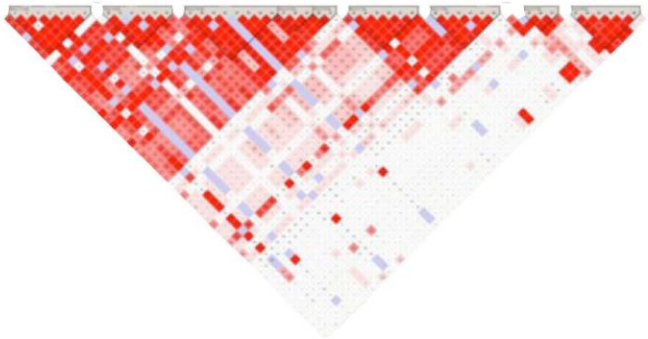
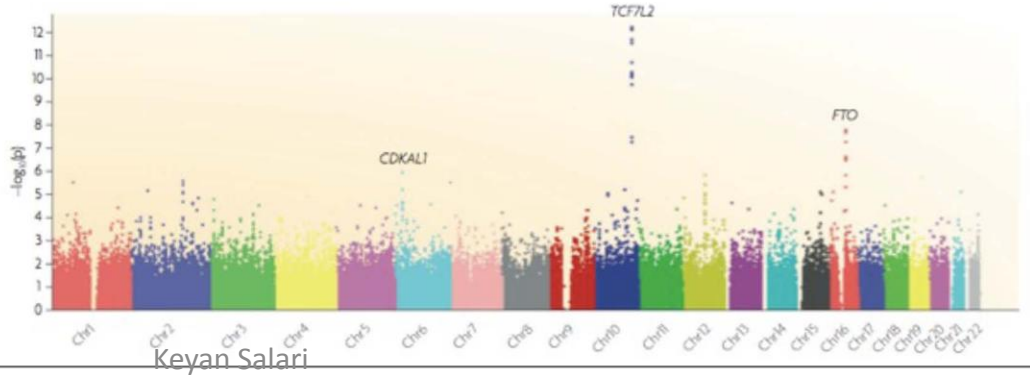
1. Natural variation in the human genome

2. Genetic Association & Linkage Disequilibrium



3. Genome-wide association studies



Hum Genet (2002) 111 : 75–87
DOI 10.1007/s00439-002-0763-x

ORIGINAL INVESTIGATION

Stephanie M. Fullerton · Andrew G. Clark
Kenneth M. Weiss · Scott L. Taylor · Jari H. Stengård
Veikko Salomaa · Eric Boerwinkle
Deborah A. Nickerson

Sequence polymorphism at the human apolipoprotein AII gene (*APOA2*): unexpected deficit of variation in an African-American sample

Sequence APOA2 in 72 people

Look at patterns of polymorphisms

Chimp		Site no.[a]													
SNP haplotype no.	Sequence haplotype no.	1 5 5 C	2 0 1 G	1 8 7 T	1 2 8 G	2 6 7 1 ?	2 0 3 8 G	2 0 8 5 C	2 1 1 5 G	2 2 3 3 C	2 8 1 8 C	2 8 6 8 C	3 9 9 4 C	3 0 2 7 T	3 0 9 2 A	2 0 8 G

Find polymorphisms at these positions.

Reference sequence is listed.

Chimp		Site no.[a]														Sample			
SNP haplotype no.	Sequence haplotype no.	1 5 5 C	2 0 1 G	1 8 7 2 T	1 2 1 8 G	2 6 7 1 ?	2 0 3 8 G	2 0 8 5 C	2 1 1 5 G	2 2 3 3 C	2 8 1 8 C	2 8 6 8 C	3 9 9 4 C	3 0 2 7 T	3 0 9 2 A	2 0 8 G	J	N	R	T
Core re-sequenced samples																			
	S9	G		C		20	●		A	●	●	●	●	●	●	●	0	0	1	1

Sequence of the first chromosome.

Circle is same as reference.

