Haplotype Blocks:

or how I learned to stop worrying and love the recombination hotspot

Benjamin Neale,

David Evans, Pak Sham

Boulder, Colorado

March 2005

http://webpages.charter.net/harshec/lego/images/simpsons/milhouse_0.jpg

Where we are going

- Multilocus mapping
- Haplotype blocks/Linkage Disequilibrium regions
  - Definitions
  - Uses
- Current data
  - HapMap
  - Other efforts
- Quick word on clades and cladistics

Multilocus Mapping

- Searching for the variant on a fine scale
- Linkage disequilibrium (LD) means redundant information
- May not parse causal variant, but through LD inferred information
- Potential epistatic effects

Linkage Disequilibrium

- Non-random assortment of alleles
- Typically occurs over kbs
- Measures based 2 loci sysem A/a & B/b:

	A	a	Total
B	p_{AB}	p_{aB}	p_B
b	p_{Ab}	p_{ab}	p_b
Total	p_A	p_a	1

Linkage Disequilibrium

- $D = p_{AB} - p_A * p_B$
- More preferable is $D' = D/D_{max}$
  - Where D_{max} is $\min(p_A p_b, p_a p_B)$ if D is positive or $\min(p_A p_B, p_a p_b)$ when D is negative

	A	a	Total
B	p_{AB}	p_{aB}	p_B
b	p_{Ab}	p_{ab}	p_b
Total	p_A	p_a	1

Linkage Disequilibrium

- $D = p_{AB} - p_A * p_B$
- $r^2 = D^2 / p_A p_a p_B p_b$
- which is the correlation coefficient between alleles A and B

	A	a	Total
B	p_{AB}	p_{aB}	p_B
b	p_{Ab}	p_{ab}	p_b
Total	p_A	p_a	1

Linkage Disequilibrium

- From $r^2=D^2/p_Ap_ap_Bp_b$
- We can test r^2 is significantly different from 0 using likelihood.
- In Haploview this is referred to as the LOD

	A	a	Total
B	p_{AB}	p_{aB}	p_B
b	p_{Ab}	p_{ab}	p_b
Total	p_A	p_a	1

What do LD regions do?

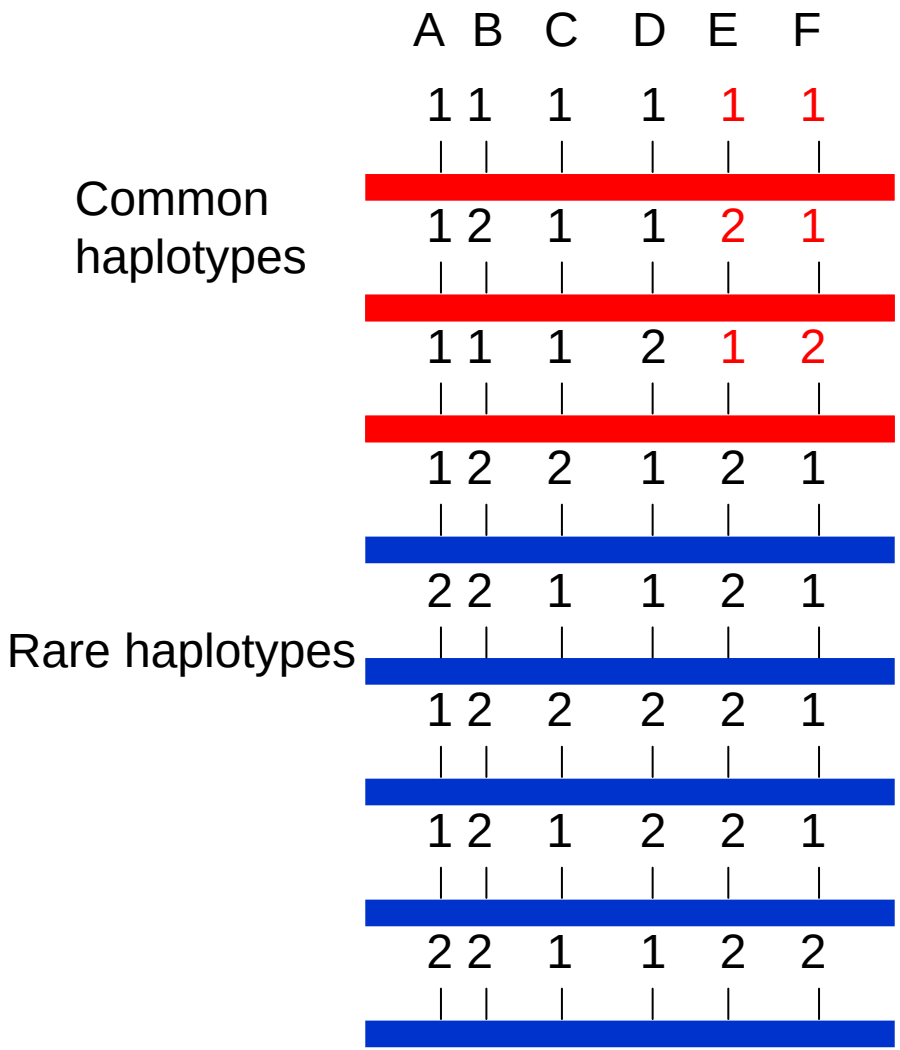
- Generate “haplotype tags” (htSNPs)
  - Tag common haplotypes
- Generate “tagging SNPs” (tSNPs)
  - Tag all variation above minor allele frequency threshold
- Parse “hidden SNPs”
  - Marginal information on untyped variants

