



From personal genomics to personal health

Daniel MacArthur

Wellcome Trust Sanger Institute

1000 Genomes Project Consortium

Genomes Unzipped Project



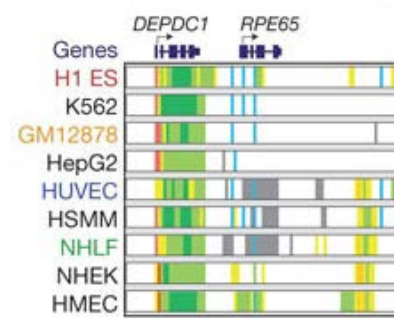
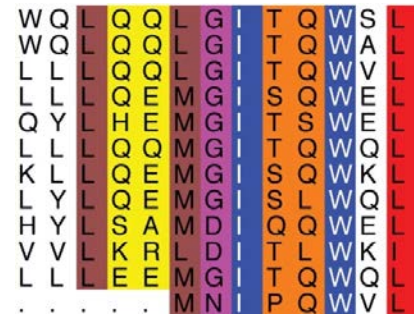
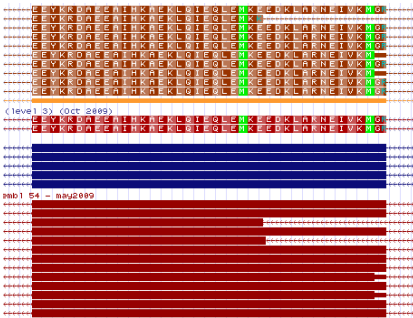
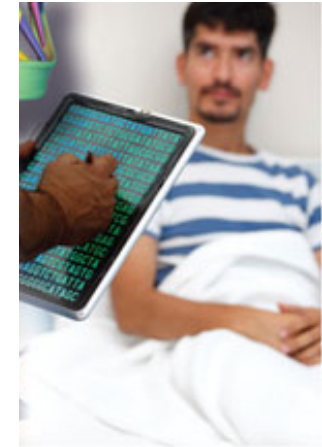
The sequence-function intersection



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

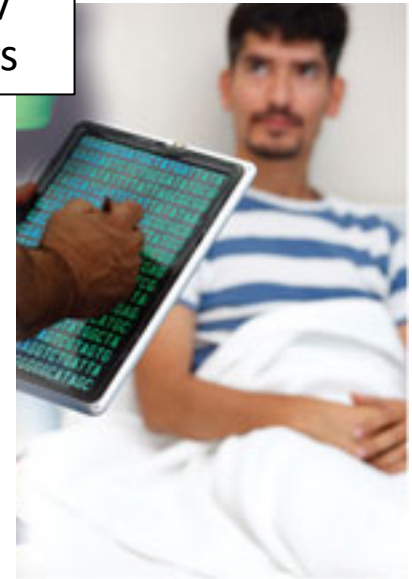
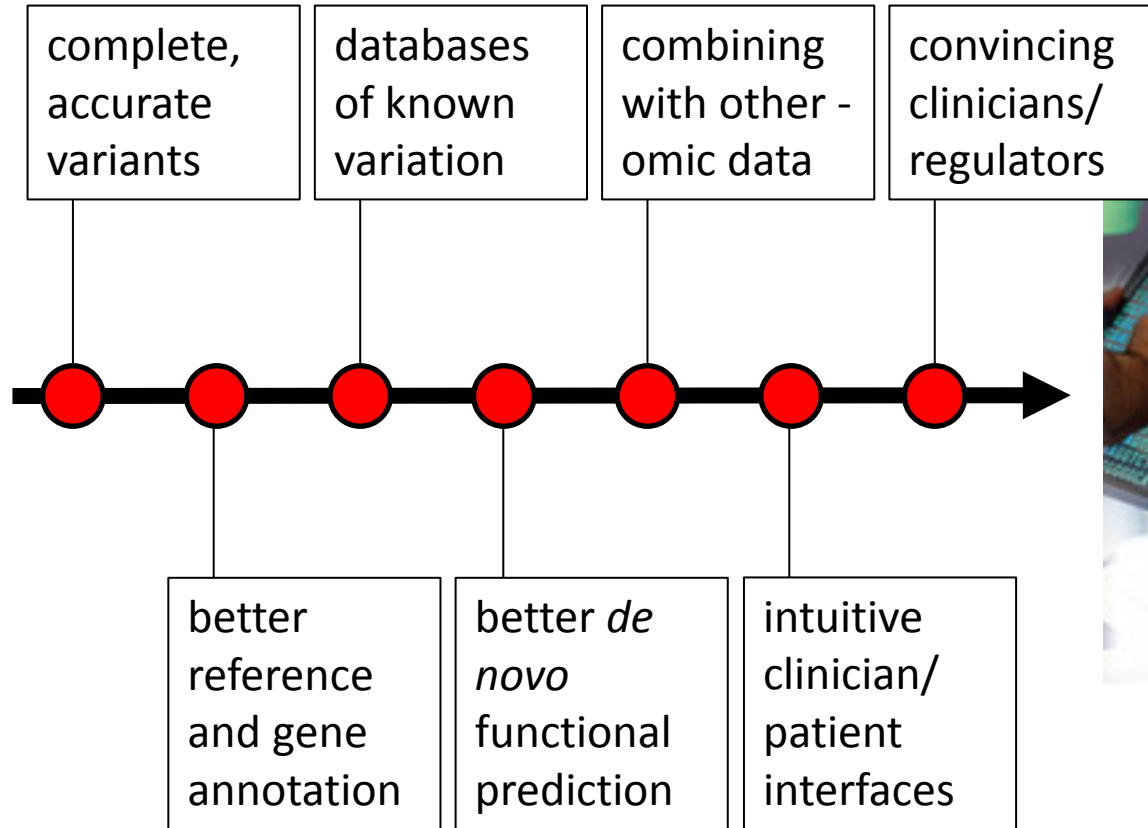


How will genomics transform healthcare?

