FIMM
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine

Stories of our past and future written in our genomes

Matti Pirinen

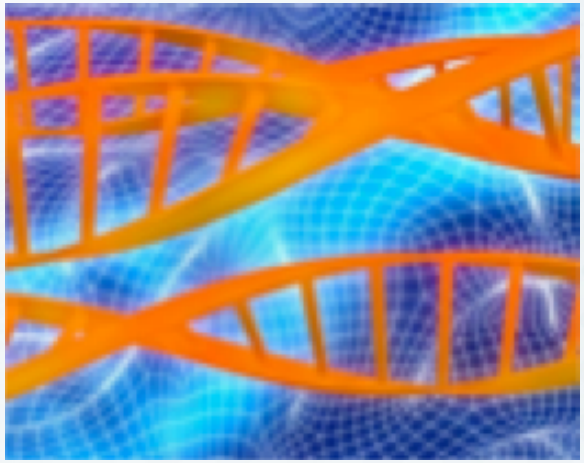
Institute for Molecular Medicine Finland

University of Helsinki

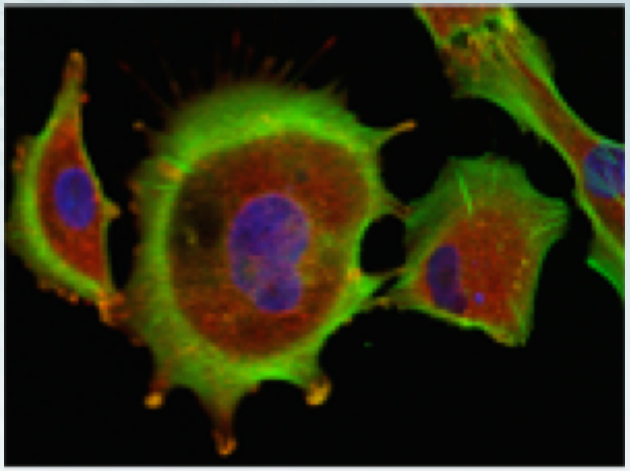
5.11.2015, Helsinki

Institute for Molecular Medicine Finland (FIMM)

- "Building a bridge from discovery to medicine"



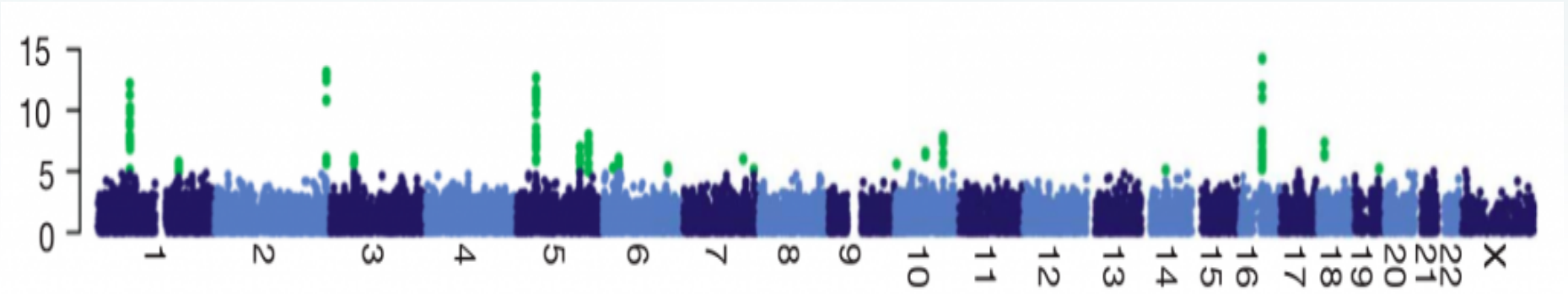
Human genomics
- Finnish cohorts + International collaboration
- Cardiovascular, Neuro-devel.
- Statistical genetics



Systems Biomedicine
- Personalized medicine
- Cancer
- Imaging & screening techn.

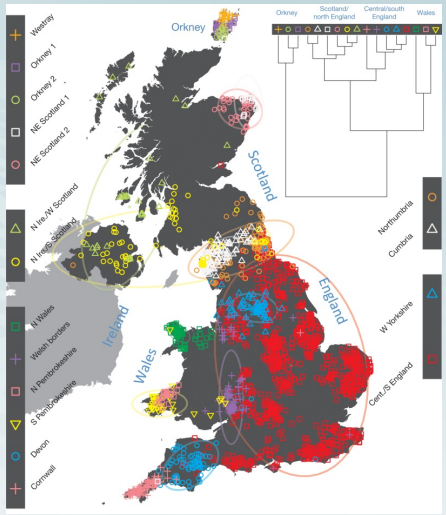
My background

- 2004 MSc Mathematics (Helsinki)
- 2009 PhD Statistical genetics (Helsinki)
- 2009-2012 Postdoc, GWAS (Oxford)
- 2012- Researcher (FIMM)



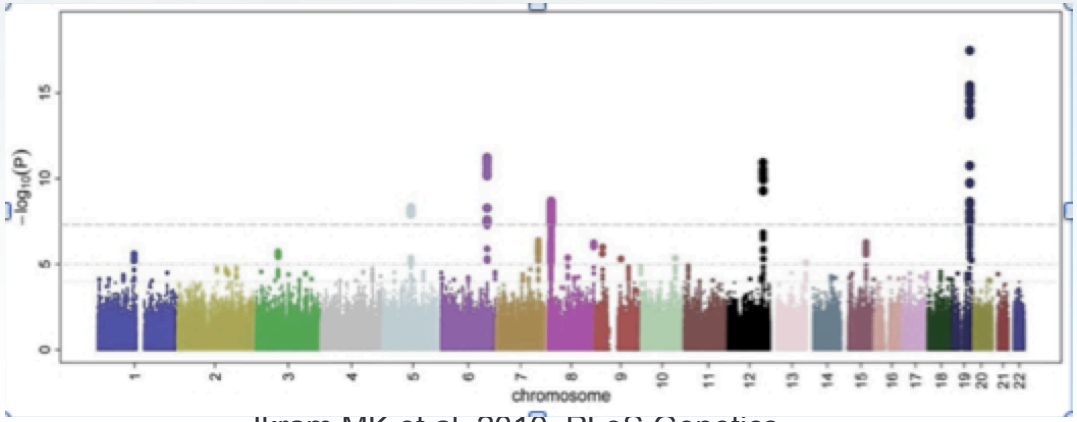
Outline

1. Fine-scale population structure in the British Isles



Leslie et al. 2015 Nature

2. Genetic studies of diseases and traits

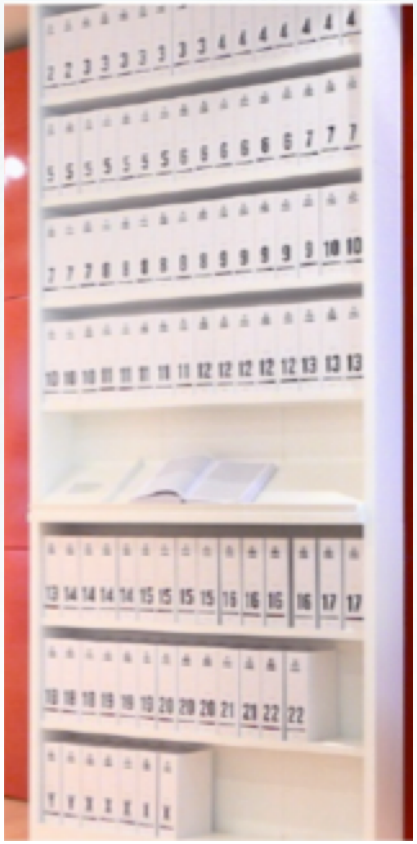


Ikram MK et al, 2010, PLoS Genetics

Human genome

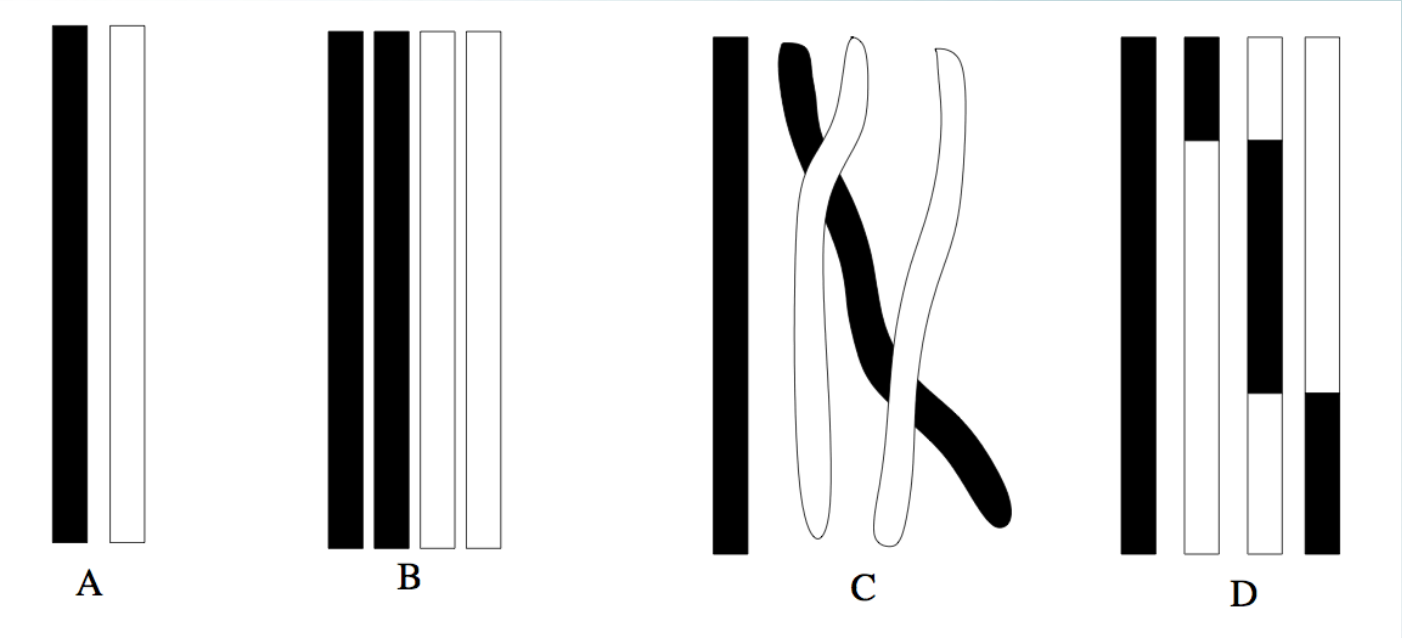
- Sequence of 3 x 10^9 letters from alphabet { A, C, G, T }

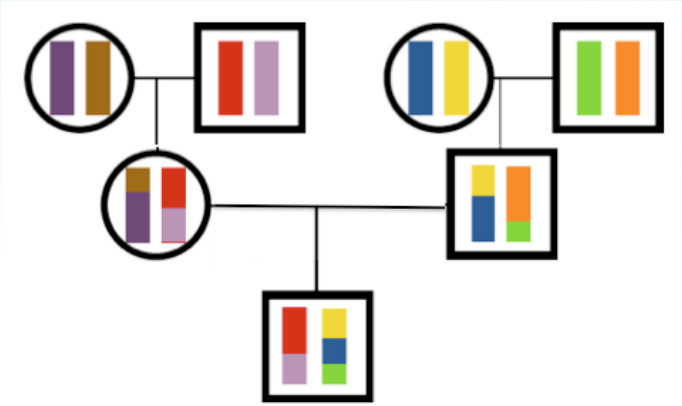
... G C G T T T A C G ...



22 chromosomes + X and Y
We carry 2 copies of genome

Inheritance of the genome

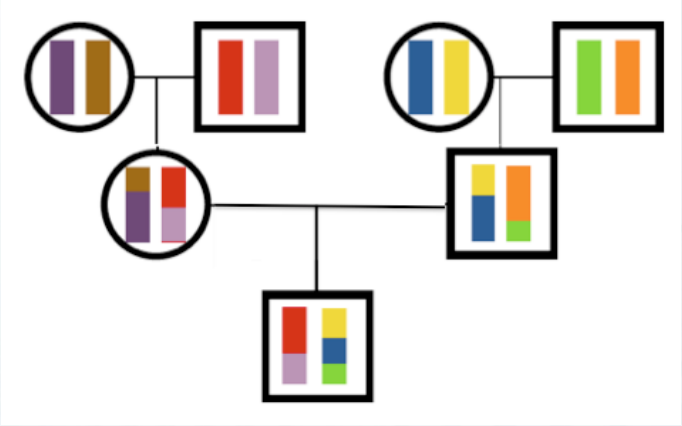


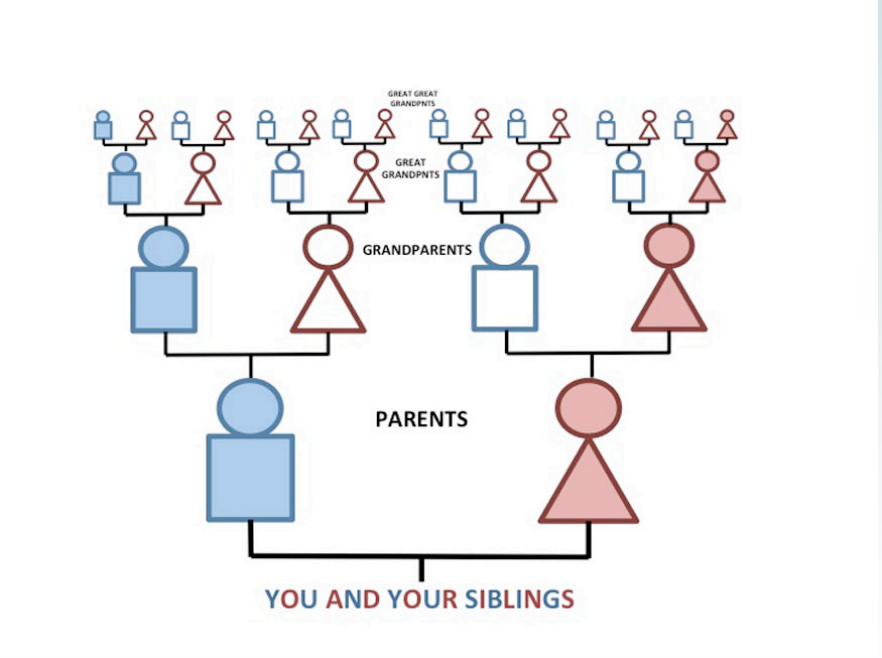


- **Recombination** shuffles the two genomes in each generation
- Genome is inherited in chunks

Lower Fig: Graham Coop lab

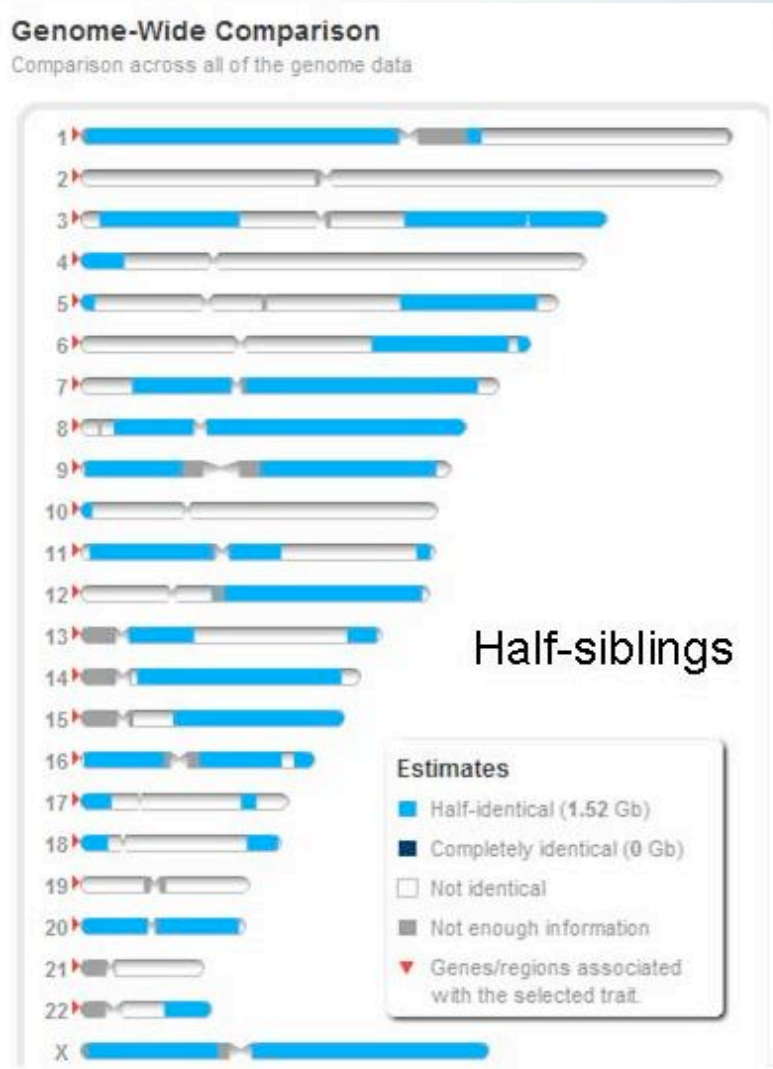
Inheritance of the genome

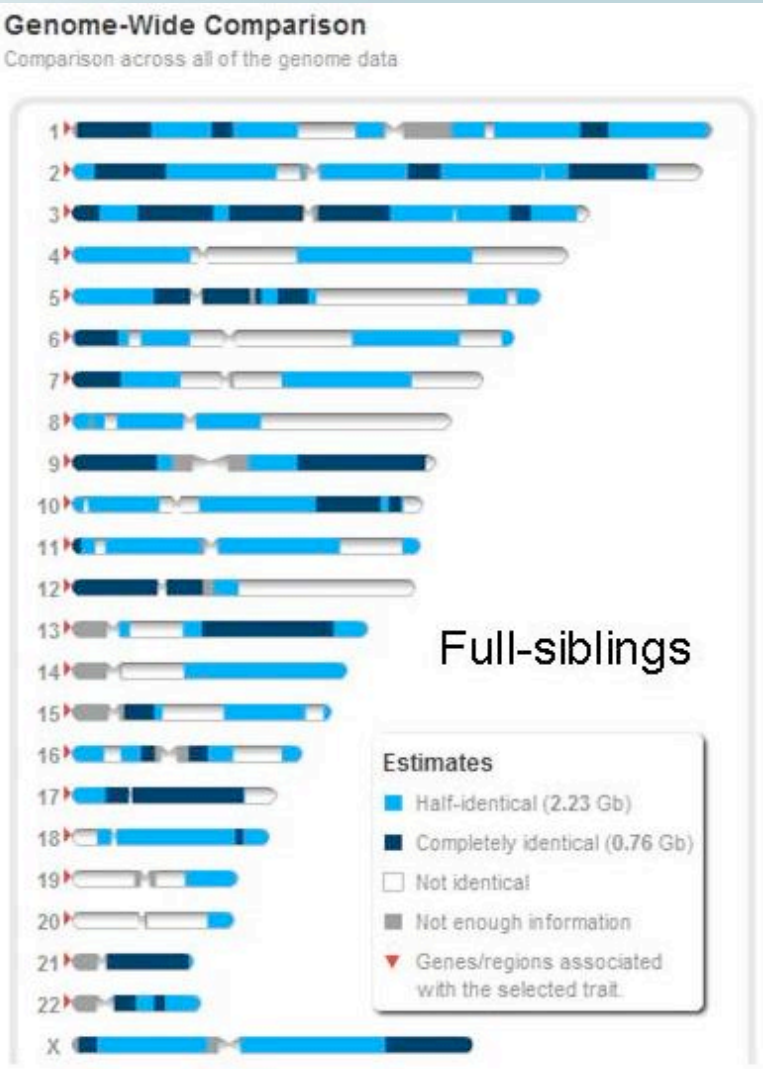




- Chunks from distant ancestors are shorter than from recent ancestors
- Close relatives share long identical chunks of genome