Haplotype Tagging

1	1	2	1	2	1	1
1	2	1	2	2	1	2
2	2	2	2	2	2	2
2	1	1	1	1	1	2
2	2	1	1	2	2	1

Visualization of blocks vs. tags

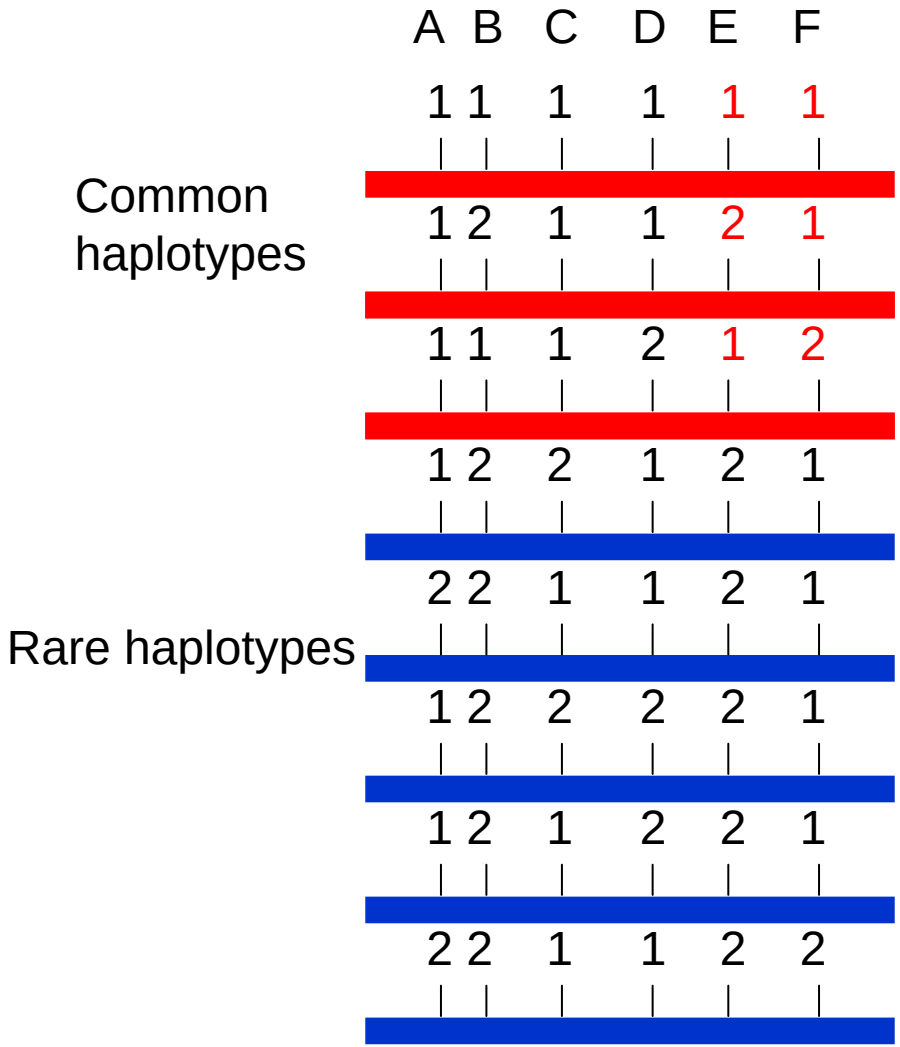
Haplotype Block methods



Tag methods

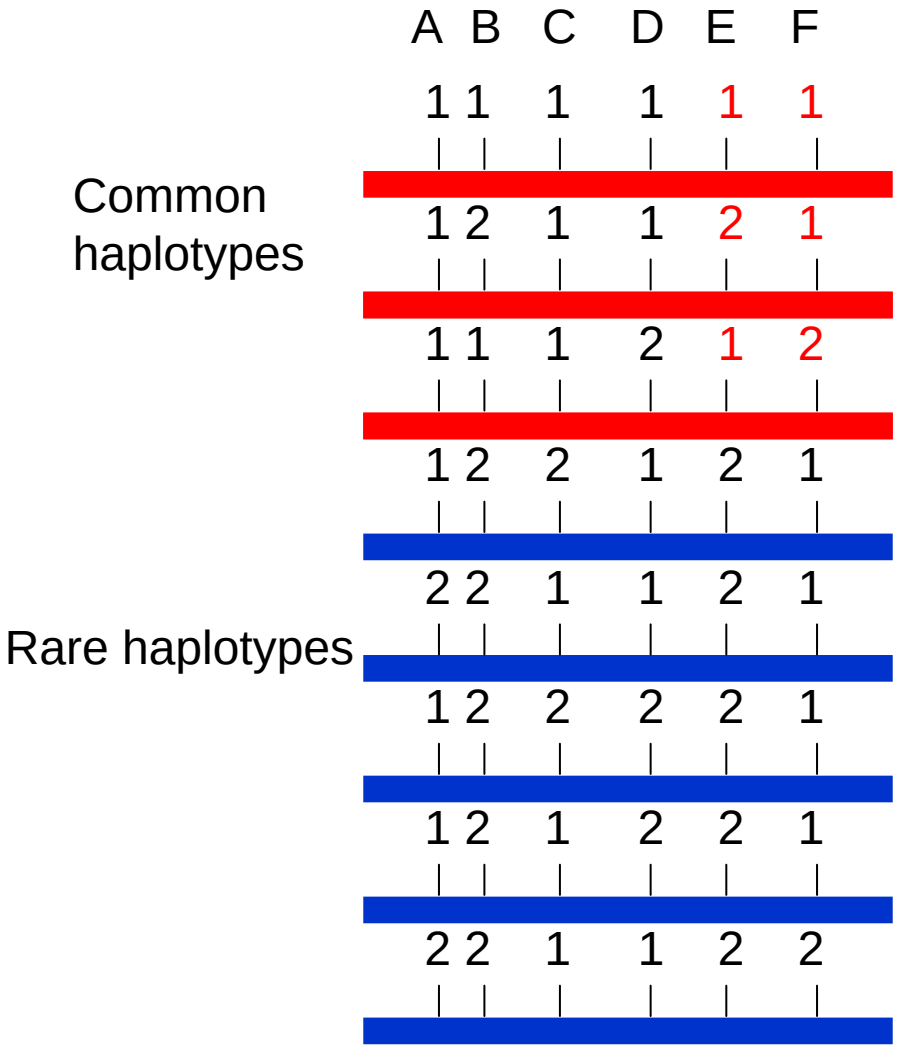
	A	B	C	D	E	F
A	1					
B	.9	1				
C	.5	.8	1			
D	.4	.6	.9	1		
E	.9	1	.8	.6	1	
F	.4	.4	.3	.4	.5	1

Haplotype Block Definitions (diversity, htSNPs)



- Patil et al. 2001: minimum SNP coverage to account for a majority of common haplotypes
- Daly et al. 2001: SNP coverage for lower haplotypic diversity

Pair-wise LD based block (htSNPs)



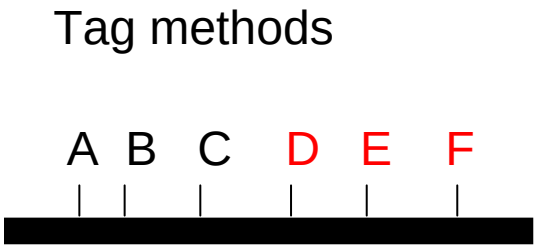
- Gabriel et al. 2002
  - Small proportion of marker pairs show evidence for historical recombination
  - Blocks are partitioned according to whether the upper and lower confidence limits on estimates of pairwise *D´* measure fall within certain threshold values
  - E.G. 80% of all pair-wise LD scores >0.7

Recombination based block (htSNPs)

- Wang et al. 2002
  - Four gamete test
  - Blocks only where there is no evidence of recombination
  - Out of following pairs only 3 are observed:
    - 11
    - 12
    - 21
    - 22

Prediction based tagging (tSNPS)

- Prediction at a certain pre-defined R^2
- Stram et al. 2003
  - Prediction of haplotypes
- Weale et al 2003
  - Prediction of all SNPs



	A	B	C	D	E	F
A	1					
B	.9	1				
C	.5	.8	1			
D	.4	.6	.9	1		
E	.9	1	.8	.6	1	
F	.4	.4	.3	.4	.5	1

General LD map questions

- How well do tag SNPs inform ‘hidden SNPs’
- How does allele frequency affect results
- How does marker density affect results
- How well do tag SNPs perform in the same population as sampled
- How well do tag SNPs perform in different populations

How well do all the prior methods do?

