Royal Veterinary College
University of London

A comparison of methodologies to locate an autosomal recessive genetic diseases using SNP chip genotypes

Geoff Pollott

Background

- Autosomal recessive conditions arise in most breeds from time to time and may cause genetic diseases
- Genome-wide SNP data may be used to locate this area by looking for characteristic patterns of homozygosity
- Several methods have been proposed
- This paper compares 5 methods which were used to locate the well-documented Lavender Foal Syndrome (Brooks *et al*., 2010. *PLoS Genetics*, 6: e1000909).

RVC

Lavender foal syndrome (Brooks et al., 2010)

- Lethal autosomal recessive condition
- Single base pair deletion in Exon 30 of the MYO5A gene
- Located at position 138,235,715 on ECA1
- Used 6 affected cases and 30 unaffected relatives
- Genotyped using the EquineSNP50 chip

RVC

Methods to be compared

Chi-squared - genotypic test in PLINK (Purcell et al., 2007. *American Journal of Human Genetics*, 81:559-575).

ASSIST and ASSHOM (Charlier et al., 2008. *Nature Genetics*, 40:449-454).

Autozygosity by difference (ABD) (Pollott, 2012. EAAP Bratislava)

PLINK –homozyg option (Purcell et al., 2007. *American Journal of Human Genetics*, 81:559-575).

