UNIVERSITY OF BOLOGNA

Ph. D. Program in
" Biodiversity and Evolution"
(XIX cycle)
BIO/08 - Anthropology

HUMAN GENETIC VARIABILITY OF TWO GENES INVOLVED IN IRON HOMEOSTASIS, HAPTOGLOBIN AND HEPCIDIN, AND IN A GENELESS REGION OF CHROMOSOME 22

Paolo Garagnani

Coordinator:

Prof. Giovanni Cristofolini

Supervisors:

Prof. Donata Luiselli

Dr. Vilma Mantovani

Academic Dissertation, 2007

INTRODUCTION

The latest human genome data reveals that the total number of genes for our species is around 30,000, far fewer than previously predicted. The data confirm that humans of all ethnic backgrounds share 99.99 percent of the same genetic code. Individual variations comprise only 0.01 percent of the entire sequence, which includes 2.91 billion base pairs.

So far it appears that genes tend to cluster along just a few of the chromosomes (e.g. 17, 19, 22) while others appear quite barren. Researchers have long been puzzled by that fact that nearly one quarter of the human genome contains sequences that lack protein-coding genes, so-called junk DNA. About one-fourth of the genome could be considered deserts, with lengthy gene-free segments. More than a third of the genome (35.3 percent) contains repetitive sequences, suggesting that this so-called "junk DNA" deserves further study. In addition to repeated segments, there is also renewed interest in the 2.1 million scraps of DNA known as SNPs, or single nucleotide polymorphisms. While the new data suggest most of these are harmless, some may underlie certain disease processes.

The discovery that the human genome may only contain around 30,000 genes is one of the most surprising findings of the world wide research effort. The good news is that this may make it easier to identify the functions of the genes, any of which could be promising targets for drug development. However researchers are already moving away from the 'one gene- one protein' idea to the study of complex interactions between genes that appear to have different functions in different contexts. Now that the basic structure of the genome has been identified, it opens the way for an even more complex area of study functional genomics, dedicated to finding out the function of single and multiple interactive genes.

Genetic variation in the human genome

The completion of the initial draft of the human genome sequence (38) provided the first in-depth description of the different components of the human genome (table1) and considerable information about its gene content and gene positions. The genome sequence has been used extensively as a starting point for many subsequent studies

focusing on the function of genes or the functional elements of particular loci linked to disease. Recently, human genetics shifted to large-scale studies of human genome variation. The large-scale polymorphism data that are made available can now be used to efficiently study genetic differences among humans.

The post-genomic research in human and population genetics is centred around the study of the patterns of genetic variation across the genome and among populations with the aim of understanding the origin of complex disease and the evolutionary history of our species.

Variation in the human genome sequence plays a powerful but poorly understood role in the etiology of common medical conditions. Because the vast majority of heterozygosity in the human population is attributable to common variants and because the evolutionary history of common human diseases (which determined the allele spectrum for causal alleles) is not yet known, one promising approach is to comprehensively test common genetic variation for association to medical conditions (1–3). This approach is increasingly practical because 4 million (4, 5) of the estimated 10 million (6) common single nucleotide polymorphisms (SNPs) are already known.

Single nucleotide polymorphisms (SNPs) most often result from base substitutions occurring through a nonrepaired error that occurs during DNA replication; other causes of DNA polymorphisms include gene conversion and duplication.

The mutation rate at most positions of the human genome is relatively low (on the order of 10_8 substitution per base pair per generation) compared with the most recent common ancestor of any two individuals (on the order of 10,000 generations). Therefore, the vast majority of allelic differences between individuals are inherited rather than de novo mutations. Two individuals sharing the same allele at one position are most likely identical by descent (for this specific portion of the DNA) rather than through two independent mutations (i.e., homoplasy).

After a mutant allele is introduced in the gene pool, its fate is determined by the interaction of two evolutionary forces: random genetic drift and natural selection. Drift affects the distribution of a SNP in the population by random sampling of different alleles at each generation (only a small fraction of all possible gametes are transmitted to the next generation). If only drift acts on a particular region of the genome, a SNP frequency varies randomly from generation to generation until one allele eventually reaches fixation (either 100% or 0%), and the time before fixation occurs is mainly determined by the population size. In other words, if only genetic drift acts,

polymorphisms are only transient products of random fluctuations. However, natural selection affects the probability that a particular variant is passed to the next generation. It can either increase the probability and speed of fixation of an newly arisen allele if the mutant allele confers a fitness advantage (i.e., selective sweep or positive selection), remove new deleterious variants from the gene pool (i.e., negative selection), or maintain several alleles in the gene pool over extended periods of time (i.e., balancing selection).

Table 1. Genetic variation in the human genome

Variation type	Definition	Frequency (if known) in the human genome
SNP	Single base pair variation found in >1% of chromosomes in a given population	~10 million SNPs in the human population
Insertion/deletion variant (InDel)	Deletion or insertion of a segment of DNA. Includes small polymorphic changes and large chromosomal aberrations. InDels >1 kb in size are often also called CNVs	~1 million insertion/deletions polymorphisms >1 bp in the human genome
Microsatellite (e.g. CA_n repeats)	Sequences containing variable numbers of 1–6 bp repeats totaling <200 bp in length	>1 million microsatellites in the human genome, accounting for ~3% of the sequence
Minisatellite and variable numbers of tandem repeats (VNTRs)	Polymorphic sequence containing 20–50 copies of 6–100 bp repeats	~150 000 minisatellites, of which ~20% are polymorphic
Multisite variant (MSV)	Single nucleotide variant with complex characteristics due to CNV or gene conversion	The number of MSVs is currently unknown
Intermediate-sized structural variant (ISV)	Gain or loss of a DNA sequence >8 kb in size also includes inversion breakpoints	297 ISVs were identified using a fosmid library from a single genome
CNV; copy number polymorphism (CNP); large-scale CNV	Copy number change >1 kb. If the frequency is >1%, it is called a CNP. LCVs are CNVs ~50 kb in size or greater	The frequency of CNVs in the human genome is unknown. Estimates of larger CNVs (>50 kb)
Inversion	Rearrangement causing a segment of DNA to be present in reverse orientation	Estimates of microscopically detectable inversion frequencies are 0.12–0.7% (pericentric) and 0.1–0.5% (paracentric); sub-microscopic unknown
Translocation	Rearrangement in which a DNA fragment is attached to different chromosome	1/500 is heterozygous for a reciprocal translocation and 1/1000 for Robertsonian translocations
Unbalanced rearrangements	Rearrangements which lead to a net gain or loss of DNA are referred to as unbalanced	Unbalanced rearrangements occur in ~1/1500 live births

da Feuk et al., 2006

Whereas mutations are generated essentially randomly in DNA molecules and random genetic drift affects chromosomes as a whole, selection, in recombining genomes such as humans, acts differently on distinct regions of the genome. Consequently, the level of polymorphism varies greatly between different portions of the genome. For example, regions of noncoding DNA (introns and intergenic regions) typically harbor a much higher diversity than coding DNA (25), regulatory elements (such as promoters or

splicing sites), or conserved nongenic elements (2, 17). Similarly, within exons, synonymous and nonsynonymous positions differ in their diversity, as do, at a lower scale, twofold from fourfold degenerated sites (i.e., nucleotides that can be changed into, respectively, one or any other nucleotide and still code for the same amino acid) (25). Diversity also varies along the genome over several Megabases, even after correcting for gene or Guanine-Cytosine (GC) content (27). This large-scale variation probably represents differences in recombination rates that affect the frequency of SNPs (either directly through a putative mutagenic effect or indirectly via background selection) (27). If one considers the distribution of genetic diversity, not across regions of the genome, but across individuals, the frequencies of DNA polymorphisms also differ greatly. For one locus this is often summarized by the frequency spectrum that displays the frequency of all SNPs in a population. In a random mating population of constant size and without selection acting at the locus considered, most SNPs are expected to be present at low frequency and only found in one or few individuals, whereas very few SNPs will be common (Figure 1).

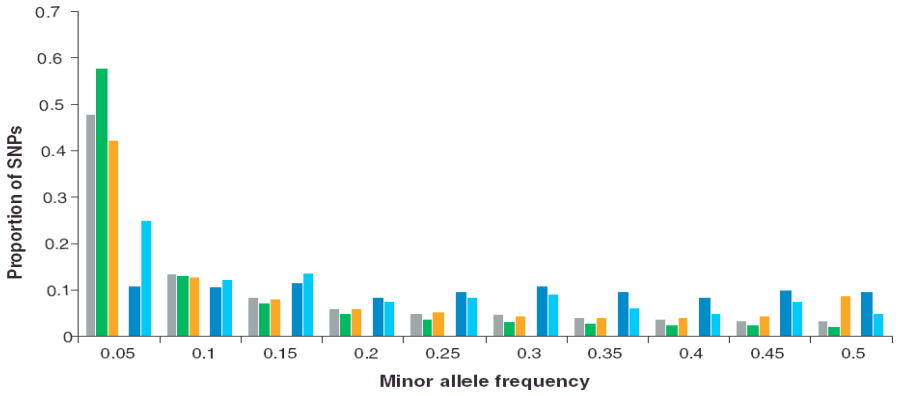


Fig. 1

This figure shows the distribution of polymorphic sites according to their minor allele frequency (on the x axis). The expected distribution under the standard neutral model (i.e., a pan-mictic population of constant size in which genetic variation is not affected by natural selection) is displayed in gray. The distributions of single nucleotide polymorphisms under a model of population growth and population substructure are displayed respectively, in green and orange. The observed distributions in the HapMap and ENCODE data set are shown in deep blue and light blue (for the individuals of European ancestry), respectively.

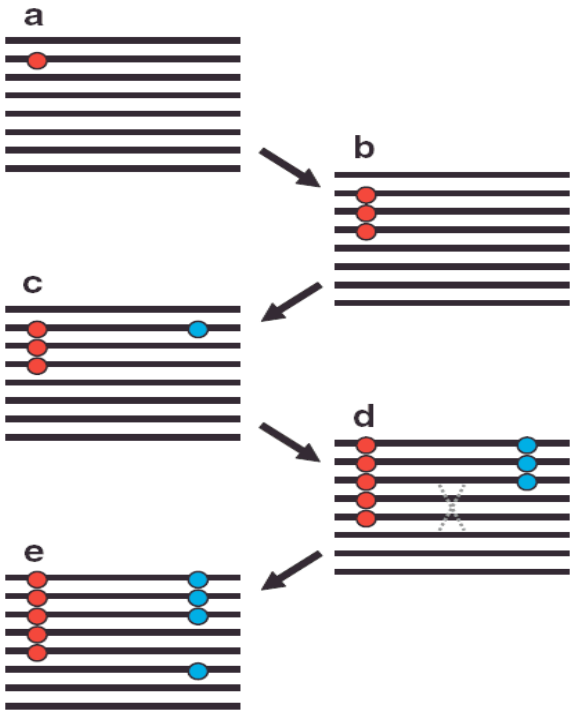
Deviations from this pattern may reflect violation of either the neutrality of the loci or of the demographic assumptions. For example, population growth and positive selection increase the proportion of rare alleles (i.e., alleles with low frequency), whereas balancing selection and population substructure increase the proportion of intermediate alleles (Figure 1). Analyzing the frequency spectrum is one of the most common tools in population genetics, and many tests of neutrality rely on it directly or indirectly. Most of the recent large-scale projects that generated genome-wide polymorphism data used genotyping of selected SNPs (rather than more expensive and less high-throughput resequencing), which introduces a skew in the frequency spectrum, especially if the project focuses on intermediate frequency polymorphisms, as in the HapMap project (Figure 1). Frequency spectra obtained from such data can be adjusted to take into account the ascertainment bias (see, e.g., 49, 50), but a reliable reconstruction of the evolutionary history of a particular locus currently still requires resequencing data that produce unbiased diversity estimates and allow more rigorous testing of neutrality.

Recombination, Haplotypes and Haplotype Blocks

One of the most amazing developments in our understanding of the organization of genetic diversity concerns the phenomenon of recombination and its consequences on the allelic correlations between neighboring SNPs. Recombination refers to crossovers and the exchange of chromatid fragments during meiosis.

In designing and interpreting association studies of genotype and phenotype, it is necessary to understand the structure of haplotypes in the human genome. Haplotypes are the particular combinations of alleles observed in a population. When a new mutation arises, it does so on a specific chromosomal haplotype. The association between each mutant allele and its ancestral haplotype is disrupted only by mutation and recombination in subsequent generations.

Thus, it should be possible to track each variant allele in the population by identifying (through the use of anonymous genetic markers) the particular ancestral segment on which it arose.



Effect of genetic drift and recombination on haplotypes. (*a*) A polymorphism is introduced by a mutation in a population of identical chromosomes. (*b*) The haplotype carrying the red allele increases in frequency in the population by genetic drift. (*c*) A second mutation, occurring on a haplotype carrying the red allele, introduces a new polymorphism. (*d*) The haplotype carrying both derived alleles increases in frequency in the population by genetic drift. At this point all chromosomes carrying the blue allele at the second polymorphic site also carry the red allele at the first site [resulting in high linkage disequilibrium (LD)]. (*e*) A crossover involving two different haplotypes (*red–blue* and ancestral) occurs between the two single nucleotide polymorphisms, leading to a reassortment of the alleles: The blue allele no longer occurs exclusively on haplotypes carrying a red allele (reducing the LD).

Fig. 2

Because each position can be hit only once by a mutation, one should end up with three different possible combinations if no haplotype is lost by genetic drift or eliminated by selection (Figure 2d). This nonrandom association of alleles (on the figure the blue allele is always associated with the red) is referred to as linkage disequilibrium (LD). If a crossover occurs between the two polymorphic positions, it may generate a fourth haplotype and break down LD between the two markers (Figure 2e). Interestingly, recombination does not occur evenly but happens much more frequently in a small fraction of the genome. These hot spots of recombination (typically 1–2 kb long) are frequent across the genome (roughly 1 every 50 kb) and account for more than 80% of

all recombination (15, 42, 46). Hot spots of recombination appear to be highly dynamic and at least some of them differ in intensity and/or location between humans and chimpanzees as well as possibly between human populations (15, 58, 59, 76). This may influence the design of association studies and the choice of markers for studying a given population, but because recombination rates seem conserved among human populations over a larger scale (>1 Mb), variations in recombination pattern among humans are unlikely to affect linkage studies (34, 64).

A popular and simple way to describe patterns of genetic diversity is to partition chromosomes into “blocks” of SNPs in high LD with each other (23). This usually allows summarizing polymorphism diversity into three to five common haplotypes that account for most of the sample variability. However, one should keep in mind that the definition and particularly the limits of these entities are highly arbitrary, as the resulting block depend both on the SNP density and on the number of individuals studied. These “blocks” should not be a priori considered as stretches of DNA sequence where no recombination occurs, separated from each other by recombination hot spots (even if this will be true in some instances). Note that the LD patterns observed in a sample result both from recombination events and the population history of this sample (56, 74, 75).

Haplotype blocks refer to sites of closely located SNPs which are inherited in blocks. Regions corresponding to blocks have a few common haplotypes which account for a large proportion of chromosomes. Identification of haplotype blocks is a way of examining the extent of LD in the genome, which generally provides useful information for the planning of association studies. The aim is to identify a minimal subset of SNPs that can characterize the most common haplotypes. In each block, a small fraction of single-nucleotide polymorphisms (SNPs), referred to as "tag SNPs," can be used to distinguish a large fraction of the haplotypes. These tag SNPs can potentially be extremely useful for association studies, in that it may not be necessary to genotype all SNPs.

The "Hap-Map" project (http://www.hapmap.org/index.html.en) describes haplotype blocks in the human genome.

The rationale of such major enterprise is that this information will make it possible to conduct disease gene association studies more quickly and efficiently than ever before, resulting in the more rapid elucidation of the variants that predispose to disease and influence drug response. Recent surveys of haplotype diversity in a number of genomic segments, and up to a chromosome, support the notion of haplotype "blocks" of limited diversity. While it is not clear how these finding will translate to other regions or populations, it seems imminent that large studies to elucidate the patterns of genomic variation will be carried out.

With the imminent flood of data from the Haplotype Map project, there is an urgent need to extend haplotype-related computational methods to the whole-genome scale. There are many fundamental computational and statistical problems to solve: how to calculate genome-wide haplotypes and determine haplotype block boundaries on sequence, how to find the optimal number of SNPs per haplotype block, how to account for the variation in LD across different populations, how to find the most informative SNPs for each haplotype block, how to deal with uncertainty in annotations and map locations, development of controls for spurious statistical association, the effects of population substructure, estimation of statistical significance, effects of genotyping error, how to manage and represent the large amount of data, methods for modeling the genotyping process in silico to avoid artifacts and failures in the massive genotyping projects, etc. Finally, methodologies to use the haplotype information for more efficient design and analysis of whole-genome association studies are in their infancy and deserve attention.

Gabriel et al. (2002) examines whether there is similar haplotype block-structure between and within populations (Nigeria/Yoruba, Asia, African Americans, Europeans), Fig.3.

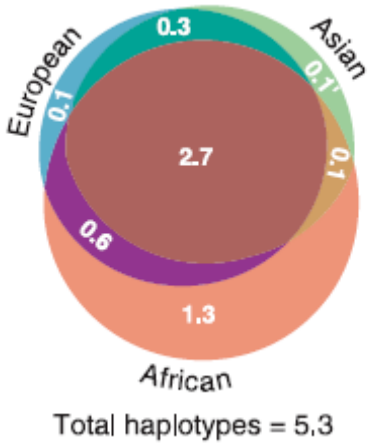


Fig. 3- Distribution of haplotypes across populations (Gabriel et al., 2002)

Haplotype methods have contributed to the identification of genes for Mendelian diseases (7–9) and, recently, disorders that are both common and complex in inheritance (10– 12).

Many studies have examined allelic associations [also termed “linkage disequilibrium” (LD)] across one or a few gene regions. These studies have generally concluded that linkage disequilibrium is extremely variable within and among loci and populations [reviewed in (13– 15)]. Recently, examination of a higher density of markers over contiguous regions (16–18) suggested a surprisingly simple pattern: blocks of variable length over which only a few common haplotypes are observed punctuated by sites at which recombination could be inferred in the history of the sample.

The major attraction of haplotype methods is the idea that common haplotypes capture most of the genetic variation across sizable regions and that these haplotypes (and the undiscovered variants they contain) can be tested with the use of a small number of haplotype tag SNPs (“htSNPs”) (16, 18, 19, 39). A number of reports (39–41), however, have suggested that many SNPs fail to conform to the underlying haplotype structure and would be missed by haplotype-based approaches.

Recent genomic surveys have produced high-resolution haplotype information, but only in a small number of human populations.

Conrad et al., (2006) report haplotype structure across 12 Mb of DNA sequence in 927 individuals representing 52 populations. The geographic distribution of haplotypes reflects human history, with a loss of haplotype diversity as distance increases from Africa. Although the extent of linkage disequilibrium (LD) varies markedly across

populations, considerable sharing of haplotype structure exists, and inferred recombination hotspot locations generally match across groups. The study provides the first data on global haplotype variation across multiple megabases of sequence in multiple genomic regions. Global patterns of haplotype variation accord well with a population model in which genetic variation passed through a serial dilution process as humans spread progressively further from our African source (fig 4).

In the future, autosomal haplotype data will surely provide an important tool for unravelling the history of human migrations.

A related application of haplotype data is identifying genomic regions that have been targets of natural selection2,6,7. Such tests, which attempt to identify unusually long common haplotypes likely to have reached high frequencies as a result of recent natural selection, have reduced power in populations that are strongly bottlenecked owing to the property that common haplotypes are often extremely long even in neutral regions.

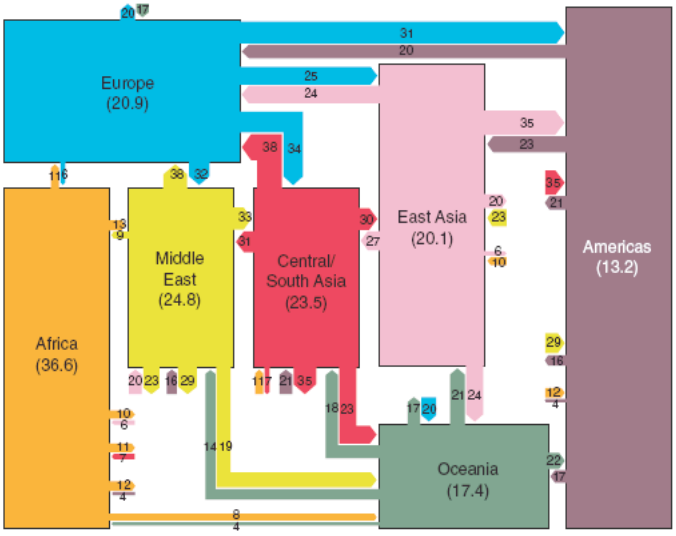


Fig 4 Schematic world map of haplotype diversity. Colored boxes represent regions of the world, positioned geographically. The average number of haplotypes per genomic core region in a sample size of 54 chromosomes is written for each geographic region. Links entering a geographic region indicate the percentages of distinct haplotypes from the geographic region found in other regions (and are drawn proportionately in width). For example, on average 11% of haplotypes observed in Europe in a given part of the genome are found in Africa, whereas 6% of African haplotypes are found in Europe. The links can be viewed as a description of haplotype 'flow': for example, 11% gives a measurement of the proportion of distinct European haplotypes that could have come from Africa (without mutation or recombination), and 6% gives the proportion of African haplotypes that could have come from Europe.

Available resources for studying human genetic diversity

The SNP Database

The single nucleotide polymorphism database (dbSNP) serves as a central depository of SNPs in the public domain. It provides a description of the SNP and its flanking regions and links to multiple National Center for Biotechnology Information (NCBI) Internet pages. The latest release (Build 125) contains more than 27 million SNPs, more than 4 million of which lie within genes. More than 6 million SNPs have been validated and are likely relatively common in the population. Frequency descriptions are available for about 600,000 SNPs, but this figure will dramatically increase with the incorporation of the HapMap phase II data in dbSNP.

The HapMap and Perlegen Projects

Both the International HapMap Consortium (2) and Perlegen Sciences (29) recently produced and released genome-wide genotyping data in samples from several human populations. In both cases, the goal was to cover the entire genome with common DNA polymorphisms that are evenly spaced. The density obtained is of roughly 1 SNP every 2 kb for the Perlegen data set and of 1 every 3 kb for the first phase of the HapMap project (and the latter density will increase rapidly up to 1 SNP every kilobase when phase II is completed). Both data sets are freely available on the Internet. The International HapMap Consortium (2) genotyped individuals from four populations: Centre d'Etude du Polymorphisme Humain (CEPH) (Utah residents with Northern or Western European ancestry), Yoruba from Ibadan (Nigeria), Han Chinese from Beijing (China), and Japanese from Tokyo (Japan). The sampling scheme varies for different populations: The Yoruba and CEPH individuals consist of 30 unrelated trios while the Han Chinese and Japanese are represented by 45 and 44 unrelated individuals, respectively.

The SNPs were chosen in order to obtain a high proportion of intermediate frequency polymorphisms (i.e., with a minor allele frequency higher than 5%). This makes this data set especially valuable for choosing markers for medical studies as they will likely maximize the information for each individual, but it complicates their use in population genetic analyses because the allele frequency spectra are biased (see above). Perlegen Sciences analyzed three populations (CEPH, African Americans, and Han Chinese from the Los Angeles area) represented by 23 or 24 individuals each.

From natural variation to phenotypic differences

Our knowledge of “natural” genetic diversity in humans and its organization has greatly progressed, notably with the results of genome-wide projects. The next step for human geneticists will be to use this tremendous amount of information on natural variation to unravel the still mostly mysterious connections between DNA and phenotypic differences. Using the HapMap and Perlegen data sets, several groups have identified regions with high levels of population differentiation, low levels of diversity, or unusually long stretches of DNA sequence in high or complete LD (2, 8, 50, 72). All these loci are promising candidates to further study the influence of natural selection on the human genome.

Validating these regions, as well as integrating information from other species [such as the chimpanzee genome sequence that was recently released (43)], will shed light on where and how natural selection acted during human evolution to shape current human genome diversity.

Additional studies are necessary to validate these potentially interesting loci as it is still unclear whether these observations are actual genomic differences or artifacts of the ascertainment in the choice of the SNPs.

One cannot reject the possibility that some of these loci may be chance events, given the large amount of data analyzed. Resequencing the regions of interest in a large worldwide sample (e.g., using the CEPH-HGDP panel) would help solve some of these uncertainties and could contribute to the understanding of what makes or made humans human. These results can also have important medical implications by identifying alleles involved in common diseases. Under the assumption that at least some common diseases are due to common variants, one would expect that the disease alleles may, in particular environmental circumstances, confer a benefit to the carrier that would balance the deleterious effect of the disease. Such loci would evolve under balancing selection and thus could be identified by searching the entire genome for loci with a local excess of diversity and highly differentiated haplotypes (70). Alternatively, some common diseases may result from a change in environment during the last phases of human evolution such that the derived protective alleles (formerly deleterious) have not yet reached fixation in humans (16). These alleles should harbor the footprints of a recent (in fact, ongoing) selective sweep [i.e., low diversity, excess of rare alleles,

extended regions of high LD (48)] and could also be identified by a genome-wide search for selected loci performed in healthy individuals.

Medical Genetics: Toward Genome-Wide Studies

The most important contribution from these large resources of human genetic diversity will likely concern the medical field, with, hopefully, the identification of genes (and their variants) involved in genetic disorders and a better understanding of the molecular mechanisms leading to disease. Both Perlegen and the International HapMap data sets can now be used to identify risk or protective alleles involving a particular disease. The HapMap project favored this application by preferentially selecting intermediate allele frequency polymorphisms that are more informative for linkage or association studies (compared to rare alleles). Additionally, these data lead to a very precise description of the LD patterns in the human genome.

Using a set of markers evenly distributed on the physical map (e.g., one marker every ten kb) may be inefficient in testing a significant fraction of the genome with high recombination rates, given that markers in these regions are poorly correlated and thus may have not been “tagged” by the markers used.

Incorporating information about the recombination patterns while choosing the markers allows one to use the LD information to aptly select SNPs [tag-SNPs (tSNPs)] for efficient coverage of the genome. This will lead to a higher marker density in regions of high recombination rate and fewer markers (and less redundancy) in regions of low recombination rates. Thus, one can anticipate that these data will be successful to both narrow down previously identified candidate regions using a denser map of markers (37, 44) and to identify new candidate loci through whole genome association scans (36).

The information and resources resulting from the recent completion of genome-wide diversity projects will surely catalyze the discovery of common genetic variants affecting disease risk, and increase our understanding of gene expression regulation. This is particularly promising for medical genetics, given the availability of numerous genetic markers and adequate technologies to analyze thousands of individuals.

