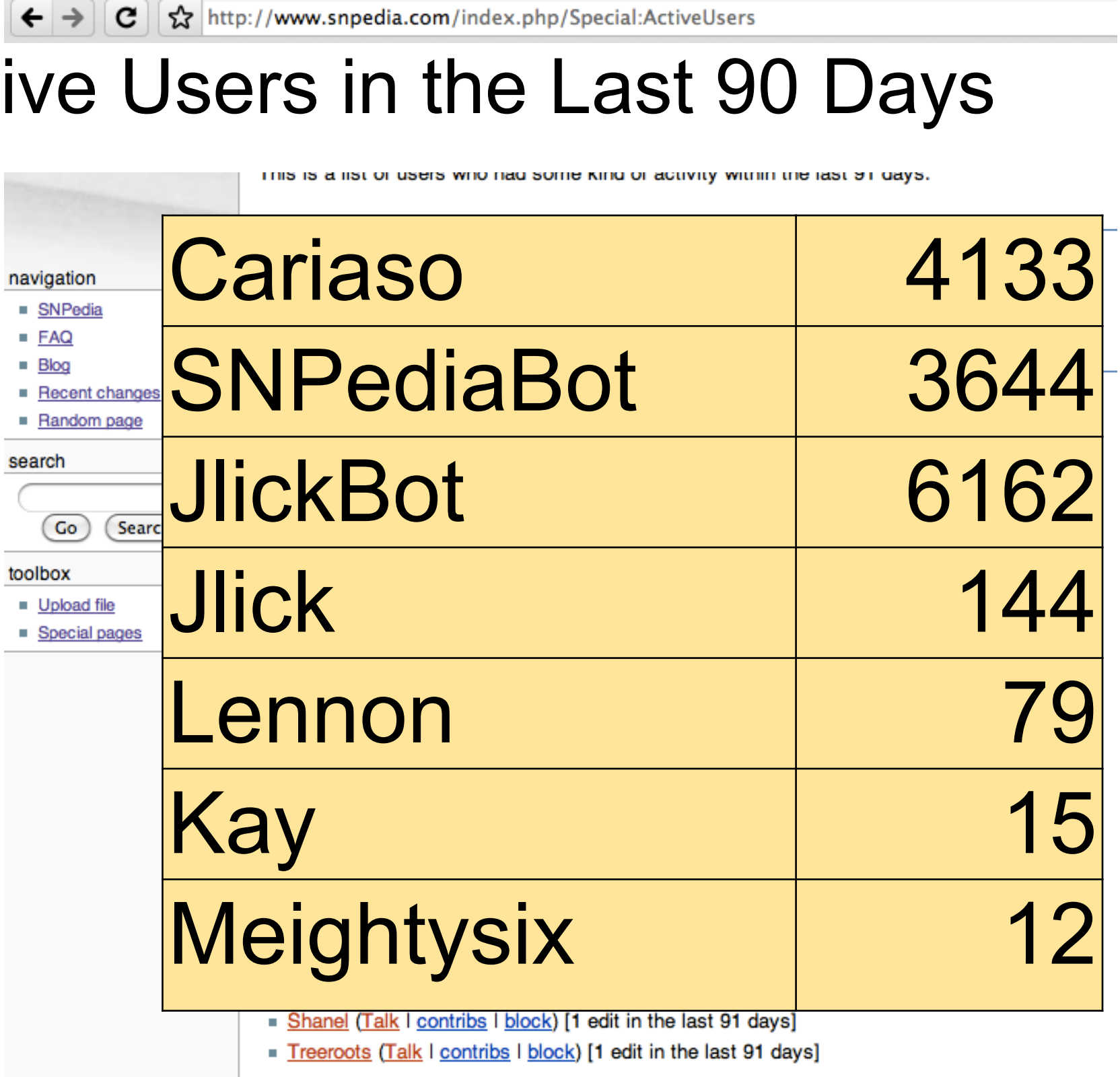
Go Search

toolbox

- Upload file
- Special pages

- [2 edits in the last 91 days]
- [1 edit in the last 91 days]
- Blainebettinger (Talk | contribs | block) [5 edits in the last 91 days]
- [1 edit in the last 91 days]
- Cariaso (Talk | contribs | block) (Bureaucrat, Administrator) [4,133 edits in the last 91 days]
- [1 edit in the last 91 days]
- Elizabeth Craig (Talk | contribs | block) [1 edit in the last 91 days]
- [1 edit in the last 91 days]
- Freeseek (Talk | contribs | block) [3 edits in the last 91 days]
- Jlick (Talk | contribs | block) [144 edits in the last 91 days]
- JlickBot (Talk | contribs | block) (Bot) [6,612 edits in the last 91 days]
- Kay (Talk | contribs | block) [15 edits in the last 91 days]
- Lennon (Talk | contribs | block) (Administrator) [79 edits in the last 91 days]
- Max von Brabant (Talk | contribs | block) [1 edit in the last 91 days]
- Meightysix (Talk | contribs | block) [12 edits in the last 91 days]
