Identifying Genetic Variants that Influence Atopic Dermatitis

Lavinia Paternoster

Senior Lecturer in Genetic Epidemiology

MRC Integrative Epidemiology Unit, University of Bristol, UK

19th November 2018

Outline

- Atopic dermatitis as a complex genetic trait
- Statistical methods to identify associations
- How to conduct a Genome-Wide Association study (GWAS)
- What GWAS of AD have found
- Next steps

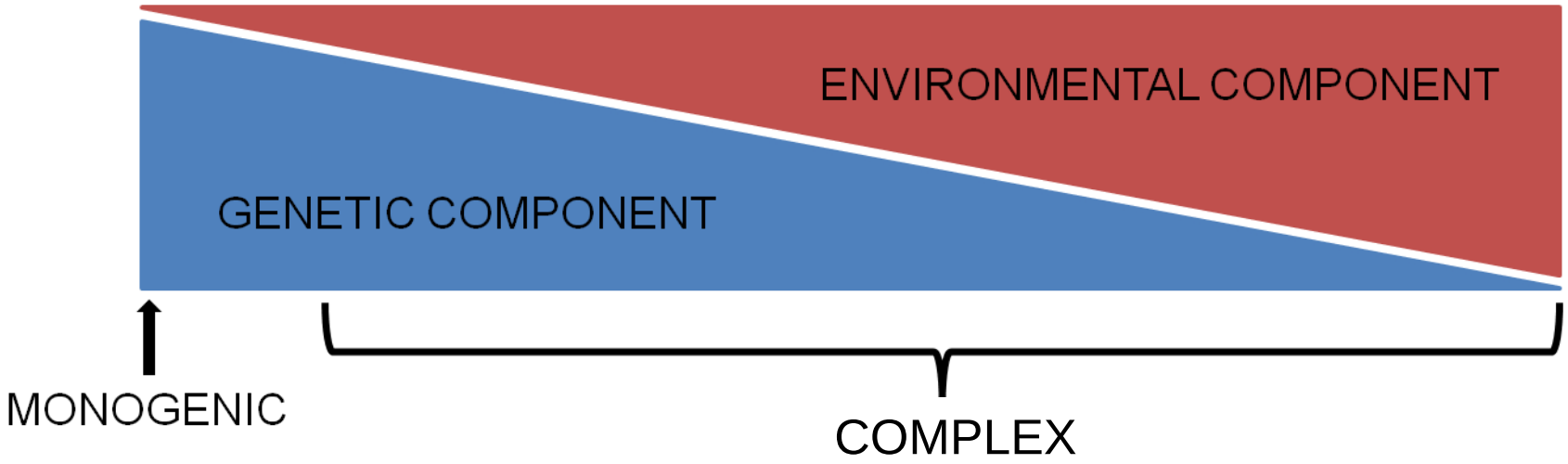
Complex Trait Genetic Associations

Monogenic

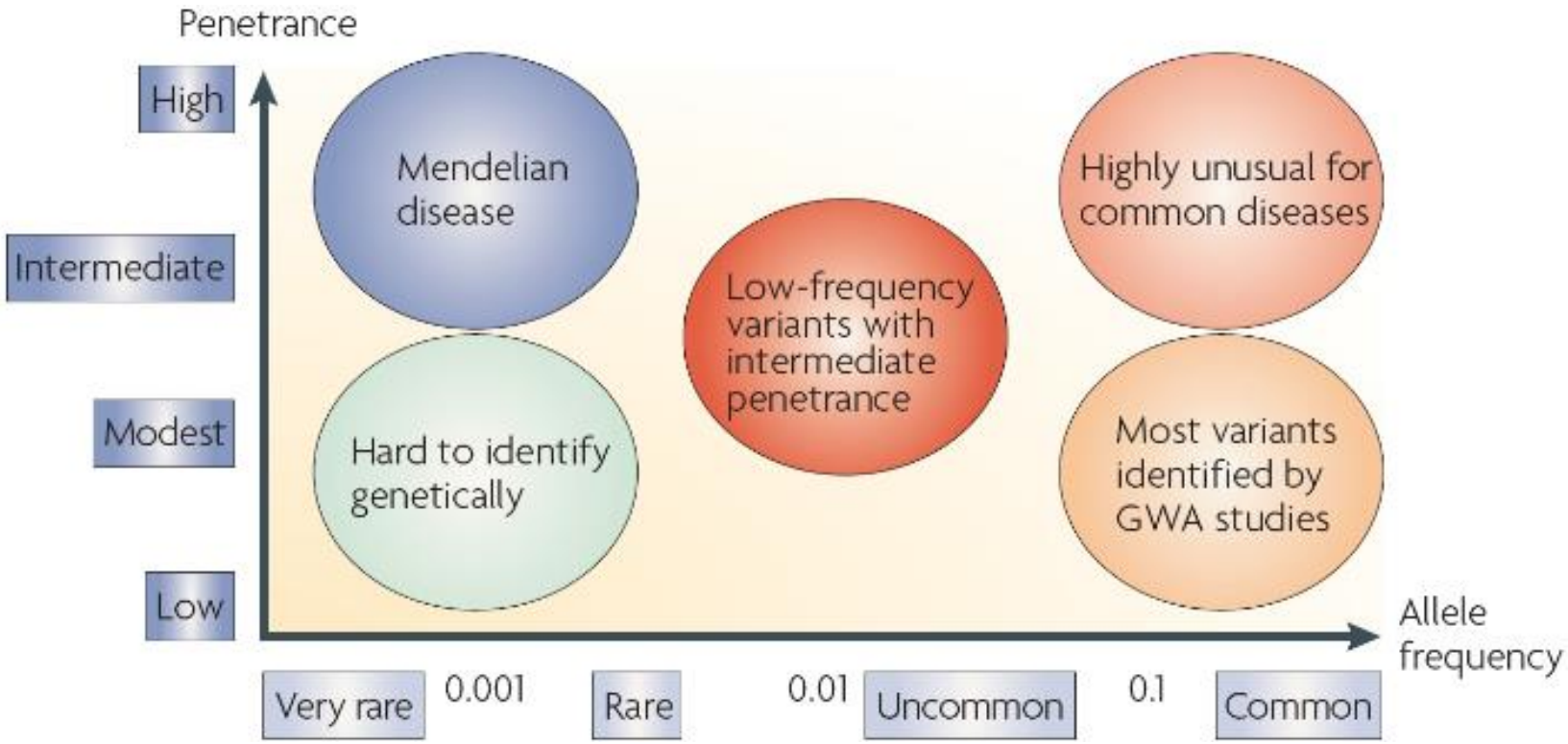
- Distinct model of inheritance
- Often rare
- Single gene - high penetrance
- Limited environmental influences