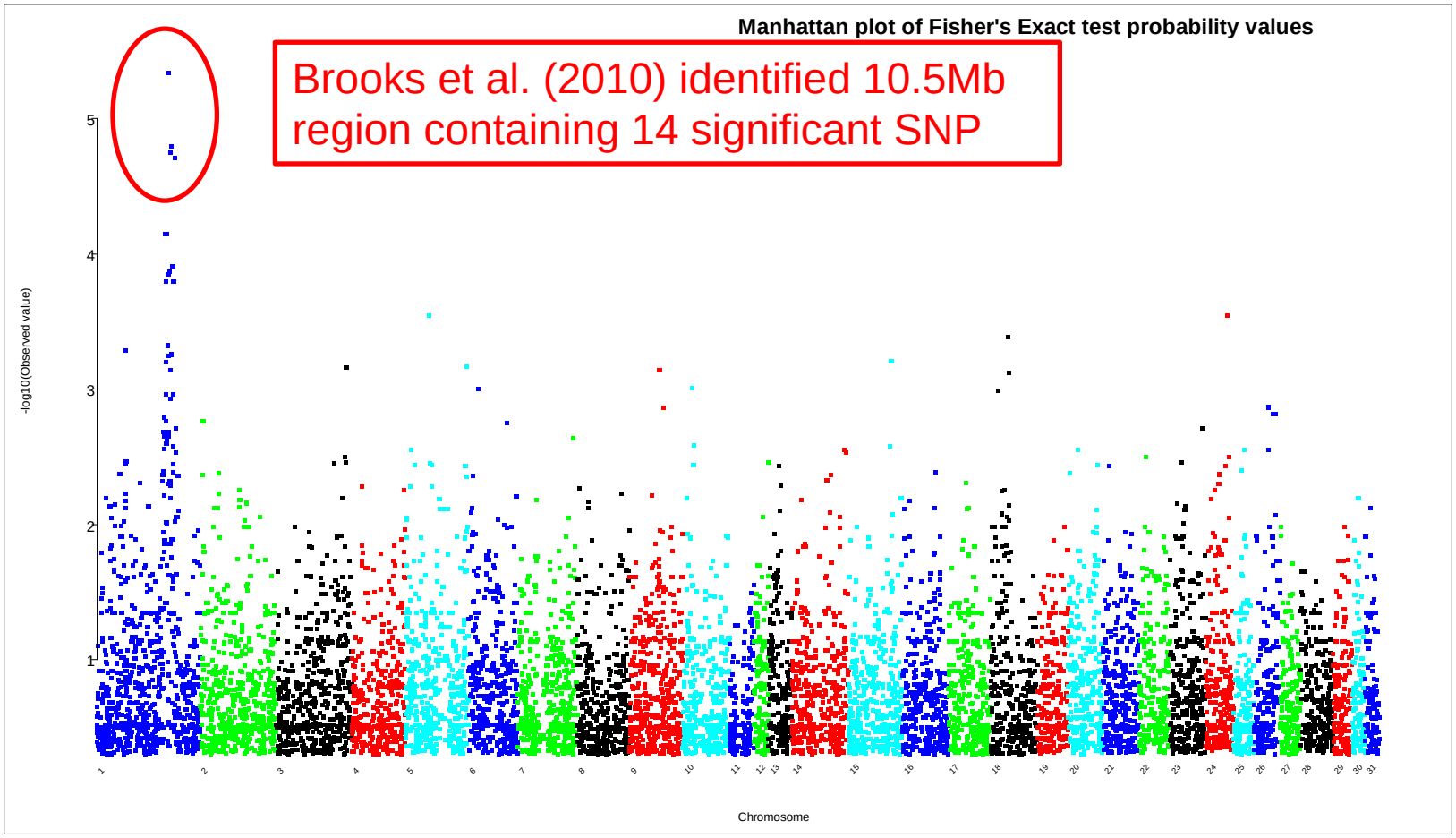
RVC

RESULTS

Output from all 5 methods

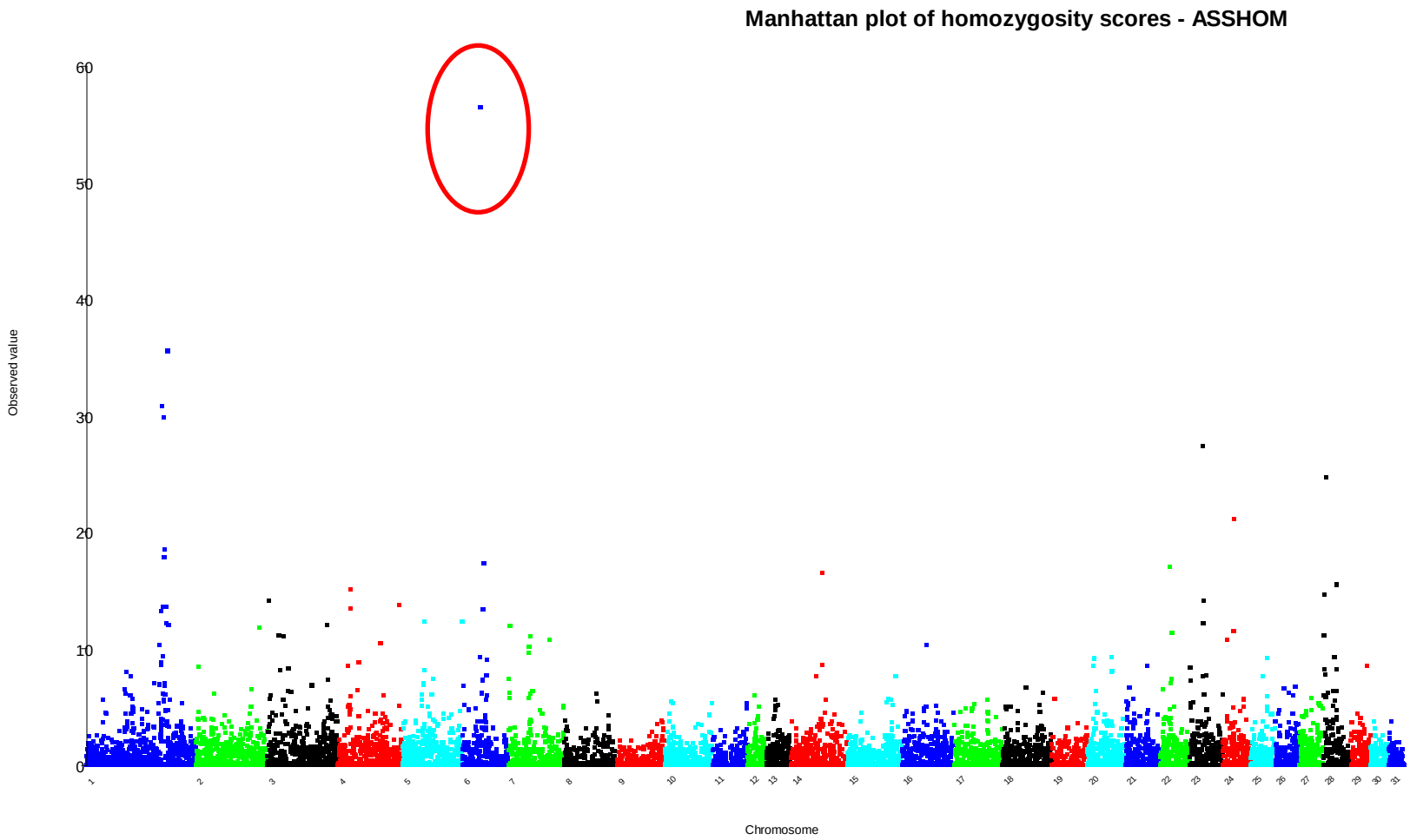
RVC

Chi-squared; Fisher's Exact Test



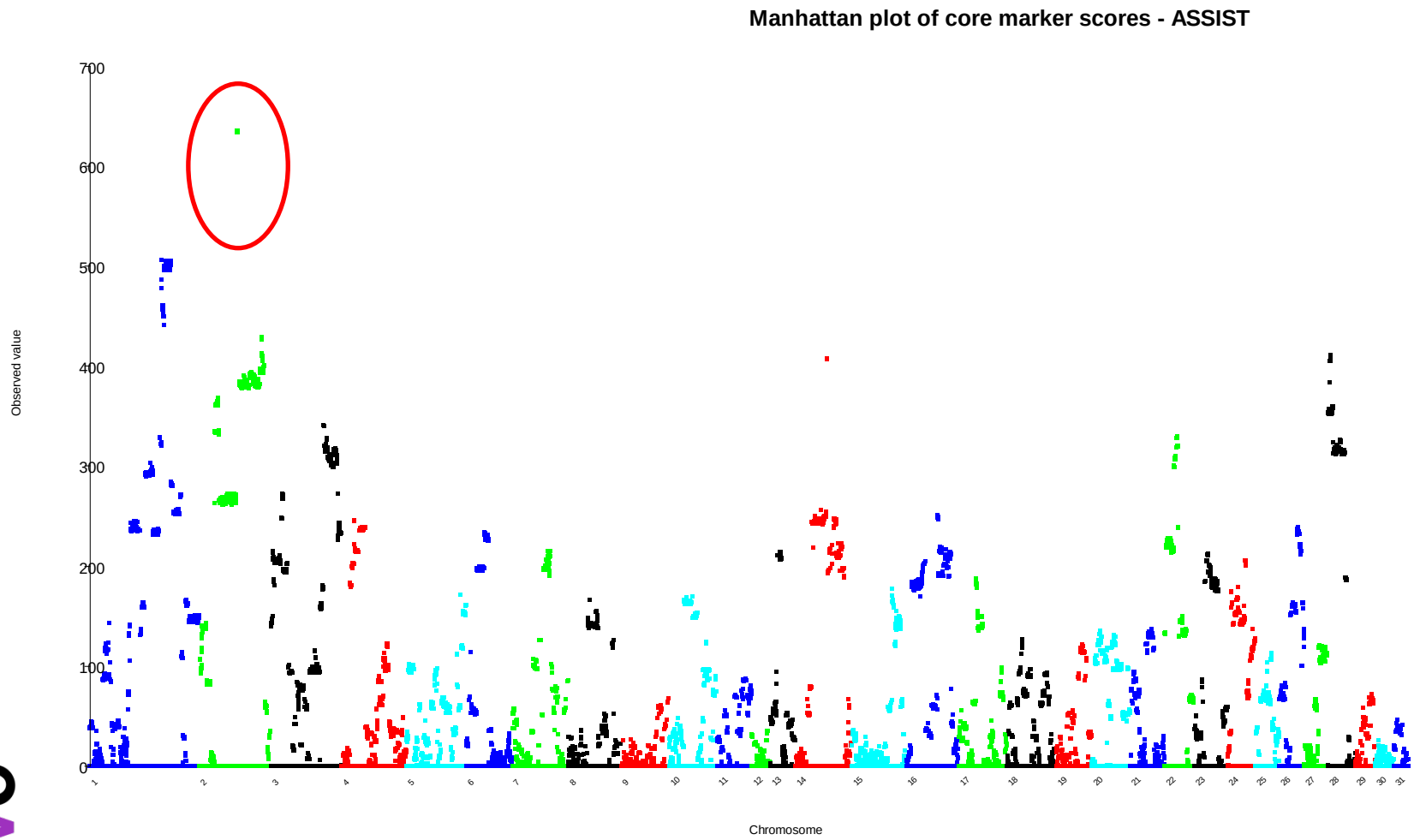
RVC

Charlier et al. (2008) - ASSHOM



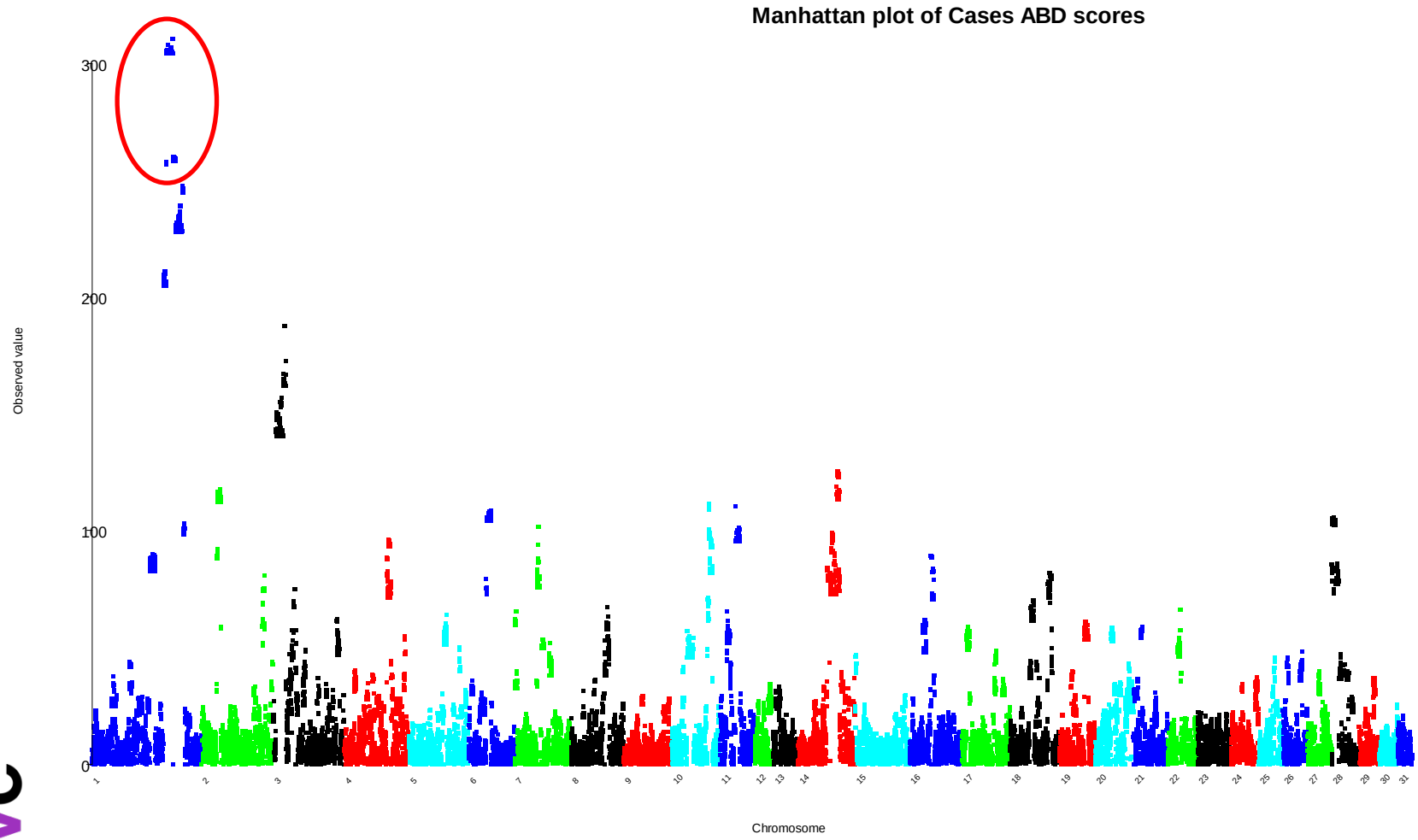
RVC

Charlier et al. (2008) - ASSIST



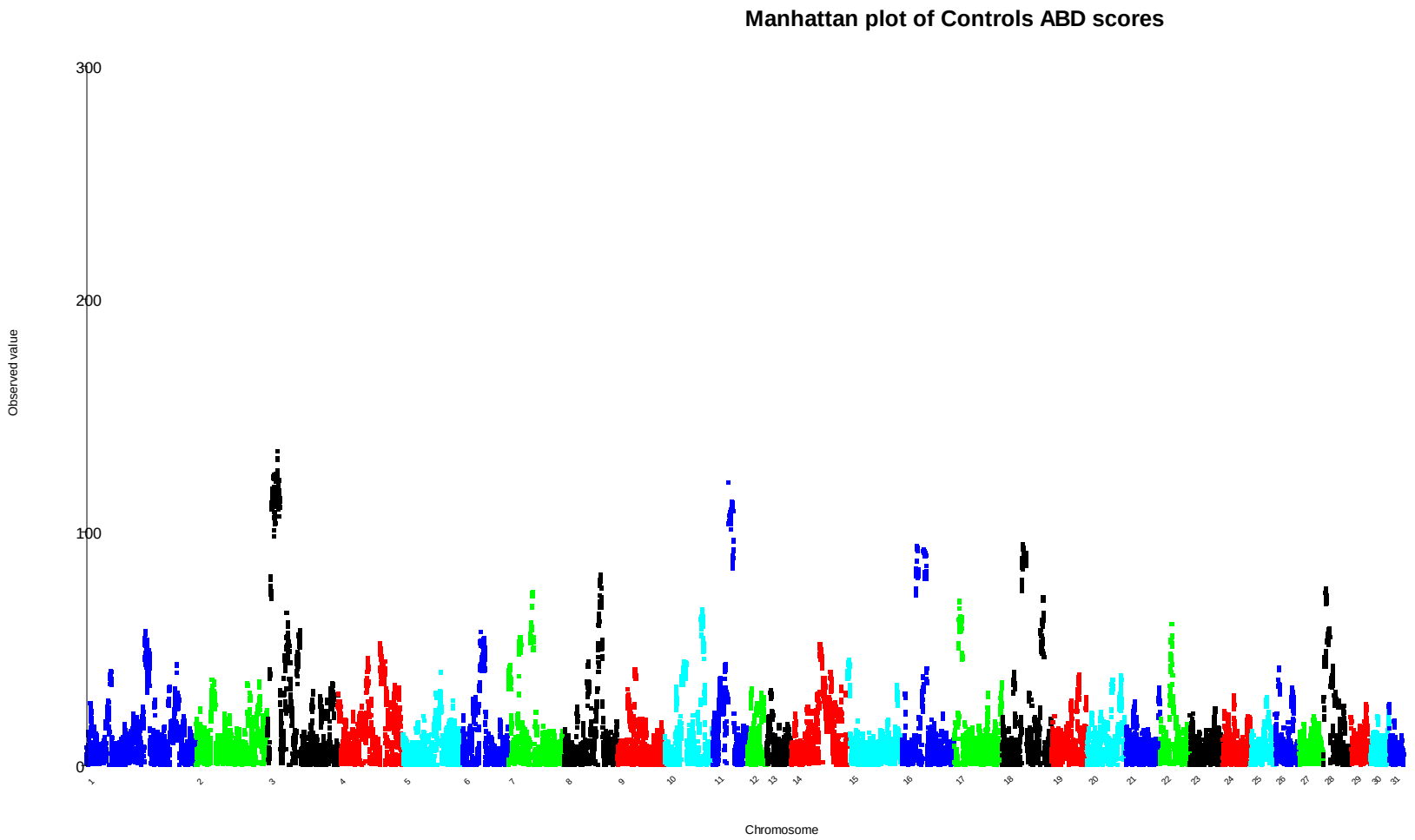