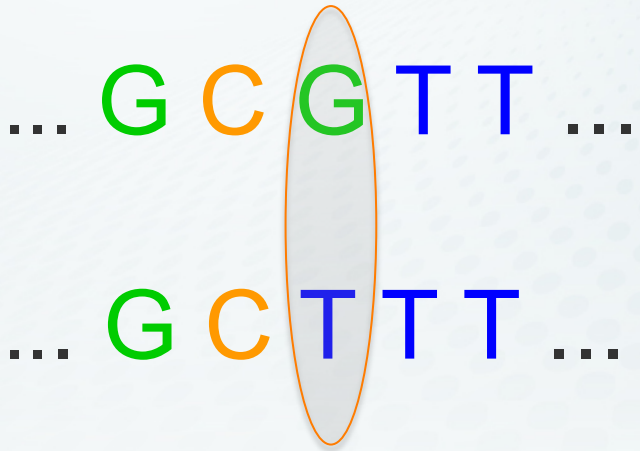
More identical chunks = more recent shared ancestors





Single-nucleotide polymorphism (SNP)

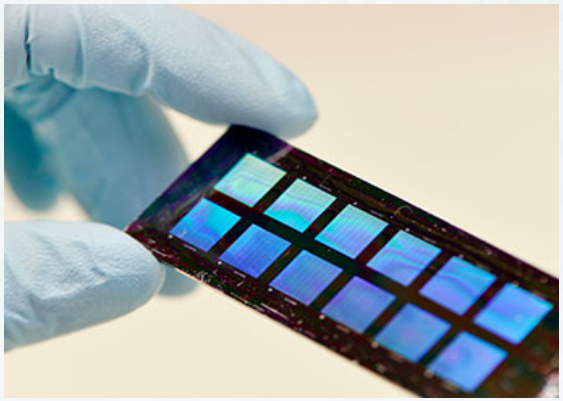
- Roughly 1:300 positions is a SNP, i.e., shows common variation in the population

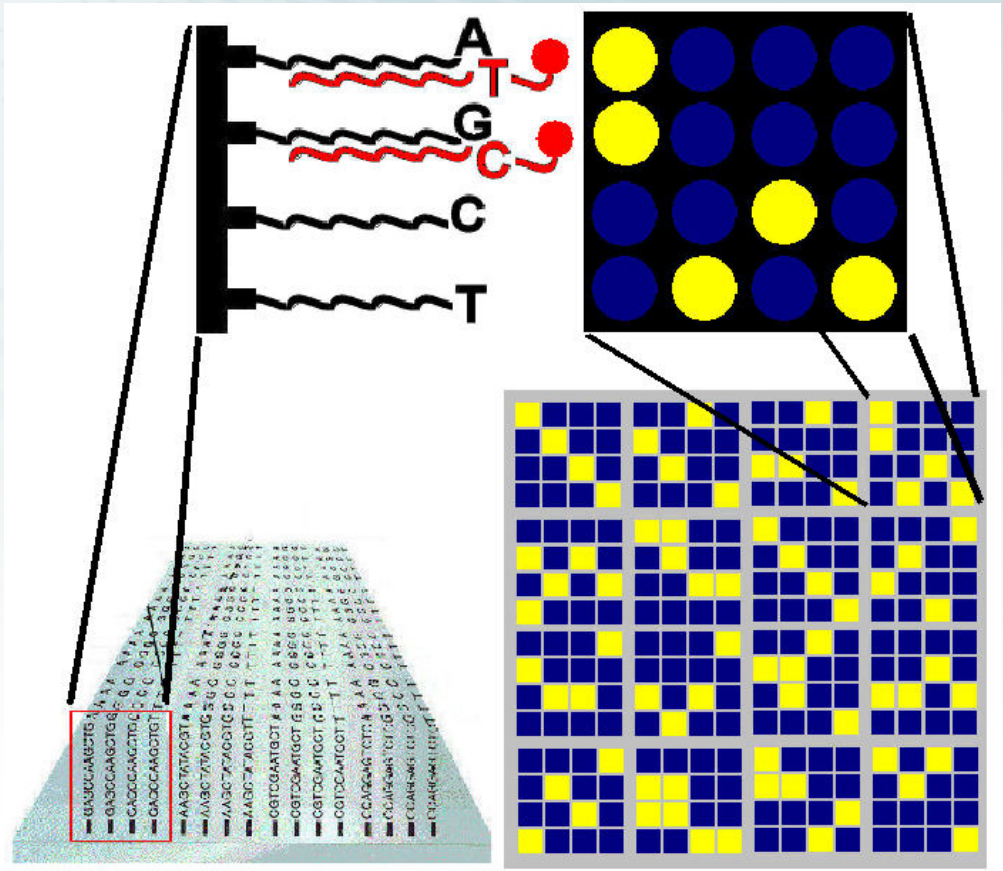


Genomes carry one of two ***alleles*** at this SNP: G or T

Reading SNPs

- Human SNP array can measure 10^6 SNPs
- Cost per individual 50…100 euros

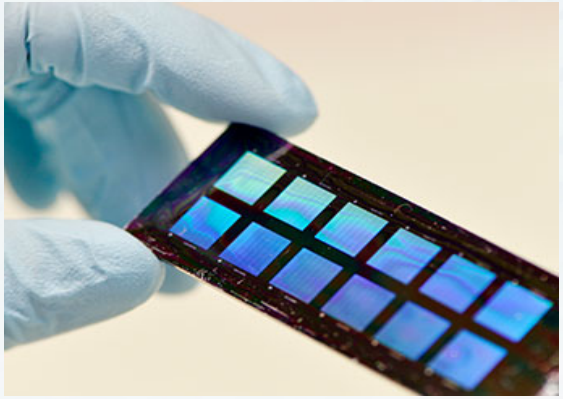




Steven M. Carr
www.mun.ca/biology/scarr/DNA_Chips.html

Reading SNPs

- Human SNP array can measure 10^6 SNPs
- Cost per individual 50…100 euros

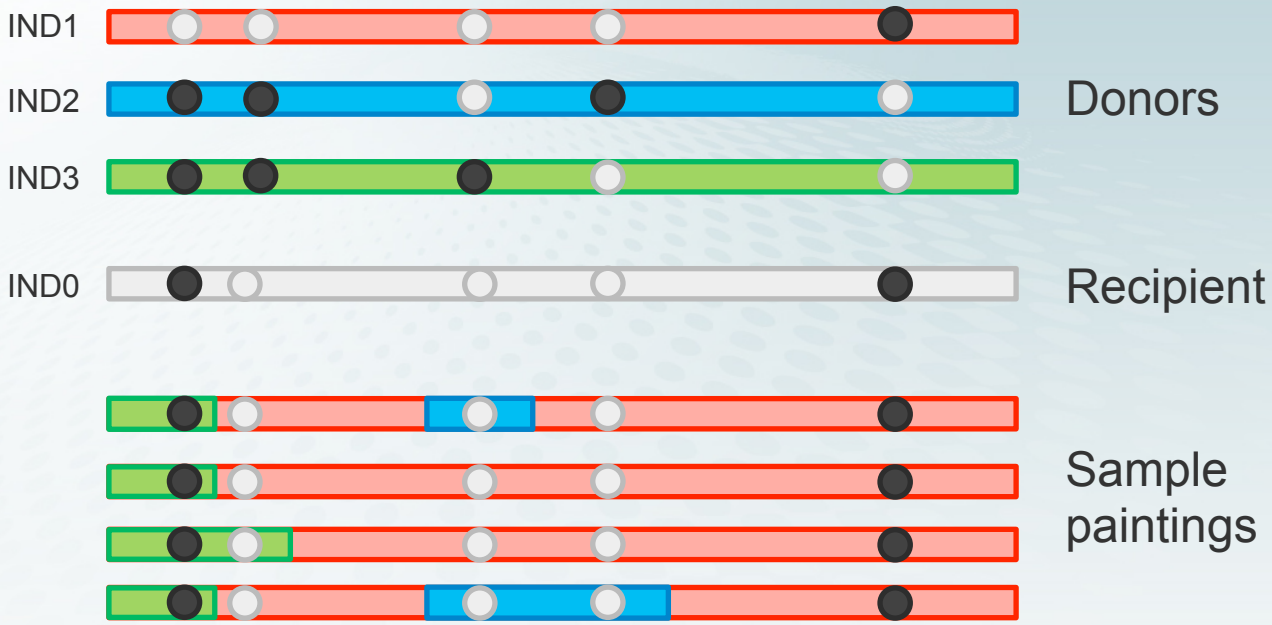


$n = 10^3$

IND1	M	A/A	A/C	A/C	C/C	A/A	A/C	A/C	C/C	C/C
IND2	M	A/A	A/C	A/A	C/C	A/C	A/C	A/C	C/C	A/C
IND3	M	A/C	C/C	A/C	C/C	A/A	C/C	C/C	C/C	C/C
IND4	F	A/A	A/C	A/A	C/C	A/C	A/C	A/C	C/C	C/C
IND5	F	A/A	C/C	A/C	C/C	A/A	A/A	C/C	A/A	A/C
IND6	M	A/C	A/C	C/C	C/C	A/A	A/C	A/C	C/C	C/C
IND7	F	A/A	A/A	A/C	C/C	A/A	C/C	A/C	C/C	A/C
IND8	M	A/C	A/C	C/C	C/C	C/C	A/C	C/C	C/C	C/C
IND9	F	A/A	A/A	A/C	C/C	A/A	A/C	C/C	A/C	C/C

$m = 10^6$

Estimating genetic similarity by Chromosome painting



- Count the number or the length of the chunks

  → Coancestry matrix

- ChromoPainter by G. Hellenthal

	IND1	IND2	IND3
IND0	4.2	1.2	1.9

Sini Kerminen

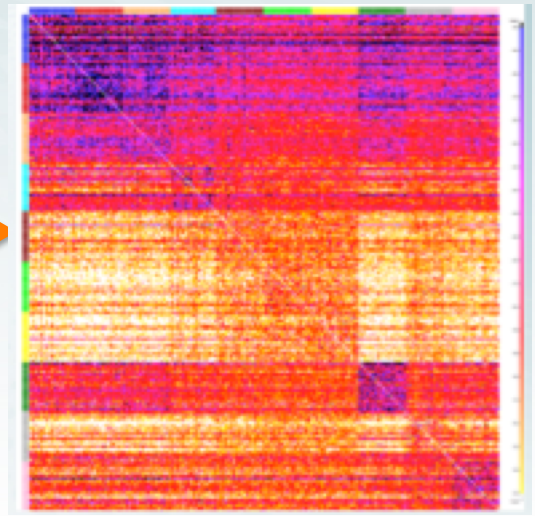
ChromoPainter

- Data: $10^5 \ldots 10^6$ genome-wide SNPs on 1000s of individuals
- Summarized by n x n matrix of genetic similarities

$n=10^3$

```
IND1 M A/A A/C A/C C/C A/A A/C A/C C/C C/C
IND2 M A/A A/C A/A C/C A/C A/C A/C C/C A/C
IND3 M A/C C/C A/C C/C A/A C/C C/C C/C C/C
IND4 F A/A A/C A/A C/C A/C A/C A/C C/C C/C
IND5 F A/A C/C A/C C/C A/A A/A C/C A/A A/C
IND6 M A/C A/C C/C C/C A/A A/C A/C C/C C/C
IND7 F A/A A/A A/C C/C A/A C/C A/C C/C A/C
IND8 M A/C A/C C/C C/C C/C A/C C/C C/C C/C
IND9 F A/A A/A A/C C/C A/A A/C C/C A/C C/C
```

$m=10^6$

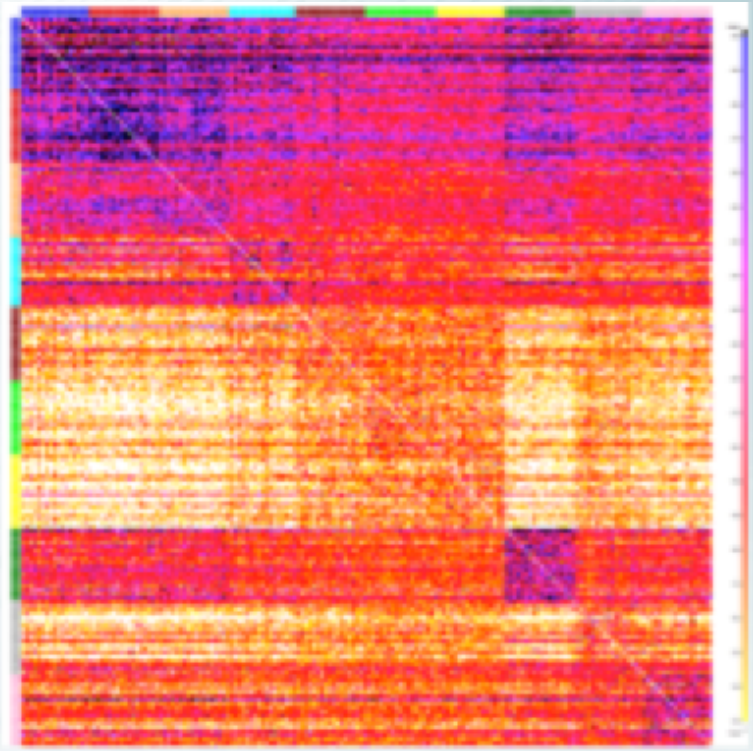


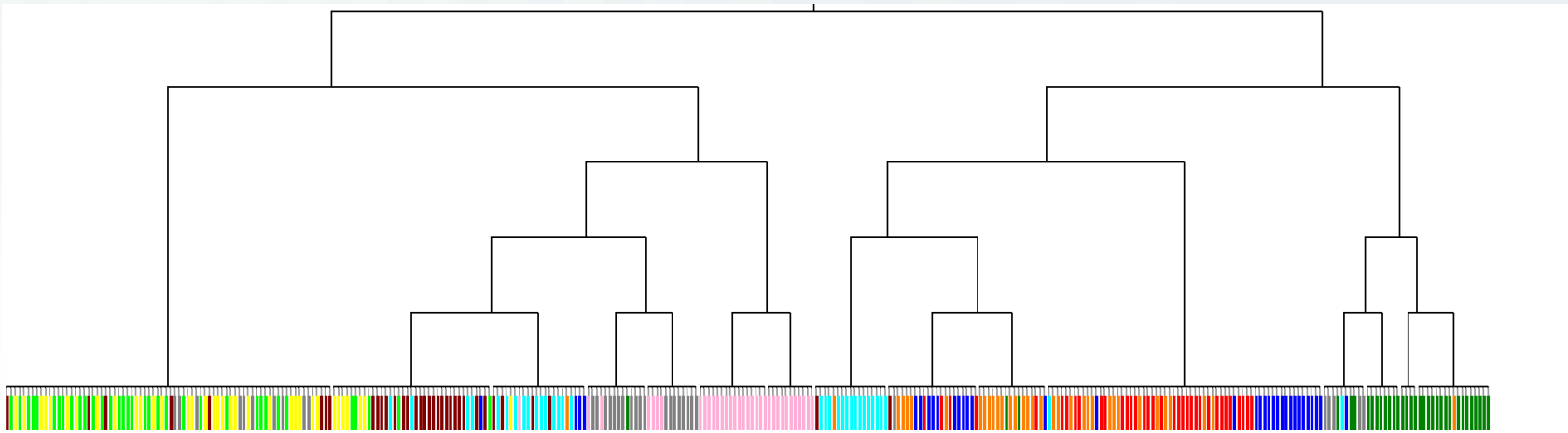
$n=10^3$

$n=10^3$

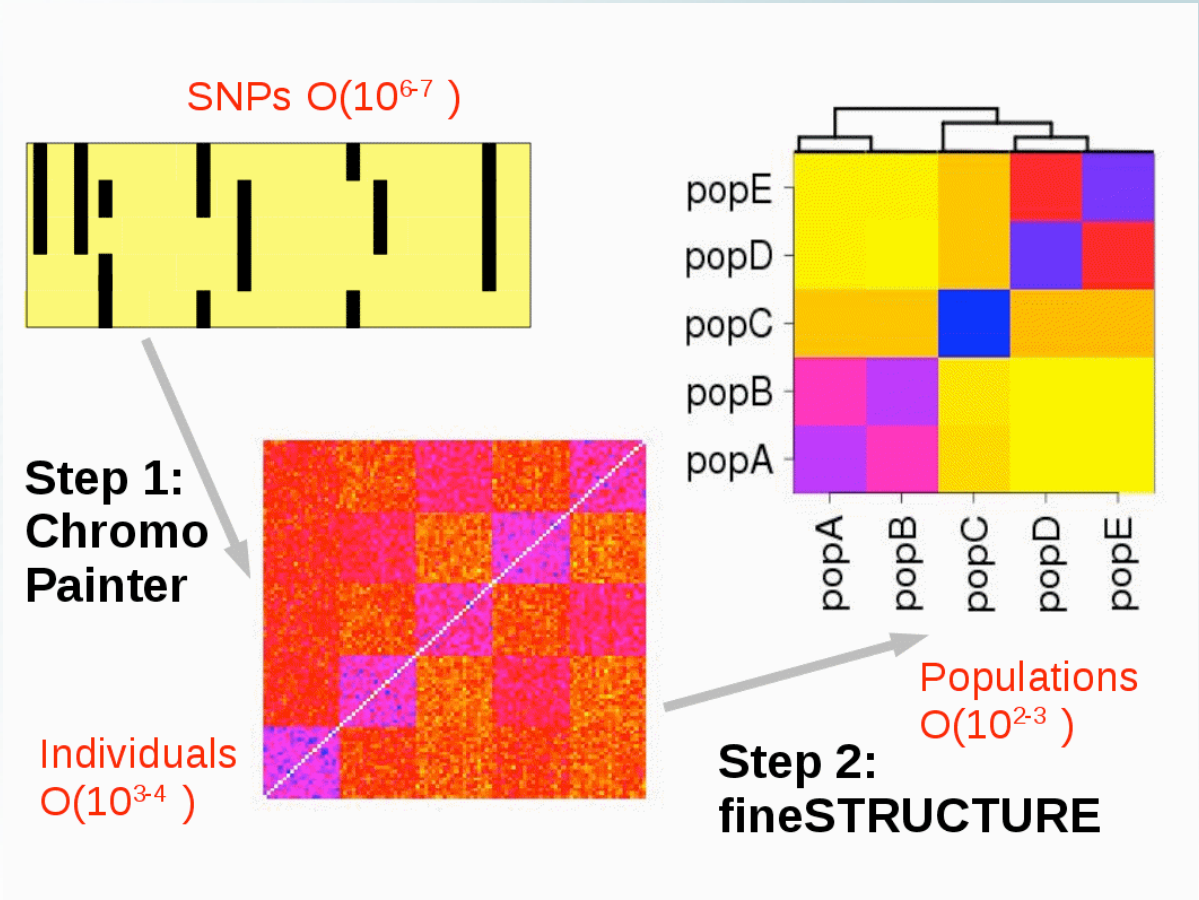
Similarities

Groups





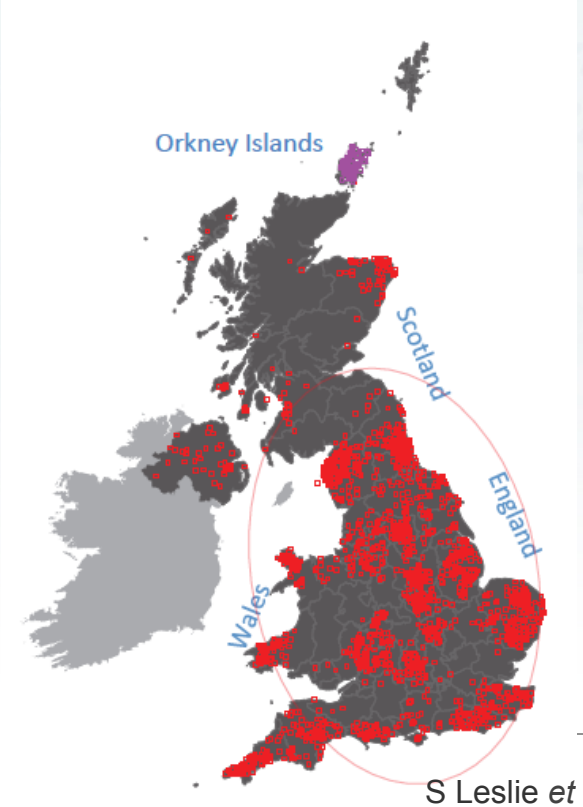
ChromoPainter and fineSTRUCTURE



www.paintmychromosomes.com

People of the British Isles (POBI)

- Samples from rural Britain with grandparents born within 80 km from each other (pairwise)
  - Takes us to geographical distribution of genetics of late 1800s
  - 520,000 SNPs genotyped on 2,039 indivuals



Bruce Winney
Stephen Leslie
Garrett Hellenthal
.
.
.
Walter Bodmer
Peter Donnelly