Chimp		Site no.[a]															Sample			
SNP haplotype no.	Sequence haplotype no.	15 5 C	20 1 G	18 7 2 T	12 1 8 G	26 7 1 ?	20 3 8 G	20 8 5 C	21 1 5 G	22 3 3 C	28 1 8 C	28 6 8 C	39 9 4 C	30 2 7 T	30 9 2 A	20 8 G	J	N	R	T
Core re-sequenced samples																				
	S9	G		C		20	●		A	●	●	●	●	●	●	●	0	0	1	1
	S9a	G		C		18	●		A	●	●	●	●	●	●	●	0	1	0	1
	S2	G		C		19	●		●	●	●	●	●	●	●	●	15	10	12	37
	S2a	G		C		20	●		●	●	●	●	●	●	●	●	0	2	3	5
	S2b	G		C		18	●		●	●	●	●	●	●	●	●	0	2	1	3
	S2c	G		C		21	●		●	●	●	●	●	●	●	●	1	0	1	2
	S1d	G		●		19	●		●	●	●	●	●	●	●	●	5	0	0	5
	S1	G		●		16	●		●	●	●	●	●	●	●	●	17	19	14	50
	S1a	G		●		18	●		●	●	●	●	●	●	●	●	5	1	0	6
	S1b	G		●		15	●		●	●	●	●	●	●	●	●	2	0	0	2
	S1c	G		●		17	●		●	●	●	●	●	●	●	●	1	0	0	1
	S6	●		●		16	●		●	●	●	●	●	●	●	●	1	2	0	3
	S5	●		●		14	●		●	T	●	A	●	●	●	●	1	4	2	7
	S3	●		●		14	●		●	T	●	A	●	C	G	A	0	3	6	9
	S7	●		●		13	C		●	●	T	●	●	●	●	●	0	2	0	2
	S8	●		●		13	C		●	●	T	●	●	C	G	●	0	1	1	2
	S4	●		●		13	C		●	●	T	●	T	C	G	●	0	1	6	7
	S4a	?		●		14	C		●	●	T	●	T	C	G	●	0	0	1	1

Chimp		Site no.[a]														Sample			
SNP haplotype no.	Sequence haplotype no.	15 5 C	20 1 G	18 7 2 T	12 1 8 G	26 7 1 ?	20 3 8 G	20 8 5 C	21 1 5 G	22 3 3 C	28 1 8 C	28 6 8 C	39 9 4 C	30 2 7 T	30 9 2 A	20 8 G	J	N	R	T
Core re-sequenced samples																			
	S9	G		C		20	●		A	●	●	●	●	●	●	●	0	0	1	1
	S9a	G		C		18	●		A	●	●	●	●	●	●	●	0	1	0	1
	S2	G		C		19	●		●	●	●	●	●	●	●	●	15	10	12	37
	S2a	G		C		20	●		●	●	●	●	●	●	●	●	0	2	3	5
	S2b	G		C		18	●		●	●	●	●	●	●	●	●	0	2	1	3
	S2c	G		C		21	●		●	●	●	●	●	●	●	●	1	0	1	2
	S1d	G		●		19	●		●	●	●	●	●	●	●	●	5	0	0	5
	S1	G		●		16	●		●	●	●	●	●	●	●	●	17	19	14	50
	S1a	G		●		18	●		●	●	●	●	●	●	●	●	5	1	0	6
	S1b	G		●		15	●		●	●	●	●	●	●	●	●	2	0	0	2
	S1c	G		●		17	●		●	●	●	●	●	●	●	●	1	0	0	1
	S6	●		●		16	●		●	●	●	●	●	●	●	●	1	2	0	3
	S5	●		●		14	●		●	T	●	A	●	●	●	●	1	4	2	7
	S3	●		●		14	●		●	T	●	A	●	C	G	A	0	3	6	9
	S7	●		●		13	C		●	●	T	●	●	●	●	●	0	2	0	2
	S8	●		●		13	C		●	●	T	●	●	C	G	●	0	1	1	2
	S4	●		●		13	C		●	●	T	●	T	C	G	●	0	1	6	7
	S4a	?		●		14	C		●	●	T	●	T	C	G	●	0	0	1	1

Commonly Used Descriptors

- Haplotype Frequencies
  - The frequency of each type of chromosome
  - Contain all the information provided by other summary measures
- Commonly used summaries
  - D
  - D'
  - r^2 or Δ^2

Haplotype Frequencies

		Locus B		Totals
		B	b	
Locus A	A	p_{AB}	p_{Ab}	p_A
	a	p_{aB}	p_{ab}	p_a
Totals		p_B	p_b	1.0

Fill out this table.