- Napobo3 (Talk | contribs | block) [2 edits in the last 91 days]
- Redmed2000 (Talk | contribs | block) [3 edits in the last 91 days]
- SNPediaBot (Talk | contribs | block) (Bot) [3,644 edits in the last 91 days]
- Samsnyder (Talk | contribs | block) [2 edits in the last 91 days]
- Scigenom (Talk | contribs | block) [1 edit in the last 91 days]
- Shanel (Talk | contribs | block) [1 edit in the last 91 days]
- Treeroots (Talk | contribs | block) [1 edit in the last 91 days]

# Active Users in the Last 90 Days




| | |
|---|---|
| Cariaso | 4133 |
| SNPediaBot | 3644 |
| JlickBot | 6162 |
| Jlick | 144 |
| Lennon | 79 |
| Kay | 15 |
| Meightysix | 12 |

# Your raw data costs $500, $250, $48k...$1k

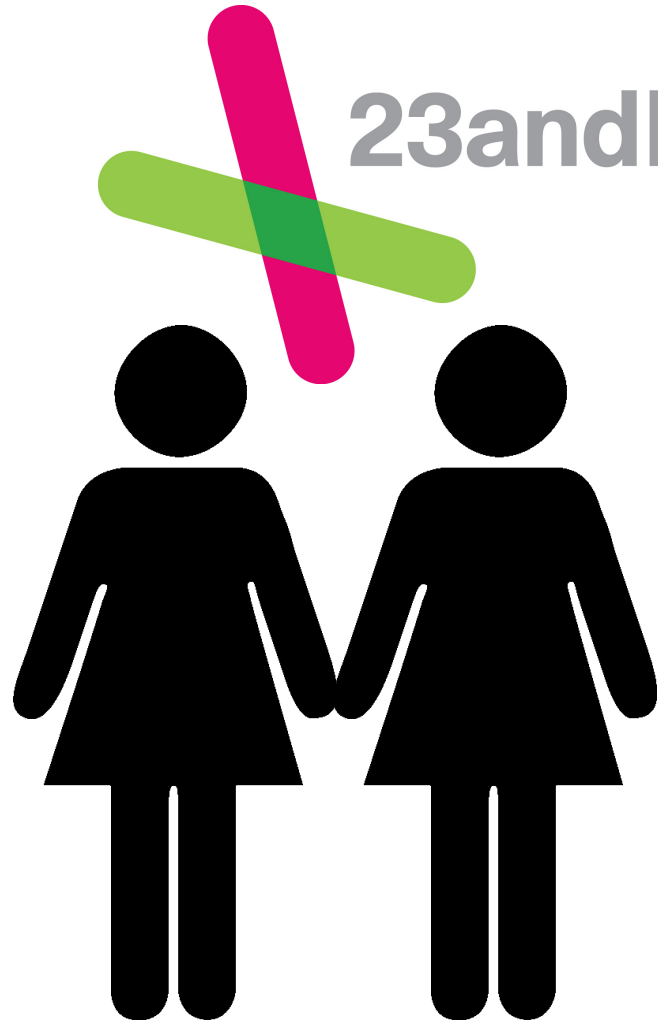
| snp# | chromosome | position | genotype |
|---|---|---|---|
| rs3057 | 8 | 44603386 | GG |
| rs3934834 | 1 | 995669 | CC |
| rs11807848 | 1 | 1051029 | CC |
| rs334 | 11 | 5204808 | TT |
| rs1426654 | 15 | 46213776 | AA |
| rs9651273 | 1 | 1021403 | AG |
| rs1799945 | 6 | 26199157 | GG |