- Carrier testing for severe recessive disease mutations (prenatal counselling)
- Cancer sequencing to determine prognosis/treatment
- Pharmacogenomics and nutrigenomics
- Risk prediction for complex diseases: both common and rare variants
- **Routine inclusion of *useful* genetic results in healthcare decisions**

Getting from here to there

AGCGCTGGAT
AGGCTCGGAT
CGGCTAGAGA
TTGCTAGCCA
TATAGGGATC
GCCTGGATAG
GCCGATCGGC
TAGGATCTAC
CATTAGGATA



Data-sets discussed



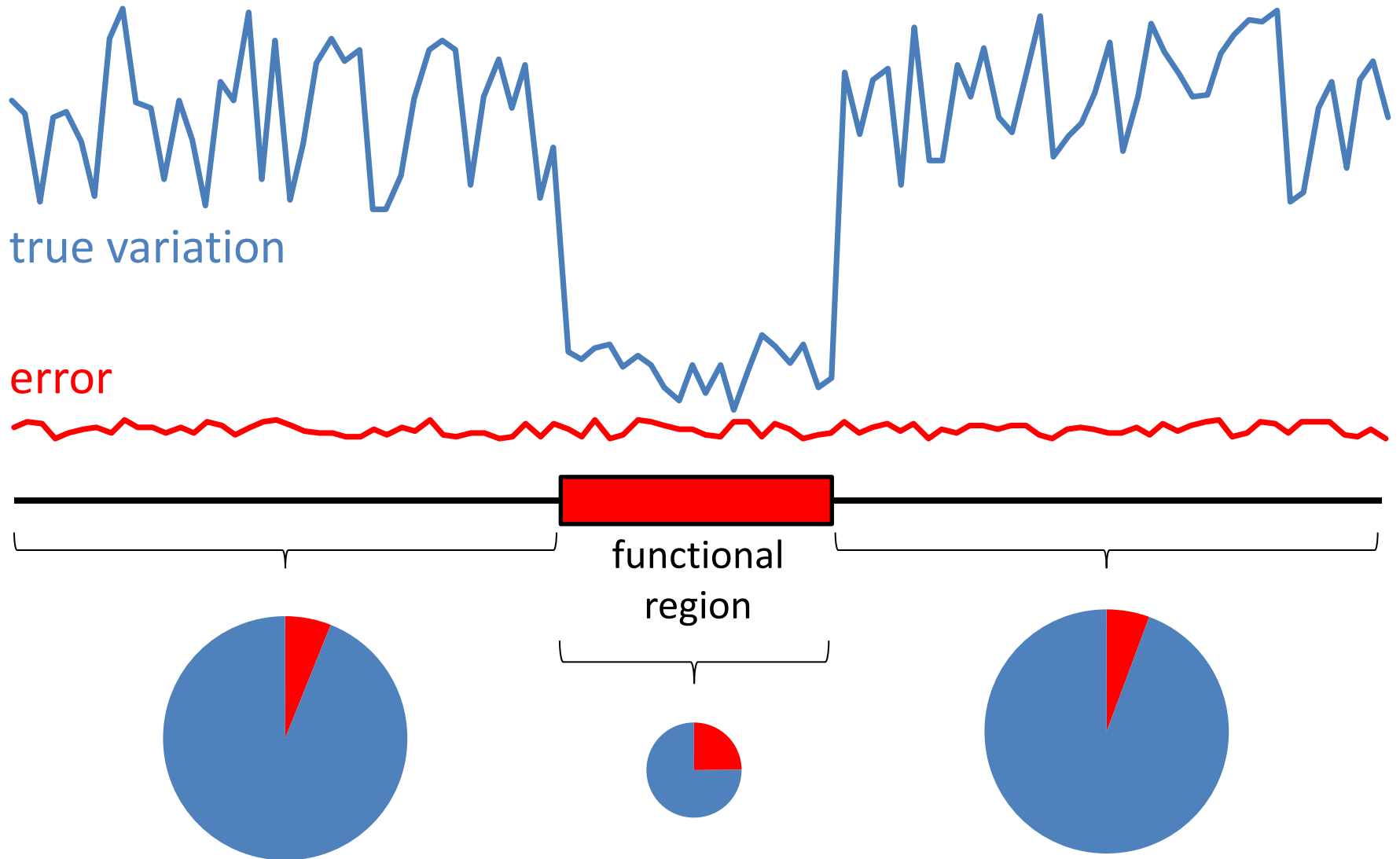
**Functional annotation
group**



**Genomes
Unzipped**

Sequencing is not perfect
Errors and challenges ahead in the
functional interpretation of
sequence data