X11 is number of times that haplotype is seen.

	2818 C	**2818 T**	
3027 T	X11	X21	# 3027 T alleles
3027 C	X12	x22	#3027 C alleles
	# 2818 C Allele	# 2818 T allele	

	2818 C	2818 T	
3027 T	125/146	2/146	127/146 T alleles
3027 C	9/146	10/146	19/146 C alleles
	134/146 C Allele	12/146 T allele	

Convert to fractions

	2818 C	**2818 T**	
3027 T	.86	.013	.87 T alleles
3027 C	.061	.068	.13 C alleles
	.92 C Allele	.08 T allele	

Linkage Equilibrium Expected for Distant Loci

$$p_{AB} = p_A p_B$$

$$p_{Ab} = p_A p_b = p_A(1 - p_B)$$

$$p_{aB} = p_a p_B = (1 - p_A) p_B$$

$$p_{ab} = p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium Expected for Nearby Loci

$$p_{AB} \neq p_A p_B$$

$$p_{Ab} \neq p_A p_b = p_A(1 - p_B)$$

$$p_{aB} \neq p_a p_B = (1 - p_A) p_B$$

$$p_{ab} \neq p_a p_b = (1 - p_A)(1 - p_B)$$

Disequilibrium Coefficient D_{AB}

$$D_{AB} = p_{AB} - p_A p_B$$

$$p_{AB} = p_A p_B + D_{AB}$$

$$p_{Ab} = p_A p_b - D_{AB}$$

$$p_{aB} = p_a p_B - D_{AB}$$

$$p_{ab} = p_a p_b + D_{AB}$$

Calculate D_{AB}

$$D_{AB} = P_{AB} - P_A P_B$$
$$= .86 - (.87)(.92)$$
$$= .86 - /80$$
$$= .06$$

D_{AB} is hard to interpret

- Sign is arbitrary …
  - A common convention is to set A, B to be the common allele and a, b to be the rare allele
- Range depends on allele frequencies
  - Hard to compare between markers

D' – A scaled version of D

$$D'_{AB} = \begin{cases} \dfrac{D_{AB}}{\min(p_A p_B, p_a p_b)} & D_{AB} < 0 \\ \dfrac{D_{AB}}{\min(p_A p_b, p_a p_B)} & D_{AB} > 0 \end{cases}$$

- Ranges between –1 and +1
  - More likely to take extreme values when allele frequencies are small
  - ±1 implies at least one of the observed haplotypes was not observed

Calculate D’

If $D_{AB} > 0$

$D' = D_{AB} / \min(P_A P_b, P_a P_B)$

$= .06 / \min[(.87 * .08), (.13 * .92)]$

$= .06 / \min (.069 , .12)$

$= .06 / .069 = .87$

If DAB < 0

$D' = D_{AB} / \min (P_A P_B, P_a P_b)$

More on D'

- Pluses:
  - D' = 1 or D' = -1 means no evidence for recombination between the markers
  - If allele frequencies are similar, high D' means the markers are good surrogates for each other
- Minuses:
  - D' estimates inflated in small samples
  - D' estimates inflated when one allele is rare

Δ^2 (also called r^2)

$$\Delta^2 = \frac{D_{AB}^2}{p_A(1-p_A)p_B(1-p_B)}$$

$$= \frac{\chi^2}{2n}$$

- Ranges between 0 and 1
  - 1 when the two markers provide identical information
  - 0 when they are in perfect equilibrium
- Expected value is $1/2n$