ACKOWLEDGMENTS

1. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K.,Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. Nat. Genet., 36, 949–951.

2. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. et al. (2004) Large-scale copy number polymorphism in the human genome. Science, 305, 525–528. 3. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R. et al. (2005) Segmental duplications and copy-number variation in the human genome. Am. J. Hum. Genet., 77, 78–88.

4. Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D. et al. (2005) Fine-scale structural variation of the human genome. Nat. Genet., 37, 727–732.

5. Feuk, L., Macdonald, J.R., Tang, T., Carson, A.R., Li, M., Rao, G., Khaja, R. and Scherer, S.W. (2005) Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet., 1, e56.

6. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. and Pritchard, J.K. (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet., 38, 75–81.

7. Hinds, D.A., Kloek, A.P., Jen, M., Chen, X. and Frazer, K.A. (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat. Genet., 38, 82–85.

8. McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J. et al. (2006) Common deletion polymorphisms in the human genome. Nat. Genet., 38, 86–92.

9. Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J. and Donnelly, P. (2005) A haplotype map of the human genome. Nature, 437, 1299–1320.

10. The International HapMap Consortium (2003) The International HapMap Project. Nature, 426, 789–796.

11. Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M. et al. (1994) A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). Science, 265, 2049–2054.

12. Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G. et al. (2002) A high-resolution recombination map of the human genome. Nat. Genet., 31, 241–247.

13. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. et al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science, 235, 1616–1622.

14. Buckley, P.G., Mantripragada, K.K., Piotrowki, A., Diaz de Stahl, T. and Dumanski, J.P. (2005) Copy-number polymorphisms: mining the tip of an iceberg. Trends Genet., 21, 315–317.

15. Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Althshuler, D.M., Aburatani, H., Jones, K., Tyler-Smith, C., Hurles, M.E. et al.(2006) Copy number variation: new insights in genome diversity. Genome Res., submitted.

16. Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet., 18, 74–82.

17. Bugge, M., Bruun-Petersen, G., Brondum-Nielsen, K., Friedrich, U., Hansen, J., Jensen, G., Jensen, P.K., Kristoffersson, U., Lundsteen, C., Niebuhr, E. et al. (2000) Disease associated balanced chromosome rearrangements: a resource for large scale genotype–phenotype delineation in man. J. Med. Genet., 37, 858–865.

18. Lupski, J.R. and Stankiewicz, P. (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet., 1, e49.

19. Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. Nat. Rev. Genet., 7, 85–97.

20. Carter, N.P. (2004) As normal as normal can be? Nat. Genet., 36, 931–932.

21. Eichler, E.E. (2006) Widening the spectrum of human genetic variation. Nat. Genet., 38, 9–11.

22. Lee, C. (2005) Vive la difference! Nat. Genet., 37, 660–661.

23. Nadeau, J.H. and Lee, C. (2006) Genetics: copies count. Nature, 439, 798–799.

24. Check, E. (2005) Human genome: patchwork people. Nature, 437, 1084–1086.

25. van Ommen, G.J. (2005) Frequency of new copy number variation in humans. Nat. Genet., 37, 333–334.

26. Cawthon, R.M., Weiss, R., Xu, G.F., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. et al. (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. Cell, 62, 193–201.

27. Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H. et al. (1990) Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell, 60, 509–520.

28. Royer-Pokora, B., Kunkel, L.M., Monaco, A.P., Goff, S.C., Newburger, P.E., Baehner, R.L., Cole, F.S., Curnutte, J.T. and Orkin, S.H. (1986) Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. Nature, 322, 32–38.

29. Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M. and Kunkel, L.M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature, 323, 646–650.

30. Burghes, A.H., Logan, C., Hu, X., Belfall, B., Worton, R.G. and Ray, P.N. (1987) A cDNA clone from the Duchenne/Becker muscular dystrophy gene. Nature, 328, 434–437.

31. Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M. and Dryja, T.P. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature, 323, 643–646.

32. Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H. and Bruns, G.A. (1990) Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. Nature, 343, 774–778.

33. Wallace, M.R., Marchuk, D.A., Andersen, L.B., Letcher, R., Odeh, H.M., Saulino, A.M., Fountain, J.W., Brereton, A., Nicholson, J., Mitchell, A.L. et al. (1990) Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science, 249, 181–186.

34. Vissers, L.E., Veltman, J.A., van Kessel, A.G. and Brunner, H.G. (2005) Identification of disease genes by whole genome CGH arrays. Hum. Mol. Genet., 14 (Spec no. 2), R215–R223.

35. Rosenberg, C., Knijnenburg, J., Bakker, E., Vianna-Morgante, A.M., Sloos, W., Otto, P.A., Kriek, M., Hansson, K., Krepischi-Santos, A.C., Fiegler, H. et al. (2006) Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents. J. Med. Genet., 43, 180–186.

36. Schoumans, J., Ruivenkamp, C., Holmberg, E., Kyllerman, M., Anderlid, B.M. and Nordenskjold, M. (2005) Detection of chromosomal imbalances in children with

idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). J. Med. Genet., 42, 699–705.

37. de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N. et al. (2005) Diagnostic genome profiling in mental retardation. Am. J. Hum. Genet., 77, 606–616.

38. Shaw-Smith, C., Redon, R., Rickman, L., Rio, M., Willatt, L., Fiegler, H., Firth, H., Sanlaville, D., Winter, R., Colleaux, L. et al. (2004) Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J. Med. Genet., 41, 241–248.

39. Miyake, N., Shimokawa, O., Harada, N., Sosonkina, N., Okubo, A., Kawara, H., Okamoto, N., Kurosawa, K., Kawame, H., Iwakoshi, M. et al. (2006) BAC array CGH reveals genomic aberrations in idiopathic mental retardation. Am. J. Med. Genet. A, 140, 205–211.

40. Bauters, M., Van Esch, H., Marynen, P. and Froyen, G. (2005) X chromosome array-CGH for the identification of novel X-linked mental retardation genes. Eur. J. Med. Genet., 48, 263–275.

41. Lugtenberg, D., de Brouwer, A.P., Kleefstra, T., Oudakker, A.R., Frints, S.G., Schrander-Stumpel, C.T., Fryns, J.P., Jensen, L.R., Chelly, J., Moraine, C. et al. (2005) Chromosomal copy number changes in patients with non-syndromic X-linked mental retardation detected by array CGH. J. Med. Genet., doi:10.1136/jmg.2005.036178.

42. Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C. et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet., 34, 27–29.

43. Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., Devon, R.S., Clair, D.M., Muir, W.J., Blackwood, D.H. et al. (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum. Mol. Genet., 9, 1415–1423.

44. Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H.,Muller,K., Kaaranen,M., Lindsberg, P.J., Hannula-Jouppi, K. et al. (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. Proc. Natl Acad. Sci. USA, 100, 11553–11558.

45. Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H. and Kere, J. (2005) The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. PLoS Genet., 1, e50.

46. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. and Monaco, A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature, 413, 519–523.

47. Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M. et al. (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science, 310, 317–320.

48. Somerville, M.J., Mervis, C.B., Young, E.J., Seo, E.J., del Campo, M., Bamforth, S., Peregrine, E., Loo, W., Lilley, M., Perez-Jurado, L.A. et al. (2005) Severe expressive-language delay related to duplication of the Williams–Beuren locus. N. Engl. J. Med., 353, 1694–1701.

49. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R. et al. (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science, 302, 841.

50. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M. et al. (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat. Genet., 38, 24–26.

51. Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E. et al. (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature, 439, 851–855.

52. Groot, P.C., Mager, W.H. and Frants, R.R. (1991) Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family. Genomics, 10, 779–785.

53. Hollox, E.J., Armour, J.A. and Barber, J.C. (2003) Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster.Am. J. Hum. Genet., 73, 591–600.

54. Aldred, P.M., Hollox, E.J. and Armour, J.A. (2005) Copy number polymorphism and expression level variation of the human falphag-defensin genes DEFA1 and DEFA3. Hum. Mol. Genet., 14, 2045–2052.

55 Zondervan, K.T. & Cardon, L.R. The complex interplay among factors that influence allelic association. Nat. Rev. Genet. 5, 89–100 (2004)

56 The International HapMap Consortium. A haplotype map of the human genome. Nature 437, 1299–1319 (2005).

57 Tishkoff, S.A. et al. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271, 1380–1387 (1996).

58 Reich, D.E. et al. Linkage disequilibrium in the human genome. Nature 411, 199–204 (2001).

59 Plagnol, V. & Wall, J.D. Possible ancestral structure in human populations. PLoS Genet. 2, 972–979 (2006).

60 Sabeti, P.C. et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837 (2002).

61 Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. PLoS Biol. 4, e72 (2006).

62 McVean, G.A.T. et al. The fine-scale structure of recombination rate variation in the human genome. Science 304, 581–584 (2004).

63 Ptak, S.E. et al. Fine-scale recombination patterns differ between chimpanzees and humans. Nat. Genet. 37, 429–434 (2005).

64 Fearnhead, P. & Smith, N.G. A novel method with improved power to detect recombination hotspots from polymorphism data reveals multiple hotspots in human genes. Am. J. Hum. Genet. 77, 781–794 (2005).

65 Patil, N. et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294, 1719–1723 (2001).

Hinds, D.A. et al. Whole-genome patterns of common DNA variation in three human populations. Science 307, 1072–1079 (2005).

66 De Bakker, P.I.W., Graham, R.R., Altshuler, D., Henderson, B.E. & Haiman, C.A. Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations. Pac. Symp. Biocomput. 11, 478–486 (2006).

67 Huang, W. et al. Linkage disequilibrium sharing and haplotype-tagged SNP portability between populations. Proc. Natl. Acad. Sci. USA 103, 1418–1421 (2006).

68 Montpetit, A. et al. An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. PLoS Genet. 2, e27 (2006).

69 Service, S. et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat. Genet. 38, 556–560 (2006).

70 Willer, C.J. et al. Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. Genet. Epidemiol. 30, 180–190 (2006).

71 Yoo, Y.K. et al. Fine-scale map of Encyclopedia of DNA Elements regions in the Korean population. Genetics 174, 491–497 (2006).

72 Rosenberg, N.A. et al. Genetic structure of human populations. Science 298, 2381–2385 (2002).

73 Barrett, J.C. & Cardon, L.R. Evaluating coverage of genome-wide association studies. Nat. Genet. 38, 659–662 (2006).

74 Sawyer, S.L. et al. Linkage disequilibrium patterns vary substantially among populations. Eur. J. Hum. Genet. 13, 677–686 (2005).

75 Bonnen, P.E. et al. Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. Nat. Genet. 38, 214–217 (2006).

76 Gonzalez-Neira, A. et al. The portability of tagSNPs across populations: a worldwide survey. Genome Res. 16, 323–330 (2006).

77 Cann, H.M. et al. A human genome diversity cell line panel. Science 296, 261–262 (2002).

78 Cavalli-Sforza, L.L. The Human Genome Diversity Project: past, present and future. Nat. Rev. Genet. 6, 333–340 (2005).

79 Rosenberg, N.A. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann. Hum. Genet. published online 29 March 2006 (doi:10.1111/j.1469-1809.2006.00285.x).

80 Kong, A. et al. A high-resolution recombination map of the human genome. Nat. Genet. 31, 241–247 (2002).

81 Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644 (2006).

82 Ramachandran, S. et al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA 102, 15942–15947 (2005).

83 Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. & Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15, 1496–1502 (2005).

84 Mountain, J.L. & Cavalli-Sforza, L.L. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. Proc. Natl. Acad. Sci. USA 91, 6515–6519 (1994).

85 Nielsen, R. Population genetic analysis of ascertained SNP data. Hum. Genomics 1, 218–224 (2004).

86 Rogers, A.R. & Jorde, L.B. Ascertainment bias in estimates of average heterozygosity. Am. J. Hum. Genet. 58, 1033–1041 (1996).

87 Bowcock, A.M. et al. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368, 455–457 (1994).

88 Crawford, D.C. et al. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am. J. Hum. Genet. 74, 610–622 (2004).

89 Stephens, J.C. et al. Haplotype variation and linkage disequilibrium in 313 human genes. Science 293, 489–493 (2001).

90 Nielsen, R., Hubisz, M.J. & Clark, A.G. Reconstructing the frequency spectrum of ascertained single-nucleotide polymorphism data. Genetics 168, 2373–2382 (2004).

91 Winckler, W. et al. Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308, 107–111 (2005).

92 De Bakker, P.I.W. et al. Efficiency and power in genetic association studies. Nat. Genet. 37, 1217–1223 (2005).

93 Carlson, C.S. et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J. Hum. Genet. 74, 106–120 (2005).

94 Rosenberg, N.A. et al. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 1, 660–671 (2005).

95 Marchini, J. A comparison of phasing algorithms for trios and unrelated individuals. Am. J. Hum. Genet. 78, 437–450 (2006).

96 Hurlbert, S.H. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52, 577–586 (1971).

97 Kalinowski, S.T. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv. Genet. 5, 539–543 (2004).

98 Nei, M. Molecular Evolutionary Genetics (Columbia Univ. Press, New York, 1987).

99 Weir, B.S. Genetic Data Analysis II (Sinauer, Sunderland, Massachusetts, 1996).

Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. Genetics 155, 945–959 (2000).

CHAPTER 1

MOLECULAR VARIABILITY IN HAMP GENE PROMOTER AND CODING REGION IN A COHORT OF HEMOCHROMATOSIS PATIENT

INTRODUCTION

Iron is ubiquitous in nature. During evolution, virtually all forms of life developed various methods and means to keep iron at adequate levels and prevent deficiency. Iron is essential for many intracellular functions by virtue of its ability to accept or donate electrons. This same characteristic, however, allows free iron in solution to form highly reactive free radicals that can lead to cell damage. The harmful effects are related to increased oxidative stress and production of reactive oxygen species causing oxidative damage to lipids, proteins, and nucleic acids. The appropriate regulation of systemic iron homeostasis, therefore, is crucial for the survival and wellbeing of all complex organisms, including humans. The average male human contains approximately 4 grams of iron (Brittenham 1994). Under normal conditions approximately 1–2 mg of iron per day enters the body via the enterocytes of the proximal small intestine. This newly absorbed dietary iron is released into the circulation and binds to the serum protein transferrin. Approximately 3 mg of iron circulates bound to transferrin and, although this represents only a small proportion of total body iron, evidence is accumulating that this iron compartment is vital for the maintenance of body iron homeostasis. Transferrin bound iron is taken up by cells by transferrin receptor 1 (TfR1)-mediated endocytosis (Huebers and Finch 1987). Most of the transferrin bound iron in the circulation is destined for the developing erythrocytes of the bone marrow, where it is taken up at a rate of approximately 22 mg of iron per day (Brittenham 1994), and used in the production of haemoglobin. About 65–70% of body iron exists in this form in circulating red blood cells. Old or damaged red blood cells are removed from the circulation by the macrophages of the reticuloendothelial (RE) system, where iron is released from haemoglobin and either stored in the intracellular iron storage protein ferritin, or released back into the circulation as transferrin-bound iron. The cells of the RE system release about 22 mg of iron per day, thus replacing the amount taken up by the bone marrow. Other body cells also take up iron from the circulation via TfR1-mediated endocytosis although in far lower

amounts. This iron can be used to synthesize a wide range of iron containing intracellular proteins, such as the haem containing cytochromes, and proteins containing iron-sulfur clusters. Approximately 10–15% of body iron is present in such proteins, with up to 80% of this found in muscle cell myoglobin. The remaining 20% of body iron is present as storage iron, predominantly located in the macrophages of the RE system and the hepatocytes of the liver. The movement of iron between these compartments is a tightly regulated process that can be modulated according to the body's iron requirements. Individual cells maintain appropriate intracellular iron levels by altering the expression of TfR1 on the cell surface (Huebers and Finch 1987). The more iron each cell requires, the higher the expression of TfR1. While the uptake of iron by cells is predominantly controlled locally by intracellular iron levels, iron release, particularly from the cells of the RE system, liver and proximal small intestine, appears to be regulated by systemic signals. Evidence now suggests that the control of cellular iron efflux is the major regulatory point for the maintenance of systemic iron homoestasis (Frazer and Anderson 2003). For example, the stimulation of erythropoiesis following blood loss increases the iron requirements of the bone marrow. To cope with this demand, the body increases cellular iron release, making available iron stored in the macrophages of the RE system and hepatocytes of the liver. At the same time, the increase in iron release by intestinal enterocytes allows iron taken up from the diet to enter the circulation and replenish the body's iron stores. Once erythropoiesis and the demand for iron have reduced to normal levels, cellular iron release also decreases to maintain iron homeostasis. The importance of tightly regulated cellular iron release is evident from a number of pathological conditions in humans. Patients with the genetic disorder HFE-associated haemochromatosis may develop severe iron loading that can often result in tissue damage and possibly organ failure (Fleming et al. 2005). This appears to be caused by inappropriately high iron release by enterocytes and macrophages. As a result, iron normally stored in the RE system or lost from the alimentary canal is released into the circulation, overwhelming the iron binding capacity of transferrin, and allowing the formation non-transferrin bound iron (Batey et al. 1980). This form of iron is very rapidly taken up by cells, particularly hepatocytes, leading to iron overload (Brissot et al. 1985). Another pathological condition involving altered iron release from cells is the anaemia of inflammation. During an inflammatory response, iron release from cells decreases, lowering the level of transferrin-bound iron in the circulation (Weiss and Goodnough 2005). This response

probably evolved to assist the immune system to overcome infection by withholding iron (a well recognized virulence factor for microorganisms) from invading pathogens. However, in chronic inflammatory conditions, such as rheumatoid arthritis, the transferrin saturation is reduced for long enough to compromise iron delivery to the developing erythrocytes in the bone marrow, resulting in anaemia (Weiss and Goodnough 2005). In both of the above pathological states, cellular iron release does not reflect the body's iron requirements. Due to its physiological importance, the mechanism by which the body regulates iron release is the focus of intense scrutiny.

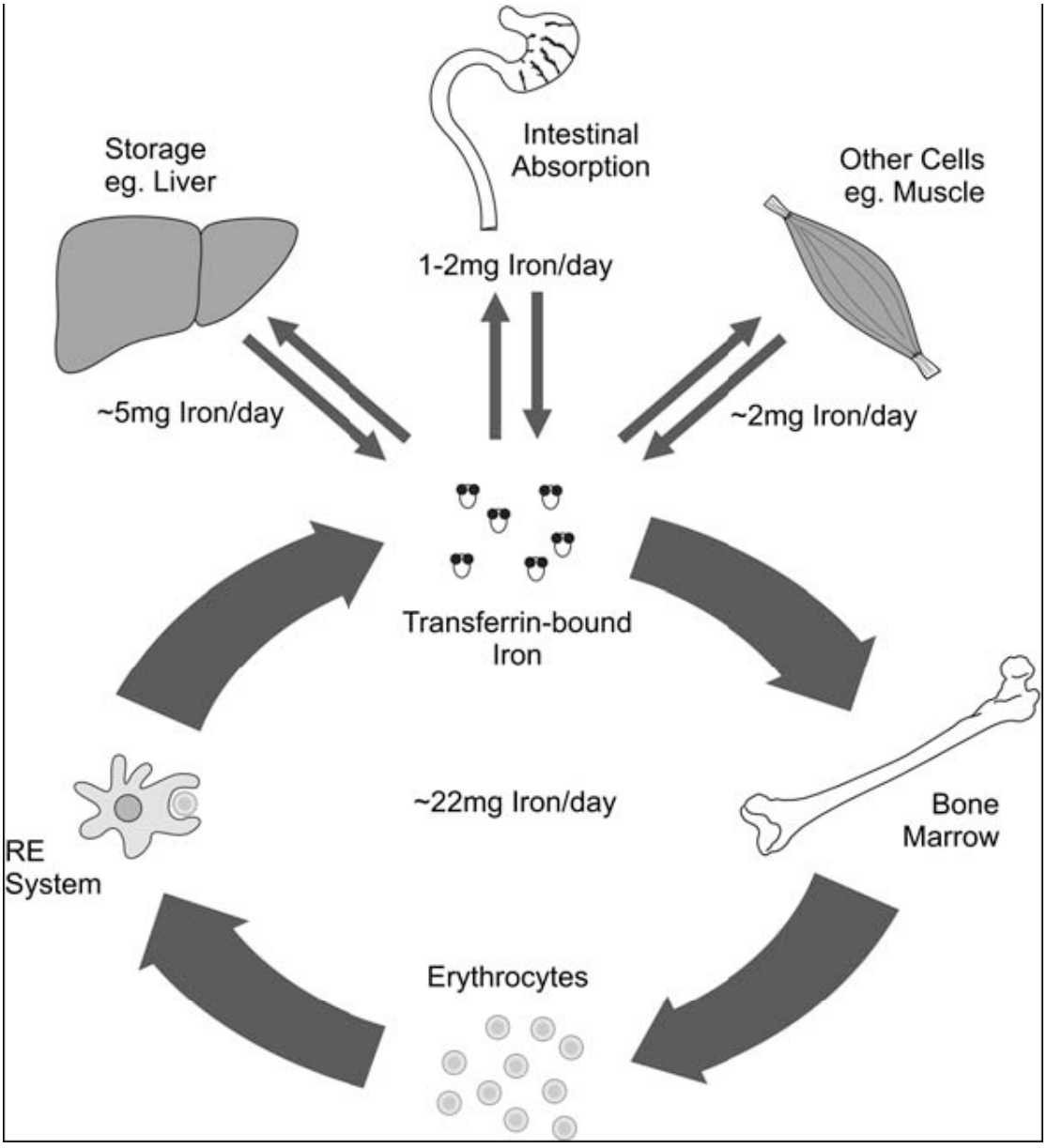


Fig.1 scheme of daily iron movement in the different body compartment

HEREDITARY HEMOCHROMATOSIS

This term was introduced to define the association of widespread tissue injury with massive tissue iron deposition1 and likely referred to a clinical entity named ‘‘bronze diabetes’’2 and ‘‘cirrhose pigmentaire’’3 first reported in France in the second half of the 19th century. After only one century the term was associated with an hereditary disease4 and linked to the major histocompatibility class I complex A3, on the short arm of chromosome 6.5 6 In 1996, the most prevalent HC gene, HFE, was cloned.7 However, once the HFE gene was identified, it appeared immediately clear that HFE mutations accounted for most but not all cases of HC.8. Since then, unprecedented progress in the field of iron genetics has led to the identification of new genes involved in iron metabolism whose mutations are responsible for cases of hereditary iron storage disorders. The term haemochromatosis (HC) (synonymous for hereditary or idiopathic or primary haemochromatosis) defines an autosomal recessive disorder of iron metabolism characterized by tissue iron overload potentially leading to multiorgan disease, such as liver cirrhosis, endocrinopathy, and cardiomyopathy. The syndrome is the result of a genetically determined failure to stop iron from entering the circulatory pool when it is not needed. It is associated with pathogenic mutations of at least four HC genes (that is, HFE, TfR2, HJV, and HAMP) and it is likely due to a regulatory defect in iron homeostasis in the liver.9 Four basic features define HC and are characteristic of the classic disorder related to HFE C282Y homozygosity (the prototype for this subset and by far the most common form) and the rare disorders more recently attributed to loss of TfR2, HAMP, or HJV . These features include: hereditary, usually autosomal recessive, trait; iron overload initially involving the plasma compartment (reflected in increasing serum transferrin saturation); iron overload subsequently involving parenchymal cells (reflected in increasing serum ferritin) with the potential for organ damage and disease; unimpaired erythropoiesis and optimal response to phlebotomy.9 Established concepts regarding primary cause and molecular pathogenesis of HC have been more or less all challenged by recent new discoveries, particularly those pertaining to the genetic field. Based on this new information and on circumstantial evidence provided by recent human and animal studies, a unifying pathogenic model for HC is presented in this article.

HFE

HFE is a major histocompatibility class I-like protein whose ancestral peptide binding groove is too narrow to allow classic antigen presentation10 while a possible non-classic immunological activity has been recently proposed.11 It is incapable of binding iron.12 Interaction between HFE and the transferrin receptor, TfR1, which mediates transferrin bound iron uptake by most cells,12 13 has been fully documented although its biological effects are still uncertain. However, it is unclear whether the interaction of HFE with TfR1 is key for the pathogenesis of HC.14–16 The C282Y mutation (substitution of tyrosine for cysteine at position 282 due to a single base,
845G-.A), the most common pathogenic mutation of HFE, is associated with disruption of a disulfide bond in HFE that is critical for its binding to b2 microglobulin.17 The latter interaction is necessary for the stabilization (intracytoplasmic), transport, and expression of HFE on the cell surface and endosomal membranes where HFE interacts with TfR1. The H63D mutation, a common HFE mutation whose pathogenic significance is still uncertain, does not impair HFE-TfR1 interaction. While the biological function of HFE is still unknown, circumstantial evidence indicate that it might be required for the synthesis of hepcidin, the iron hormone secreted by hepatocytes.

TRANSFERRIN RECEPTOR 2 (TFR2)

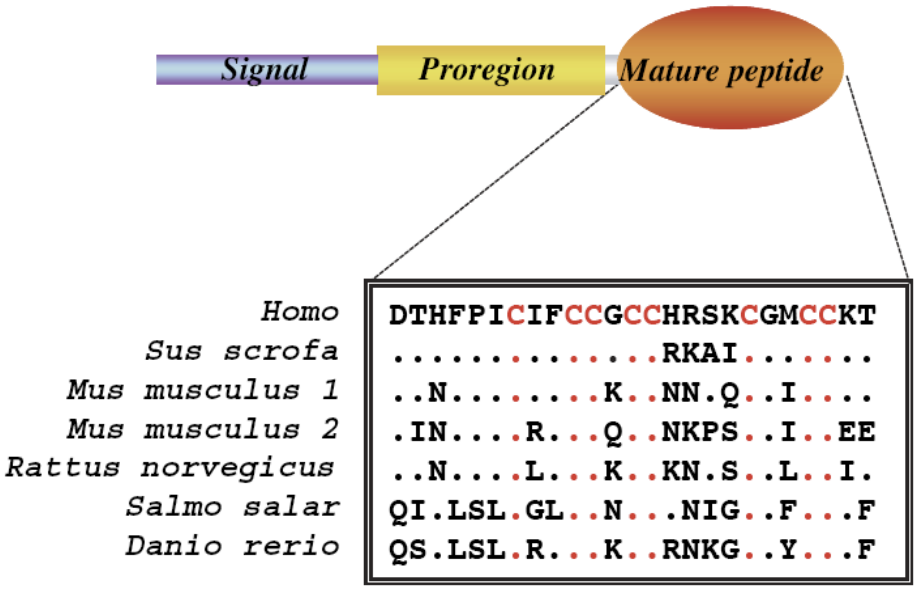
In 1999, the gene for a second human transferrin receptor (TfR2) was cloned.18 Unlike TfR1, the new receptor was found to be highly expressed in the liver and it was not regulated by intracellular iron status.19 TfR2 mediates the uptake of transferrin bound iron by hepatocytes,18 possibly through the mechanism of receptor mediated endocytosis similar to that described for TfR1, but its in vitro affinity for transferring is 25–30-fold lower than that of TfR1.20 Yet, TfR2 mediated transferrin iron uptake may be of importance in hepatocytes, which express a low number of TfR1. The biological role and function of TFR2 remain unknown, but recent studies
suggests a role for TfR2 in hepcidin synthesis in the liver (see below). In fact, its putative role in hepatocyte uptake of iron18 is difficult to reconcile with the HC phenotype observed in humans with pathogenic TfR2 mutations21 and in TfR2 knockout mice.22 TfR2 does not seem to interact with HFE,20 but its persistent hepatic expression during iron overload might conceivably reflect a contribution to the modulation of hepcidin synthesis in this setting.

