Heritability in humans: MZ twins

Mean each pair = z_i

Each individual = z_{ij}

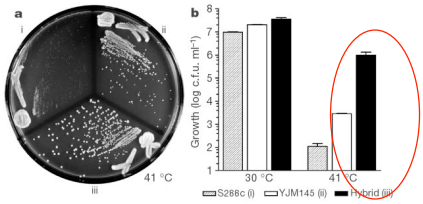
Total mean sq = $\Sigma \Sigma(z_{ij} - z)^2 / T = \sigma_t^2$

Within pairs mean sq = $\Sigma \Sigma(z_{ij} - z_i)^2 / N = \sigma_w^2$

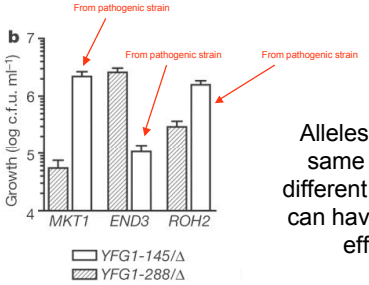
Between pairs mean sq = $2 \Sigma(z_i - z)^2 / N\text{-}1 = \sigma_b^2$

$h^2 = (\sigma_b^2 - \sigma_w^2) / \sigma_t^2$

NO progeny as extreme as diploid hybrid

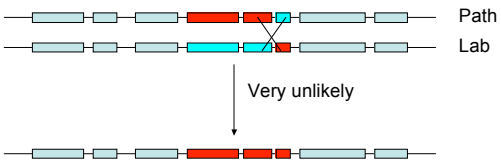


Three mutant genes

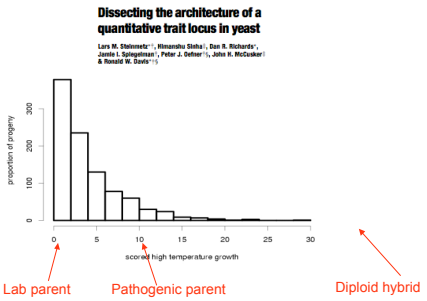


Alleles from the same strain at different genes/loci can have different effects.

Linked mutations of opposite effect

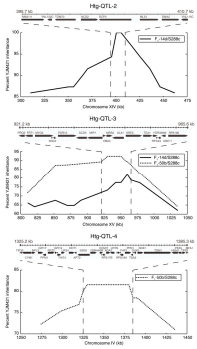


Why is distribution of progeny so skewed?

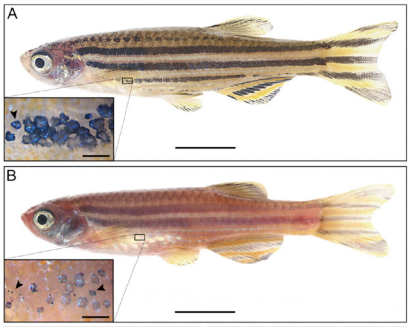


Why is distribution of progeny so skewed?

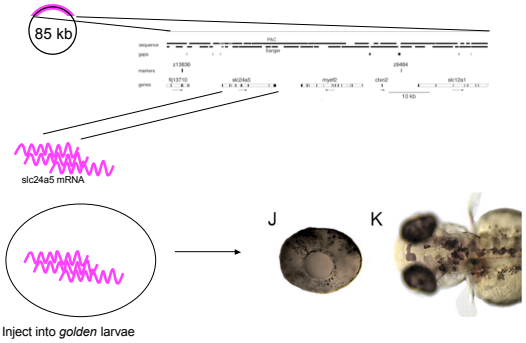
Hypothesis: interaction between loci (see problem set)



Golden mutation



Fine-mapping

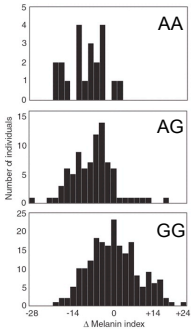


No truncation in humans, but…

No other species have the Thr allele: what does this mean?
Could be deleterious, just an accidental mutation.
Could be advantageous for some humans, no other species.

Correlates with human differences

Note that this is not linkage analysis. Individuals are unrelated.



Genetic association studies

What is a haplotype?

Fig. 11.25

What is a haplotype?

Fig. 11.25

What is a haplotype?

(a) Formation of haplotypes over evolutionary time

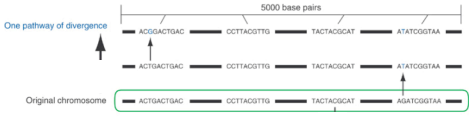


Fig. 11.25

What is a haplotype?

(a) Formation of haplotypes over evolutionary time

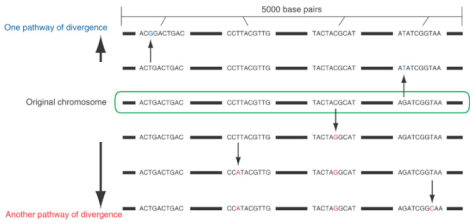
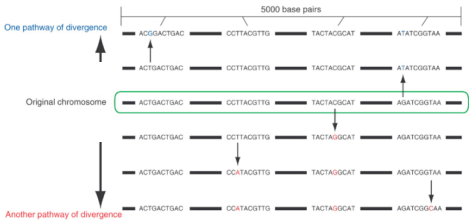


Fig. 11.25

What is a haplotype?