- No one knows
- Lots of method and not a huge amount of clear data
- Still a bit questionable about what the implications of haplotype tests are

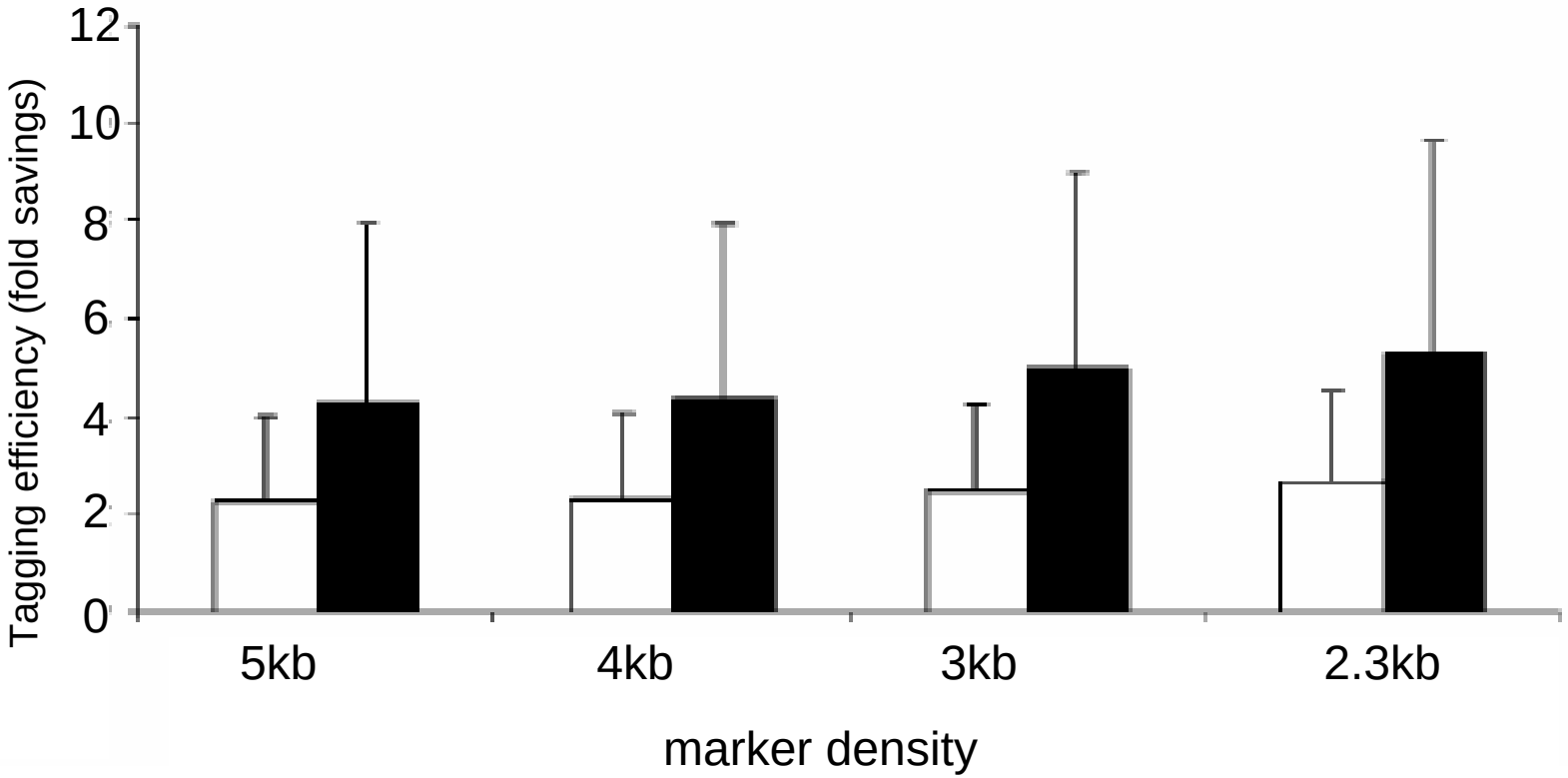
Data—Ke et al.

- SNP per 2.3 kb for 10 Mb of chromosome 20
- 96 UK Caucasians, 48 CEPH founders, and 97 African Americans
- Wellcome Trust in Oxford and Sanger Centre

Results from Ke

- ~3 fold savings from LD in European descent
- ~2 fold savings from LD in African descent
- $r^2 > .85$ with ‘hidden SNP’ with freq > 20%
- As MAF of hidden SNP decreases as compared to the tag SNP r^2 decreases

Savings from different marker densities from Ke et al.

