Free Factories: from the Quantum Coreworld to the Personal Genome Project

(a tour of Genomics, Computing and Economics topics in my research)

Alexander Wait Zaranek
awaitz@post.harvard.edu

Biophysics 101 seminar
Thursday, September 17th, 2009

Overview

- “Big Data” and Free Factories

Applications of Free Factories:

Swift, Editing, Four Individual Genomes

The Aim of Free Factories:

Enable the Personal Genome Project to bring together physicians, scientists and the general public in a single endeavor

Nucleic acids are the largest repository of digital information on Earth

NCBI Trace Archive - 10^{15}

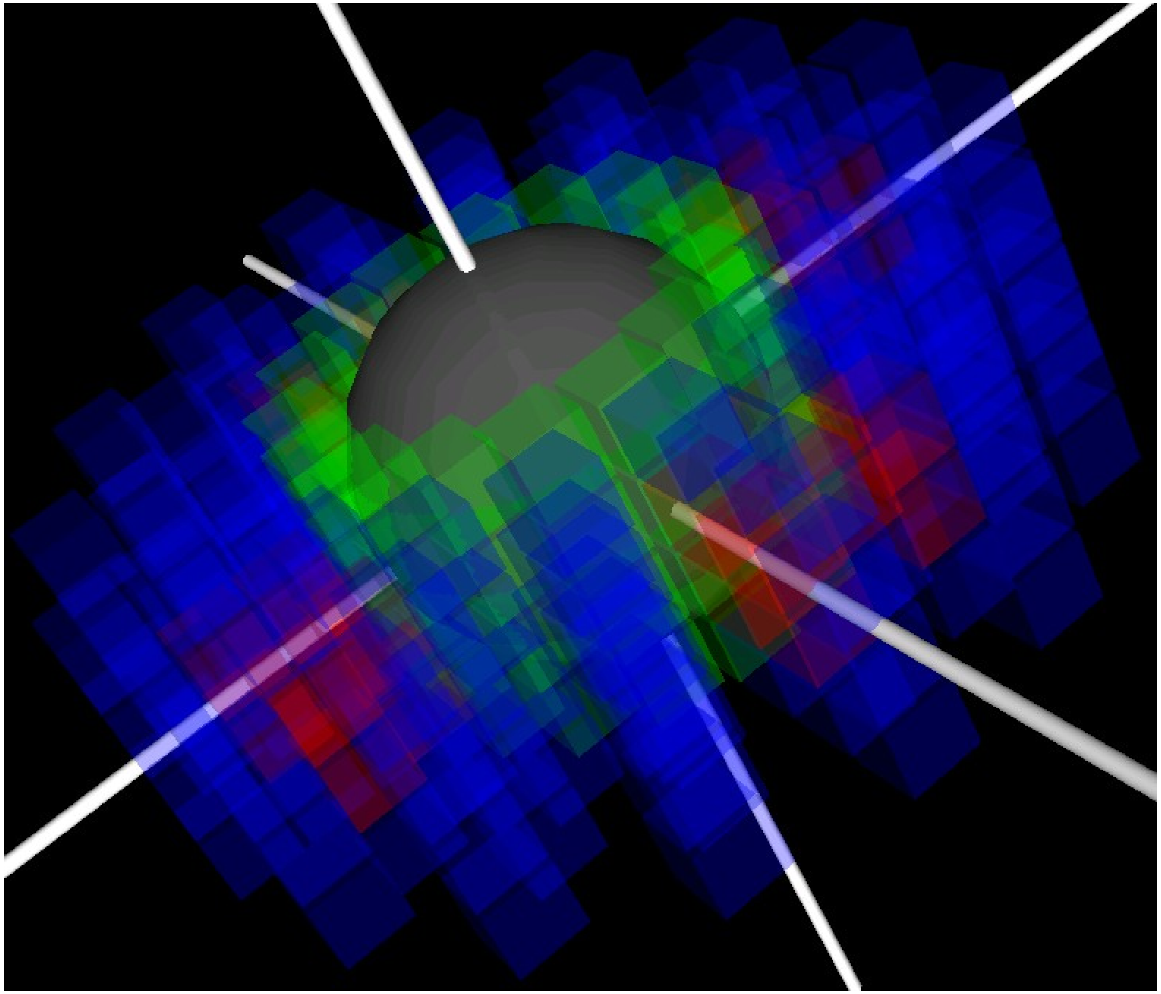
“Ordinary” digital universe - 10^{21}

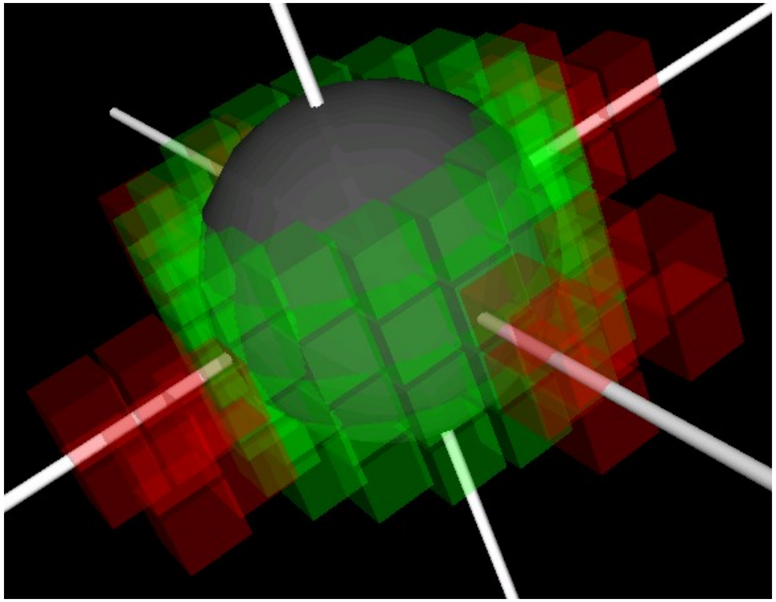
Human associated digital universe - 10^{32}

Biological digital universe - 10^{36}

Pedulla et al. (2003) Cell
http://www.emc.com/digital_universe
http://www.ncbi.nlm.nih.gov/Traces/home/

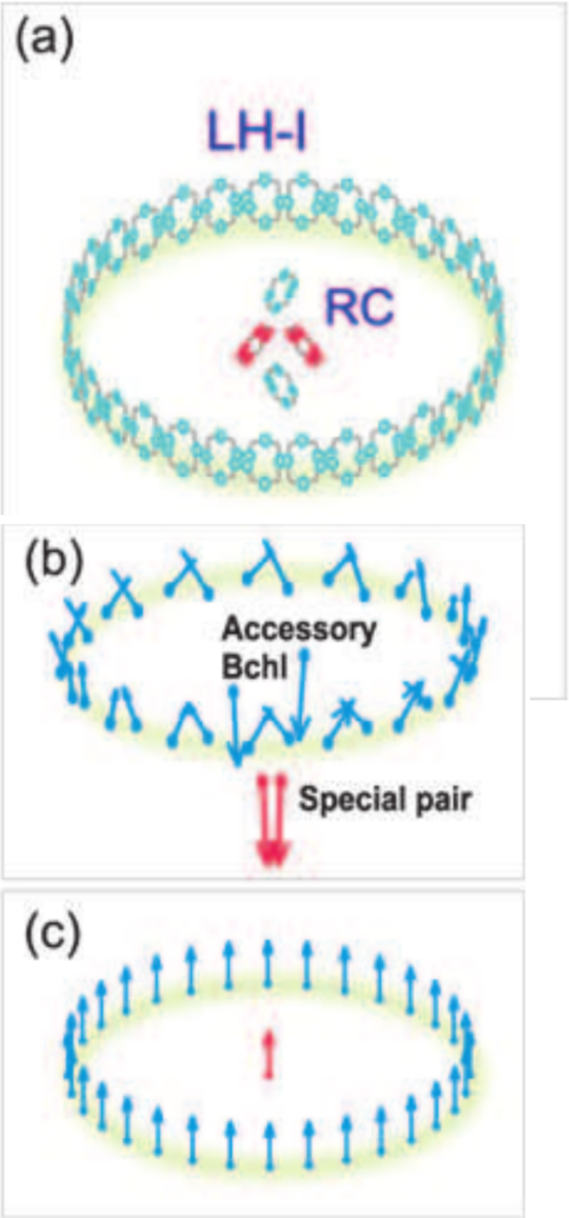
The Quantum Coreworld is a 3D digital evolution system consistent with the rules of quantum information processing



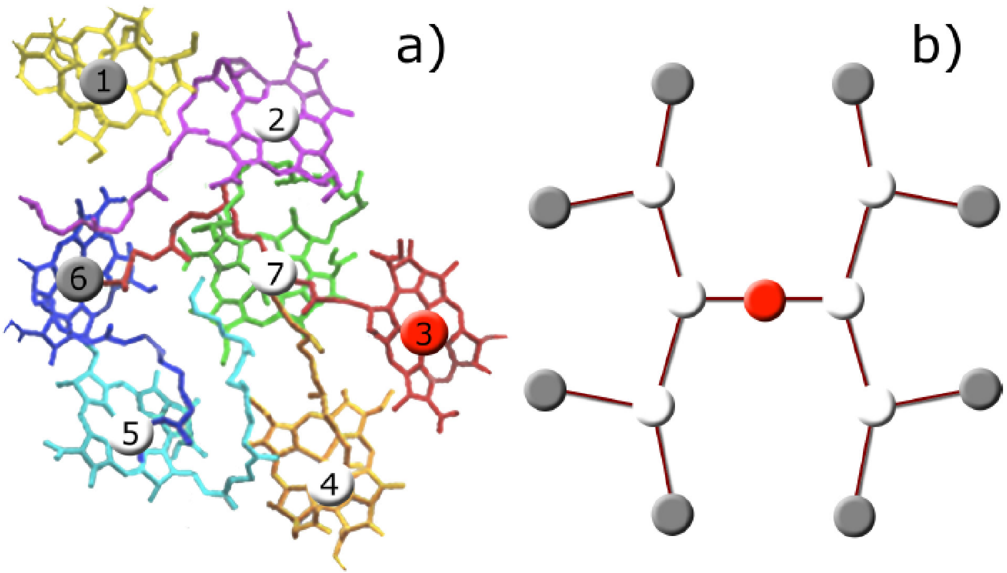


Translucent cubes represent one femtoliter of water (a cube one micron on each side)
Digital organisms—written in a simple assembly language—evolve in the Coreworld
Regions of the world periodically exchange material at well defined boundaries (red)
New digital lifeforms enter the world nearest to the center (green)

Quantum Information in Real Biology

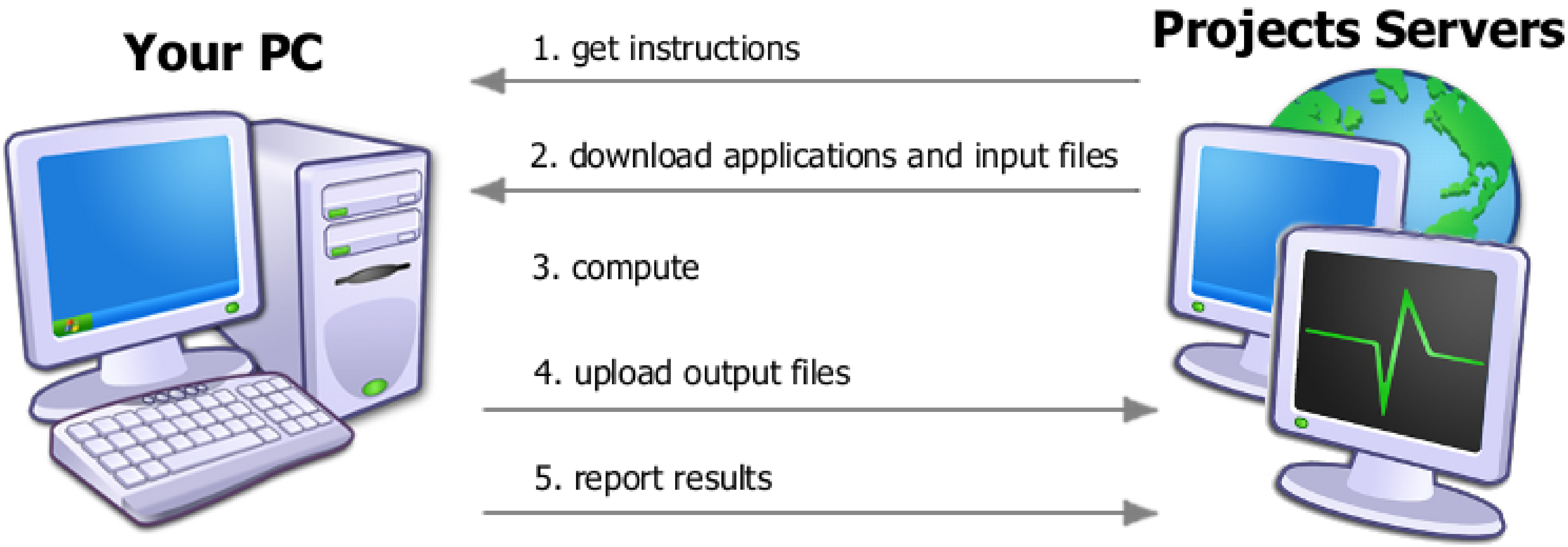


Schematic of the LH1-RC core of purple bacteria *Rhodobacter sphaeroides*. Olaya-Castro (2007) Phys Rev B.