(a) Formation of haplotypes over evolutionary time



(b) Comparison of two final haplotypes with a compilation of five SNPs spread over 5000 base pairs

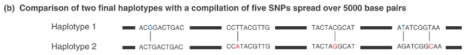


Fig. 11.25

Association mapping (qualitative)

(a) New mutation (M) in ancestral population

Fig. 11.26

Association mapping (qualitative)

(a) New mutation (M) in ancestral population

(b) Three regions of haplotypes

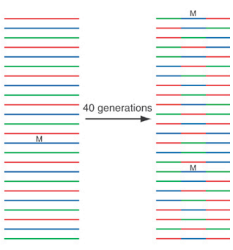


Fig. 11.26

Association mapping (qualitative)

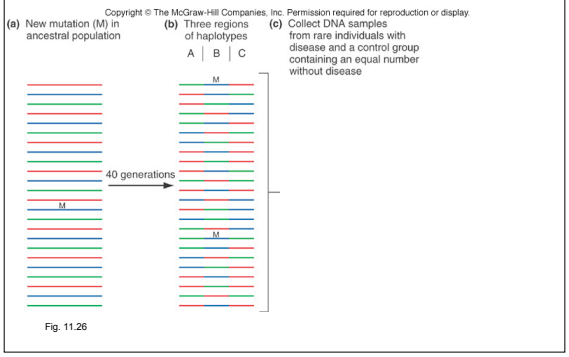


Fig. 11.26

Association mapping (qualitative)

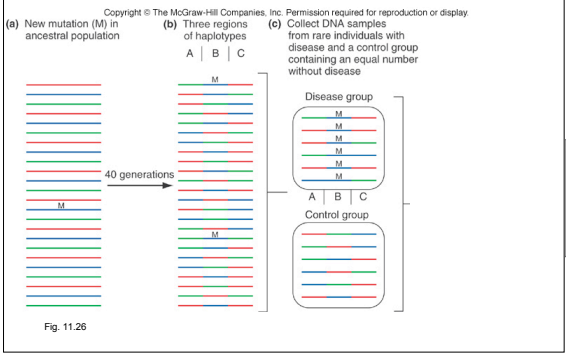


Fig. 11.26

Association mapping (qualitative)

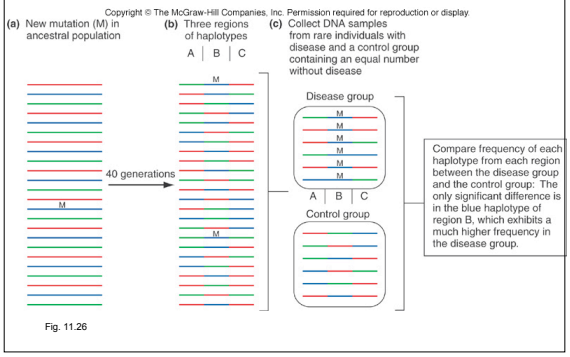
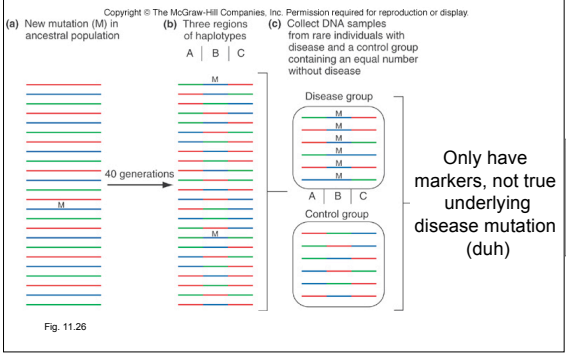


Fig. 11.26

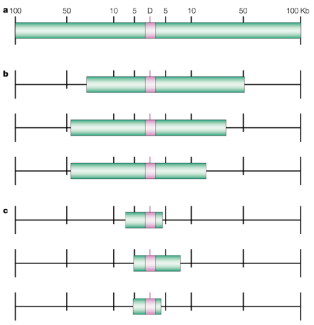
Association mapping (qualitative)



Only have markers, not true underlying disease mutation (duh)

Fig. 11.26

Linkage disequilibrium

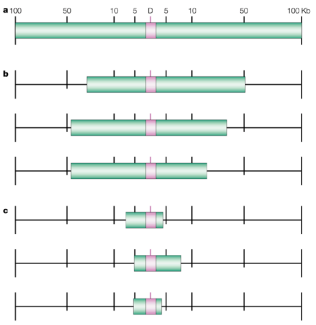


GENETIC ANALYSIS OF AFRICAN POPULATIONS: HUMAN EVOLUTION AND COMPLEX DISEASE

Nature Reviews | Genetics

Linkage disequilibrium

Marker alleles appear together in disease population more than you would expect

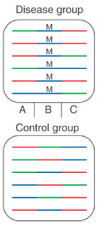


GENETIC ANALYSIS OF AFRICAN POPULATIONS: HUMAN EVOLUTION AND COMPLEX DISEASE

Nature Reviews | Genetics

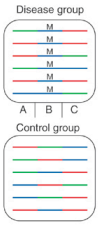
Association mapping (qualitative)

In association, we don't calculate a recombination fraction; we aren't counting recombinants.



Association mapping (qualitative)

In association, we don't calculate a recombination fraction; we aren't counting recombinants. Each individual could represent a different number of generations (and recomb) since mutation arose.