Functional variants are enriched for artefacts

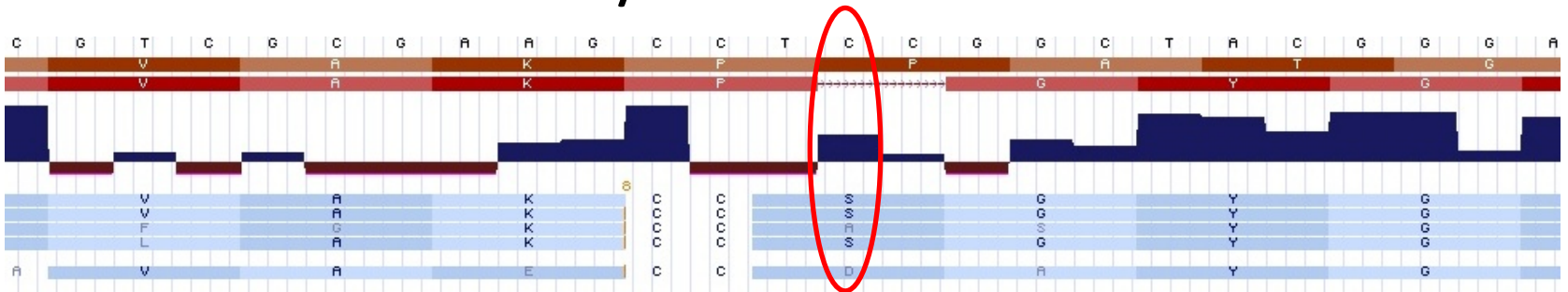


Hunting mistakes

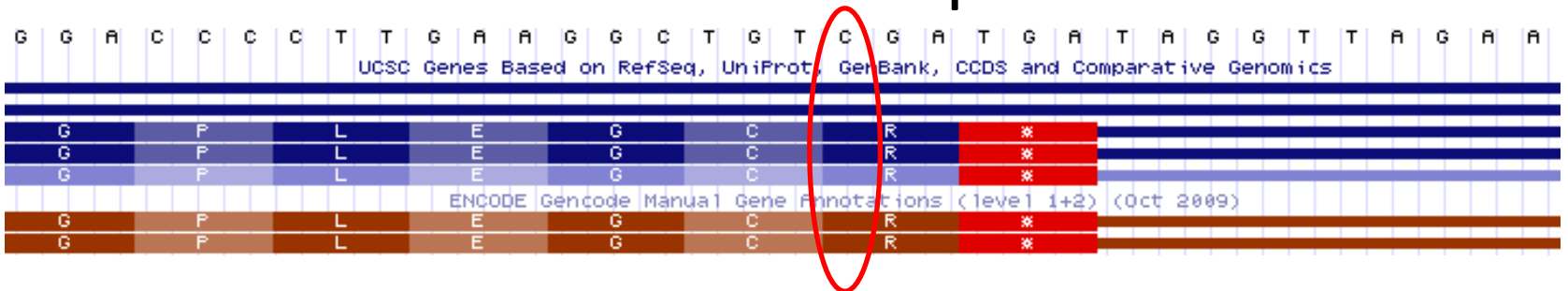
Sequencing/genotype calling error



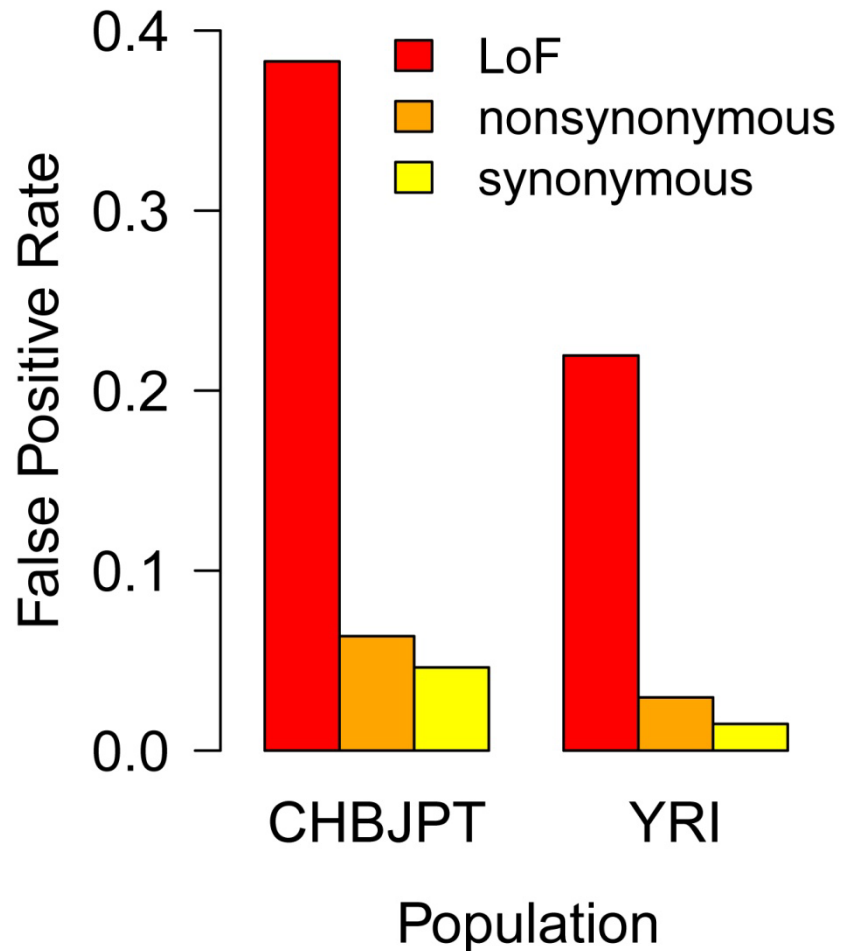
Reference/annotation artefact



Incorrect functional prediction



Functional variants are massively enriched for sequencing errors...