ArXiv:0807.0929v1—Aspuru-Guzik group (Harvard)

Volunteer Computing



Your PC

Projects Servers

1. get instructions
2. download applications and input files
3. compute
4. upload output files
5. report results

BOINC

Active: 335,473 volunteers, 562,937 computers.
24-hour average: 1,309.43 TeraFLOPS.

http://openwetware.org/wiki/PersonalGenomes@Home

A Free Factory is inspired by Free Software

Free Software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software.

(http://www.gnu.org/philosophy/free-sw.html)

A Free Factory should protect the freedom of its user community to:

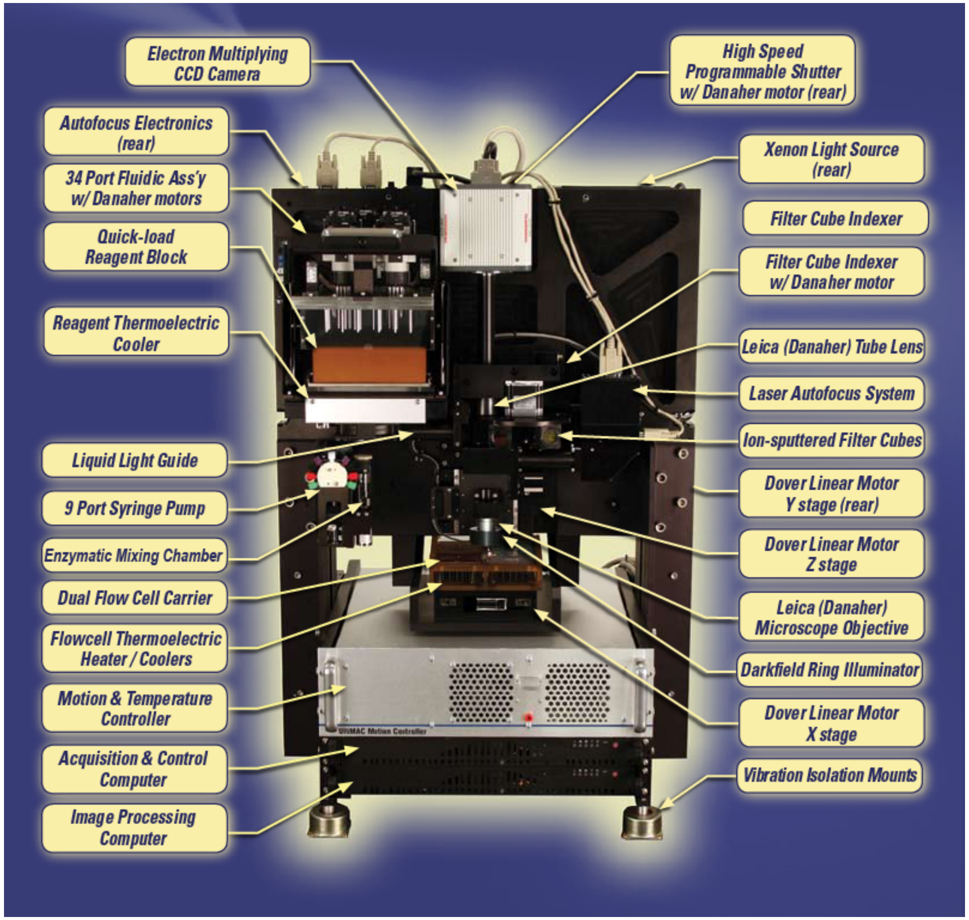
1) operate their own identical factory;

2) operate a modified factory;

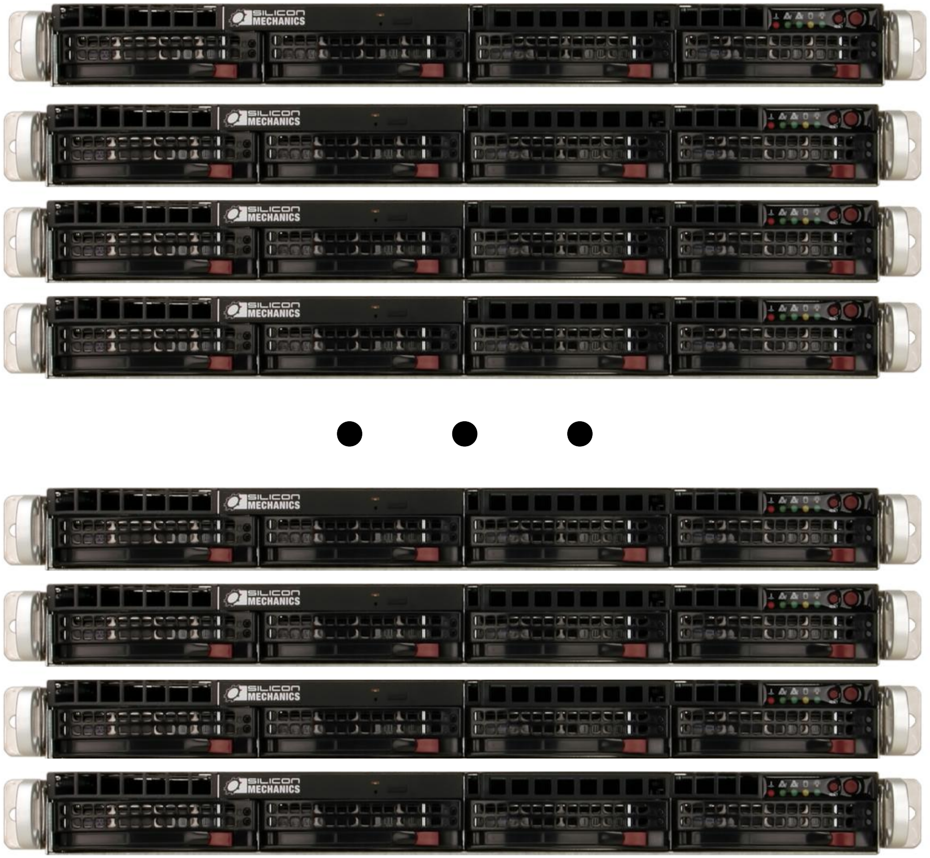
3) distribute the information required to operate and modify the factory to others, and;

4) study and improve all factory equipment, methods, software, raw materials, and so on.

A Free DNA Sequencing Factory could be built by combining the “Polonator” with commodity computers running Free and Open Source Software



Courtesy — Rich Terry and Greg Porreca



I designed the Free Factories computational infrastructure to complement the Polonator.

With sequencing costs falling more rapidly than computing costs, this is a formidable challenge.

A recent budget for a Free (DNA Sequencing) Factory

Setup costs ($490,000 total)
- $150,000 - Polonator
- $80,000 - local storage & computation cluster
- $160,000 - remote storage & computation cluster (deployed in second year)
- $100,000 - cluster maintenance, support, training (2 years, 1/2 FTE)

Operational costs ($181,000 in first year)
- $30,000 - power, cooling, network for Polonator and local cluster
- $40,000 - bioinformatics technician (1/2 FTE)
- $40,000 - lab technician (1/2 FTE)
- $71,000 - reagents ($7,085 per run, w/ 2x13bp reads, 100Gbp)

Operational costs ($315,000 in second year)
- $30,000 - power, cooling, network for Polonator and local cluster
- $30,000 - power, cooling, network for remote cluster
- $80,000 - bioinformatics technician (1 FTE)
- $80,000 - lab technician (1 FTE)
- $95,000 - reagents ($4,769 per run, w/ 2x48bp reads, 740Gbp)

Total costs ($986,000 two years w/ 840Gbp at $1174 per gigabase.)

In this budget, computing costs are 38% of the total while they were only 11% for a similar project last year.

To realize the benefits of cheap sequencing, free and open source methodologies help us consider the whole ecosystem of scientists, physicians and the general public as a single system.

Overview

"Big Data" and Free Factories

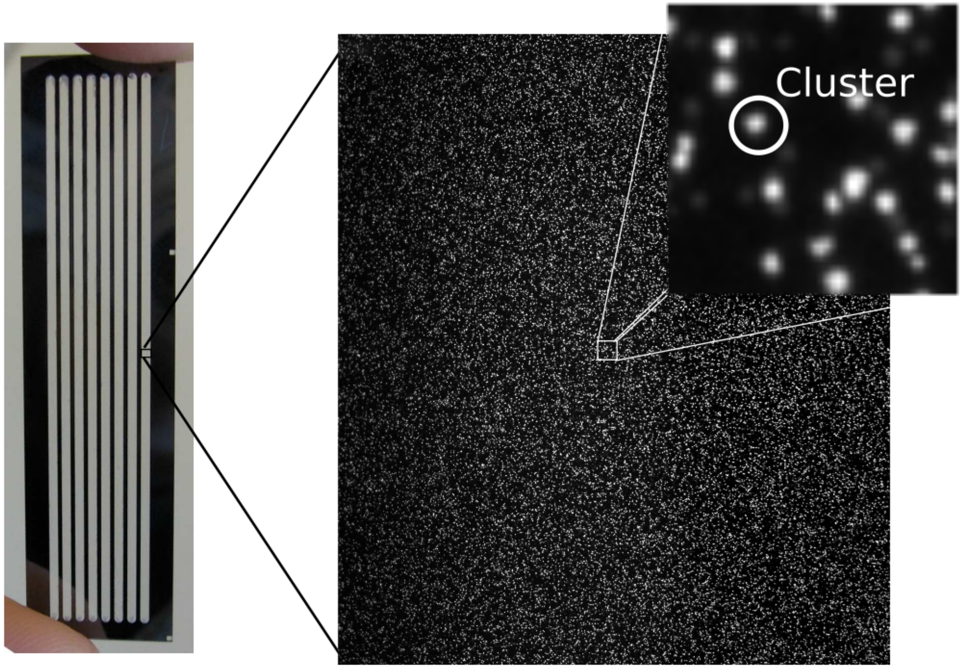
Applications of Free Factories:

- Swift, Editing, Four Individual Genomes

The Aim of Free Factories:

Enable the Personal Genome Project to bring together physicians, scientists and the general public in a single endeavor

Swift—the open source data processing pipeline for the Illumina GA achieves a 14% improvement in placed reads for PGP2 data

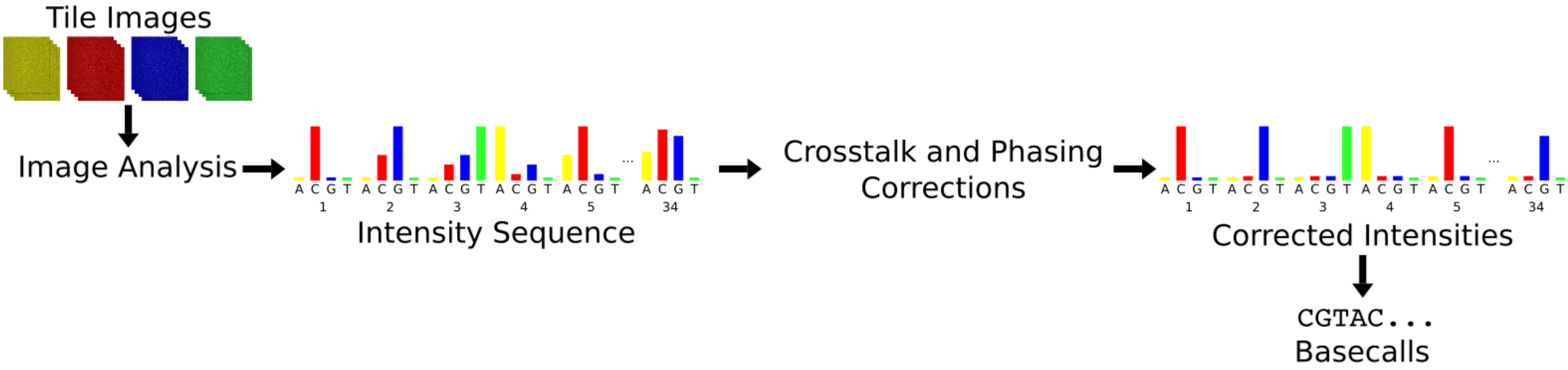


An Illumina GA2 flow-cell consists of 8 lanes which are imaged in 100 tiles (see left)

The Swift pipeline examines all images in a tile and makes basecalls. (see below)