HEMOJUVELIN (HJV)

Hemojuvelin (also called HFE2, or repulsive guidance molecule C (RgmC)) is transcribed from a gene of 4265 bp into a full length transcript with five spliced isoforms.23 Analyses of hemojuvelin in human tissues detect substantial expression in adult and fetal liver, heart, and skeletal muscle.23 The putative full length protein is 426 amino acids with a large RGM motif, homologous to repulsive guidance molecules involved in neuronal cells migration; it contains a C terminal putative transmembrane domain characteristic of a glycosylphosphatidylinositol linked membrane anchor (GPI anchor). Removal of the GPI anchor or proteolysis would be expected to generate a soluble form of hemojuvelin, suggesting that it can be present in either a soluble or a cell associated form. The function of hemojuvelin is presently unknown. However, hepcidin levels are depressed in individuals with HJV mutations,23 and in HJV knockout mice,24 and a recent study in vitro indicated that HJV is a transcriptional regulator of hepcidin25 (see below). In this study, cellular hemojuvelin positively regulated hepcidin mRNA expression, and recombinant soluble hemojuvelin suppressed hepcidin mRNA expression in primary human hepatocytes in a log linear dose dependent manner, suggesting binding competition between soluble and cell associated hemojuvelin.

HEPCIDIN (HAMP)

Hepcidin, the long awaited iron hormone, is an antimicrobial peptide produced by hepatocytes in response to inflammatory stimuli and iron.26–28 It is the product of the HAMP gene, consisting of three exons and two introns located on chromosomes 7 and 19 in mouse and humans, respectively. Humans and rats have a single HAMP gene 28, whereas two functional genes, Hamp 1 and 2 are present in the mouse genome.29 Hepcidin mRNA is nearly confined to the liver. The transcript encodes a precursor protein of 84 amino acids, including a putative 24 amino acid leader peptide, while the circulating forms consist of only the C terminal portion (20 and 25 amino acid peptides).



In solution, the small cysteinerich hepcidin peptides form a distorted beta sheet with an unusual vicinal disulphide bridge found at the turn of the hairpin, which is probably of functional significance.30 Due to significant antibacterial and antifungal activities of the C terminal peptide, hepcidin has been classified as a member of the cysteine-rich, cationic, antimicrobial peptides, including the thionins and defensins. Evidence from transgenic mouse models indicates that hepcidin is the principal downregulator of the transport of iron across the small intestine and the placenta, and its release from macrophages. In vivo injection of hepcidin into mice significantly reduced mucosal iron uptake and transfer to the carcass, independently of iron status or presence of HFE,31 or induces hypoferraemia in humans.32 The present view is that hepcidin downregulates iron efflux from the intestine and macrophages by interacting with the main iron export protein in mammals, ferroportin (FPN). In fact, it has been recently shown that hepcidin binds to FPN in cultured cells stably expressing FPN and, following complex internalisation, leads to FPN degradation.33

A UNIFYING PATHOGENIC MODEL FOR HAEMOCHROMATOSIS

The first biochemical manifestation of HC is an increase in transferrin saturation which reflects an uncontrolled influx of iron into the bloodstream from enterocytes and macrophages. Duodenal transfer of iron to plasma is inappropriately high for body iron stores,49 suggesting downregulation failure in HC. Phlebotomy normally triggers sharp

transient increases in absorption (from 1–2 mg/day to 5 mg/day), mainly to ensure bone marrow supplies, but in HC this response is exaggerated (8–10 mg/day) and the rate remains high for years.50 The end result is intestinal absorption of iron that generally exceeds loss by approximately 3 mg/day in HC.51 While the only way that total body iron can be increased is through increased intestinal iron absorption, macrophages (normally a much more important source of plasma iron than either enterocytes or hepatocytes)52 are also important in the pathogenesis of HC. They are invariably iron poor in HC and seem to release more iron or to retain less transferrin bound iron than their normal counterparts.9 Historically, the intestine has been seen as the primary site of the defect in HC and studies showing that HFE is normally expressed in intestinal crypts reinforced this idea leading to the development of a specific pathogenic model.9 This model attributed the relative iron deficiency of mature absorptive HC enterocytes and increased intestinal iron absorption to an abnormal interaction between TfR1 and mutant HFE in intestinal crypt cells. The presence in this model of C282Y mutant HFE which is unable to interact with TfR1, leads to iron deficient crypt cells, which give rise to iron deficient daughter cells. These cells are ‘‘programmed’’ to react to iron starvation by hyperactively and persistently absorbing iron from the intestinal lumen and transferring virtually all of it

into the bloodstream, regardless of actual erythropoietic needs. More recently, however, this model has been challenged with the discovery of hepcidin and its central involvement in iron homeostasis. Consequently, attention has moved to the liver as the primary site of the defect in HC. The progressive expansion of the plasma iron pool in HC, which occurs at a much faster rate in ‘‘juvenile’’ forms of the disorder (HJV and HAMP related HC) compared with late onset forms (HFE and TfR2-related HC), is likely the result of increased transfer of iron to the blood compartment from enterocytes (that is, increased intestinal absorption) and from reticuloendothelial macrophages.9 As mentioned above, the main regulator of iron efflux from enterocytes and macrophages in humans is hepcidin. In HFE, TfR2, and HJV related HC, hepatic expression or serum/urine levels of this peptide are inappropriately low.42 43 53 54 Its expression in the liver is also significantly impaired in HFE, TfR2, and HJV knockout mice24 55–57 and hepatic deposition of iron in HFEKO animals can be prevented by hepcidin overexpression.58 These findings suggests a unifying pathogenic model for all forms of HC in which HFE, TfR2, and HJV are all independent but complimentary regulators of

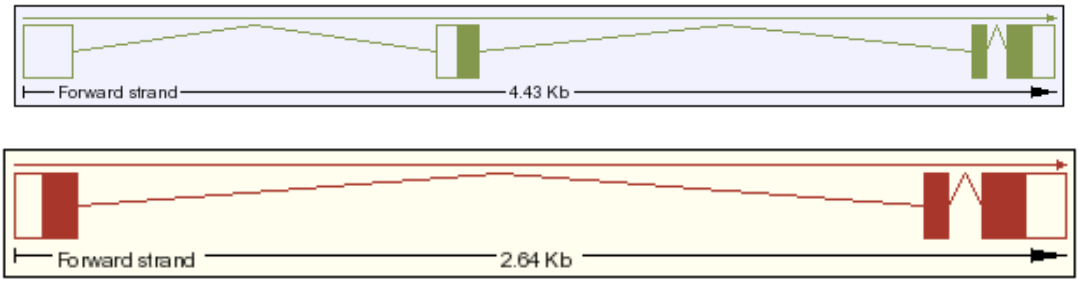
hepcidin synthesis in the liver (fig 1). Hypothetically, these three proteins may be important for

sensing circulatory iron and turning on hepcidin gene transcription, albeit with different mechanisms and functional impact: HFE might have a role in endosomes and/or plasma membranes iron traffic even independently of TfR1; TfR2, highly expressed in hepatocytes even during iron overload, might signal transferrin bound or unbound iron; soluble and cell associated hemojuvelin might reciprocally regulate hepcidin expression in response to changes in extracellular iron concentration: soluble HJV might even signal the iron status of peripheral tissues, such as skeletal muscles where HJV also seems to be expressed. When all three proteins function correctly (and the HAMP gene that encodes hepcidin is normal), the amount of iron transferred

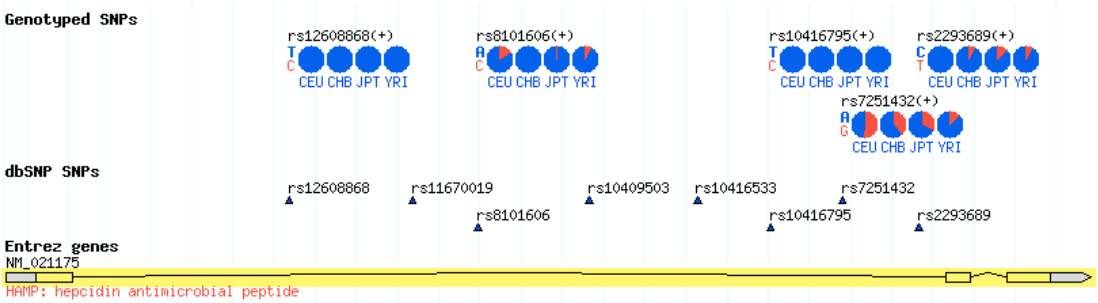
into the blood will be appropriate to body needs, and excessive iron deposition in tissues will be avoided (fig 1A). The relative contributions of the three genes to this modulatory process may be different, with a more substantial role assigned to HJV based on the more severe iron overload phenotype associated with HJV mutations. Loss of one of the minor regulatory proteins (HFE or TfR2 related HC) will result in an appreciable increase in iron influx into the bloodstream but residual hepcidin activity will be sustained by the second minor regulator and the major regulator, HJV

gene (fig 1B, C). The result is a mild ‘‘adult’’ HC phenotype, with gradual plasma iron loading and gradual accumulation of iron in tissues. Loss of the ‘‘major’’ hepcidin regulator, HJV, will produce a more dramatic effect on iron into the bloodstream and result in a more severe, ‘‘juvenile’’, HC. Combined loss of HFE and TfR2 (HFE+TfR2 related HC) would theoretically result in much more rapid and substantial increases in plasma iron and, consequently, greater iron overload in tissues: in short, a more severe juvenile phenotype, not unlike that produced by loss of HJV. In fact, a recent study has described patients with severe juvenile HC phenotype associated with combined mutations of HFE and TfR2.59 Finally, complete loss of hepcidin (HAMP related HH), in spite of normal HFE, TfR2, and HJV, will inevitably lead to massive uncontrolled release of iron into the circulation and severe HC phenotype.

HAMP GENE



HAMP gene is located in chromosome 19 q13.12. In Vega database (http://vega.sanger.ac.uk/Homo_sapiens/) two different transcripts are reported, one longer with an additional intron and an additional 5' UTR that spans in a region of 4,43 Kb and one shorter characterized by only two introns that spans in a region of 2.64 Kb. Considering the central role of this small peptide and the different stimuli to which it has to respond the hypothesis that the gene has two promoters in order to get different pathway of activation seems quite logical. Although logical the longer form of the gene is actually observed in only one reports and in any other databases the protein is reported only in the shorter version, thus in our study we considered the shortest one.



From deep researches in different databases merged that in the coding region of the HAMP gene no polymorphic variants are reported.

THE STUDY

This data is consistent with the reported central role of the protein. Aim of our study is to screen for new mutation the coding regions and the promoter of HAMP gene in 83 unrelated subjects characterized by high levels of iron in blood. As previously reported HAMP gene plays a pivotal role in iron homeostasis, so we selected subjects that presents physiological picture that can be related to juvenile hereditary hemochromatosis. All the subjects participating the study where previously typed for

the three major HFE mutations C282Y, H63D, S65C in order to exclude for the high iron serum level an involvement of those variants.

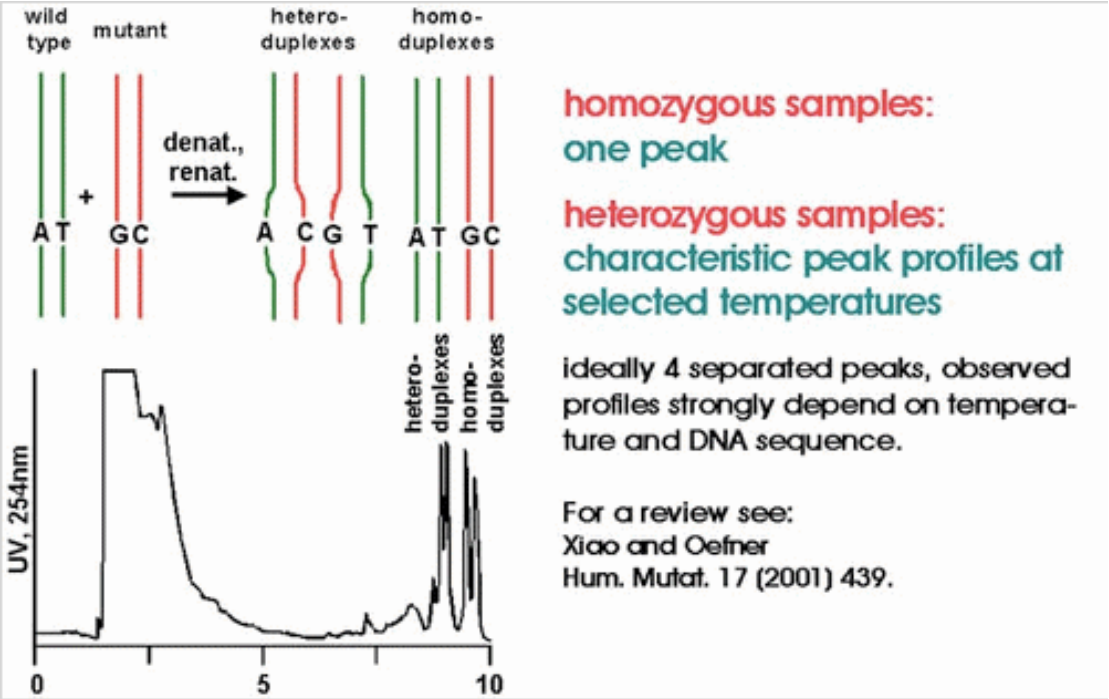
MATERIAL AND METHODS

Subjects: A total of 84 unrelated probands between 7 and 47 years of age with serum transferrin saturation values between 40% to 99% were selected. The probands were previously typed for the three most prevalent mutations related to hereditary hemochromatosis: C282Y, H63D, S65C in the Hfe gene, and resulted negative. The patients were referred from Clinical Hematologic, Interne Medicine from S. Orsola-Malpighi hospital of Bologna.

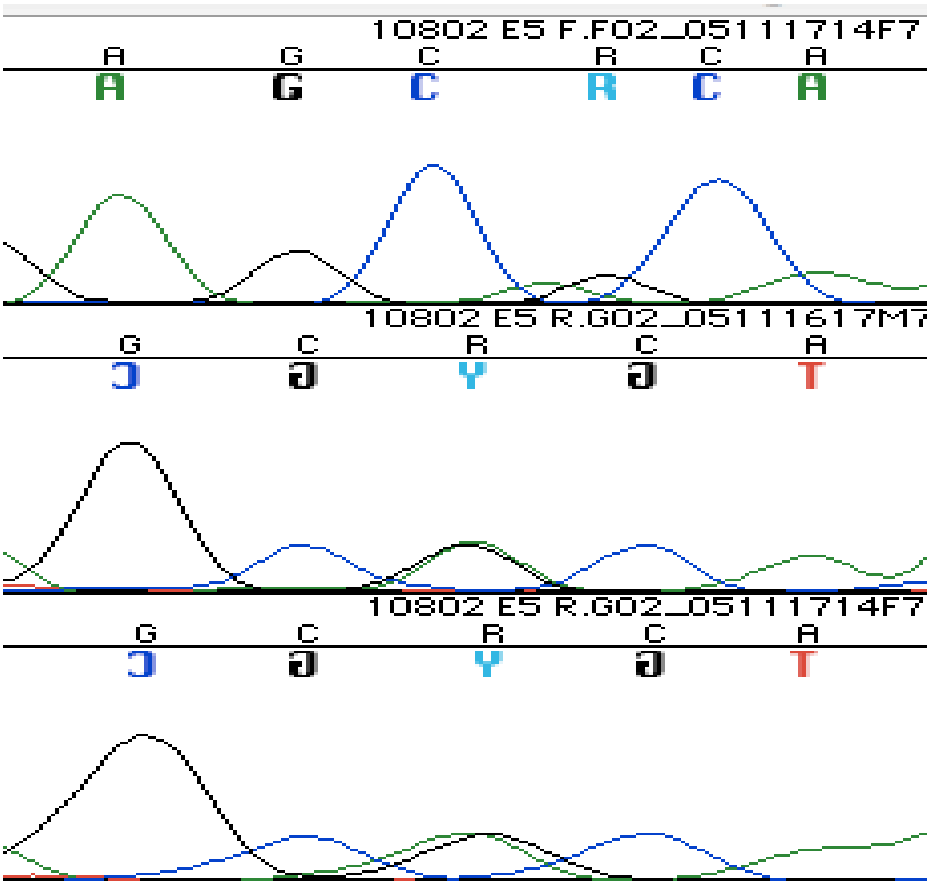
Genetic analysis: All patients underwent mutation's analysis in the HAMP gene.

DNA was isolated from peripheral blood lymphocytes using QI Amp DNA Blood Kit (Quiagen Inc. Valencia, CA). All 3 exons flanking intronic sequences and the promoter of the HAMP gene (Gene Bank AF041021), were amplified in 50 µl standard PCR containing 50 ng DNA using Ampli-Taq Gold and Gene-Amp PCR system 9700 thermocycler (Applied Biosystems).

Following verification of product size by agarose gel electrophoresis, amplicons were denaturated at 95°C for 10 minutes and allowed to renature over 10 minutes to create heteroduplices. Each fragment was run on an automated WAVE dHPLC (denaturated High Performance Liquid Chromatography) instrument equipped with a DNA Sep® column (transgenomic, San Jose, CA).



All variant identified by the denaturing HPLC scanning were examined by direct sequencing of both strands using the CEQ 8000 protocol for Taq-dye terminator cycle sequencing on an automated CEQ 8000 DNA Sequencer (Beckman Coulter Inc., Fullerton, CA) after a purification with Seq-Prep Kit (Gene Dia, Lammari, LU, Italy). Results were analyzed with Sequencer software (Gene Codes Corp., Ann Arbor, MI).



Wherever possible the novel mutations detected were confirmed by specific restriction analysis using commercial enzymes and digestion conditions suggested by the producer (New England BioLabs Inc., Beverly, MA).

Nomenclature for the description of human sequence variations was referred to J.T. dan Dunnan and Antonorakis.

RESULTS

A total of 84 unrelated probands between 7 and 47 years of age with serum transferrin saturation values between 40% to 99% were included in the study. The probands resulted negative for the three most prevalent mutations related to hereditary hemochromatosis in the HFE gene. All the 84 subject were screened for new mutations in the coding region and the promoter of HAMP gene. The screening was assessed by the use of DHPLC technology. The three exons and the promoter were subdivided in 5 five amplicons, with a final score of 420 amplicons tested in the study. From the analysis of amplicons only five variants were discovered.

DNA		4675	8514	8573	10802
PROMOTORE	HEP 2 E2F+E2R DHPLC	WT	WT	ET	ET
	HEP 2 E2F+E2R SEQ	\	\	IVS1-69 C-T	2392 C-G
ESONE 1	E3F+ E3R NEW DHPLC	WT	WT	ET	ET
	E3F+ E3R NEW SEQ	\	\	IVS1-69 C-T	2392 C-G
ESONE 2	HEP 4 (ex3-4) E4Fbis+E4R DHPLC	WT	ET	ET	ET
	HEP 4(ex3) E4Fbis+E4R SEQ	\	4634 G-A	4502 C-T	4634 G-A
ESONE 3	HEP 5(ex3-4) E5F+E5R DHPLC	ET	ET	WT	ET
	HEP 5(ex3-4) E5F+E5R SEQ	4818 G-A	4634 G-A	WT	4634 G-A

One variant is in position 4502**,** is a transition C to T, this variant is reported in dbSNP database, rs2293689, and is one of the SNPs typed in the HAPMAP project and in AFD. Both on AFD and in HapMap project database the T allele is not reported in European panel, therefore is the first time that this variant appear in European population.

The other three variants are reported for the first time and probably are to be considered private. One variant is in the promoter region, in position –69, it is a transition C to T . The position –69 is in the overlapping region of amplicon 1 and 2 so both in DHPLC and sequence the signal of the transition was detected in the two different PCR products.

In position 2392 we found a transversion C to G, it is in the coding region, but is synonymous mutation, lysine to lysine. The position 2392 is in the overlapping region of the amplicons 2 and 3, so it was confirmed both in DHPLC and sequence in two different PCR products.

In position 4818 a transition G to A is detected. This variant is described for the first time, is in the coding region and is a synonymous substitution, tyrosine to tyrosine

Finally a new variants is found in two different samples, it is a transition G to A in position 4364. The 4364 G to A is in the intron two and in the overlapping region of amplicons 4 and 5, so it was confirmed both in DHPLC and sequence in two different PCR products.

The five different variants that we discovered were distributed in four individuals in this manner:

One subject presents only 4818 G to A and one subject presents only 4364 G to A.

In one sample were found the –69 C to T and the 4502 C to T. In the last positive case were found in the meanwhile the 2392 C to G and the 4364 G to A.

DISCUSSION

Hepcidin, the long awaited iron hormone, is the principal downregulator of the transport of iron across the small intestine and the placenta, and its release from macrophages. The present view is that hepcidin downregulates iron efflux from the intestine and macrophages by binding FPN and, following complex internalisation, leads to FPN degradation. In HFE, TfR2, and HJV related HC, hepatic expression or serum/urine levels of HAMP are inappropriately low.42 43 53 54 Its expression in the liver is also significantly impaired in HFE, TfR2, and HJV knockout mice24 55–57 and hepatic deposition of iron in HFEKO animals can be prevented by hepcidin overexpression.58 These findings suggests a unifying pathogenic model for all forms of HC in which HFE, TfR2, and HJV are all independent but complimentary regulators of hepcidin synthesis in the liver (fig 1). Hypothetically, these three proteins may be important for sensing circulatory iron and turning on hepcidin gene transcription, albeit with different mechanisms and functional impact. As the central role in iron homeostasis and in hemochromatosis disease development of this small peptide was established, we decided to analyze the variability in the promoter and in the coding region in 84 Italian subjects that presented symptoms of iron overload and resulted negative for the three reported HFE mutations. In the screening study a total of 420 amplicons were analyzed and 5 different variants were detected. Of these 5 variants 3 were in non-coding region and two in were found in exons, but resulted synonymous mutations. These results reflect a picture of high degree of conservations, especially if we consider that the subject that entered in the study share some defects in iron metabolism pathways and are not representative of general population. In three different studies several rare variants in HAMP gene region are reported as causative for hemochromatosis disease. In our case we can not rule out a possible involvement of the variant detected in iron overload phenotype, but it is difficult to demonstrate. From our results we can conclude that even though hepcidin plays a pivotal role in iron homeostasis regulations the molecular variability of the HAMP gene give only a marginal contribution in iron overload phenotypes.