One Million line text file

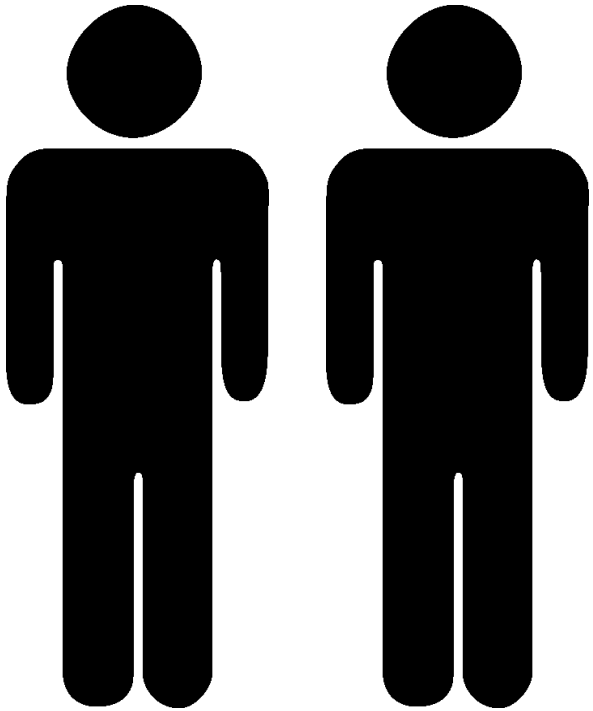
# Your raw data costs $500, $250, $48k...$1k

| snp# | chromosome | position | genotype |
|---|---|---|---|
| rs3057 | 8 | 44603386 | GG |
| rs3934834 | 1 | 995669 | CC |
| rs11807848 | 1 | 1051029 | CC |
| rs334 | 11 | 5204808 | TT |
| rs1426654 | 15 | 46213776 | AA |
| rs9651273 | 1 | 1021403 | AG |
| rs1799945 | 6 | 26199157 | GG |