Complex

- No distinct mode of inheritance
- Often common
- Multiple loci – each with a small effect
- Important environmental factors



Complex Trait Genetic Associations



McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics* **9,** 367 (2009)

Atopic Dermatitis Heritability

- Risk of child developing eczema is higher if one or both parents have AD,OR=3.4 (Dold, 1992, AoDC)

Arch Dis Child. 1992 Aug; 67(8): 1018–1022. PMCID: PMC1793604

Genetic risk for asthma, allergic rhinitis, and atopic dermatitis.

S Dold, M Wjst, E von Mutius, P Reitmeir, and E Stiepel

- Twin studies suggest a heritability of >80% (Bataille, 2012, JEADV)

J Eur Acad Dermatol Venereol. 2012 Sep;26(9):1067-73. doi: 10.1111/j.1468-3083.2011.04444.x. Epub 2012 Jan 14.

The use of the twin model to investigate the genetics and epigenetics of skin diseases with genomic, transcriptomic and methylation data.

Bataille V[1], Lens M, Spector TD.

Finding associated variants

Case/Control Study for a single SNP:

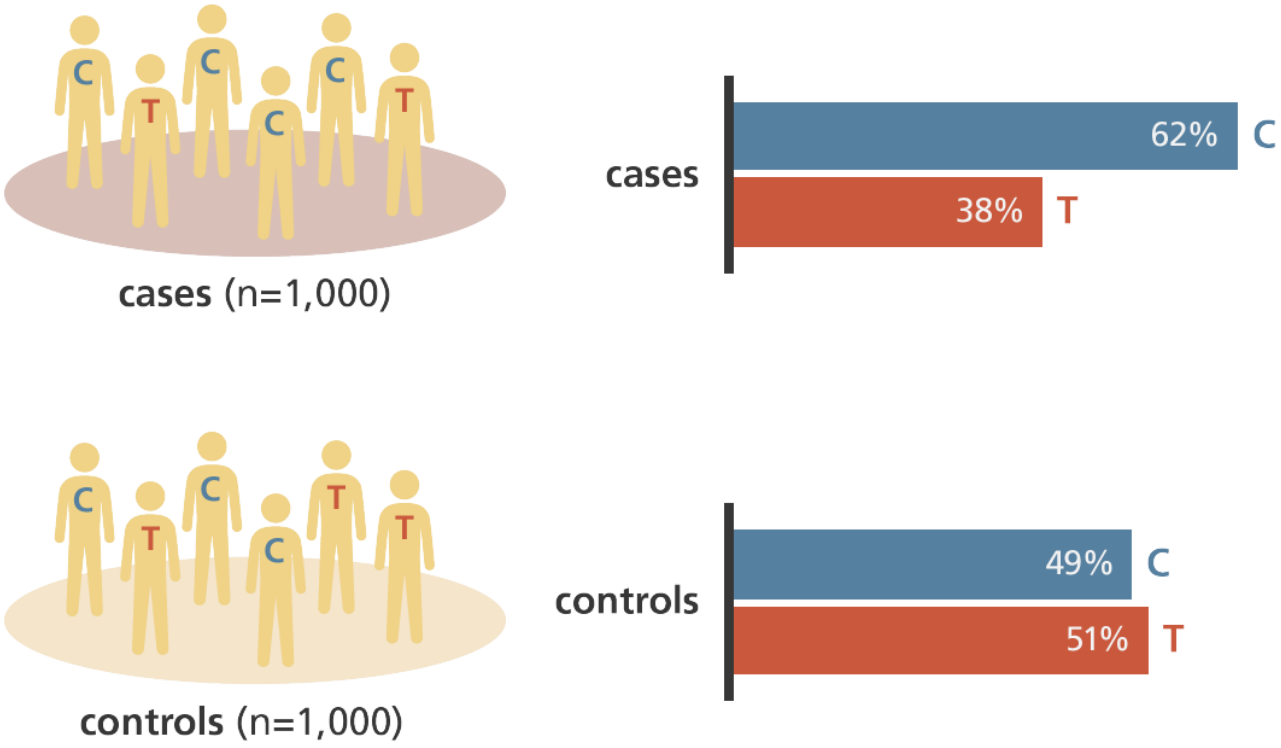


Image credit: Genome Research Limited
http://www.yourgenome.org/stories/genome-wide-association-studies

Statistical methods

- Logistic regression
  - Exposure = SNP – coded as 0, 1, 2 (i.e. # of alt. alleles)
  - Outcome = Eczema (case/control) – coded as 0,1 (binary)
  - Covariates = Age, Sex, Ancestry – various coding

$$Disease = \beta_0 + \beta_1 SNP + \beta_2 sex + \beta_3 \text{age} + \beta_4 ancestryPC1 \ldots\ldots$$

- Extend to Genome-wide ---->>>> Just repeat this simple statistical test for each SNP (>1 million times)

- Get 1 result per SNP tested

GWAS software – Plink (cog-genomics.org/plink/2.0)