The association revolution

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1

A Genome-Wide Association Study Identifies *IL23R* as an Inflammatory Bowel Disease Gene

Get markers by (re-)sequencing

Chromosome	SNP	Insertion /deletion	Total variations	Screened bases	Frequency of variation (bp)	Length of chromosome* (Mb)	Density of identified variations (bp)
1	15 166	1181	16 347	14 344 460	877	220	13 458
2	12 772	1126	13 898	11 608 613	835	240	17 269
3	9052	805	9857	8 813 579	894	200	20 290
4	6 336	638	6974	5 826 029	835	186	26 670
5	8 284	777	9061	7 648 904	844	182	20 086
6	11 609	1023	12 632	11 252 798	891	172	13 616
7	11 969	931	12 900	11 528 580	894	146	11 318
8	4 545	364	4 909	3 887 257	792	146	29 741
9	5 561	441	6002	4 802 482	800	113	18 827
10	5 819	470	6 289	5 014 774	797	130	20 671
11	5 995	430	6 425	6 109 367	951	132	20 545
12	7 108	570	7 678	7 190 880	937	134	17 452
13	2 632	242	2 874	2 236 249	778	99	34 447
14	5 827	436	6 263	5 231 516	835	87	13 891
15	4 118	324	4 442	4 001 428	901	80	18 010
16	5 154	314	5 468	5 517 492	1 009	75	13 716
17	5 673	434	6 107	6 403 592	1 049	78	12 772
18	2 263	203	2 466	1 947 367	790	79	32 036
19	6 952	425	7 377	6 301 849	854	58	7 862
20	5 075	363	5 438	4 688 878	862	61	11 237
21	3 452	220	3 672	2 970 778	809	33	8 087
22	5 198	261	5 459	5 814 927	1 065	36	6 595
X	3 105	246	3 351	6 169 474	1 841	128	38 198
Y	24	0	24	556 613	23 192	19	791 667

SNP, Single-nucleotide polymorphism
*Data from Venter et al. 2001

Hisanori Haga · Ryo Yamada · Yoav Ohnishi Yusuke Nakamura · Toshihiro Tanaka

Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190 562 genetic variations in the human genome

Association scan, qualitative

Table 1. Association of rs3773472 with knee osteoarthritis in the genome-wide analysis

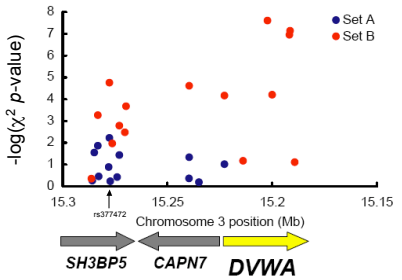
	Case					Control					
	Genotype					Genotype					
Population	CC	CG	GG	Sum	Allele G frequency	CC	CG	GG	Sum	Allele G frequency	*P* value for allele frequency
Set A	52	37	5	94	0.250	259	279	77	615	0.352	0.0059
Set B	324	272	50	646	0.288	240	314	74	628	0.368	0.000017
Set A+B combined	376	309	55	740	0.283	499	593	151	1243	0.360	0.00000065

χ^2 test

	osteoarthritis	controls
C's	141	797
G's	47	433

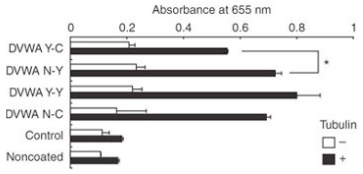
Common variants in *DVWA* on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis

Fine-mapping



Common variants in *DVWA* on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis

Beginnings of molecular confirmation

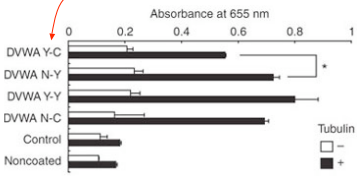


Common variants in *DVWA* on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis

Yoshinori Miyamoto[1], Dongquan Shi[1], Masahiro Nakajima[1], Koichi Ozaki[2], Akihiro Sudo[3], Akihiro Kotani[3], Atsumasa Uchida[3], Yoshihiro Tanaka[3], Naoshi Fukui[4], Tatsuhiko Tsunoda[5], Atsushi Takahashi[5], Yusuke Nakamura[5], Qing Jiang[1] & Shiro Ikegawa[1]

Beginnings of molecular confirmation

coding polymorphisms



Common variants in *DVWA* on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis

Yoshinori Miyamoto[1], Dongquan Shi[1], Masahiro Nakajima[1], Koichi Ozaki[2], Akihiro Sudo[3], Akihiro Kotani[3], Atsumasa Uchida[3], Yoshihiro Tanaka[3], Naoshi Fukui[4], Tatsuhiko Tsunoda[5], Atsushi Takahashi[5], Yusuke Nakamura[5], Qing Jiang[1] & Shiro Ikegawa[1]

Another example: qualitative

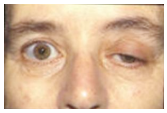
Table 1 | Association of rs16862847 G with early onset of disease in French and UK myasthenia gravis patients

			G allele			Genotype			AG+GG/AA	
Population	Group	*N*	Frequency	Odds ratio (95% CI)	*P*-value	AA	AG	GG	Odds ratio (95% CI)	*P*-value
France	P_{EO}	50	0.23	2.35 (1.31–4.23)	0.0049	27	23	0	3.3 (1.67–6.52)	0.0008
	P_{LFEO}	130	0.11			104	24	2		
	C	151	0.11			120	28	3		
United Kingdom	P_{EO}	46	0.21	2.01 (1.04–3.87)	0.029	29	15	2	2.08 (0.98–4.4)	0.044
	P_{LFEO}	104	0.11			87	12	5		
	C	109	0.11			85	23	1		
Combined	P_{EO}	96	0.22	2.19 (1.41–3.39)	0.00048	56	38	2	2.66 (1.6–4.41)	0.00015
	P_{LFEO}	234	0.11			191	36	7		
	C	260	0.11			205	51	4		

P_{EO} indicates patients with early onset of myasthenia gravis, defined by age at the onset of disease in the lower quartile of the distribution in each cohort (≤21 yr in French patients, including 8 patients of age 21, and ≤18 yr in UK patients, including 10 patients of age 18). P_{LFEO} indicates patients in the upper three quartiles of the distribution. C indicates matched controls. Odds ratios were determined relative to the control populations. The exact *P*-values were tested two-sided in the French cohort, and one-sided in the replication UK cohort. The combined data set was tested two-sided using stratified 2 × 2 tables. The data also fitted a partially (or semi-) dominant model (Cochran–Armitage test with strata: $P = 4 \times 10^{-5}$; odds ratio for AG versus AA genotypes, 2.27, 95% CI, 1.4–3.69; for GG versus AA genotypes, 5.15, 95% CI, 1.96–13.6).

