Whole Genome Analyses (WGA)

John Hardy (hardyj@mail.nih.gov)

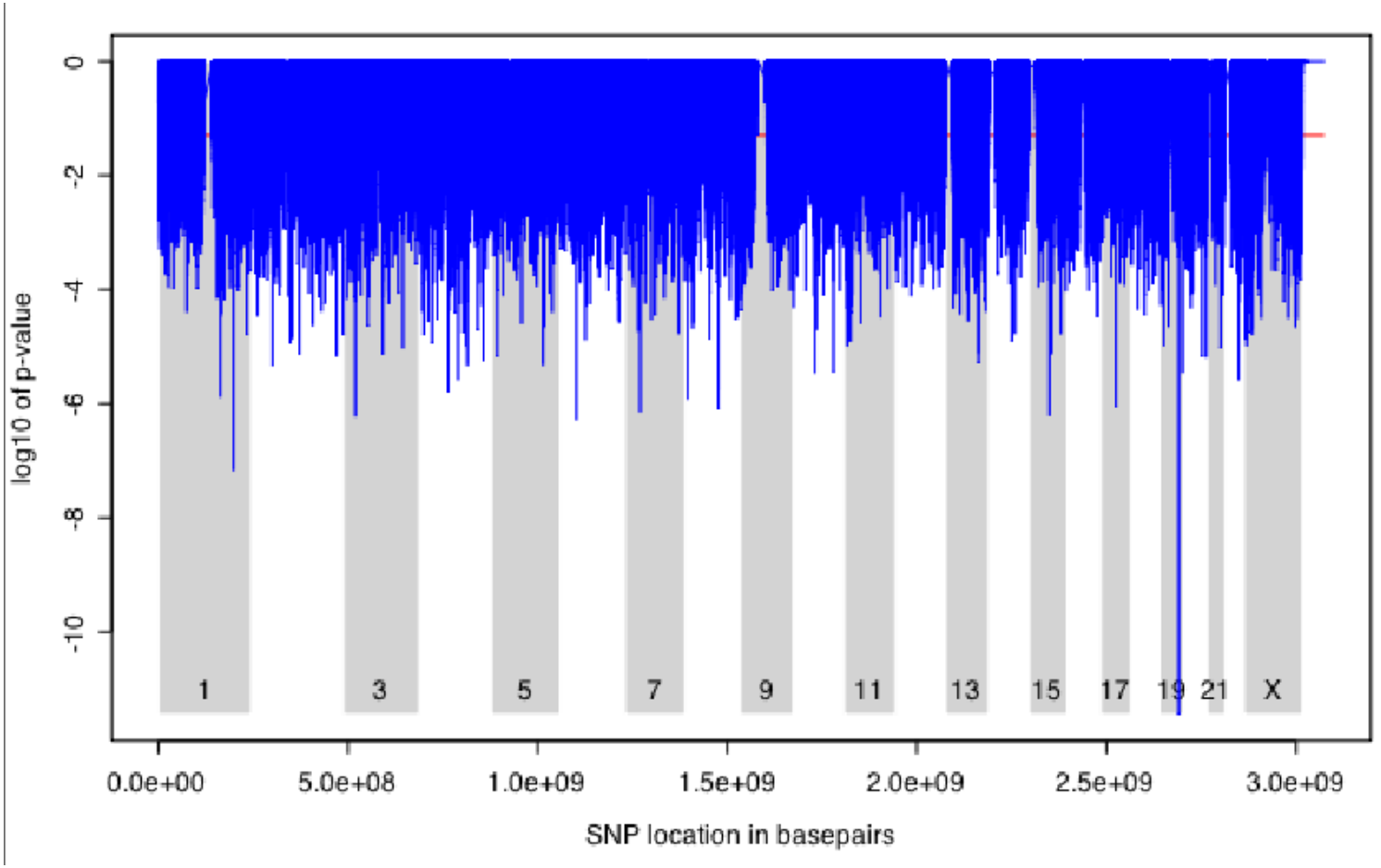
Andy Singleton (singleta@mail.nih.gov)

Laboratory of Neurogenetics, NIA/NIH,
Building 35, Bethesda Campus

Whole Genome Study For Alzheimer's Disease

500K Affymetrix Array

(presently underpowered ~180 cases and controls: collaboration with TGen: now 800 cases and controls: data analysis ongoing)



Our Studies (Illumina Bead Station)

- Whole Genome Association Analysis of Parkinson's Disease
  - 276 Cases and Controls NIA/NINDS funded (completed: in follow up)
- Whole Genome Association Analysis of Ischemic Stroke
  - 276 Cases and Controls NIA/NINDS funded (completed: in follow up)
- Whole Genome Association Analysis of ALS
  - 276 Cases and Controls NIA/NINDS/ALSA funded (in progress)
- Whole Genome Analysis of Haplotypic Brain Expression
  - 300 Control Brains NIA/TGen funded (in progress)
- Whole Genome Analysis of African Americans
  - 200 from the HANDLs Study (Baltimore, NIA/Michele Evans PI) (in progress: more planned)