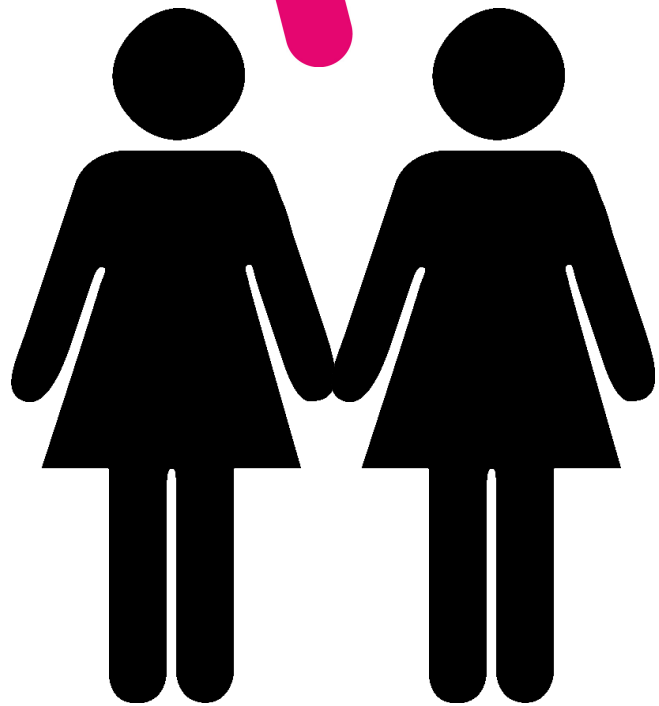
One Million line text file

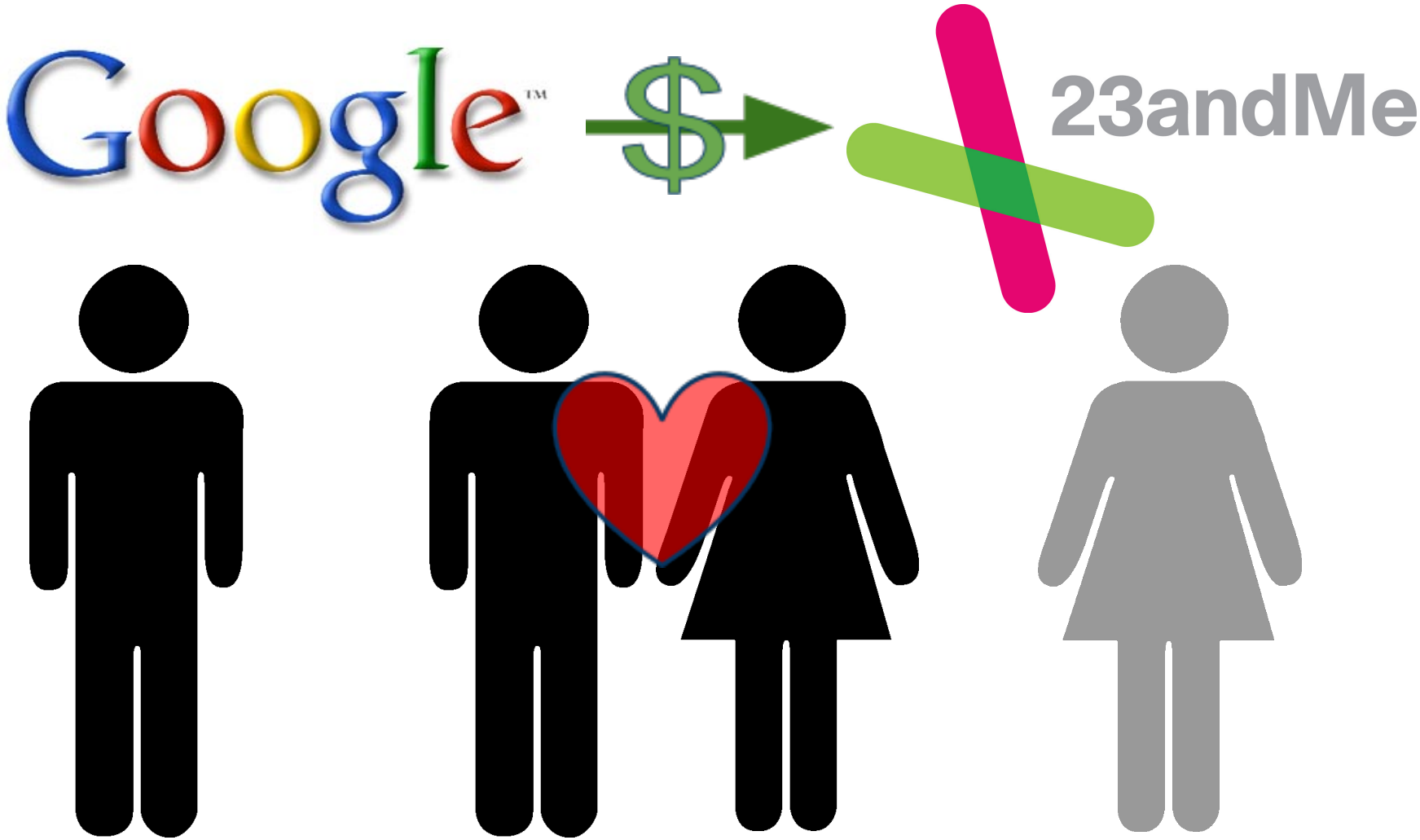

23andMe

Google™



23andMe




Google™
23andMe

# Promethease takes this file

| snp# | chromosome | position | genotype |
|---|---|---|---|
| rs3934834 | 1 | 34628821 | CC |
| rs9651273 | 1 | 39028816 | AG |
| rs334 | 1 | 65330201 | TT |
| rs3057 | 2 | 766134 | GG |
| rs11807848 | 2 | 8115765 | CC |
| rs1799945 | 2 | 10002342 | CG |
| rs1426654 | 2 | 12453229 | AA |