This process is very similar for the Polonator and for ABI SOLiD

The community is eager to adapt the pipeline to new instruments



Tile Images

Image Analysis →

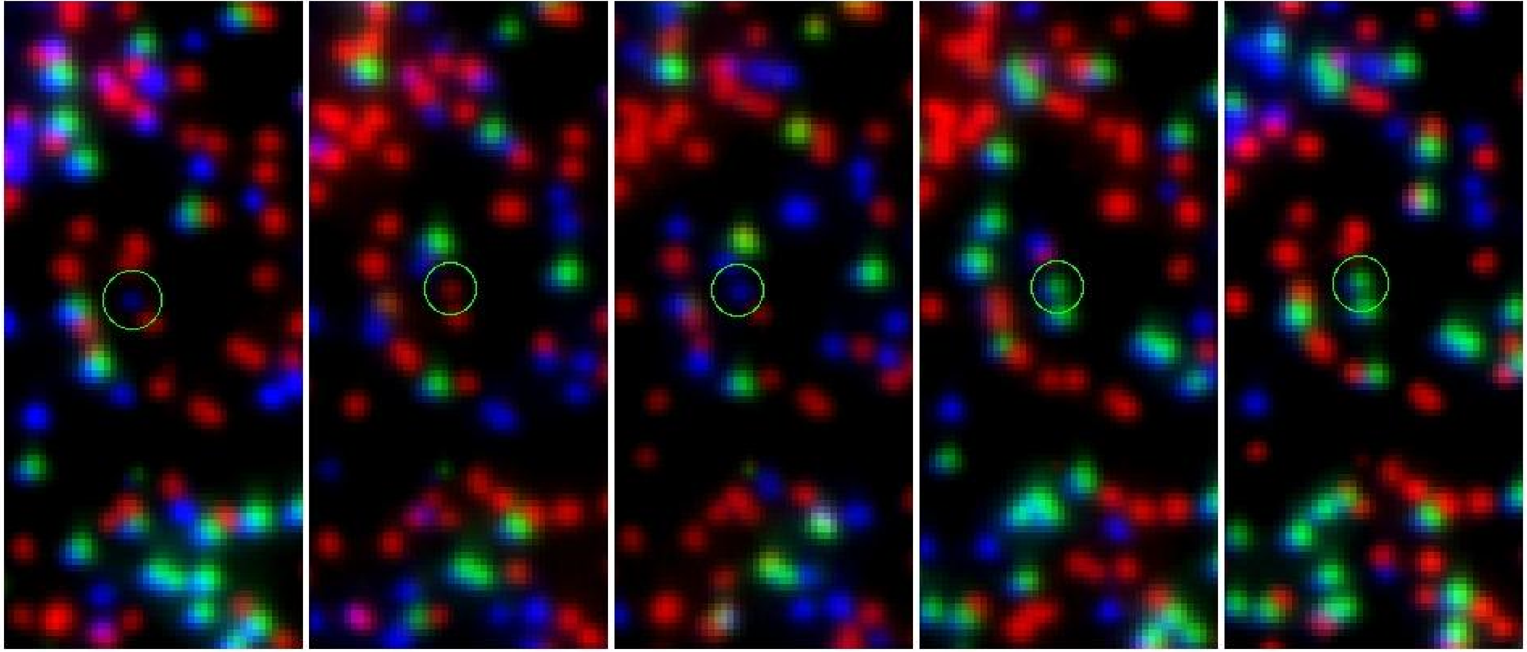
Intensity Sequence →

Crosstalk and Phasing Corrections →

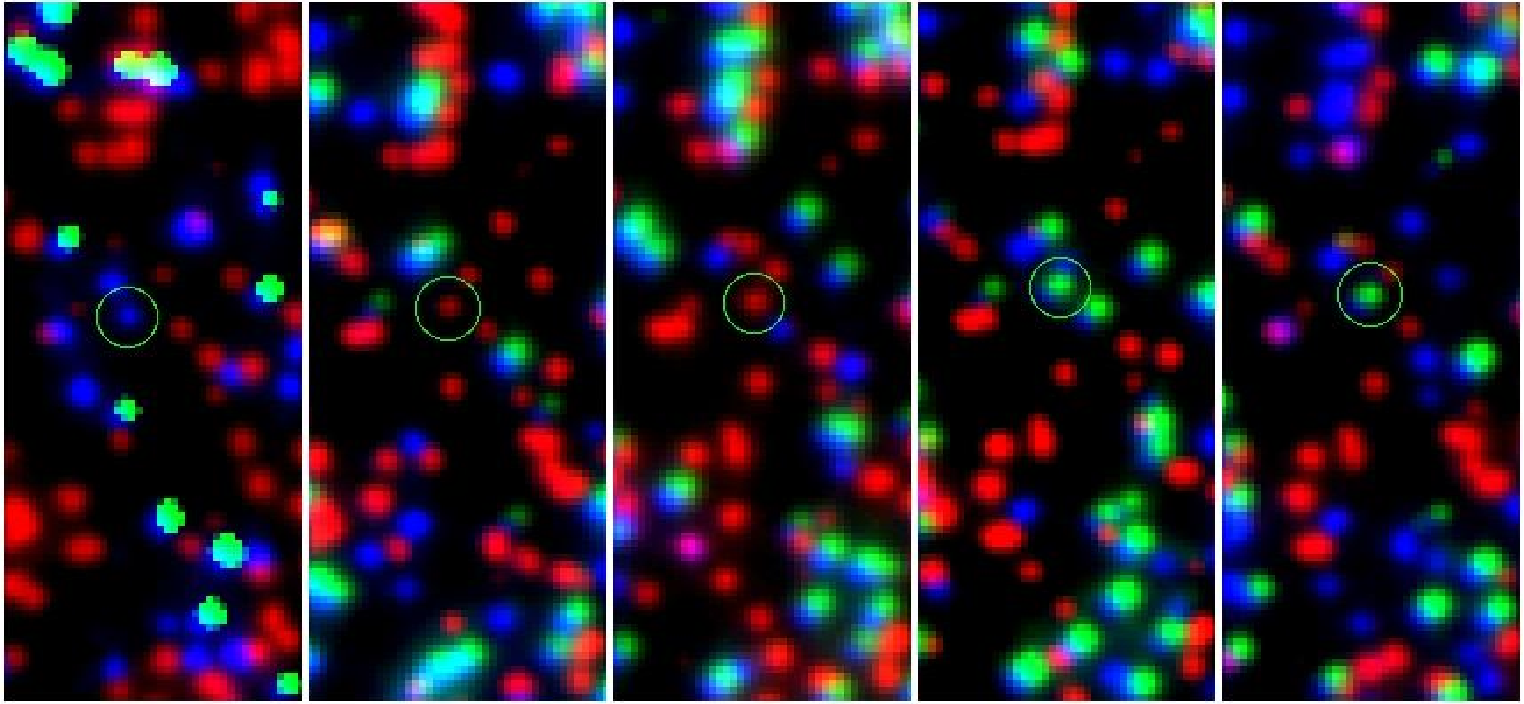
Corrected Intensities

CGTAC...
Basecalls

GCACACGG**TCTGG**GCCAAGCAGATTGCAGAGGCGGg



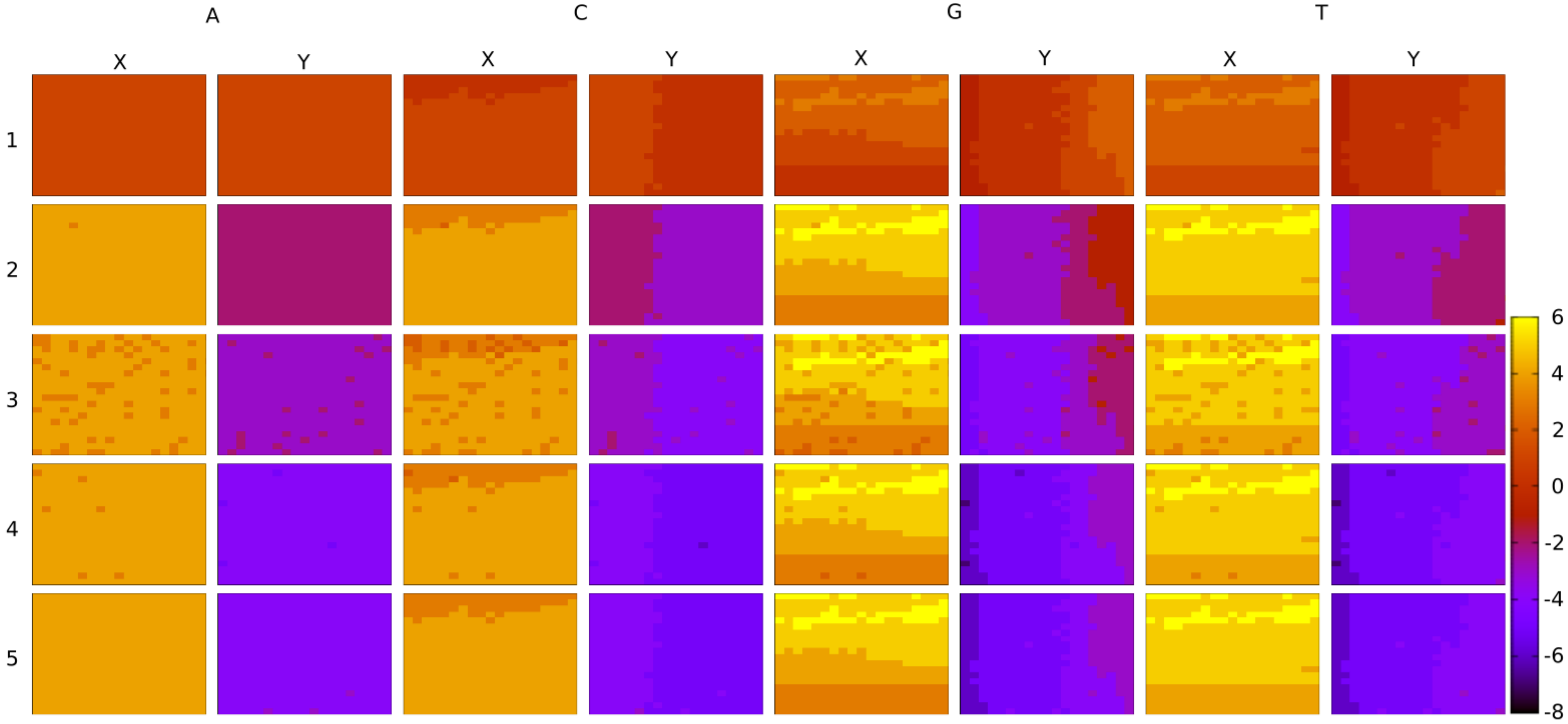
CAGCACACGG**TCCGG**GCCAAGCAGATTGCAGAGGCG



Red = C Green = G Blue = T

Courtesy – Andrea Loehr

Open-source encourages innovations on one platform to be adopted on others



Plot shows X,Y offsets—separately for A,C,G,T florescence channels—in 400 subregions from 5 tiles of Illumina data (processed by the Swift pipeline). The Swift community is eager to try such innovations in the Polonator.

Overview

“Big Data” and Free Factories

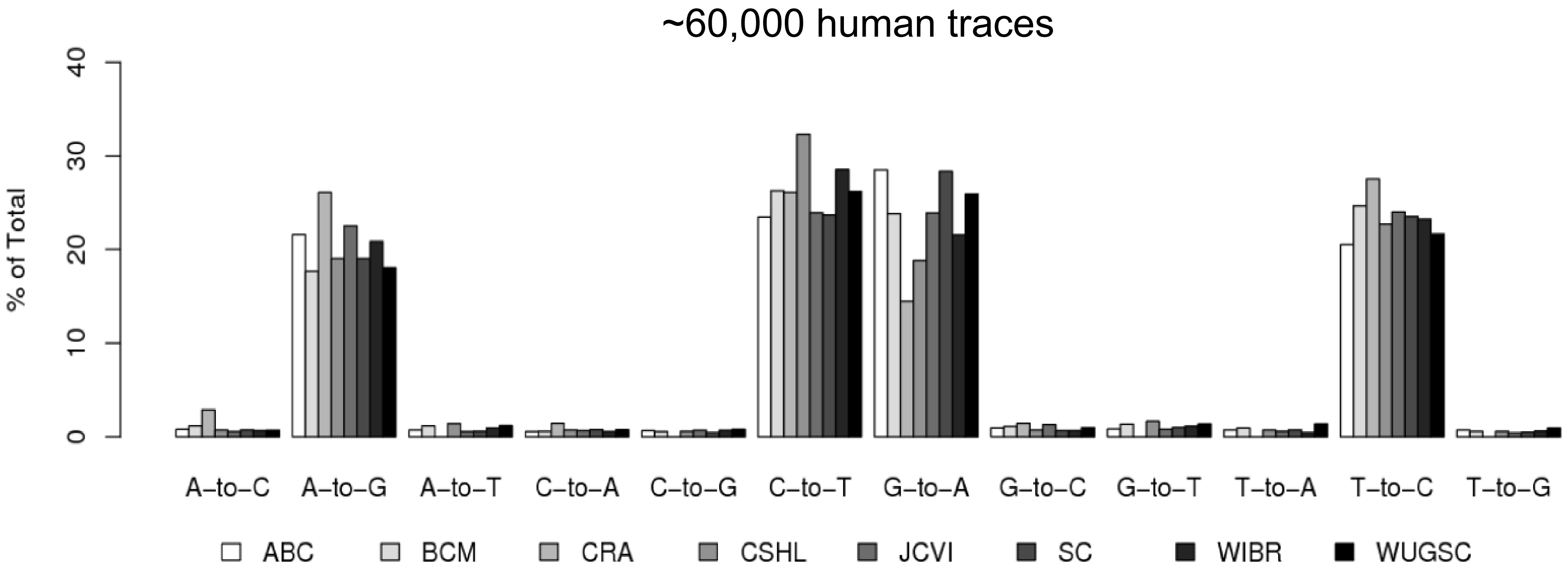
Applications of Free Factories:

- Swift, Editing, Four Individual Genomes

The Aim of Free Factories:

Enable the Personal Genome Project to bring together physicians, scientists and the general public in a single endeavor

Search for DNA and RNA editing candidates



Align 600 million traces (200 million human) to their reference genome

Extract 20.7 million w/ three consecutive mismatches

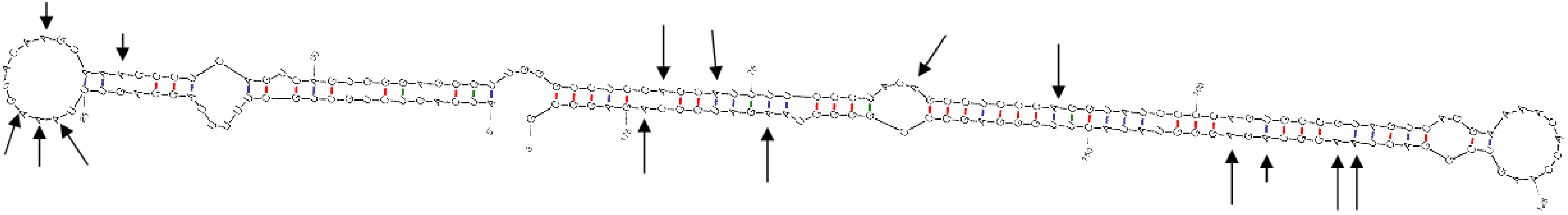
Report mismatches with phred 40 or greater and spanning 100bp or more

Consumed 5.8 terabytes of disk and 5 years of computer time

```
Genome 421  CGGTCCTGAAGGCACAAGTAAGTTACATGAGGAAGTGGCTCAAGTGCCCATGGTCTCTAC  480
Trace  421  ............................................................  480

Genome 481  TCCTGCCACCTGCCTTCTCTCCCCTAGCCTGCACCGATGGCCTCATGGGGAGTTCCCTGT  540
Trace  481  ............................................................  540

Genome 541  GATCAGTTGACAGAGGAAGGGAAGACTAGGCCCTGGTTCAGAGATGGTTCTACATGATAT  600
Trace  541  ............................................................  600

Genome 601  GCAGGCACCACCCGGAAGTGGACAGCTGCAGGACTACAGCCCTTTCTAGGACATCCCTGA  660
Trace  601  .............A................A.............................  660

Genome 661  AGGACAGCGGTGGAGGGAACTTCCCAGTGGGCAGAACTTCGAGCAGTGCACCTGGTTATG  720
Trace  661  ..............AA............................................  720

Genome 721  CACTTTGCATGGAAGGAGAAATGGCCAGATGTCTGATTATATACTGATTCATGGGCTGCA  780
Trace  721  .............................................A..............  780

Genome 781  GCCAATGGTTTGGCTGGATGGTCAGGGACTTGGAAGAAGCATGATTGGAAAATGTGTGAC  840
Trace  781  ......A............A....A......A..............AA.........A..  840

Genome 841  AAAGAAATCTAGGGAAGAAGTATGTGGATGGACCTCTCTGAGAGGTCAAAAACTGTGAAG  900
Trace  841  ...........A................................................  900

Genome 901  ATATTTGTATCCCATGTGAGTGCTCACCAATGGGTGACCTCAGCAGAGGGGGATTTTAAC  960
Trace  901  .................................................A.A........  960

Genome 961  AATCAAGTGGATAGGAT  977
Trace  961  .................  977
```

A DNA editing candidate at the locus of a human retrotransposon. TI#1735626615 aligns uniquely to Chr2 where the known retrotransposon HERVL-A1 is located. A cluster of mismatches (worst mismatch phred 35; best mismatch phred 49) suggests that the trace originates from an edited version of the element. GG-to-AG (11/15) and GA-to-AA (4/15) motifs suggest the preferred dinucleotide context for APOBEC3G and APOBEC3F.

```
X.T Genome      ATCAGTCTGCTGCTTTTTAGGAGTTTAAAGGACAAGTAAAGCCTCAGTCAGTGGGAGGGT
1810455972      ..........................ggg.....g....g.................... 234
1810477902      ............................g..........g.................... 334
1065483469      ..........................ggg............................... 283
1065490247      ..........................ggg............................... 328
1065466398      ..........................ggg............................... 524
1065471353      ..........................g................................. 535
1669879253      ............................g.....g......................... 464

X.T Genome      TGGGCCTCCACCATTTTCCCTACAGCCTCCCAGGTATCCCCAGTGCCGTAGTCAGGAAAA
1810455972      .........g..g..........g.................................... 294
1810477902      .........g.................................................. 394
1065483469      .........g..g..........g.......g............................ 343
1065490247      .........g.............g.................................... 388
1065466398      .........g..g..........g.......g............................ 584
1065471353      .........g.............g.................................... 595
1669879253      .........g.................................................. 524

X.T Genome      CACCAAGTCGGACTAAGGCAGAGGGTATACTTGGGAGGCCGGGGTAAGATGGCAGAGGCG
1810455972      ..............gg.....g........................g......g...... 354
1810477902      ..............gg..............................g............. 454
1065483469      ..............gg.....g........................g......g...... 403
1065490247      ..............gg.....g........................g......g...... 448
1065466398      ..............gg.....g........................g......g...... 644
1065471353      ..............gg...g.g........................g......g...... 655
1669879253      ..............g......g........................g............. 584
```



Evidence for RNA editing in X. Tropicalis. Multiple traces (of RNA origin) show numerous A-to-G mismatches. The predicted RNA structure shows a long dsRNA structure which is the preferred target of editing by ADARs.

Overview

“Big Data” and Free Factories

Applications of Free Factories:

- Swift, Editing, <u>Four Individual Genomes</u>

The Aim of Free Factories:

Enable the Personal Genome Project to bring together physicians, scientists and the general public in a single endeavor

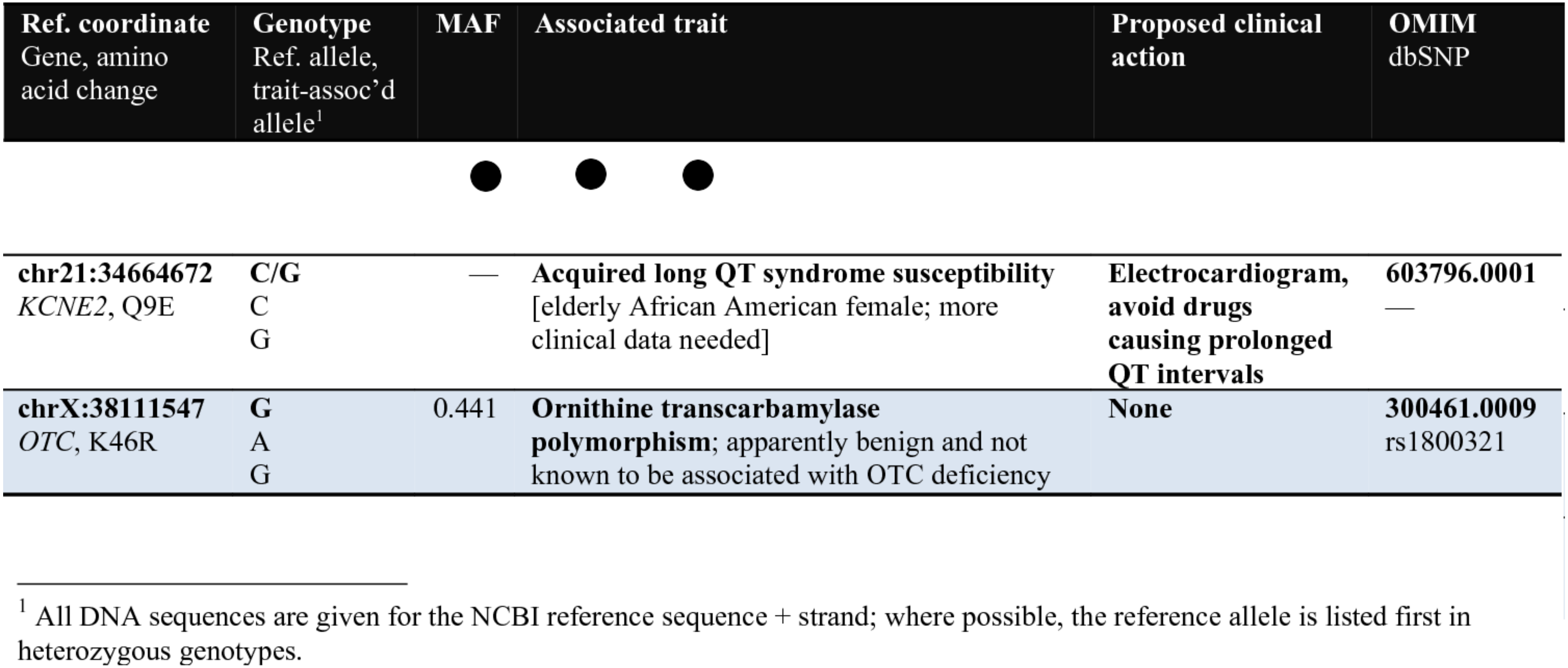
Analysis of individual genomes

Use GeneTests to focus on genes where clinical action is already taken

Convert variants in HG18 coordinates into gene/protein coordinates

Cross-reference with OMIM to obtain a list of known variants with pointers into the literature

Obtain allele frequencies when available (typically not available for rare variants)



Ref. coordinate Gene, amino acid change	**Genotype** Ref. allele, trait-assoc'd allele[1]	**MAF**	**Associated trait**	**Proposed clinical action**	**OMIM** dbSNP
		●	● ●		
chr21:34664672 *KCNE2*, Q9E	**C/G** C G	—	**Acquired long QT syndrome susceptibility** [elderly African American female; more clinical data needed]	**Electrocardiogram, avoid drugs causing prolonged QT intervals**	**603796.0001** —
chrX:38111547 *OTC*, K46R	**G** A G	0.441	**Ornithine transcarbamylase polymorphism**; apparently benign and not known to be associated with OTC deficiency	**None**	**300461.0009** rs1800321

---

[1] All DNA sequences are given for the NCBI reference sequence + strand; where possible, the reference allele is listed first in heterozygous genotypes.