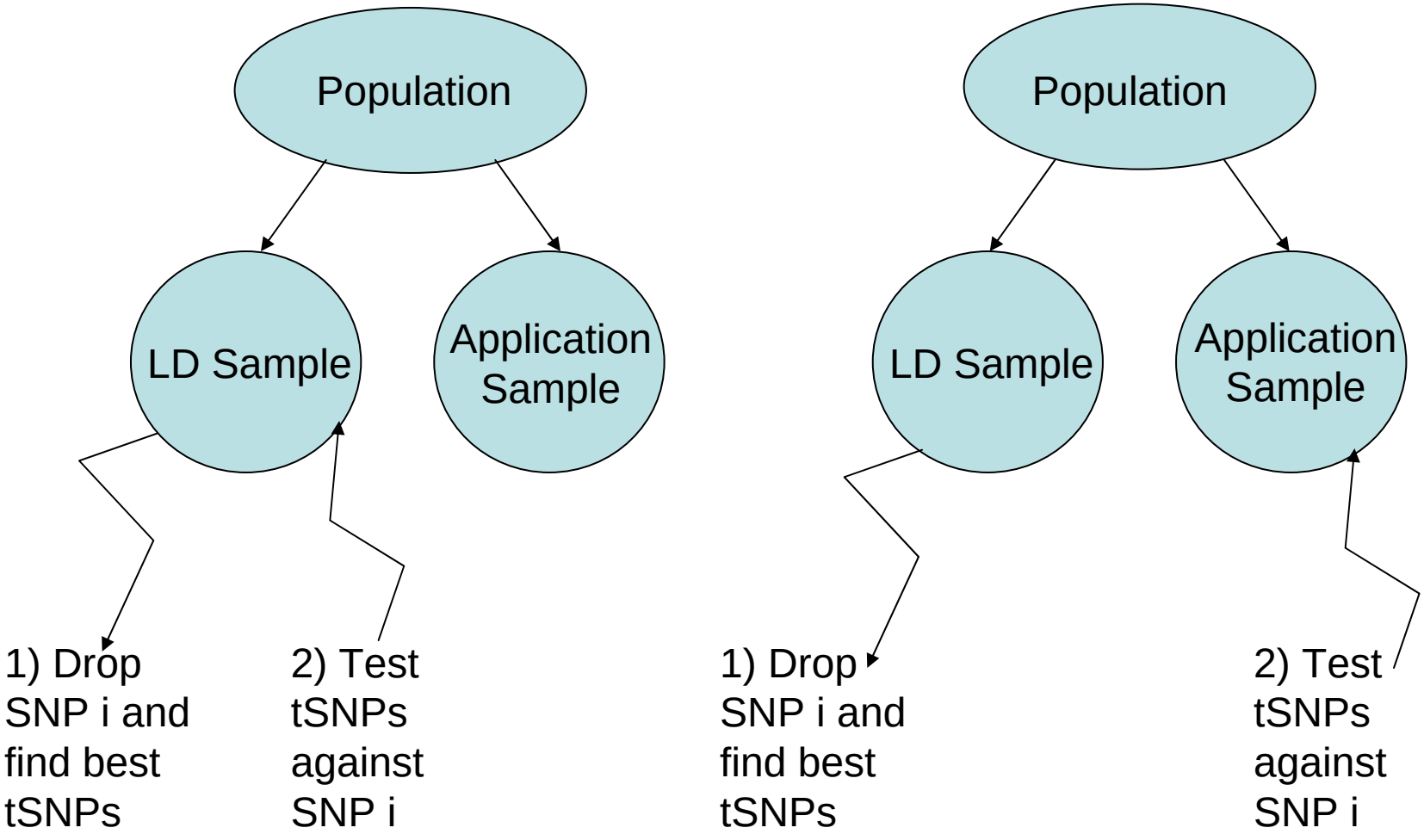


Dark bars: 100% hap diversity, Light bars: 80% hap diversity

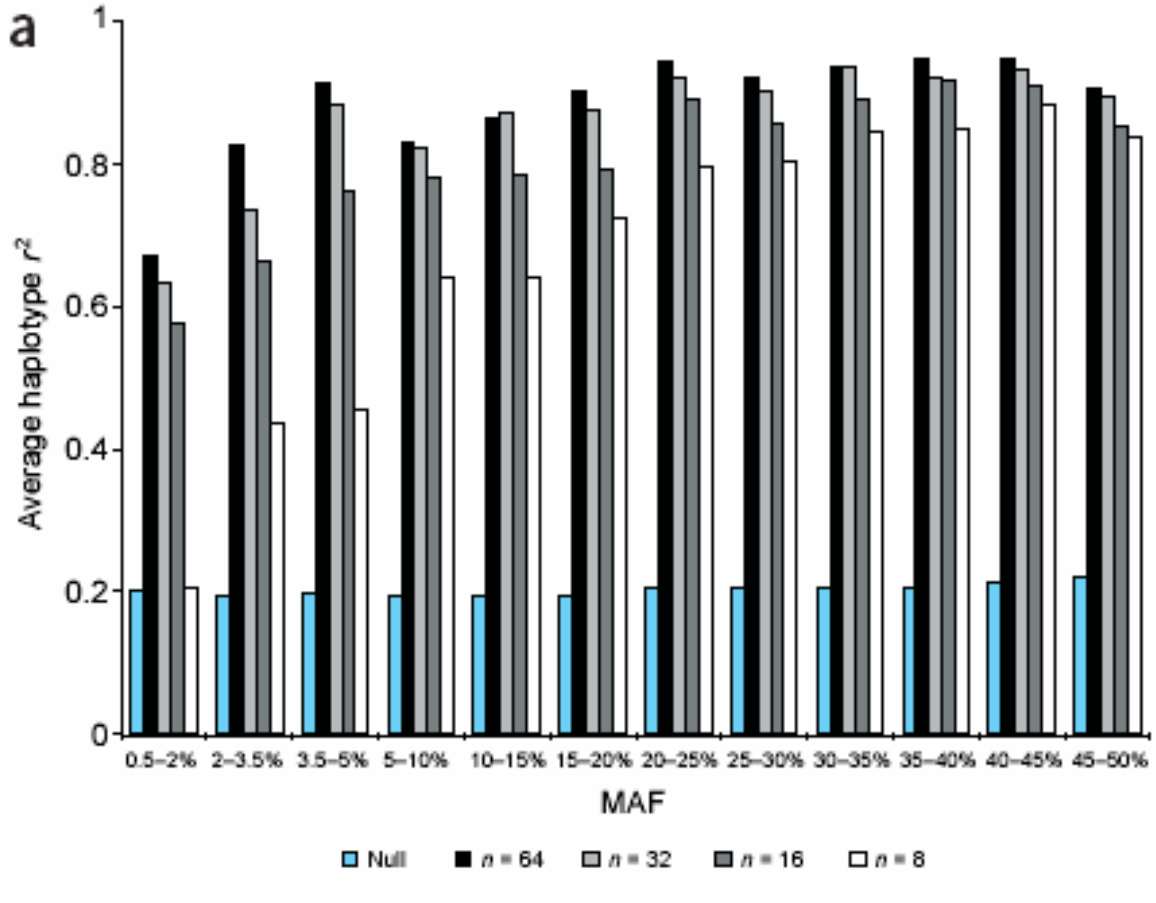
Ahmadi et al. sample

- 55 genes: 2,123 kb with 1 SNP/3.5 kb
- 2 samples: Caucasian (CEPH) and Japanese—64 individuals
- Haplotype r^2 approach
- UCL in conjunction with GSK