One Million line text file

# Promethease takes this file

| **snp#** | **genotype** |
|---|---|
| rs3934834 | CC |
| rs9651273 | AG |
| rs334 | TT |
| rs3057 | GG |
| rs11807848 | CC |
| rs1799945 | CG |
| rs1426654 | AA |

One Million line text file

# Promethease adds data

| snp# | Population Frequency | genotype |
|---|---|---|
| rs3934834 | 100% | CC |
| rs9651273 | 32% | AG |
| rs334 | 100% | TT |
| rs3057 | 18% | GG |
| rs11807848 | 100% | CC |
| rs1799945 | 9% | CG |
| rs1426654 | ?% | AA |

# Promethease sorts what is common

| snp# | Pop. Freq. | genotype |
|---|---|---|
| rs1799945 | 9% | CG |
| rs3057 | 18% | GG |
| rs9651273 | 32% | AG |
| rs1426654 | ?% | AA |
| rs3934834 | 100% | CC |
| rs334 | 100% | TT |
| rs11807848 | 100% | CC |

# Common SNPs are less interesting

| snp# | Pop. Freq. | genotype |
|---|---|---|
| rs1799945 | 9% | CG |
| rs3057 | 18% | GG |
| rs9651273 | 32% | AG |
| rs1426654 | ?% | AA |
| rs3934834 | 100% | CC |
| rs334 | 100% | TT |
| rs11807848 | 100% | CC |

# Magnitude = Importance

| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs1799945 | 9% | | CG |
| rs3057 | 18% | | GG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | | CC |
| rs334 | 100% | | TT |
| rs11807848 | 100% | | CC |

# Magnitude = 0 = Common = Normal

| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs1799945 | 9% | | CG |
| rs3057 | 18% | | GG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | 0 | CC |
| rs334 | 100% | 0 | TT |
| rs11807848 | 100% | 0 | CC |

# Magnitude = 2 = Hmm, interesting

| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs1799945 | 9% | | CG |
| rs3057 | 18% | 2 | GG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | 0 | CC |
| rs334 | 100% | 0 | TT |
| rs11807848 | 100% | 0 | CC |

# Sort by Magnitude, Pop. Freq

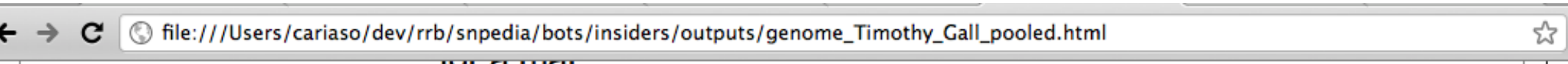
| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs3057 | 18% | 2 | GG |
| rs1799945 | 9% | | CG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | 0 | CC |
| rs334 | 100% | 0 | TT |
| rs11807848 | 100% | 0 | CC |

# Magnitude = 3.1 = More Interesting

| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs3057 | 18% | 2 | GG |
| rs1799945 | 9% | 3.1 | CG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | 0 | CC |
| rs334 | 100% | 0 | TT |
| rs11807848 | 100% | 0 | CC |

# Important stuff at the top

| snp# | Pop. Freq. | Magnitude | genotype |
|---|---|---|---|
| rs1799945 | 9% | 3.1 | CG |
| rs3057 | 18% | 2 | GG |
| rs9651273 | 32% | | AG |
| rs1426654 | ?% | | AA |
| rs3934834 | 100% | 0 | CC |
| rs334 | 100% | 0 | TT |
| rs11807848 | 100% | 0 | CC |




for a trial...

