Opinion TRENDS in Genetics Vol.21 No.11 November 2005

An evolutionary framework for common diseases: the ancestral-susceptibility model

Anna Di Rienzo[1] and Richard R. Hudson[2]

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[2]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

Unlike rare mendelian diseases, which are due to new mutations (i.e. derived alleles), several alleles that increase the risk to common diseases are ancestral. Moreover, population genetics studies suggest that some derived alleles that protect against common diseases became advantageous recently. These observations can be explained within an evolutionary framework in which ancestral alleles reflect ancient adaptations to the life-style of ancient human populations, whereas the derived alleles were deleterious. However, with the shift in environment and lifestyle, the ancestral alleles now increase the risk of common diseases in modern populations. In this article, we develop an explicit evolutionary model and use population genetics simulations to investigate the expected haplotype structure and type of disease-association signals of ancestral risk alleles.

Shifting paradigms

The general paradigms that have enabled the identification of genes for rare mendelian diseases are continually being replaced in the search for genetic variation influencing common complex diseases. For example, the limited success of linkage analysis studies of complex traits led to the recognition that methods for detecting alleles with moderate to small relative risk were needed [1].

Another paradigm broadly accepted for mendelian diseases and, more or less explicitly, extended to common diseases is the notion that alleles that increase disease risk are new mutations. These new mutations, usually referred to as 'derived' alleles, can be inferred by comparing the allele observed at any given human polymorphic site with its orthologous nucleotide position in a close outgroup species (e.g. the chimpanzee). The allele found in chimpanzee is inferred to be ancestral and the alternative allele is derived. For mendelian diseases, usually caused by rare strongly deleterious mutations, a mutation-selection-balance model in which disease alleles are continuously generated by mutation and eliminated by purifying selection is plausible in most cases. This framework has been also used to model the genetic risk to common diseases based on the possibility that, although most common diseases have a late age of onset, the susceptibility genotypes can still have weakly deleterious effects on fitness [2,3]. This model predicts that multiple rare, new alleles increasing risk will be found at trait loci, as was recently observed at the *ABCA1* gene responsible for low HDL cholesterol [4]. Indeed, a model of weak purifying selection for patterns of variation in human populations is also supported by the observation that non-synonymous variants and, in particular those occurring at evolutionarily conserved positions, tend to occur at lower frequencies compared with silent polymorphisms [5–7].

Examples of ancestral-susceptibility alleles for common diseases

Several known or probable susceptibility polymorphisms for common diseases ostensibly depart from this model in that the allele increasing risk is the ancestral allele, whereas the derived allele is protective. Examples include variants involved in biological processes such as energy metabolism and sodium homeostasis. The ε4 allele of the gene encoding Apolipoprotein E (*APOE*), which increases the risk to coronary artery disease [8–10] and Alzheimer's disease [11,12], carries the ancestral allele at two common amino acid polymorphisms. Interestingly, a re-sequencing survey showed that the haplotype class defined by the derived allele ε3 was associated with an excess of singleton variants, consistent with an expansion of this haplotype driven by positive natural selection [13]. The distribution of the ε3 allele, the most common in all populations, and its functional properties independently suggested that this allele had evolved adaptively [14]. Within the same context, it was postulated that the ancestral ε4 allele used to be a favorable 'thrifty' allele in ancient populations, and had subsequently become detrimental under contemporary environmental conditions. For example, a comparison of the polymorphism P12A at the *PPARG* gene, which was shown to influence risk to type 2 diabetes [15], with its orthologous chimpanzee sequence reveals that the risk allele (P12) is ancestral, whereas the protective and less common allele is derived. A more recent example of a variant in which the ancestral allele increases risk to a metabolic syndrome is single nucleotide polymorphism (SNP) 44 at *CAPN10* (which encodes calpain 10). The ancestral allele at this site, which was significantly associated with type 2 diabetes risk in two meta-analysis studies [16,17], occurs at low frequencies in

Corresponding author: Di Rienzo, A. (dirienzo@genetics.uchicago.edu).
Available online 8 September 2005

 0168-9525/$ - see front matter doi:10.1016/j.tig.2005.08.007

worldwide populations (1–18%) [18]. A re-sequencing survey showed that the haplotype class carrying the derived allele at SNP44 has a significant deficit of polymorphism relative to expectations under neutrality [19]; this suggests that this derived allele is younger than expected for a neutral allele at the same frequency and that it reached its current high frequency as a result of positive natural selection. The expected time for a new neutral allele to reach a frequency of 80% is $3.6N_e$ generations (where N_e is the effective population size; i.e. 720 000 years); hence, the SNP44 derived allele must have arisen much more recently.