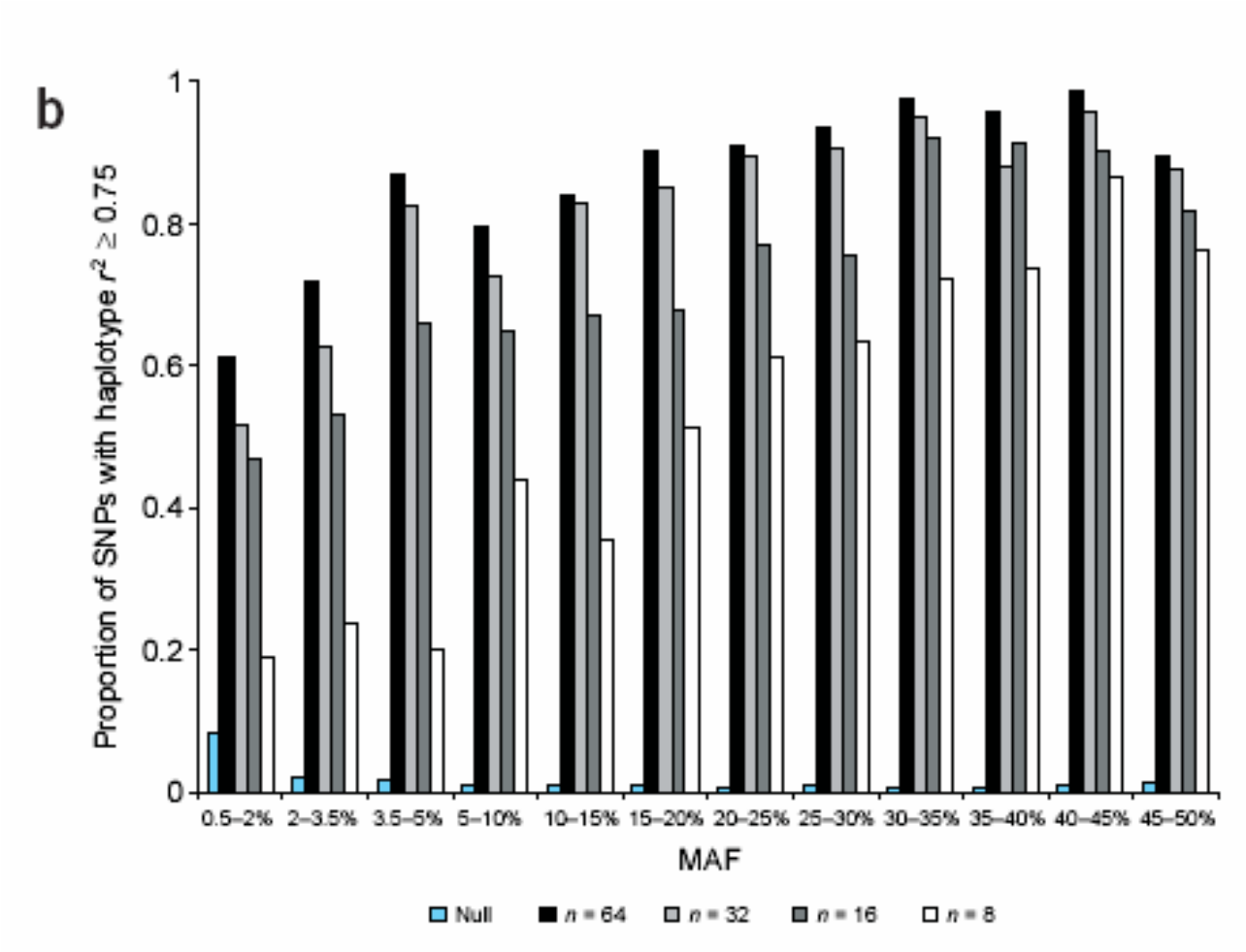
Ahmadi et al. data



Ahmadi et al.


a
Average haplotype r^2
1
0.8
0.6
0.4
0.2
0
0.5–2%
2–3.5%
3.5–5%
5–10%
10–15%
15–20%
20–25%
25–30%
30–35%
35–40%
40–45%
45–50%
MAF
Null
n = 64
n = 32
n = 16
n = 8

Ahmadi et al.


b
Proportion of SNPs with haplotype $r^2 \geq 0.75$
1
0.8
0.6
0.4
0.2
0
0.5–2%
2–3.5%
3.5–5%
5–10%
10–15%
15–20%
20–25%
25–30%
30–35%
35–40%
40–45%
45–50%
MAF
Null
n = 64
n = 32
n = 16
n = 8

Ahmadi et al. conclusions

- Echo much of Ke et al.
- Marker density improves detection, but increases SNP number
- Lower MAF, especially lower than tSNPs costs effectiveness
- Argues a global map will work (much crossover between European and Japanese populations), though questionable conclusion

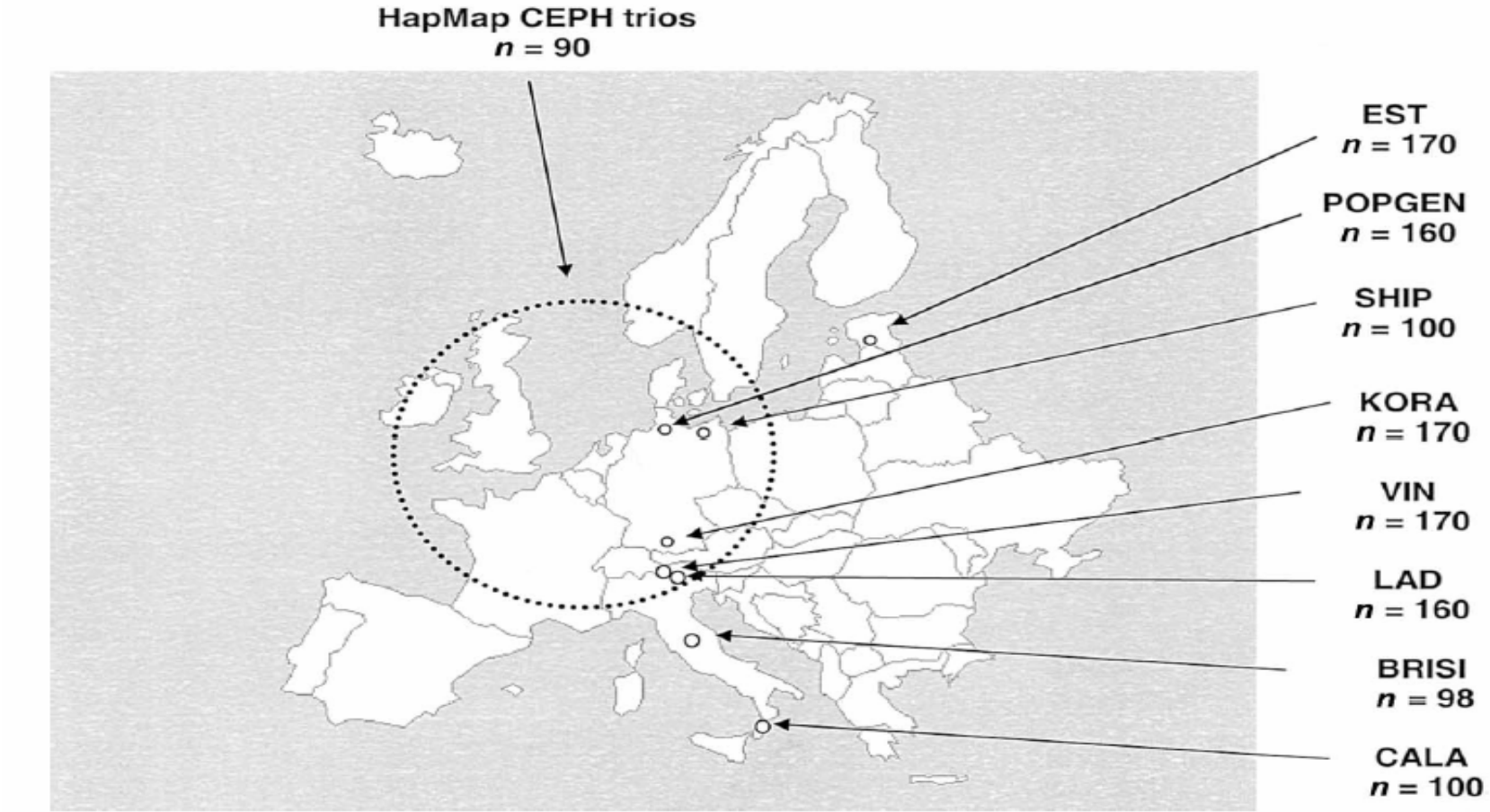
Block Boundaries

- Boundaries are hypothesized to be recombination hotspots
- Actual boundary is probably fuzzy because:
  - Demographic history
  - Differences in Recombination hotspots

Data from Mueller et al.