Three Surprises:-

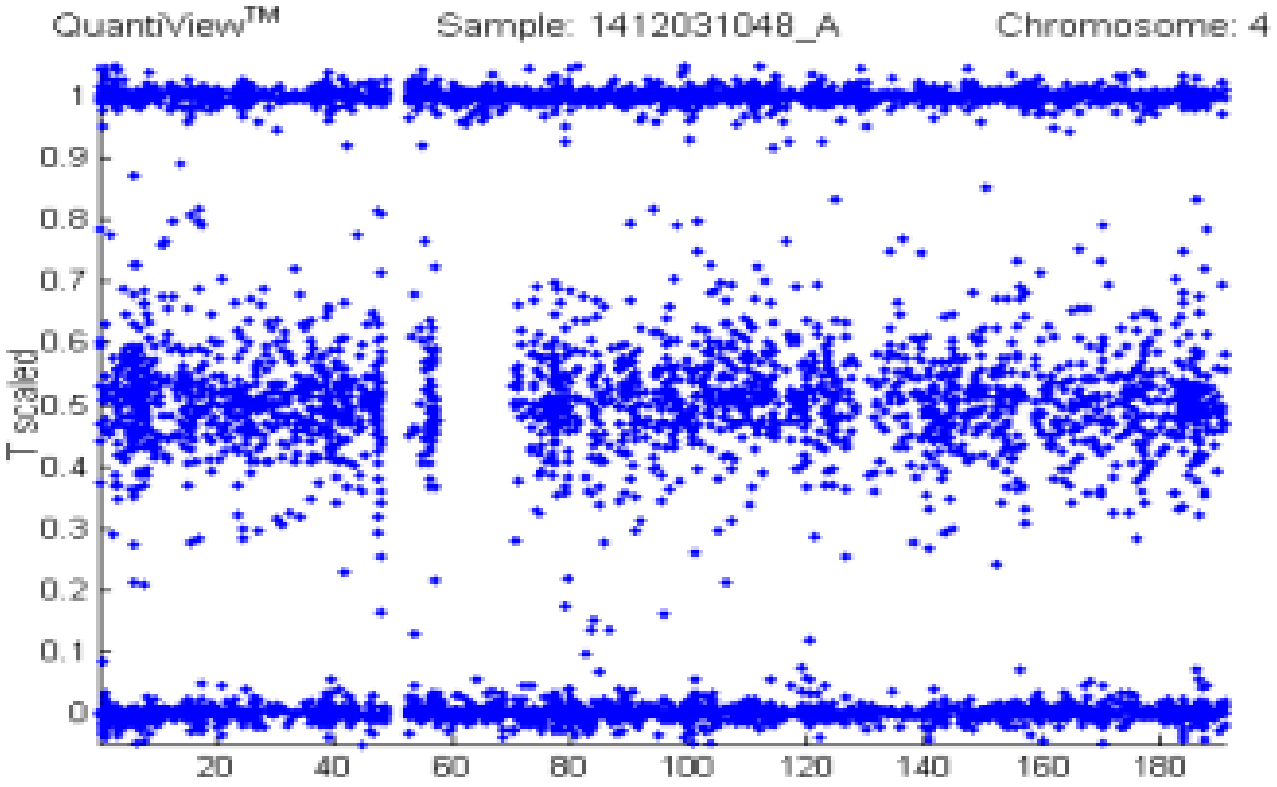
- Data quality: routinely >99% of data
- In North American Caucasian Controls, ~10% showed extensive homozygosity (not true of African Americans)
  - (parents were ~2nd-4th cousins)
- In North American Controls, ~9% had significant structural variability (large insertions and deletions)
  - (some was cell line specific, but much was not: what is "normal"?)