Input:

ID	Ecz	sex	age	ancestryPC1	...
001	0	0	6.3	0.987645	...
002	0	1	7.5	0.874624	…
003	1	0	3.2	0.967283	…
004	1	1	0.8	0.123645	…

ID	SNP1	SNP2	SNP3	SNP4	SNP…
001	0	0	2	1.1	…
002	2	1	2	1.9	…
003	0	1	2	1.5	…
004	1	1	2	0.1	…

Output:

SNP	Effect Allele	Effect allele freq	OR	Lower CI	Upper CI	P-value
SNP1	A	0.54	1.00	0.90	1.10	0.998
SNP2	G	0.2	0.95	0.89	1.01	0.089
SNP3	T	0.9	1.12	1.10	1.14	0.001

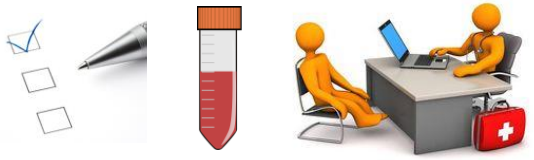
Study Design Issues

- **Genetic data – genotyping array and imputation**
  - Use genome-wide genotyping chip
  - Every study imputes to same genetic reference

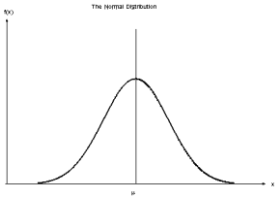
- **Confounders – only confounder is ancestry**
  - Calculate principal components of ancestry using genetic data (and software such as Eigenstrat or Plink)
  - Adjust for these components in the analysis
- **Multiple testing**
  - Conducting >1 million tests $p<0.05$ is not sufficient evidence
  - GWAS significance is $p<5x10^{-8}$
- **Sample size**
  - To achieve required p-value, large sample sizes are needed
  - Usual to conduct meta-analysis of several cohorts