- CEPH families, Estonians, 2 North German, South German, 2 Alpine, Central Italian, and Southern Italian
- Groups working together across Europe

Real example of fine-mapping

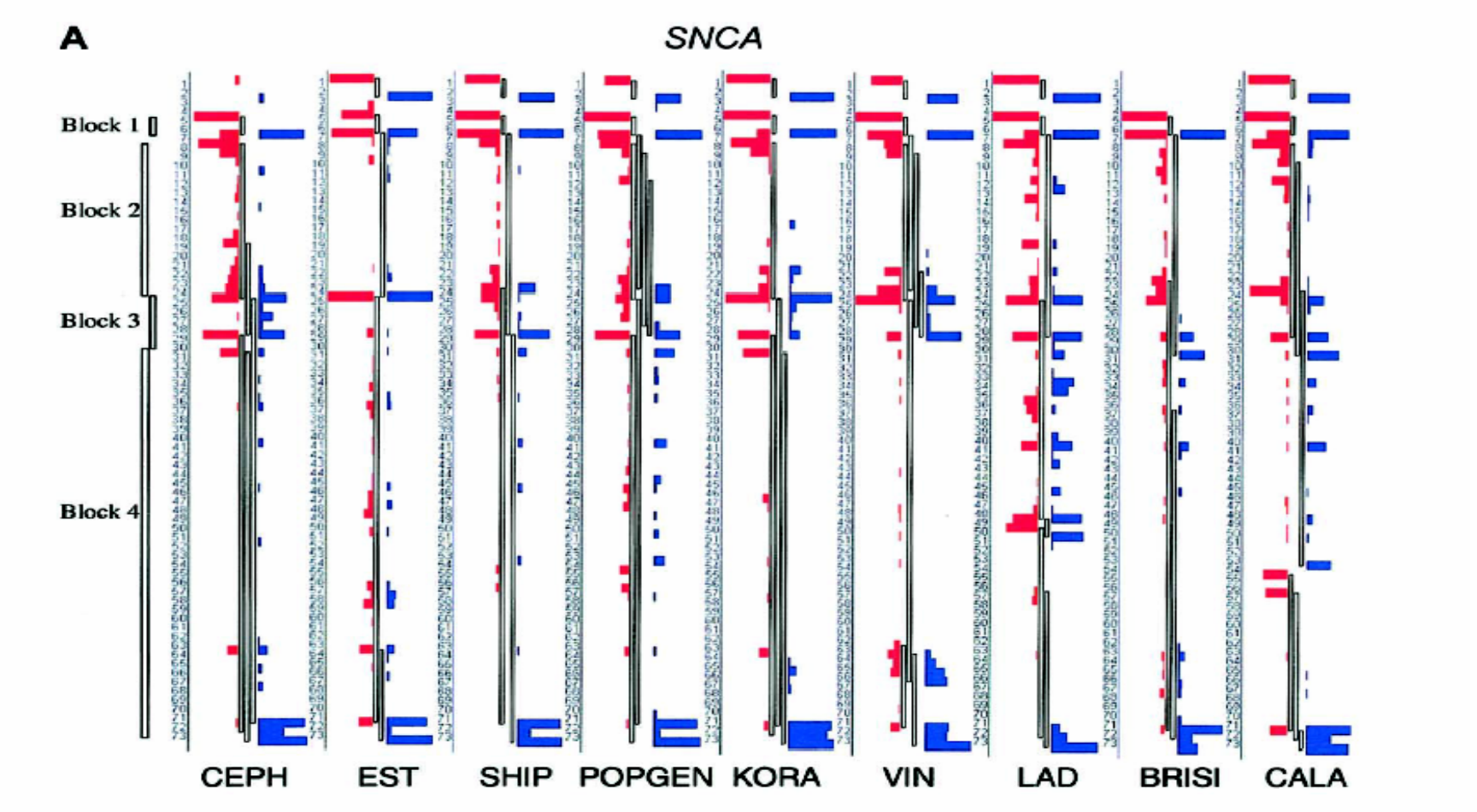


Mueller et al. AJHG 2005 Mar;76(3):387-98.

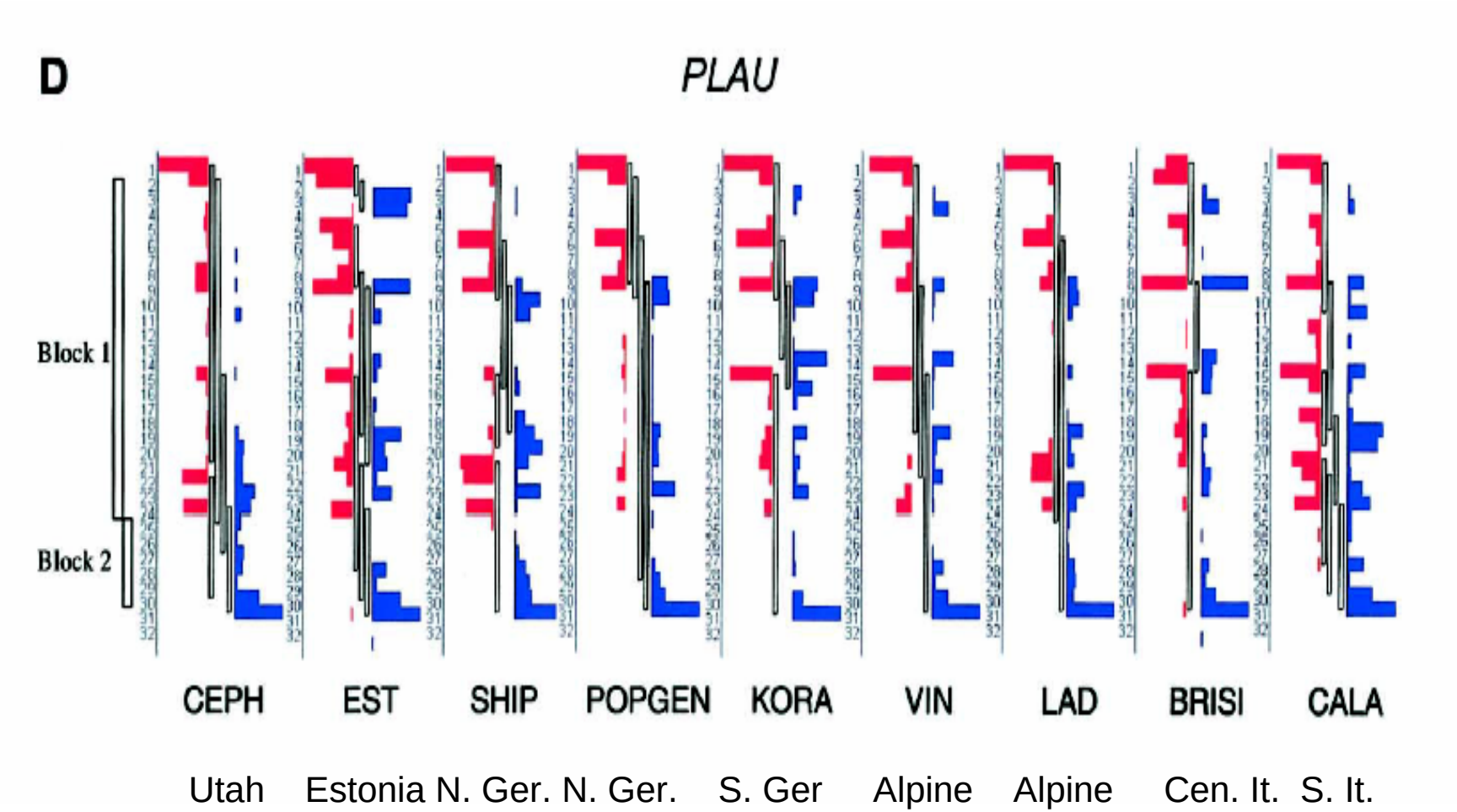
Details of mapping

- Cover gene and 76-174 kb up and downstream
- Dense mapping—SNP per 2-4 kb
- 1218 total individuals