Variants involved in the control of sodium homeostasis have shown similar patterns: ancestral alleles increase risk to hypertension and the derived protective alleles appear to carry the signature of positive natural selection at tightly linked neutral sites. The angiotensinogen gene (*AGT*) carries two variants, a promoter A–6G variant and the T235M variant, which are associated with hypertension [20]. The alleles that increase risk to hypertension, A(−6) and T235, are fixed in several primate species [21] and are assumed to be ancestral. Recently, it was proposed that the derived G(−6) allele had quickly risen to high frequency as a result of positive natural selection [22]. Another example is the *CYP3A5*3* variant, which results in a non-functional protein [23]. The ancestral functional allele was reported to be associated with increased systolic blood pressure and mean arterial pressure in African Americans [24]. The proposed mechanism involves the conversion of cortisol to 6-β-hydroxycortisol by cytochrome P450 (CYP) 3A5 in the kidney, leading to higher reabsorption of sodium and water retention [25,26]. A recent survey of sequence variation in this gene revealed that the derived protective allele (*CYP3A5*3*) is associated with low levels of haplotypic variation, which is expected if recent positive selection had quickly driven it to high frequency [27].

Several additional examples of ancestral alleles proposed to increase risk to common diseases or disease-related phenotypes are listed in Table 1; these variants show a broad range of allele frequencies and between-population differentiation. Further population genetics studies will be necessary to determine how many derived alleles carry the signature of positive natural selection.

Models for the evolution of genetic susceptibility to (some) common diseases

It is possible that the recent increase in life expectancy uncovered the latent genetic susceptibility to diseases with post-reproductive age of onset. Hence, ancestral disease risk variants might be expected not to have had fitness consequences; this scenario, in which ancestral or derived alleles are equally likely to increase disease risk, might easily fit within the common disease–common variant hypothesis. However, the evidence for natural selection favoring derived protective alleles might require more complex models for the evolution of the genes influencing the susceptibility to some common diseases. More specifically, it suggests not only that these genes did not evolve neutrally, but also that the environmental pressures acting on them changed during human evolution. Indeed, it was hypothesized [28,29] that disease-susceptibility genotypes conferred a selective advantage in ancestral human populations.

In its simplest incarnation, the thrifty genotype hypothesis [28] posits that the genetic predisposition to type 2 diabetes is the consequence of metabolic adaptations to an ancient lifestyle characterized by fluctuating and unpredictable food supply and high levels of physical activity. With the switch to a sedentary lifestyle and energy-dense diets, the thrifty genotype is no longer advantageous and gives rise to disease phenotypes, such as type 2 diabetes and obesity. An analogous evolutionary framework, sometimes referred to as the sodium-retention hypothesis, was proposed to explain the increased prevalence of essential hypertension in some ethnic groups [29]. Briefly, ancient human populations living in hot, humid areas adapted to their environment by

Table 1. A few examples of reported ancestral risk variants[a]

Gene	Record number in dbSNP[b]	Mutation	Amino acid change	Frequency of ancestral allele				Phenotypes
				African	European	Asian	Source	
ACE		D→I		0.648	0.587	0.377	ALFRED	Hypertension
ADD1	4961	G→T	G460W	0.891	0.833	0.521	Perlegen	Hypertension
ADRB2	1 042 714	G→C	E27Q	0.196	0.458	0.083	Perlegen	Obesity, asthma
ADRB3	4994	C→T	R64W	0.104	0.146	0.146	SNP500	BMI, obesity, type 2 diabetes
AGT	699	C→T	T235M	0.937	0.450	0.793	Ref. [27]	Hypertension, pre-eclampsia
	4762	C→T	T174M	0.952	0.875	0.795	Perlegen	Hypertension
	5051	A→G		0.899	0.530	0.774	Ref. [22]	Hypertension
APOE	429 358	C→T	R130C	0.017	NA	0	HapMap	Alzheimer's disease, coronary artery disease
CAPN10	2 975 760	C→T		0.020	0.141	0.107	Ref. [18]	Type 2 diabetes
CYP3A5	776 746	A→G		0.859	0.090	0.229	Ref. [27]	Hypertension
ENPP1	1 044 498	C→A	Q121K	0.227	0.167	0.042	Perlegen	Insulin resistance
GAD2	2 236 418	G→A		0.891	0.167	0.333	Perlegen	Obesity
PPARG	1 801 282	C→G	P12A	0.957	0.938	0.938	Perlegen	Type 2 diabetes
SCNN1G	5718	G→A		NA	NA	NA		Hypertension
SLC2A2	5400	T→C	I110T	0.492	0.142	0.022	HapMap	Type 2 diabetes
USF1	3 737 787	C→T		0.913	0.729	0.875	Perlegen	Familial combined hyperlipidemia
	2 073 658	G→A		0.913	0.729	0.875	Perlegen	Familial combined hyperlipidemia