S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

Peopling the British Isles 1/4

S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

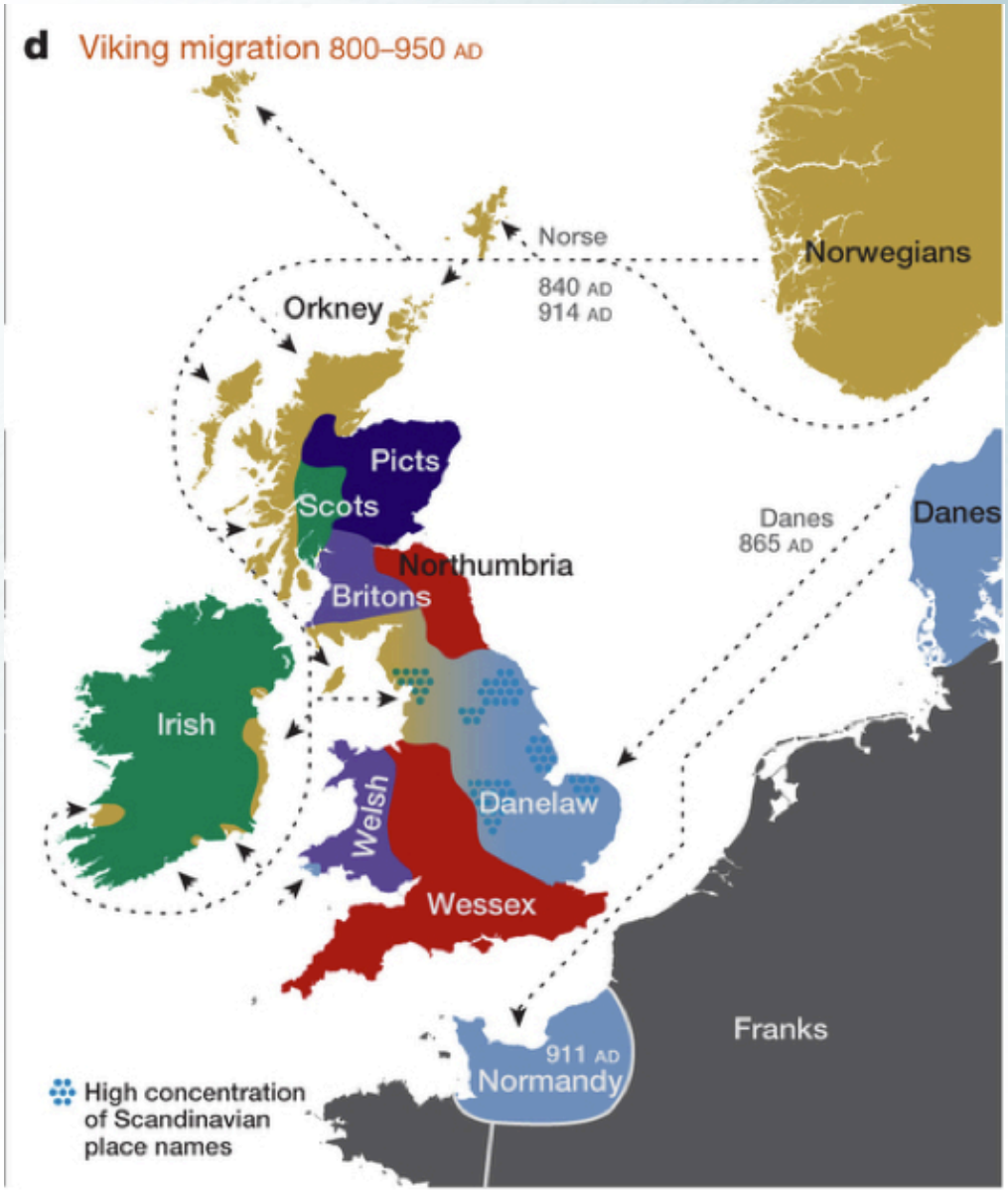
Peopling the British Isles 2/4

S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

Peopling the British Isles 3/4

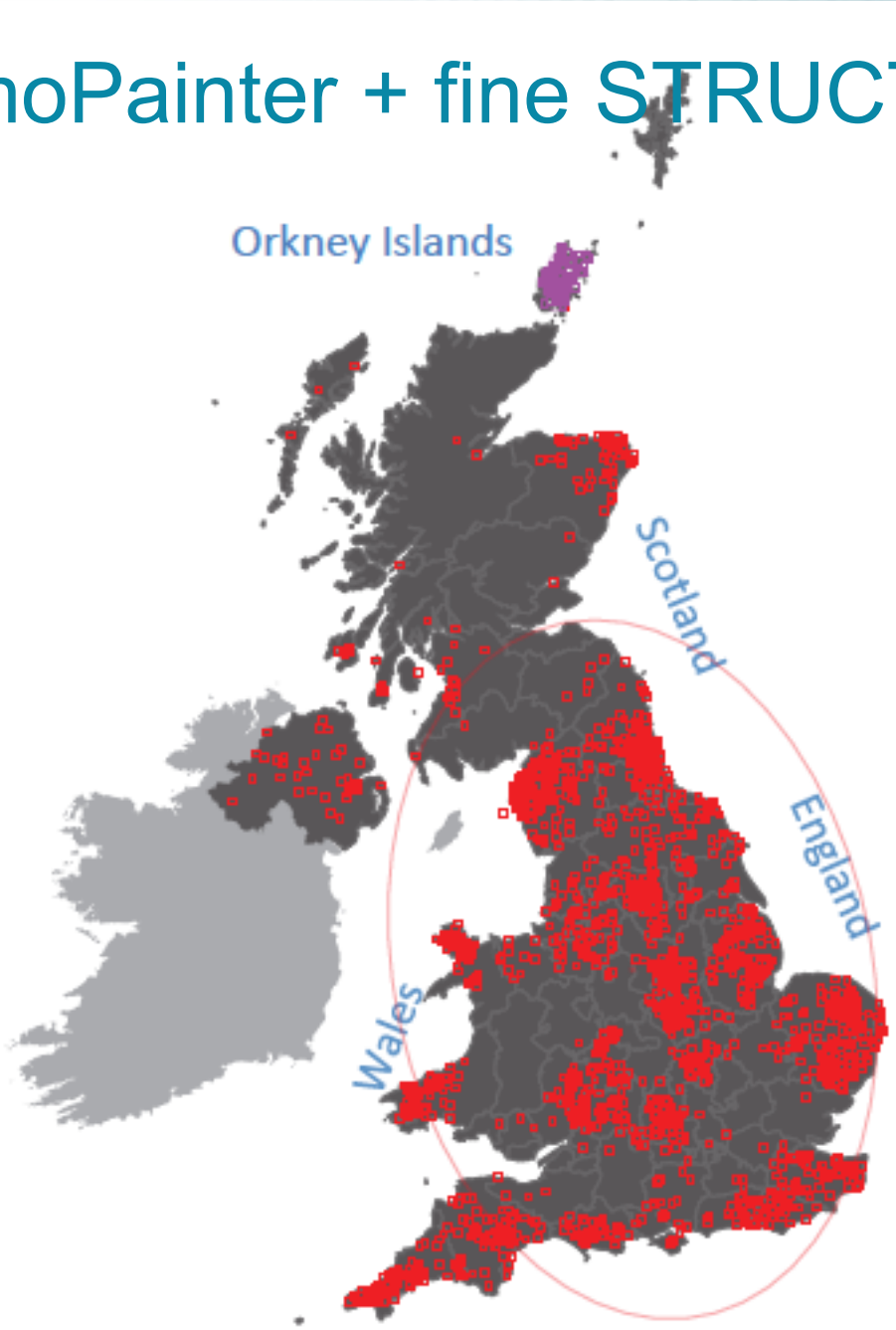
S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

Peopling the British Isles 4/4



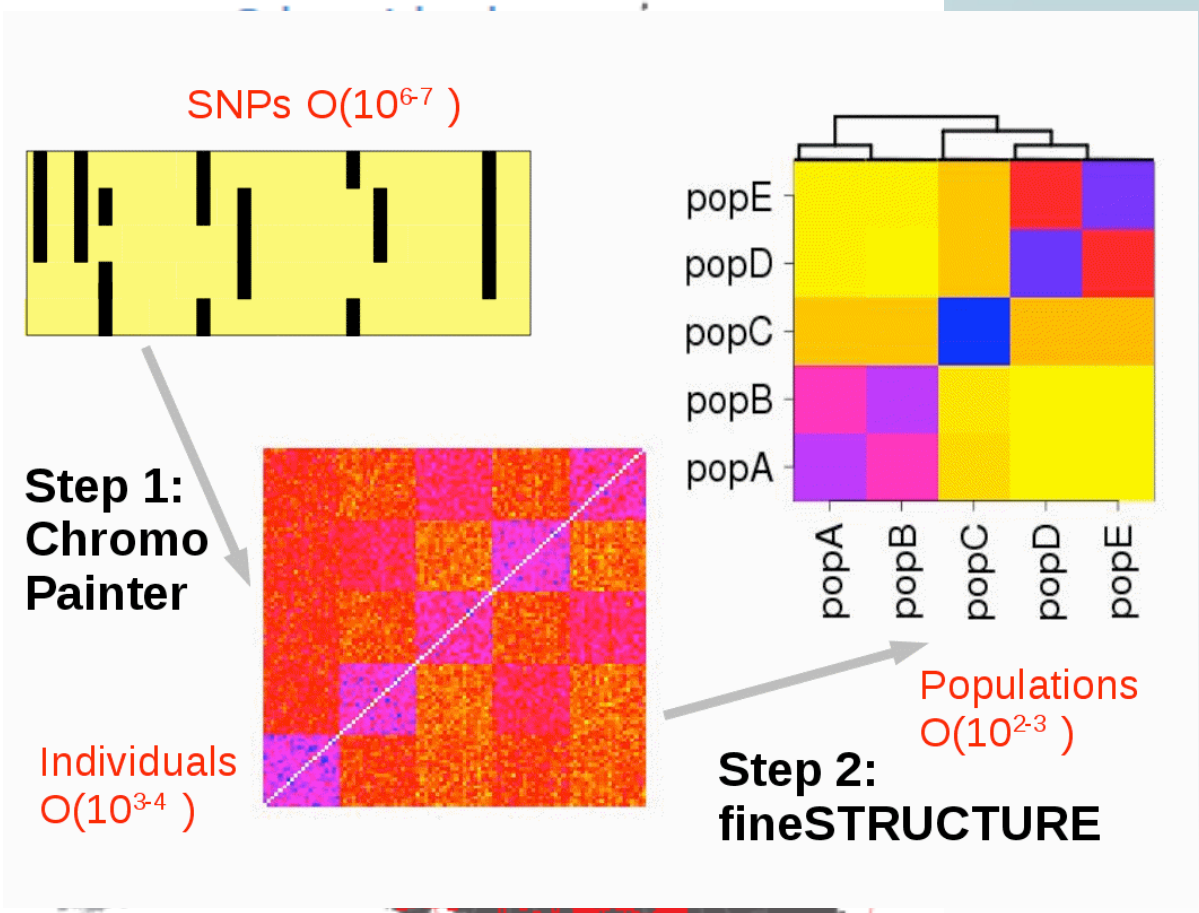
S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

ChromoPainter + fine STRUCTURE on Britain



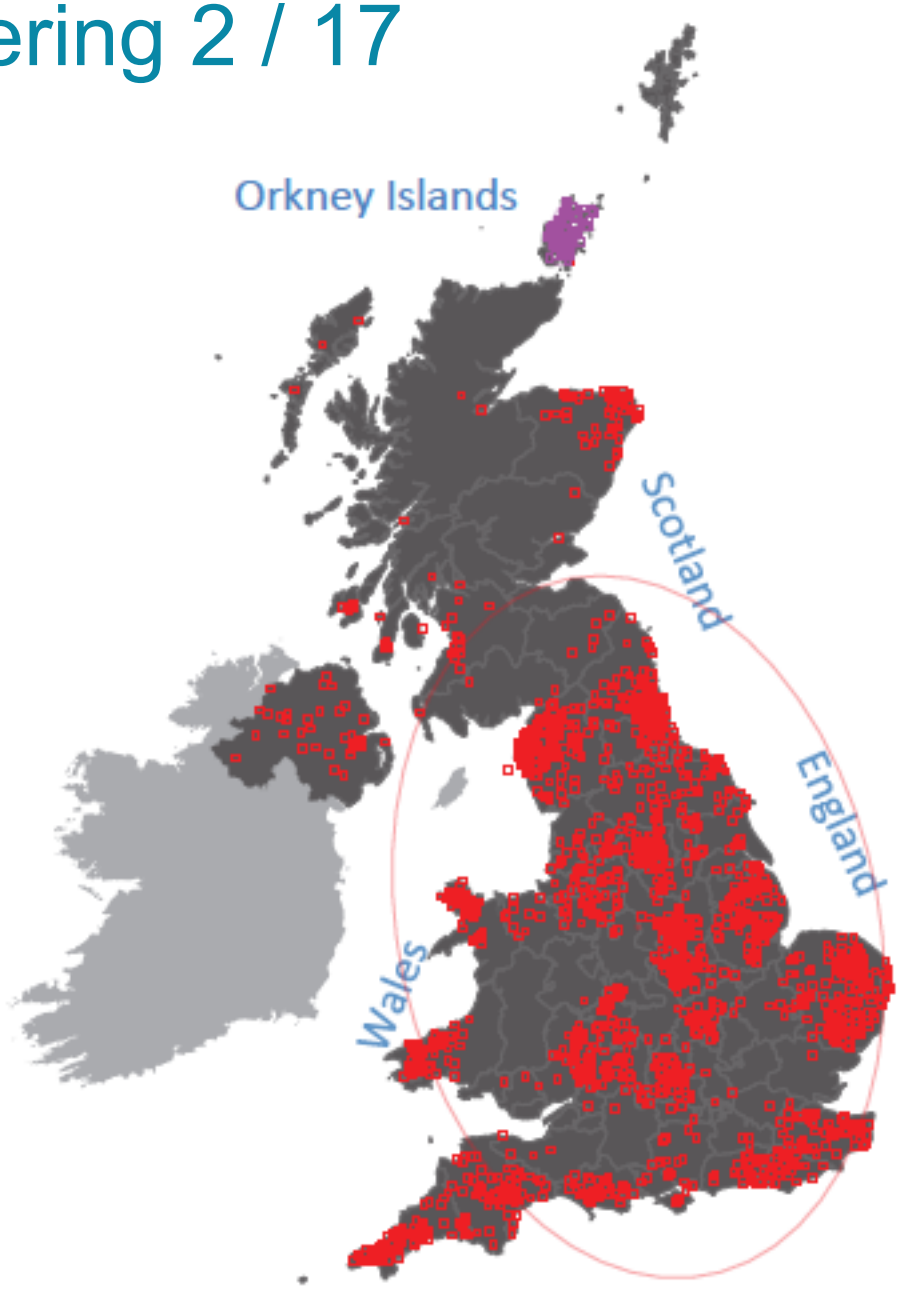
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

ChromoPainter + fine STRUCTURE on Britain



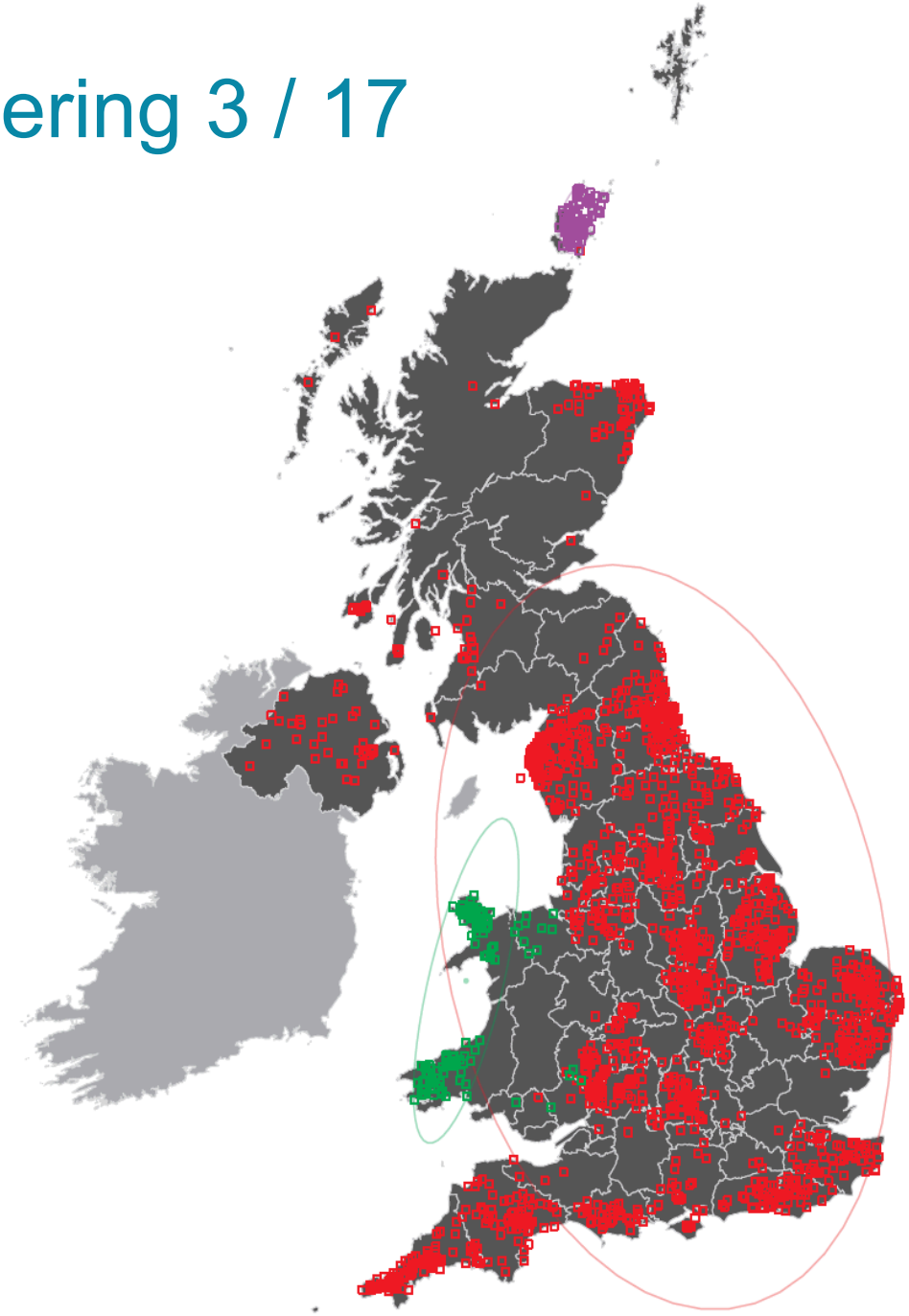
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 2 / 17