ACKNOWLEDGMENTS

1 Aboud A, Haile DJ (2000) A novel mammalian iron regulated protein involved in intracellular iron metabolism. J Biol Chem 275:19906–19912

2 Babitt JL, Huang FW, Wrighting DM, et al (2006) Bone morphogenetic protein signaling by hemojuvelin regulates hepcidin expression. Nat Genet 38:531–539

3 Bastin JM, Jones M, O'Callaghan CA, Schimanski L, Mason DY, Townsend AR (1998) Kupffer cell staining by an HFE-specific monoclonal antibody: implications for hereditary haemochromatosis. Br J Haematol 103:931–941

4 Batey RG, Lai Chung Fong P, Shamir S, Sherlock S (1980) A non-transferrin-bound serum iron in idiopathic hemochromatosis. Dig Dis Sci 25:340–346

5 Bridle KR, Frazer DM, Wilkins SJ et al (2003) Disrupted hepcidin regulation in HFE-associated haemochromatosis and the liver as a regulator of body iron homeostasis. Lancet 361:669–673

6 Brissot P, Wright TL, Ma WL, Weisiger RA (1985) Efficient clearance of non-transferrin-bound iron by rat liver. Implications for hepatic iron loading in iron overload states. J Clin Invest 76:1463–1470

7 Brittenham GM (1994) The red cell cycle. In: Brock JH, Halliday JW, Pippard MJ, Powell LW (eds) Iron metabolism in health and disease. WB Saunders Company Ltd, London, pp 31–62

8 Camaschella C, Roetto A, Cali A, et al (2000) The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. Nat Genet 25:14–15

9 Donovan A, Brownlie A, Zhou Y et al (2000) Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. Nature 403:779–781

10 Donovan A, Lima CA, Pinkus JL et al (2005) The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. Cell Metab 1:191–200

11 Feder JN, Gnirke A, Thomas W et al (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

12 Fleming RE, Britton RS, Waheed A, Sly WS, Bacon BR (2005) Pathophysiology of hereditary hemochromatosis. Semin Liver Dis 25:411–419

13 Frazer DM and Anderson GJ (2003) The orchestration of body iron intake: how and where do enterocytes receive their cues?. Blood Cells Mol Dis 30:288–297

14 Frazer DM, Inglis HR, Wilkins SJ, et al (2004a) Delayed hepcidin response explains the lag period in iron absorption following a stimulus to increase erythropoiesis. Gut 53:1509–1515

15 Frazer DM, Vulpe CD, McKie AT, et al (2001) Cloning and gastrointestinal expression of rat hephaestin: relationship to other iron transport proteins. Am J Physiol 281:G931–G939

16 Frazer DM, Wilkins, SJ, Becker EM et al (2002) Hepcidin expression inversely correlates with the expression of duodenal iron transporters and iron absorption in rats. Gastroenterology 123:835–844

17 Frazer DM, Wilkins SJ, Millard KN, McKie AT, Vulpe CD, Anderson GJ (2004b) Increased hepcidin expression and hypoferraemia associated with an acute phase response are not affected by inactivation of HFE. Br J Haematol 126:434–436

18 Giannetti AM, Bjorkman PJ (2004) HFE and transferrin directly compete for transferrin receptor in solution and at the cell surface. J Biol Chem 279:25866–25875

19 Heinrich PC, Behrmann I, Haan S, Hermanns HM, Muller-Newen G, Schaper F (2003) Principles of interleukin (IL)-6-type cytokine signalling and its regulation. Biochem J 374:1–20

20 Huang FW, Pinkus JL, Pinkus GS, Fleming MD, Andrews NC (2005) A mouse model of juvenile hemochromatosis. J Clin Invest 115:2187–2191

21 Huebers HA, Finch CA (1987) The physiology of transferrin and transferrin receptors. Physiol Rev 67:520–582

22 Johnson MB, Enns CA (2004) Diferric transferrin regulates transferrin receptor 2 protein stability. Blood 104:4287–4293

23 Kawabata H, Germain RS, Vuong PT, Nakamaki T, Said JW, Koeffler HP (2000) Transferrin receptor 2-a supports cell growth both in iron-chelated cultured cells and in vivo. J Biol Chem 275:16618–16625

24 Lee P, Peng H, Gelbart T, Beutler E (2004) The IL-6- and lipopolysaccharide-induced transcription of hepcidin in HFE-, transferrin receptor 2-, and beta 2-microglobulin-deficient hepatocytes. Proc Natl Acad Sci USA 101:9263–9265

25 Lee P, Peng H, Gelbart T, Wang L, Beutler E (2005) Regulation of hepcidin transcription by interleukin-1 and interleukin-6. Proc Natl Acad Sci USA 102:1906–1910

26 Lim JE, Jin O, Bennett C et al (2005) A mutation in Sec15l1 causes anemia in hemoglobin deficit (hbd) mice. Nat Genet 37:1270–1273

27 Lim JE, Jin O, Bennett C et al (2005) A mutation in Sec15l1 causes anemia in hemoglobin deficit (hbd) mice. Nat Genet 37:1270–1273

28 Lin L, Goldberg YP, Ganz T (2005) Competitive regulation of hepcidin mRNA by soluble and cell-associated hemojuvelin. Blood 106:2884–2889

29 Lou DQ, Lesbordes JC, Nicolas G, et al (2005) Iron- and inflammation-induced hepcidin gene expression in mice is not mediated by Kupffer cells in vivo. Hepatology 41:1056–1064

30 Makui H, Soares RJ, Jiang W, Constante M, Santos MM (2005) Contribution of Hfe expression in macrophages to the regulation of hepatic hepcidin levels and iron loading. Blood 106:2189–2195

31 McKie AT, Marciani P, Rolfs A et al (2000) A novel duodenal iron-regulated transporter, IREG1, implicated in the basolateral transfer of iron to the circulation. Mol Cell 5:299–309

32 Montosi G, Corradini E, Garuti C et al (2005) Kupffer cells and macrophages are not required for hepatic hepcidin activation during iron overload. Hepatology 41:545–552

33 Nakashima K, Yanagisawa M, Arakawa H, et al (1999) Synergistic signaling in fetal brain by STAT3–Smad1 complex bridged by p300. Science 284:479–482

34 Nemeth E, Rivera S, Gabayan V et al (2004a) IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. J Clin Invest 113:1271–1276

35 Papanikolaou G, Samuels ME, Ludwig EH et al (2004) Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. Nat Genet 36:77–82

37 Parkkila S, Waheed A, Britton RS et al (1997) Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary hemochromatosis. Proc Natl Acad Sci USA 94:13198-13202

38 Peyssonnaux C, Zinkernagel AS, Datta V, Lauth X, Johnson RS, Nizet V (2006) TLR4-dependent hepcidin expression by myeloid cells in response to bacterial pathogens. Blood 107:3727–3732

39 Pigeon C, Ilyin G, Courselaud B et al (2001) A new mouse liver-specific gene, encoding a protein homologous to human antimicrobial peptide hepcidin, is overexpressed during iron overload. J Biol Chem 276:7811–7819

40 Rivera S, Nemeth E, Gabayan V, Lopez MA, Farshidi D, Ganz T (2005) Synthetic hepcidin causes rapid dosedependent hypoferremia and is concentrated in ferroportin-containing organs. Blood 106:2196–2199

41 Robb A, Wessling-Resnick M (2004) Regulation of transferrin receptor 2 protein levels by transferrin. Blood 104:4294–4299

42 Roy CN, Andrews NC (2005) Anemia of inflammation: the hepcidin link. Curr Opin Hematol 12:107–111

43 Truksa J, Peng H, Lee P, Beutler E (2006) Bone morphogenetic proteins 2, 4, and 9 stimulate murine hepcidin 1 expression independently of Hfe, transferrin receptor 2 (Tfr2), and IL-6. Proc Natl Acad Sci

44 USA Jun 26, Epub ahead of print Wang RH, Li C, Xu X et al (2005) A role of SMAD4 in iron metabolism through the positive regulation of hepcidin expression. Cell Metab 2:399–409

45 Weiss G, Goodnough LT (2005) Anemia of chronic disease. N Engl J Med 352:1011–1023

46 West AP, Giannetti AM, Herr AB, et al (2001) Mutational analysis of the transferrin receptor reveals overlapping HFE and transferrin binding sites. J Mol Biol 313:385–387

47 Wilkins SJ, Frazer DM, Millard KN, McLaren GD, Anderson GJ (2006) Iron metabolism in the hemoglobin-deficit mouse: correlation of diferric transferrin with hepcidin expression. Blood 107:1659–1664

48 Zhang AS, West AP Jr, Wyman AE, Bjorkman PJ, Enns CA (2005) Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. J Biol Chem 280:33885–33894

49 Zhang A, Xiong S, Tsukamoto H, Enns CA (2004) Localization of iron metabolism-related mRNAs in rat liver indicate that HFE is expressed predominantly in hepatocytes. Blood 103:1509–1514

50 Zhang AS, Sheftel AD, Ponka P (2006) The anemia of ‘‘haemoglobin-deficit’’ (hbd/hbd) mice is caused by a defect in transferrin cycling. Exp Hematol 34:593–598

51 von Recklinghausen FD. Uber Haemochromatose. Taggeblatt der (62) Versammlung deutscher. Naturforscher and Aerzte in Heidelberg 1889:324–25.

52 Trousseau A. Glycosurie; diabete sucre. Clinique Med de l’Hotel de Paris 1865;2:663–98.

53 Troisier M. Diabe`te sucre´. Bull Soc Anatomique Paris 1871;44:231–5.

54 Sheldon J. Haemochromatosis. London: Oxford University Press, 1935.

55 Simon M, Pawlotsky Y, Bourel M, et al. Idiopathic haemochromatosis associated with HL-A 3 tissular antigen] Nouv Presse Med 1975;4:1432.

56 Simon M, Le Mignon L, Fauchet R, et al. A study of 609 HLA haplotypes marking for the hemochromatosis gene: (1) mapping of the gene near the HLA-A locus and characters required to define a heterozygous population and (2) hypothesis concerning

the underlying cause of hemochromatosis-HLA association. Am J Hum Genet 1987;41:89–105.

57 Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 1996;13:399–408.

58 Pietrangelo A. Hemochromatosis 1998: is one gene enough? J Hepatol 1998;29:502–9.

59 Pietrangelo A. Hereditary hemochromatosis—a new look at an old disease.N Engl J Med 2004;350:2383–97.

60 Lebron JA, Bennett MJ, Vaughn DE, et al. Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. Cell 1998;93:111–23.

61 Rohrlich PS, Fazilleau N, Ginhoux F, et al. Direct recognition by alphabeta cytolytic T cells of Hfe, a MHC class Ib molecule without antigen-presenting function. Proc Natl Acad Sci U S A 2005;102:12855–60.

62 Feder JN, Penny DM, Irrinki A, et al. The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. Proc Natl Acad Sci U S A 1998;95:1472–7.

63 Gross CN, Irrinki A, Feder JN, et al. Co-trafficking of HFE, a nonclassical major histocompatibility complex class I protein, with the transferrin receptor implies a role in intracellular iron regulation. J Biol Chem 1998;273:22068–74.

64 Zhang AS, Davies PS, Carlson HL, et al. Mechanisms of HFE-induced regulation of iron homeostasis: Insights from the W81A HFE mutation. Proc Natl Acad Sci U S A 2003;100:9500–5.

65 Davies PS, Zhang AS, Anderson EL, et al. Evidence for the interaction of the hereditary haemochromatosis protein, HFE, with the transferrin receptor in endocytic compartments. Biochem J 2003;373:145–53.

66 Carlson H, Zhang AS, Fleming WH, et al. The hereditary hemochromatosis protein, HFE, lowers intracellular iron levels independently of transferrin receptor 1 in TRVb cells. Blood 2005;105:2564–70.

67 Waheed A, Parkkila S, Zhou XY, et al. Hereditary hemochromatosis: Effects of C282Y and H63D mutations on association with beta(2)-microglobulin, intracellular processing, and cell surface expression of the HFE protein in COS-7 cells. Proc Natl Acad Sci U S A 1997;94:12384–9.

68 Kawabata H, Yang R, Hirama T, et al. Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family. J Biol Chem 1999;274:20826–32.

69 Kawabata H, Germain RS, Vuong PT, et al. Transferrin receptor 2-alpha supports cell growth both in iron-chelated cultured cells and in vivo. J Biol Chem 2000;275:16618–25.

70 West AP Jr, Bennett MJ, Sellers VM, et al. Comparison of the interactions of transferrin receptor and transferrin receptor 2 with transferrin and the hereditary hemochromatosis protein HFE. J Biol Chem 2000;275:38135–8.

71 Camaschella C, Roetto A, Cali A, et al. The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. Nat Genet 2000;25:14–15.

72 Fleming RE, Ahmann JR, Migas MC, et al. Targeted mutagenesis of the murine transferrin receptor-2 gene produces hemochromatosis. Proc Natl Acad Sci U S A 2002;99:10653–8.

73 Papanikolaou G, Samuels ME, Ludwig EH, et al. Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. Nat Genet 2004;36:77–82.

74 Huang FW, Pinkus JL, Pinkus GS, et al. A mouse model of juvenile hemochromatosis. J Clin Invest 2005;115:2187–91.

75 Lin L, Goldberg YP, Ganz T. Competitive regulation of hepcidin mRNA by soluble and cell-associated hemojuvelin. Blood 2005;106:2884–9.

76 Krause A, Neitz S, Magert HJ, et al. LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity. FEBS Lett 2000;480:147–50.

77 Park CH, Valore EV, Waring AJ, et al. Hepcidin, a urinary antimicrobial peptide synthesized in the liver. J Biol Chem 2001;276:7806–10.

78 Pigeon C, Ilyin G, Courselaud B, et al. A new mouse liver-specific gene, encoding a protein homologous to human antimicrobial peptide hepcidin, is overexpressed during iron overload. J Biol Chem 2001;276:7811–19.

79 Nicolas G, Bennoun M, Porteu A, et al. Severe iron deficiency anaemia in transgenic mice expressing liver hepcidin. Proc Natl Acad Sci U S A 2002;99:4596–601.

80 Hunter HN, Fulton DB, Ganz T, et al. The solution structure of human hepcidin, a peptide-hormone with antimicrobial activity that is involved in iron uptake and hereditary hemochromatosis. J Biol Chem 2002;277:37597–603.

81 Laftah AH, Ramesh B, Simpson RJ, et al. Effect of hepcidin on intestinal iron absorption in mice. Blood 2004;103:3940–4.

82 Rivera S, Nemeth E, Gabayan V, et al. Synthetic hepcidin causes rapid dosedependent hypoferremia and is concentrated in ferroportin-containing organs. Blood 2005;106:2196–9.

83 Nemeth E, Tuttle MS, Powelson J, et al. Hepcidin regulates iron efflux by binding to ferroportin and inducing its internalization. Science 2004;306:2090–93.

84 Nicolas G, Chauvet C, Viatte L, et al. The gene encoding the iron regulatory peptide hepcidin is regulated by anaemia, hypoxia, and inflammation. J Clin Invest 2002;110:1037–44.

85 Nemeth E, Valore EV, Territo M, et al. Hepcidin, a putative mediator of anemia of inflammation, is a type II acute-phase protein. Blood 2003;101:2461–3.

86 Nemeth E, Rivera S, Gabayan V, et al. IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. J Clin Invest 2004;113:1271–6.

87 Lee P, Peng H, Gelbart T, et al. The IL-6- and lipopolysaccharide-induced transcription of hepcidin in HFE-, transferrin receptor 2-, and beta2-microglobulin-deficient hepatocytes. Proc Natl Acad Sci U S A 2004;101:9263–5.

88 Roy CN, Custodio AO, de Graaf J, et al. An Hfe-dependent pathway mediates hyposideremia in response to lipopolysaccharide-induced inflammation in mice. Nat Genet 2004;36:481–5.

89 Frazer DM, Wilkins SJ, Millard KN, et al. Increased hepcidin expression and hypoferraemia associated with an acute phase response are not affected by inactivation of HFE. Br J Haematol 2004;126:434–6.

90 Montosi G, Corradini E, Garuti C, et al. Kupffer cells and macrophages are not required for hepatic hepcidin activation during iron overload. Hepatology 2005;41:545–52.

91 Ganz T. Hepcidin, a key regulator of iron metabolism and mediator of anemia of inflammation. Blood 2003;102:783–8.

92 Bridle KR, Frazer DM, Wilkins SJ, et al. Disrupted hepcidin regulation in HFEassociated haemochromatosis and the liver as a regulator of body iron homoeostasis. Lancet 2003;361:669–73.

93 Gehrke SG, Kulaksiz H, Herrmann T, et al. Expression of hepcidin in hereditary hemochromatosis: evidence for a regulation in response to serum transferrin saturation and non-transferrin-bound iron. Blood 2003;102:371–6.

94 Lou DQ, Lesbordes JC, Nicolas G, et al. Iron- and inflammation-induced hepcidin gene expression in mice is not mediated by Kupffer cells in vivo. Hepatology 2005;41:1056–64.

95 Wang RH, Li C, et al. A role of SMAD4 in iron metabolism through the positive regulation of hepcidin expression. Cell Metab 2005;2:2399–409.

96 Nicolas G, Bennoun M, Devaux I, et al. Lack of hepcidin gene expression and severe tissue iron overload in upstream stimulatory factor 2 (USF2) knockout mice. Proc Natl Acad Sci U S A 2001;98:8780–5.

97 Courselaud B, Pigeon C, Inoue Y, et al. C/EBPalpha regulates hepatic transcription of hepcidin, an antimicrobial peptide and regulator of iron metabolism. Cross-talk between C/EBP pathway and iron metabolism. J Biol Chem 2002;277:41163–70.

98 Roetto A, Papanikolaou G, Politou M, et al. Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. Nat Genet 2003;33:21–2.

99 Lynch SR, Skikne BS, Cook JD. Food iron absorption in idiopathic hemochromatosis. Blood 1989;74:2187–93.

100 Smith PM, Godfrey BE, Williams R. Iron absorption in idiopathic haemochromatosis and its measurement using a whole-body counter. Clin Sci 1969;37:519–31.

101 Crosby WH. The control of iron balance by the intestinal mucosa. Blood 1963;22:441–9.

102 Bothwell TH. Overview and mechanisms of iron regulation. Nutr Rev 1995;53:237–45.

103 Nemeth E, Roetto A, Garozzo G, et al. Hepcidin is decreased in TFR2-hemochromatosis. Blood 2004;105:1803–6.

104 Papanikolaou G, Tzilianos M, Christakis JI, et al. Hepcidin in iron overload disorders. Blood 2005;105:4103–5.

105 Muckenthaler M, Roy CN, Custodio AO, et al. Regulatory defects in liver and intestine implicate abnormal hepcidin and Cybrd1 expression in mouse hemochromatosis. Nat Genet 2003;34:102–7.

106 Kawabata H, Fleming RE, Gui D, et al. Expression of hepcidin is downregulated in TfR2 mutant mice manifesting a phenotype of hereditary hemochromatosis. Blood 2005;105:376–81.

107 Niederkofler V, Salie R, Arber S. Hemojuvelin is essential for dietary iron sensing, and its mutation leads to severe iron overload. J Clin Invest 2005;115:2180–6.

108 Nicolas G, Viatte L, Lou DQ, et al. Constitutive hepcidin expression prevents iron overload in a mouse model of hemochromatosis. Nat Genet2003;34:97–101.

109 Pietrangelo A, Caleffi A, Henrion J, et al. Juvenile hemochromatosis associated with pathogenic mutations of adult hemochromatosis genes. Gastroenterology 2005;128:470–9.

110 Pietrangelo A. The ferroportin disease. Blood Cells Mol Dis 2004;32:131–8.

111 OMIM. Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/ entrez/query.fcgi?db =OMIM (accessed 16 January 2006).

CHAPTER 2
MOLECULAR VARIABILITY IN THE CHROMOSOMIC REGION OF HAPTOGLOBIN GENE

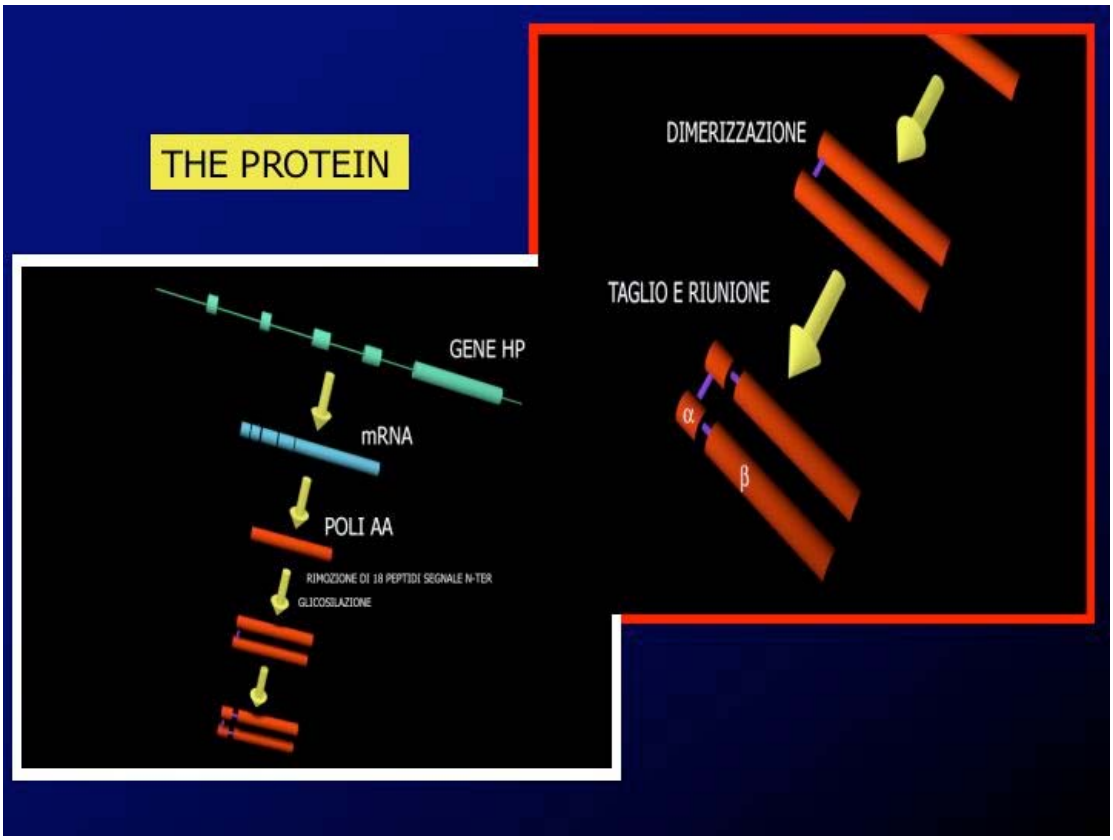
INTRODUCTION

Haptoglobin (Hp) is a plasma protein with hemoglobin-binding capacity. It is a well-known marker of hemolysis. Hp is also an acute-phase protein that functions as a bacteriostatic agent, an inhibitor of prostaglandin synthesis and angiogenesis [1]. However, the best-known biological function of Hp is capture of hemoglobin (Hb). After destruction of erythrocytes, free Hb in the circulation passes through the glomerular filter and renal damage may occur. The binding of Hp with Hb prevents both iron loss and kidney damage during intravascular hemolysis [2].

The identification of functional differences in haptoglobin molecules resulting from relatively common polymorphisms has further elucidated the importance of haptoglobin in iron homeostasis and in disease processes influenced by iron metabolism including hemochromatosis atherosclerosis and cardiovascular disease, and infectious diseases.

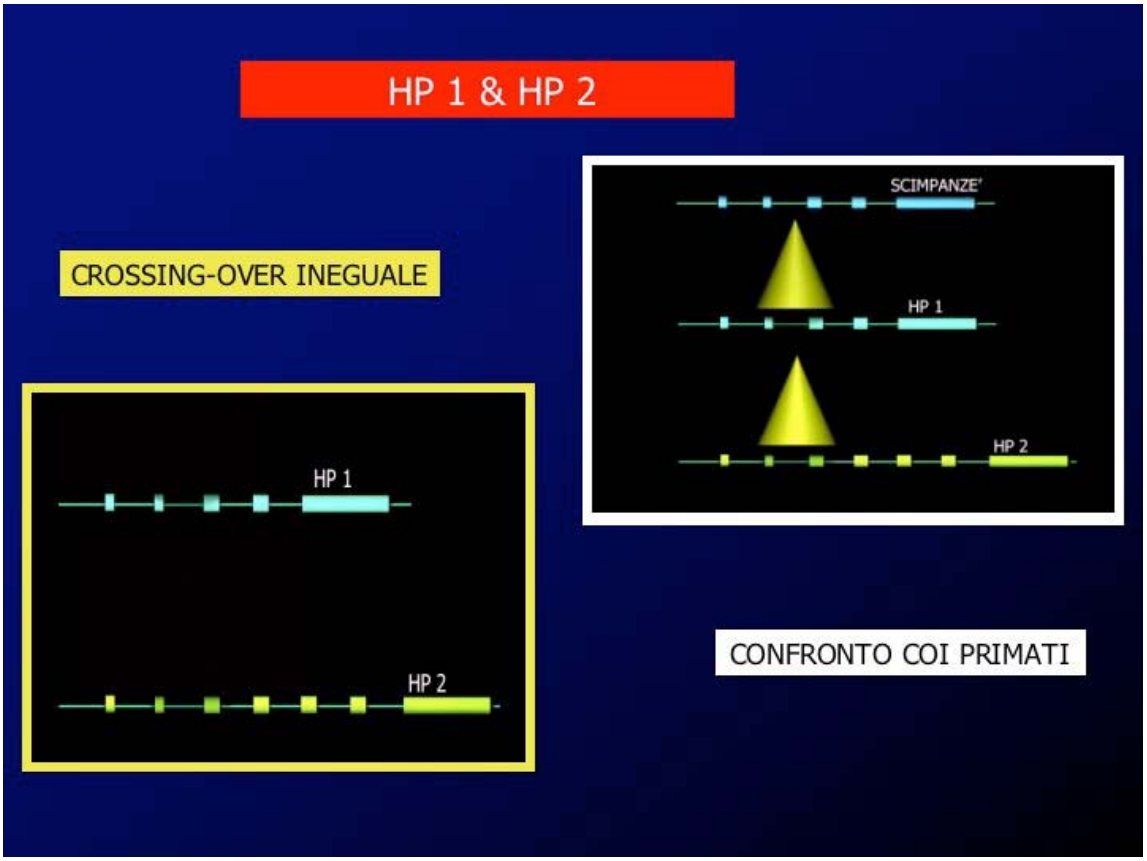
HAPTOGLOBIN POLYMORPHISMS AND DIFFERENCE IN PROPERTIES BETWEEN THE PHENOTYPES

Haptoglobin (Hp) is an α2-sialoglycoprotein with hemoglobin (Hb)-binding capacity [3 and 4] and is characterized by a molecular heterogeneity with three major genotypes: Hp 1-1, Hp 2-1 and Hp 2-2 [2, 3, 4 and 5].



These genotypes are molecularly determined by two alleles: Hp1 and Hp2 [3 and 4]. The homozygote, Hp1/Hp1 shows a single fast-migrating Hp 1-1 protein band on starch gel electrophoresis. The homozygote Hp2/Hp2 shows a series of slower migrating bands. The heterozygote Hp1/Hp2 displays another series of slow bands and a weak Hp 1-1 band. Hp consists of two different polypeptide chains, the α-chain and the β-chain [3 and 4]. The β-chain (40 kDa) is heavier than the α-chain and is identical in all Hp types. The α-chain shows three major forms: α-1s, α-1f (s=slower, F=faster) and the slow migrating α-2-chain. The Hp 1-1 phenotype has α-1-chains, while α-2-chains are present in Hp from individuals with the Hp 2-1 or Hp 2-2 phenotype [3 and 4]. The loci involved for the Hp synthesis are located on chromosome 16q22. The Hp 1-1 protein is a small molecule (86 kDa) with formula (α1β)2. Heterozygote Hp 2-1, (α1β)2+(α2β)n (n=0, 1, 2,…), is characterized by polymerization. Hp 2-2 comprises higher molecular mass forms (>200 kDa) with formula (α2β)n (n=3, 4, 5,…) [5 and 7] .