[a]Abbreviation: NA, not available.
[b]For more details on dbSNP, see http://www.ncbi.nlm.nih.gov/SNP/.

retaining salt, whereas populations in cooler, temperate climates adapted to conditions of higher sodium levels. This hypothesis recently received support from the finding of a latitudinal cline of allele frequency for variants likely to influence sodium homeostasis and/or hypertension, including those in *AGT* [27], *CYP3A5* [27], angiotensin I converting enzyme (*ACE*), *ENAC*α (also known as *SCNN1A*) and *ENAC*γ, (also known as *SCNN1G*), which encode sodium channels (J.H. Young, personal communication).

What these two hypotheses have in common is a radical and relatively recent change in the selective pressures acting on biological processes responsible for maintaining the correct balance between the organism and its environment. The recent environmental change disrupts this balance leading, in turn, to new detrimental phenotypes. Thus, these hypotheses, originally based only on disease physiology and epidemiology, can be translated into testable population genetics models of disease susceptibility. One such model could envision that the ancestral versions of genes that affect susceptibility to common diseases today reflect ancient adaptations. In early human populations, these alleles were maintained by purifying selection, whereas weakly deleterious derived alleles were prevented from reaching intermediate frequency, similar to the rare amino acid polymorphisms observed at the genome-wide level [5,6]. With the switch in lifestyle and environmental conditions, the ancestral alleles no longer confer a selective advantage and increase disease risk, whereas some derived alleles protect against disease and become either neutral or advantageous. Whether the selective advantage inferred for some of the derived variants is the direct consequence of disease protection or results from pleiotropic effects is currently a matter of speculation [27].

Comparing patterns of LD in derived and ancestral alleles

The shift from linkage mapping of mendelian diseases to association mapping of complex diseases has focused attention on the decay of linkage disequilibrium (LD) around risk alleles and on their haplotype structure as a prerequisite for the design of optimal association studies. Some investigators have turned to coalescent simulations of evolutionary neutrality as a means to develop expectations for the extent of LD decay and haplotype structure of disease alleles [30]. Extrapolating from mendelian diseases, these simulation studies considered only the case in which the allele that increases disease risk is derived. However, the haplotype structure and the pattern of LD are expected to differ substantially for ancestral and derived alleles under neutrality, and even more if selection acted on derived alleles. This is because each mutation event occurs, by necessity, on a particular haplotypic background. Hence, derived alleles (especially recent or low-frequency alleles) will tend to be associated with a subset of all haplotypes. Conversely, because all extant variation arose on the ancestral haplotype background, the ancestral allele will tend to be distributed across a larger repertoire of haplotypes.

We performed population genetics simulations to quantify the difference in the haplotype diversity and the number of haplotypes associated with ancestral and derived alleles at any given frequency. A complete haplotype map of the human genome will soon be available [31], therefore, many disease mapping studies will be designed to detect haplotype associations using haplotype tag SNPs, rather than single-site associations. Thus, the diversity and the number of haplotypes carrying a particular risk allele are informative summaries for comparing ancestral and derived alleles. Coalescent simulations were used to model evolutionary neutrality for all variation throughout the history of the population. In addition, to model the ancestral-susceptibility scenario outlined above, we performed forward simulations of a mutation-selection-balance model in which selective pressures change at a point in the past so that a portion of derived deleterious alleles become advantageous. As expected, the difference in haplotype diversity and number of haplotypes associated with ancestral versus derived alleles is much greater in the model with a shift in selective pressures compared with the neutral model (Figure 1 in the supplementary material online).