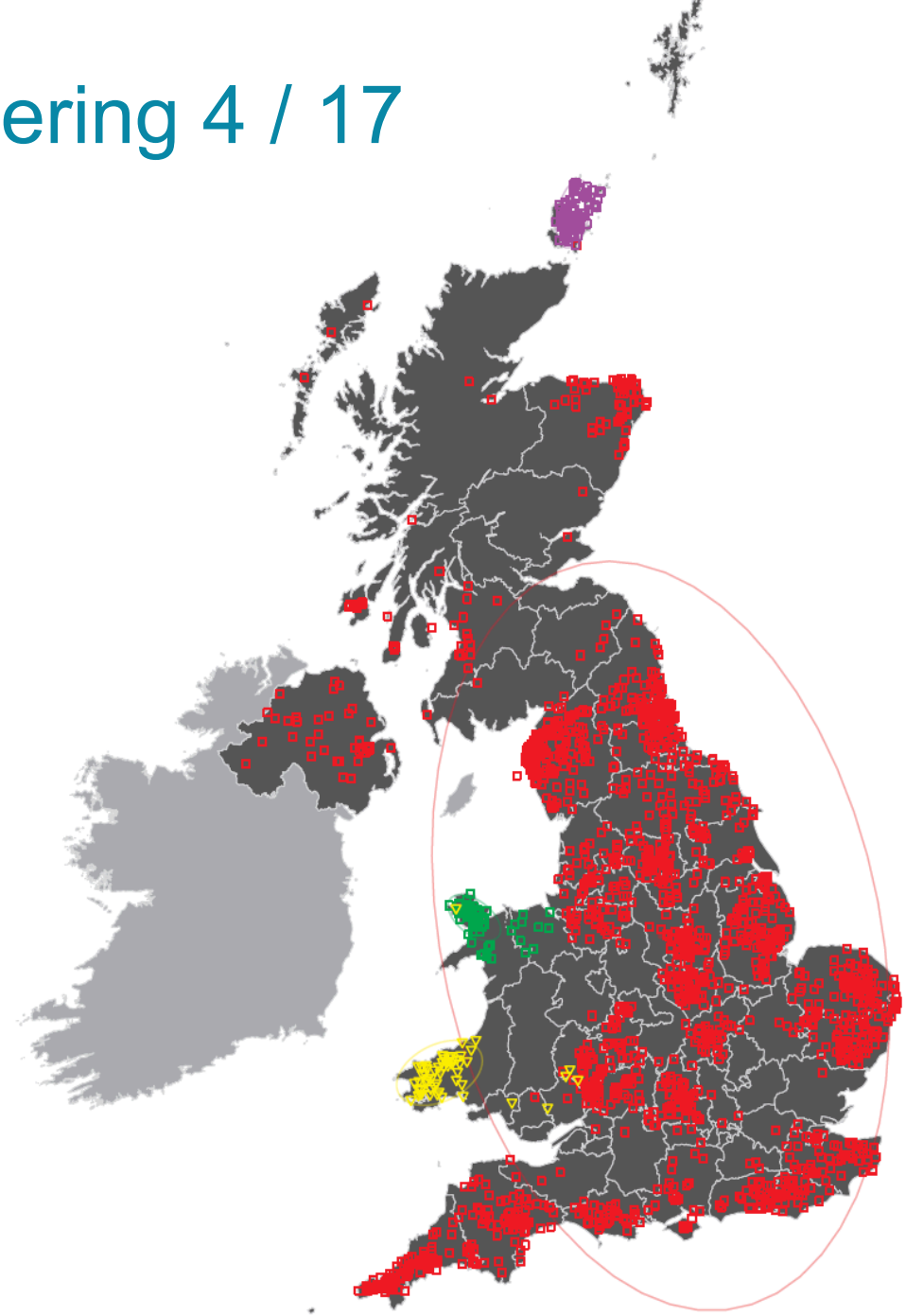
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 3 / 17



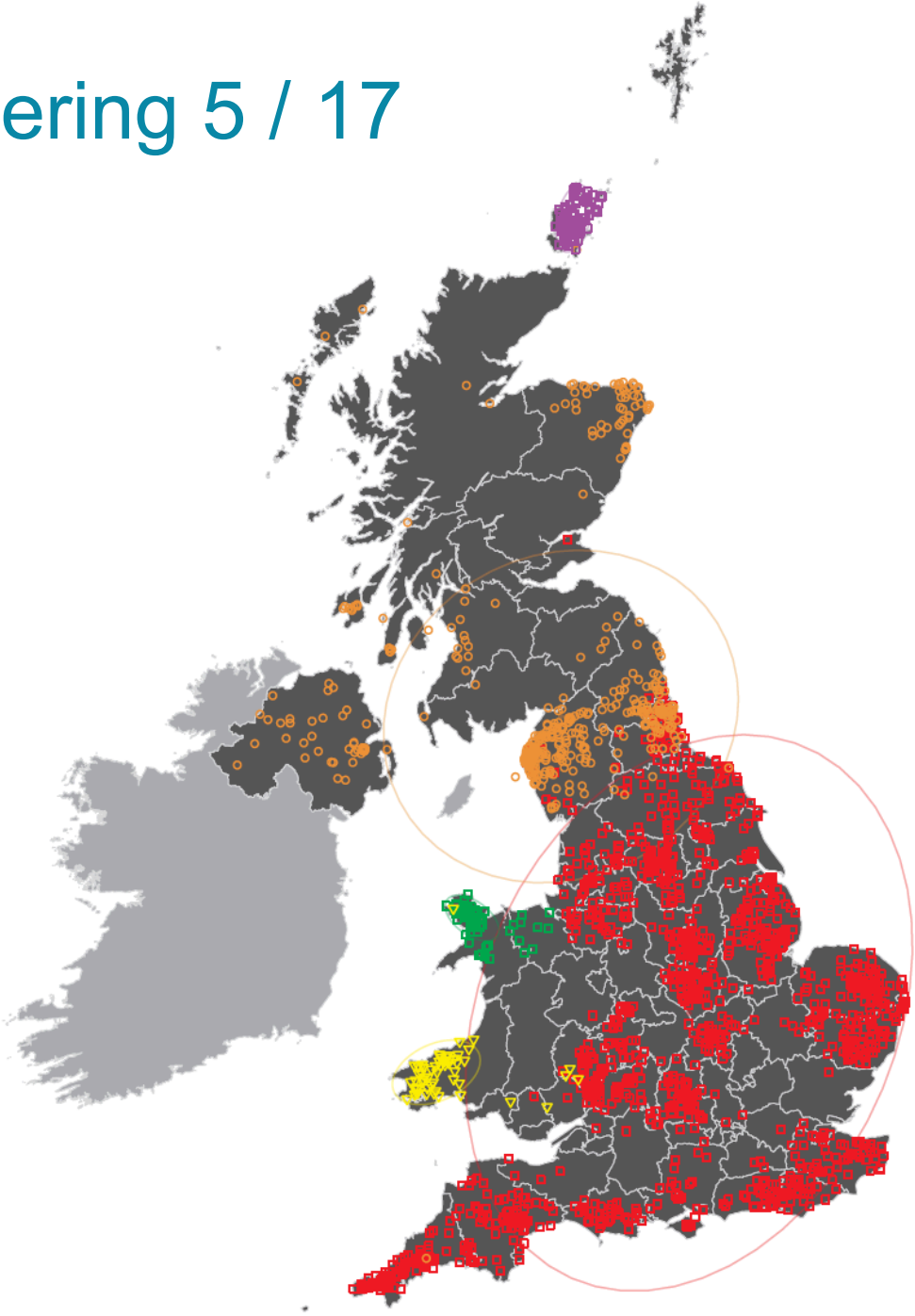
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 4 / 17



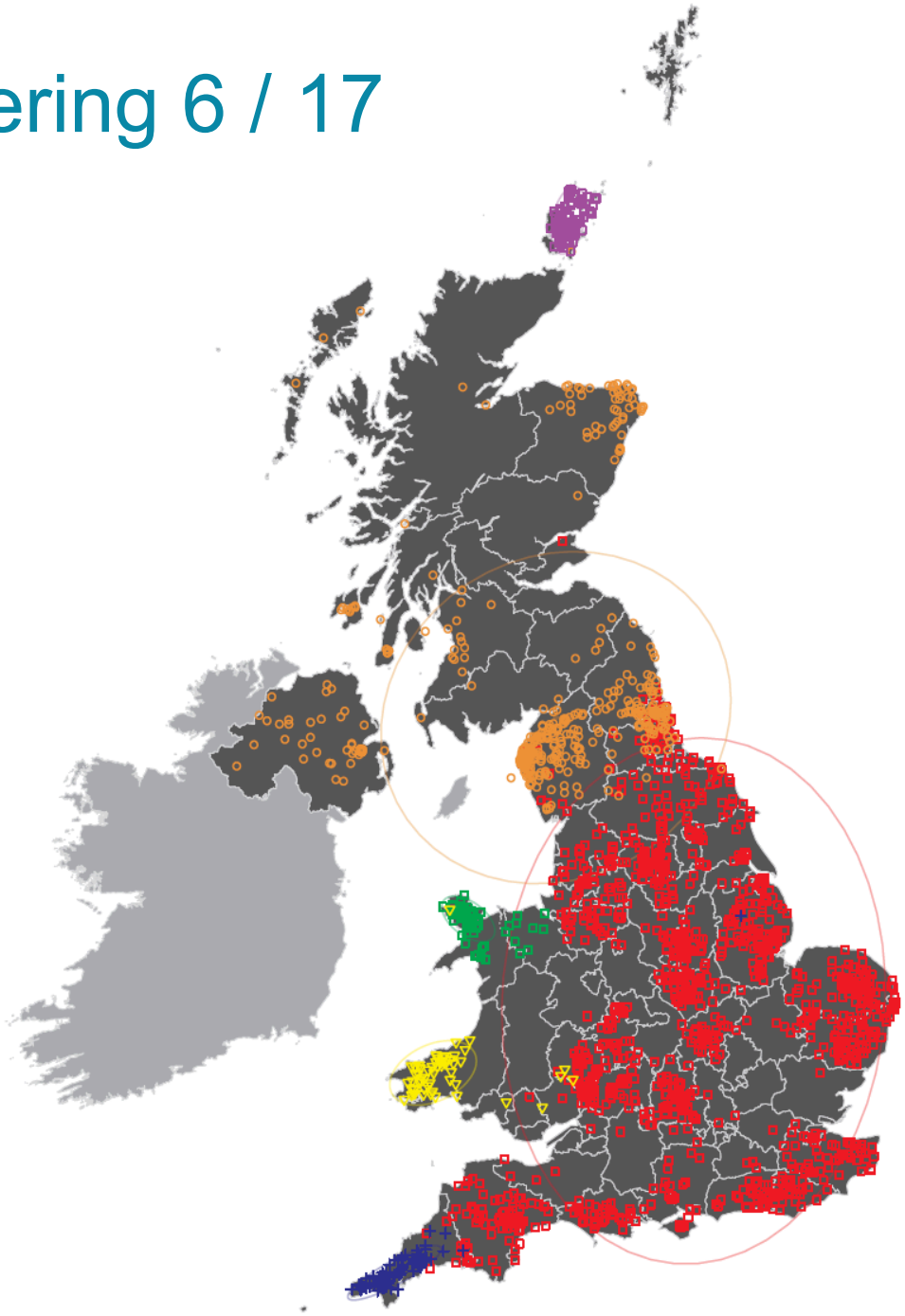
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 5 / 17



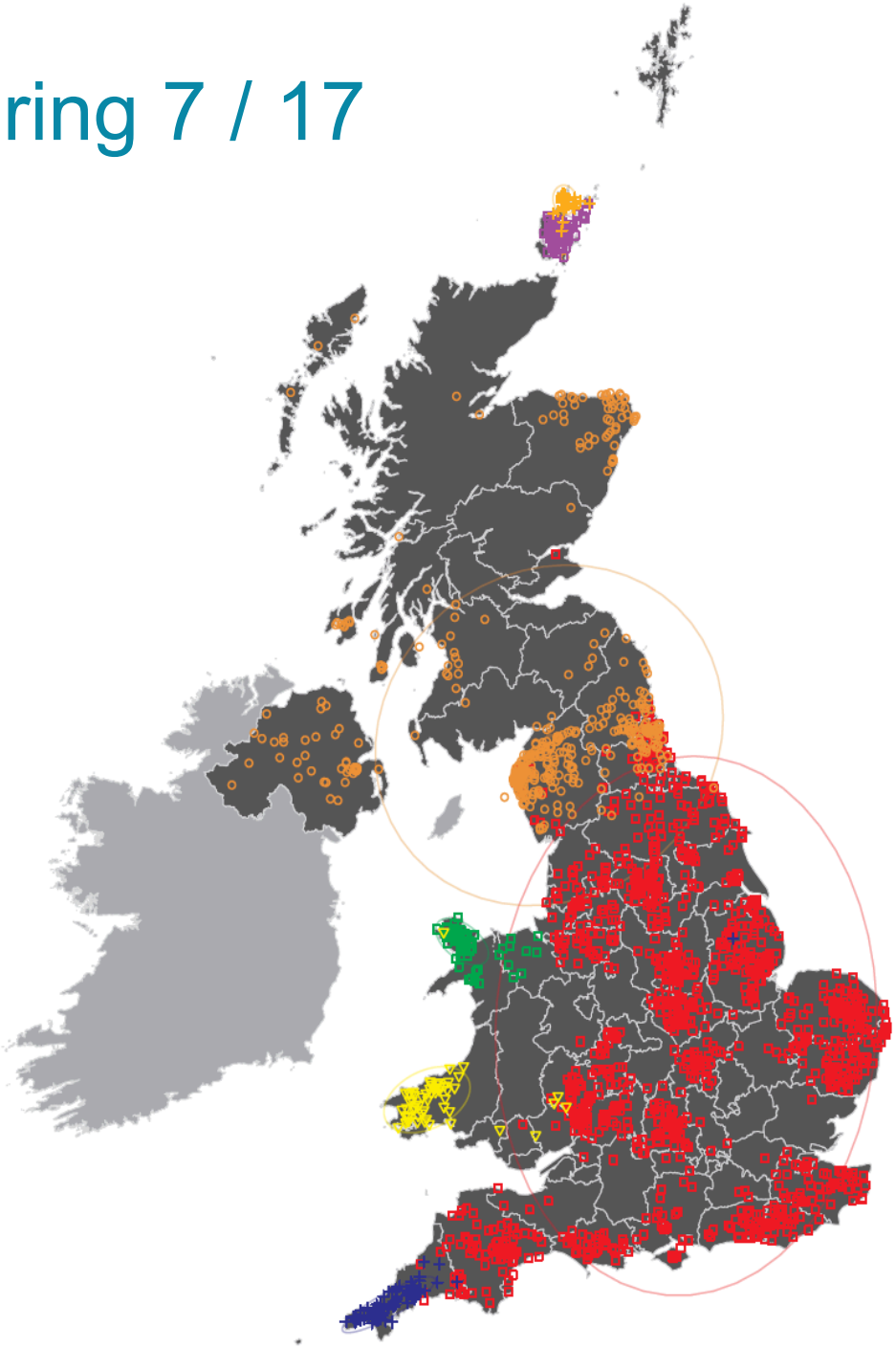
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 6 / 17



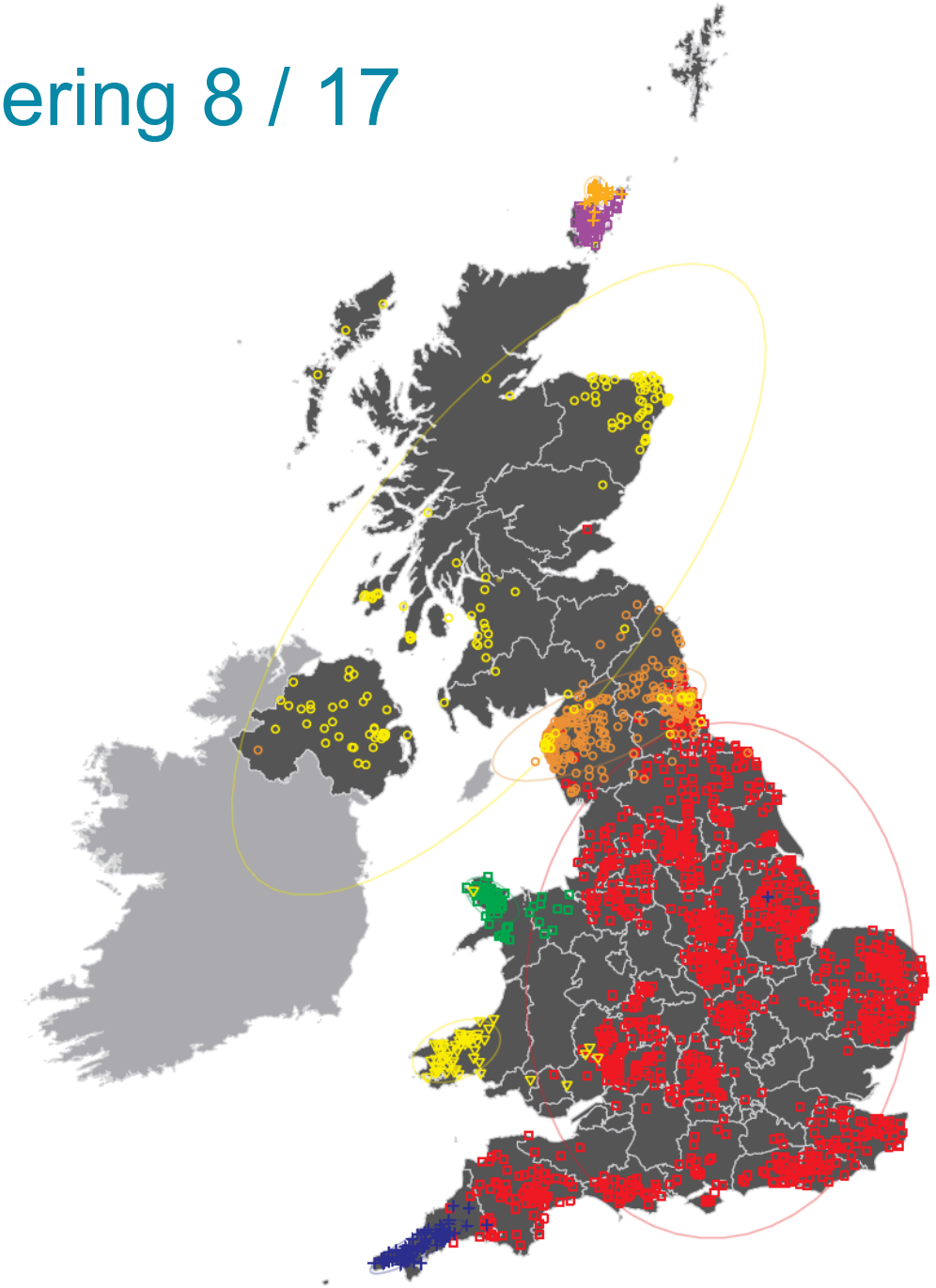
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 7 / 17



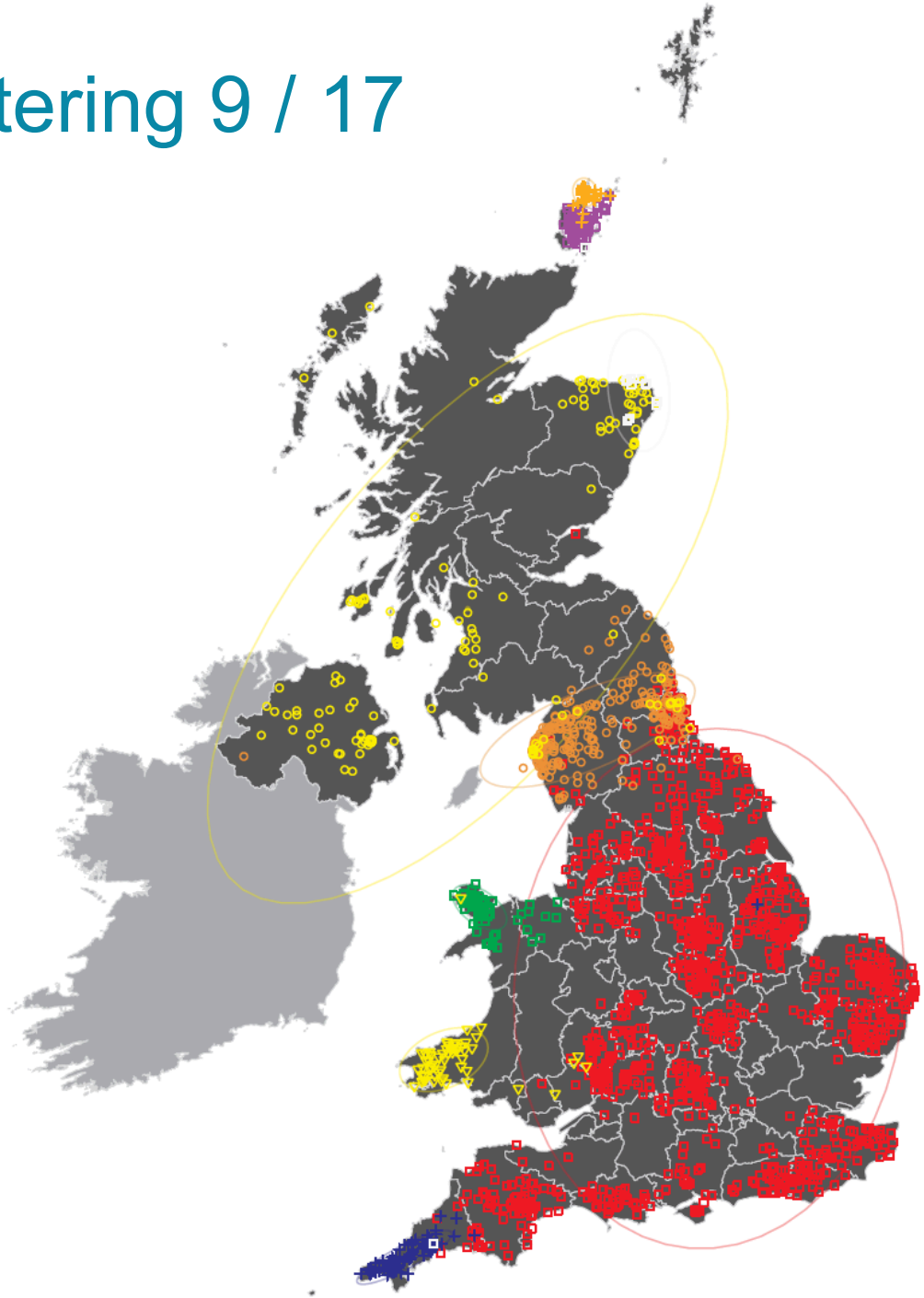
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 8 / 17