Conducting a GWAS study

Sample recruitment



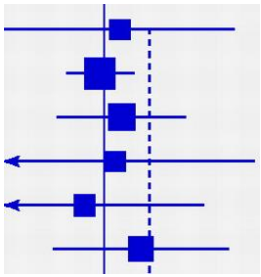
Data collection

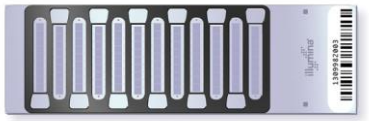


Phenotype data preparation

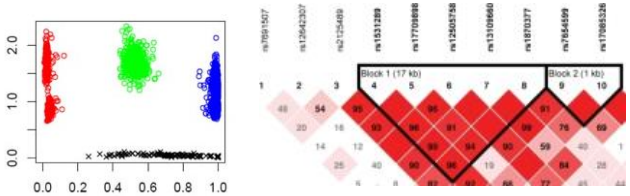
GWAS

Meta-analysis

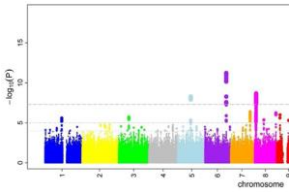




Genotyping



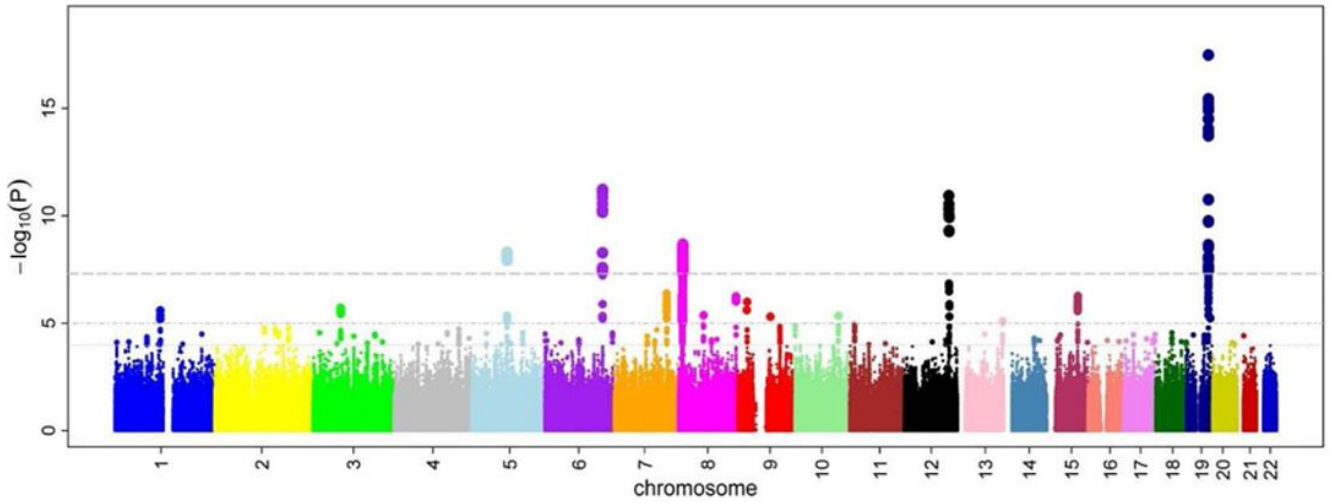
Genotype quality control & imputation



Interpret

GWAS results summary

Manhattan plot

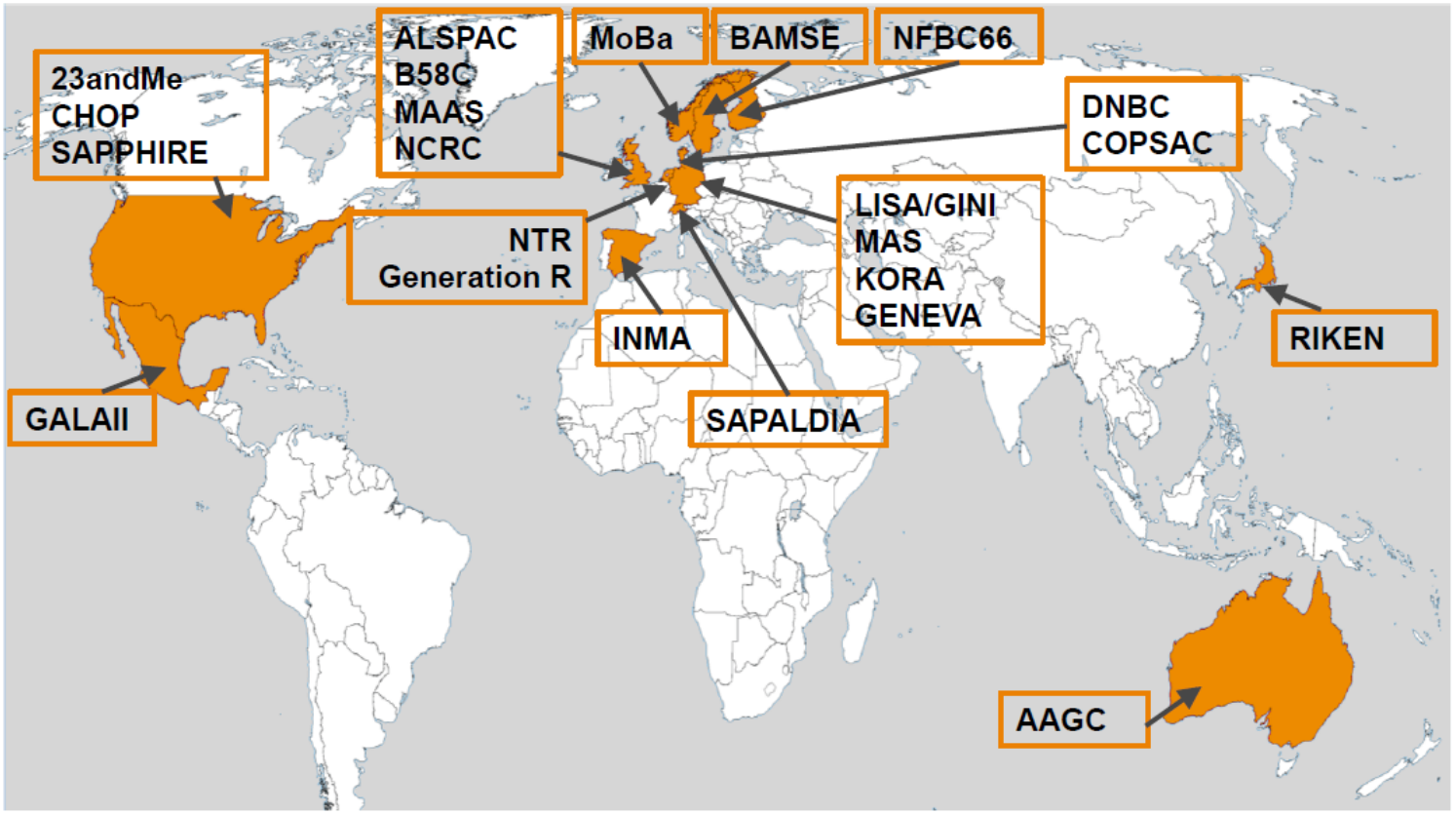


Results table

SNP	Effect Allele	Effect allele freq	OR	Lower CI	Upper CI	P-value
SNP1	A	0.54	1.00	0.90	1.10	0.998
SNP2	G	0.2	0.95	0.89	1.01	0.089
SNP3	T	0.9	1.12	1.10	1.14	0.001

What have Atopic Dermatitis GWAS found?

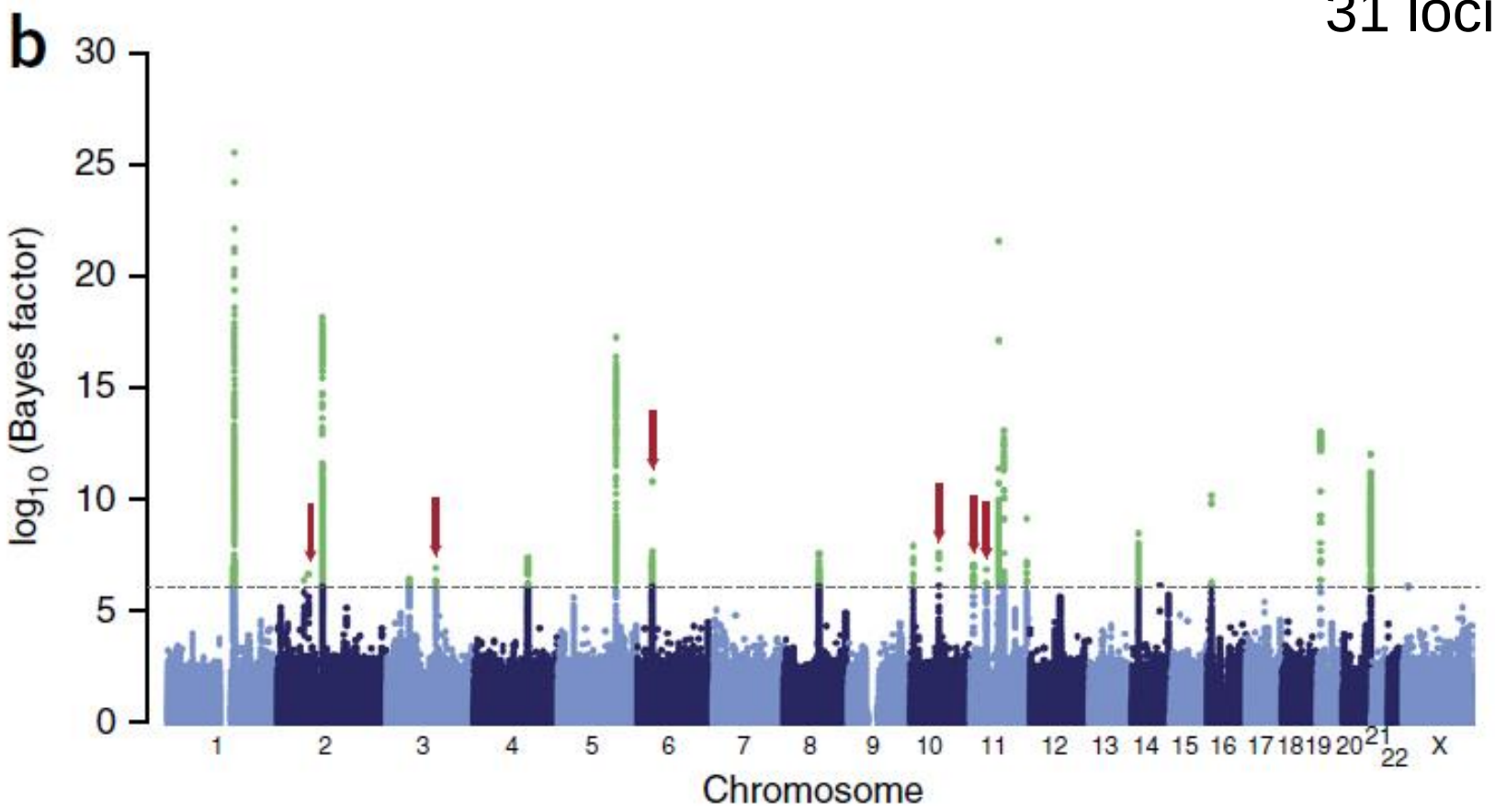
EAGLE GWAS 2015: 21,399 cases, 95,464 controls
all imputed to 1000 Genomes