It was previously shown that a model of mutation-selection balance in a long-term steady environment can lead to increased allelic heterogeneity [2]. However, our simulations of the ancestral-susceptibility model suggest that, if selective pressures change so that a portion of previously deleterious mutations becomes advantageous, the expected allelic heterogeneity is much lower. More specifically, for the parameter values used in our simulations, the average number of such (protective) alleles is 13.6 immediately after the change in selective pressures. However, by the time their overall frequency reaches 10% or more, the average number of these alleles is only two. A more thorough exploration of the parameter space is currently under way to determine how broadly this observation applies to the ancestral-susceptibility model.

Implications for association studies of common diseases

If at least one SNP that is in perfect, or near-perfect, LD with the causative variant is typed, and the analysis is performed using information from single SNPs rather than haplotypes, an unambiguous association signal is expected regardless of whether the risk allele is ancestral or derived (we refer to an unambiguous association signal as the excess of an allele or haplotype in cases and of the alternative in controls). However, if information from multi-allelic systems (i.e. microsatellites or haplotypes) is used, the signal can be more complex. This is because, compared with a derived allele at the same frequency, the ancestral allele will tend to be distributed over a larger number of haplotypes none of which could be sufficiently over-represented in cases to yield a significant signal. If the derived allele protects against the disease, this might lead to an excess of one haplotype in controls without a concomitant excess of another haplotype in cases. In the context of a transmission disequilibrium test (TDT), this would correspond to the under-transmission of one haplotype to cases without an over-transmission of another haplotype. Note, however, that a significant test result can still be generated by either test.

To illustrate this point and quantify this effect, we performed another set of forward-population-genetics simulations in which a single deleterious variant becomes advantageous after the change in selective pressures. Therefore, we simulated a contiguous sequence segment in which a short terminal portion (meant to represent the trait locus) experienced temporally varying selective pressures while the remainder (i.e. the marker segment) evolved neutrally. For each simulated sample, the phenotypes were randomly assigned based on the genotypes at the trait site and a set of corresponding penetrance values; in all examples, the derived allele at the trait site was assumed to be protective. To investigate the expected association signals for haplotype-based tests, the full frequency distribution of marker haplotypes was reduced to a 2×2 contingency table, in which each chromosome in cases and controls was classified as either carrying a specified haplotype or not; this was repeated for each marker haplotype present at sufficient frequency (i.e. 20 times). For each contingency table, the haplotype counts in cases and controls were calculated and a χ^2 test performed. To compare the results of haplotype-based and single-site tests, the contingency table χ^2 test was also performed for the trait variant in addition to a variant in LD with the trait variant. The ultimate goal of this exercise was to illustrate quantitatively the types of association signal expected under the assumption of an ancestral risk allele. For the haplotype-based tests, we asked how often a significant test result is associated with an excess of different marker haplotypes in cases and in controls compared with how often it is associated with an excess only in cases or only in controls. The same procedure was repeated using coalescent simulations of evolutionary neutrality and forward simulations in which deleterious alleles become advantageous with probability 0.5 (Tables 1 and 2 in the supplementary material online).

Although the results shown in Table 2 are not meant to be exhaustive with regard to the full spectrum of disease models, they suggest some general conclusions. For relatively high frequencies of the ancestral risk allele (0.8–0.9, as in the *PPARG* variant), a significant test result will often be associated with an excess of at least one marker haplotype only in controls, without an excess of another haplotype in cases. For the examples in which the ancestral risk allele is less common (0.2–0.3 and 0.1–0.2, as in the *CAPN10* and *APOE* variants), a marker haplotype excess in cases and controls is the most frequent outcome.

The observation of a marker haplotype excess in controls in the absence of an excess in cases is usually interpreted as being due to the presence of a protective allele. However, the risk allele counterpart can remain elusive (i.e. no single allele is over-represented in cases). Under our model, a possible follow-up analysis could focus on the SNPs with ancestral alleles in perfect or near-perfect LD with the haplotype over-represented in controls. At these SNPs, the ancestral states represent candidate risk alleles that can be further prioritized for single-site association tests or functional *in vitro* studies.

Concluding remarks

Certainly, ancestral alleles represent only a subset of all risk variants. However, the number of such variants that have already been proposed and, in some examples, replicated is sufficiently large to conclude that they account for a non-trivial portion of all genetic susceptibility to common diseases. Furthermore, in this model, protective alleles are expected to generate more clear association signals than risk alleles. Indeed, several examples of derived protective alleles have recently been identified (e.g. Boutin *et al.* [32] and Pajukanta *et al.* [33]).