An IRF8-binding promoter variant and AIRE control *CHRNA1* promiscuous expression in thymus

Matthieu Giraud[1,2], Richard Taubert[1], Claire Vandiedonck[1], Xiayi Ke[1], Matthieu Lévi-Strauss[1], Franco Pagani[3], Francisco E. Baralle[3], Bruno Eymard[4], Christine Tranchant[4], Philippe Gajdos[4], Angela Vincent[5], Nick Willcox[5], David Beeson[5], Bruno Kyewski[2] & Henri-Jean Garchon[1]



http://www.wrongdiagnosis.com/c/cholinesterase_inhibitors.htm

Another example: qualitative

Early onset

Table 1 | Association of rs16862847 G with early onset of disease in French and UK myasthenia gravis patients

			G allele			Genotype			AG+GG/AA	
Population	Group	*N*	Frequency	Odds ratio (95% CI)	*P*-value	AA	AG	GG	Odds ratio (95% CI)	*P*-value
France	P_{EO}	50	0.23	2.35 (1.31–4.23)	0.0049	27	23	0	3.3 (1.67–6.52)	0.0008
	P_{LFEO}	130	0.11			104	24	2		
	C	151	0.11			120	28	3		
United Kingdom	P_{EO}	46	0.21	2.01 (1.04–3.87)	0.029	29	15	2	2.08 (0.98–4.4)	0.044
	P_{LFEO}	104	0.11			87	12	5		
	C	109	0.11			85	23	1		
Combined	P_{EO}	96	0.22	2.19 (1.41–3.39)	0.00048	56	38	2	2.66 (1.6–4.41)	0.00015
	P_{LFEO}	234	0.11			191	36	7		
	C	260	0.11			205	51	4		

P_{EO} indicates patients with early onset of myasthenia gravis, defined by age at the onset of disease in the lower quartile of the distribution in each cohort (≤21 yr in French patients, including 8 patients of age 21, and ≤18 yr in UK patients, including 10 patients of age 18). P_{LFEO} indicates patients in the upper three quartiles of the distribution. C indicates matched controls. Odds ratios were determined relative to the control populations. The exact *P*-values were tested two-sided in the French cohort, and one-sided in the replication UK cohort. The combined data set was tested two-sided using stratified 2 × 2 tables. The data also fitted a partially (or semi-) dominant model (Cochran–Armitage test with strata: $P = 4 \times 10^{-5}$; odds ratio for AG versus AA genotypes, 2.27, 95% CI, 1.4–3.69; for GG versus AA genotypes, 5.15, 95% CI, 1.96–13.6).

An IRF8-binding promoter variant and AIRE control *CHRNA1* promiscuous expression in thymus

Matthieu Giraud[1,2], Richard Taubert[1], Claire Vandiedonck[1], Xiayi Ke[1], Matthieu Lévi-Strauss[1], Franco Pagani[3], Francisco E. Baralle[3], Bruno Eymard[4], Christine Tranchant[4], Philippe Gajdos[4], Angela Vincent[5], Nick Willcox[5], David Beeson[5], Bruno Kyewski[2] & Henri-Jean Garchon[1]

Another example: qualitative

Normal onset

Table 1 | Association of rs16862847 G with early onset of disease in French and UK myasthenia gravis patients

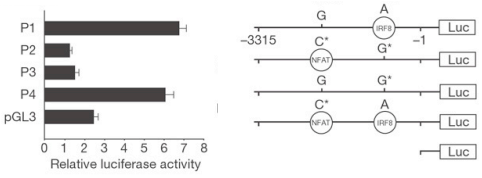
			G allele			Genotype			AG+GG/AA	
Population	Group	*N*	Frequency	Odds ratio (95% CI)	*P*-value	AA	AG	GG	Odds ratio (95% CI)	*P*-value
France	P_{EO}	50	0.23	2.35 (1.31–4.23)	0.0049	27	23	0	3.3 (1.67–6.52)	0.0008
	P_{LFEO}	130	0.11			104	24	2		
	C	151	0.11			120	28	3		
United Kingdom	P_{EO}	46	0.21	2.01 (1.04–3.87)	0.029	29	15	2	2.08 (0.98–4.4)	0.044
	P_{LFEO}	104	0.11			87	12	5		
	C	109	0.11			85	23	1		
Combined	P_{EO}	96	0.22	2.19 (1.41–3.39)	0.00048	56	38	2	2.66 (1.6–4.41)	0.00015
	P_{LFEO}	234	0.11			191	36	7		
	C	260	0.11			205	51	4		

P_{EO} indicates patients with early onset of myasthenia gravis, defined by age at the onset of disease in the lower quartile of the distribution in each cohort (≤21 yr in French patients, including 8 patients of age 21, and ≤18 yr in UK patients, including 10 patients of age 18). P_{LFEO} indicates patients in the upper three quartiles of the distribution. C indicates matched controls. Odds ratios were determined relative to the control populations. The exact *P*-values were tested two-sided in the French cohort, and one-sided in the replication UK cohort. The combined data set was tested two-sided using stratified 2 × 2 tables. The data also fitted a partially (or semi-) dominant model (Cochran–Armitage test with strata: $P = 4 \times 10^{-5}$; odds ratio for AG versus AA genotypes, 2.27, 95% CI, 1.4–3.69; for GG versus AA genotypes, 5.15, 95% CI, 1.96–13.6).