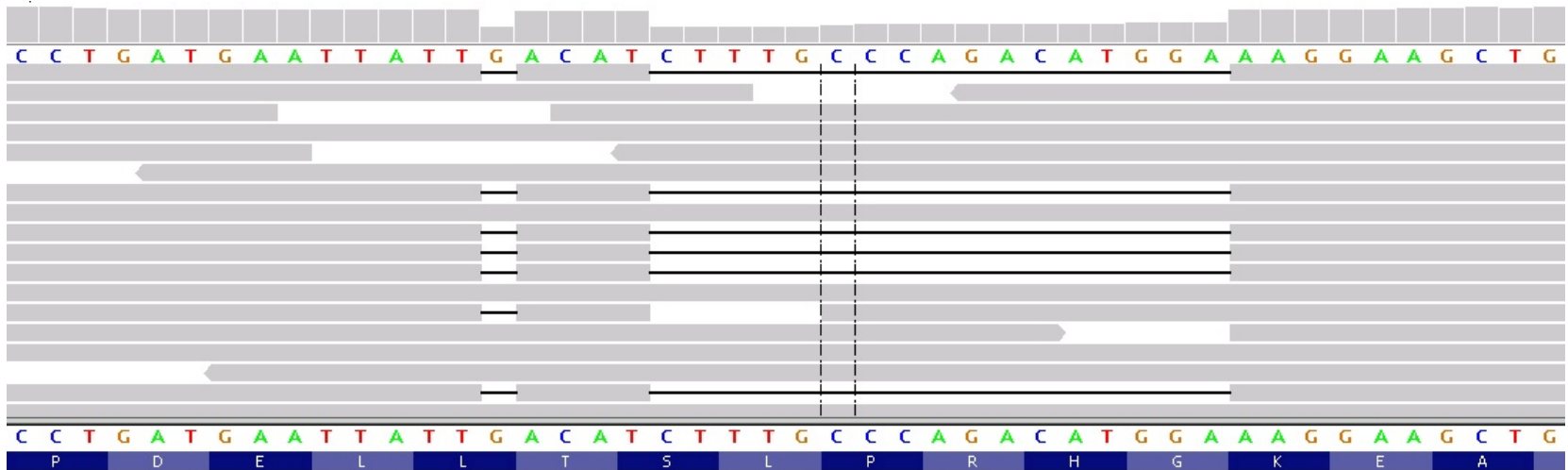


- Final low-coverage calls vs Illumina 1KGP12 chip
- Errors are mostly (but not entirely!) removed by new read-based filters

...and also massively enriched for annotation artefacts

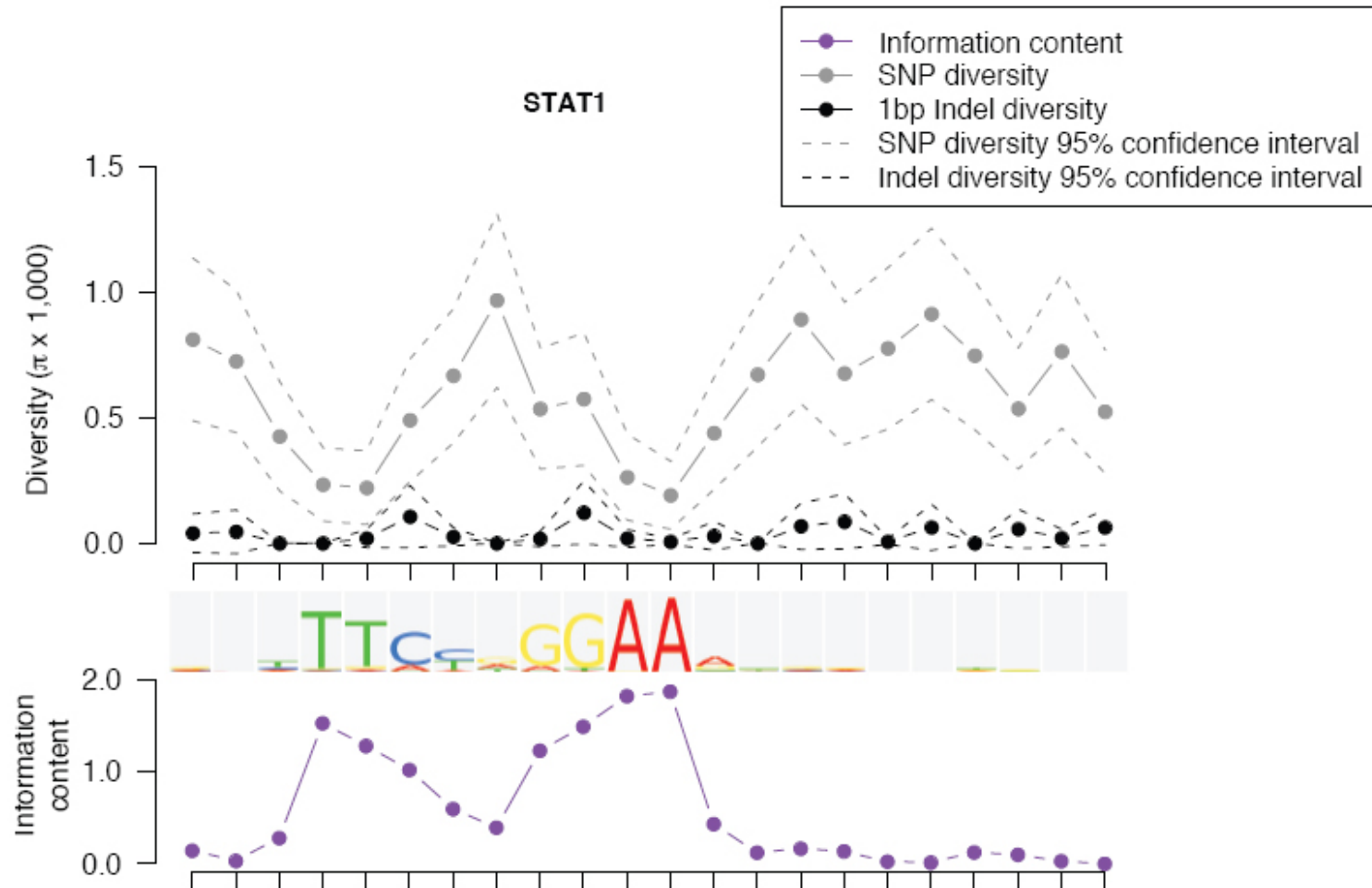
- Complete manual reannotation of ~700 LoF-containing genes by HAVANA team
- Developed automated pipeline to find other common errors
- Identified many artefacts:
 - 244 likely reference errors
 - 566 likely annotation errors
 - 313 variants unlikely to cause real LoF
- Errors fixed in Gencode, submitted to GRC