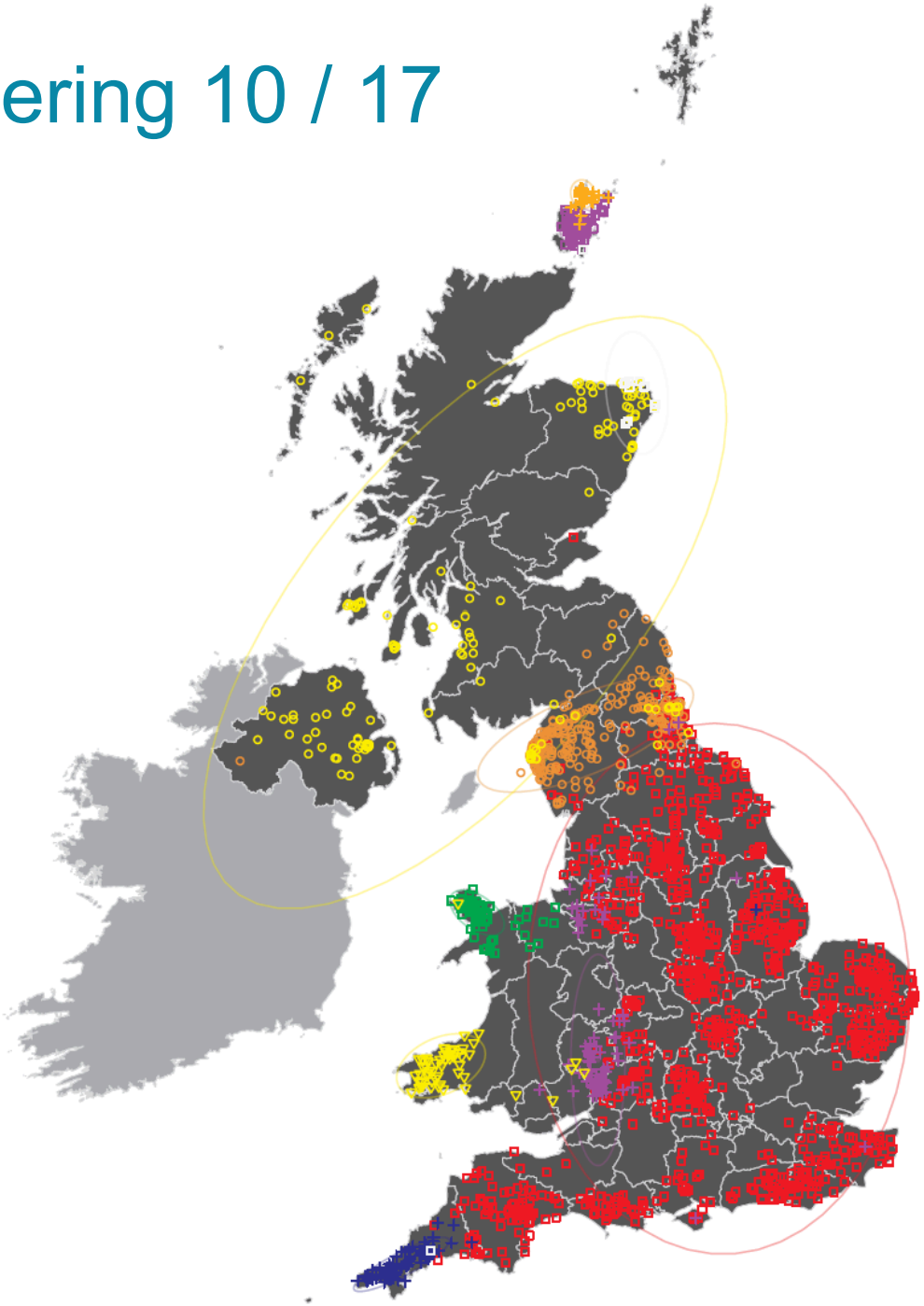
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 9 / 17



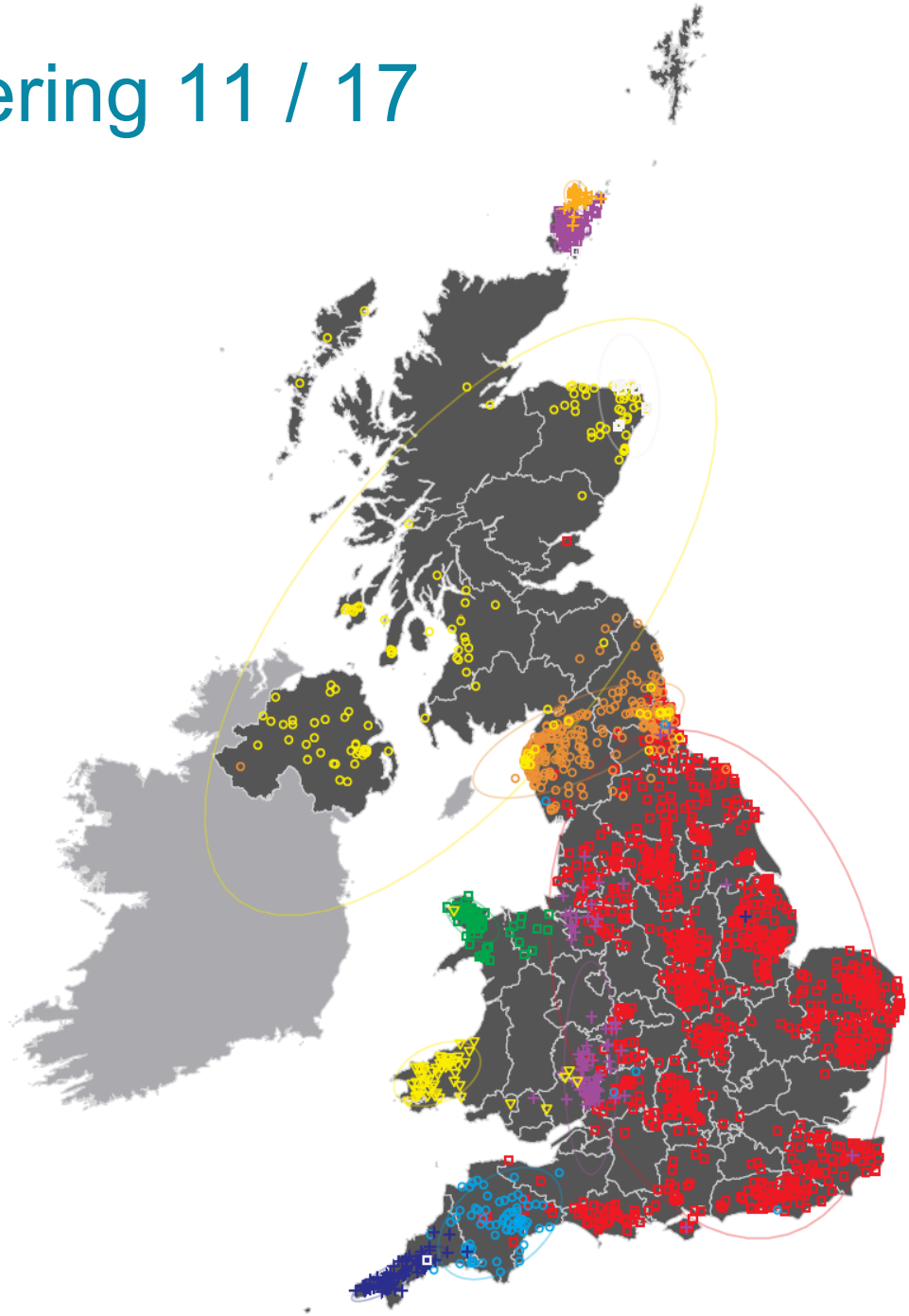
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 10 / 17



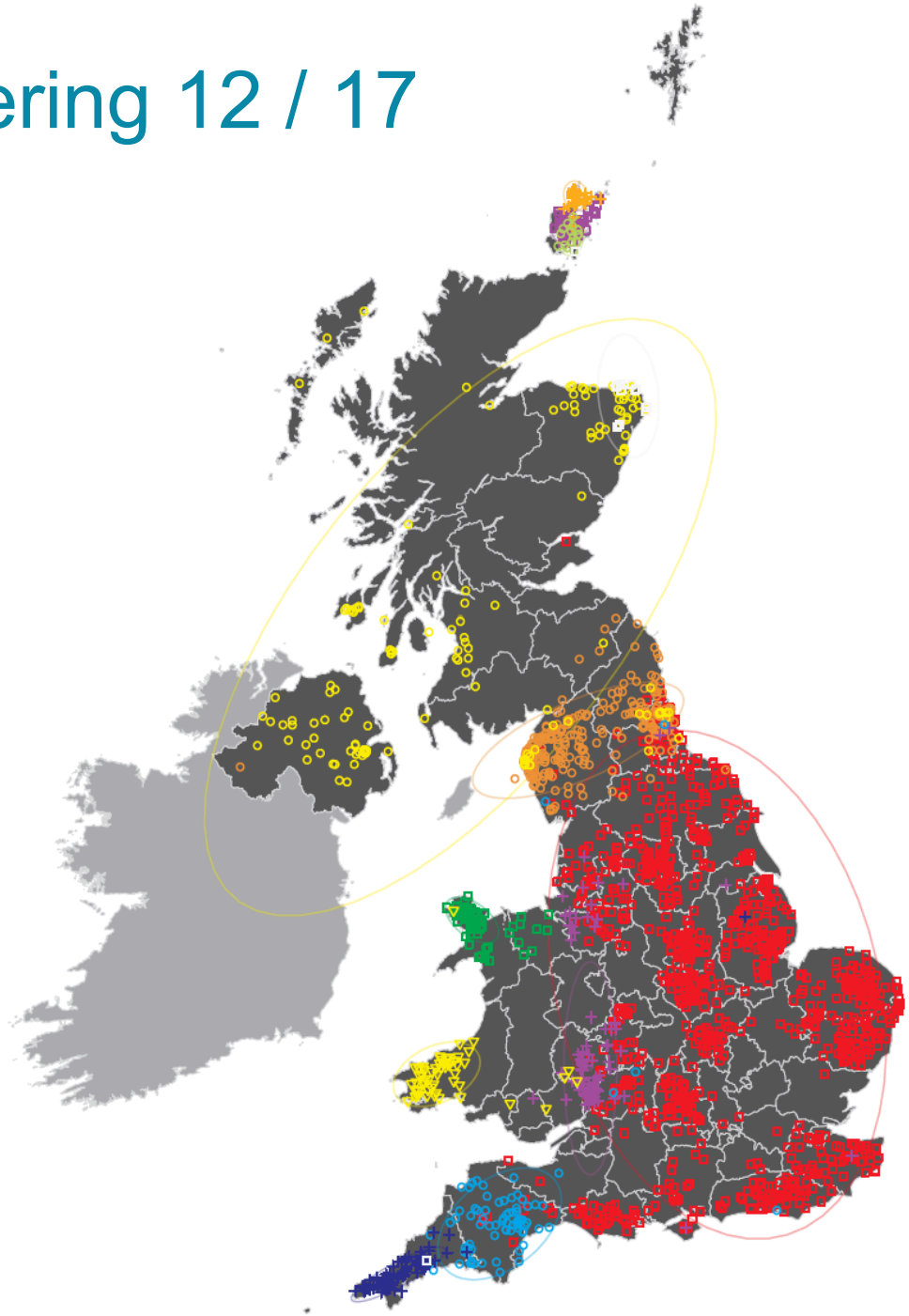
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 11 / 17



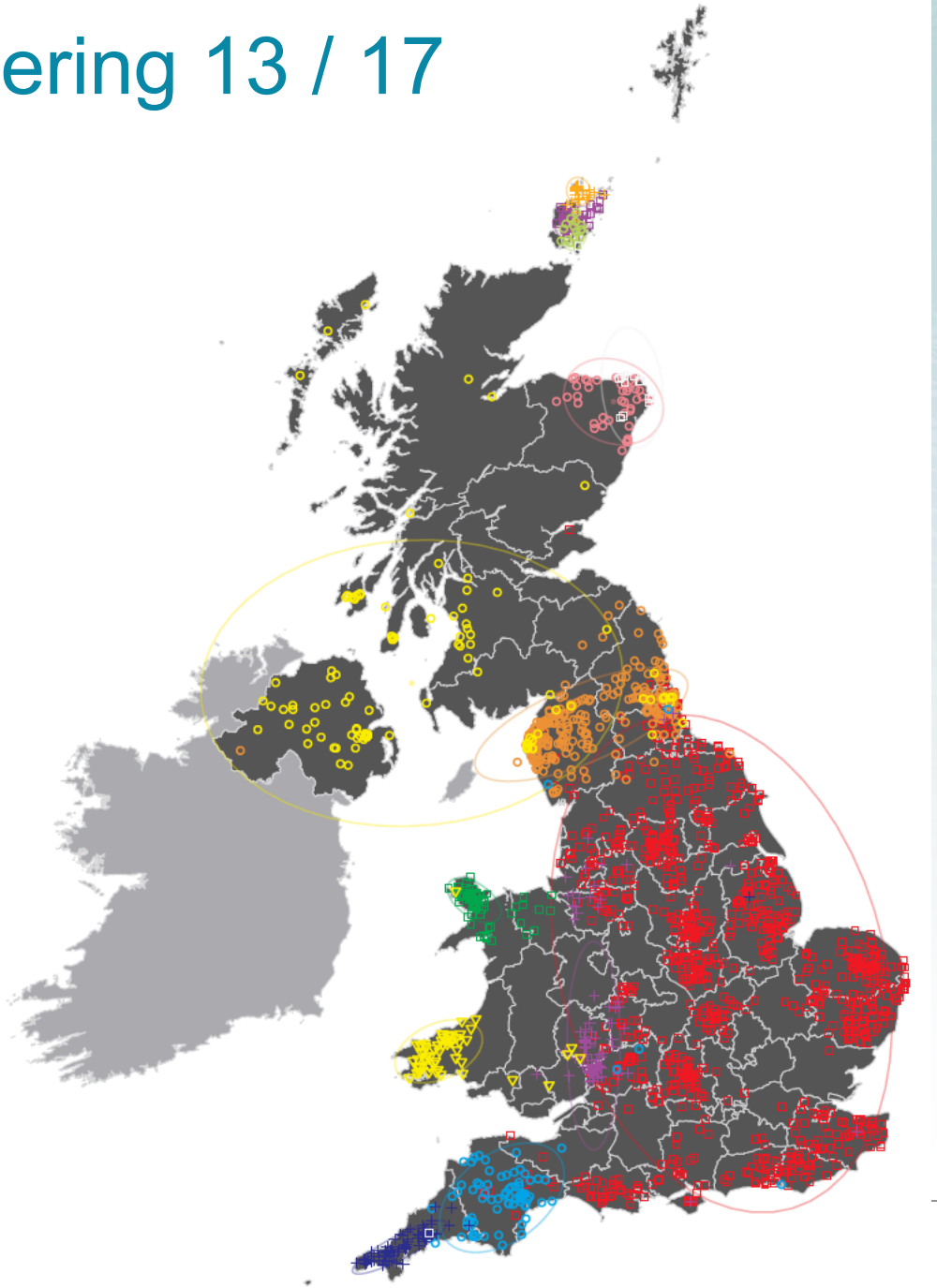
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 12 / 17



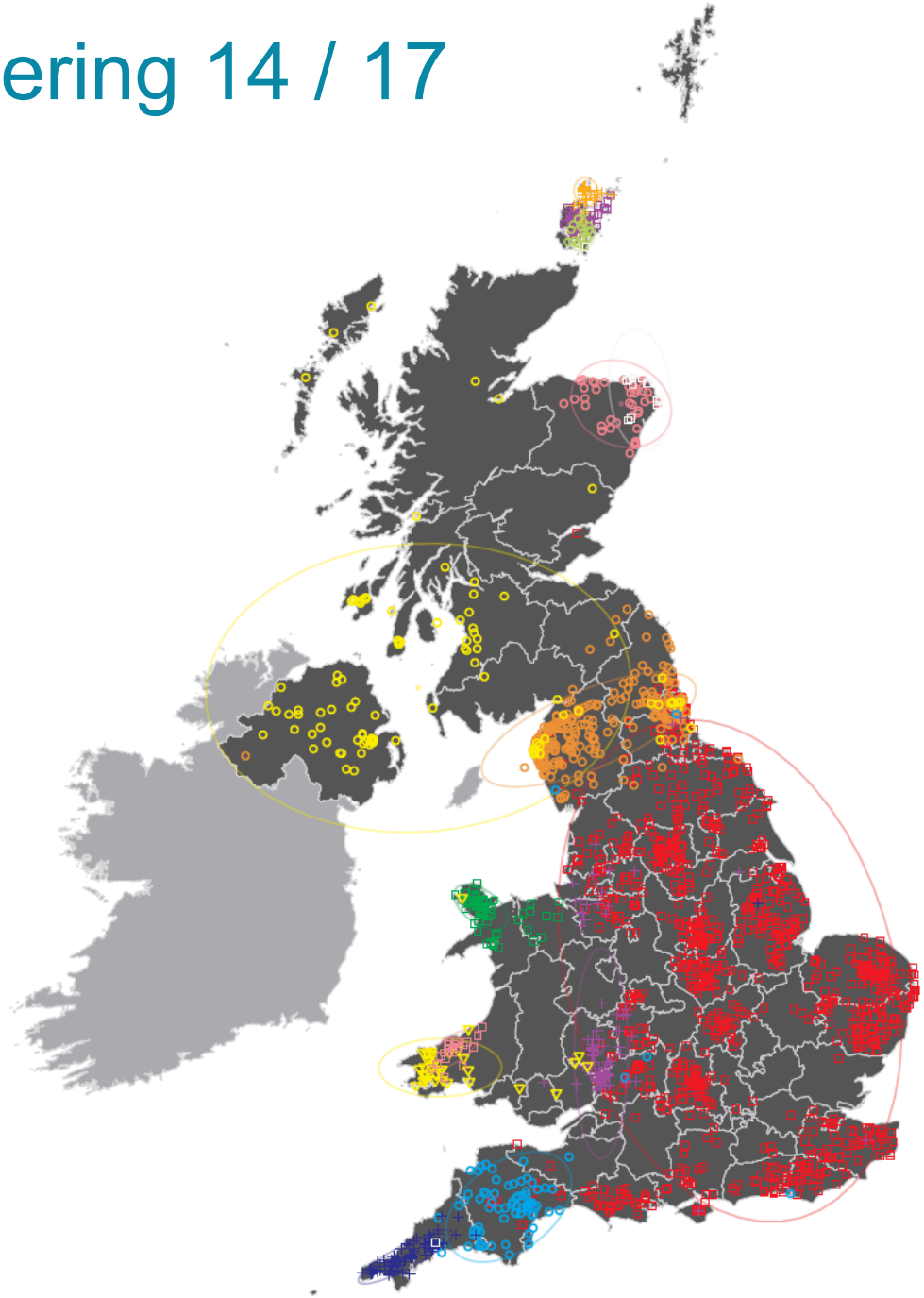
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 13 / 17