An IRF8-binding promoter variant and AIRE control *CHRNA1* promiscuous expression in thymus

Matthieu Giraud[1,2], Richard Taubert[1], Claire Vandiedonck[1], Xiayi Ke[1], Matthieu Lévi-Strauss[1], Franco Pagani[3], Francisco E. Baralle[3], Bruno Eymard[4], Christine Tranchant[4], Philippe Gajdos[4], Angela Vincent[5], Nick Willcox[5], David Beeson[5], Bruno Kyewski[2] & Henri-Jean Garchon[1]

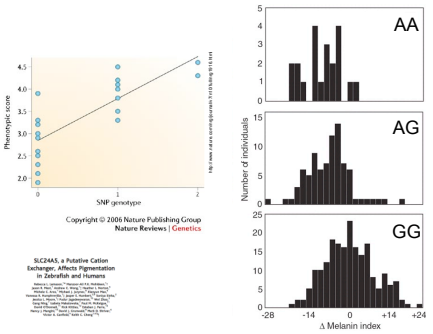
A promoter SNP (at last)



An IRF8-binding promoter variant and AIRE control *CHRNA1* promiscuous expression in thymus

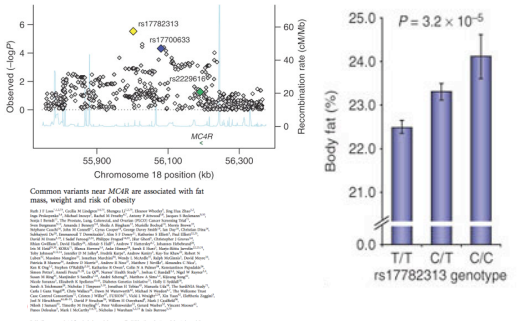
Matthieu Giraud[1,2], Richard Taubert[1], Claire Vandiedonck[1], Xiayi Ke[1], Matthieu Lévi-Strauss[1], Franco Pagani[3], Francisco E. Baralle[3], Bruno Eymard[4], Christine Tranchant[4], Philippe Gajdos[4], Angela Vincent[5], Nick Willcox[5], David Beeson[5], Bruno Kyewski[2] & Henri-Jean Garchon[1]

Quantitative test for association



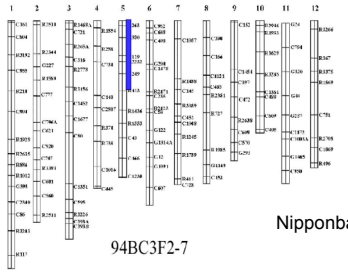
SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans

Association scan, quantitative



Common variants near *MC4R* are associated with fat mass, weight and risk of obesity

Rice yield: start with linkage

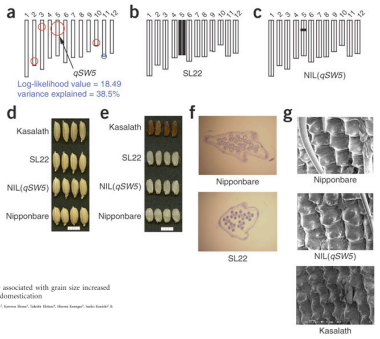


94BC3F2-7

Nipponbare x Kasalath
↓
F1
↓
F2

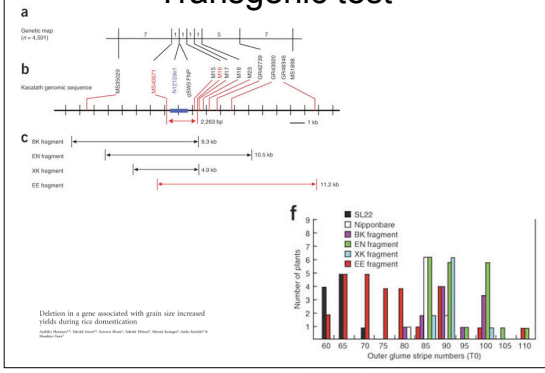
Deletion in a gene associated with grain size increased yields during rice domestication

Narrow down by backcross



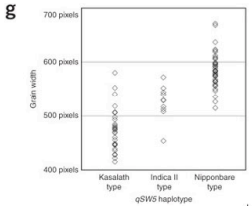
Deletion in a gene associated with grain size increased yields during rice domestication

Transgenic test



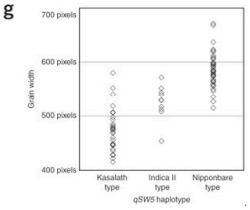
Deletion in a gene associated with grain size increased yields during rice domestication

Association across 100 cultivars (quant)



Deletion in a gene associated with grain size increased yields during rice domestication

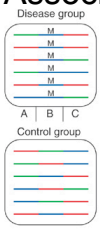
Association across 100 cultivars (quant)

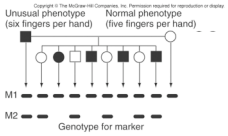


Conclude that these alleles are common across many cultivars, not just in linkage cross.