- Realized both homozygosity and structural variability could be disease-related

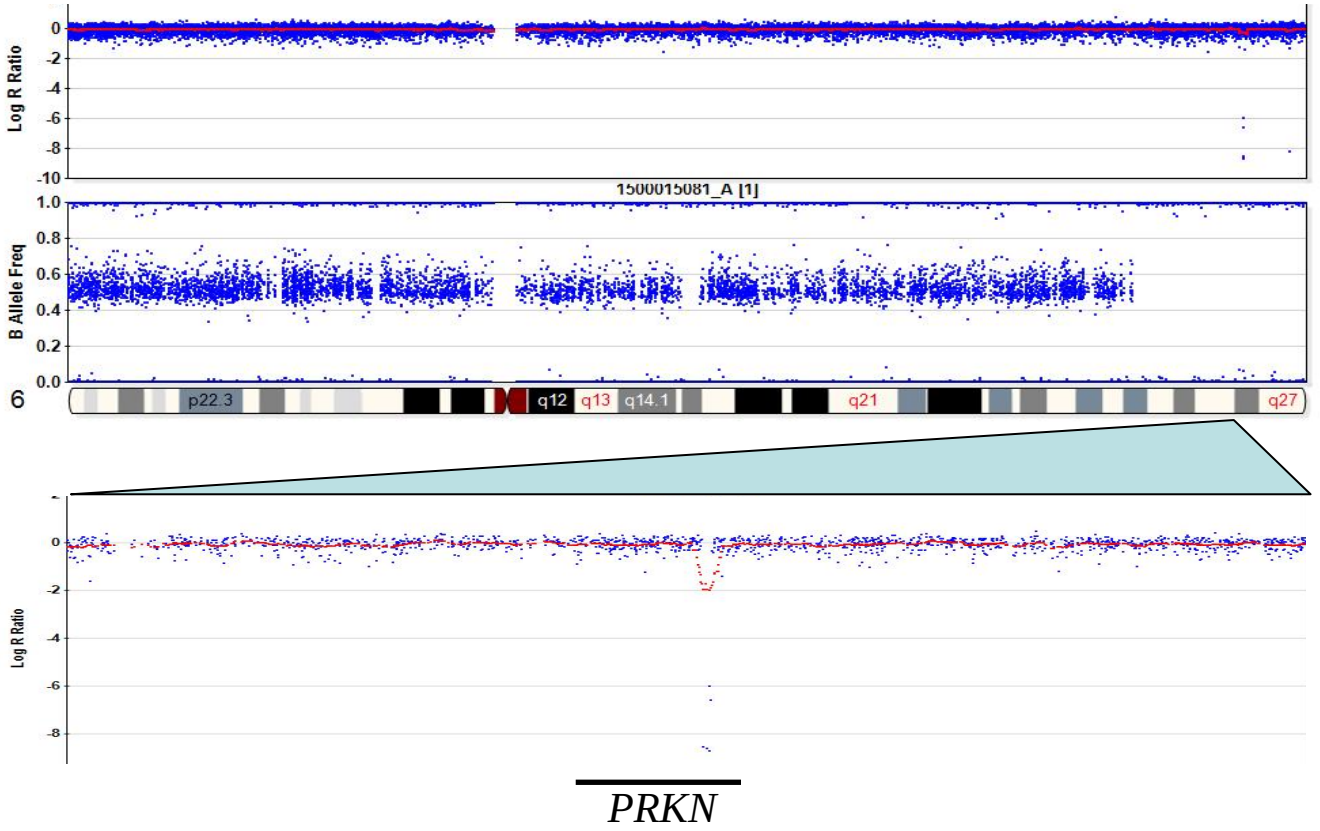
http://ccr.coriell.org/ninds/

Loss of Heterozygosity

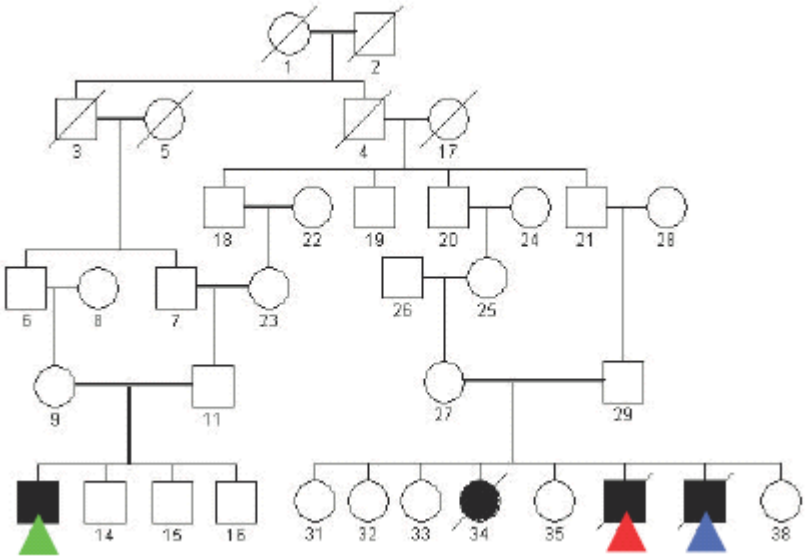
(10% North American Controls Show Evidence for Consanguinity