RVC

Autozygosity by difference - cases



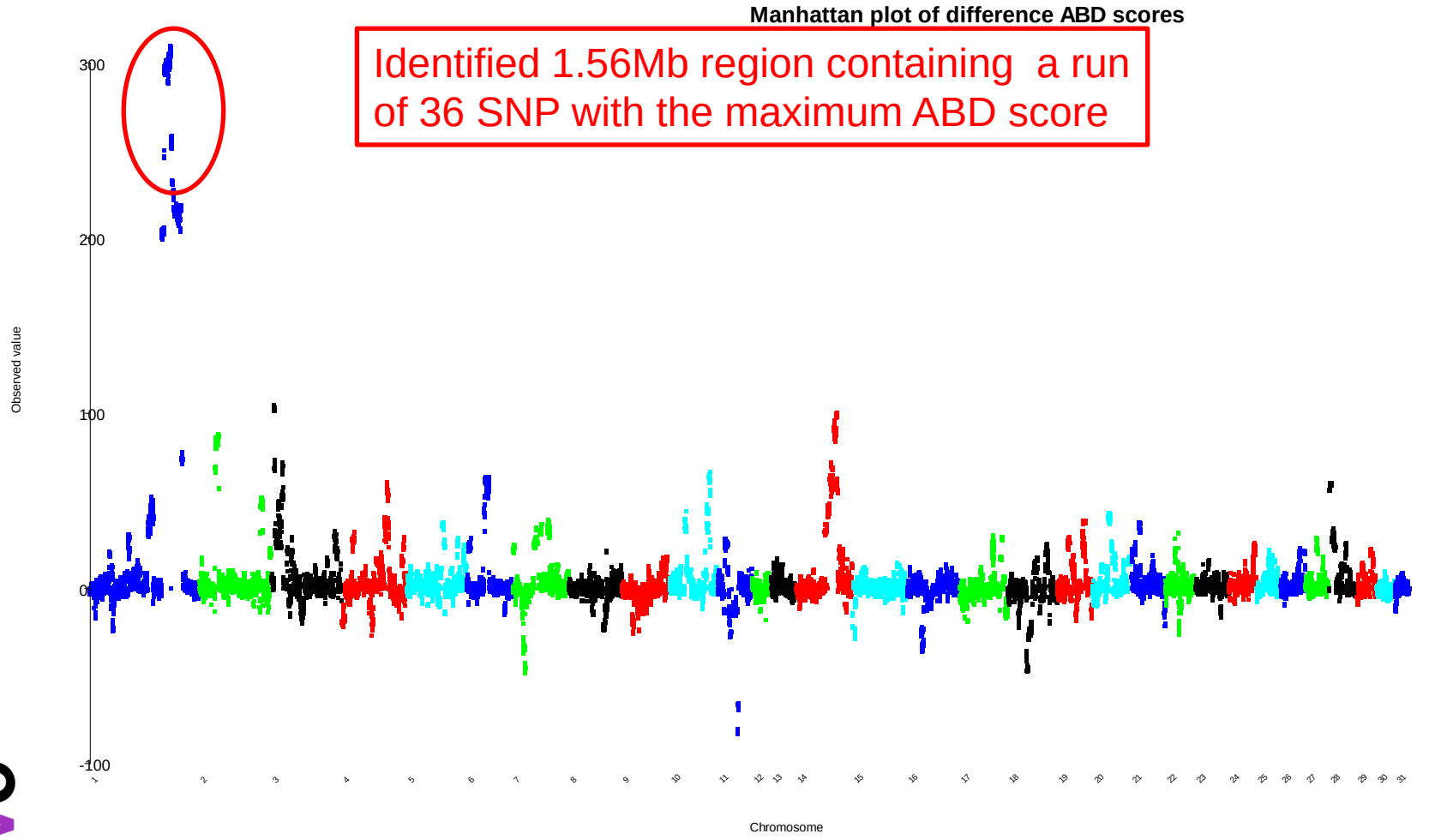
RVC

Autozygosity by difference - controls



RVC

Autozygosity by difference – ABD scores



RVC

Consensus line from PLINK -homozyg

Identified 1.56Mb region containing a run of 32 SNP

Chromosome	SNP Start	SNP end	BP1	BP2	Length (kb)	No. SNP	Mean kb length of all 6 cases	Mean SNP no in cases