It is evident that there is a diverse spectrum of plausible evolutionary models for the genetic bases of complex traits and common diseases. Importantly, a single evolutionary model is unlikely to account for the genetic susceptibility to all common diseases. Furthermore, different risk variants for a single common disease – even within the same gene – might have followed different evolutionary trajectories.

Different evolutionary models have different implications for the frequency spectrum of disease alleles and their geographic distribution, which in turn determine the optimal strategies for finding susceptibility variants. For

Table 2. Association signals in cases and controls simulated using the model with a shift in selective pressures[a]

D[b] allele frequency	Prevalence	Penetrances (genotype relative risk)			Haplotype-based tests				Single-site tests	
					Proportion of replicates in which at least one haplotype is in excess with $\chi^2>3.84$			Number of replicates	Proportion of replicates with $\chi^2>3.84$	
		AA[b]	AD[b]	DD[b]	Controls and cases[c]	Controls only[c]	Cases only[c]		Trait variant (number of replicates)	Marker in LD[d] with trait variant (number of replicates)
0.1–0.2	10.8%	0.12 (2)	0.08 (1.3)	0.06 (1)	0.26	0.29	0.05	117	0.67 (138)	0.39 (18)
0.1–0.2	11.1%	0.12 (2)	0.09 (1.5)	0.06 (1)	0.20	0.23	0.06	111	0.37 (123)	NA[e]
0.1–0.2	7.4%	0.08 (2)	0.06 (1.5)	0.04 (1)	0.15	0.18	0.16	102	0.35 (136)	NA[e]
0.1–0.2	6.9%	0.08 (2)	0.04 (1)	0.04 (1)	0.34	0.25	0.04	106	0.97 (105)	NA[e]
0.1–0.2	7.9%	0.08 (2)	0.08 (2)	0.04 (1)	0.03	0.12	0.11	130	0.08 (104)	NA[e]
0.7–0.8	7.1%	0.12 (2)	0.08 (1.3)	0.06 (1)	0.61	0.22	0.09	131	0.92 (138)	0.40 (45)
0.8–0.9	6.6%	0.12 (2)	0.08 (1.3)	0.06 (1)	0.53	0.03	0.03	117	0.82 (138)	0.49 (47)

[a]In these simulations only one deleterious variant becomes advantageous after the environmental change.
[b]D denotes the derived protective allele and A the ancestral risk allele.
[c]In all replicates, 400 cases and 400 unaffected controls were compared.
[d]$0.75\leq r^2\leq0.85$.
[e]Not available (NA). Fewer than ten replicates contained a marker with the specified r^2 value with the trait variant.

example, the predictions of the ancestral-susceptibility model proposed here are clearly different from those of the mutation-selection-balance model in which disease variants are derived and slightly deleterious [2,3]. Although a thorough investigation of the ancestral-susceptibility model has not been performed yet, it seems possible that it will result in relatively common susceptibility alleles. Indeed, our simulations suggest a limited allelic heterogeneity for those deleterious variants that become advantageous following the environmental change. Consequently, standard association studies, especially if aimed at detecting protective alleles, might be powerful. In addition, a variety of approaches based on detecting the signature of positive natural selection might offer an appealing independent framework for honing in on disease variants. The action of natural selection on disease variants can generate detectable patterns against the genome-wide background of neutrally evolving loci; these patterns can be used to identify interesting variants. For example, depending on the geographic distribution of the selective pressures acting on risk variants, their allele frequencies can differ across human populations more than expected for neutrally evolving loci [18,34,35]. Similarly, disease variants can show specific geographic patterns that mirror the spatially varying selective pressures that acted on them. For example, a latitudinal cline of allele frequency was shown for variants involved in sodium homeostasis and hypertension [27].

Importantly, the selective pressures entailed in the ancestral-susceptibility model did not target the disease phenotype *per se*, but rather phenotypes resulting from normal biological processes, such as energy metabolism or sodium homeostasis. This raises the possibility that the same model might apply to other biological processes and corresponding traits, such as height, endurance, cognition and drug response, which might have been exposed to changing selective pressures during human evolutionary history.