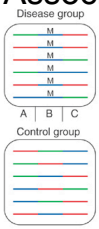
Deletion in a gene associated with grain size increased yields during rice domestication

Association *vs.* linkage

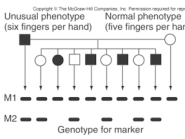




Association *vs.* linkage

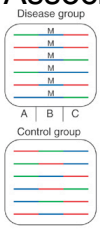


Unrelated individuals (usually)

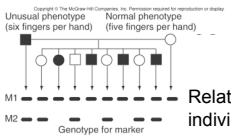


Related individuals

Association *vs.* linkage



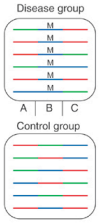
Unrelated individuals (usually)



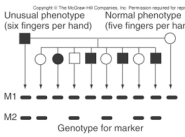
Related individuals

Extreme of linkage study is one large family; less likely that phenotype has multiple genetic causes (locus heterogeneity).

Association *vs.* linkage



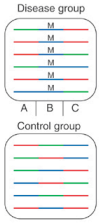
Unrelated individuals



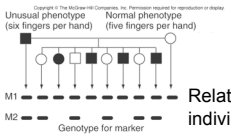
Related individuals

Strong, easy to detect

Association *vs.* linkage



Unrelated individuals

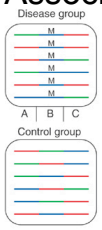


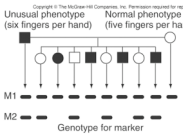
Related individuals

Strong, easy to detect, but rare in population

Association *vs.* linkage

Unrelated individuals



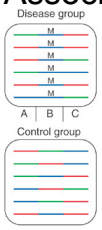


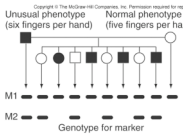
Related individuals

Strong, easy to detect, but rare in population; may not be reflective of common disease.

Association *vs.* linkage

Unrelated individuals



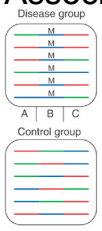


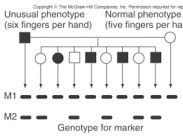
Related individuals

Strong, easy to detect, but rare in population; may not be reflective of common disease.
Also, hard to collect family data.

Association *vs.* linkage

Unrelated individuals





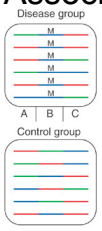
Related individuals

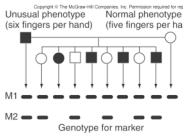
Common but weak effects

Strong, easy to detect, but rare in population; may not be reflective of common disease.
Also, hard to collect family data.

Association *vs.* linkage

Unrelated individuals





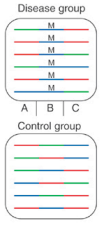
Related individuals

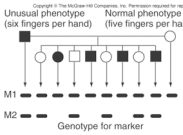
Common but weak effects; need 1000's of samples to detect

Strong, easy to detect, but rare in population; may not be reflective of common disease.
Also, hard to collect family data.

Association *vs.* linkage

Unrelated individuals





Related individuals

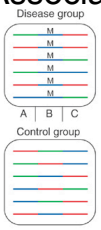
Common but weak effects; need 1000's of samples to detect. If no common cause, can fail.

Strong, easy to detect, but rare in population; may not be reflective of common disease.
Also, hard to collect family data.

Another key feature of association mapping: resolution

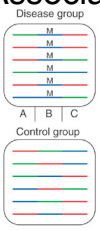
Association *vs.* linkage

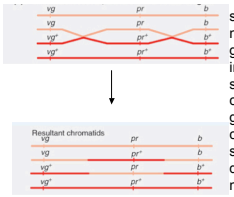
many recombinations have happened since common ancestor; shared region is small; no signal for distant markers



Association *vs.* linkage

many recombinations have happened since common ancestor; shared region is small; no signal for distant markers

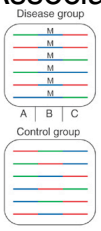


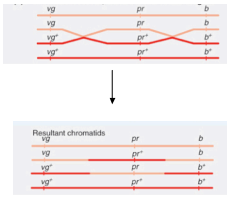


small number of generations; individuals share big chunks of genome; can get signal at distant markers

Association *vs.* linkage

many recombinations have happened since common ancestor; shared region is small; no signal for distant markers



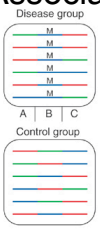


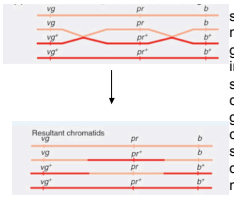
small number of generations; individuals share big chunks of genome; can get signal at distant markers

So you need very high density of markers to get signal in an association study, but you get very high spatial resolution.

Association *vs.* linkage

many recombinations have happened since common ancestor; shared region is small; no signal for distant markers





small number of generations; individuals share big chunks of genome; can get signal at distant markers

In the "old days" of sparse markers, linkage analysis was the best strategy.

But there is a pitfall of association tests: "population structure"

Diabetes in Native Americans

Diabetes in Native Americans

DIABETES MELLITUS IN AMERICAN (PIMA) INDIANS

PETER H. BENNETT
THOMAS A. BURCH*
MAX MILLER†
Southwestern Field Studies Section, Epidemiology and Field Studies Branch, National Institute of Arthritis and Metabolic Diseases, Phoenix, Arizona

Summary The prevalence of diabetes mellitus among the Pima Indians, who live in a hot desert environment in Arizona, U.S.A., has been determined by means of systematic glucose-tolerance tests. Using conservative conventional criteria the prevalence was 50%, among those aged 35 years and over. Frequency distributions of the two-hour post-carbohydrate-load plasma-glucose levels were clearly bimodal above 35 years of age. The findings indicate that the Pima Indians have the highest prevalence of diabetes mellitus yet recorded, and that in this population normal and hyperglycæmic groups may be logically separated on the basis of the bimodality of the frequency distributions of two-hour post-load glucose levels.