**Help interpreting your results**

**Chat**

Show Everything

| | |
|---|---|
| gs144<br>Magnitude: 4 | Male. |
| gs162<br>Magnitude: 4 | CYP2C9 poor metabolizers require<br>as tamoxifen, warfarin, fluvastin, an<br>inflammatory agents such as aspiri |
| 20.4% rs1333049(C;C)<br>Magnitude: 4<br>ambig | 1.9x increased risk for CAD...more... |
| gs191<br>Magnitude: 3.1 | impaired NSAID drug metabolism, which is a risk factor for gastrointestinal bleeding when taking any of these medications: aceclofenac, celecoxib, diclofenac, ibuprofen, indomethazine, lornoxicam, meloxicam, naproxen, piroxicam, tenoxicam and valdecoxib. You have one of these *CYP2C8*3 (rs11572080 and rs10509681) *CYP2C9*2 (rs1799853) *CYP2C9*3 (rs1057910) |
| gs192<br>Magnitude: 3 | You have 2 variations in MTHFR which influence homocystine levels. people with gs193 are more strongly affected. |
| 10.6% rs7903146(T;T)<br>Magnitude: 3<br>Count: 2 | 2x increased risk for diabetes (and perhaps colon cancer) This genotype greatly increases the risk of Type-2 diabetes. rs7903146(T;T) strongly predicted future type-2 diabetes. ...more... |

file:///Users/cariaso/dev/rrb/snpedia/bots/insiders/outputs/genome_Russell_Letkeman_20081023161035.html

Leprosy...more... 8 snps
Leukemia...more... 8 snps
Liver cancer

(hide)

## Liver cancer

| | | |
|---|---|---|
| 54.9% | rs6983267(G;T) | 1.2x risk of prostate cancer 1.2x risk of prostate cancer...more... |
| | Magnitude: 2.1 | |
| None | rs4444903(A;A) | average...more... |
| 67.3% | rs1800629(G;G) | complex; generally normal risk...more... |
| | Magnitude: 0.1 | |
| 89.4% | rs1800562(G;G) | Not a C282Y hemochromatosis carrier....more... |
| | Magnitude: 0 | |

4 snps
Longevity...more... 35 snps
Lumbar disc disease...more... 2 snps
Lung cancer...more... 36 snps
Lupus...more... 56 snps
Lyme disease...more... 1 snp
Lymphoma...more... 7 snps
Malaria...more... 9 snps

# A Real Report

## These seem to be the most interesting snps.

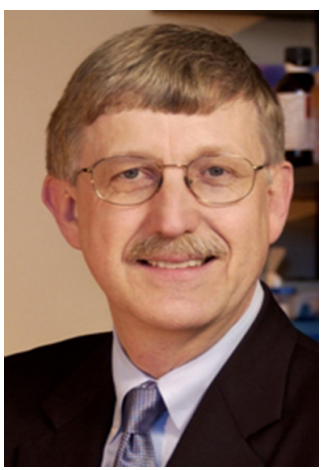
(hide)

| | |
|---|---|
| 0.9% rs4363657(C;C)<br>Magnitude: 4<br>Count: 2 | 17x increased myopathy risk for statin users...more... |
| gs231<br>Magnitude: 3.5 | You have the genotype which is suspected of not responding to multivitamins in the article at http://www.nature.com/nm/journal/v16/n9/full/nm0910-953.html There is a study seeking participants with your genotype at http://diygenomics.pbworks.com/MTHFR |
| gs191<br>Magnitude: 3.1 | impaired NSAID drug metabolism, which is a risk factor for gastrointestinal bleeding when taking any of these medications: aceclofenac, celecoxib, diclofenac, ibuprofen, indomethazine, lornoxicam, meloxicam, naproxen, piroxicam, tenoxicam and |

# Promethease Report

rs2395029 "exciting to see the Abacavir example come forward this past year where hypersensitivity reactions are almost entirely predictable by a genetic test. It was good to see that that was well demonstrated in a carefully controlled clinical trial… [and that] led to a clear change in the FDA labeling."

Francis Collins MD PhD
Director of the NIH



# Questions?

Thank you to

- MediaWiki
- Semantic MediaWiki
- Semantic Forms
- pywikipediabot
- Mediawiki::Bot