Block Boundaries in SNCA



Block Boundaries in PLAU



High LD regions

- Use public data to define blocks and tag SNPs—HapMap
- Generate from own data
  - Sample size
  - Measure of LD
  - Ethnic population
  - Ascertainment

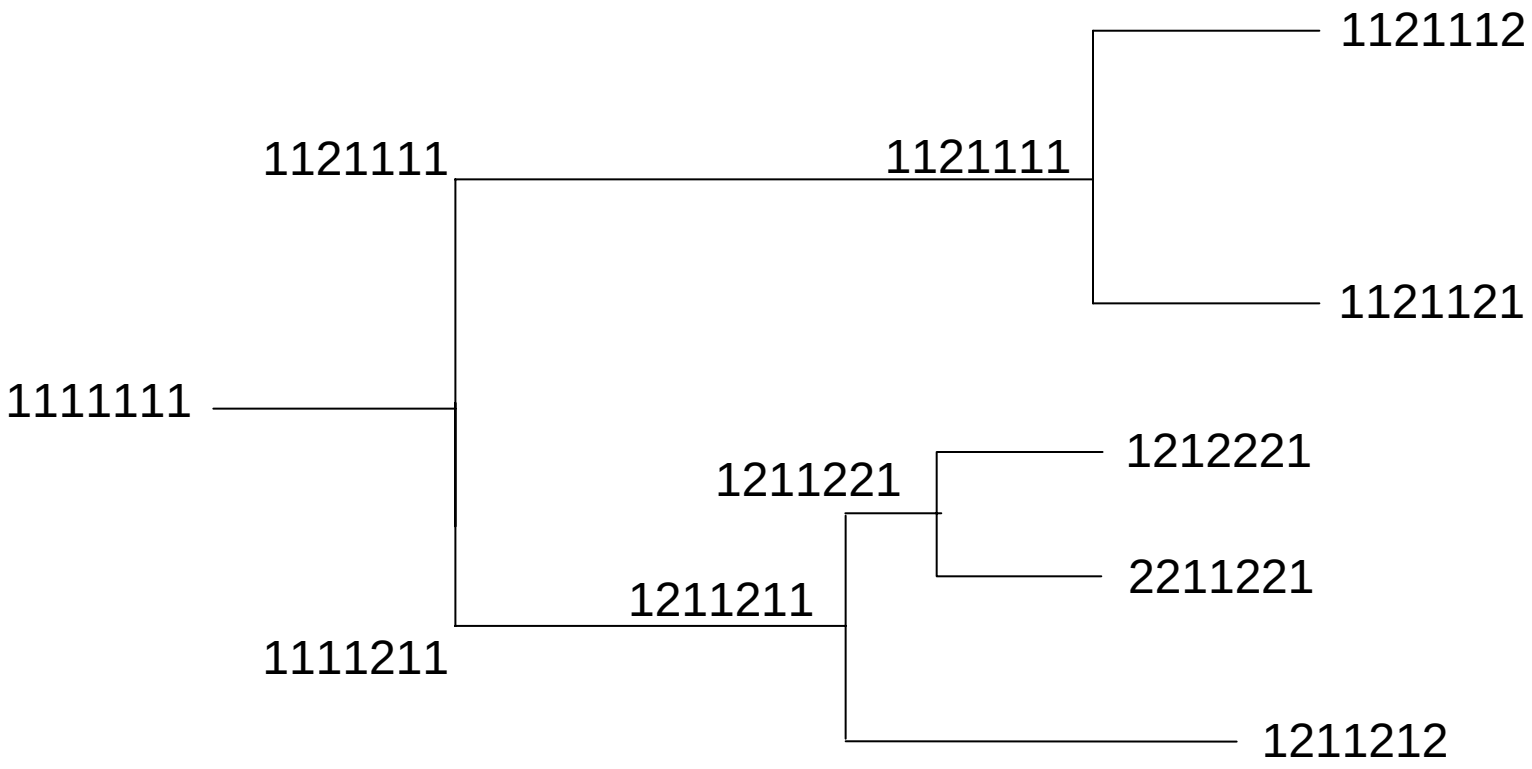
Summary

- Ongoing projects, few clear answers
- LD is useful, but just how much is unknown
- Blocks as firm concepts seems unlikely at this point
- Methods exist that ignore this altogether, and just use genotypes

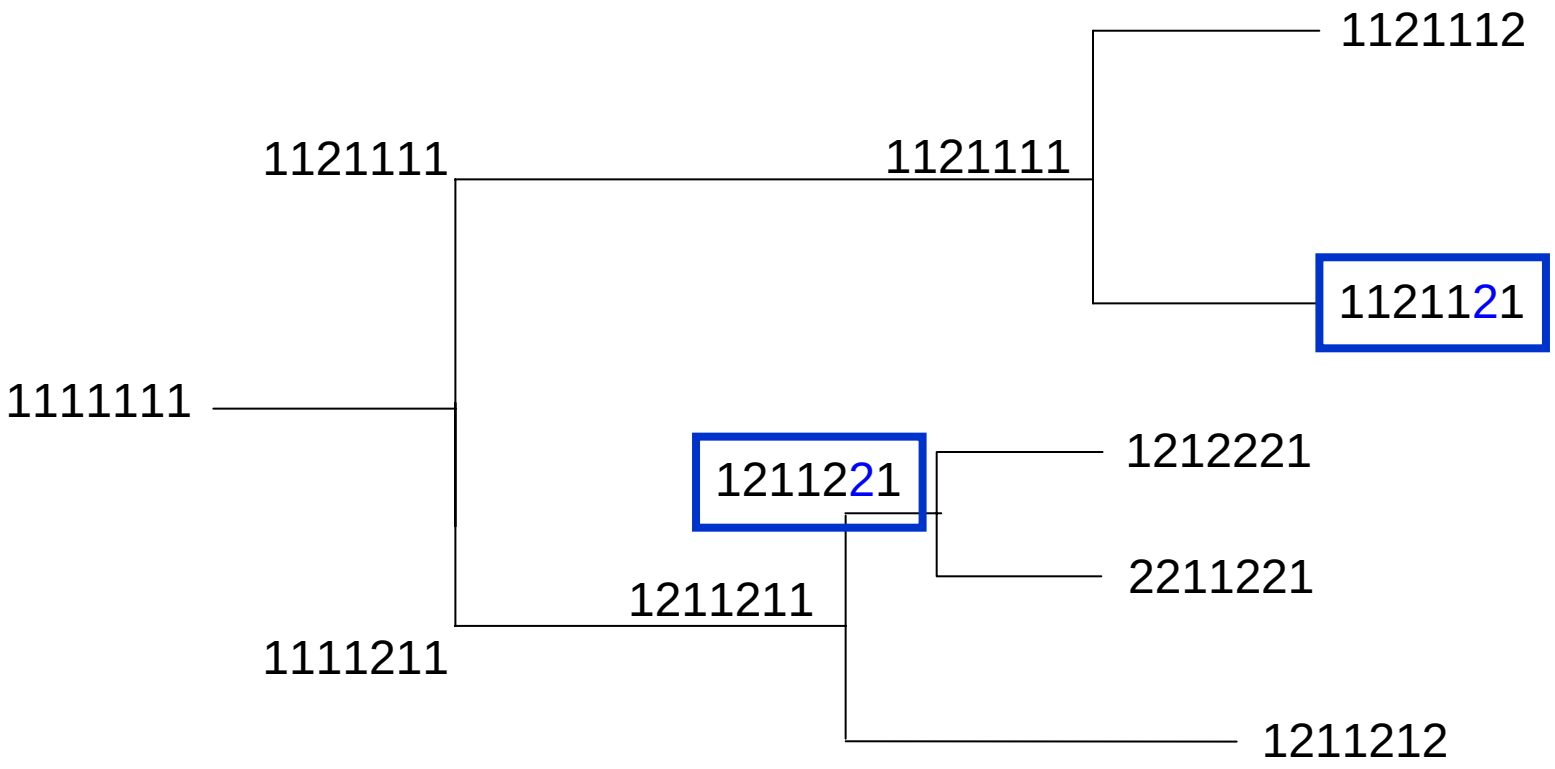
How do we get new haplotypes?