(1971)

Diabetes in Native Americans

DIABETES MELLITUS IN AMERICAN (PIMA) INDIANS

PETER H. BENNETT
THOMAS A. BURCH*
MAX MILLER†
Southwestern Field Studies Section, Epidemiology and Field Studies Branch, National Institute of Arthritis and Metabolic Diseases, Phoenix, Arizona

Summary The prevalence of diabetes mellitus among the Pima Indians, who live in a hot desert environment in Arizona, U.S.A., has been determined by means of systematic glucose-tolerance tests. Using conservative conventional criteria the prevalence was 50%, among those aged 35 years and over. Frequency distributions of the two-hour post-carbohydrate-load plasma-glucose levels were clearly bimodal above 35 years of age. The findings indicate that the Pima Indians have the highest prevalence of diabetes mellitus yet recorded, and that in this population normal and hyperglycæmic groups may be logically separated on the basis of the bimodality of the frequency distributions of two-hour post-load glucose levels.

(1971)

Family studies indicate it is at least partly genetic, not environmental.

Association mapping causal loci

Typed IgG heavy chains with protein assay.
Phenotypes can serve as markers too…

$Gm^{2;5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

Typed IgG heavy chains with protein assay.
Phenotypes can serve as markers too…

(Multiple proteins from chr 14 region: haplotype)

$Gm^{2;5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

Association Between Diabetes and the Haplotype $Gm^{2;5,13,14}$ Among Residents of the Gila River Indian Community—No Restrictions on Age or Fraction of Indian Heritage

$Gm^{2;5,13,14}$	No. of Subjects	No. (%) with Diabetes
Present	293	23 (8)
Absent	4,627	1,343 (29)

NOTE.—Diabetes was inversely associated with the haplotype $Gm^{2;5,13,14}$ ($\chi^2 = 61.6$; df = 1; $P < .001$). Prevalence ratio = 0.27 (95% confidence interval = 0.18–0.40).

$Gm^{2;5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

Association Between Diabetes and the Haplotype $Gm^{2;5,13,14}$ Among Residents of the Gila River Indian Community—No Restrictions on Age or Fraction of Indian Heritage

$Gm^{2;5,13,14}$	No. of Subjects	No. (%) with Diabetes
Present	293	23 (8)
Absent	4,627	1,343 (29)

NOTE.—Diabetes was inversely associated with the haplotype $Gm^{2;5,13,14}$ ($\chi^2 = 61.6$; df = 1; $P < .001$). Prevalence ratio = 0.27 (95% confidence interval = 0.18–0.40).

	diabetes	control
Gm	23	270
no Gm	1343	3284

$Gm^{2;5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

Association Between Diabetes and the Haplotype $Gm^{3,5,13,14}$ Among Residents of the Gila River Indian Community—No Restrictions on Age or Fraction of Indian Heritage

$Gm^{3,5,13,14}$	No. of Subjects	No. (%) with Diabetes
Present	293	23 (8)
Absent	4,627	1,343 (29)

NOTE.—Diabetes was inversely associated with the haplotype $Gm^{3,5,13,14}$ [$\chi^2 = 61.6$; df = 1; $P < .001$]. Prevalence ratio = 0.27 (95% confidence interval = 0.18–0.40).

	diabetes	control
Gm	23	270
no Gm	1343	3284

$Gm^{3,5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

"Gm is protective against diabetes?"

Association Between Diabetes and the Haplotype $Gm^{3,5,13,14}$ Among Residents of the Gila River Indian Community—No Restrictions on Age or Fraction of Indian Heritage

$Gm^{3,5,13,14}$	No. of Subjects	No. (%) with Diabetes
Present	293	23 (8)
Absent	4,627	1,343 (29)

NOTE.—Diabetes was inversely associated with the haplotype $Gm^{3,5,13,14}$ [$\chi^2 = 61.6$; df = 1; $P < .001$]. Prevalence ratio = 0.27 (95% confidence interval = 0.18–0.40).

	diabetes	control
Gm	23	270
no Gm	1343	3284

$Gm^{3,5,13,14}$ and Type 2 Diabetes Mellitus: An Association in American Indians with Genetic Admixture

William C. Knowler,* Robert C. Williams,†‡ David J. Pettitt,* and Arthur G. Steinberg§

Association mapping causal loci

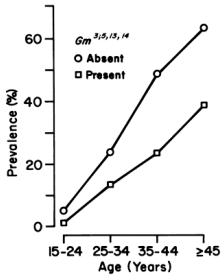


Figure 1 Prevalence of diabetes by age and the presence of the haplotype $Gm^{3,5,13,14}$ among residents of the Gila River Indian Community.