Analysis of an individual African genome reveals a rare mutation—KCNE2 Q9E—not present in dbSNP. Since the individual is anonymous, recommendations cannot be returned to them.

```
chr21 34664672 34664672

aggagggaagcatgtctactttatccaatttcacaG            #$'+&#*,-..$35<$4+<9IC=9EGE?/%IICI2+
aggagggaagcatgtctactttatccaatttcacaC            1(*+),,48029*22<=:?44AIIIIIIGI5IIIII
 ggagggaagcatgtctactttatccaatttcacaCa            IIIIIIIIIIIIII7EIIIIIII?4+:;I>;I05)
   agggaagcatgtctactttatccaatttcacaGaga            IIIIIIIIIIIIII>II<IIII20:IIB++,)3:/<
    gggaagcatgtctactttatccaatttcagaGagac            IIIAIIIIIIII:I<3III+III)1III/1%%1%0,/
    gggaagcatgtctactttatccaatttcacaGagac            II*I@G.7IIIII,IIIIIIIIIIIIIIIIIIBI5I
     ggaagcatgtctactttatccaatttcacaGagacg            IIIIIIIIIII@;IIII8I>C?'IIDI*I9+-H-8-
      gaagcatgtctactttatccaatttcacaGagacgctggaa       /CI-(@379*58+A+@I7)III9+6BCIIIIIIIIIIIIIII
         gcatgtctactttatccaatttcacaGagacgctgg            IIIIIDIF;I@EE2<I/2&5<9:.<+&&3+.(++&(
         gcatgtatactttatccaatttcacaGagacgctgg            ","#"'"%#%*-$4$&/,(,3":59%+I2;I#C003
          catgtctactttatccaatttcacaGagacgctggaa           +%,/'1(&2(++7)/I1I(-&@>IB8I6<III+EH?D
           atgtagactgtatccaatttcacaCagacgctggaa            4683$#A,9'$;%I5II4$+BII-IIIIIIIIIIII
           atgtcttctttatccaatttcacaCagacgctggaa            ,++1'1#52-09/+:,I6/+I3I)=?I9IIF3<6II
           atgtctactttatccaatttcacaCagacgctggaa            "%'&"%"$'4,3*+*A/*2"B/01H9C?ICIAIIII
            tgtctactttatccaatttcacaCagacgctggaag            IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9@I-,I
            tgtctactttatccaatttcacaGagacgctggann            IIIIIGIIII?III2CIIII*<29-94;<2<&4"!!
            tgtctactttatccaatttcacaCagacgctggaagacgtt       ;I*;IF5GCICG97?,>I4422.22+.--.-&-%&')%'%$
            tgtctactttatccaatttcacaCagacgctggaag            5B7=I5=H<B2I83,/8:C7+3/4***-4/+/4")4
              tctactttatccaatttcacaCagacgctggaagac            &&#*&'+&6&-*?:-+&2<4EIII?III2IIIIIII
              tctactttatccaatttcacaGagacgctggaagac            .),%*+,-.#/.065-3>4:I596&I+IBG,III8I
              tgtactttagccaatttcacaTagacgctggaagac            &"#&&+#%,%,%&/&$&,.(8%=$-IG:78I1II=I
              gctggggaattaatgcagattTtgtcagtctttaaa            %"&'$"$%#"#"#-0)&$$%6',%+$&1)'+10.(@
              tctactttatccaatttcacaCagacgctggaagac            4*12(24+I0;=II0+4I9CIIIIIIIIIIIIIIIF
               ctactttatccaatttcacaCagacgctggaagacg            &$%&*()+(-)/1%*&&%+&>2&(4532<=(4DI@.
               ctactttatccaatttcacaGagacgcttgcagacg            IIIIIIIIIIIIIIIIIIIGIII*734*.2"36-%4
                 actttatccaatttcacaGagacgctggaagacgtc            &+.+(+)-%25&5'4.<88I;+AAI)II5II@IGII
                 actttatccaatttcacaGagacgctggaagacgtt            IIIIIIIIIIIII@IIIIIIIIIHIII&EIF$II*
                  ctttatccaatctcacaCagacgctggaagacgtct            IIIIIIIIIIIIIII:IIEIGIIE.7+0-**05+6
                   tttatccaatttcacaGagacgctggaagacgtctt            ++(('8/'.1+4F9@IH;54I4CI*I*>E)IIC>=@
                    ttatccaatttcacaGagacgctggaagacgtcttc            6,2$I2HC3+,;<IBHII2IBIAIIIIIIIIIIIII
                    ttatccaatttcacaCagacgctggaagacgtcttc            IICI-(I-AII?211131*+;114/.-+5&$&$/+"
                      atccaatttcacaCagacgctggaagacgtcttccg            +**+54))&,'2*6-.)26643IB<<7II6IHI>II
                       tccaatttgacaGagacgctggaagacgtcttccgaa           "%&+,%(#"&#/+1(+-=$0:3IC76%</IF=ADII?
                       tccaatttcacaCagacgctggaagacgtcttccga            %/&BI+'%(-)1?I2=<687==IIIBIAIIGDIIII
                       tccaatttcacaCagacgctggaagacgtcttccga            I>>5H11A:5205,;>3<3+%-)**+*6&(/,*(,'
                        ccaatttcacaCagacgctggaagacgtcttccgaa            +%&($%$.%*&11.%340-6'5C84AAI;IIIIIII
                         caagttcacaCagacgcgggaagacgactttcggag            5I2A9.+ICIII;@$1'+&2'2-&*0$$&+"&&#$&
                          aatttcacaGagacgctggaagacgtcttccgaagg            IIIIIBIIIIIIII>I7HII0CDA*D)*?:'&2+80:
                          aatttcacaCagacgctggaagacgtcttccgaagga           <4IHIIII<<E<+53>-,8+*2(*((+&%%%("%&0"
                           atttcacaGagacgctggaagacgtctttcgtagga            IIIIIII;I=IG/C5-@8%)/(,./#.2$'%$'*&"
                             ttcacaCagacgctggaagacgtcttccgtaggatt            IIIFI>IIIBFE?I96-/;+/,..78*)$##%)".:
                              tcacaCagacgctggcagacgtcttccgaaggattt           IIIIIIIIIIIIIII'II6IIIIHI/I+>&II6III
                                                                               %*%++5/7=98+70II>1II5IDIIIIIIIIIEIII
                                                                                '%"%(&"&&'+%%,3)2&&/$+%$/*'1/)*;)(31
                                                                                 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
                                                                                 IIIIIIIIIII60@4705,=D+05)&45)2/-2;'
                                                                                 +,++++2-5(5-62<.+97164*1I9@FIC.=IEII
                                                                                  >IIIIHICGFB0283'/*,6/&)(#),,")30-3"0
                                                                                   ++'#*',*$)5.+,*,/5,8>8)3;/C8I61;IIII
                                                                                   *
```

```
*
C 25 sum(q)=676
G 22 sum(q)=607
T 2 sum(q)=10
```

The Free Factories infrastructure was used to assemble the raw reads—120 gigabases—from HapMap NA18507. The alignment for KCNE2 Q9E is shown above. SNPs for this genome did not accompany Bentley et al. (2008) Nature.

Overview

“Big Data” and Free Factories

Applications of Free Factories:

Swift, Editing, Four Individual Genomes

The Aim of Free Factories:

- Enable the Personal Genome Project to bring together physicians, scientists and the general public in a single endeavor

PersonalGenomes.org

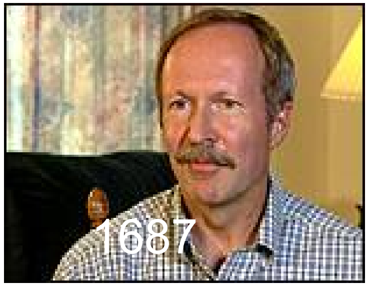
Subject & public access (not just research elite)

Entrance exam to ensure highly informed consent

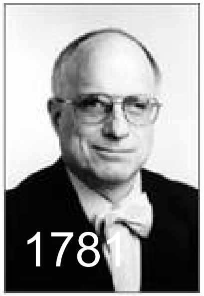
Scalable to millions of research subjects, budget $1,000/person for DNA & trait data

Highly integrated, holistic, systems-biology

Cells available for personal functional genomics

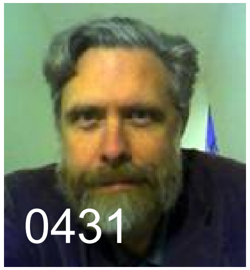


1687

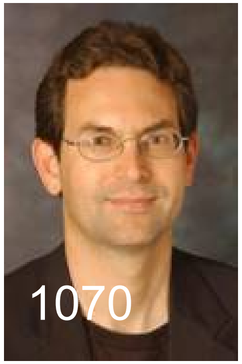


1781

1833

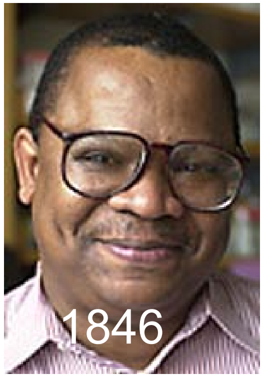


0431



1070

1731



1846

1660

How do we organize computational resources to serve the combined needs of scientists, physicians and the general public?

The Idea

Public

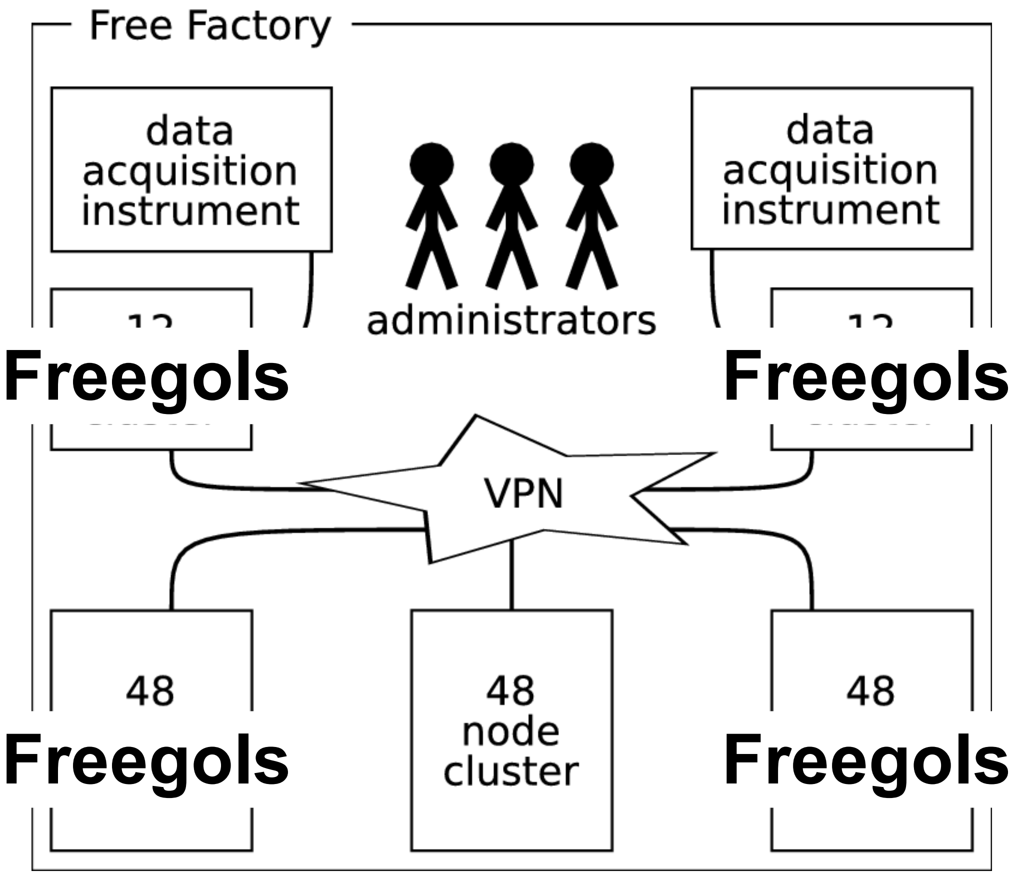
Public

Scientists

Scientists

Physicians

Physicians



Public

Public

Scientists

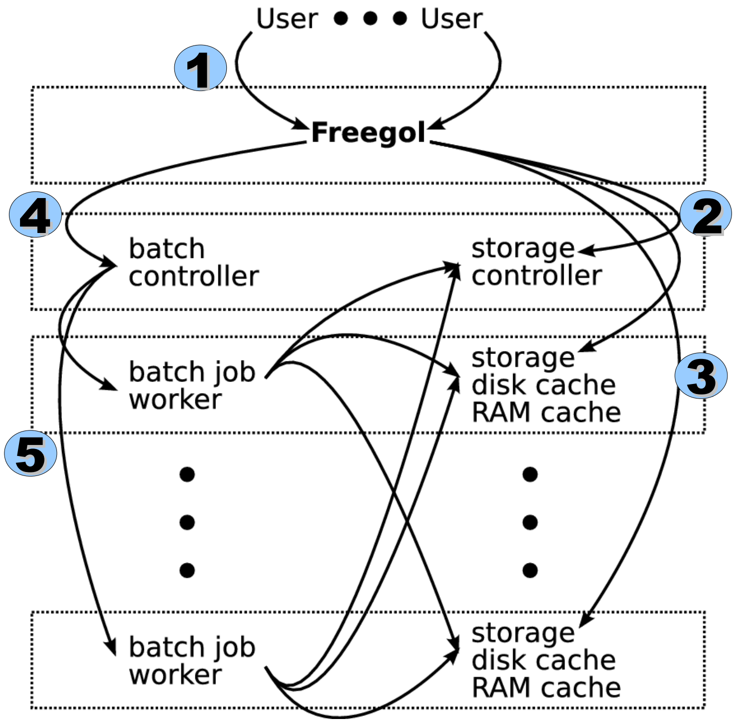
Scientists

Physicians

Physicians

A shared infrastructure for web service virtual machines, which I call “Freegols”.

Freegols—or Free Golems (another word for robot)—operate in independent virtual machines running on the Free Factories infrastructure.