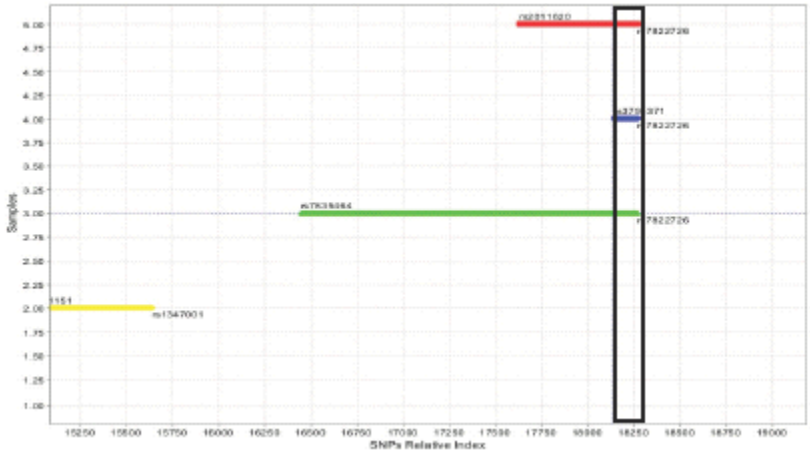


Detection of homozygous parkin deletion causing Parkinson's disease



Homozygosity Mapping on Infinium 300K: one hit linkage for a new disease

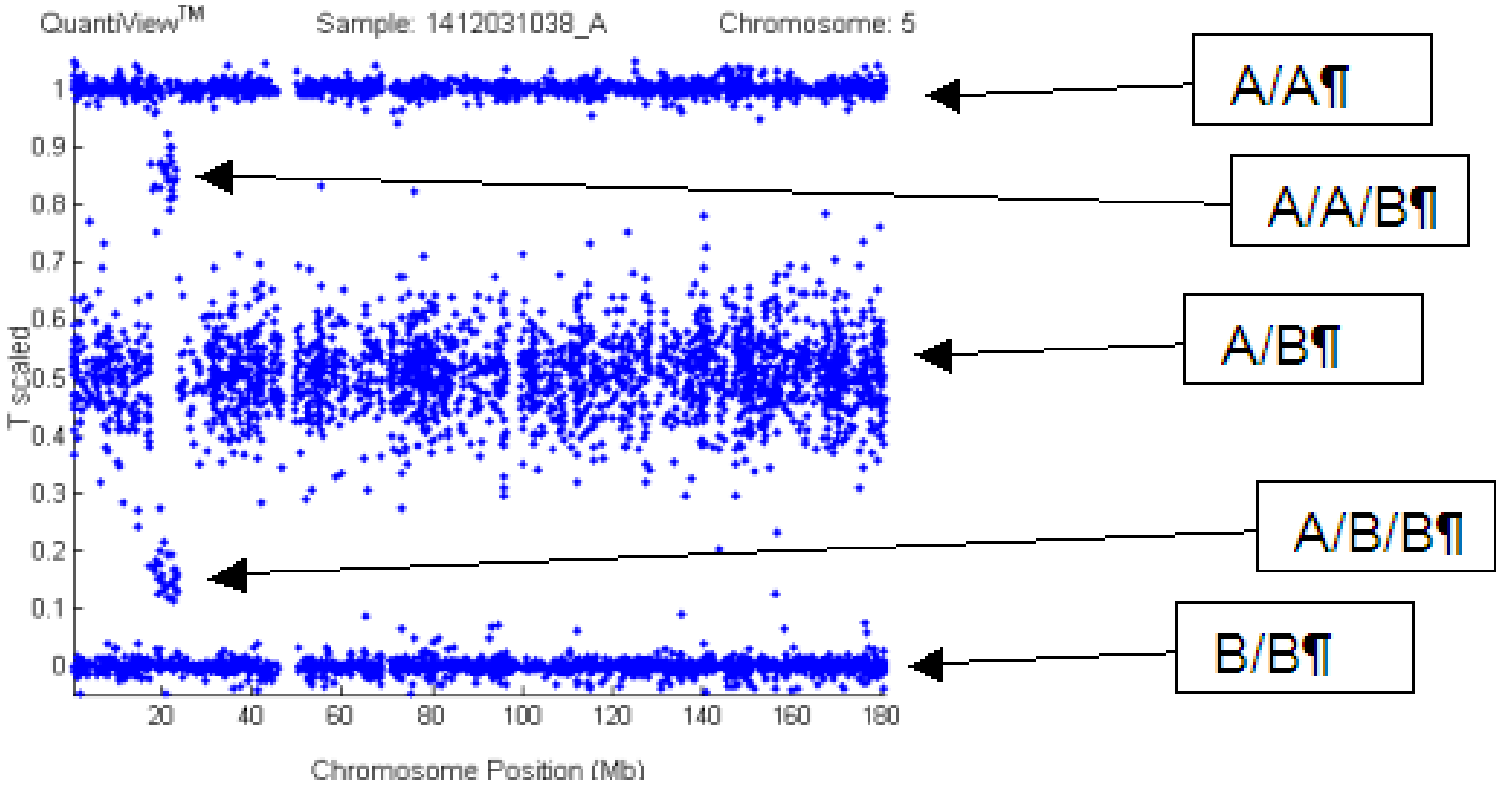




- Recessive young onset ataxia
- Single segregating region ~1Mb

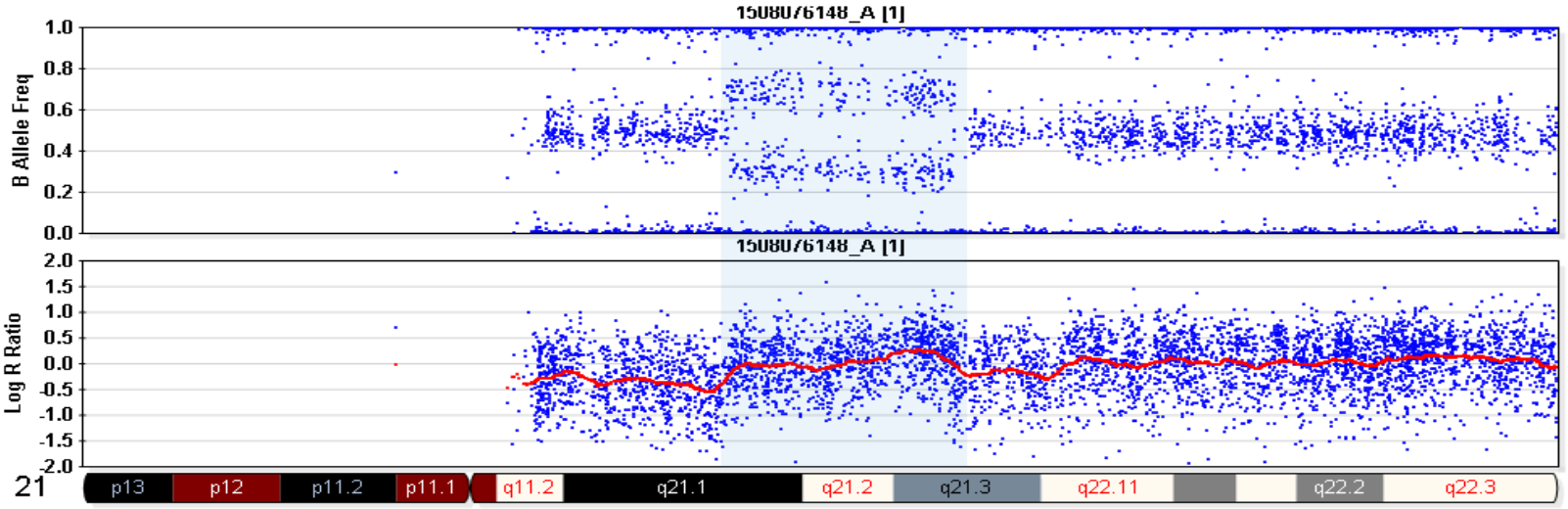
Structural Alterations

Chromosome 5 Control Male of 65 years



APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy

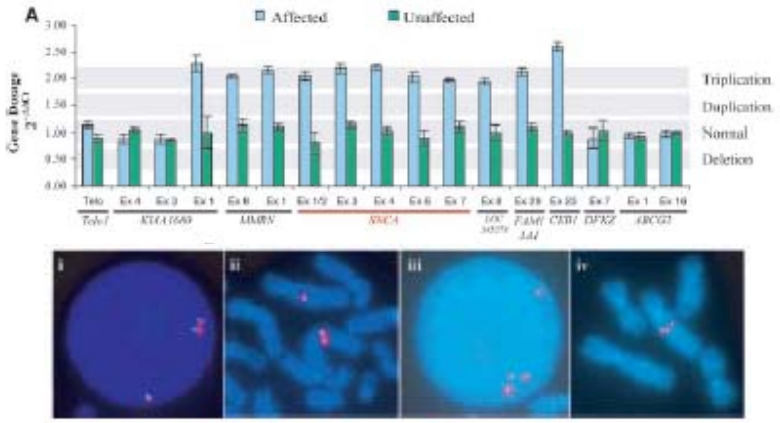
Anne Rovelet-Lecrux[1], Didier Hannequin[1,2], Gregory Raux[1], Nathalie Le Meur[3], Annie Laquerrière[4], Anne Vital[5], Cécile Dumanchin[1], Sébastien Feuillette[1], Alexis Brice[6], Martine Vercelletto[7], Frédéric Dubas[8], Thierry Frebourg[1] & Dominique Campion[1,9]