The principal difference between alleles Hp 1 and Hp 2 is the presence of a duplicated DNA segment of 1700 bp in Hp 2 but not Hp 1 (2). Most likely, formation of the Hp 2 allele is the result of a breakage and reunion event at non-homologous positions within the fourth and second introns of two Hp 1 genes. As a consequence of this illegitimate crossing-over event, exons 5 and 6 of allele Hp 2 originate from exons 3 and 4, respectively, of one of these Hp 1 genes



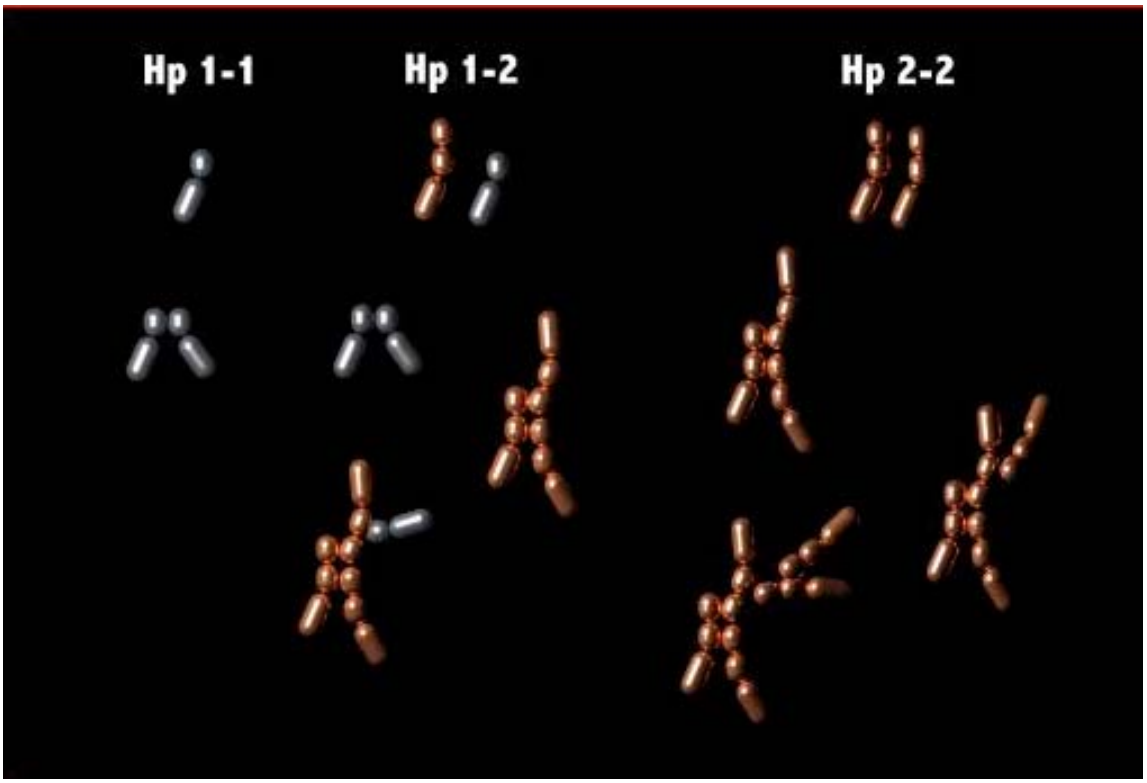
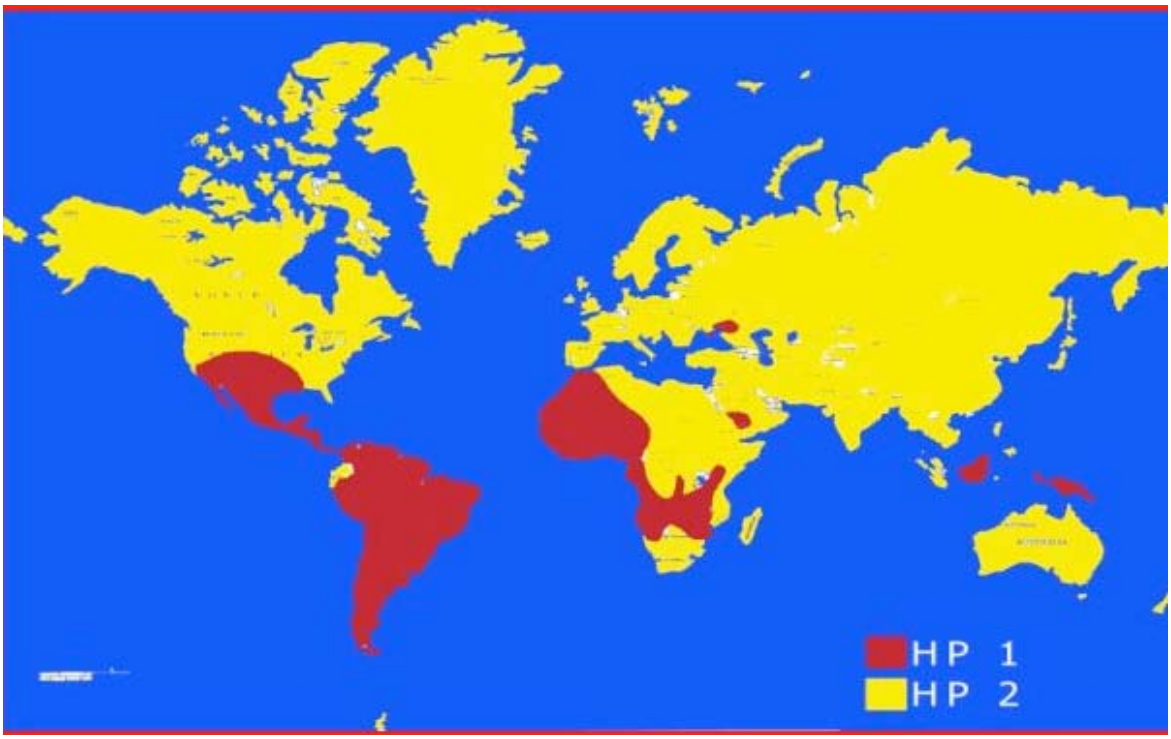


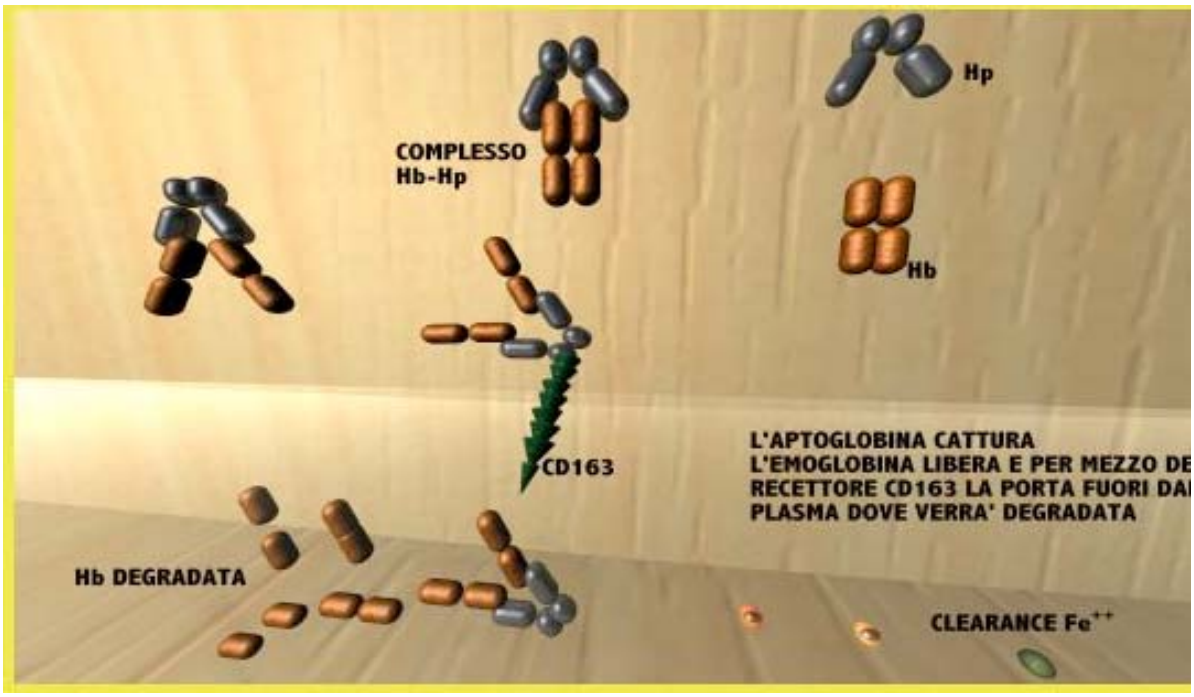
Fig. 1.
Iconographic presentation of the differences in molecular weight between the different haptoglobin phenotypes.

The synthesis of Hp is considerably lower in fetal than in adult liver [3]. The hepatic synthesis of Hp is induced by cytokines such as interleukin-6, interleukin-1 and tumor necrosis factor [3 and 8]. The haptoglobin concentration is Hp phenotype-dependent. The reference range for haptoglobin concentration is lower in individuals carrying the Hp 2-2 phenotype than individuals carrying the Hp 1-1 and Hp 2-1 phenotype [9].
The haptoglobin phenotype distribution differs according to geographical localisation of the population studied [1]. The haptoglobin allele frequencies show marked geographical differences, with the lowest Hp1 allele frequency (0.10) in Southeast Asia and the greatest Hp1 frequency (0.80) in indigenous populations of South America [1]. The phenotypic distribution in European populations shows that 15% individuals are Hp 1-1, 50% Hp 2-1, and 35% Hp 2-2, corresponding with a Hp1 allele frequency of 0.40 [1 and 10]



Haptoglobin polymorphisms, directly or indirectly, influence pathways involved in iron metabolism:

Hp forms a soluble complex with Hb. The binding of Hp with Hb is the strongest known noncovalent interaction among the plasma transport proteins, with the complex having a very high affinity and stability [11]. Circulating Hp is saturated when 500–1500 mg/l free Hb is present.



The half-life of Hb–Hp complexes in plasma is 20 min [12]. Hepatocellular uptake of Hp–Hb complexes reduces the loss of haem iron through the kidney [1]. Unlike hemopexin and transferrin, Hp is not recycled after endocytosis but the Hb–Hp complex is instead degraded by lysosomes [13]. Hb binding depends not only on the serum concentration of Hp but also on the Hp phenotype [14]. The hemoglobin binding capacity is lowest among Hp 2-2 subjects [15] due to lower serum concentrations, as discussed above, as well as to a lower ability to bind Hb [11]. After destruction of erythrocytes, free Hb in the circulation passes through the glomerular filter and renal damage may occur. Hp reduces the loss of Hb and iron, because the Hb–Hp complex is not filtered through the glomeruli but is transported to the liver. The lower binding capacity of Hb in individuals with the Hp 2-2 phenotype results in more renal damage and higher serum iron levels [1].

Free Hb promotes the accumulation of hydroxyl radicals [16] and harmful reactive oxygen species (free radicals), because iron (Fe2+) can generate extremely reactive hydroxyl radicals in the presence of H2O2 (Fenton reaction). Haem iron catalyses the oxidation of low-density lipoproteins, which can damage vascular endothelial cells [17]. In addition, the breakdown of erythrocytes in the interstitial fluid results in Hb-mediated hydroxyl radical formation. Plasma Hp can be regarded as a major antioxidant protecting against Hb-driven lipid peroxidation [18 and 19]. The ability of haptoglobin to reduce hemoglobin-induced free radical damage is phenotype-dependent [14 and 18].

The distribution of highly polymeric Hp 2-2 proteins in extravascular fluids is restricted by their molecular mass [5]. Consequently, the antioxidative capacity of body fluids is less effective in Hp 2-2 individuals [5].

An increased frequency of Hp 2-2 and Hp0 was observed in familial and posttraumatic epilepsy and was explained by a less efficient inhibition of Hb-driven brain-lipid peroxidation after hemorrhage within the central nervous system [22].

CD163 has been identified as the monocyte–macrophage receptor that binds the Hb–Hp complex. Hp and Hb do not bind to the CD163 receptor separately.

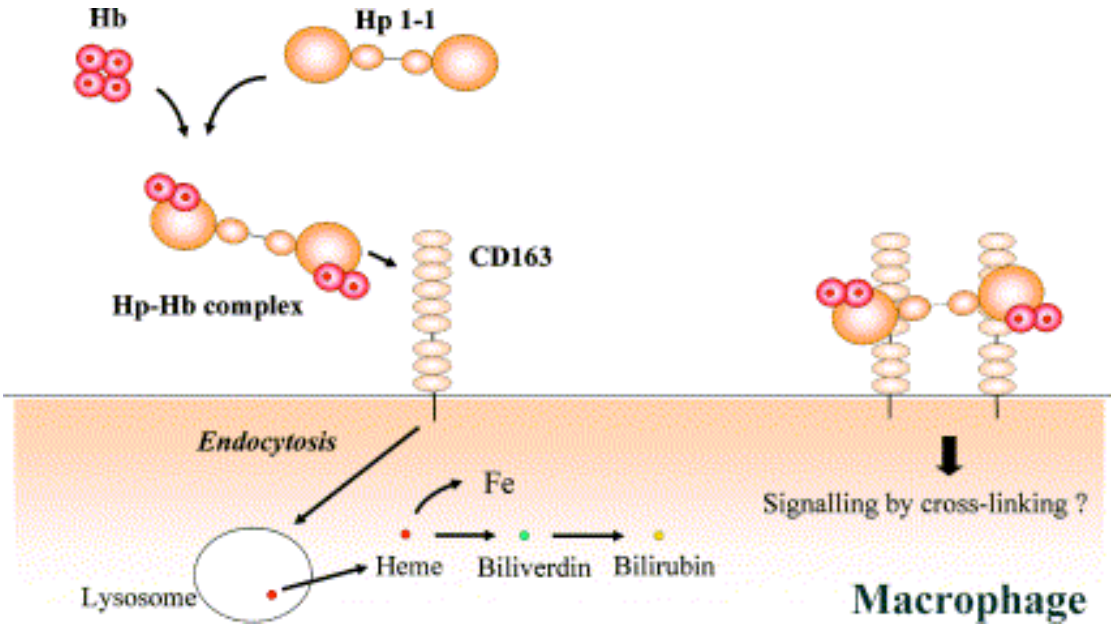


Fig. 2. Schematic outline of the Hp–Hb scavenging function of CD163/HbSR. It is feasible to envision that binding of Hp–Hb by CD163/HbSR on macrophages can lead to cross-linking. Only the Hp 1-1 phenotype is shown in the figure.

CD163 binds only hemoglobin and haptoglobin in complex, which indicates the exposure of a receptor-binding neoepitope. In its function as a Hb scavenger, the CD163 receptor accounts for a substantial transfer of iron into the macrophages. Complexes of hemoglobin and multimeric Hp 2-2 haptoglobin exhibit higher functional affinity for CD163 than do complexes of hemoglobin and dimeric Hp 1-1 haptoglobin [23]. In vitro experiments with radiolabeled (125I) Hb showed that human peripheral blood monocytes take up Hb–Hp 2-2 complexes whereas, even in zymosan-activated monocytes, free Hb and Hb bound to Hp 1-1 or 2-1 are not internalised [24]. When intracellular concentrations of haem increase with the endocytosis of Hb–Hp 2-2 complexes, there is a rapid induction of ferritin synthesis. In mammals, the cytosolic ferritins are ubiquitous and made of two subunit types, the H- and L-chains, with about

50% sequence identity and very similar three-dimensional (3D) structures. H-chains have ferroxidase activity, which accelerates Fe(II) oxidation, the rate-limiting step of ferritin iron incorporation, in a reaction that consumes one dioxygen molecule per two Fe(II) ions with the production of hydrogen peroxide. The L-subunit has no catalytic activity on its own, but it assists the activity of the H-subunits by offering sites for iron nucleation and mineralisation and increasing the turnover at the ferroxidase centers [25]. Higher cytosolic L-ferritin levels are present in peripheral blood monocytes from Hp 2-2 subjects (687±152 _g/g protein) compared to Hp 1-1 and 2-1 subjects (326±83 and 366±109 _g/g protein, respectively) [24]. This effect is presumably due to iron released from haem affecting the iron regulatory protein (IRP) which regulates ferritin mRNA translation by binding to the iron responsive element (IRE) [26 and 27]. Cytosolic H-ferritin content in monocytes is not different between Hp phenotypes, similar to what is observed in iron-loaded livers where only the L-ferritin form is upregulated [28]. In males, but not in females, the Hp 2-2 phenotype is associated with higher serum iron, transferrin saturation and ferritin concentrations than in the Hp 1-1 and Hp 2-1 phenotype, whereas soluble transferrin concentrations are lower. Serum ferritin correlated with monocyte L-ferritin content which is also highest in the male Hp 2-2 subgroup [24]. A positive correlation was observed between serum and monocyte ferritin levels suggesting that increased L-ferritin synthesis in monocyte–macrophages results in higher ferritin levels in the circulation [24].

1.2.4. Influence on the immune system

Haptoglobin has been proposed to be involved in a highly interactive ensemble of lymphocytes, neutrophils, and monocytes participating in inflammatory processes [29, 30, 31, 32, 33 and 34]. Kristiansen et al. [23] speculated that the Hp–Hb complex, like antibodies, may cross-link several CD163 molecules on the surface of macrophages, triggering an internal signalling cascade that results in increased secretion of anti-inflammatory cytokines. Human serum from Hp 2-2 and Hp 2-1 individuals agglutinates the bacterial cells of the Streptococcus pyogenes group A strain, carrying the T4 antigen. The Hp 2-2 serum has higher agglutination titres than the Hp 2-1 serum. In contrast, serum from individuals with the Hp 1-1 phenotype has no agglutination effect. Hp is not a true antibody because it does not possess the highly variable antigen-binding sites characteristic of the Fab moiety of immunoglobulins. The agglutination is probably mediated via binding with lectin-like structures [35 and 36].

Comparison of reference values for lymphocyte subsets in the peripheral blood and bone marrow show significant differences between haptoglobin phenotypes. Individuals with the Hp 2-2 phenotype have higher peripheral B-lymphocyte counts and CD4+ T lymphocytes counts than individuals with the Hp 1-1 phenotype. In contrast, in the bone marrow, CD4+ T cell percentages are high but B cell percentages are low in individuals with the Hp 2-2 phenotype. Flow cytometric analysis demonstrates that Hp binds to the CD22 receptor on human B lymphocytes. Although the affinity of the binding is the same for the three phenotypes, the number of free CD22 binding sites in the circulation is estimated to be higher in Hp 2-2 individuals. No significant Hp binding has been detected for T cells and NK cells [33].

A controversial issue is the association of increased body iron stores with increased cardiovascular risk [43]. In this context, Hp polymorphism has been proposed as a risk factor for developing atherosclerotic vascular disease, independent of the classical risk factors such as smoking, hypertension, diabetes mellitus, or serum lipid concentrations. In essential hypertension, coronary artery lesions and target organ damage are more common among Hp 2-2 patients. It is suggested that hypertensives with a Hp 2-2 phenotype need more complex combinations of antihypertensive drugs to reduce blood pressure than individuals with the Hp 1-1 phenotype. The hypertensive patient carrying Hp 2-2 is more likely to accumulate atherosclerotic lesions of the coronary or peripheral arteries, despite comparable lipid levels, smoking habits and BMI [44]. When compared with individuals carrying the Hp 1-1 phenotype, males carrying the Hp 2-2 phenotype are at risk for developing premature coronary artery disease [45] and peripheral vascular disease [46]. Following acute myocardial infarction, the severity of infarction and outcome are worse for the Hp 2-2 patients [47]. Diabetic patients who are homozygous for the Hp1 allele are provided increased protection against the development of diabetic vascular complications such as nephropathy and retinopathy [48 and 49]. In a recent study, the Hp phenotype seems to be highly predictive of adverse cardiac events, particularly myocardial infarction, in the 1-year period after percutaneous transluminal coronary angioplasty in individuals with diabetes. Patients with the Hp 1-1 phenotype had a 0% risk of myocardial infarction, in contrast to patients carrying the Hp 2-2 phenotype who had a risk of 8.4% ($p<0.001$). No difference in risk was seen in individuals without diabetes [50]. The same authors demonstrated a severe impairment in the ability of the Hp 2-2 phenotype to prevent oxidation by glycosylated hemoglobin. They proposed that the specific interaction between diabetes, cardiovascular disease and

Hp phenotype is the result of the impaired clearing capacity of glycosylated hemoglobin–Hp complexes from the subendothelial space. A delay in the clearing of these complexes results in oxidation of low-density lipoprotein to atherogenic oxidized low-density lipoprotein [51].

Hp acts as a natural bacteriostat by preventing the utilisation of Hb by pathogenic bacteria which require iron for their growth [52]. The iron-restrictive environment in body fluids established by Hp–Hb binding is part of the nonspecific defense mechanism against bacterial invasion. Rats inoculated intraperitoneally with pathogenic Escherichia coli and Hb are fully protected against lethality by simultaneous administration of Hp [52]. L. pneumophila is an intracellular pathogen which has a definite requirement for iron, reflected by an inability to grow on medium in the absence of iron supplementation. In vitro, intracellular multiplication in macrophages is inhibited in the absence of iron. In a cohort of infected Dutch patients, 117 patients with Legionella disease (LD), 95 matched controls for age, gender and residency and 61 asymptomatic L. pneumophila seropositive controls were included. In this cohort, the Hp phenotype did not influence susceptibility for LD. However, Hp 2-1 was associated with severe LD, which in this cohort could not be explained by differences in iron status [53].

Clinical studies have demonstrated that Hp polymorphism may play a role in a number of viral infections including HIV and hepatitis C [10, 54, 55 and 56]. An initial study in HIV-1 infected adults demonstrated the Hp 2-2 phenotype was associated with: (i) higher mortality rates; (ii) lower median survival time since diagnosis of HIV seropositivity (Hp 2-2, 7.33; Hp 2-1/1-1, 11.0 years); (iii) higher baseline plasma HIV-1 RNA levels among antiretroviral drugs-naïve subjects (Hp 2-2, 5.26±0.82; Hp 2-1, 4.64±0.73; and Hp 1-1, 3.75±1.01 log10 RNA copies/ml); (iv) the greatest increase in plasma HIV-1 RNA levels over a 1-year period [55]; and (v) more accumulation of iron and oxidation of vitamin C, suggesting that less efficient protection against hemoglobin/iron-driven oxidative stress may be a direct mechanism for stimulating viral replication. Oxidative stress induced by reactive oxygen radicals stimulates HIV replication through the activation of the nuclear transcription factor-_B, and contributes to the development of cell damage and immunodeficiency [56 and 57]. A more recent case-control study revealed that Hp 2-2 was associated with significantly lower CD4+ levels among HIV-1-infected subjects in Ghana [58]. The Hp0 phenotype has been identified as a protective factor in HIV-1 seropositive Ghanaians [59].

In liver transplant patients following chronic hepatitis C infection, graft survival was shown to be dependent on donor Hp phenotype, Hp 2-2 being associated with lowest graft survival [60].

Although conflicting data are available, the haptoglobin polymorphism seems to influence iron recycling in healthy men and patients with hereditary hemochromatosis. This is demonstrated by the CD163 receptor mediated mechanism through which hemoglobin–Hp complexes are internalised by monocyte–macrophages. The uptake of hemoglobin–Hp complexed by the CD163 receptor is more potent for the hemoglobin–Hp 2-2 complex than for the complex carrying the Hp 1-1 phenotype. In addition, more evidence has become available demonstrating the deleterious effect of the Hp 2-2 phenotype in diabetic patients and the evolution toward complications of atherosclerotic processes. The haptoglobin polymorphism clearly influences outcome in the host after viral and bacterial infection. Differences in clinical presentation in patients with iron storage disease, atherosclerotic processes, and infections are explained, in part, by the different properties of the haptoglobin pheno/genotypes.

HAPTOGLOBIN AND MALARIA

In malaria-endemic countries of Africa, anaemia is very common in pregnant women and in children under five. Although anaemia is multifactorial—causative factors include iron deficiency and other nutritional deficiencies, helminth infection, and HIV—malaria is clearly an extremely important factor. Over half of malaria-related deaths are attributed to severe malaria anaemia (which is defined as malaria parasitaemia and a haemoglobin (Hb) concentration less than 50 g/l) [1]. Several antimalarial interventions have been shown to prevent anaemia, including insecticide treated nets, residual spraying, malaria chemoprophylaxis, and, more recently, intermittent presumptive treatment of infants (i. e., antimalarials coadministered with childhood immunization). Insecticide-treated nets have been shown to decrease all-cause mortality [2]. The pathogenesis of malaria anaemia remains incompletely understood. Dyserythropoiesis (disordered red cell development, which is, at least in part, due to inflammatory cytokines acting on erythroid precursors), intravascular hemolysis of infected red cells, and destruction of both parasitized and uninfected erythrocytes by splenic macrophages are all important [3,4]. Interestingly, it has been estimated that ten or more uninfected erythrocytes may be lost for each infected one [5], presumably because malaria infection alters uninfected erythrocytes. The probable causes of red cell loss include oxidation of band 3 (the anion transporter of the erythrocyte membrane) or membrane lipids, and deposition of IgG, complement, or immune complexes on the erythrocyte surface. Hp2-2 is less able than Hp1-1 to penetrate extravascular spaces, and Hp2-2 is less able to prevent oxidative damage or immune activation than the other phenotypes [7].

Given the importance of intravascular hemolysis in malaria infection, studies of Hp in malaria are important, but results to date have been inconclusive. Earlier studies based on phenotyping suggested Hp1-1 was associated with severe or symptomatic infection [8,9]. For example, in one study of 72 patients in Sudan with cerebral malaria, the percentage of patients with Hp phenotypes 1-1, 2-1, and 2-2 were 63.9%, 29.2%, and 6.9%, respectively [8]. But the low prevalence of Hp2-2 and the high prevalence of Hp1-1 in the patients in this study may reflect lack of sensitivity of the electrophoretic technique used in the study to detect the less abundant (and multiple) bands of Hp2-2 rather than the single abundant Hp1-1. Circulating levels of Hp1-1 are normally significantly higher than those of Hp2-2, but levels fall during malaria infection due to clearance of Hp/Hb complexes; age may influence the degree of fall in

Hp with malaria [10]. More recently, PCR-based genotyping studies have examined associations between Hp genotype and severe malaria or associations between Hp genotype and mild disease [11,12]. These studies did not find clear associations with Hp genotype; although, in one study, there was a suggestion of an association with severe malaria, especially severe anaemia. In a recent study Atkinson and colleagues have taken a different approach [13]. Working in The Gambia, where malaria transmission is intense but highly seasonal and anaemia is a common complication, they followed a group of children over the malaria season. The authors used Hb concentration as their main readout. Although baseline Hb levels were similar across Hp genotypes, children who were Hp2/2 had a greater fall in Hb (by 4 g/l) over the malaria season than did other children. Hp genotype significantly predicted Hb levels at the end of the malaria season in multivariate analysis, together with iron status and several other variables. A separate analysis examining factors influencing the magnitude of the fall in Hb gave similar results. Additionally, children with Hp2/2 genotype were more likely to be parasitaemic. The authors suggest that the effect of Hp genotype on the fall in Hb (and, thus, possible protection from anaemia) may be more significant than that of sickle cell trait (HbAS) or glucose-6-phosphate dehydrogenase deficiency. On the other hand, sickle cell trait and glucose- 6-phosphate dehydrogenase deficiency have been more compellingly associated with protection from severe malaria than have Hp genotypes [14]. How Does Hp Affect Hb Levels? How does Hp genotype influence Hb fall? The authors offer several interesting suggestions. Because destruction of uninfected erythrocytes may be especially important in development of malaria anaemia, this mechanism is one plausible route. Oxidative damage to uninfected cells might be more marked in Hp2-2 individuals since Hp2-2 proteins bind less efficiently to Hb, and since levels of Hp2-2 are lower and more easily depleted, increasing premature destruction of erythrocytes. Another reason Hp genotype may influence Hb levels might be because complexes of Hp2-2 with Hb, but not other Hp/Hb complexes, enter monocytes (which express less CD163 than macrophages), stimulating cytokine release by these circulating cells. Thus, different Hp types inducing different cytokine responses may affect the duration of inflammatory, marrow-suppressive cytokine production following malaria. Persistent marrow suppression, as judged by inappropriately low reticulocyte responses after resolution of infection, is a feature of malaria [3,4]. And Hp2/2 genotype has been associated with increased iron accumulation in haemochromatosis— in one study but not in others (reviewed in [7])—suggesting Hp genotype may have some role

in iron homeostasis. Interestingly, the 17% prevalence of Hp2/2 in Atkinson and colleagues' study was much lower than that which was reported in European populations [7], suggesting a possible protective role for Hp1 genotypes against malaria. Public Health Implications Is a 4 g/l difference in Hb fall clinically relevant? Probably. Although the effects of this fall on the proportion of children suffering moderate or severe anaemia are not given in the study, the population distribution of Hb levels is wide. A fall in mean Hb may translate into increased numbers of individuals with symptomatic disease, although this effect is not demonstrated in this study. From a public health perspective, interventions that result in a rise in Hb of similar magnitude—such as intermittent presumptive treatment in pregnancy [15]—are endorsed by the World Health Organization. And small changes in group levels may hide subgroups with more dramatic changes in Hb. How one might identify individuals at risk for the more dramatic falls in Hb, or whether antimalarial measures would prevent such falls, are unresolved questions.