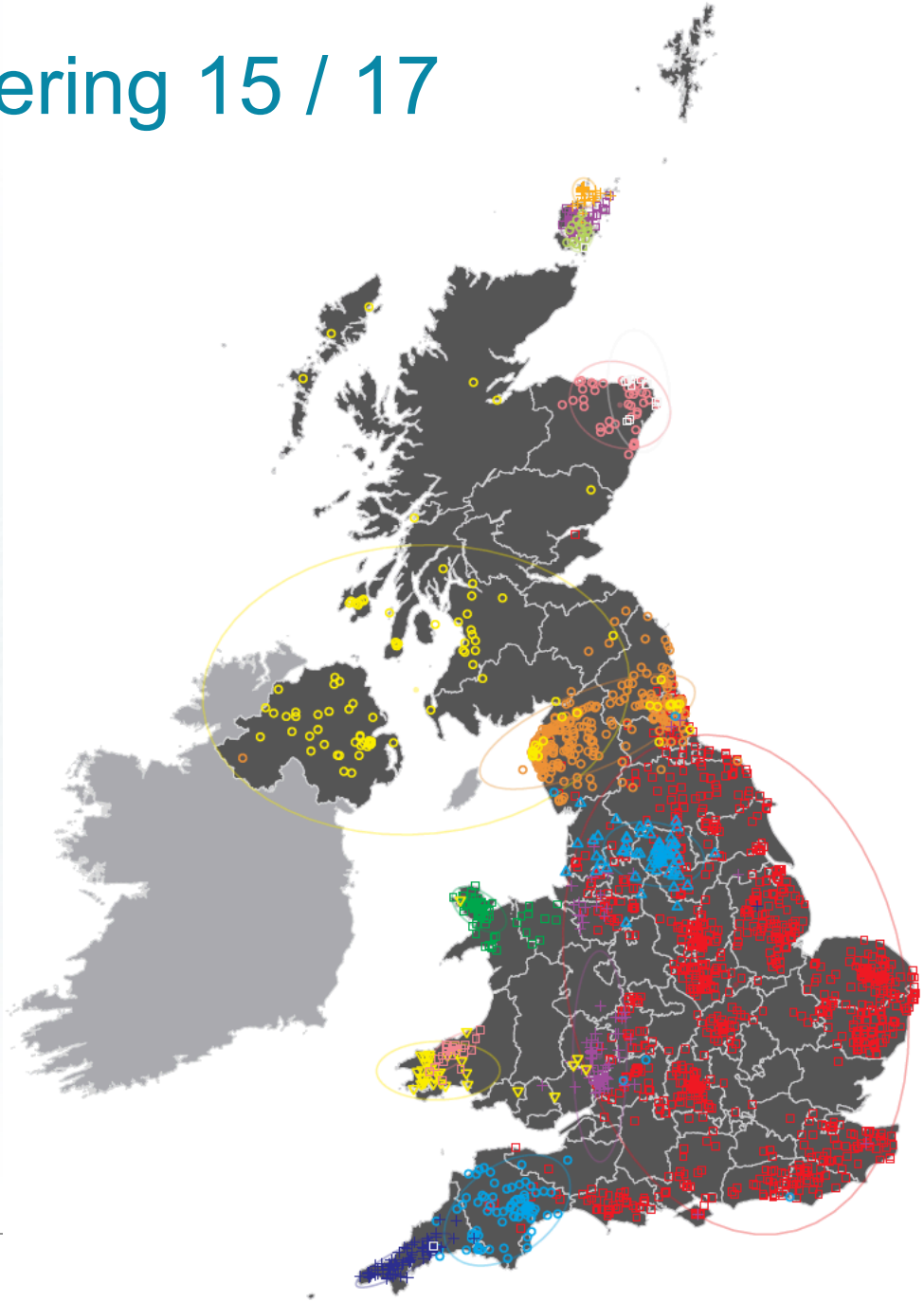
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 14 / 17



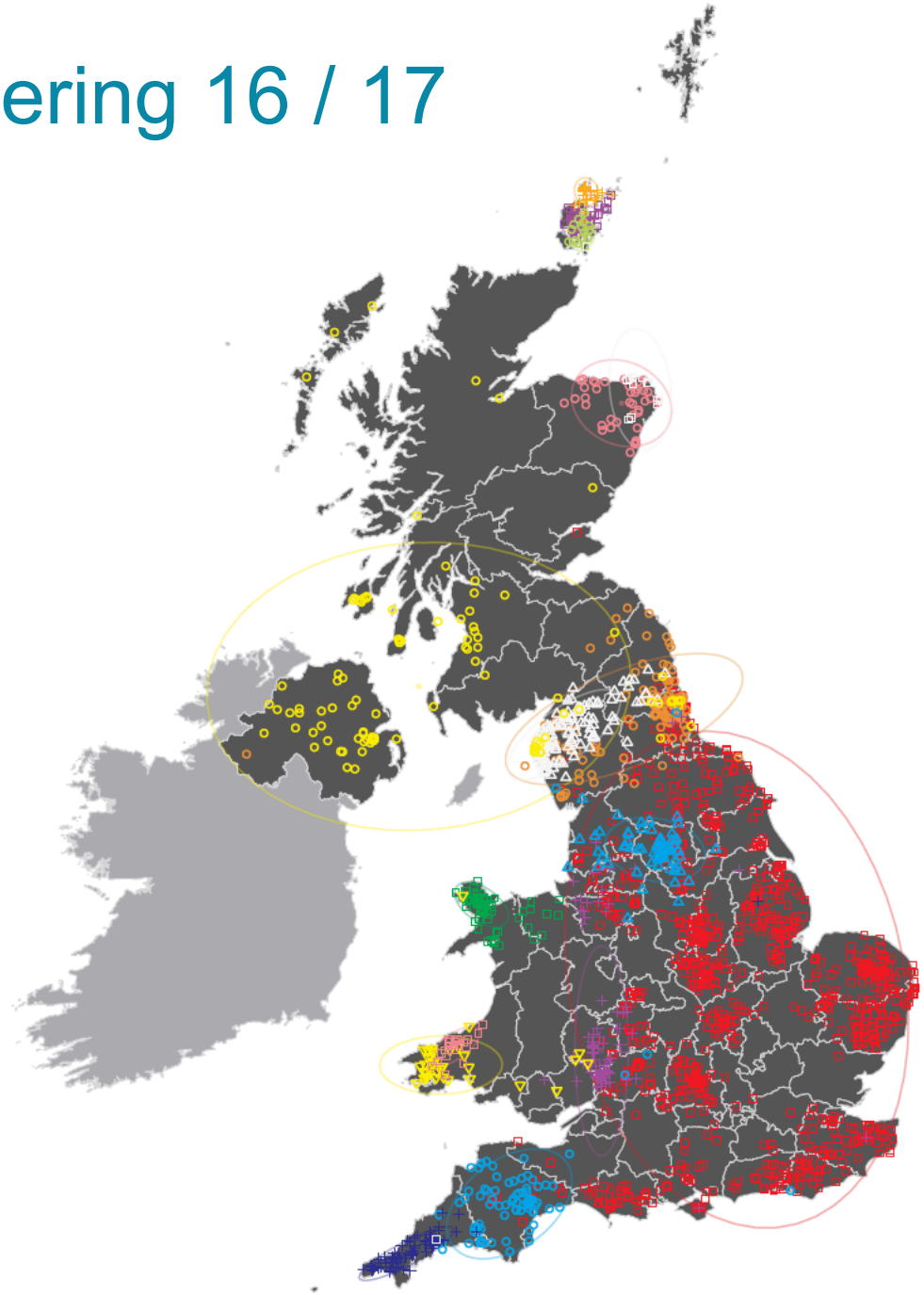
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 15 / 17



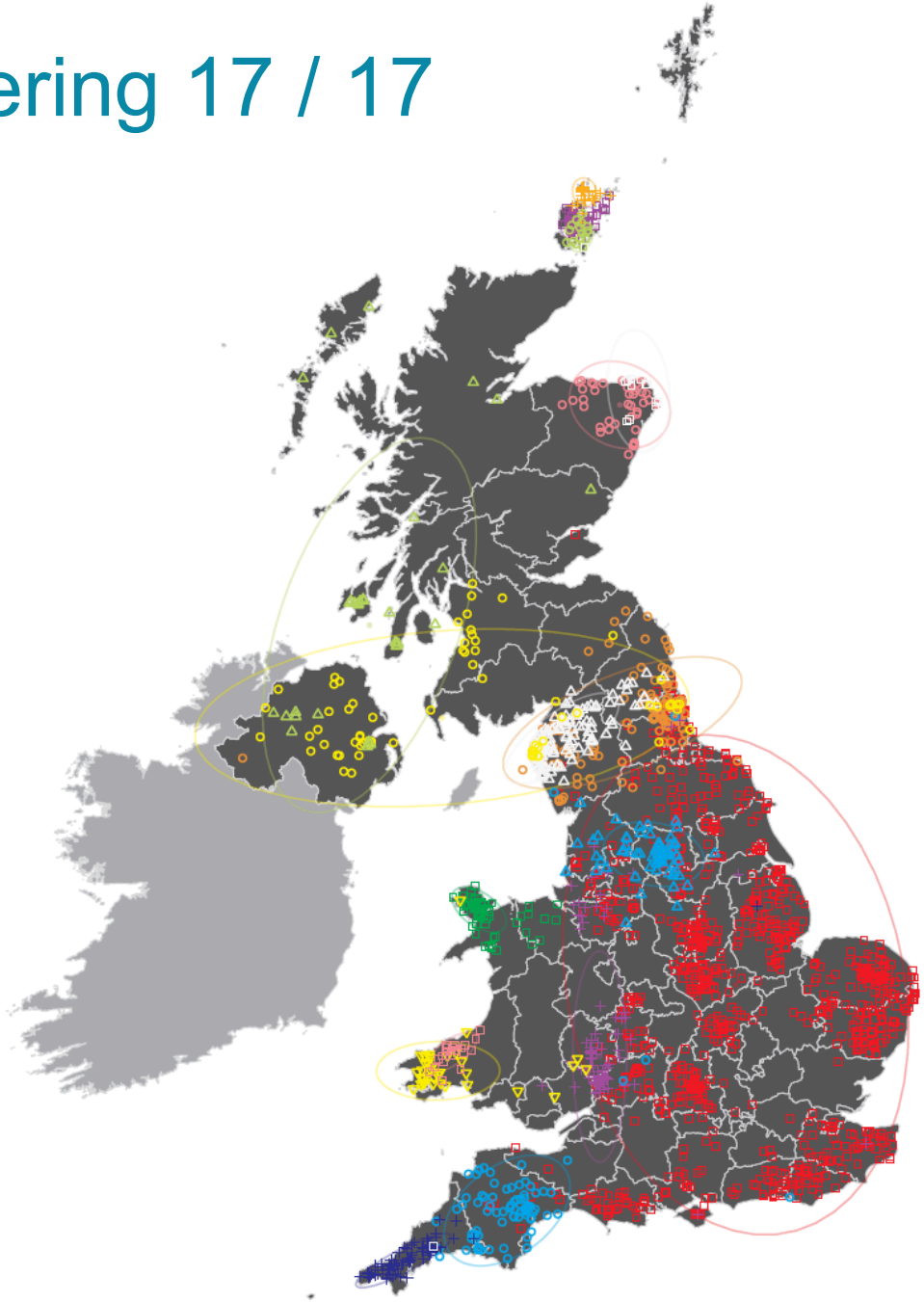
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 16 / 17



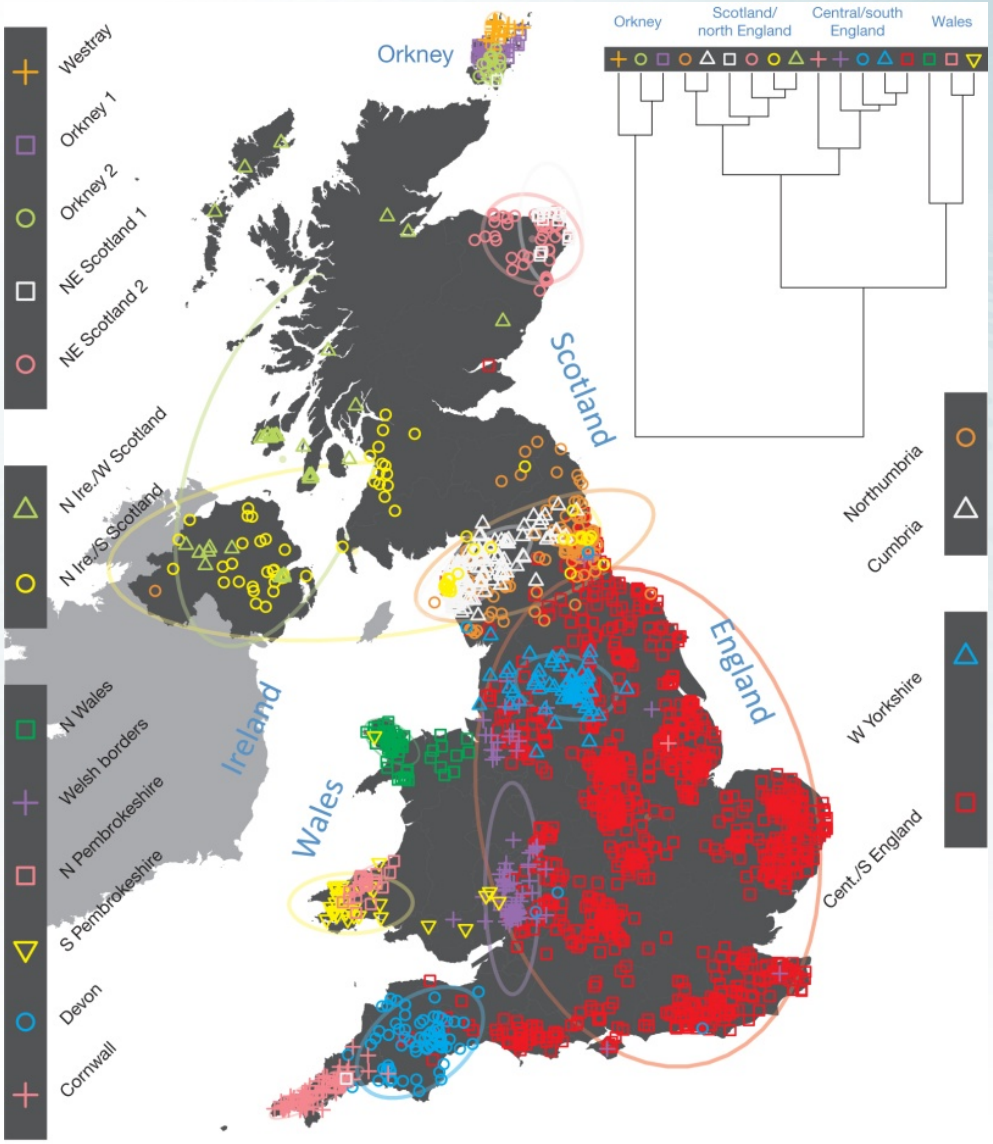
S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

Clustering 17 / 17



S Leslie *et al. Nature* **519**, 309-314 (2015)
doi:10.1038/nature14230

17 fineStructure clustering of the UK



1. Orkney
2. Wales
3. Scotland & N Ireland
4. Corwall and Devon
5. England (“Anglo-Saxon”)

Conclusion:

- Clear geographic clustering, but not monotonically
- Evidence of Saxon migration
- No single “Celtic” population
- Fairly small genetic trace from Vikings

S Leslie *et al. Nature* **519**, 309-314 (2015) doi:10.1038/nature14230

ARTICLE

doi:10.1038/nature14230

The fine-scale genetic structure of the British population

Stephen Leslie[1,2,3]*, Bruce Winney[3]*, Garrett Hellenthal[4]*, Dan Davison[5], Abdelhamid Boumertit[3], Tammy Day[3], Katarzyna Hutnik[3], Ellen C. Royrvik[3], Barry Cunliffe[6], Wellcome Trust Case Control Consortium 2†, International Multiple Sclerosis Genetics Consortium†, Daniel J. Lawson[7], Daniel Falush[8], Colin Freeman[9], Matti Pirinen[10], Simon Myers[11], Mark Robinson[12], Peter Donnelly[9,11]§ & Walter Bodmer[3]§

UNIVERSITY OF OXFORD

Nature 19 March 2015

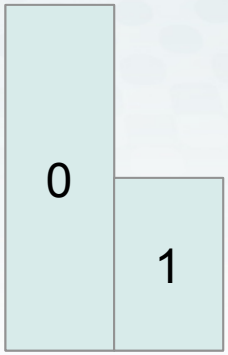
Genetic association study of a disease

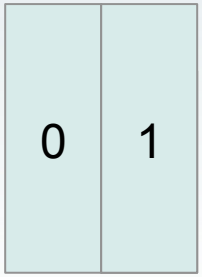
Individuals

Cases

Controls

Genotypes Single nucleotide polymorphisms (SNPs)





Question Are the genotype distributions different between cases and controls?

Population structure and association studies

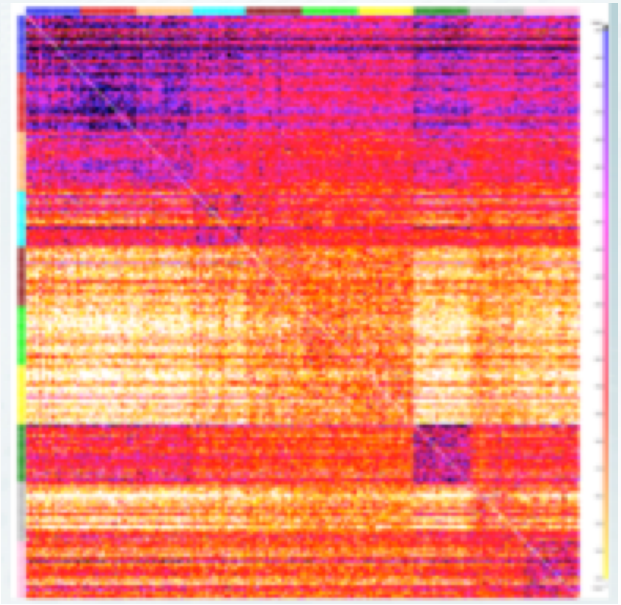
- SNP that has no effect on heart disease but has different regional frequencies
  - Allele 1 frequency 0.23 in Helsinki region
  - Allele 1 frequency 0.35 in Oulu region
- Does not show association with disease in Helsinki or in Oulu
- What happens if we do not match well regions of case and control ?

Problem with population structure and association studies

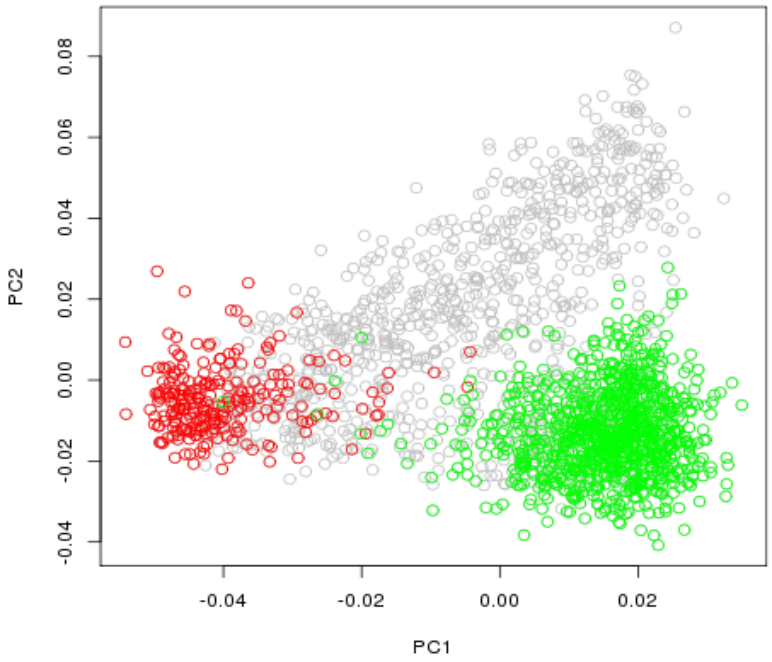
- SNP that has no effect on heart disease but has different regional frequencies
  - Allele 1 frequency 0.23 in Helsinki region
  - Allele 1 frequency 0.35 in Oulu region
- Consider sampling
  - 2000 cases (500 from H and 1500 from O). Allele 1 freq is 0.32
  - 2000 controls (1500 from H and 500 from O).
  - Allele 1 freq is 0.26
- False association that allele 1 increases risk for heart disease !
- Cases and controls must be matched !