Predicting function will require including all variants



- Two apparent frameshift indels in the *CASP8AP2* gene (one 17 bp, one 1 bp) are in fact on the same haplotype
- Overall effect is in-frame deletion of six amino acids

Moving beyond coding regions



Translating genomics

Some lessons from the personal
genomics industry

Accurate genome interpretation is just the beginning

- Most individuals (including clinicians) have very little understanding of genetics
- We need good interfaces for effectively conveying genetic data to non-experts
- To date, academics have done a fairly poor job of building such interfaces

Lumigenix™



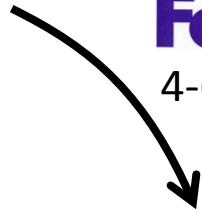
How it works



FedEx



FedEx
4-6 weeks



23andMe genetics just got personal.

welcome | ancestry | health | how it works | store

Choose the DNA test that's right for you.

Fill in your family tree.
Ancestry Edition, \$199 | Learn more

Buy Now

Take charge of your health.
Health Edition, \$429 | Learn more

Buy Now

Choose to have it all.
23andMe Complete, \$499

Buy Now

Find a disease or trait that we cover:

Select a Disease or Trait

Popular Topics:

- Type 2 Diabetes
- Rheumatoid Arthritis
- Prostatitis
- Breast Cancer
- Colorectal Cancer
- Prostate Cancer
- Celiac Disease
- Child's Disease
- Alzheimer's Disease
- Restless Legs Syndrome
- Age-related Macular Degeneration
- Parkinson's Disease
- Coumadin®/Warfarin Sensitivity
- Fragile X Carriers

News and Press

- 23andMe Makes New Discoveries in Genetics Using Novel, Web-based, Participant driven Methods
June 24, 2010
- Sergiy Birn's Search for a Parkinson's Cure
June 22, 2010
- 23andMe Enlists Informed Medical Decisions to Make Genetic Counseling

Scientific Resources and Principles

Research Initiatives

- Parkinson's Disease
- Pregnancy
- Sarcoma

Physician Resources

- Read our open letter to the medical community

Our Science

- Scientific Advisory Board
- Read our open letter to the scientific community

23andMe

Search

paternal line

Your Y chromosome DNA determines your paternal haplogroup. What is a haplogroup?

Map | History | Haplogroup Tree

Paternal Haplogroup: R1a1a*
R1a1a* is a subgroup of R1a1a, which is described below.
Locations of haplogroup R1a1a circa 500 years ago, before the era of intercontinental travel.

Haplogroup: R1a1a, a subgroup of **G1a1**
Age: 12,000 years
Region: Eastern Europe, Scandinavia, Southwestern Asia, India
Populations: Ukrainians, Indians, Poles
Highlight: R1a1a is the most common haplogroup in eastern Europe.

Your Family and Friends

- D2a1b** Japanese Person
- E1b1a1a1** Nigerian Person
- G2a** Lincoln De Kalb
- I1** Greg Mendel (Dad), Alan Mendel (Son), Ian Mendel (Son), Fred Mendel (Grandpa)
- I2b1** Ron Fisher (Grandpa)
- N** Chinese Person
- N1c1** Luke Justins
- O3a1** Ed Yong
- R1a1a*** Daniel MacArthur
- R1b1h2a1** Mark Henderson
- R1b1h2a1** Don Conrad

Sharing & Community

Compare Genes

Family Inheritance

23andMe Community

Genome Sharing

23andMe

Research Surveys (10)

Research Snippets

Research Initiatives

Research Discoveries

Human Prehistory Videos

Human Prehistory: Prologue

Out of (Eastern) Africa

How much is your genome worth?