BREVIA

α-Synuclein Locus Triplication Causes Parkinson's Disease

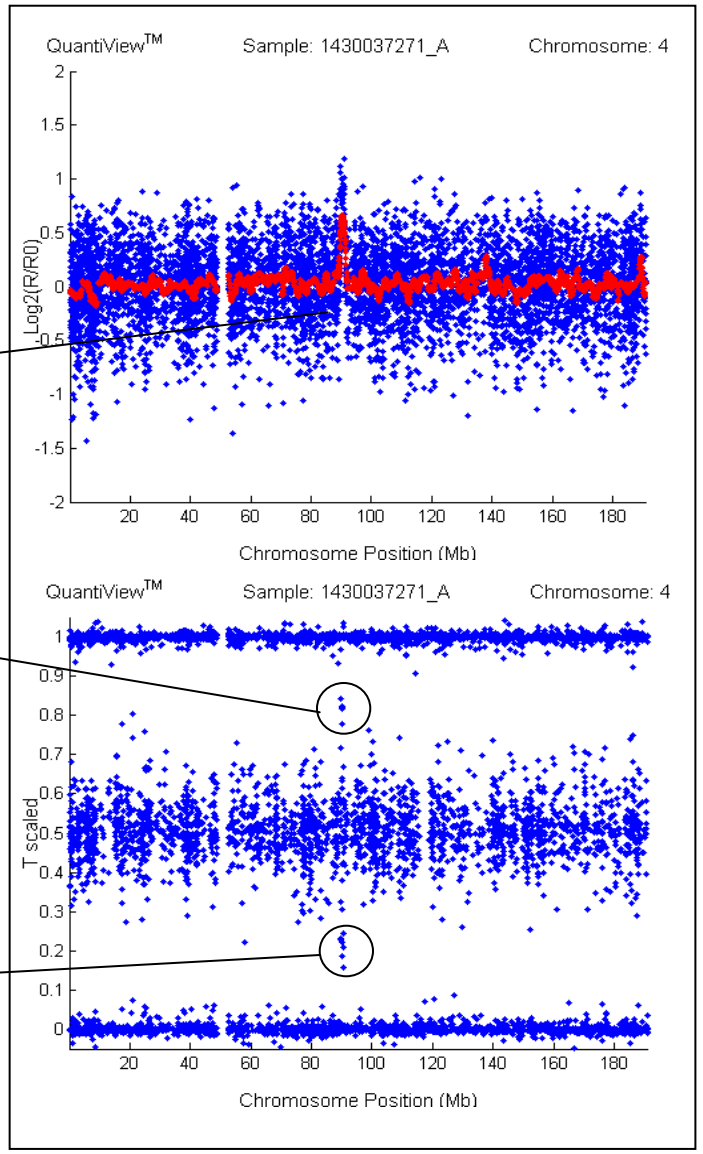
A. B. Singleton,[1]*† M. Farrer,[4]† J. Johnson,[1] A. Singleton,[2] S. Hague,[1] J. Kachergus,[4] M. Hulihan,[4] T. Peuralinna,[1] A. Dutra,[3] R. Nussbaum,[2] S. Lincoln,[4] A. Crawley,[2] M. Hanson,[1] D. Maraganore,[5] C. Adler,[6] M. R. Cookson,[1] M. Muenter,[6] M. Baptista,[1] D. Miller,[1] J. Blancato,[7] J. Hardy,[1] K. Gwinn-Hardy[2]



TRIP

A/A/A/B

A/B/B/B



Whole Genome Association Analyses

- Population choice is important
- Data handling is not trivial
  - (our lab has been generating ~6,000,000 a day for 6 months) and now has ~1,000,000,000 genotypes)

Population Choice: LRRK2 and PD

Genetic screening for a single common *LRRK2* mutation in familial Parkinson's disease

*William C Nichols, Nathan Pankratz, Dena Hernandez, Coro Paisán-Ruíz, Shushant Jain, Cheryl A Halter, Veronika E Michaels, Terry Reed, Alice Rudolph, Clifford W Shults, Andrew Singleton, Tatiana Foroud, for the Parkinson Study Group-PROGENI investigators**

A common *LRRK2* mutation in idiopathic Parkinson's disease

William P Gilks, Patrick M Abou-Sleiman, Sonia Gandhi, Shushant Jain, Andrew Singleton, Andrew J Lees, Karen Shaw, Kailesh P Bhatia, Vincenzo Bonifati, Niall P Quinn, John Lynch, Daniel G Healy, Janice L Holton, Tamas Revesz, Nicholas W Wood

A frequent *LRRK2* gene mutation associated with autosomal dominant Parkinson's disease

*Alessio Di Fonzo, Christan F Rohé, Joaquim Ferreira, Hsin F Chien, Laura Vacca, Fabrizio Stocchi, Leonor Guedes, Edito Fabrizio, Mario Manfredi, Nicola Vanacore, Stefano Goldwurm, Guido Breedveld, Cristina Sampaio, Giuseppe Meco, Egberto Barbosa, Ben A Oostra, Vincenzo Bonifati, and the Italian Parkinson Genetics Network**

Am. J. Hum. Genet. 76:000–000, 200.

Report

Identification of a Novel *LRRK2* Mutation Linked to Autosomal Dominant Parkinsonism: Evidence of a Common Founder across European Populations

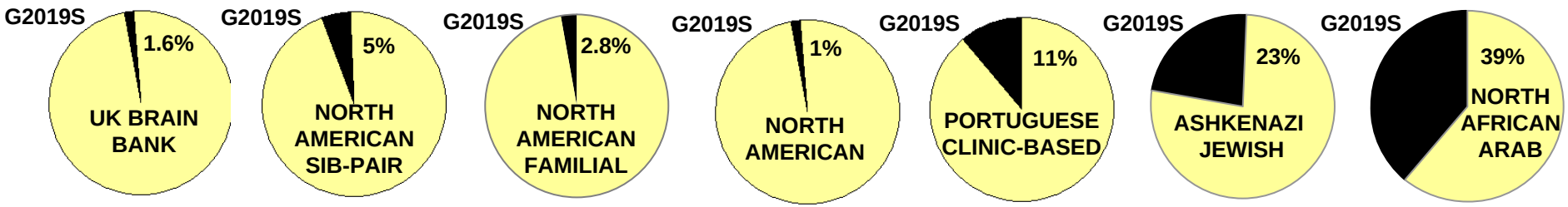
Jennifer Kachergus,[1,*] Ignacio F. Mata,[1,*] Mary Hulihan,[1] Julie P. Taylor,[1] Sarah Lincoln,[1] Jan Aasly,[3] J. Mark Gibson,[5] Owen A. Ross,[1,6] Timothy Lynch,[7,8] Joseph Wiley,[7,8] Haydeh Payami,[9] John Nutt,[10] Demetrius M. Maraganore,[11] Krzysztof Czyzewski,[12] Maria Styczynska,[13] Zbigniew K. Wszolek,[2] Matthew J. Farrer,[1] and Mathias Toft[1,4]