HEREDITARY HEMOCHROMATOSIS

The phenotypic expression of hereditary hemochromatosis associated with the Cys282Tyr mutation of the HFE gene, varies from a fully penetrant clinical syndrome characterised by bronze pigmentation, cirrhosis, arthritis, endocrinopathy and cardiomyopathy to a mere laboratory finding of increased serum iron status [37]. This phenotypic heterogeneity could be influenced by additional non-genetic (environmental) causes, additional mutations or additional genes. The Hp gene has been proposed as one of the modifier genes in the phenotypic expression of HFE-linked hemochromatosis [38].

Three studies have examined the possible influence of the Hp gene on the phenotypic expression of HFE-linked hereditary hemochromatosis. In one study of 167 Cys282Tyr homozygous patients, identified as having hereditary hemochromatosis by clinical symptoms, a low Hp1 allele frequency (0.30) was found due to an overrepresentation of the Hp 2-2 type (49%), suggesting that the Hp 2-2 phenotype is associated with an accelerated iron overload [38]. Male patients carrying Hp 2-2 had higher serum ferritin levels (median: 1747 _g/l) than those with a Hp 1-1 or 2-1 type (median 1100 _g/l) ($p<0.03$). The volume of blood removed with phlebotomy to achieve iron depletion (defined as serum ferritin <50 _g/l) was also higher in Hp 2-2 patients ($p<0.03$) [38].

Two other studies have not shown an association [39 and 40]. Beutler et al. [40] found no overrepresentation of the Hp 2-2 phenotype in 115 hemochromatosis patients and the Hp polymorphism did not influence the iron status of patients and 117 healthy subjects. Carter et al. [39] found no influence of the Hp phenotype on transferrin saturation and ferritin concentration in 265 randomly selected control subjects, in 66 blood donors homozygous for the HFE Cys282Tyr mutation lacking clinical features of hemochromatosis, and in 68 patients presenting clinically with hemochromatosis, who were homozygous for Cys282Tyr. Both studies [39 and 40] concluded that Hp type neither influences iron status in normal subjects nor predicts clinical presentation of hereditary hemochromatosis in individuals from South Wales and the United States.

The difference between our study and the studies by Carter and Beutler can be explained by: (1) the moderate effect on iron homeostasis of the haptoglobin phenotype in comparison with the hemochromatosis gene. The rate of iron uptake through haptoglobin–hemoglobin complexes is relatively small as compared to major iron regulatory pathways. However, as in hemochromatosis, a long-term effect may have important consequences on iron status. This is further supported by the finding in liver

transplant patients, where donor Hp phenotype determines the iron status post-transplantation [41]. (2) Differences in patient selection and populations may also explain these contradictory findings. For example most Cys282Tyr homozygotes found by screening [40] were asymptomatic, whereas Cys282Tyr homozygotes in our study [38] were selected due to their clinical presentation of hemochromatosis. This is clearly reflected by serum iron markers in these trials. In Beutler's cohort, [40] low concentrations of serum ferritin (<100 _g/l) and serum transferrin saturation (<50%) were observed in a large number of cases. Higher values of serum ferritin and serum transferrin saturation in Cys282Tyr homozygous individuals were seen in our trial [38]. (3) It is possible that the conflicting results might reflect geographical differences between study populations. It is of interest, in this regard, that no relationship between serum iron indices and Hp phenotype could be found in reports from Africa [42]. If these geographical differences are the consequence of differences in genetic background or differences in the environment, is unclear.

MATERIALS AND METHODS

SAMPLES

A total of 96 unrelated probands between 7 and 47 years of Italian ancestry, with serum transferrin saturation values between 40% to 99% were selected. The probands were previously typed for the three most prevalent mutations related to hereditary hemochromatosis: C282Y, H63D, S65C in the Hfe gene, and resulted negative. The patients were referred from Clinical Hematologic, Interne Medicine from S. Orsola-Malpighi hospital of Bologna.

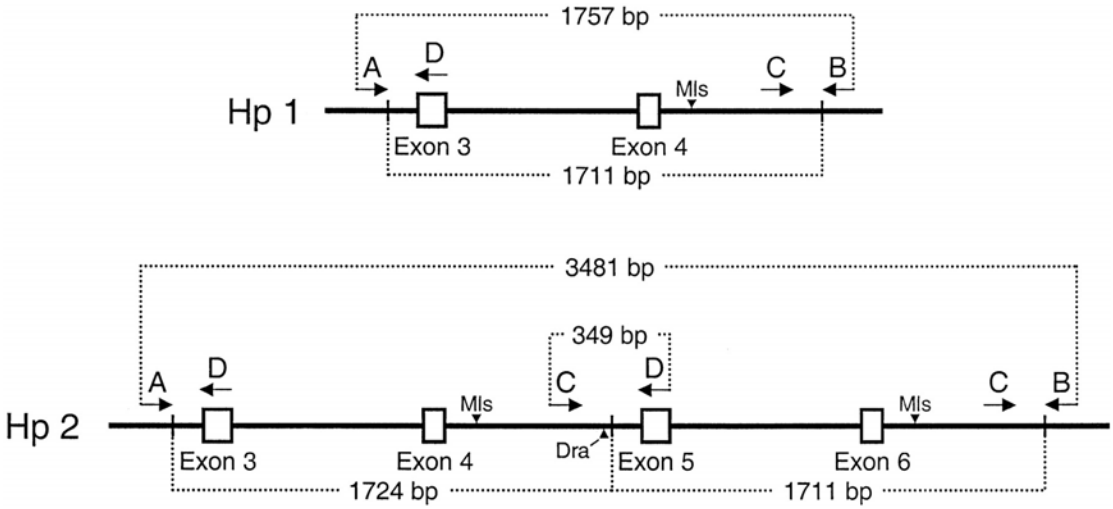
A total of 54 unrelated subjects were collected in Castelmassa village (Rovigo, Italiy), along Po river, in an area that until the 1930 was characterized by pandemic malaria.

Subjects were selected by age, individuals younger than 60 years old were excluded, this is to collect only individual that grew up when malaria was still acting in the area or immediately after its removal.

A total of 96 unrelated healthy subject of Italian origin were included in the study and used as control pool.

DUPLICATION GENOTYPING

DNA sequences specifying haptoglobin alleles Hp 1 and Hp 2 and the haptoglobin-related gene. The Hp 1- and Hp 2-specific sequences are contained in the EMBL/GenBank Data Libraries under accession numbers AC004682 and M69197, respectively (6)(7). In these sequences, the haptoglobin gene is represented by the allele subtypes Hp 1S and Hp 2FS (2). According to the sequence present in AC004682, the Hp 1-specific DNA region has a length of 1711 bp (Fig. 1); it extends, in an inverse orientation, from nucleotide position 188616 to nucleotide position 186906. In the sequence present in M69197, the 3435-bp Hp 2-specific DNA segment starts at position 2804 and ends at position 6238; it contains two units of similar sequences, consisting of 1724 and 1711 bp (Fig. 1). The sequences of the 1711-bp segments present in AC004682 and M69197 are complementary with the exception of one divergence located at position 188141 in AC004682, which corresponds to position 5003 in M69197. The sequence of the haptoglobin-related gene is also contained in M69197.



Hp 1 is represented by subtype Hp 1S, as indicated by the presence of the 1711-bp element. Hp 2 is represented by subtype Hp 2FS, as shown by the presence of the 1724-bp element followed by the 1711-bp element. The arrows, representing oligonucleotide primers A, and B are located at positions next to the binding sites of the primers within the DNA sequence. Also shown are the sizes of the PCR products obtained with primer pairs A/B (1757 and 3481 bp).

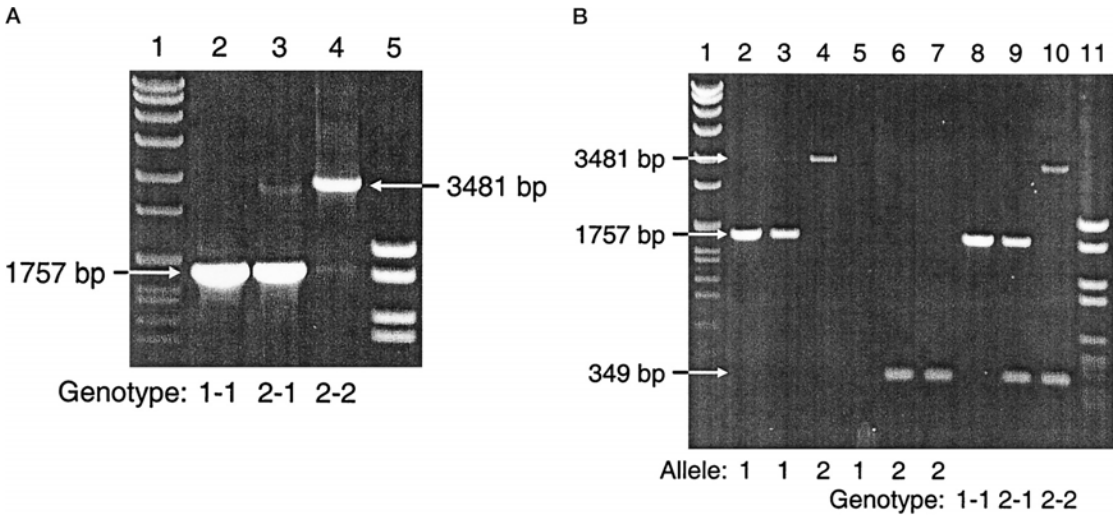
OLIGONUCLEOTIDE

Primers A (5'-GAGGGGAGCTTGCCTTTCCATTG-3') and B (5'-GAGATTTTTGAGCCCTGGCTGGT-3') were used for amplification of a 1757-bp Hp 1 allele-specific sequence and a 3481-bp Hp 2 allele-specific sequence (Fig. 1).

In Hp 1 and Hp 2, the annealing sites for primer A are located immediately upstream of the 1711-bp unit and the 1724-bp unit, respectively (Fig. 1). The nucleotide at the 5' end of primer A corresponds to position 188639 in AC004682 (Hp 1) and position 2781 in M69197 (Hp 2). Primer B has binding sites just downstream of the 1711-bp elements of Hp 1 and Hp 2 (Fig. 1). The nucleotide at the 5' end of primer B corresponds to position 186883 in AC004682 (Hp 1) and position 6261 in M69197 (Hp 2).

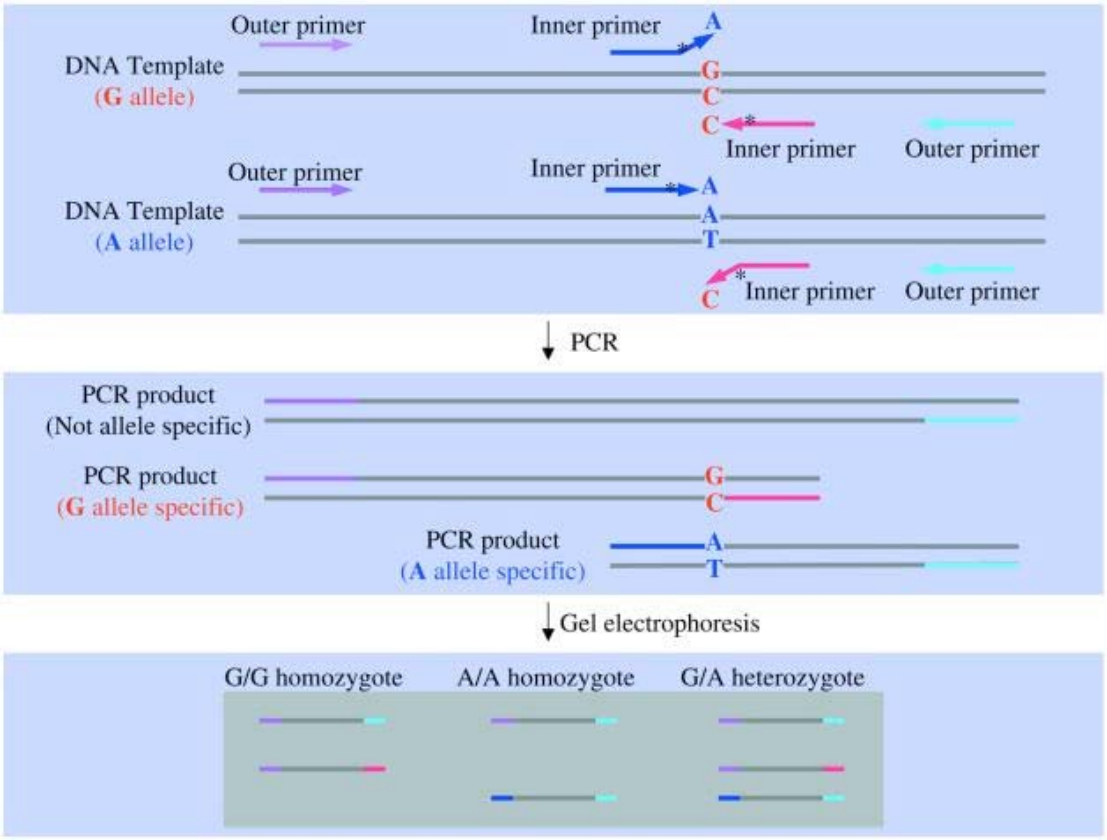
The 20-μL reactions contained 2 U of Taq polymerase (Qiagen), 1–100 ng of DNA, and 200 μM each of dATP, dCTP, dGTP, and dTTP (Invitrogen); PCR buffer was used as suggested by the supplier (Qiagen) with no supplements added. After initial denaturation at 95 °C for 2 min, the two-step thermocycling procedure consisted of denaturation at 95 °C for 1 min and annealing and extension at 69 °C for 2 min (in the presence of primers A and B or primers A, B, C, and D) or 1 min (in the presence of primers C and D only), repeated for 35 cycles, and followed by a final extension at 72 °C for 7 min. The thermocyclers used were GeneAmp PCR systems 9600 and 9700

(Applied Biosystems). For genotype assignments, the PCR products were separated in 0.7% agarose gels. Eight percent polyacrylamide gels (Invitrogen) were also suitable in cases where primers C and D were used for the determination of the Hp 2 allele instead of primers A and B. Genotype determinations were done without knowledge of the phenotyping results.



TETRARMS

A simple and economical SNP genotyping method involving a single PCR reaction followed by gel electrophoresis is reported here. The technique, named tetra-primer ARMS-PCR, adopts certain principles of the tetra-primer PCR method (5) and the amplification refractory mutation system (ARMS) (6; Fig. 1). Differences between the tetra-primer ARMS-PCR method, the original tetra-primer PCR method and the Bi-PASA (bidirectional PCR amplification of specific alleles) method reported by Liu et al. (7) are summarised in Table 1. In contrast to Bi-PASA, both inner primers of the tetra-primer ARMS-PCR method encompass a deliberate mismatch at position –2 from the 3_-terminus. An extra destabilizing mismatch has been found to increase the specificity of classical ARMS-PCR (6,8–11). Rules for selecting a nucleotide for the additional mismatch in classical ARMS PCR have been described previously (8): a ‘strong’ mismatch (G/A or C/T mismatches) at the 3_-terminus of an allele-specific primer will likely require a ‘weak’ second mismatch (C/A or G/T) and vice versa, whereas a ‘medium’ mismatch (A/A, C/C, G/G or T/T) at the 3_-terminus will likely require a ‘medium’ second mismatch.



Schematic presentation of the tetra-primer ARMS-PCR method. The single nucleotide polymorphism used here as an example is a G_A substitution, but the method can be used to type other types of single base substitutions. Two allele-specific amplicons are generated using two pairs of primers, one pair (indicated by pink and red arrows, respectively) producing an amplicon representing the G allele and the other pair (indicated by indigo and blue arrows, respectively) producing an amplicon representing the A allele. Allele specificity is conferred by a mismatch between the 3_-terminal base of an inner primer and the template. To enhance allelic specificity, a second deliberate mismatch (indicated by an asterisk) at position –2 from the 3_-terminus is also incorporated in the inner primers. The primers are 26 nt or longer, so as to minimize the difference in stability of primers annealed to the target and non-target alleles, ensuring that allele specificity results from differences in extension rate, rather than hybridisation rate. By positioning the two outer primers at different distances from the polymorphic nucleotide, the two allele-specific amplicons differ in length, allowing them to be discriminated by gel electrophoresis.

It is also demonstrated here that high genotyping throughput can be achieved by combining this method with the microplate array diagonal gel electrophoresis (MADGE) technique (12). As primer design is a critical part of this method and is time consuming, we have developed a primer design computer program and made it accessible to other users through the Internet.

Computer software to design primers for tetra-primer ARMS-PCR

As primer design is a critical part of this method and is time-consuming, we developed a primer design computer program to facilitate this task. The program, outlined in Figure 6, is accessible through the Internet at http://cedar.genetics.soton.ac.uk/public_html/primer1.html. Users need to input the target DNA sequence, specify the polymorphic site and define criteria for the primers (Tm, %GC, length and complementarity) and product sizes.

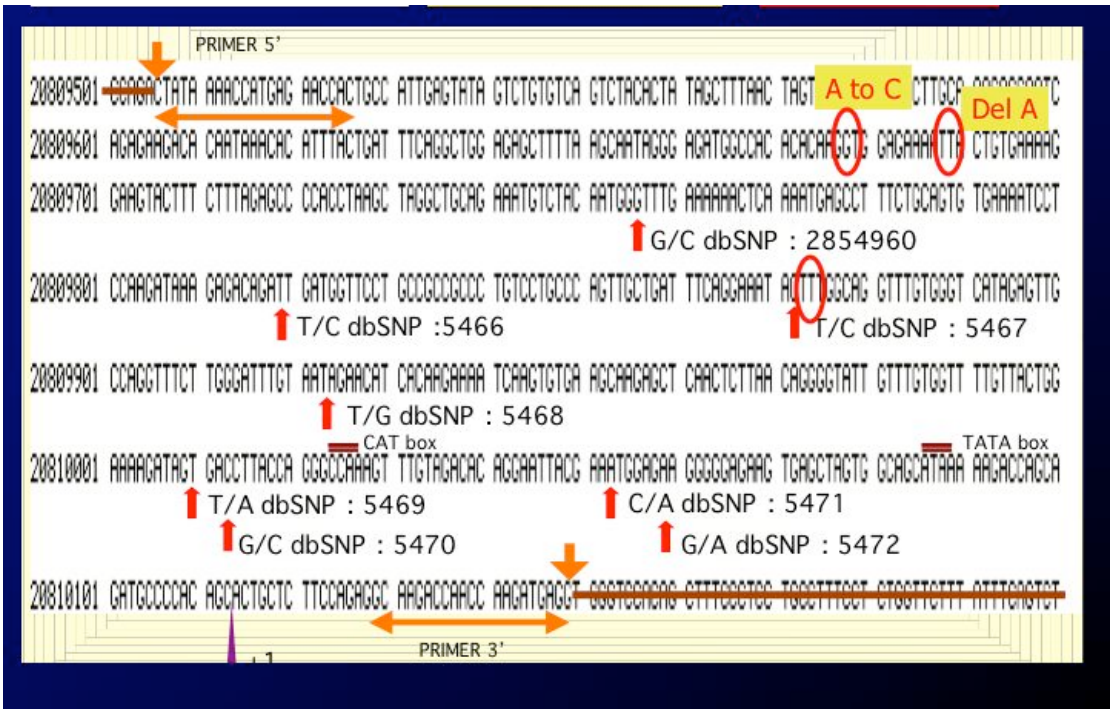
PROMOTER RESEQUENCING.

We performed PCR amplification to determine HP genotypes and the presence of the HP*del allele, as described previously (Koda et al. 1998, 2000). We also examined HP*1 variants by amplifying HP exons 2 to 4 (Koda et al. 1998). Using restriction digests with XbaI, we were able to differentiate the HP*1S and HP*1F alleles (Maeda et al. 1984). The promoter region of HP was amplified using PCR, purified, and sequenced directly with an ABI Prism dye terminator cycle sequencing ready reaction kit (Applied Biosystems, Tokyo, Japan).

PRELIMINARY RESULTS

HP PROMOTER RESEQUENCING IN A WORLDWIDE POPULATION

We evaluated the distribution of HP alleles and base substitutions at the promoter region among Africans (Ghana), Europeans (South Africa), and Chinese to determine the population differences and associations. The distribution of the four HP alleles among the study populations showed marked geographic differences (Table 1). The HP*2 allele was the most prevalent in all three populations but was significantly higher in the Chinese population (P 0.01). The HP*1F allele was absent from the Chinese population and was high in the African population (P 0.02). The HP*1S allele was low in the African population (P 0.005). The HP*del allele was present only in the Chinese population, as observed previously (Koda et al. 2000). All genotype distributions were in Hardy-Weinberg equilibrium. There was distinct variation in the distribution of HP promoter SNPs (55AVG, 61AVC, 101CVG, 104TVA, 191TVG, and 242CVT) across the study populations.

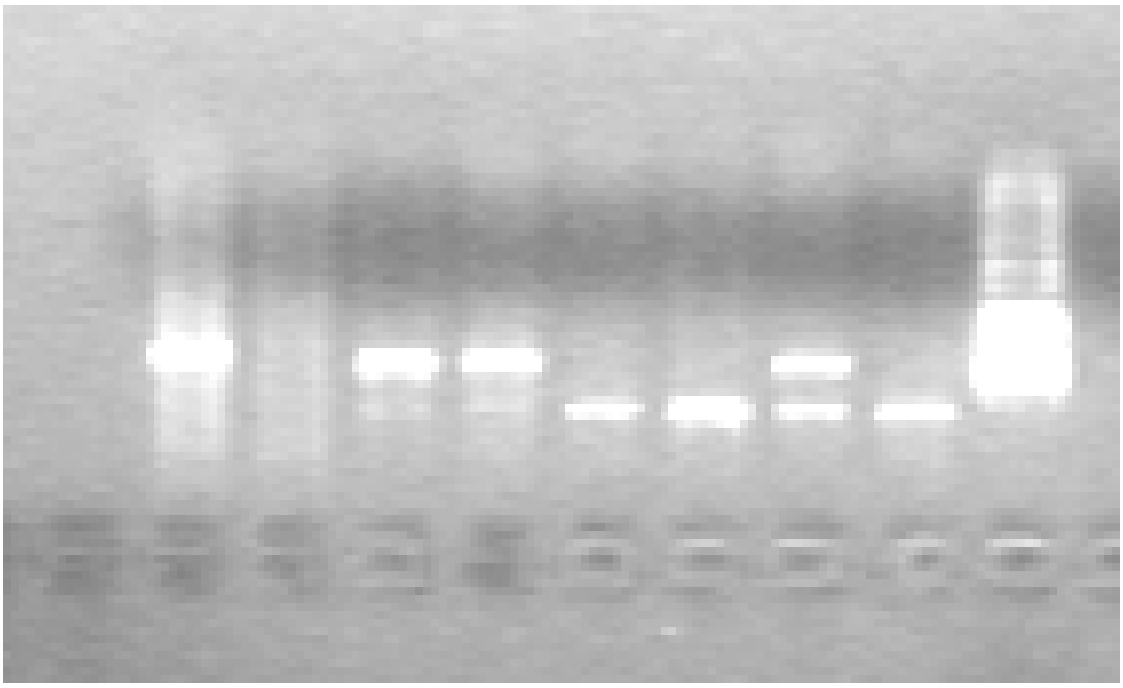


Strikingly, the 61C and 101G substitutions were found only in the African population (Table 2). The 242T substitution was generally high in Europeans (P 0.01). The 55G substitution was predominant in all populations but was elevated in Africans compared to the Chinese population (P 0.007). The 104A substitution was significantly high in

Africans. These observations reveal that four SNPs (55G, 61C, 101G, 104A) occurred significantly in the African population and that the 242T mutation was significant in the Europeans. SNPs occurred less frequently in the Chinese population. FST estimates suggest that the African population is moderately differentiated from the Chinese and Europeans, who appear less differentiated from each other (Table 2). Analysis of the SNPs with the SNPs of two out-groups (chimpanzee and rhesus monkey) showed that 55A, 61A, 101C, 104T, 191T, and 242C are the ancestral nucleotides and that the HP promoter region is highly conserved (data not shown). The most likely haplotypes determined by Arlequin software are shown in. We found that the 61C and 101G substitutions were strongly in linkage disequilibrium with the HP*2 and HP*1S alleles, respectively. The 55G and 104A substitutions are almost always associated with the HP*1F allele in both the African and European populations. The 55G substitution was found associated with HP*2 in all populations, with frequencies that varied extensively according to the population. The 55G, 104A, HP*2 haplotype was found only in the African population, which also had a high frequency of the 55G, 104A, HP*1F haplotype. Overall, 18 haplotypes were observed among the study populations. Of this number, only three haplotypes were shared among all three populations. With the exception of the 55G, HP*2 haplotype, all the shared haplotypes were of the wild-type allele, and even then visible variations existed between the frequency distributions.

HP DUPLICATION ALLELE GENOTYPING IN CASTELMASSA POPULATION.

The haptoglobin genotypes of 249 consecutive patients were determined with genomic DNA prepared from blood samples. With PCR protocol 1, we successfully genotyped 244 of the 249 samples, whereas we could not establish the genotype of 5 samples. In these five cases, the 1757-bp Hp 1-specific product was present as a relatively weak band, probably because of low concentrations of genomic DNA.