1	BIEC2-59910	BIEC2-60700	136,812,666	138,375,254	1,563	32	18,515	435
3	BIEC2-773565	BIEC2-773833	17,973,786	19,831,439	1,858	48	14,447	350
3	BIEC2-775015	BIEC2-775091	27,687,621	28,000,837	313	10	11,523	280
3	BIEC2-776679	BIEC2-777162	34,703,671	36,615,659	1,912	41	5,473	129
3	BIEC2-777732	BIEC2-777904	36,840,265	38,157,802	1,318	35	2,220	56
3	BIEC2-806850	BIEC2-806892	100,552,940	101,088,438	535	12	10,966	261
6	BIEC2-946284	BIEC2-946484	30,618,147	31,501,172	883	29	9,438	227
6	BIEC2-947221	BIEC2-947850	35,047,269	36,256,430	1,209	29	8,883	212
7	BIEC2-974785	BIEC2-975013	2,259,854	2,620,805	361	16	2,950	73
7	BIEC2-996517	BIEC2-996677	39,560,106	41,340,178	1,780	44	5,502	124
16	BIEC2-343301	BIEC2-343301	41,231,085	41,231,085	0	1	3,365	79
22	BIEC2-583367	BIEC2-583748	16,163,673	16,652,941	489	9	11,709	275
24	BIEC2-636782	BIEC2-636946	17,289,305	17,573,926	285	7	7,170	169

RVC

Comparing the methods

Chi-squared, ABD and PLINK all identified the same region on ECA1.

PLINK and ABD found the same narrow area.