1 million SNPs
\$99 + \$9/month
disease/ancestry



whole genome
\$9,500
raw data only

genomes unzipped

public personal genomics



www.genomesunzipped.org

◀ [Last chance to submit comments to the FDA about DTC genetics](#)

[My Genome Online - A Challenge To You](#) ▶

Calculating your Alzheimer's risk

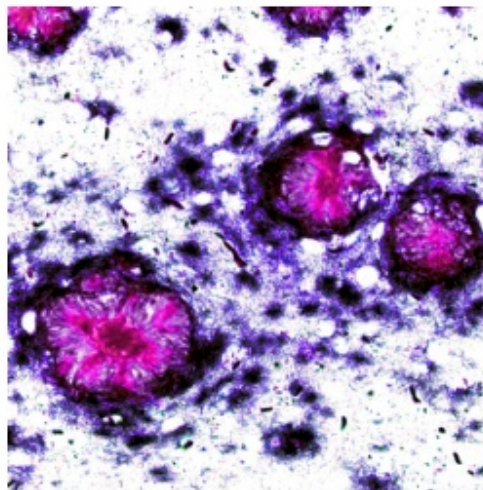
13/05/2011

Categories: [Disease Risk](#) and [DIY](#)

Written by [Luke Jostins](#)

[Edit](#)

For many diseases we have very little ability to determine who is at high or low risk; the risk factors are unreplicated, complicated, or understudied. However, for other diseases we can do much better. [Alzheimer's disease](#) is a form of senile dementia that is characterised by abnormal clustering of proteins in the brain (right). We know a number of important risk factors for Alzheimer's, and knowing your own risk factors may seriously change your estimate of the chance of developing the disease. But how can you calculate this risk?




About

Genomes Unzipped is a group blog providing expert, independent commentary on the personal genomics industry.

[About The Project](#)

[About The Contributors](#)

Subscribe

 [RSS](#)

 [Email](#)

 [Twitter](#)

Recent Posts

[Cracking non-coding variation, carrying cystic fibrosis, and more Alzheimer's prediction](#)

[Notes on the evidence for extensive RNA editing in humans](#)