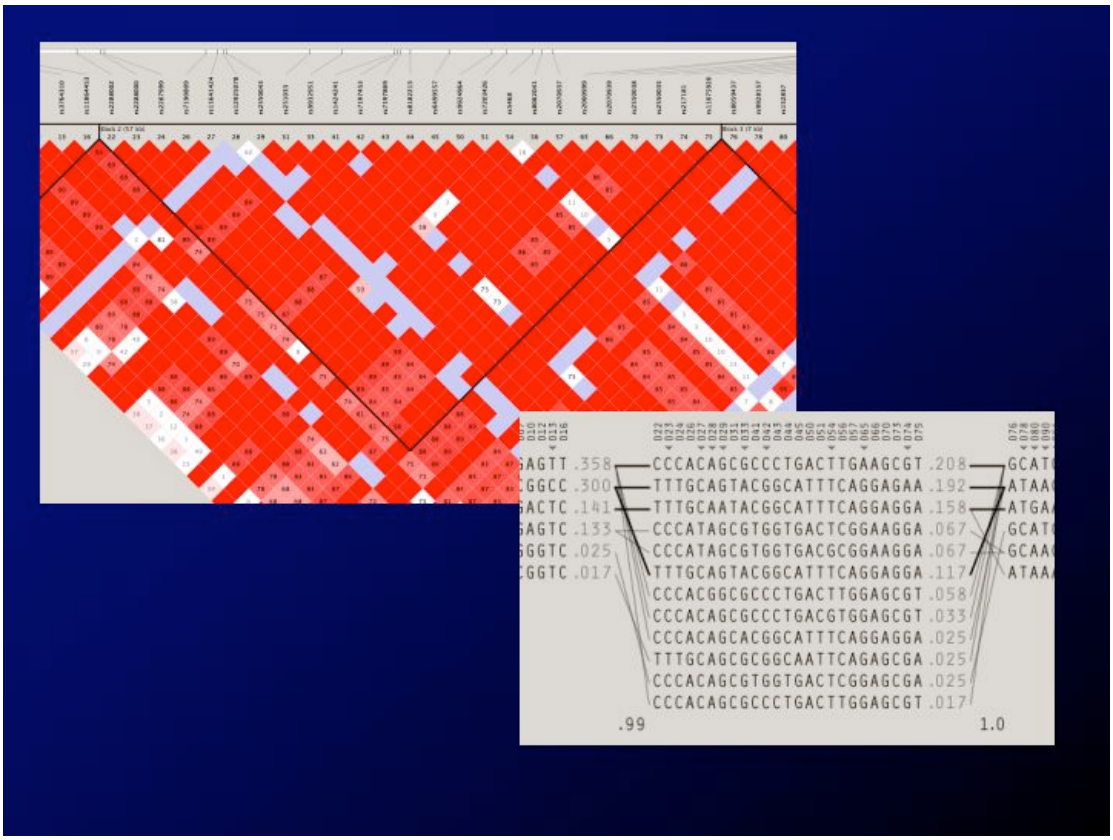


In the same lanes, the 3481-bp Hp 2-specific band was not visible, and it was unclear whether individual samples did not contain the Hp 2 allele or whether the 3481-bp product was present at concentrations too low to be detected in the gel. Thus, in these cases, it was not possible to decide whether the genotype was Hp 1-1 or Hp 2-1. Subsequent genotyping of the five samples with PCR protocols 2 and 3 independently demonstrated that two samples were Hp 1-1 and three samples were Hp 2-1.

GENOTYPING OF 9 TAGSNPS IN THE GENOMIC REGION OF HP GENE

We reconstructed haplotypes for all three genes, using the PLEM algorithm (52) implemented in the tagSNPs program (53), and selected tagSNPs on the basis of the R2 coefficient, which quantifies how well the tagSNP haplotypes predict the SNPs or the number of copies of haplotypes an individual carries. We chose tagSNPs so that common SNP genotypes (minor allele frequency 0.03) and common haplotypes (frequency 0.03) were predicted with R2 0.8 (54). In order to evaluate our tagSNPs' performance in capturing unobserved SNPs within the genes and to assess whether we needed a denser set of markers, we performed a SNP-dropping analysis (25,55). In brief, each of the genotyped SNPs was dropped in turn and tagSNPs were selected from the remaining SNPs so that their haplotypes predicted the remaining SNPs with an R2 value of 0.85. We then estimated how well the tagSNP haplotypes of the remaining

SNPs predicted the dropped SNP, an evaluation that can provide an unbiased and accurate estimate of tagSNP performance (25,55).

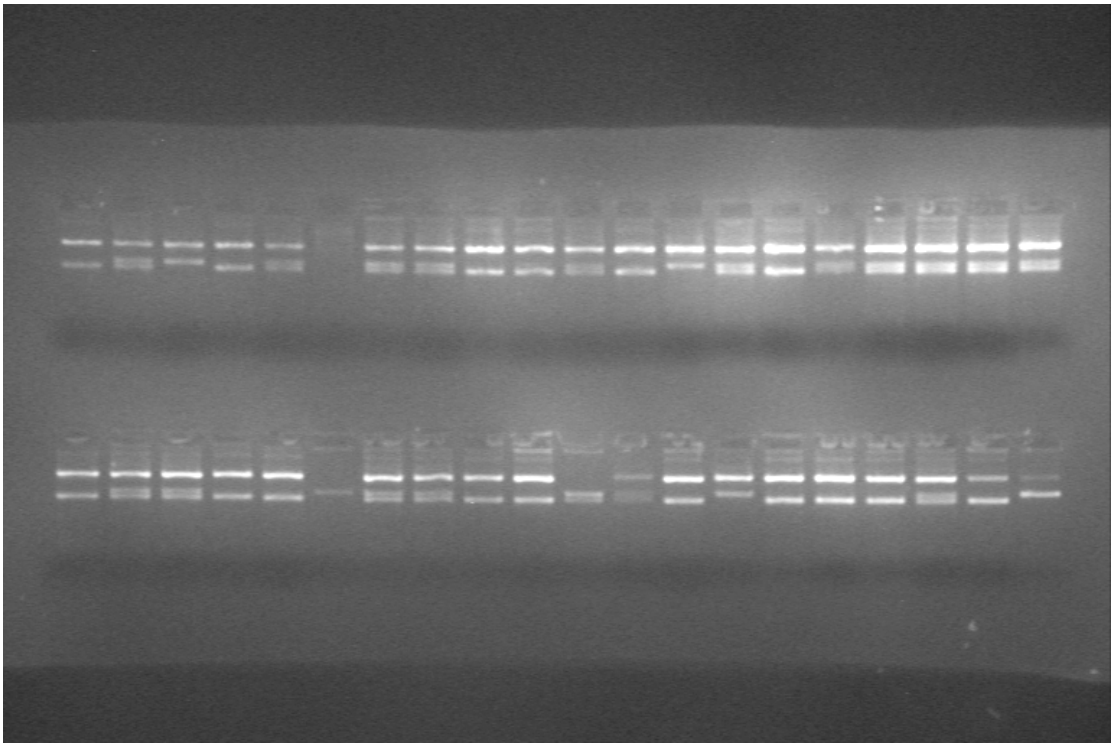


TAGSNPS GENOTYPING

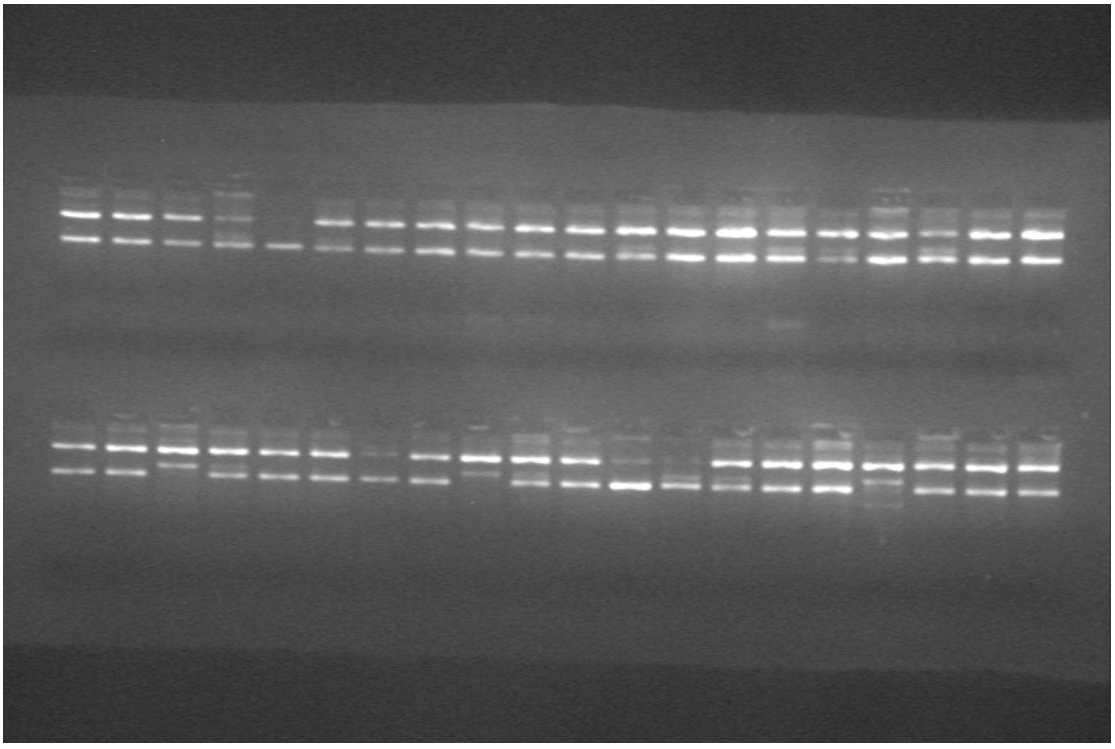
A simple and economical SNP genotyping method involving a single PCR reaction followed by gel electrophoresis is reported here. The technique, named tetra-primer ARMS-PCR, adopts certain principles of the tetra-primer PCR method (5) and the amplification refractory mutation system (ARMS) (6; Fig. 1). Differences between the tetra-primer ARMS-PCR method, the original tetra-primer PCR method and the Bi-PASA (bidirectional PCR amplification of specific alleles) method reported by Liu et al. (7) are summarised in Table 1. In contrast to Bi-PASA, both inner primers of the tetra-primer ARMS-PCR method encompass a deliberate mismatch at position –2 from the 3_-terminus. An extra destabilizing mismatch has been found to increase the specificity of classical ARMS-PCR (6,8–11). Rules for selecting a nucleotide for the additional mismatch in classical ARMS PCR have been described previously (8): a ‘strong’ mismatch (G/A or C/T mismatches) at the 3_-terminus of an allele-specific primer will likely require a ‘weak’ second mismatch (C/A or G/T) and vice versa, whereas a

‘medium’ mismatch (A/A, C/C, G/G or T/T) at the 3_-terminus will likely require a ‘medium’ second mismatch.

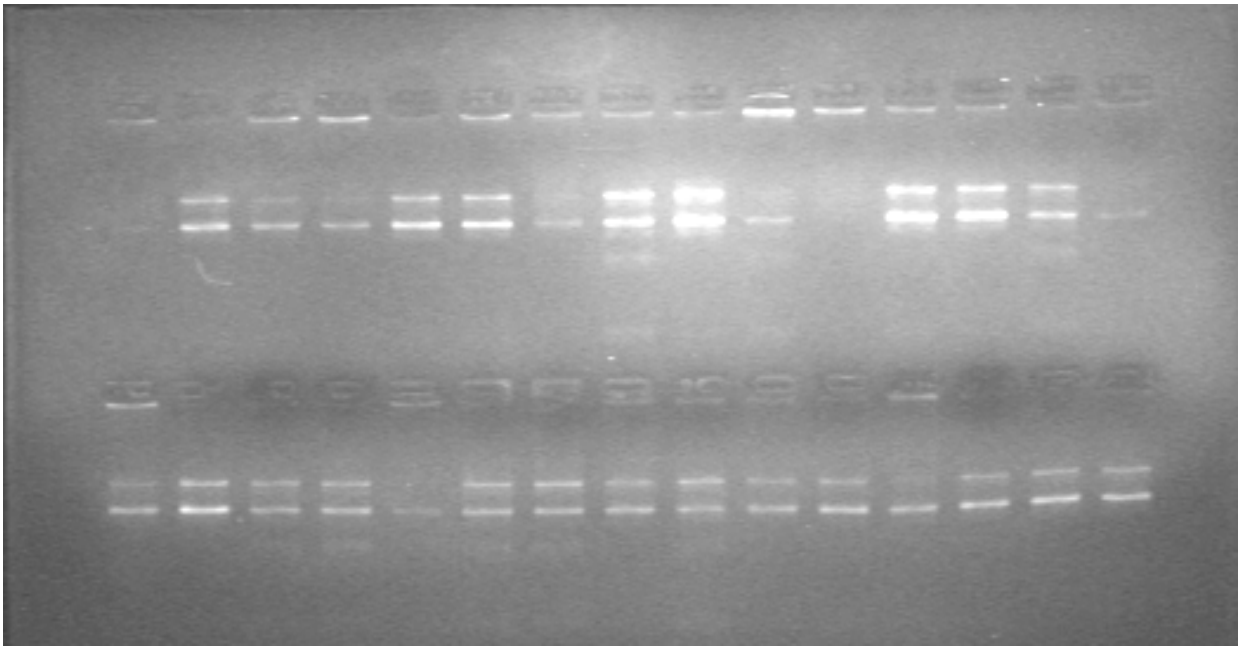
rs2288000



rs11641424



rs12925078



DISCUSSION

The present data suggest that HP promoter polymorphism can be used as a marker in human population genetic studies. We determined the frequency distribution of HP alleles and promoter polymorphism using PCR and direct DNA sequencing of PCR products. We sequenced a 300-bp genomic DNA fragment of the HP promoter region and found population specificity of the distribution of not only the HP alleles but also the base substitutions at the promoter region and the haplotypes.

The 61C and 101G substitutions appear to be specific to individuals of African origin, because they were not detected in Europeans or Asians. Interestingly, the 61C substitution, located in one of three interleukin 6 responsive elements (Oliviero and Cortese 1989), was also found in African Americans, with the HP 2,1mod phenotype having a lower serum Hp level (Maeda 1991).

We have also seen this substitution in South African Xhosas at frequencies that are similar to those observed in Ghanaians (results not shown). We previously reported that the 61AVC and 101CVG substitutions were strongly associated with ahaptoglobinemia and hypohaptoglobinemia, respectively, which are frequent in African populations (Teye et al. 2003). The uniqueness of the 61C and 101G substitutions in the African population is highlighted by the fact that they are associated with the HP*2 and HP*1S alleles, respectively. The 104A substitution also appeared to be specific to African populations because we detected only one (1%) allele out of 100 European alleles, whereas the Ghanaian population had as many as 92 (37.4%) out of 246 alleles and the Chinese population had none. The South African white population is estimated to be 93% European and 7% non-European (a mix of people of southern African and Asian descent) (Botha et al. 1975). Therefore the probability of the one 104A substitution coming from an African gene cannot be excluded. The European population showed a higher frequency of the 242T substitution, which was the only major marker for this population. The 55G and 191G substitutions occurred at high and low frequencies, respectively, in all study populations, so they may not be suitable for population genetic studies. However, the 61C, 101G, 104A, and 242T substitutions could be useful for population genetic studies in African and European populations.

The Chinese population appears to be less polymorphic with respect to the SNPs than the African or European populations. However, the presence of the HP*del allele and the absence of the HP*1F allele serve as unique markers within the Chinese population.

The data revealed a dense SNP map within the HP promoter region, containing as many as six polymorphic sites within a short fragment of 300 bp. Thus, together with other markers, HP polymorphism at the allele and promoter levels may provide new pieces to the anthropogenetic puzzle of human migration and population differentiation. All the substitutions can be detected in a single PCR and sequencing reaction because the DNA fragment is less than 400 bp, making it simple and convenient to analyze.

ACKOWLEDGEMENT

1. M.R. Langlois and J.R. Delanghe, Biological and clinical significance of haptoglobin polymorphism in humans. Clin. Chem. 42 (1996), pp. 1589–1600.

2. E.R. Giblett, The haptoglobin system. Ser. Haematol. 1 (1968), pp. 3–20.

3. B.H. Bowman, Haptoglobin. In: B.H. Bowman, Editor, Hepatic Plasma Proteins, Academic Press, San Diego (1993), pp. 159–167.

4. H.E. Schultze, J.F. Heremans et al.Molecular Biology of Human ProteinsNature and Metabolism of Extracellular Proteins vol. 1, Elsevier, Amsterdam (1966), pp. 384–402.

5. V. Lange, Der Haptoglobin Polymorphismus-nicht nur eine genetische Markierungshilfe (Review) . Anthropol. Anz. 50 (1992), pp. 281–302.

6. O. Smithies, Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. Biochem. J. 61 (1955), pp. 629–641.

7. L. Thomas, A. Fateh-Moghadam, W.G. Guder, W. Hofmann, H. Reiber and M. Lammers, Editors, Proteindiagnostik, Behringwerke, Frankfurt a.M (1991), pp. 92–93.

8. J.G. Raynes, S. Eagling and K.P. McAdam, Acute-phase protein synthesis in human hepatoma cells: differential regulation of serum amyloid A (SAA) and haptoglobin by interleukin-1 and interleukin-6. Clin. Exp. Immunol. 83 (1991), pp. 488–491.

9. L. Thomas, Haptoglobin/hämopexin. In: L. Thomas, Editor, Labor und Diagnose (4th ed.), Medizinische Verlagsgesellschaft, Marburg (1992), pp. 813–820.

10. H.K. Louagie, J.T. Brouwer, J.R. Delanghe, M.L. De Buyzere and G.G. Leroux-Roels, Haptoglobin polymorphism and chronic hepatitis C. J. Hepatol. 25 (1996), pp. 10–14.

11. T. Okazaki, Y. Yangisawa and T. Nagai, Analysis of the affinity of each haptoglobin polymer for hemoglobin by two-dimensional affinity chromatography electrophoresis. Clin. Chim. Acta 258 (1997), pp. 137–144.

12. L. Garby and W.D. Noyes, Studies on hemoglobin metabolism: I. The kinetic properties of the plasma hemoglobin pool in normal man. J. Clin. Invest. 38 (1959), pp. 1479–1483.

13. J.R. Delanghe and M.R. Langlois, Hemopexin: a review of biological aspects and the role in laboratory medicine. Clin. Chim. Acta 312 (2001), pp. 13–23. SummaryPlus

14. J. Javid, The effect of haptoglobin-polymer size on hemoglobin binding capacity. Vox Sang. 10 (1965), pp. 320–325.

15. J.R. Delanghe, K. Allcock, M.R. Langlois, L. Claeys and M.L. De Buyzere, Fast determination of haptoglobin phenotype and calculation of hemoglobin-binding

capacity using high pressure gel permeation chromatography. Clin. Chim. Acta 291 (2000), pp. 43–51.

16. S.M.H. Sadrzadeh, E. Graf, S.S. Panter, P.E. Hallaway and J.W. Eaton, Hemoglobin—a biologic Fenton reagent. J. Biol. Chem. 259 (1984), pp. 14354–14356.

17. J.M.C. Gutteridge, Lipid peroxidation and antioxidants as biomarker of tissue damage. Clin. Chem. 41 (1995), pp. 1819–1828.

18. J.M.C. Gutteridge, The antioxidant activity of haptoglobin towards haemoglobin-stimulated lipid peroxidation. Biochim. Biophys. Acta 917 (1987), pp. 219–223.

19. Y. Miller, S. Altamentova and N. Shaklai, Oxidation of low-density lipoprotein by hemoglobin stems from a heme-initiated globin radical: antioxidant role of haptoglobin. Biochemistry 36 (1997), pp. 12189–12198.

20. S.K. Lim, H. Kim, S.K. Lim, A. bin Ali, Y.K. Lim, Y. Wang et al., Increased susceptibility in Hp knockout mice during acute hemolysis. Blood 92 (1998), pp. 1870–1877.

21. Y.K. Lim, A. Jenner, A.B. Ali, Y. Wang, S.I. Hsu, S.M. Chong et al., Haptoglobin reduces renal oxidative DNA and tissue damage during phenylhydrazine-induced hemolysis. Kidney Int. 58 (2000), pp. 1033–1044.

22. S.S. Panter, S.M.H. Sadrzadeh, P.E. Hallaway, J.L. Haines, V.E. Andersen and J.W. Eaton, Hypohaptoglobinemia associated with familial epilepsy. J. Exp. Med. 161 (1985), pp. 748–754.

23. M. Kristiansen, J.H. Graversen, C. Jacobsen, O. Sonne, H.-J. Hoffman, S.K.A. Lax et al., Identification of the haemoglobin scavenger receptor (Letter) . Nature 409 (2001), pp. 198–201.

24. M.R. Langlois, M.-E. Martin, J.R. Boelaert, C. Beaumont, Y.E. Taes, M.L. De Buyzere et al., The haptoglobin 2-2 phenotype affects serum markers of iron status in healthy males. Clin. Chem. 46 (2000), pp. 1619–1625.

25. A. Cozzi, B. Corsi, S. Levi, P. Santambrogio, G. Biasiotto and P. Arosio, Analysis of the biological functions of H- and L-ferritins in HeLa cells by transfection with siRNAs and cDNAs: evidence for a proliferative role of L-ferritin. Blood 103 (2004), pp. 2377–2383.

26. R.D. Klausner, T.A. Rouault and J.B. Harford, Regulating the fate of mRNA: the control of cellular iron metabolism. Cell 72 (1993), pp. 19–28.

27. M. Hentze and L.C. Kühn, Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc. Natl. Acad. Sci. U. S. A. 93 (1996), pp. 8175–8182.

28. P. Ponka, C. Beaumont and D.R. Richardson, Function and regulation of transferrin and ferritin. Semin. Hematol. 35 (1998), pp. 35–54.

29. S.M. El Ghmati, E.M. Van Hoeyveld, J.A.G. Van Strijp, J.L. Ceuppens and E.A.M. Stevens, Identification of haptoglobin as an alternative ligand for CD11b/CD18. J. Immunol. 156 (1996), pp. 2542–2552.

30. L. Wagner, A. Gessl, S. Baumgartner Parzer, W. Base, W. Waldhäusl and M.S. Pasternack, Haptoglobin phenotyping by newly developed monoclonal antibodies. Demonstration of haptoglobin uptake into peripheral blood neutrophils and monocytes. J. Immunol. 156 (1996), pp. 1989–1996.

31. N. Berkova, C. Gilbert, S. Goupil, J. Yan, V. Korobko and P.H. Naccache, TNF-induced haptoglobin release from human neutrophils: pivotal role of the TNF p55 receptor. J. Immunol. 162 (1999), pp. 6226–6232.

32. M. Maes, W. Cooreman, J. Delanghe, S. Scharpé, A. Wauters, H. Neels et al., Components of biological variation in plasma haptoglobin: relationships to plasma fibrinogen and immune variables, including interleukin-6 and its receptor. Clin. Chim. Acta 239 (1995), pp. 23–35.

33. M.R. Langlois, J.R. Delanghe, J. Philippé, J. Ouyang, D. Bernard, M.L. De Buyzere et al., Distribution of lymphocyte subsets in bone marrow and peripheral blood is associated with haptoglobin type. Binding of haptoglobin to the B-cell lectin CD22. Eur. J. Clin. Chem. Clin. Biochem. 35 (1997), pp. 199–205.

34. K. Hanasaki, L.D. Powell and A. Varki, Binding of human plasma sialoglycoproteins by the B-cell specific lectin CD22. Selective recognition of immunoglobulin A and haptoglobin. J. Biol. Chem. 270 (1995), pp. 7543–7550.

35. W. Köhler and O. Prokop, Relationship between haptoglobin and Streptococcus pyogenes T4 antigens (Letter) . Nature 271 (1978), p. 373.

36. O. Prokop and W. Köhler, Haptoglobins act similar to antibodies. Zent.bl. Gynakol. 101 (1979), pp. 1111–1116.

37. V.R. Gordeuk, G. McLaren and W. Samowitz, Etiologies, consequences, and treatment of iron overload. Crit. Rev. Clin. Lab. Sci. 31 (1994), pp. 89–133.

38. H. Van Vlierberghe, M. Langlois, J. Delanghe, Y. Horsmans, P. Michielsen, J. Henrion et al., Haptoglobin phenotype 2-2 overrepresentation in Cys282Tyr hemochromatotic patients. J. Hepatol. 35 (2001), pp. 707–711.

39. K. Carter, D.J. Bowen, A. McCune and M. Worwood, Haptoglobin type neither influences iron accumulation in normal subjects nor predicts clinical presentation in HFE C282Y haemohromatosis: phenotype and genotype analysis. Br. J. Haematol. 22 (2003), pp. 326–332.

40. E. Beutler, T. Gelbart and P. Lee, Haptoglobin polymorphism and iron homeostasis. Clin. Chem. 48 (2002), pp. 2232–2235.

41. J. Delanghe, B. de Hemptinne, H. Vandenbussche, R. Troisi and J. Decruyenaere, Donor organ haptoglobin phenotype determines outcome in liver transplantation following hepatitis [Abstract] . Clin. Chem. 44 Suppl. (1998), p. A159.

42. I. Kasvosve, V.R. Gordeuk, J.R. Delanghe, Z.A. Gomo, I.T. Gangaidzo, H. Khumalo et al., Iron status in black persons is not influenced by haptoglobin polymorphism. Clin. Chem. Lab. Med. 40 (2002), pp. 810–813.

43. T.P. Tuomainen, K. Punnonen, K. Nyyssönen and J.T. Salonen, Association between body iron stores and the risk of acute myocardial infarction in men. Circulation 97 (1998), pp. 1461–1466.

44. J. Delanghe, D. Duprez, M. De Buyzere, B. Bergez, B. Callens, G. Leroux-Roels et al., Haptoglobin polymorphism and complications in established essential hypertension. J. Hypertens. 11 (1993), pp. 861–867.

45. J. Delanghe, B. Cambier, M. Langlois, M. De Buyzere, H. Neels, D. De Bacquer et al., Haptoglobin polymorphism, a genetic risk factor in coronary artery bypass surgery. Atherosclerosis 132 (1997), pp. 215–219. SummaryPlus

46. J. Delanghe, M. Langlois, D. Duprez, M. De Buyzere and D. Clement, Haptoglobin polymorphism and peripheral arterial occlusive disease. Atherosclerosis 145 (1999), pp. 287–292. SummaryPlus

47. J.P. Chapelle, A. Albert, J.P. Smeets, C. Heusghem and H.E. Kulbertus, Effect of the haptoglobin phenotype on the size of a myocardial infarct. N. Engl. J. Med. 307 (1982), pp. 457–463.

48. F.M. Nakhoul, S. Marsh, I. Hochberg, R. Leibu, B.P. Miller and A.P. Levy, Haptoglobin genotype as a risk factor for diabetic retinopathy. J. Am. Med. Assoc. 284 (2000), pp. 1244–1245.

49. A.P. Levy, A. Roguin, I. Hochberg, P. Herer, S. Marsh, F.M. Nakhoul et al., Haptoglobin phenotype and vascular complications in patients with diabetes. N. Engl. J. Med. 343 (2000), pp. 969–970.