LRRK2 mutations and Parkinsonism

Mathias Toft, Ignacio F Mata, Jennifer M Kachergus, Owen A Ross,

- Gly2019Ser, alters conserved aa within the kinase activation loop
- Could this be constitutively active? Involved in target recognition?

```
H. sapiens      AAIIAKIADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIY
R. norvegicus   AAIIAKIADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIY
M. musculus     AAIIAKIADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIY
X. laevis       SAIIAKIADYGIAQYCCRMGIKTSEGTPGFRAPEVARGNVIY
D. melanogaster NLVHIKIADYGISRQTAPSGAKGFGGTEGFMAPEIIRYNG--
A. mellifera    HPVHVKVADYGISRLTLPTGAKGFGGTEGFMAPEIIKYNGEE
```



Whole Genome Analyses of 276 US PD cases

Raphael Gibbs; gibbsr@mail.nih.gov

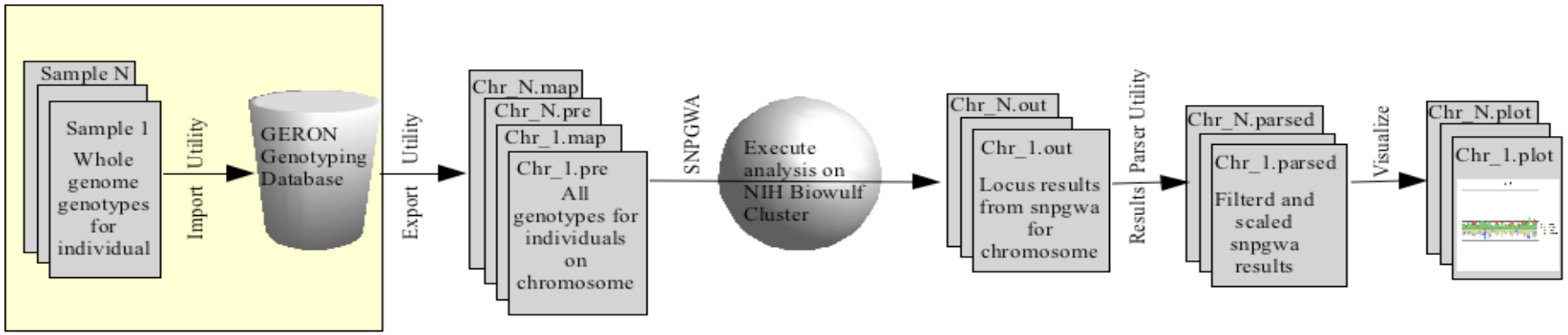
LNG Data Management Software

GERON (Raph Gibbs, Andy Singleton):

- Clinical data, sample tracking and genotype storage

SNP GWA (Carl Langefeld, Matt Stiegert at Wake Forest):

- Analysis (Hardy Weinberg; dominant; recessive; additive; haplotypic)



Whole Genome Data

- Association… sure…
  - Probably OK down to ~ λs of 1.5-2.0
- Homozygosity mapping
  - In kindreds and in populations
- Insertion/deletion cataloguing
  - What is “normal” and what is pathologic?
- Genetic Ancestry
  - Different populations
- Whole genome diplotype/expression correlation
  - 300 Human Cortical Samples (500K Affy SNP data from TGen: 24K Illumina Expression data)
- Cell lines
  - What are (stem) cell lines like? (a mess)

Whole Genome Analyses

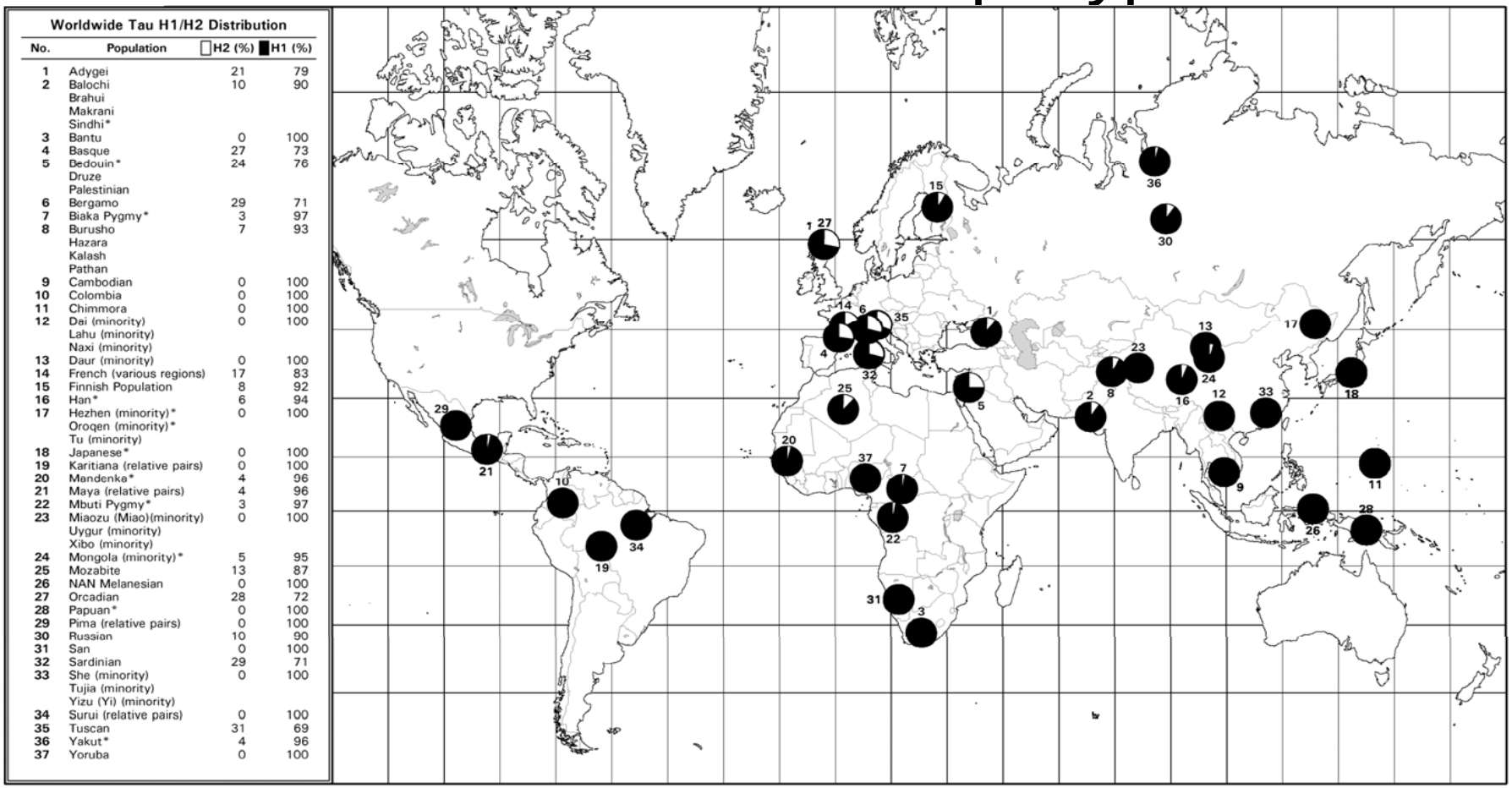
- 1) Whole genome associations will pick up alleles of large effect
- 2) Tell you what is *NOT* there (valuable to NIH) (within the bounds of the study design and population of course)
- 3) They are *ADDITIVE:* allowing studies to be pooled easily
- 4) Identify quickly and easily insertions/deletions (need a normal catalogue)
- 5) Enables homozygosity mapping in “outbred” populations (parkin example)
- 6) Enables cell lines to be characterized (stem cells etc)
- 7) Enables genotype/expression correlations for cis/(trans?) correlations of gene expression (valuable for complex trait genetic associations)