Lumigenix™



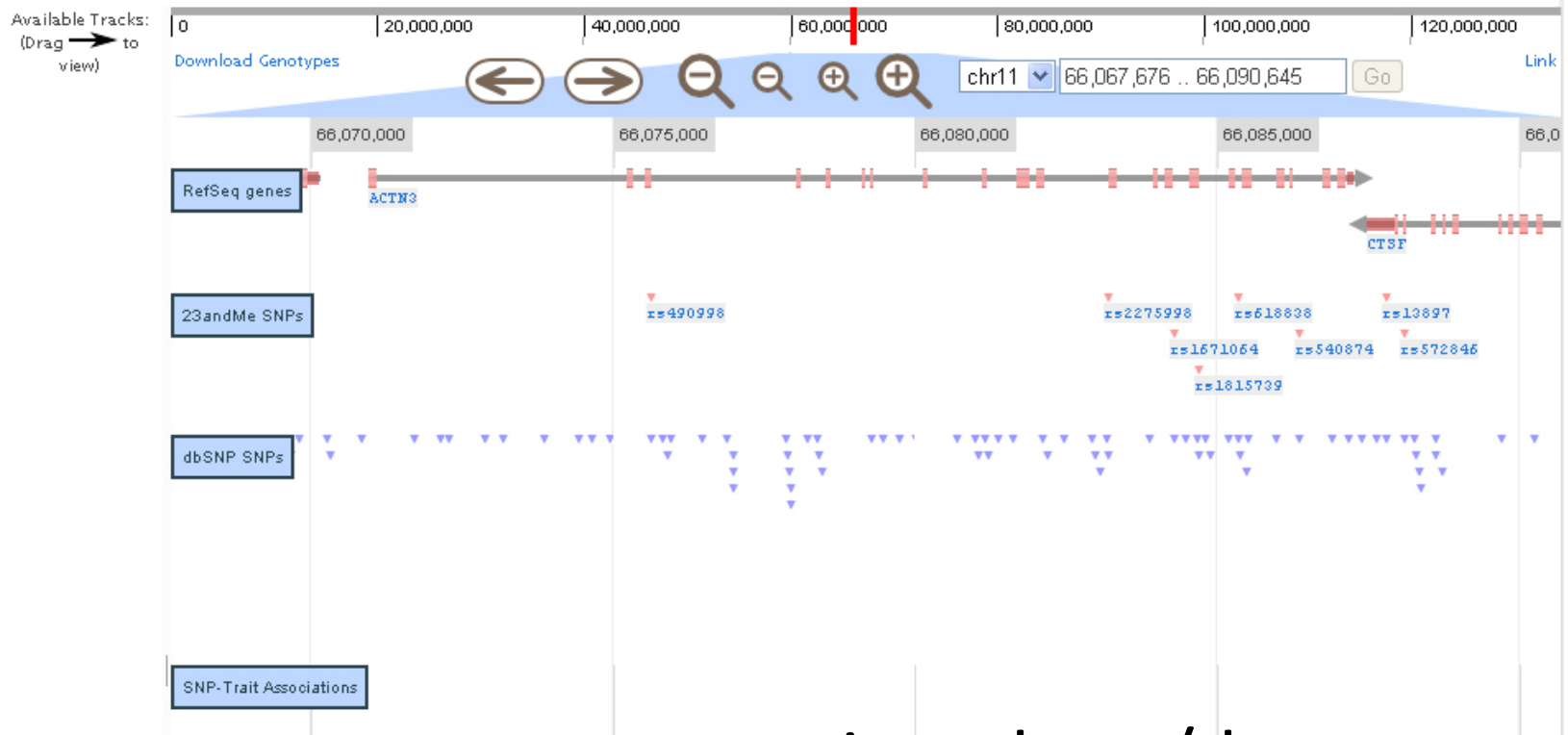
www.genomesunzipped.org

chr	pos	rsID	DGM001	LXJ001	DBV001	CFW001	KIM001	VXP001	JCB001	JXA001	JKP001	DFC001
chr2	1000713	rs6707252	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1003294	rs11894899	GG	GT	GG	GG	GG	GG	GT	GG	GT	GG
chr2	1004757	rs7574097	GG	GG	GG	GG	AG	GG	GG	AG	AG	GG
chr2	1004837	rs7574207	GG	AG	GG	GG	AG	GG	AG	AG	AA	GG
chr2	1009753	rs9309723	CC	CT	CT	CT	CC	CC	CT	CT	CC	CT
chr2	1017830	rs4971336	AA	CC	AC	AC	AC	AA	CC	CC	CC	AC
chr2	1027731	rs8179874	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1027783	rs4971329	AA	CC	AC	AC	AA	AC	CC	CC	CC	AC
chr2	1033352	rs4246558	CC	CC	CC	CC	CC	CC	CC	CT	CC	CC
chr2	1034617	rs4488692	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1036996	rs9309724	AA	AG	AA	AA	AA	AG	AG	AA	AG	AA
chr2	1038118	rs4629189	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1048688	rs4640416	CC	TT	CT	TT	CC	TT	TT	CT	CT	TT
chr2	1054700	rs7426276	GG	GG	GG	GG	AG	GG	GG	AG	GG	GG
chr2	1054885	rs4555377	AA	AG	AG	GG	AA	AA	AG	AA	AA	GG
chr2	1059949	rs4586665	TT	CC	CC	CT	TT	CC	CC	CT	CT	CC
chr2	1069320	rs4971432	TT	CT	CT	CT	TT	TT	CT	TT	TT	CC
chr2	1071594	rs10865542	TT	GT	GT	GT	TT	TT	GT	TT	TT	GG
chr2	1073995	rs7578498	TT	CT	CT	TT	TT	CT	CT	CT	CT	TT
chr2	1081787	rs11686452	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1087795	rs6716119	GG	AG	AG	AG	GG	GG	AG	GG	GG	AA
chr2	1089597	rs4971441	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1091080	rs4482521	CC	AA	AA	AC	CC	AA	AA	CC	AA	AA
chr2	1091127	rs12469446	AA	GG	GG	AG	AA	GG	GG	AA	GG	GG
chr2	1106648	rs4971463	GG	AG	AG	AG	GG	GG	AG	GG	GG	AA
chr2	1109365	rs7565112	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1110566	rs4971465	AA	GG	GG	AG	AA	GG	GG	AA	GG	GG
chr2	1118570	rs4491751	CC	AC	AC	AC	CC	AC	AC	CC	CC	AA
chr2	1133477	rs4998208	AA	AG	GG	GG	AA	AG	AG	GG	AG	GG
chr2	1140095	rs7579535	CC	CC	CC	AC	CC	CC	CC	AC	AC	CC
chr2	1142434	rs4246561	TT	TT	CT	TT	CT	TT	TT	CT	CT	TT
chr2	1143552	rs4971396	GG	GG	GT	GT	GG	GG	GG	TT	GT	GG
chr2	1147068	rs4971397	TT	TT	GT	GT	GT	TT	TT	GG	GG	TT
chr2	1189920	rs4461289	AG	GG	AG	GG	AG	AA	GG	GG	AA	GG
chr2	1193337	rs4446093	GG	AG	AG	GG	GG	GG	AG	GG	GG	AA
chr2	1194796	rs6751090	CC	CT	CT	CC	CC	CC	CC	CC	CC	CC
chr2	1206395	rs4971418	AG	GG	AG	AA	AG	AA	AG	AG	AA	GG
chr2	1257125	rs7568265	CT	CT	TT	TT	CC	TT	TT	CT	TT	TT
chr2	1258693	rs7369772	CC	CC	CC	CC	CT	CC	CC	CT	CC	CC
chr2	1262038	rs11674954	TT	TT	TT	TT	TT	GG	TT	TT	GG	TT
chr2	1263027	rs7424980	CC	TT	CT	CC	CC	TT	CT	CC	TT	TT

www.genomesunzipped.org/data

The Genomes Unzipped Browser

The Unzipped Browser was developed using [JBrowse](#) by Joe Pickrell and Luke Jostins. [Click here](#) for information on how to use it. All chromosome positions are relative to build 36.



www.genomesunzipped.org/data

Lessons learned so far

- Raw data: extremely accurate for SNP genotyping tests
 - same individual, 3 tests: **1.4-4.5 errors/100K SNPs**
 - 23andMe trios: **2.6-4.5 errors/100K SNPs**
- Risk prediction: good, but could be improved
 - some potentially worrying variation in algorithms between companies
 - larger variation in background risk estimates
 - SNP lists can be overly conservative/out of date
- Interfaces: generally intuitive, responsible

Risks and benefits

Table 3. Primary Outcome Measures before and after Receipt of Results of Genetic Testing for 2037 Subjects Who Completed Follow-up.*

Outcome Measure	Baseline Score	Follow-up Score	P Value†
Anxiety	35.2±9.6	34.6±10.0	0.80
Dietary fat intake	16.0±7.9	15.2±7.5	0.89
Exercise‡	28.6±23.0	28.6±22.9	0.61

* Plus-minus values are means ±SD. The assessment tools and ranges of scores for each category are listed in Table 2.

† All P values were calculated with the use of the Wilcoxon signed-rank test after adjustment for covariates.

‡ A total of 1943 subjects were included in this analysis.