User • • • User

1

Freegol

4

2

batch
controller

storage
controller

batch job
worker

storage
disk cache
RAM cache

3

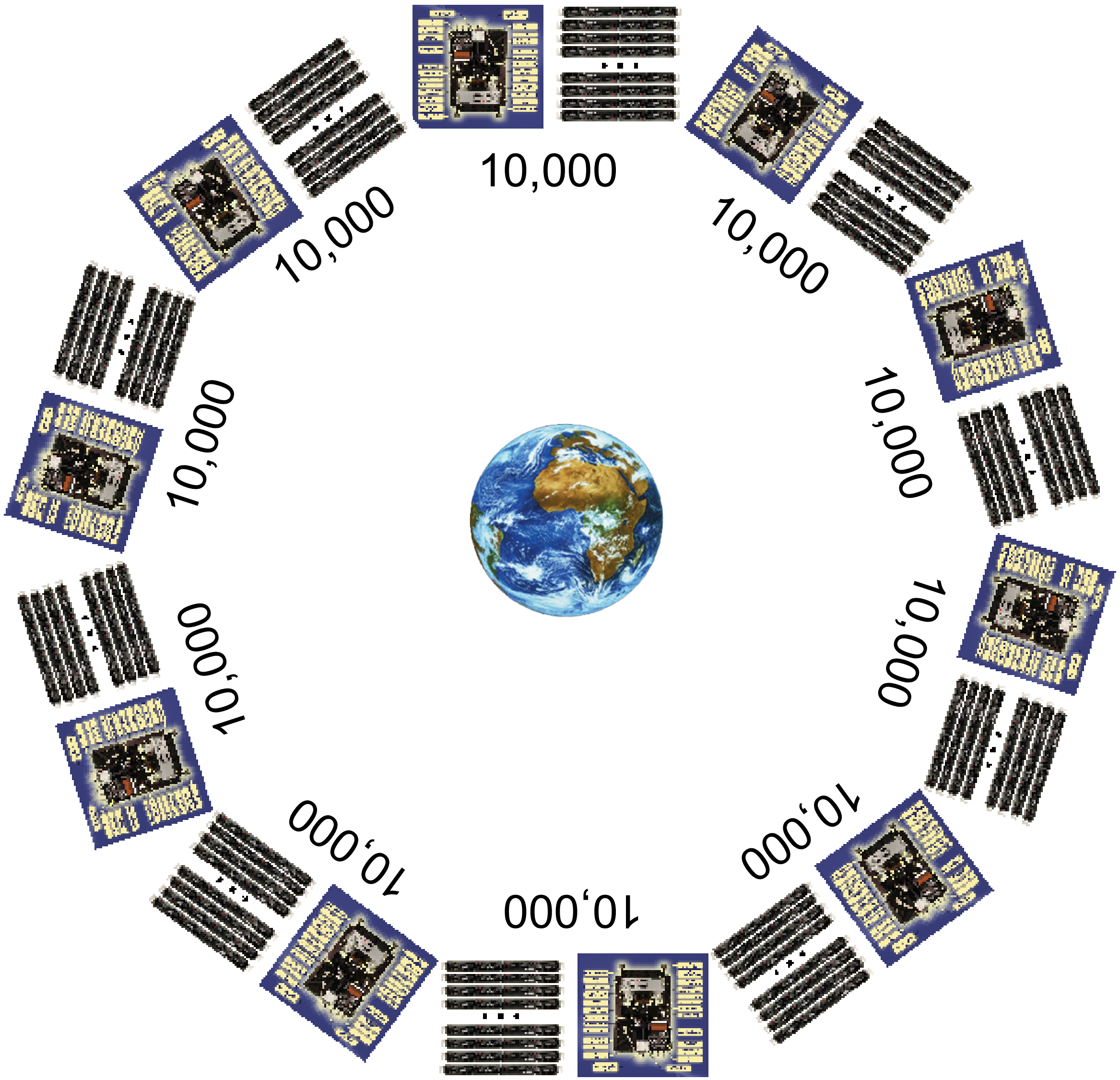
5

batch job
worker

storage
disk cache
RAM cache

As a Freegol services many simultaneous user requests, it continually supervises "workflows" that process terabytes of data and consume many thousands of CPU hours

Scalable Infrastructure for 100,000 people



Maintain infrastructure close to participants

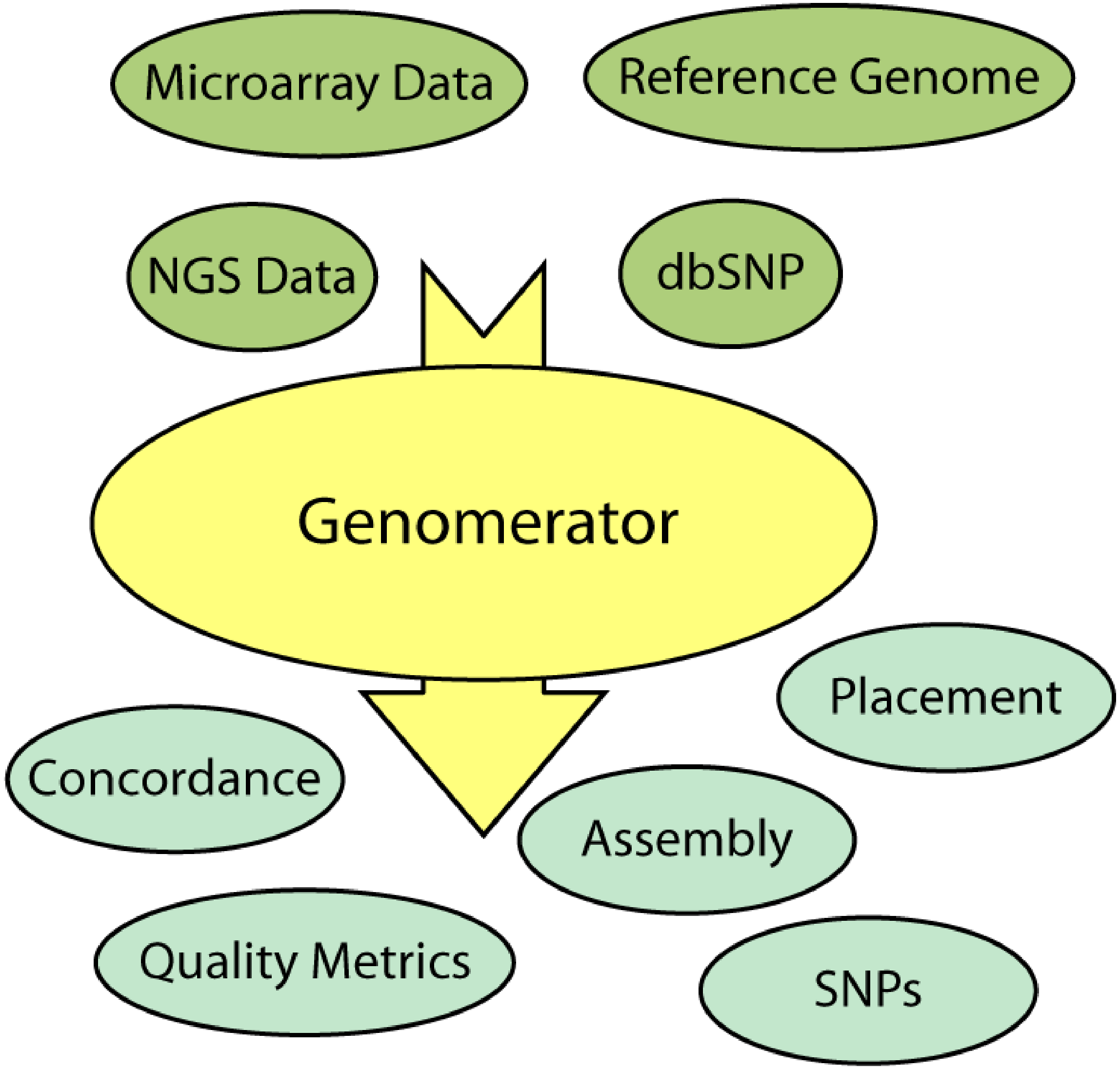
Add sequencing instruments, computational clusters, and storage independently

Freegols can use storage and compute resources from any Free Factory

Fault-tolerant to hardware and software failures

Built-in provenance tracking

Genomerator manages Next-Generation-Sequencing data, launches workflows, and, generates quality metrics


Microarray Data

Reference Genome

NGS Data

dbSNP

Genomerator

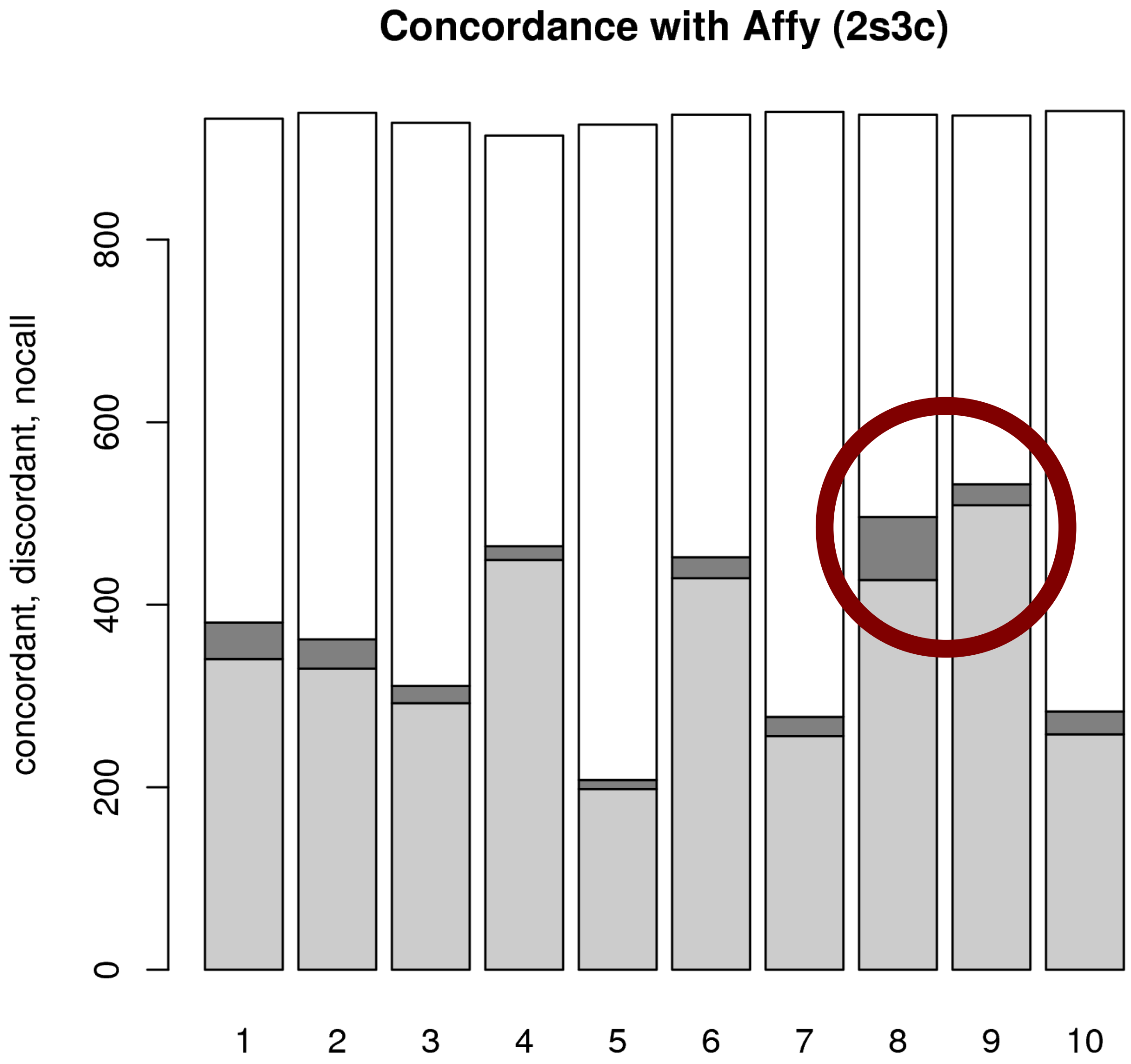
Placement

Concordance

Assembly

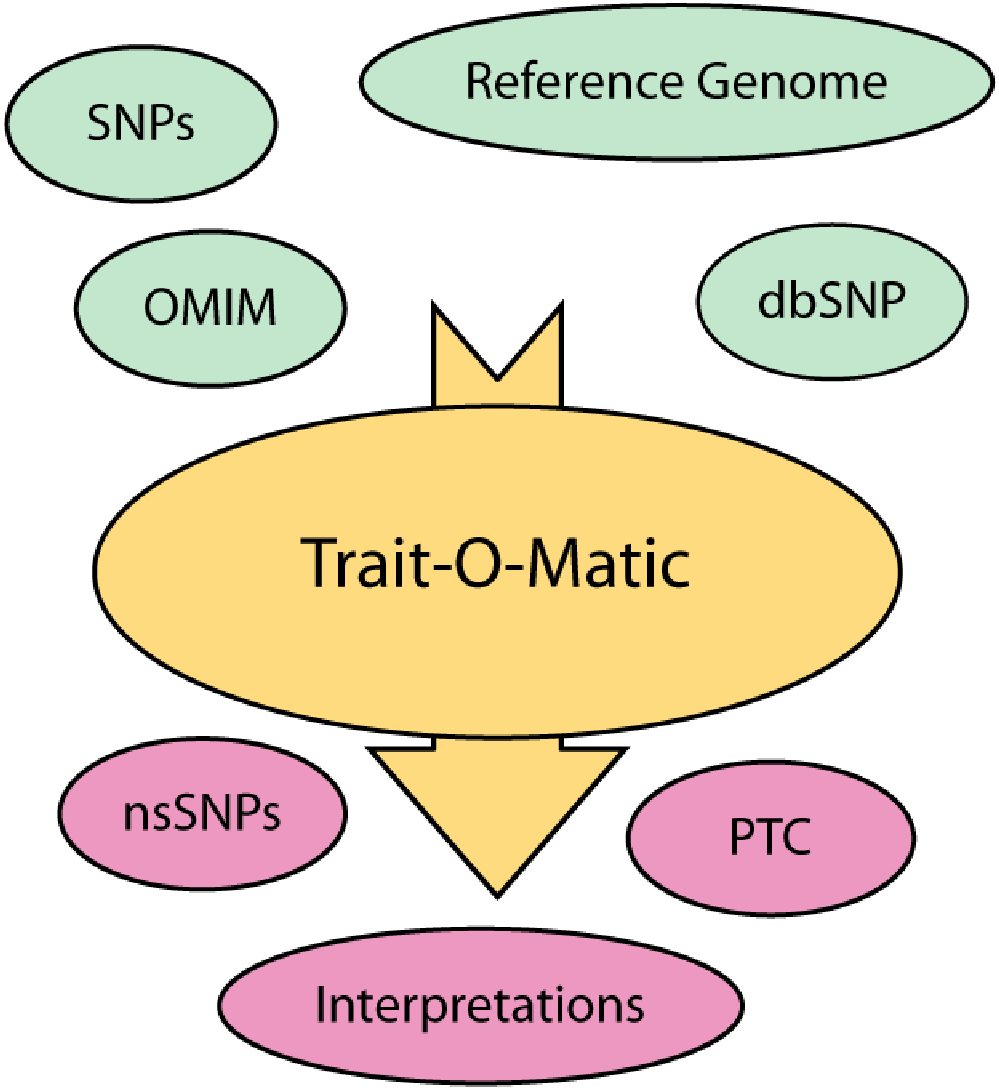
Quality Metrics

SNPs



By reprocessing the data, from images, we can improve the accuracy of PGP8 (the worst sample) to be equivalent to PGP9 (the best) while increasing the number of called variants.