Paternoster et al. 2015, Nature Genetics 47:1449-56

What have Atopic Dermatitis GWAS found?

31 loci



Paternoster et al. 2015, Nature Genetics 47:1449-56

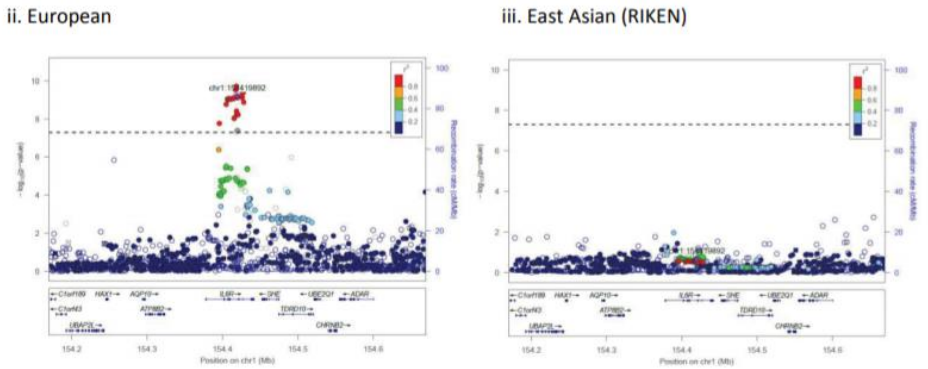
What have Atopic Dermatitis GWAS found?

Variant	Locus	Nearest gene[a]	EA/OA	European, fixed effects				All cohorts, MANTRA		Known atopy loci?	
				n (studies)	EAF	OR (95% CI)	*P*	*n* (studies)	$\log_{10}$ (BF)	Trait	Reference
Known loci											
rs61813875	1q21.3	*CRCT1*/***LCE3E*** (*FLG*)[b]	G/C	93,326 (18)	0.02	1.61 (1.48–1.75)	**5.6×10^{-29}**	96,419 (20)	**25.53**	AD	3,4,5
rs10791824	11q13.1	*OVOL1*	G/A	102,761 (21)	0.57	1.12 (1.09–1.15)	**2.1×10^{-19}**	116,556 (25)	**21.56**	AD	9
rs12188917	5q31.1	*RAD50*/***IL13***	C/T	102,761 (21)	0.21	1.14 (1.10–1.17)	**4.0×10^{-17}**	116,554 (25)	**17.24**	AD, A, IgE	9,18,58
rs6419573	2q12.1	*IL18R1*/***IL18RAP***	T/C	102,760 (21)	0.26	1.11 (1.08–1.14)	**1.5×10^{-13}**	116,557 (25)	**18.10**	AD, A, AS, SRA	8,14,18,21
rs2212434	11q13.5	***C11orf30***/*LRRC32*	T/C	102,761 (21)	0.45	1.09 (1.07–1.12)	**4.6×10^{-13}**	116,557 (25)	**13.02**	AD, AS, SRA, AR, A	11,14,15,21,59
rs4809219	20q13.33	***RTEL1/TNFRSF6B***	C/A	102,760 (21)	0.27	0.90 (0.87–0.93)	**7.0×10^{-13}**	116,555 (25)	**11.98**	AD	7,10
rs2918307	19p13.2	*ADAMTS10*/***ACTL9***	G/A	100,707 (20)	0.16	1.12 (1.08–1.16)	**4.6×10^{-12}**	114,504 (24)	**12.98**	AD	9
rs2041733	16p13.13	*CLEC16A*	C/T	103,066 (22)	0.55	0.92 (0.90–0.94)	**2.5×10^{-11}**	116,862 (26)	**10.11**	AD, A+HF	7,53
rs12730935[c]	1q21.3	*IL6R*	A/G	102,760 (21)	0.39	1.08 (1.05–1.11)	**6.1×10^{-11}**	116,556 (25)	**7.15**	AD, A	12,15
4:123243592[d]	4q27	*KIAA1109* (*IL2*)[b]	R/I	102,761 (21)	0.37	1.08 (1.05–1.10)	**4.2×10^{-9}**	107,119 (24)	**7.32**	AD, AS, SRA	7,14,21
rs4713555	6p21.32	***HLA-DRB1***/*HLA-DQA1*	T/G	91,217 (15)	0.27	0.91 (0.89–0.94)	**5.4×10^{-9}**	105,014 (19)	**10.76**	AD, AS, SRA, A	6,8,14,18,21