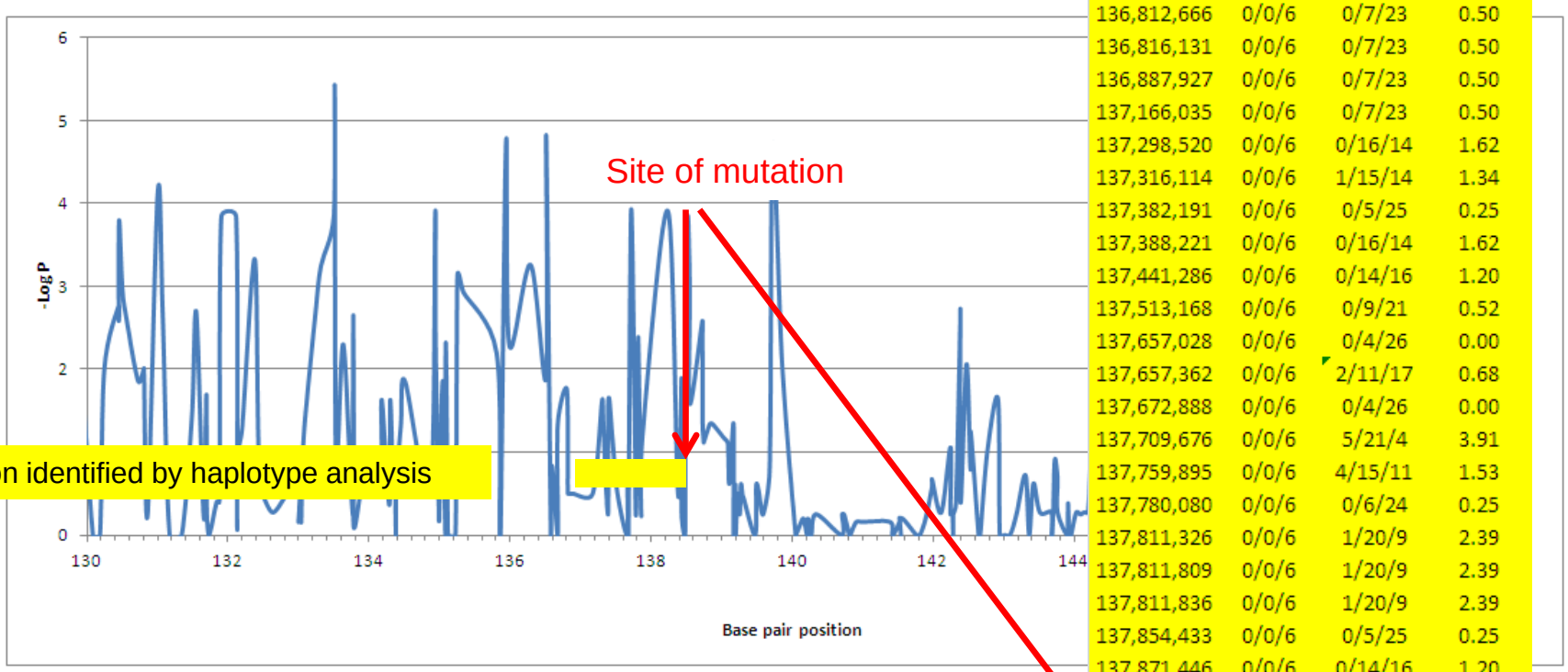
Brooks et al (2010) only found this area after haplotype analysis.

ASSIST and ASSHOM found different regions on different chromosomes.

What about the detail?