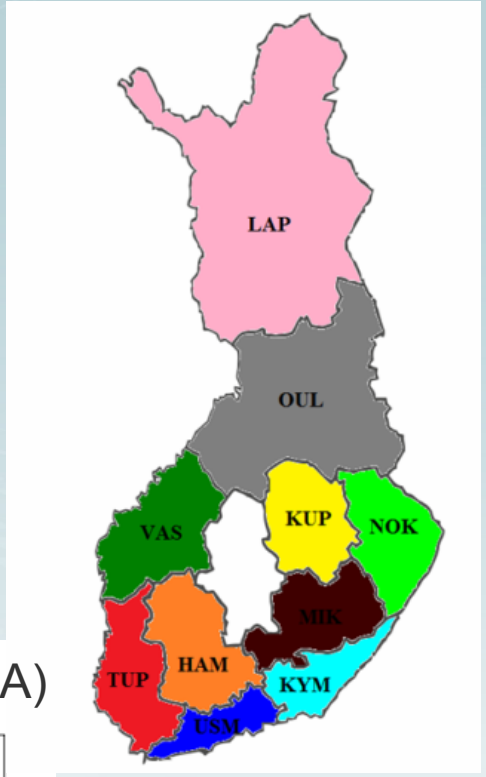
Matching cases and controls

- Often we do not know regional origins of samples or they may not be informative of genetic background
- But we can infer genetic similarity and adjust the analyses for that



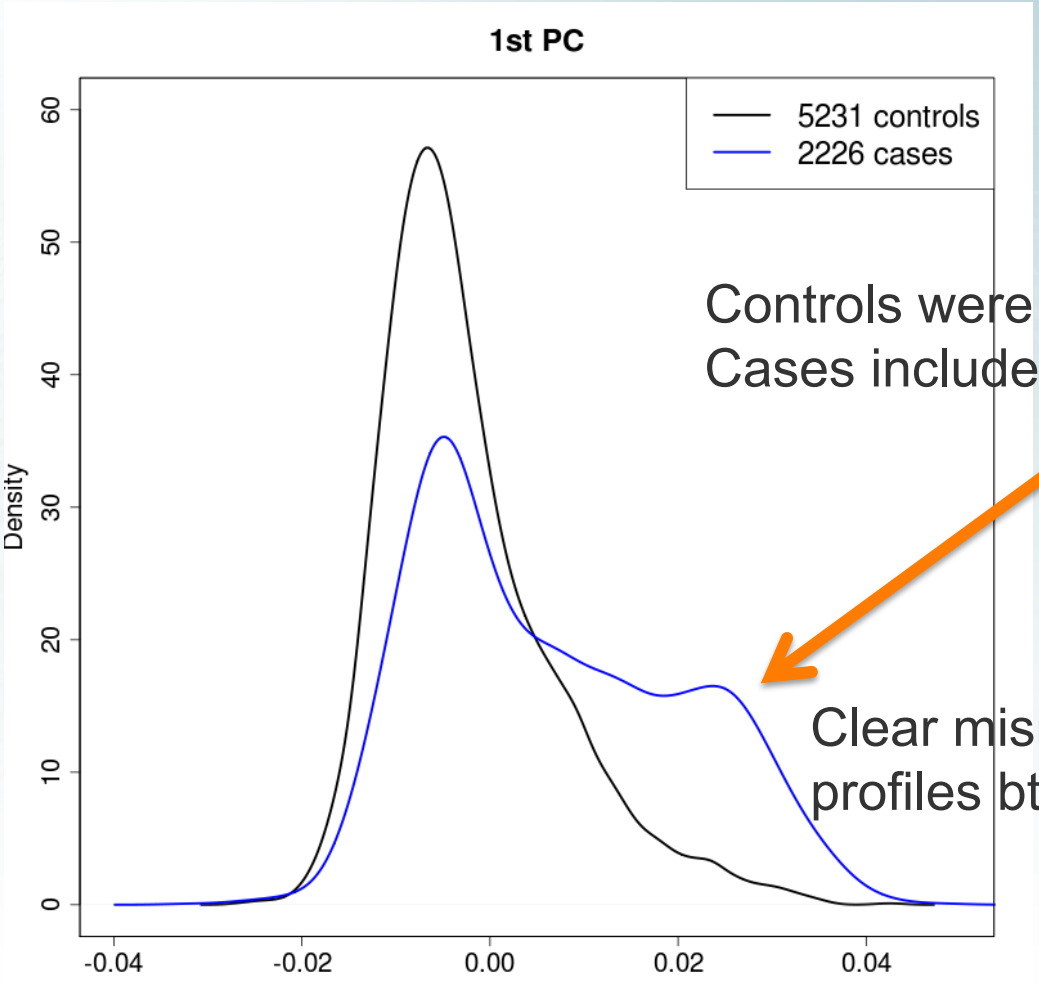
Principal components analysis (PCA)





Sini Kerminen

Psoriasis and the British Isles data

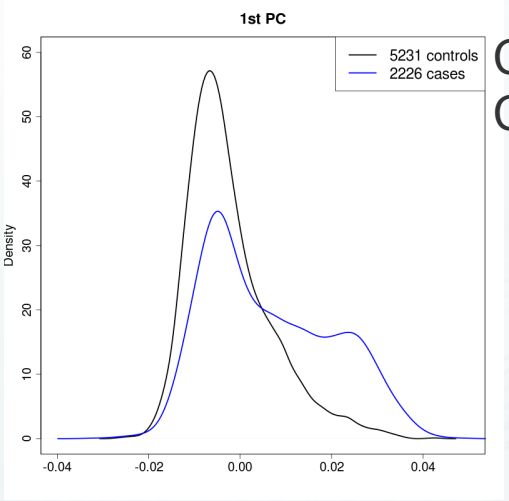


Controls were all from the UK
Cases included 500 Irish samples

Clear mismatch in ancestry profiles btw cases / controls!

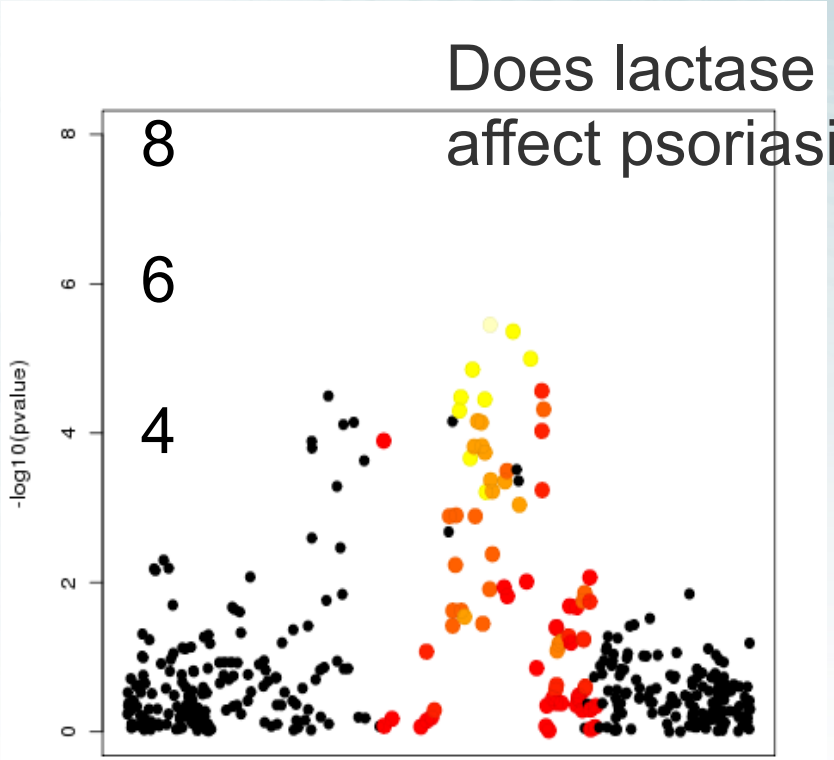
Strange et al. 2010 Nat Genet

Psoriasis and the British Isles data



Controls were all from the UK
Cases included 500 Irish samples

Does lactase gene really affect psoriasis susceptibility ?

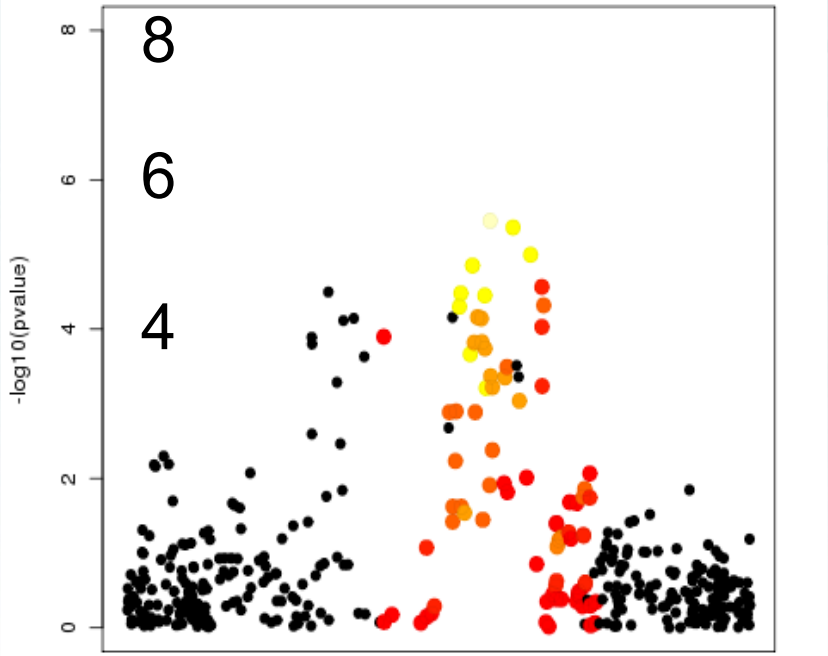


Region around LCT

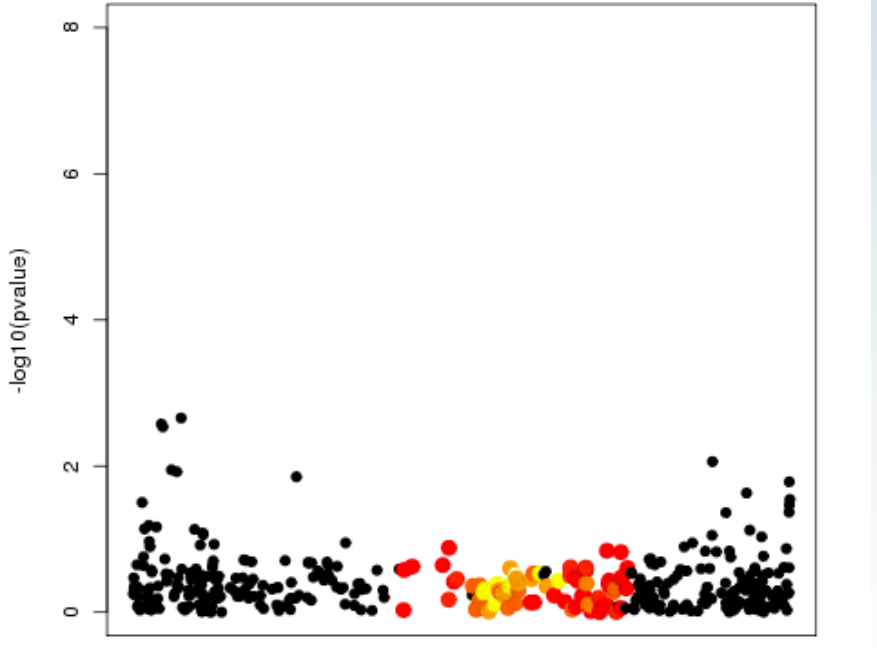
Psoriasis and the British Isles data

Controls were all from the UK
Cases included 500 Irish samples

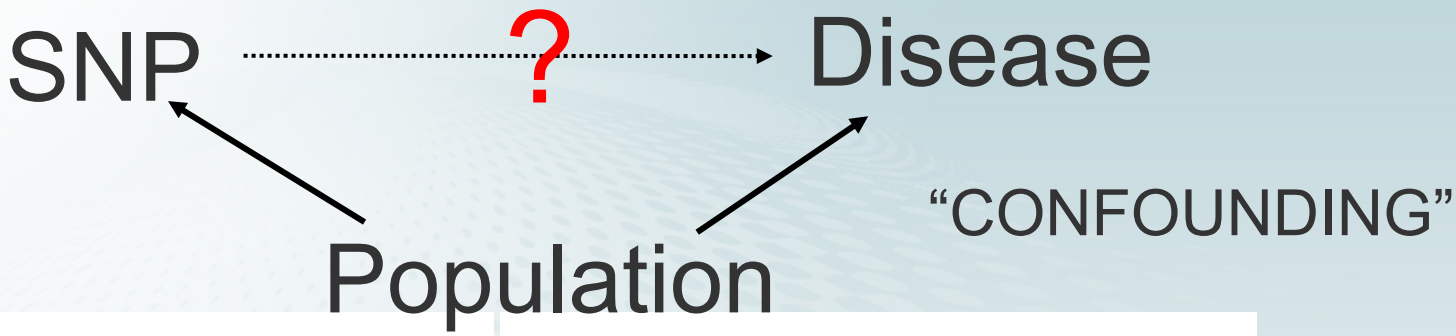
Does lactase gene really affect psoriasis susceptibility?

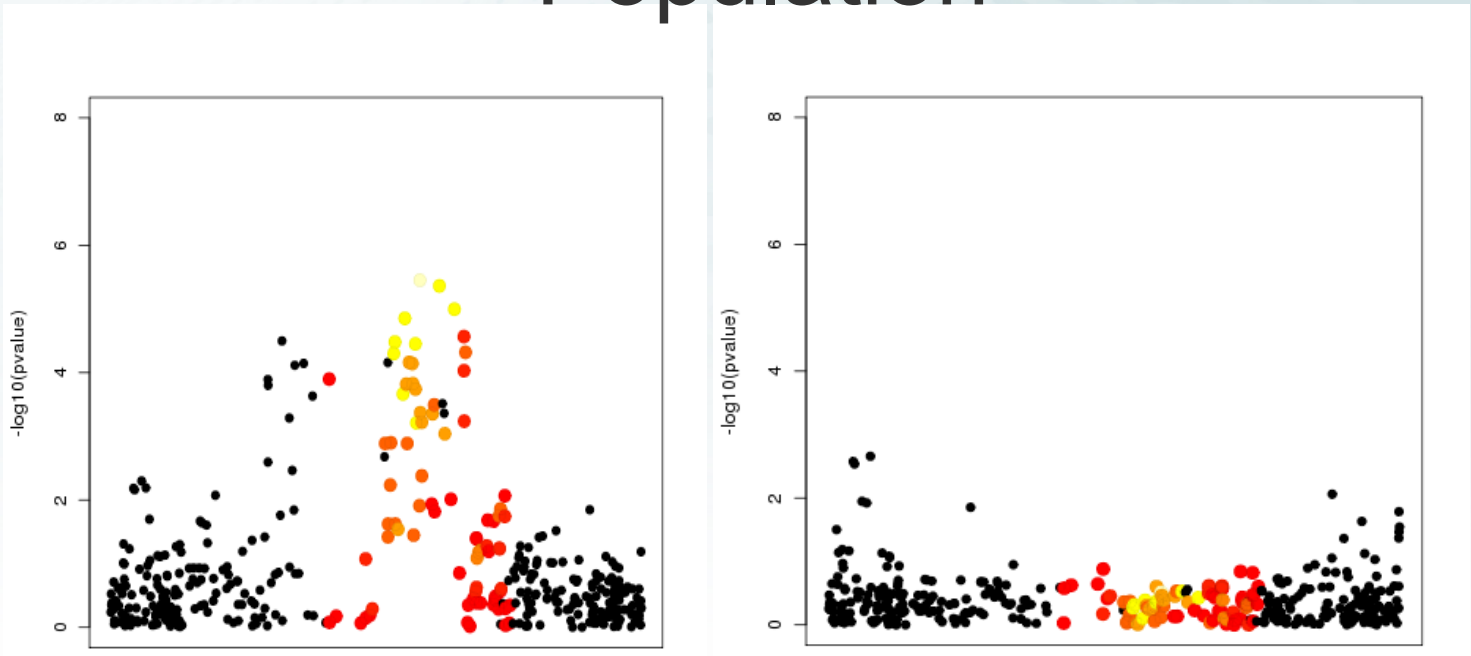


Probably not, since the signal can be explained by ancestry (1st PC).