Trait-o-matic interprets variants



SNPs

Reference Genome

OMIM

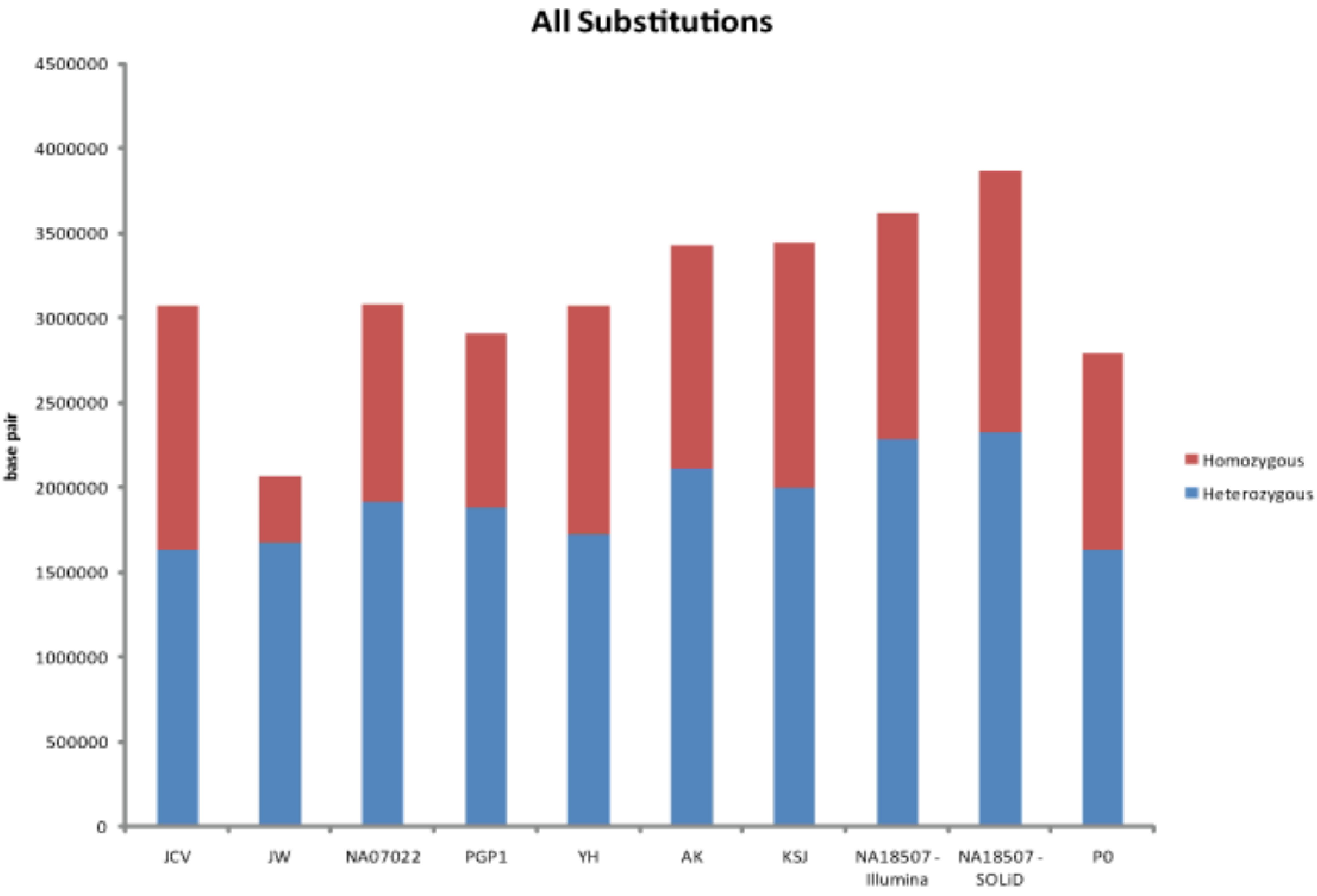
dbSNP

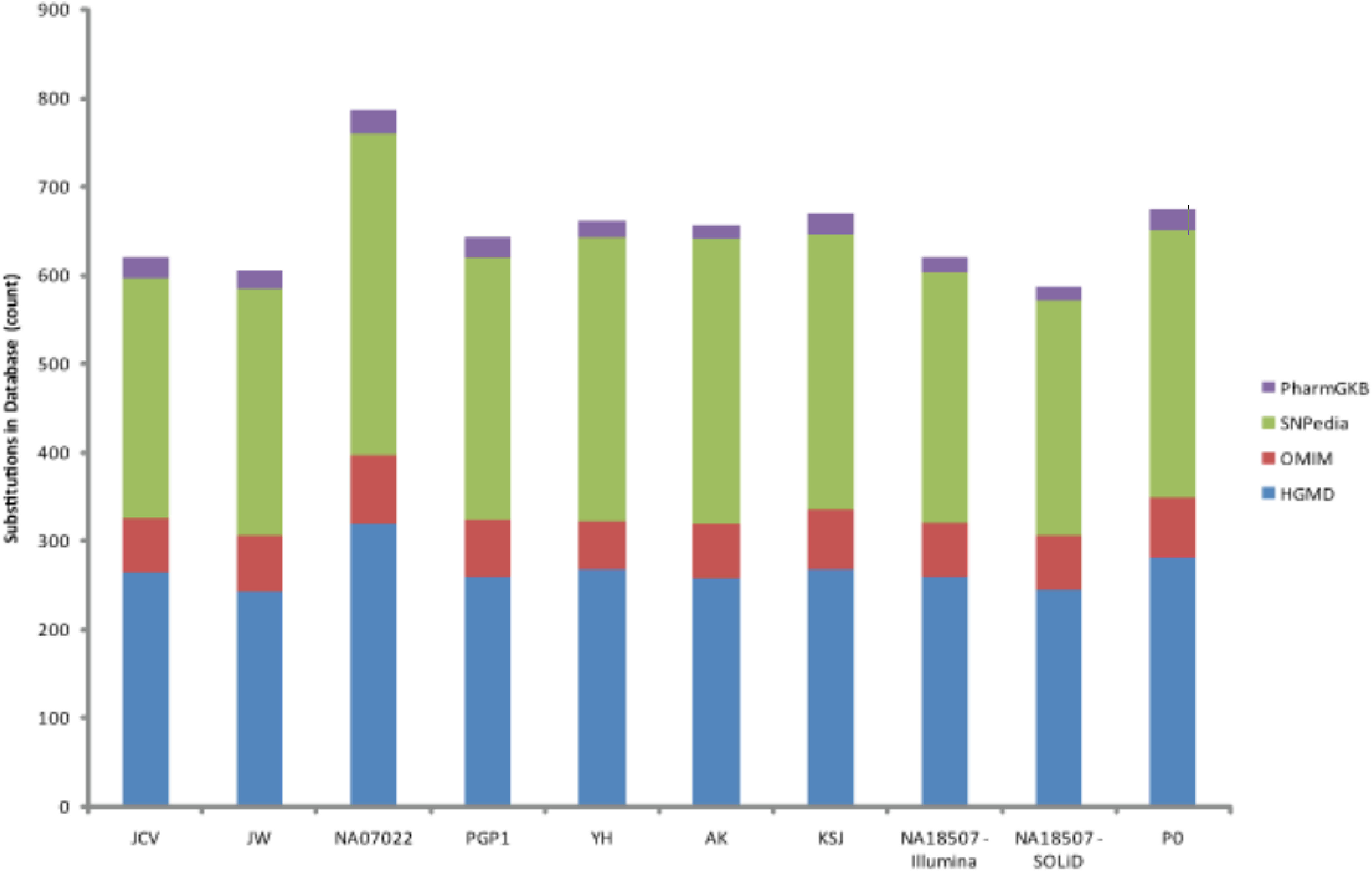
Trait-O-Matic

nsSNPs

PTC

Interpretations

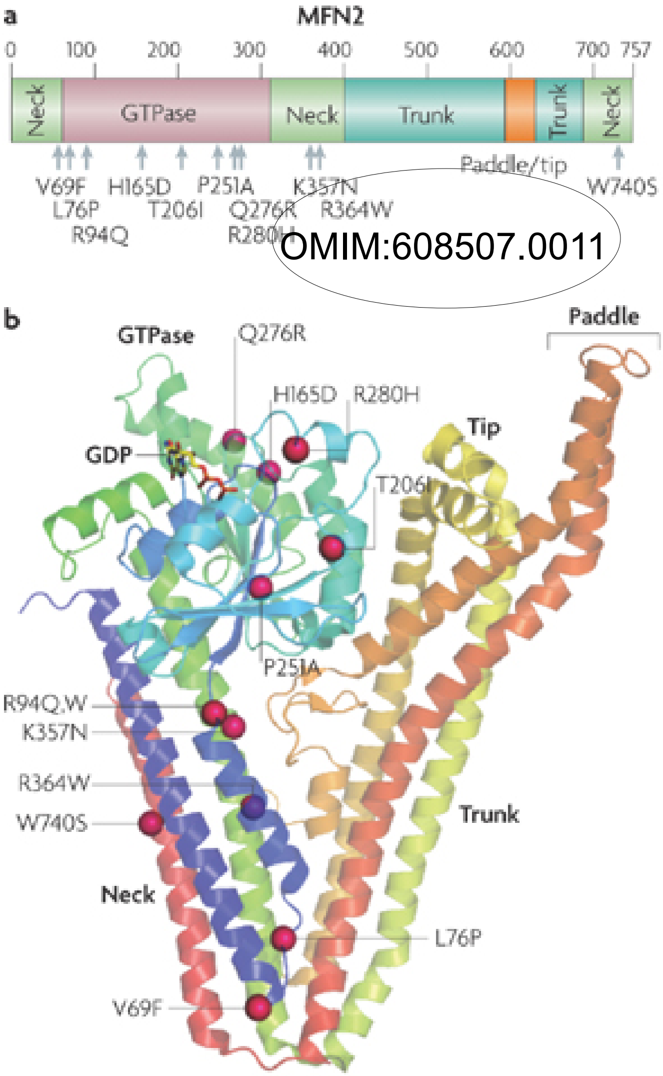

All Substitutions
base pair
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
JCV
JW
NA07022
PGP1
YH
AK
KSJ
NA18507 - Illumina
NA18507 - SOLiD
P0
Homozygous
Heterozygous


Substitutions in Database (count)
900
800
700
600
500
400
300
200
100
0
JCV
JW
NA07022
PGP1
YH
AK
KSJ
NA18507 - Illumina
NA18507 - SOLiD
P0
PharmGKB
SNPedia
OMIM
HGMD

Trait-o-matic cross-references variants with major databases and looks for damaging coding changes

PGP2 – MFN2 R364W – HEREDITARY MOTOR AND SENSORY NEUROPATHY VI

?