Paternoster et al. 2015, Nature Genetics 47:1449-56

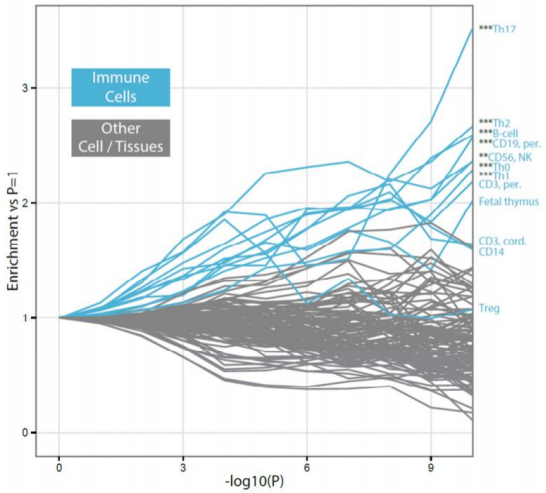
Some interesting observations

- Some loci aren't associated in all ethnicities tested



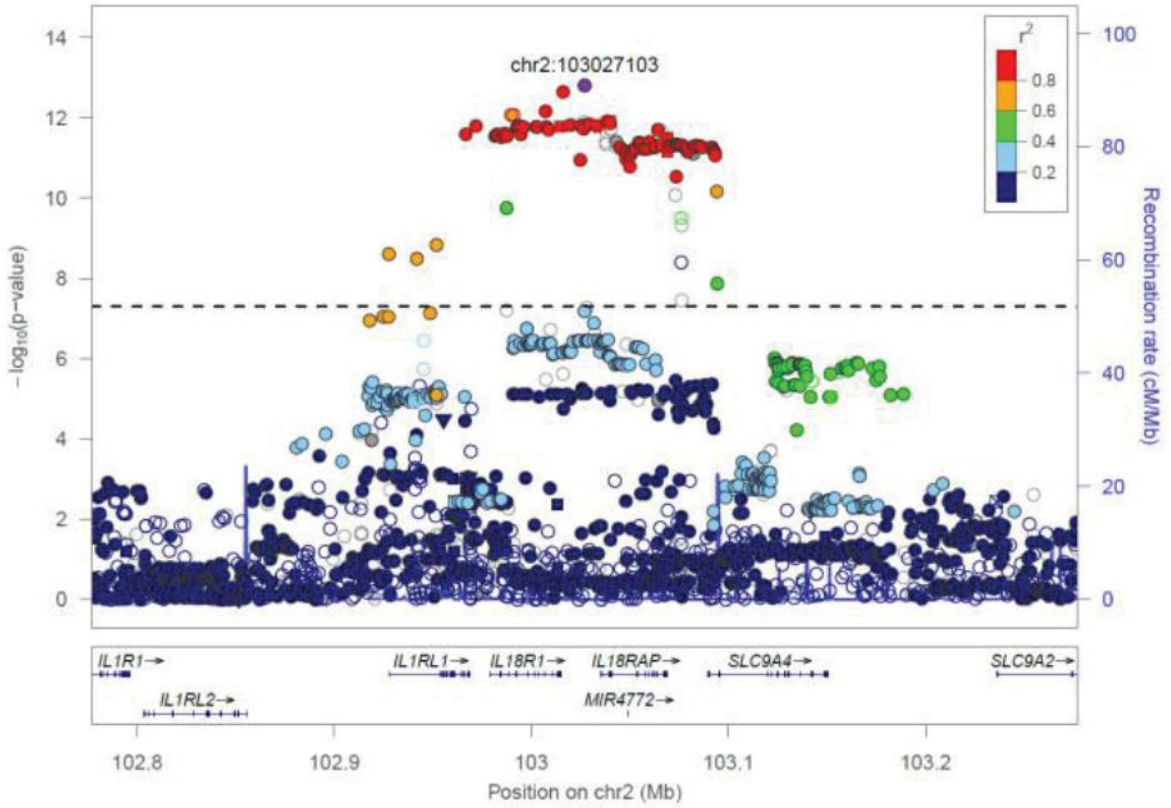
- AD genetic loci overlap with autoimmune disease, such as inflammatory bowel disease

- GWAS loci are enriched for DNAse hypersensitivity in immune celltypes, especially T_H17



Next steps

GWAS identify loci not causal genes…….



In tomorrow's Bioinformatics session we'll discuss how to identify genes!