RVC

Fisher's exact test detail – EC



Region identified by haplotype analysis

bp	AFF	UNAFF	LogP
133,508,742	1/0/5	8/22/0	5.34
135,848,841	0/1/5	0/7/23	0.00
135,860,233	0/1/5	0/7/23	0.00
135,938,654	0/0/6	7/21/2	4.75
135,983,312	0/0/6	4/17/9	2.29
136,284,874	0/0/6	7/18/5	3.26
136,504,755	0/1/5	2/22/6	1.89
136,506,161	0/1/5	9/21/0	4.80
136,812,666	0/0/6	0/7/23	0.50
136,816,131	0/0/6	0/7/23	0.50
136,887,927	0/0/6	0/7/23	0.50
137,166,035	0/0/6	0/7/23	0.50
137,298,520	0/0/6	0/16/14	1.62
137,316,114	0/0/6	1/15/14	1.34
137,382,191	0/0/6	0/5/25	0.25
137,388,221	0/0/6	0/16/14	1.62
137,441,286	0/0/6	0/14/16	1.20
137,513,168	0/0/6	0/9/21	0.52
137,657,028	0/0/6	0/4/26	0.00
137,657,362	0/0/6	2/11/17	0.68
137,672,888	0/0/6	0/4/26	0.00
137,709,676	0/0/6	5/21/4	3.91
137,759,895	0/0/6	4/15/11	1.53
137,780,080	0/0/6	0/6/24	0.25
137,811,326	0/0/6	1/20/9	2.39
137,811,809	0/0/6	1/20/9	2.39
137,811,836	0/0/6	1/20/9	2.39
137,854,433	0/0/6	0/5/25	0.25
137,871,446	0/0/6	0/14/16	1.20
138,230,294	0/0/6	5/21/4	3.91
138,375,254	0/0/6	0/7/23	0.50
138,425,099	0/1/5	2/22/6	1.89
138,425,445	0/5/1	0/26/3	0.26
138,475,896	0/1/5	1/5/24	0.00
138,481,053	0/0/6	1/15/14	1.34

RVC

Interpreting Chi-squared

High proportion of cases need to be homozygous for the ‘affected’ allele

Chi-squared value more influenced by the control genotype distribution

Will not pick up a significant association ($P < 0.001$) between the mutation and a SNP unless the ‘diseased’ allele has a frequency of less than ~0.5 in the controls.

RVC

Possible drawbacks of Charlier et al. (2008) method

ASSHOM

Region on ECA6 was completely homozygous (11 and 22).

Heterozygotes given very low score, made lower by calculating harmonic mean.

ASSIST

Small number of long runs of homozygosity in some cases, but not all, on ECA2.

Both methods make use of allele frequency in controls

RVC

PLINK method

Rather difficult to find the ‘right’ input parameters

Visualisation not possible

No account of controls

RVC

Discussion points

Fine mapping still required to find exact location of mutation

Homozygosity scoring using ABD most informative – particularly in unusual situations e.g. incomplete penetrance, hidden SNP situation.

Future of SNP mapping now WGS available.

RVC

Conclusions

ABD and PLINK can find narrow regions of homozygosity in small number of cases and controls

Chi-squared is only useful under certain circumstances

Visualisation of ABD method makes it useful in a wide range of genetic scenarios

RVC

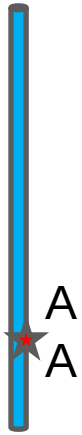
Acknowledgements

Samantha Brooks and colleagues at Cornell for making Lavender Foal Syndrome data available.

RVC

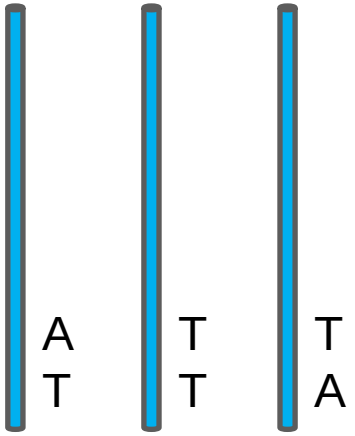
The hidden SNP problem

(Stumpf and McVean, 2003. *Nature Reviews, Genetics*)



Mutation occurs on one of the autosomes in between two SNPs

Other animals in the population have identical SNPs without the mutation



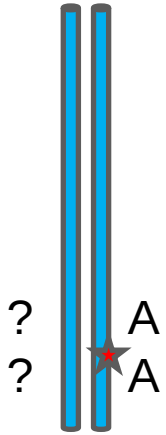
Other animals have alternative SNP combinations

RVC

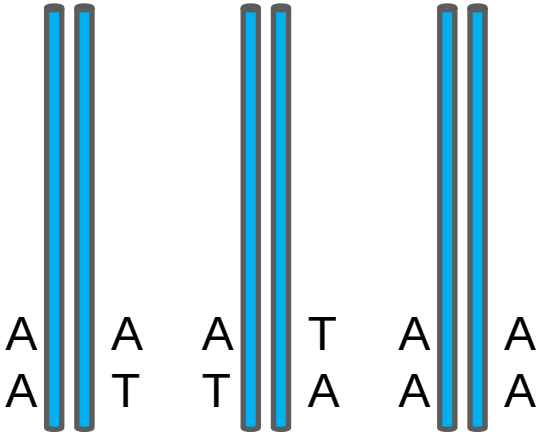
The hidden SNP problem

After a few generations we have a range of possible genotypes

An affected individual



A carrier



And unaffecteds

RVC