Self-identified heritage

Distribution of $Gm^{3,5,13,14}$ Haplotype Frequencies According to Indian Heritage in Residents of the Gila River Indian Community

No. of $Gm^{3,5,13,14}$ Haplotypes	Indian Heritage (Eighths) 0	1	2	3	4	5	6	7	8	Total (%)
0	11	0	4	19	199	4	72	123	4,195	4,627 (94.0)
1	14	0	8	4	144	0	27	13	68	278 (5.7)
2	7	0	6	0	1	0	0	0	1	15 (.3)
Total	32	0	18	23	344	4	99	136	4,264	4,920 (100.0)

Self-identified heritage

Most "full heritage" members don't have the haplotype

Distribution of $Gm^{3,5,13,14}$ Haplotype Frequencies According to Indian Heritage in Residents of the Gila River Indian Community

No. of $Gm^{3,5,13,14}$ Haplotypes	Indian Heritage (Eighths) 0	1	2	3	4	5	6	7	8	Total (%)
0	11	0	4	19	199	4	72	123	4,195	4,627 (94.0)
1	14	0	8	4	144	0	27	13	68	278 (5.7)
2	7	0	6	0	1	0	0	0	1	15 (.3)
Total	32	0	18	23	344	4	99	136	4,264	4,920 (100.0)

Self-identified heritage

Most "full heritage" members don't have the haplotype

Distribution of $Gm^{3,5,13,14}$ Haplotype Frequencies According to Indian Heritage in Residents of the Gila River Indian Community

No. of $Gm^{3,5,13,14}$ Haplotypes	Indian Heritage (Eighths) 0	1	2	3	4	5	6	7	8	Total (%)
0	11	0	4	19	199	4	72	123	4,195	4,627 (94.0)
1	14	0	8	4	144	0	27	13	68	278 (5.7)
2	7	0	6	0	1	0	0	0	1	15 (.3)
Total	32	0	18	23	344	4	99	136	4,264	4,920 (100.0)

The few without N.A. heritage are much more likely to have the haplotype

Gm haplotype is very rare in self-identified 100% Pima members.

Gm haplotype is very rare in self-identified 100% Pima members.

Gm is a marker for Caucasian ancestry.

Association and admixture

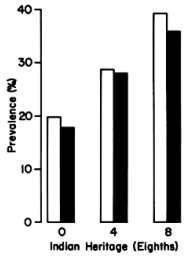


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype $Gm^{3,5,13,14}$, according to Indian heritage, among residents of the Gila River Indian Community.

Association and admixture

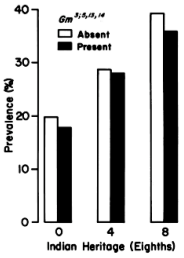


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype $Gm^{3,5,13,14}$, according to Indian heritage, among residents of the Gila River Indian Community.

Association and admixture

almost negligible

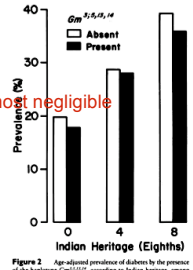


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype $Gm^{3,5,13,14}$, according to Indian heritage, among residents of the Gila River Indian Community.

huge

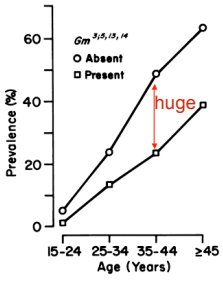


Figure 1 Prevalence of diabetes by age and the presence of the haplotype $Gm^{3,5,13,14}$ among residents of the Gila River Indian Community.

Association and admixture

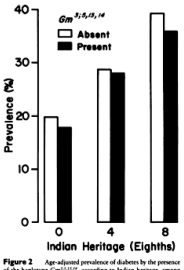


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype $Gm^{3,5,13,14}$, according to Indian heritage, among residents of the Gila River Indian Community.

these are all the Caucasians…

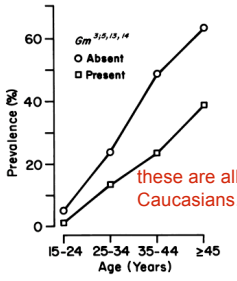


Figure 1 Prevalence of diabetes by age and the presence of the haplotype $Gm^{3,5,13,14}$ among residents of the Gila River Indian Community.

Association and admixture

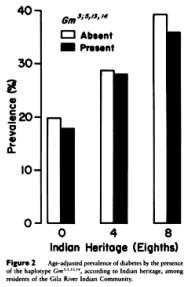


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype Gm, according to Indian heritage, among residents of the Gila River Indian Community.

Gm doesn't look like it has any additional protective effect if you stratify by familial origin first!

Association and admixture

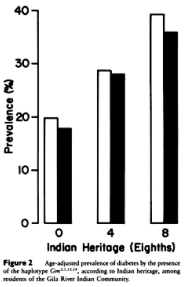


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype Gm, according to Indian heritage, among residents of the Gila River Indian Community.

Caucasian ancestry is associated with Gm haplotype.

Association and admixture

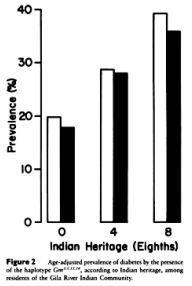


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype Gm, according to Indian heritage, among residents of the Gila River Indian Community.

Caucasian ancestry is associated with Gm haplotype.
Caucasian ancestry is associated with lower diabetes risk.

Association and admixture

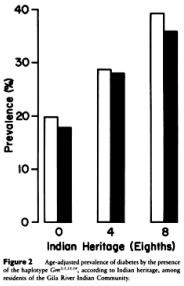


Figure 2 Age-adjusted prevalence of diabetes by the presence of the haplotype Gm, according to Indian heritage, among residents of the Gila River Indian Community.

Caucasian ancestry is associated with Gm haplotype.
Caucasian ancestry is associated with lower diabetes risk.
But Gm is not associated with lower diabetes risk!

In a genetically random sample

If disease is more prevalent in population A than B, will find more A's in cases than controls.

In a genetically random sample

If disease is more prevalent in population A than B, will find more A's in cases than controls.

Will find more A-specific alleles in cases than controls.

In a genetically random sample

If disease is more prevalent in population A than B, will find more A's in cases than controls.

Will find more A-specific alleles in cases than controls.

Will mistakenly conclude that these population-specific loci are causative for disease.