In summary, a wide range of evolutionary scenarios can underlie the genetic susceptibility to human common diseases and complex phenotypes. Furthermore, multiple models are likely to apply to the same phenotype and even to different risk variants within the same gene. Our simulation results suggest that the interpretation of disease mapping signals can be facilitated if the genetic susceptibility to common diseases is considered within an evolutionary framework.

Acknowledgements

We thank V. Clark for help with data base searches, M. Abney, N. Freimer, C. Gilliam, D. Nicolae, C. Ober and J. Pritchard for comments on the manuscript and, especially, N. Cox for many helpful discussions. We are grateful to L. Kruglyak for sharing information regarding his simulations. This research was partially supported by NIH grant DK56670.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2005.08.007

References

1 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* 273, 1516–1517
2 Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137
3 Reich, D.E. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet.* 17, 502–510
4 Cohen, J.C. *et al*. (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869–872
5 Cargill, M. *et al*. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231–238
6 Halushka, M.K. *et al*. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* 22, 239–247
7 Leabman, M.K. *et al*. (2003) Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5896–5901
8 Stengard, J.H. *et al*. (1995) Apolipoprotein E polymorphism predicts death from coronary heart disease in a longitudinal study of elderly Finnish men. *Circulation* 91, 265–269
9 de Knijff, P. *et al*. (1994) Genetic heterogeneity of apolipoprotein E and its influence on plasma lipid and lipoprotein levels. *Hum. Mutat.* 4, 178–194
10 Davignon, J. *et al*. (1988) Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 8, 1–21
11 Corder, E.H. *et al*. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923
12 Strittmatter, W.J. *et al*. (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1977–1981
13 Fullerton, S.M. *et al*. (2000) Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am. J. Hum. Genet.* 67, 881–900
14 Corbo, R.M. and Scacchi, R. (1999) Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? *Ann. Hum. Genet.* 63, 301–310
15 Altshuler, D. *et al*. (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* 26, 76–80
16 Song, Y. *et al*. (2004) Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am. J. Hum. Genet.* 74, 208–222
17 Weedon, M.N. *et al*. (2003) Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am. J. Hum. Genet.* 73, 1208–1212
18 Fullerton, S.M. *et al*. (2002) Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am. J. Hum. Genet.* 70, 1096–1106
19 Vander Molen, J. *et al*. (2005) Population genetics of *CAPN10* and *GPR35*: implications for the evolution of type 2 diabetes variants. *Am. J. Hum. Genet.* 76, 548–560
20 Wu, S.J. *et al*. (2004) Three single-nucleotide polymorphisms of the angiotensinogen gene and susceptibility to hypertension: single locus genotype vs. haplotype analysis. *Physiol. Genomics* 17, 79–86
21 Inoue, I. *et al*. (1997) A nucleotide substitution in the promoter of human angiotensinogen is associated with essential hypertension and affects basal transcription *in vitro*. *J. Clin. Invest.* 99, 1786–1797
22 Nakajima, T. *et al*. (2004) Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am. J. Hum. Genet.* 74, 898–916
23 Kuehl, P. *et al*. (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat. Genet.* 27, 383–391
24 Givens, R.C. *et al*. (2003) CYP3A5 genotype predicts renal CYP3A activity and blood pressure in healthy adults. *J. Appl. Physiol.* 95, 1297–1300
25 Ghosh, S.S. *et al*. (1995) Renal and hepatic family 3A cytochromes P450 (CYP3A) in spontaneously hypertensive rats. *Biochem. Pharmacol.* 50, 49–54
26 Schuetz, E.G. *et al*. (1992) Expression of cytochrome P450 3A in amphibian, rat, and human kidney. *Arch. Biochem. Biophys.* 294, 206–214

27 Thompson, E.E. *et al*. (2004) CYP3A variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.* 75, 1059–1069
28 Neel, J.V. (1962) Diabetes Mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am. J. Hum. Genet.* 14, 353–362
29 Gleibermann, L. (1973) Blood pressure and dietary salt in human populations. *Ecol. Food Nutr.* 2, 143–156
30 Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22, 139–144
31 The International HapMap Consortium. (2003) The International HapMap Project. *Nature* 426, 789–796
32 Boutin, P. *et al*. (2003) GAD2 on chromosome 10p12 is a candidate gene for human obesity. *PLoS Biol.* 1, E68
33 Pajukanta, P. *et al*. (2004) Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat. Genet.* 36, 371–376
34 Akey, J.M. *et al*. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2, E286
35 Bersaglieri, T. *et al*. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74, 1111–1120