Many Genetic Associations Likely to Reflect Differences in Expression:-

- Three Common Haplotypes of Gene ω
  - ωA gives ↑
  - ωB gives →
  - ωC gives ↓
- 300 Human Control Cortices: Fully Genotyped (Affy 500K array): and Full Expression Array (Illumina 24K expression Array)
- If disease is associated with High Expression, genetic association should be seen in ωAA homozygotes and protection in ωCC with ωBB intermediate

Need to have proper diversity to know what is normal and to find admixture (CEPH Diversity series)

Distribution of Tau Haplotype

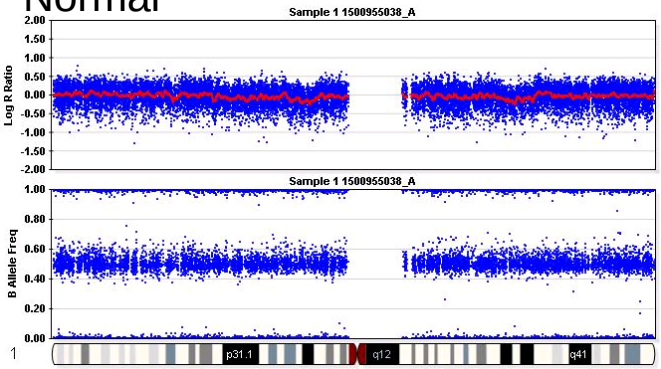


Worldwide Tau H1/H2 Distribution

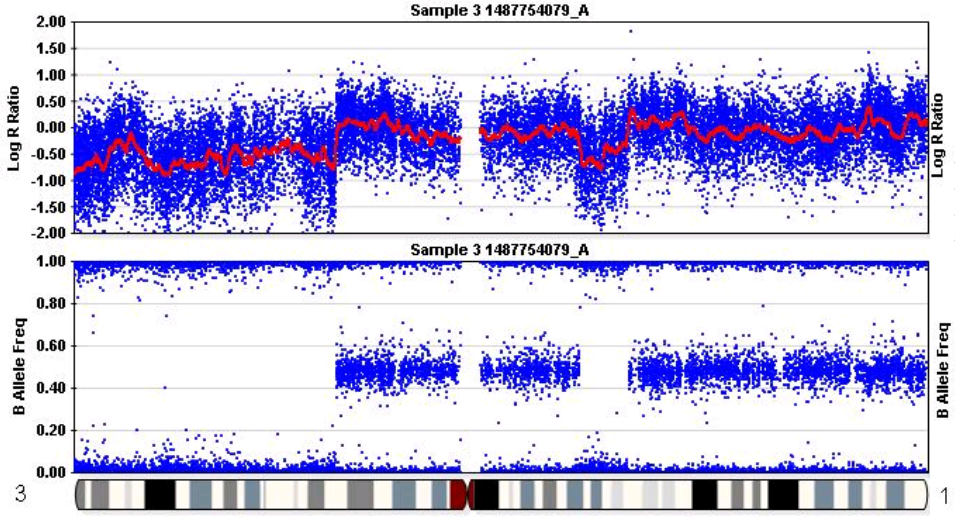
No.	Population	H2 (%)	H1 (%)
1	Adygei	21	79
2	Balochi	10	90
	Brahui		
	Makrani		
	Sindhi*		
3	Bantu	0	100
4	Basque	27	73
5	Bedouin*	24	76
	Druze		
	Palestinian		
6	Bergamo	29	71
7	Biaka Pygmy*	3	97
8	Burusho	7	93
	Hazara		
	Kalash		
	Pathan		
9	Cambodian	0	100
10	Colombia	0	100
11	Chimmora	0	100
12	Dai (minority)	0	100
	Lahu (minority)		
	Naxi (minority)		
13	Daur (minority)	0	100
14	French (various regions)	17	83
15	Finnish Population	8	92
16	Han*	6	94
17	Hezhen (minority)*	0	100
	Orogen (minority)*		
	Tu (minority)		
18	Japanese*	0	100
19	Karitiana (relative pairs)	0	100
20	Mandenka*	4	96
21	Maya (relative pairs)	4	96
22	Mbuti Pygmy*	3	97
23	Miaozu (Miao)(minority)	0	100
	Uygur (minority)		
	Xibo (minority)		
24	Mongola (minority)*	5	95
25	Mozabite	13	87
26	NAN Melanesian	0	100
27	Orcadian	28	72
28	Papuan*	0	100
29	Pima (relative pairs)	0	100
30	Russian	10	90
31	San	0	100
32	Sardinian	29	71
33	She (minority)	0	100
	Tujia (minority)		
	Yizu (Yi) (minority)		
34	Surui (relative pairs)	0	100
35	Tuscan	31	69
36	Yakut*	4	96
37	Yoruba	0	100