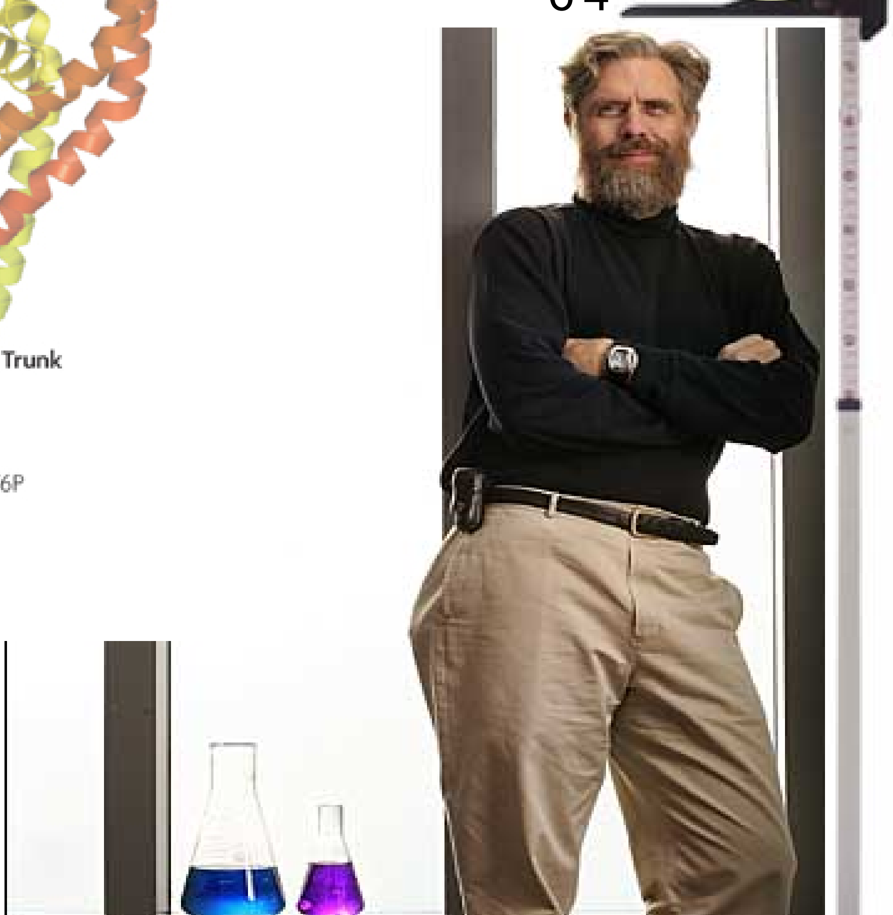


OMIM:608507.0011

PGP1 – GHR G186* – (associated with) INCREASED RESPONSIVENESS TO GROWTH HORMONE

?

6'4"



PGP1 HGR Mutation

```
chr5 42735769 42735805     GAAGCACCACGcAaTGCAGATaTTcaGAAaGGAtGG
chr5 42735776 42735812            CaCgcAATgCaGaTaTtCagaaAtgATggAtggttc
chr5 42735776 42735812            CacGCaaTGCaGatATTcaGaaATGaTggATggtTc
chr5 42735776 42735812            CAcGCAATGCAGaTaTTcagaAATgatggatggtTc
chr5 42735776 42735812            CACGCAATGCAGATATTCAGAAATGATGgATGGtTc
chr5 42735790 42735826                          ATTcAGAAAGGATGGATGGTTcTGGAGTATGAACTT
chr5 42735790 42735826                          AttcAgAAAGGATGGAtGGTtCtGGAGTATGAACtT
                                                         *
```

PGP2 MFN2 Mutation

```
chr1 11984646 11984682 AGTGAAGACCAAGTTTGAGCAGCACACGGTCCGGGC
chr1 11984658 11984694             GTTTGAGCAGCACACGGTCCGGGCCAAGCAGATTGC
chr1 11984658 11984694             GTTTGAGCAGCACAcGGTCCGGGCcAAGCaGATTGC
chr1 11984658 11984694             GTTTGAGCAGCACACGGTCCGGGCCAAGCAGATTGC
chr1 11984658 11984694             GTTTGAGCAGCACACGGTCCGGGCCAAGCAGATTGC
chr1 11984658 11984694             GTTTGAGCAGCACACGGTCCGGGCCAAGCaGATTGC
chr1 11984658 11984694             GTTTGAGCAGCACACgGTCCgGGCCaaGCAGATTgC
chr1 11984658 11984694             GTTTGAGCAGCACACGGTCCGGGCCAAGCAGATTGC
chr1 11984662 11984698                 GAGCAGCACACGGTCCGGGCCAAGCAGATTGCAGAG
chr1 11984662 11984698                 GAGCAGCACACGGTCCGGgCCAagCAgATTgCAGAg
chr1 11984662 11984698                 GAGCAGCACACGGTCCGGGCCAAGCAGaTTGCAgAG
chr1 11984662 11984698                 gAgCAGCACACgGTCCGGGCCaAGCAGATTGCAGAG
chr1 11984665 11984701                    CAGCACACGGTCCGGGCCAAGCAGATTGCAGAGGCG
chr1 11984667 11984703                      GCACACGGTCTGGGCCAAGCAGATTGCAGAGGCGGg
chr1 11984667 11984703                      GCACACGGTCTGGGCCAaGCAGATTGCAGAGGCGGg
chr1 11984667 11984703                      GCACACGGTCTGGGCCAAGCAGATTGCAGAGGCGGg
chr1 11984667 11984703                      GCACACGGTCTGGGCCAAGCAGATTGCAGAGGCGGt
chr1 11984668 11984704                       CACACGGTCCGGGCCAAGCAGATTGCAGAGGCGGTT
chr1 11984668 11984704                       CACACGGTCCGGGCCAAGCAGATTGCAGAGGCGGTT
                                                      *
```

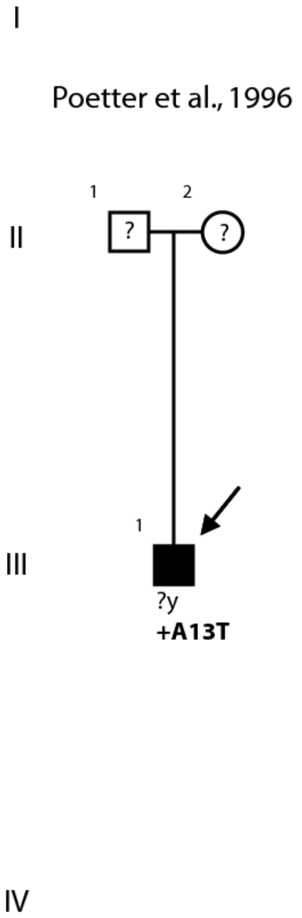
So what went wrong? The error probably occurs in an amplification step required by the capture process.

Trait-o-matic finds MYL2 A13T in PGP6

OMIM:160781.0001

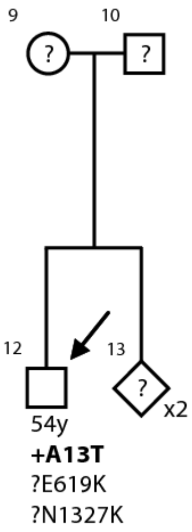
PGP6 –MYL2 A13T–
CARDIOMYOPATHY,
HYPERTROPHIC,
MID-LEFT VENTRICULAR
CHAMBER TYPE, 2
?

Poetter et al., 1996



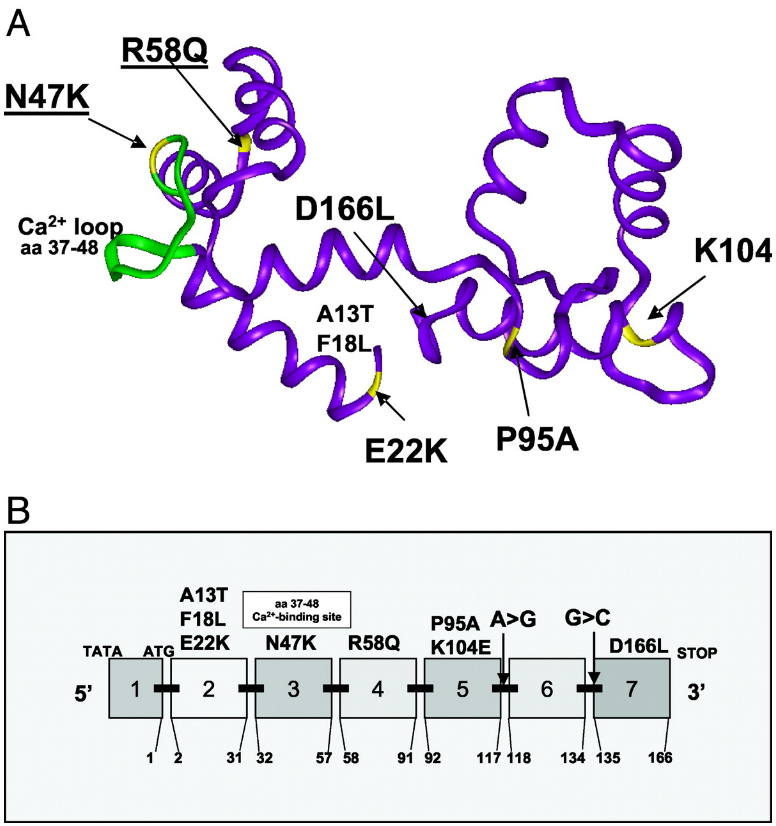
```
Ventricular          10        20        30        40        50        60
    RLC               |         |         |         |         |         |
Human       MAPKKAKKRAGGANSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN
Pig         MSPKKAKKRADGANSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN
Mouse       MAPLFAKKRIEGGTSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN
Rat         MSPKKAKKRLEGGSSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN
Consensus   MaPkkAKKRaeG-nSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN
```

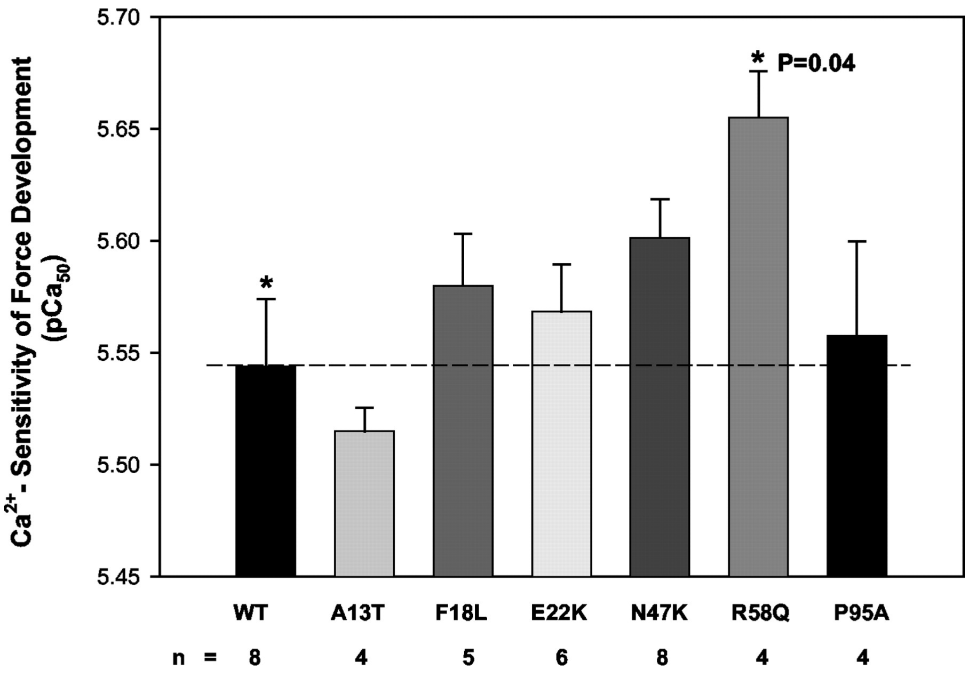
PGP6



http://genetics.med.harvard.edu/~seidman/cg3/muts/MYL2_Ala13Thr.html

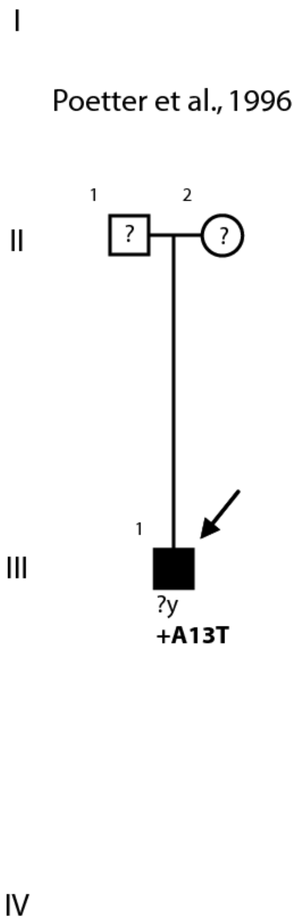
Effect of MYL2 A13T is inconclusive in functional studies



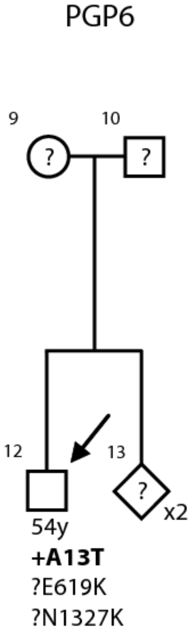


Szczesna-Cordary et al. (2004) J. Biol. Chem.

More clinical data for MYL2 A13T



Poetter et al., 1996



PGP6

Based on this data, PGP6 visited a cardiologist.
While he, thankfully, remains negative for any signs of cardiac disease, the late onset nature of this disease suggests a need for periodic followup .

Free Factories can help scientists, physicians and the general public work together to advance personalized medicine

Acknowledgments

George Church
James Hogle
Zak Kohane
Jon Seidman
Jack Szostak
Abraham Rosenbaum
Xiaodi Wu
Mike Chou
Billy Li
Wendy Chung
Heidi Rehm
John Aach
Joe Thakuria
Erez Levanon
Jason Bobe

Harris Wang
Tom Clegg
Ward Vandewege
Nava Whiteford
Chris Archibald
Andy Chute
Andrea Loehr
Irwin Jungreis
Miron Cuperman
Erik Garrison
Paras Doshi
Church lab members
Harvard Biophysics
Many others

Thank-you!