We can use the genetic data to cluster individuals.. using PCA

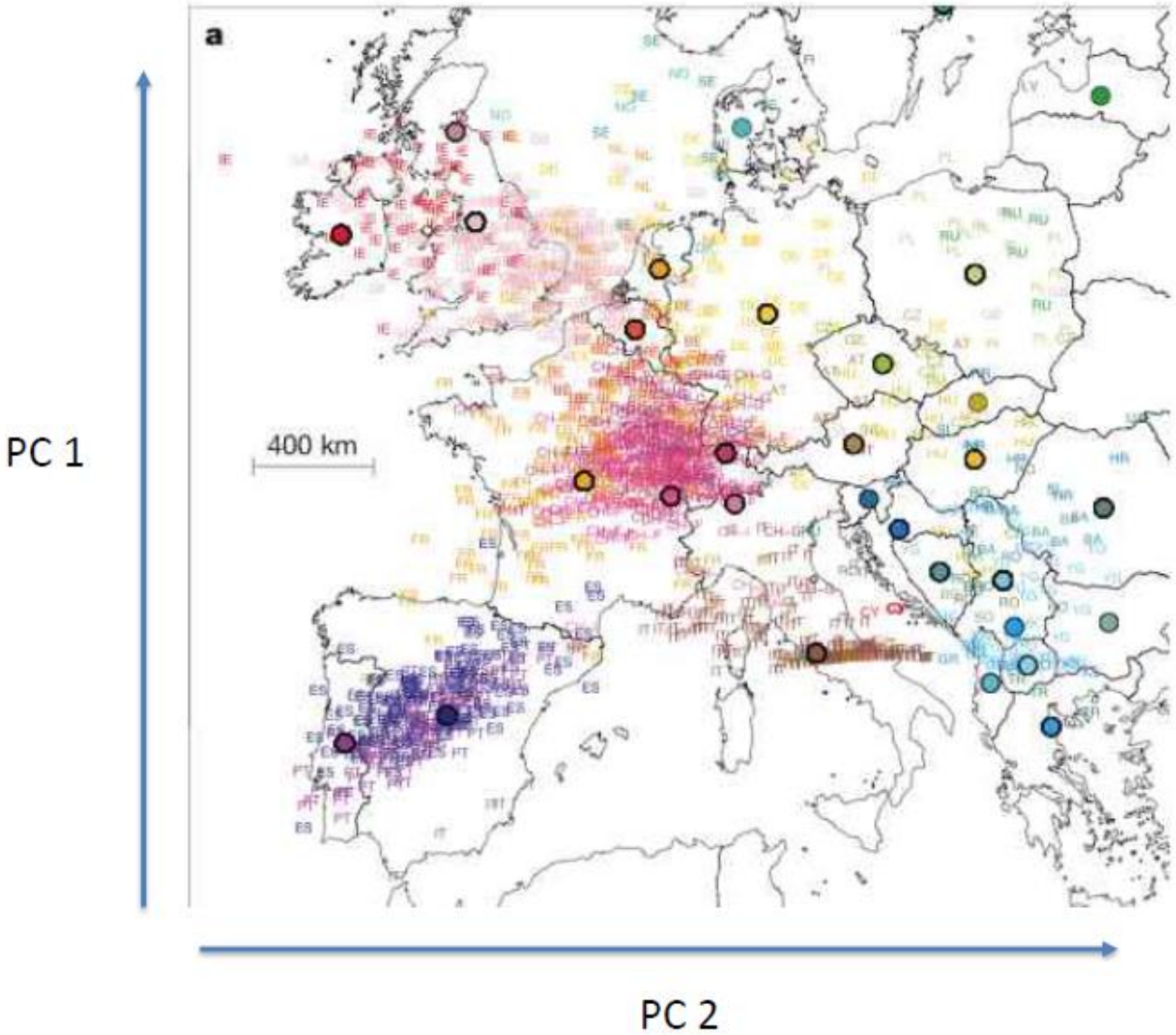
Principal components analysis (PCA) is a method for summarising high dimensional data

If you apply PCA to our genetic data it will take the 500,000 columns and turn them into far fewer columns

Those columns represent the main axes of variation that the SNPs share in common

What do these axes represent in real life…?

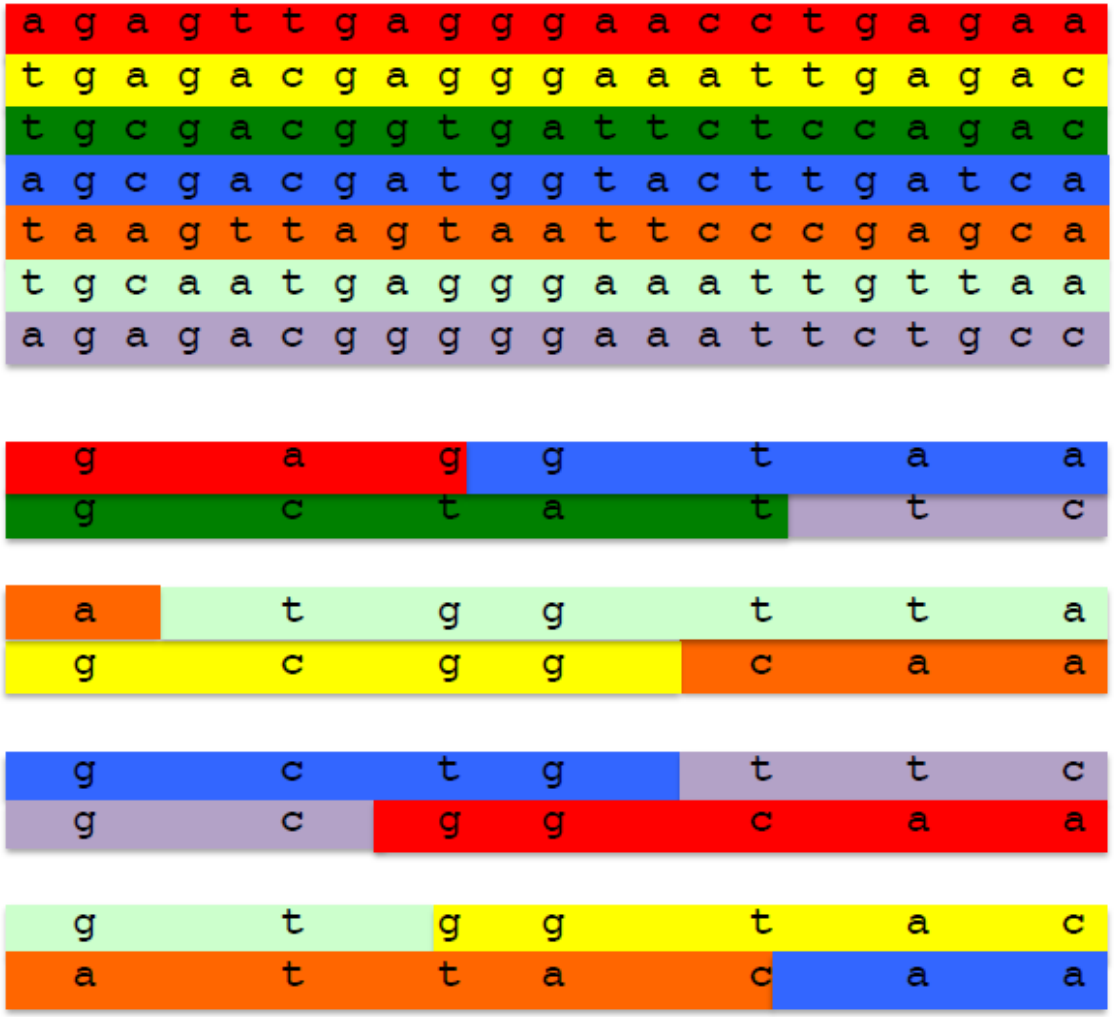
PCs across Europe



```
a g a g t t g a g g g a a c c t g a g a a
t g a g a c g a g g g a a a t t g a g a c
t g c g a c g g t g a t t c t c c a g a c
a g c g a c g a t g g t a c t t g a t c a
t a a g t t a g t a a t t c c c g a g c a
t g c a a t g a g g g a a a t t g t t a a
a g a g a c g g g g g a a a t t c t g c c

  g       a     g   g       t     a     a
  g       c     t   a       t     t     c

  a       t     g   g       t     t     a
  g       c     g   g       c     a     a

  g       c     t   g       t     t     c
  g       c     g   g       c     a     a

  g       t     g   g       t     a     c
  a       t     t   a       c     a     a
```