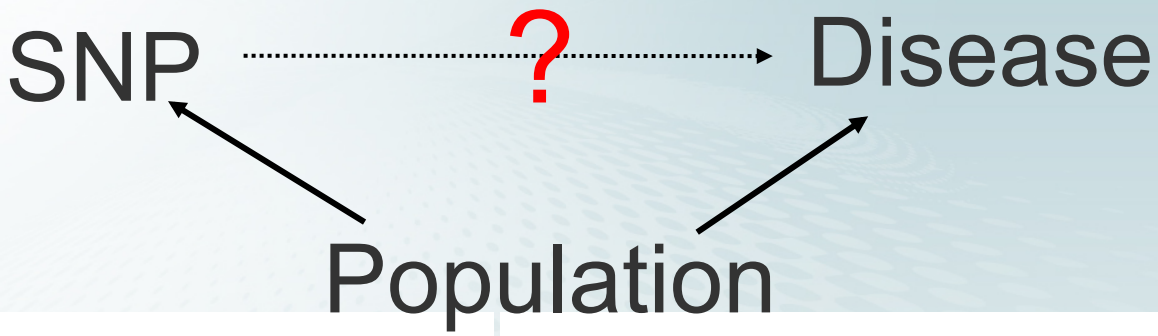


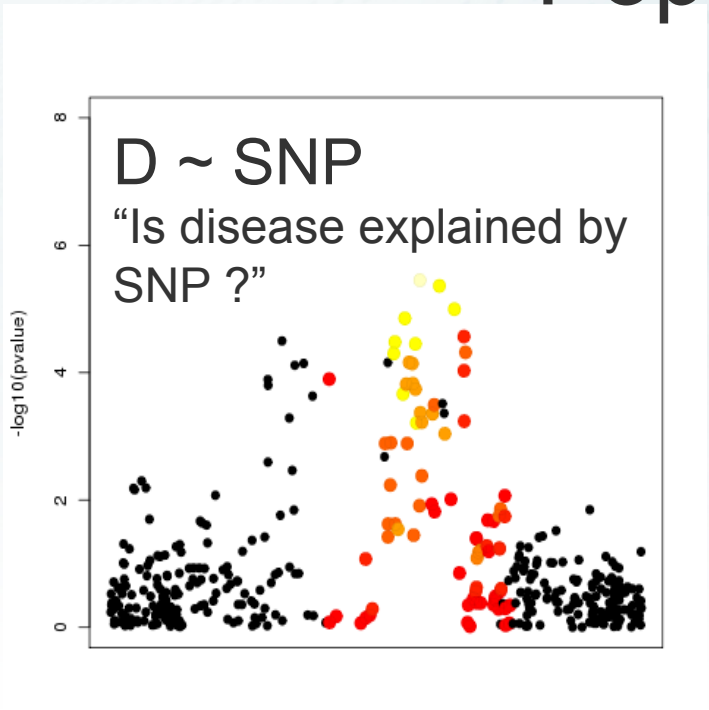
Psoriasis and the British Isles data

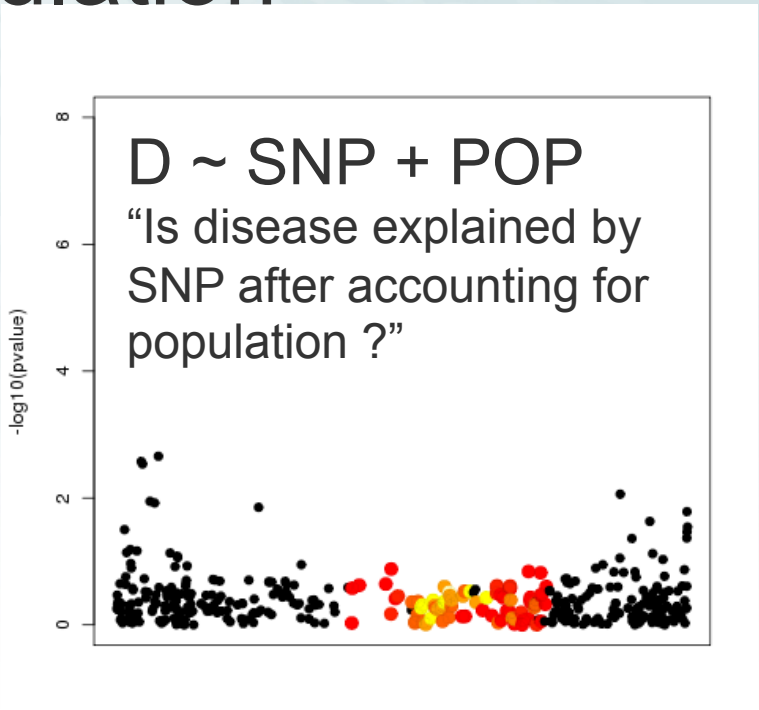




Psoriasis and the British Isles data

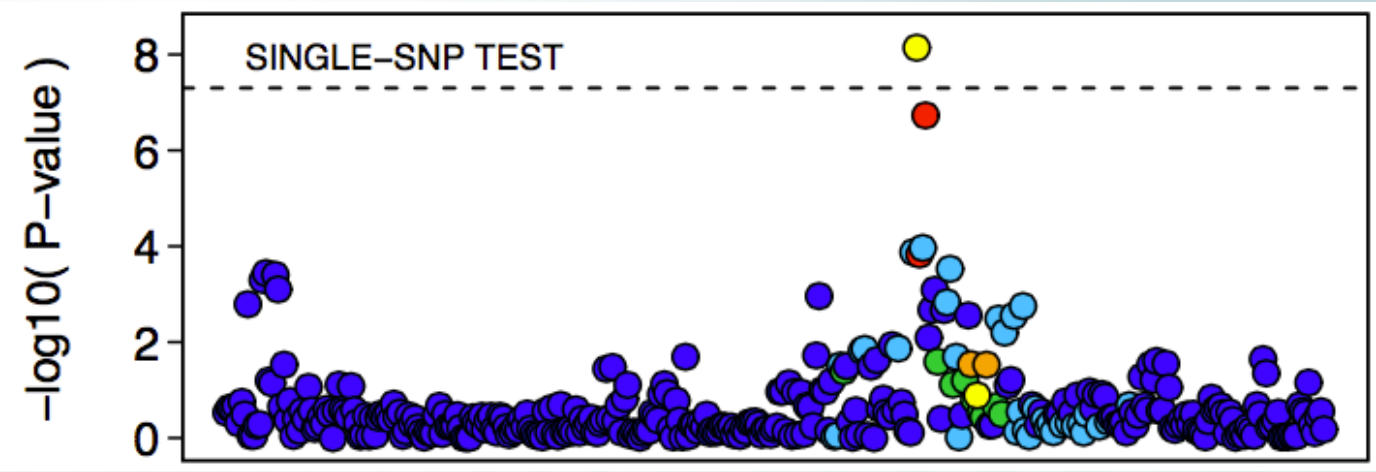






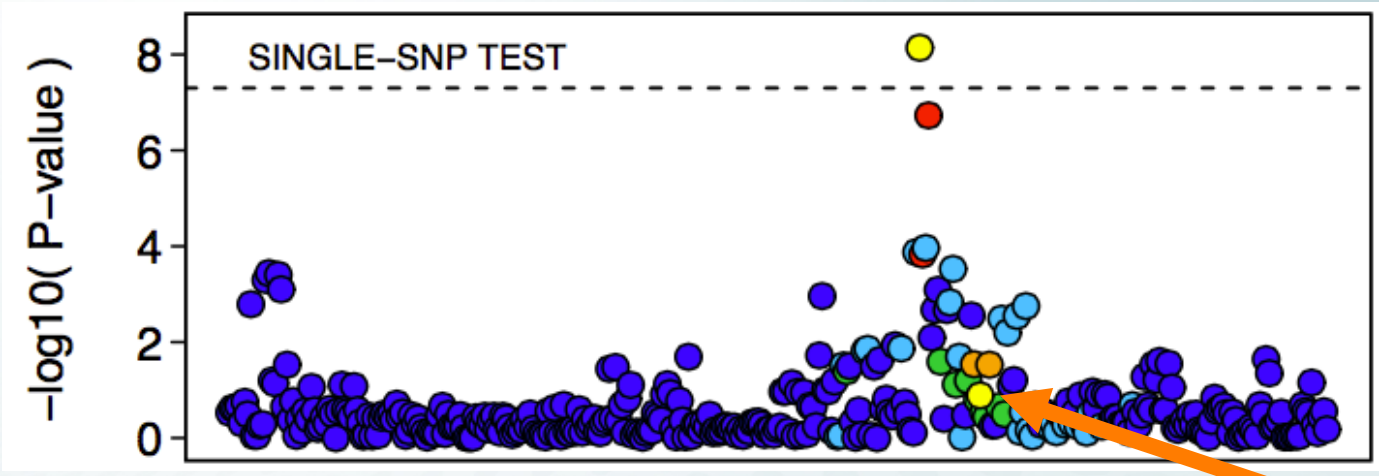
Parkinson's disease study

› Genomic region 4q22 around SNCA gene shows association

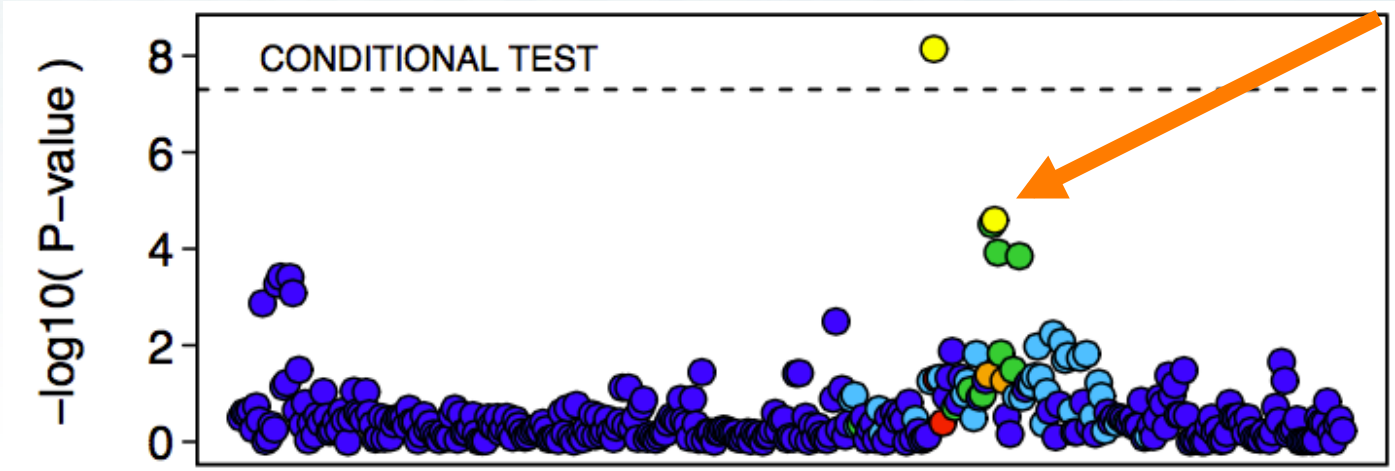


Parkinson's disease study

- Genomic region 4q22 around SNCA gene shows association



- Re-analysis conditioning on the top SNP



A new SNP pops up after conditioning on the top SNP

What's happening here?

Risk of 2-SNP combinations

		Risk
G	G	1.00
G	A	1.16
A	G	1.31
A	A	1.64

Masking effect of the second SNP

2nd SNP RISK increases →

		rs7687945 G	rs7687945 A		
rs356220	G	1 20.8%	1.16 (1.04-1.29) 41.9%	1.11	1.26 (1.16-1.37)
	A	1.31 (1.17-1.47) 27.9%	1.64 (1.41-1.90) 9.4%	1.39	
		1.18	1.25		
		1.07 (0.98-1.15)			

1st SNP RISK increases →

Spencer et al. 2011 Hum Mol Gen

Masking effect of the second SNP

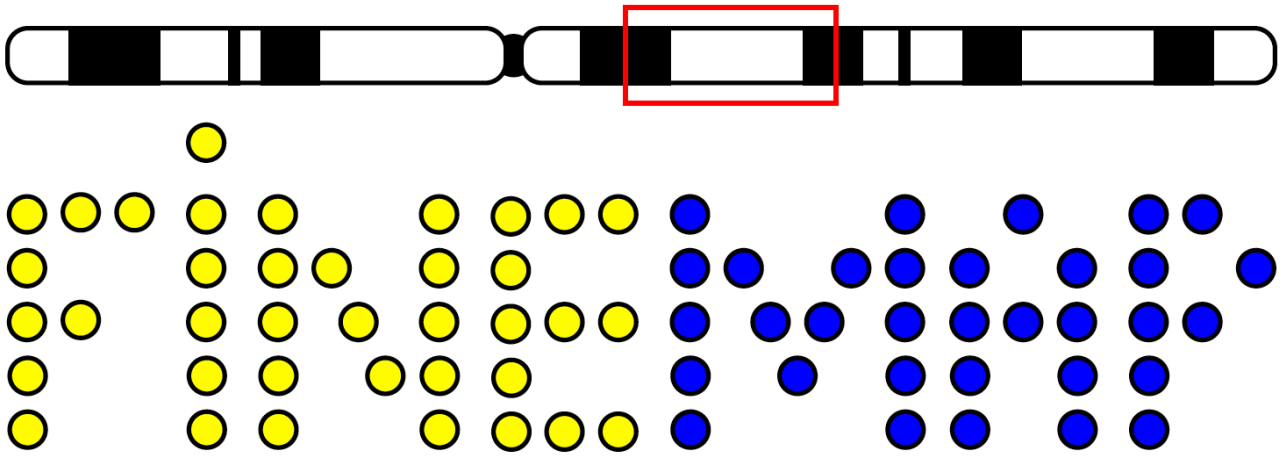
Risk allele at 2nd SNP usually go with protective allele of the 1st SNP -> masking effect

2nd SNP RISK increases →

		rs7687945 G	rs7687945 A		
rs356220	G	1 20.8%	1.16 (1.04-1.29) 41.9%	1.11	1.26 (1.16-1.37)
	A	1.31 (1.17-1.47) 27.9%	1.64 (1.41-1.90) 9.4%	1.39	
		1.18	1.25		
		1.07 (0.98-1.15)			

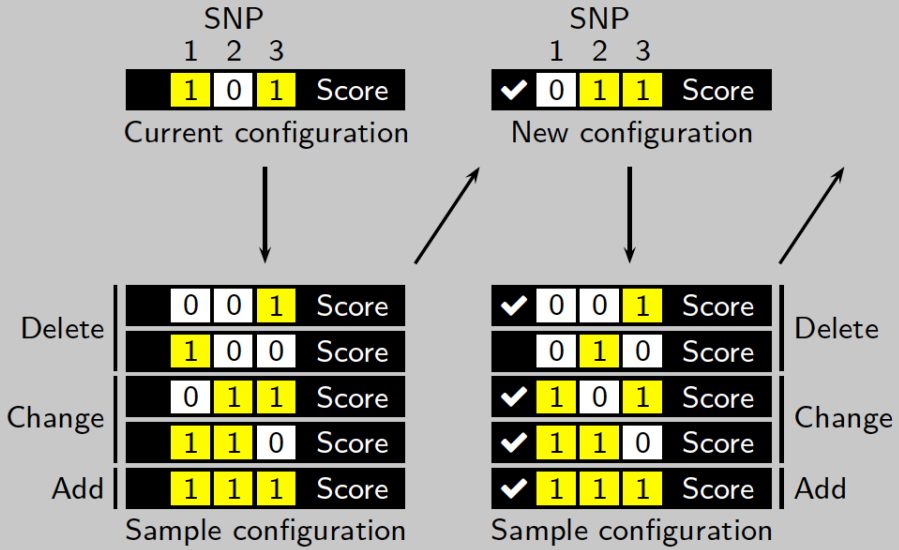
1st SNP RISK increases →

Spencer et al. 2011 Hum Mol Gen

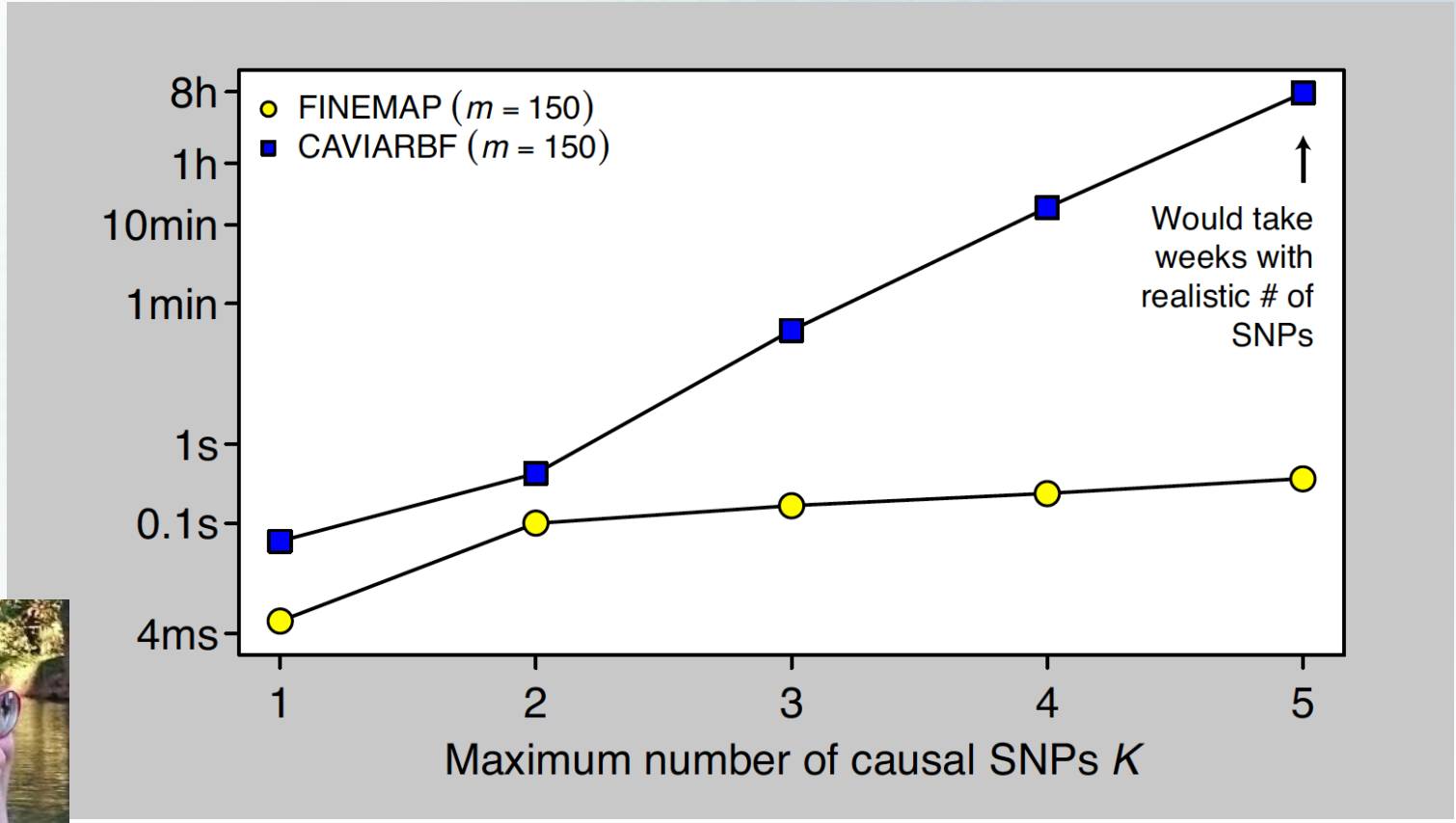


Shotgun stochastic search (Hans et al., 2007)

- Algorithm for exploring the space of causal configurations

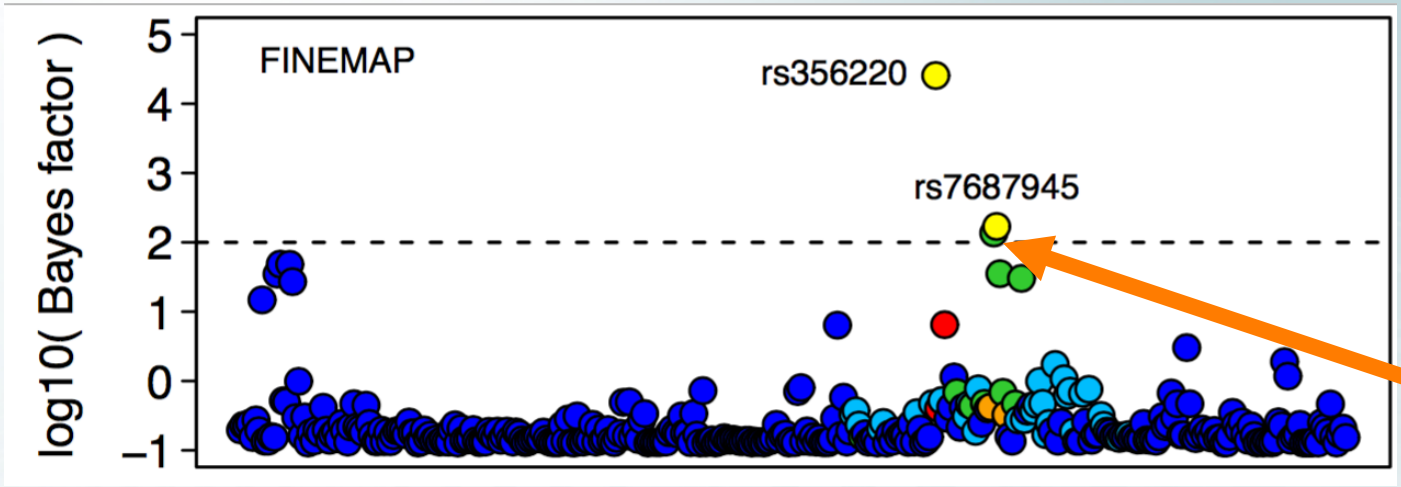


FINEMAP equally accurate but 1000s of times faster than exhaustive search

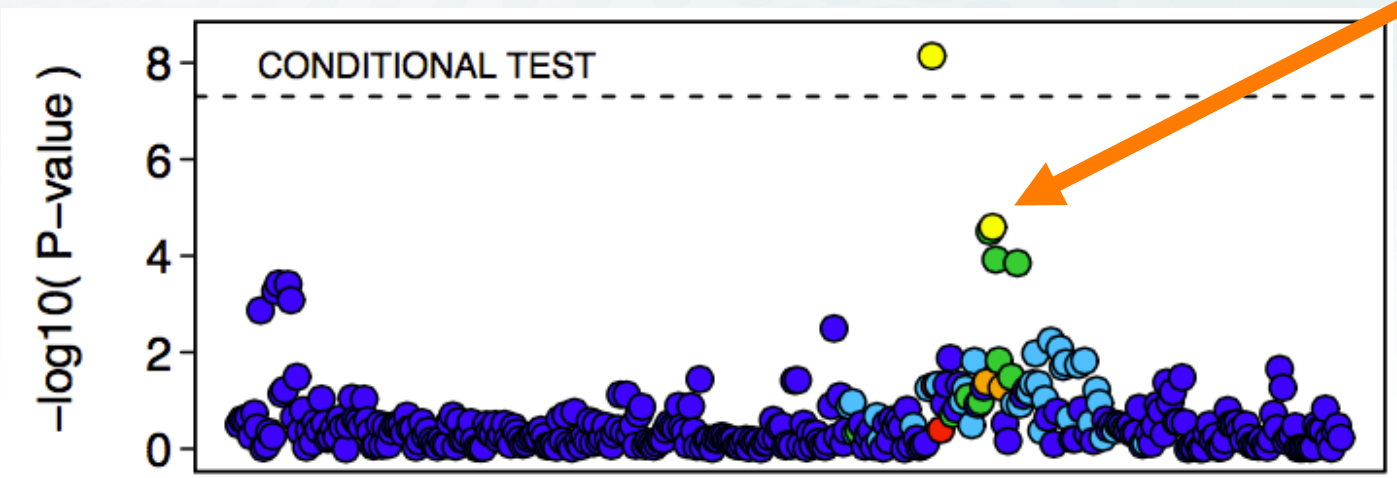


FINEMAP automates conditioning

- Genomic region 4q22 around SNCA gene using FINEMAP



- Re-analysis conditioning on the top SNP



Summary

- Reading pieces of our recent history from our genome is becoming possible
- Knowing ancestral background of samples is crucial in case-control studies
- Genome analysis is full of statistical problems
  - hidden Markov models for detecting shared segments
  - clustering to group individuals
  - variable selection to automate model search

Acknowledgements

Sini Kerminen

Samuli Ripatti

Christian Benner