Calculate R

$R = D_{AB} / SQR(P_A P_a P_B P_b)$

$= .06 / SQR(.87 * .13 * .92 * .08)$

$= .06 / SQR(7.2 \times 10^{-3})$

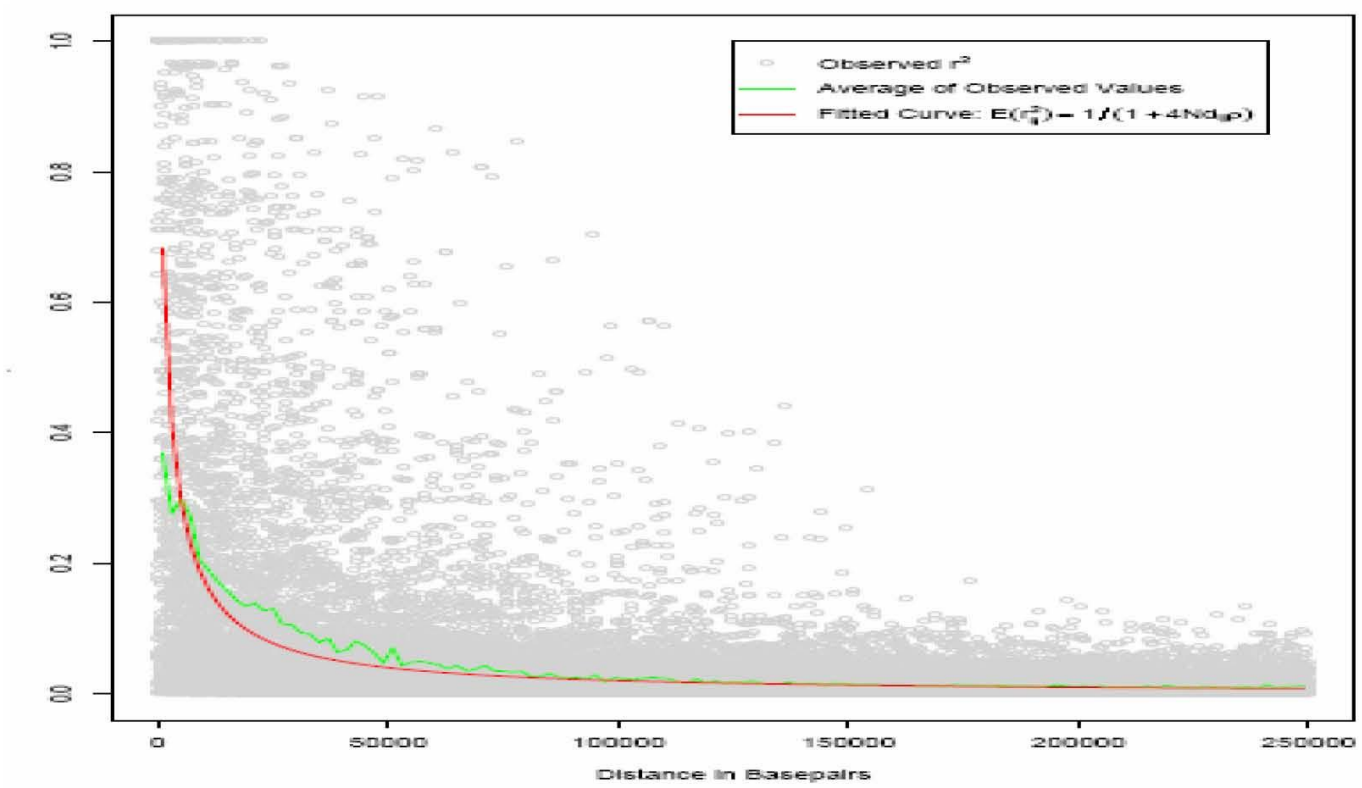
$= .06 / .085 = .706$

$R^2 = .497$

More on r^2

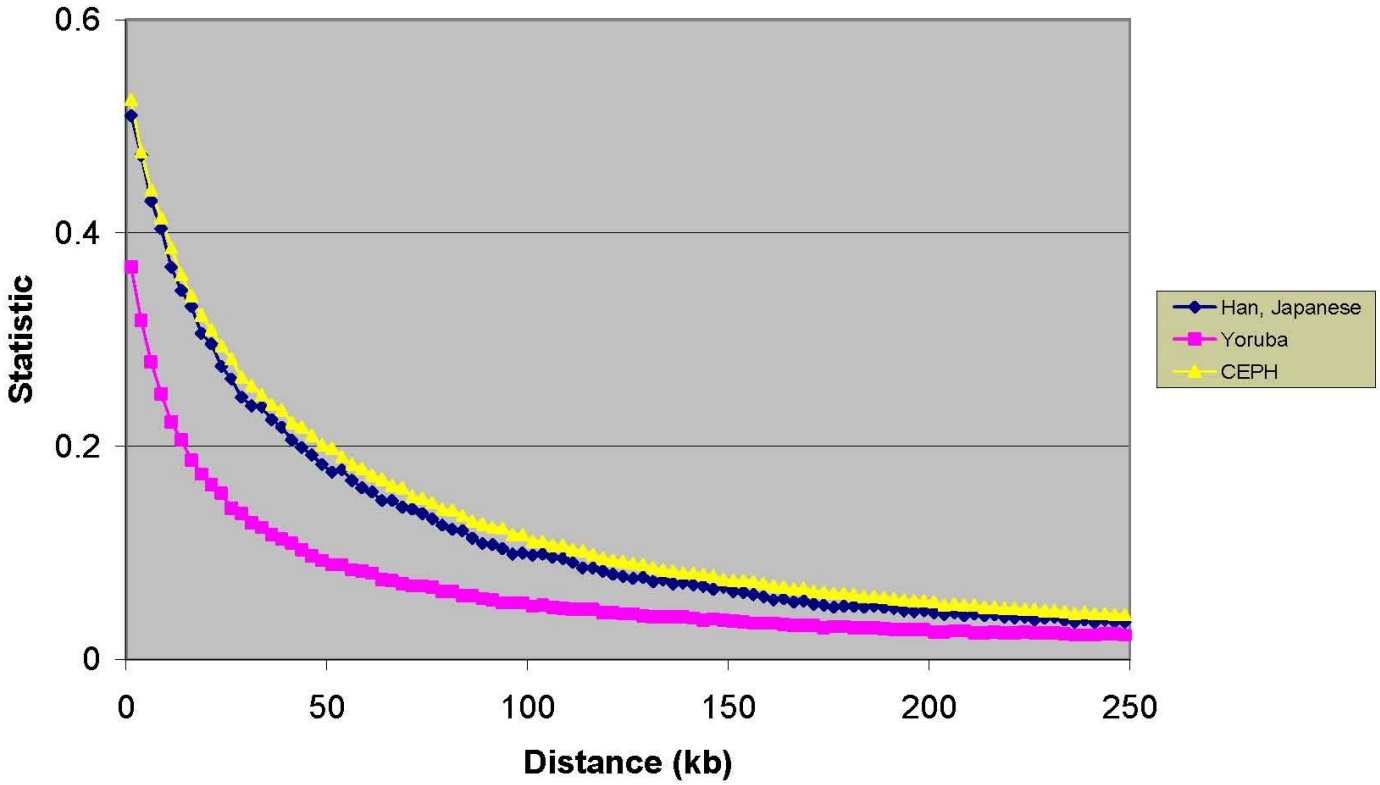
- $r^2 = 1$ implies the markers provide exactly the same information
- The measure preferred by population geneticists
- Measures loss in efficiency when marker A is replaced with marker B in an association study
  - With some simplifying assumptions (e.g. see Pritchard and Przeworski, 2001)

Summarizing Disequilibrium



slide created by Goncarlo Abecasis

Comparing Populations …



LD extends further in CEPH and the Han/Japanese than in the Yoruba