Common cell lines

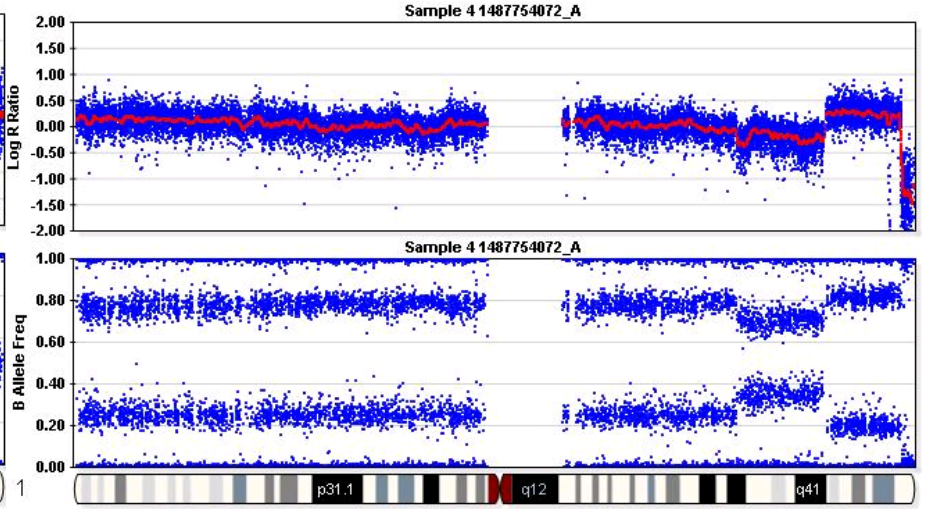
Normal



M17

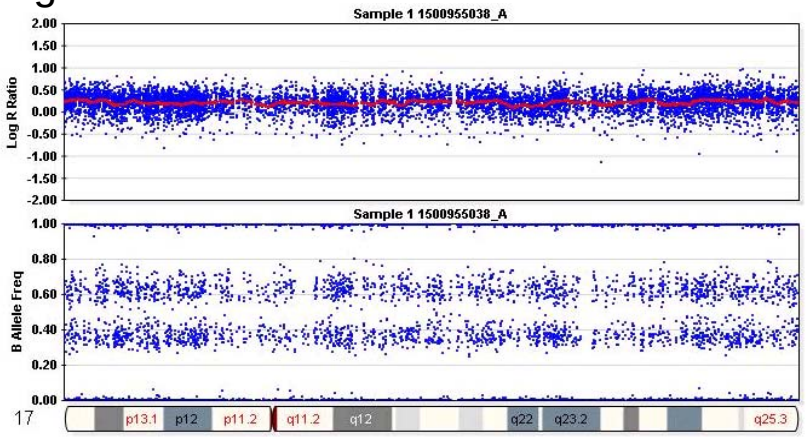


HEK293

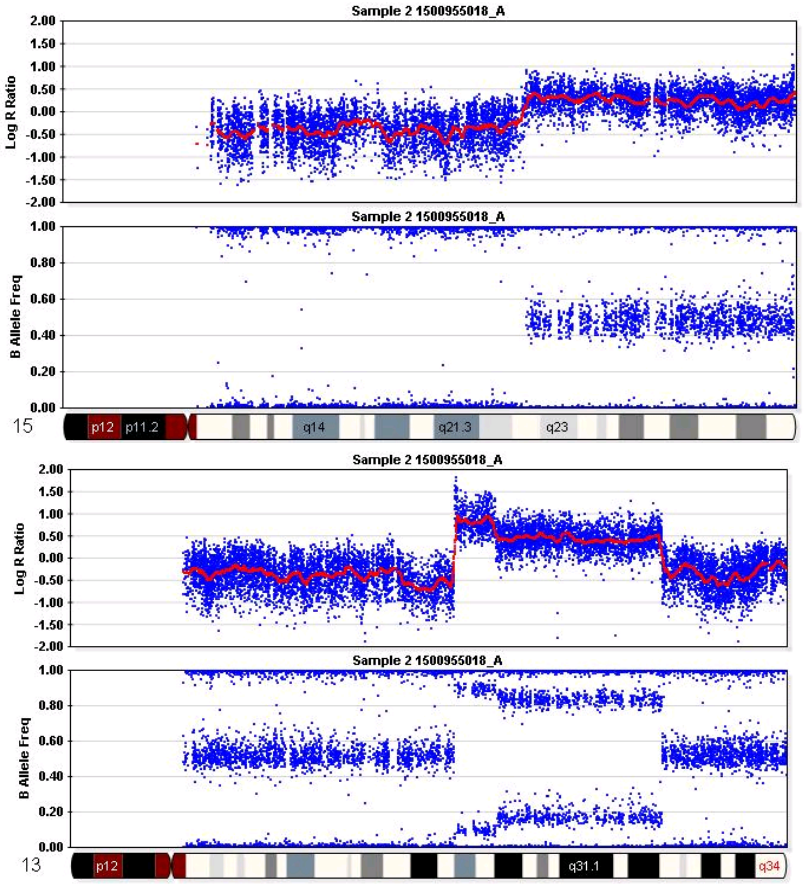


Federally approved Stem Cell Lines

Bg01P3SFF2



293F7



The Future

- All genetic samples will have WG data (now ~$700 a sample)
- All cell lines will have WG data
- Need a catalogue of variability
- Need real diversity
- Perhaps start with homogenous cohorts for whole genome association analyses