Reference haplotypes via sequencing studies

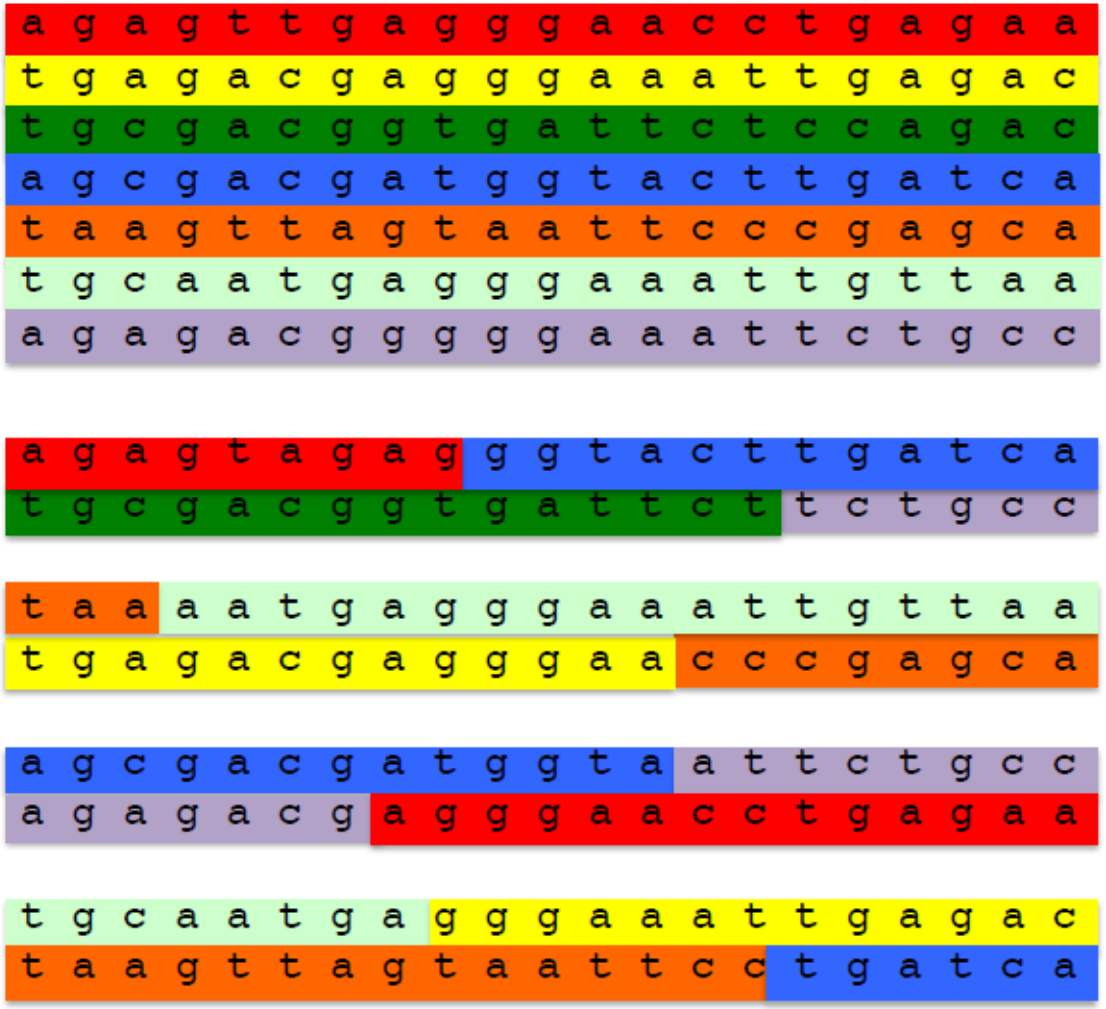
eg. 1000 Genomes Project



Reference haplotypes via sequencing studies

eg. 1000 Genomes Project

Imputation of unobserved alleles via matching of shared haplotypes



Reference haplotypes via sequencing studies

eg. 1000 Genomes Project

Imputation of unobserved alleles via matching of shared haplotypes