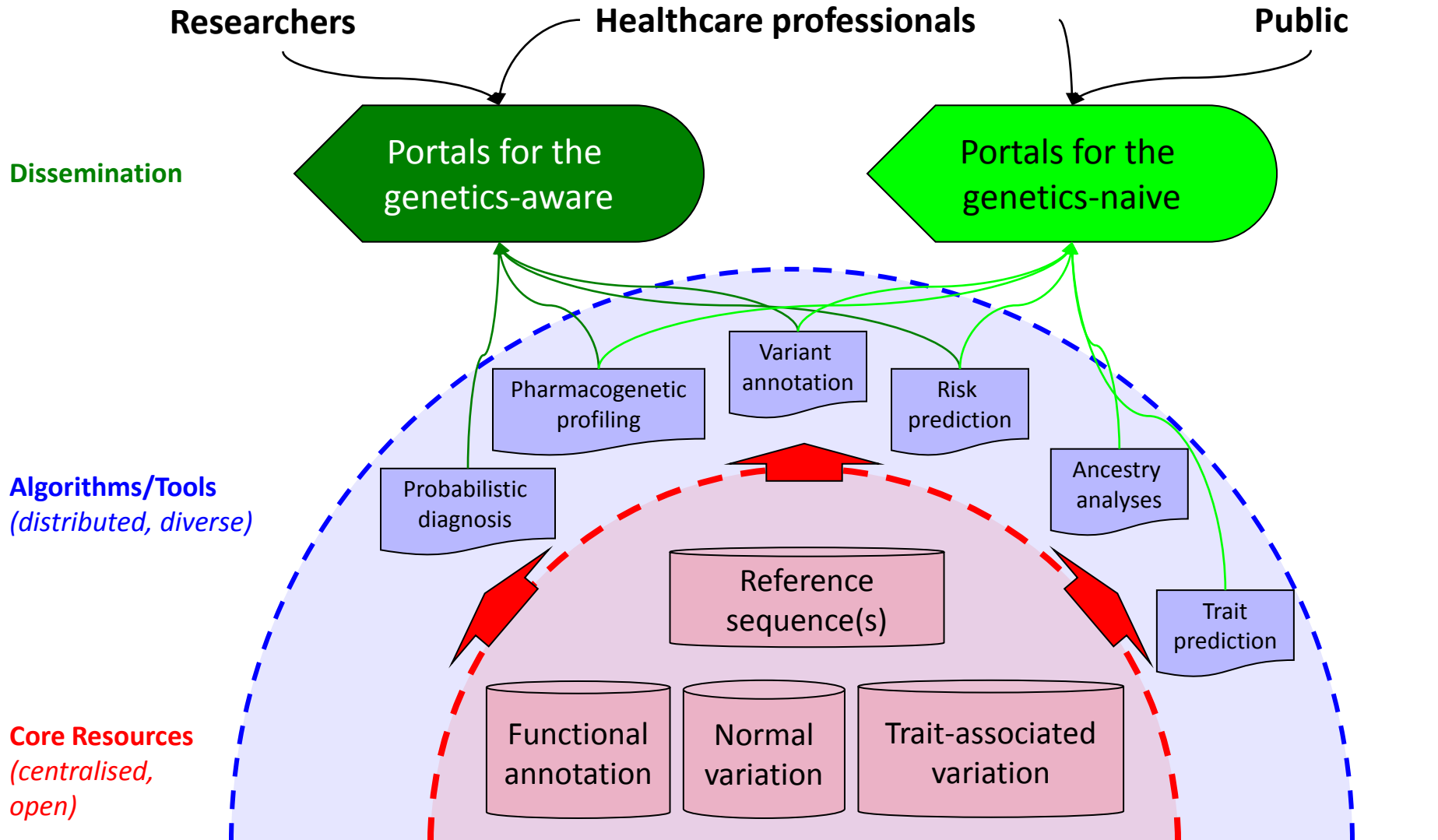
- Early adopters are well-educated about genetics
- Genetic risk predictions don't increase anxiety...
- ...but they don't seem to positively change behaviour either
- More study needed, especially for more serious conditions and later adopters

Bloss *et al.* 2011.
NEJM 473:43–49

Ethical challenges

- **Clinicians as gatekeepers vs empowering individuals**
 - doctors more likely to create lifestyle change?
 - can health system survive with gatekeeper model?
 - importance of individual autonomy
- **Regulation vs innovation**
 - some companies offer bogus products
 - strong regulation would hinder innovation
 - could just better enforce existing legislation

Moving forward: what happens next?



adapted from Matt Hurles



Thanks!

1000 Genomes Functional Annotation

Chris Tyler-Smith, **Sanger**

Suganthi Balasubramanian, Lukas Habegger, Mark Gerstein, **Yale**

Adam Frankish and Jennifer Harrow, **Sanger**

James Morris, Luke Jostins and Jeff Barrett, **Sanger**

Ni Huang, Klaudia Walter, Don Conrad and Matt Hurles, **Sanger**

Kees Albers and Richard Durbin, **Sanger**

Joe Pickrell and Jonathan Pritchard, **U. Chicago**

Stephen Montgomery and Manolis Dermitzakis, **U. Geneva**

Mark DePristo and Eric Banks, **Broad Institute**

Genomes Unzipped

Luke Jostins, Dan Vorhaus, Kate Morley, Caroline Wright,

Vincent Plagnol, Jeff Barrett, Joe Pickrell, Don Conrad,

Carl Anderson, Ilana Fisher, Jan Aerts



Daniel MacArthur

www.genomesunzipped.org

Twitter: @dgmacarthur