- Mutation events
  - Novel mutation
  - Back mutation
  - Recurrent mutation
- Recombination

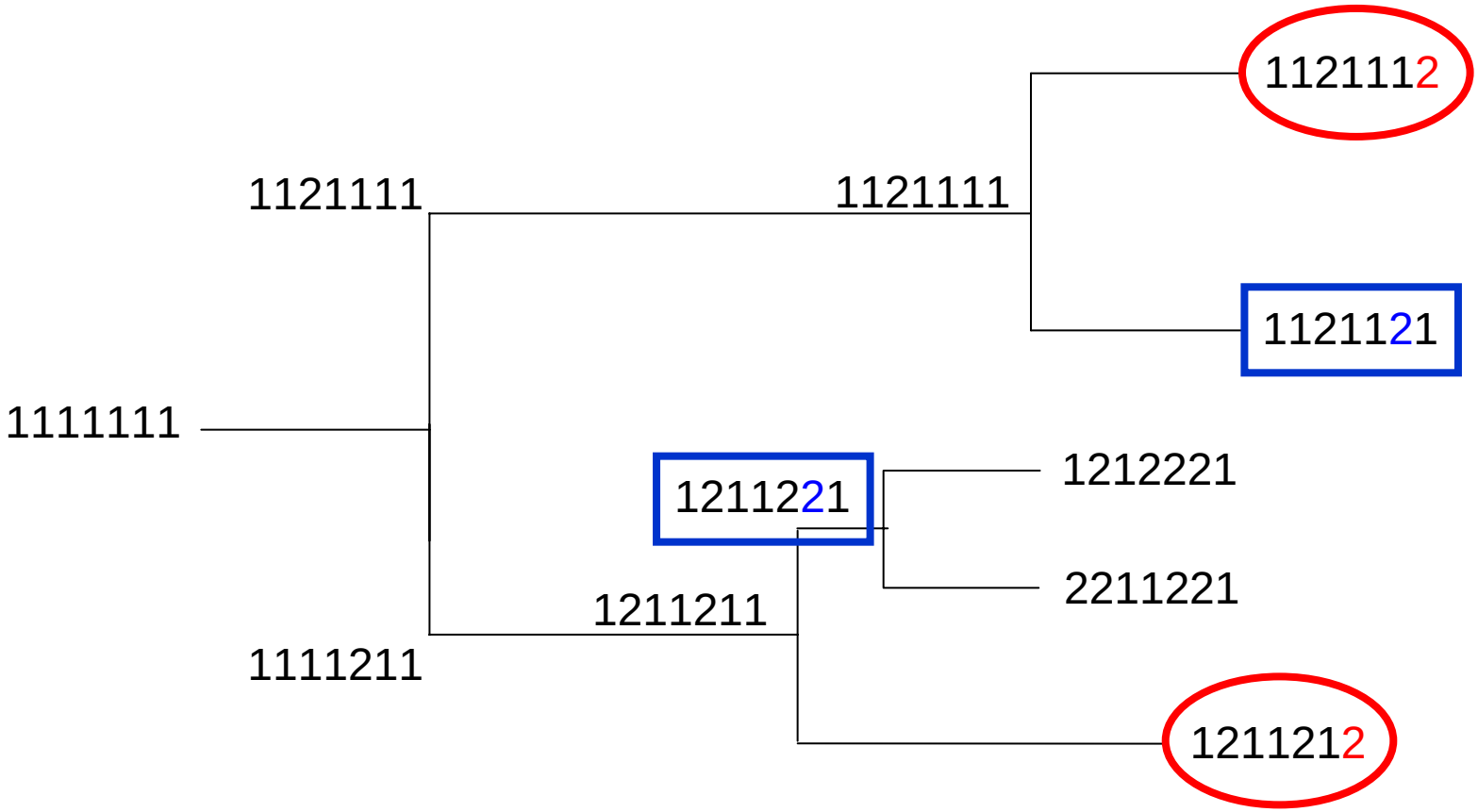
Cladograms (a.k.a. Clades)



Cladograms (a.k.a. Clades)



Cladograms (a.k.a. Clades)



Fantastic online resource for papers

- http://www.nslij-genetics.org/ld/

Bibliography

1) Ke X, Durrant C, et al. Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples.Hum Mol Genet. 2004 Nov 1;13(21):2557-65. Epub 2004 Sep 14.
2) Ke X, Hunt S, et al. The impact of SNP density on fine-scale patterns of linkage disequilibrium. Hum Mol Genet. 2004 Mar 15;13(6):577-88. Epub 2004 Jan 20.
3) Mueller JC, Lohmussaar E, et al. Linkage Disequilibrium Patterns and tagSNP Transferability among European Populations. Am J Hum Genet. 2005 Mar;76(3):387-98. Epub 2005 Jan 06.
4) Cardon, L. R. and G. R. Abecasis Implications of the initial results from the HapMap study. (2003). "Using haplotype blocks to map human complex trait loci." Trends Genet 19(3): 135-40.
5) Wall, J. D. and J. K. Pritchard Complexity of the haplotype block structure (2003). "Haplotype blocks and linkage disequilibrium in the human genome." Nat Rev Genet 4(8): 587-97.
6) Neale, B. and Sham, P. Gene based association analysis The Future of Association studies: Gene-based Analysis and Replication. (2004) AJHG 75:353-362.
7) Page, G. P., V. George, et al. Proving causation in association studies. (2003). 'Are we there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits." Am J Hum Genet 73(4): 711-9.
8) Patil, N., A. J. Berno, et al. (2001). "Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21." Science 294(5547): 1719-23.
9) Gabriel, S. B., S. F. Schaffner, et al. (2002). "The structure of haplotype blocks in the human genome." Science 296(5576): 2225-9.
10) Wang, N., J. M. Akey, et al. (2002). "Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation." Am J Hum Genet 71(5): 1227-34.
11) Stram, D. O., C. A. Haiman, et al. (2003). "Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study." Hum Hered 55(1): 27-36.
12) Weale, M. E., C. Depondt, et al. (2003). "Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping." Am J Hum Genet 73(3): 551-65.

Thanks for listening