50. A. Roguin, W. Koch, A. Kastrati, A. Schomig and A.P. Levy, Haptoglobin genotype is predictive of major adverse cardiac events in the 1-year period after percutaneous transluminal coronary angioplasty in individuals with diabetes. Diabetes Care 26 (2003), pp. 2628–2631.

51. R. Asleh, S. Marsh, M. Shilkrut, O. Binah, J. Guetta, F. Lejbkowicz et al., Genetically determined heterogeneity in hemoglobin scavenging and susceptibility to diabetic cardiovascular disease. Circ. Res. 92 (2003), pp. 1193–1200.

52. J.W. Eaton, P. Brandt, J.R. Mahoney and J.T. Lee, Haptoglobin: a natural bacteriostat. Science 215 (1982), pp. 691–693.

54. I. Kasvosve, Z.A.R. Gomo, E. Mvundura, V.M. Moyo, T. Saungweme, H. Khumalo et al., Haptoglobin polymorphism and mortality in patients with tuberculosis. Int. J. Tuberc. Lung Dis. 4 (2000), pp. 758–764.

55. J.R. Delanghe, M.R. Langlois, J.R. Boelaert, J. Van Acker, F. Van Wanzeele, G. van der Groen et al., Haptoglobin polymorphism, iron metabolism, and mortality in HIV infection. AIDS 12 (1998), pp. 1027–1032.

56. C. de Monye, D.S. Karcher, J.R. Boelaert and V.R. Gordeuk, Bone marrow macrophage iron grade and survival of HIV-seropositive patients. AIDS 13 (1999), pp. 375–380.

57. V.R. Gordeuk, J. Delanghe, M. Langlois and J.R. Boelaert, Iron status and the outcome of HIV infection: an overview. J. Clin. Virol. 20 (2001), pp. 109–113.

58. I.K. Quaye, J. Brandful, F.A. Ekuban, B. Gyan and N.A. Ankrach, Haptoglobin polymorphism in human immunodeficiency virus infection: Hp0 phenotype limits depletion of CD4 cell counts in HIV-1 seropositive individuals. J. Infect. Dis. 181 (2000), pp. 1483–1485.

59. I.K. Quaye, F.A. Ekuban, J.A. Brandful, B.A. Gyan, B.D. Akanmori and N.A. Ankrach, Haptoglobin phenotypes in HIV-1-seropositive patients in Ghana: decreased risk for Hp0 individuals. Hum. Hered. 50 (2000), pp. 382–383.

60. E. Beutler, V.J. Felitte, J.A. Koziol, N.J. Ho and T. Gelbart, Penetrance of 845G_A (C282Y) HFE hereditary haemochromatosis mutation in the USA. Lancet 359 (2002), pp. 211–218.

CHAPTER 3
DOES CULTURAL ISOLATION LEAD TO INCREASED POWER IN GENE FINDING IN ASSOCIATION STUDIES? THE BASQUE LESSON.

INTRODUCTION

Complex diseases, supposed to be caused by the interaction between multiple genes and environmental factors, are the major public health problems in developed societies. The recent breakthroughs in highthroughput genotyping technologies are rendering feasible whole genome association studies, in which the whole genome is scanned for regions where a variant modifying susceptibility to the disease may lie. This approach is based on the ability to capture an association signal by means of the underlying linkage disequilibrium (LD) between the genotyped polymorphism and the susceptibility variants. Increased linkage disequilibrium would mean, thus, and increased statistical power to capture such signals. However, whole genome association studies may be particularly prone to the problems that have plagued association studies, affecting reproducibility. Some of those problems are population substructure, which creates spurious genetic associations, and the heterogeneity in the causal variants .

It has often been suggested that genetically isolated populations would offer increased stastical power to detect association because of the impact on their genomic structure of their particular demography. LD would be higher than in other populations because of the reduced effective population size, which limits the opportunity for recombination to act and erode LD. Moreover, such populations are not expected to present internal substructure (thus reducing its impact in creating spurious associations), and to be genetically more homogeneous, which may translate into being less diverse in the genetic architecture of susceptibility to a particular complex disease.

While many populations have been proposed as isolated and ideal for association studies, empirical data that verify the assumptions mentioned above are scarce. Some studies used microsatellites in the X chromosome and found increased LD in the Saami for northern Scandinavia (lavori di Laan & Paabo), although it is unclear how LD can be compared between microsatellites and the now much more popular single nucleotide polymorphisms (SNPs), given their widely different mutation rates and models. The most comprehensive SNP-based study of LD and genetic heterogeneity on an isolated

population was performed on the Miconesian Kosrae, where indeed heterogenity was decreased and LD decayed more slowly with physical distance (ref.). However, to the best of our knowledge no such study has been performed on an isolated European population.

A candidate to being a genetically isolated population suitable for associated studies are the Basques. They live in a small area straddling the westernmost section of the French-Spanish border; Basques are ~1 million, of which a quarter are native speakers of the Basque language, an isolated, non-Indoeuropean tongue. Besides the language, many other cultural traits contribute to their cultural isolation. As for genetic isolation, they have been described as the most differentiated population in continental western Europe based on classical polymorphisms (i.e., blood groups, enzyme polymorphims, and HLA) (Calafell and Bertranpetit, 1994). Their mtDNA sequences and Y-chromosome polymorphisms do not show such relatively intense differences, and locate them at the extreme of European-wide gradients in lineage frequencies (Bertranpetit et al., 1995; Rosser et al., 2000; Bosch et al., 2001), and, in some Mendelian diseases, present reduced mutation spectra with private mutations (Cobo et al. 2005)

In the present study, we seek to verify whether the clear cultural isolation of the Basques translated into the genetic features that would make them more appropriate than other populations for association studies, namely, reduced genetic heterogeneity and extended LD. For that, we genotyped 123 SNPs in a ~1-Mb gene-free region of chromosome 22 in two Basque populations, as well as in three other non-isolated Spanish populations, which served as controls. Additionally, this data set can be used to explore the genetic heterogeneity of Spanish populations; since a major source of heterogeneity might be the the different contribution of the North African populations (a varying proportion of the Iberian Peninsula was ruled by North African Moslems from 711 AD to 1492), we also genotyped three North African populations.

Our results seem to indicate that the main (and almost only) source of genetic differentiation in our samples was Iberia vs. North Africa: the former, and the Basques in particular, were not significantly different from each other. LD decay was not increased in Basques either.

MATERIAL AND METHODS

SAMPLES

We genotyped a total of 541 individuals from eight populations (Table 1): Basques from Iparralde (the part of the Basque country lying in France) (BAF), Basques from Gipuzkoa (a province of Spain) (BAS), Catalans (in NE Iberia) (CAT), Extremadura in SW Spain (EXT), Andalusians in S Spain (AND), North Moroccans (NMO) and South Moroccans (SMO), both mostly Berber-Speakers, and Sahrauis (SAH), that is, Arabs from the Western Sahara, a former Spanish colony south of Morocco, now claimed by this country. All individuals were autochthonous and had all four grandparents born in the same region.

SNP SELECTION AND GENOTYPING.

SNPs were selected and genotyped as in González-Neira et al. (2006). Briefly, SNPs were selected from dbSNP build 115 in a region spanning 987,872 bp in chromosome 22, from bp 32600114 to bp 33587986 (NCBI Build 34). The 1-Mb region begins at the 3' end of the Glycosyltransferase-like protein LARGE, which belongs to the Glycosyltransferase family 8; no other known gene maps to this interval. Different classes of repeats have been found in the region, including SINEs, LINEs, LTRs, STRs, and others (Dunham et al. 1999).

Genotypes were produced with Spectrotyper (Sequenom), at a multiplex level of four. Initially, 211 assays were designed. Of those, 65 had call rates <70% and were dropped, as were six that failed Hardy-Weinberg equilibrium (HWE). The HWE cut off was established considering that, in a given locus, the number of populations that would fail HW follows a Poisson distribution. As the total number of tests with $P < 0.05$ gave an average of 4.8 populations with $P < 0.05$ per locus, we excluded SNPs that were not in HW equilibrium in a number of populations over the 95% tail of the Poisson distribution (in this case, four populations). Finally, 14 SNPs were rejected since they resulted monomorphic in all samples. In total, 123 SNPs were considered in the analysis dataset, with a final average distance between contiguous SNPs of 8060 bp. [non torna: 211-65-6-14=126, non 123 (???)]

For the considered loci, 27 genotype discrepancies for specific SNPs in specific individuals were found in 16,704 duplicated genotypes (0.16%), and these results were discarded before the analysis. [in questo dataset, o in quello mondiale di González-Neira 2006?]

STATISTICAL ANALYSIS

Genotypic data were managed with the SNPator platform (Morcillo et al., in preparation; www.snpator.org). SNPator allows easy database building starting from an excel sheet and provides a friendly-use filter system for data management and analysis. It directly performs several statistical analysis and generate input files ready-to-use for different statistical analysis software.

Average heterozigosities and the average number of alleles was directly calculated using the Arlequin package ver 2.00 (Schneider, 1996; http://anthropologie.unige.ch/arlequin/). Arlequin was also used to compute F_{ST} distances among populations. The F_{ST} distance matrix was visulazed by means of multidimensional scaling (MDS) computed with Statistica ver 6.0. MDS plots in two dimension the data points (in this case, populations) which coordinates that minimize the difference between the distances among populations in the plot compared to those in the input matrix. Such differences are measured with the stress parameter, which provides a measure of the quality of the representation. MDS is considered acceptable if the stress is less than 0.15. Since FST distances, as computed by Arlequin, can take negative values, we scaled the distance matrix by adding a small positive number to every value in the matrix.

The significance of the differences in allele frequencies among populations or among groups of populations was tested by means of the analysis of the molecular variance (AMOVA, Excoffier et al. 1992), which is also implemented in Arlequin.

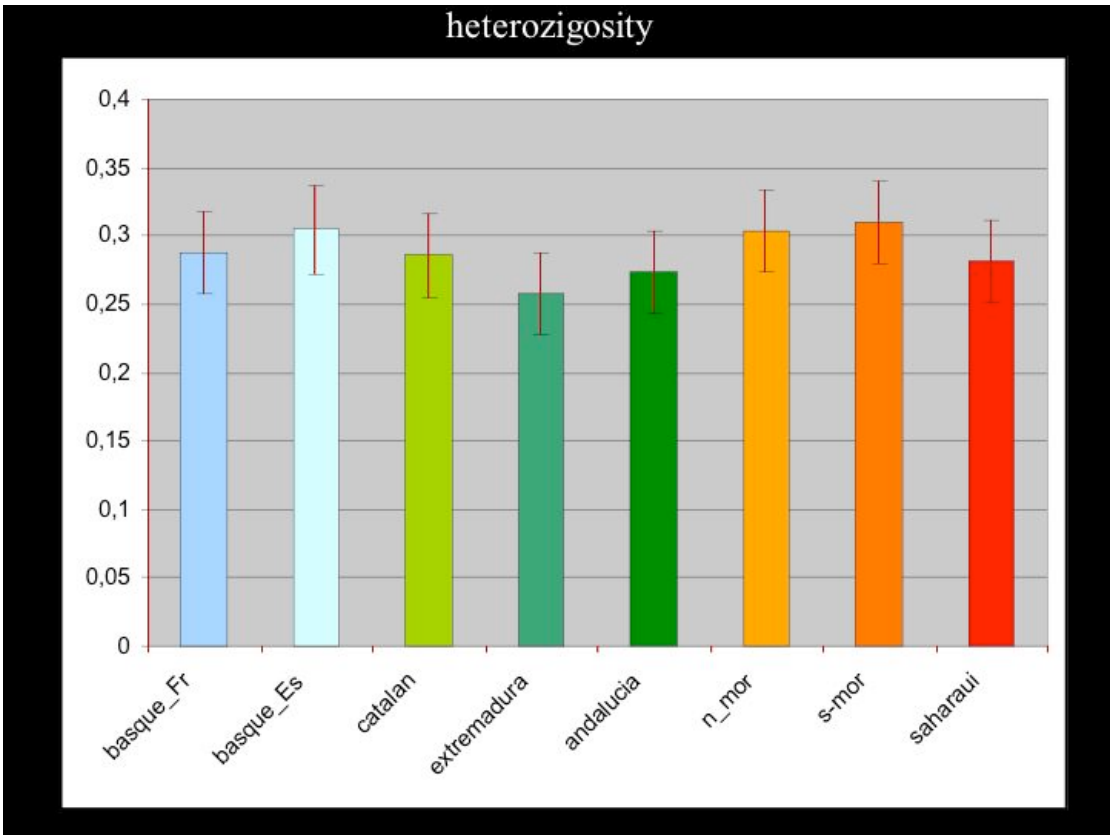
Genetic substructure within samples was investigated with a model-based clustering method developed by Pritchard et Al.(Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun;155(2):945-59. PMID: 10835412 [PubMed - indexed for MEDLINE]) based on a Bayesian statistical approach. This method infers the presence of population structure from multilocus molecular datasets attributing individuals to K arbitrary, unknown ancestral populations. In our analysis we considered our set of data under an admixture model.

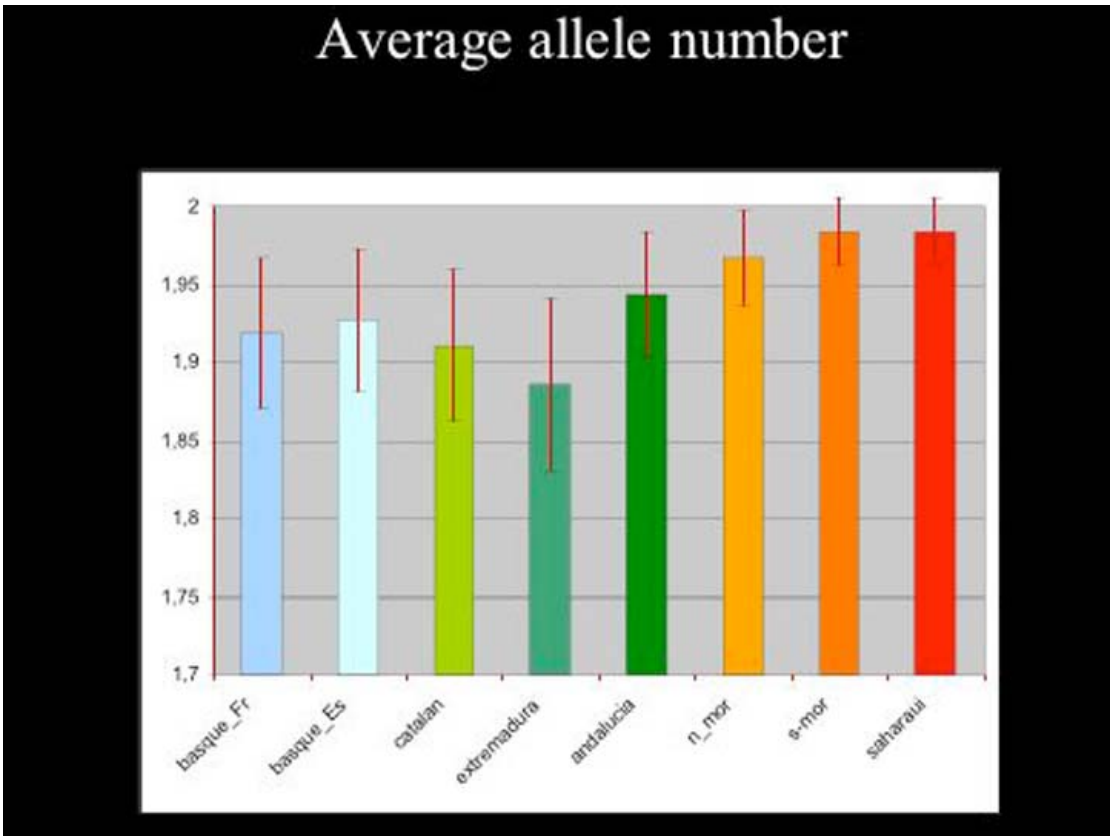
LD decay was described by counting the proportion of SNP pairs that had r2 or |D'| values ≥ 0.8 for different bins of physical distance. LD parameters were computed with Haploview, which was also used to find tagSNPs by means of the tagger algorithm (ref.).

RESULTS

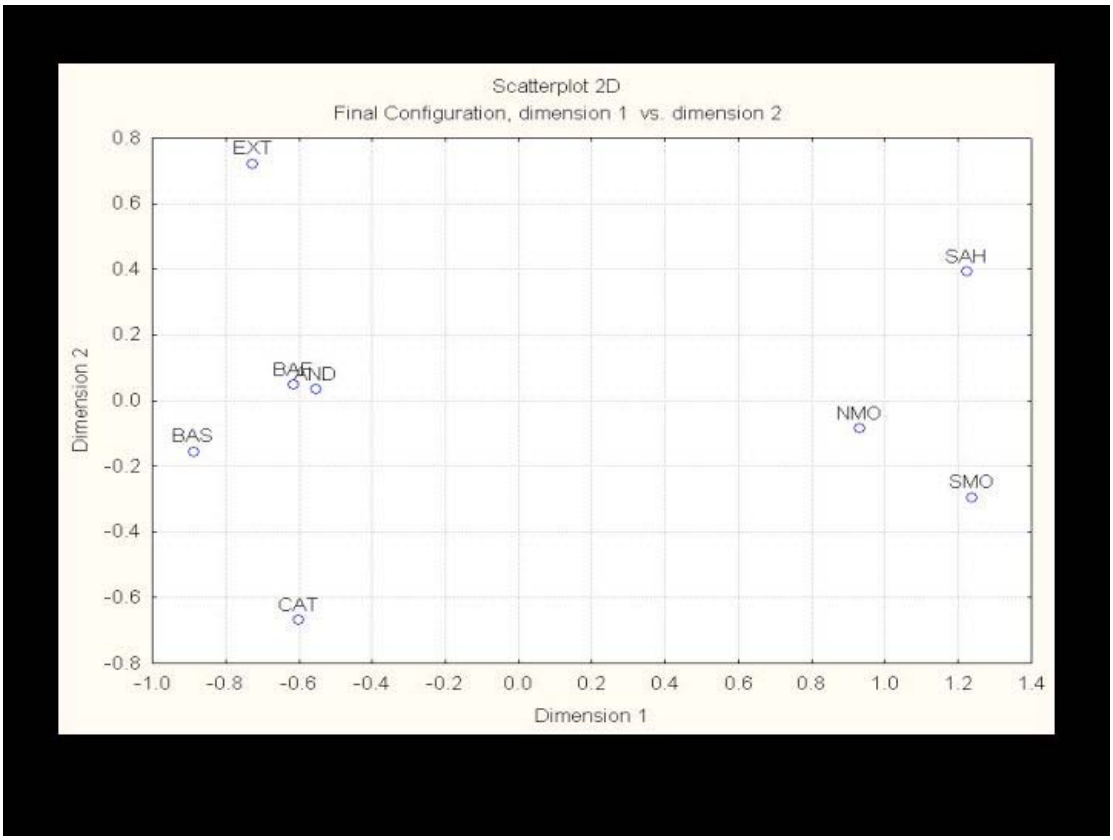
We obtained and analyzed genotype frequencies for 123 SNPs in a ~1-Mb geneless region of chromosome 22 with a mean distance of 8232 bp from each other. Genotypes were available for 541 individuals from: two Basque populations, three general Iberian populations and three north African populations.

Mean heterozigosity plotted for each population in Fig. 1. Two conclusions can be readily drawn: confidence intervals widely overlap, and so differences in heterozygosty among these populations are not statiscally significant, but, in any case, Basques are not the most genetically homogeneous population in the study. This pattern is repeated when considering the mean number of alleles per locus (Fig. 2).





F_{ST} genetic distances were computed among populations and plotted by means of MDS. Since it can be argued that LD implies that SNPs do not behave as independent variables, we performed this analysis both with the whole 123-SNP dataset, or with the set of the 56 SNPs that were tagSNPs in each of the eight populations. Both data sets produced very similar results, and only the latter is presented in Fig. 3. Stress was 0.0072, which is quite low and indicates a good fit between the plot and the initial matrix. Two features stand out from the MDS plot: the main source of genetic variation in our sample seems to be Iberian vs. North African populations, and the Basques cluster together with the other Iberian populations. Actually, it is the small Extremadura sample that seems to depart from the general Iberian cluster. Southern Iberian populations would be expected to be closer to Northern Africans than any other Iberians, if the North African invasion of 711 AD had a significant demographical impact. However, as seen in Fig. 3, this is not the case.

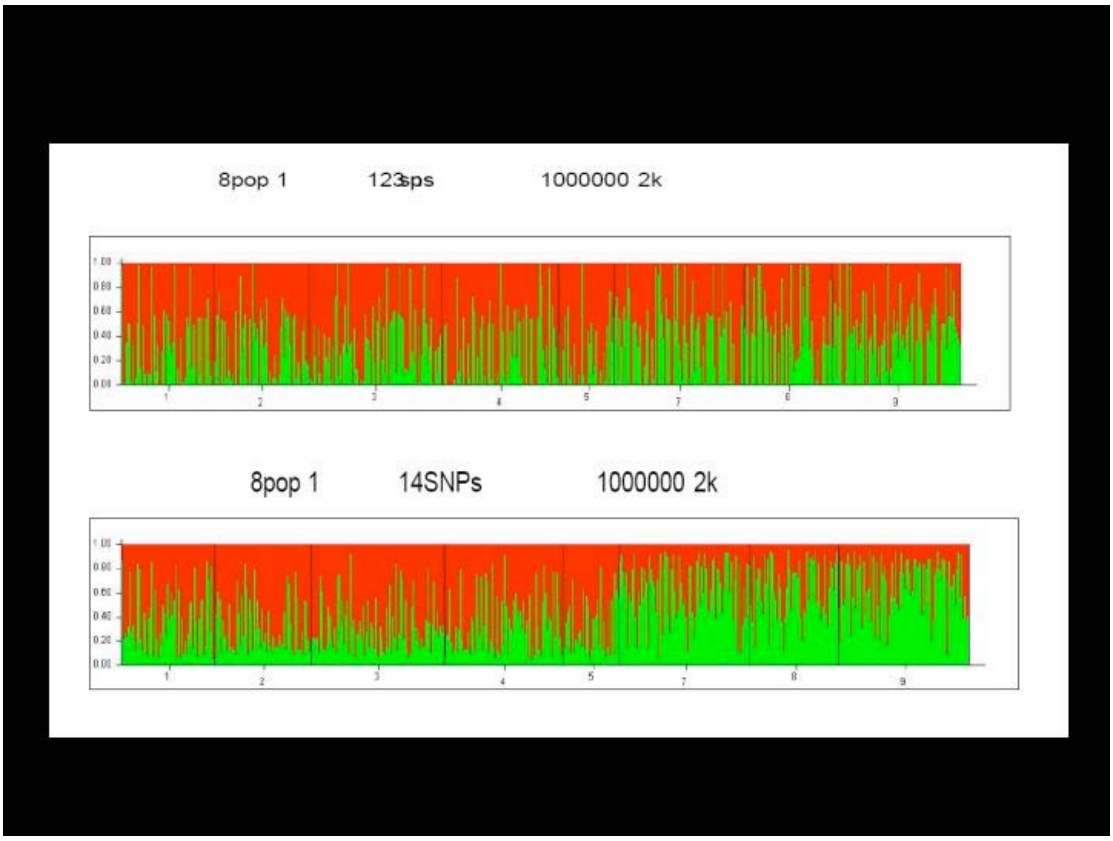


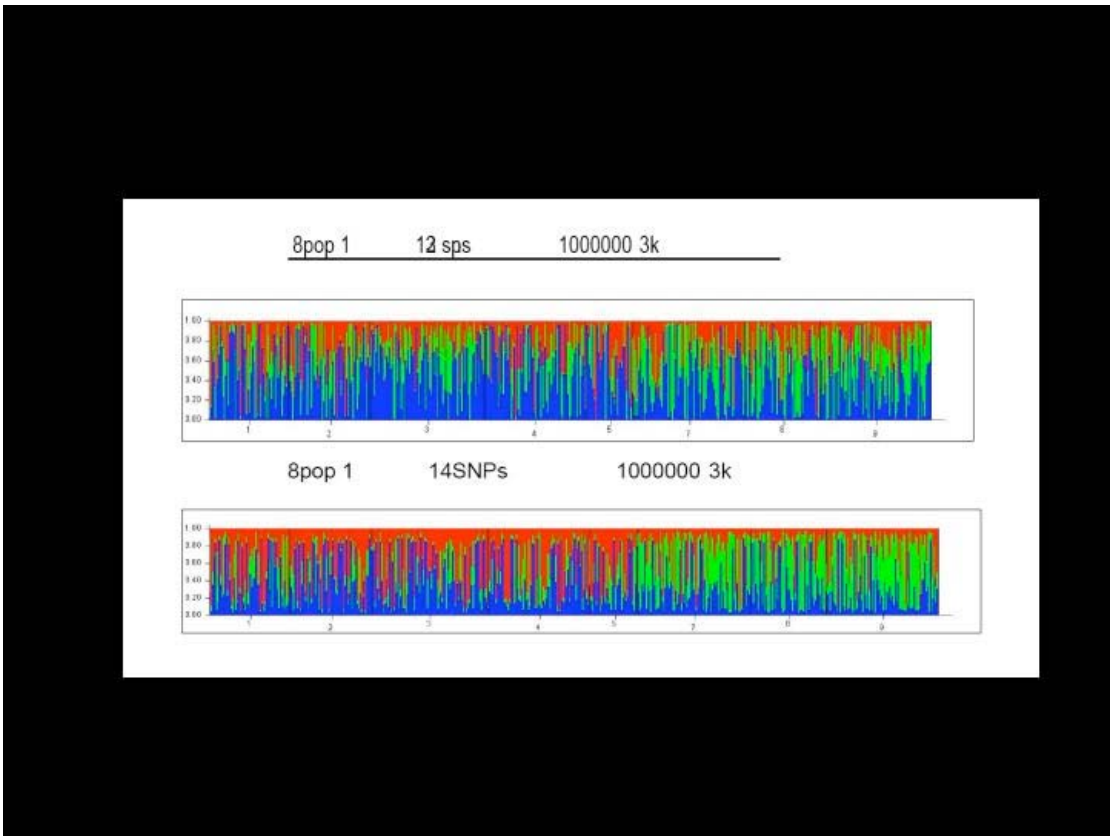
Next, we tested for the significance of the genetic differences apparent in the F_{ST} matrix by means of AMOVA (Tab. 2). As above, we used the whole data set or the set of tagSNPs, with very similar results. F_{ST} among all populations is 1.03% ($p<0.001$), and F_{SC} (that is, the proportion of the total genetic variance explained by differences between groups of populations) is 2.05% ($p=0.02$) among the Iberians and North Africans. F_{SC} was much smaller among Basques and Iberian non-Basques is -0.02%, which has to be taken as 0. That is, allele frequencies for 123 SNPs show on average no statistically significant difference among Basques and non-Basques.

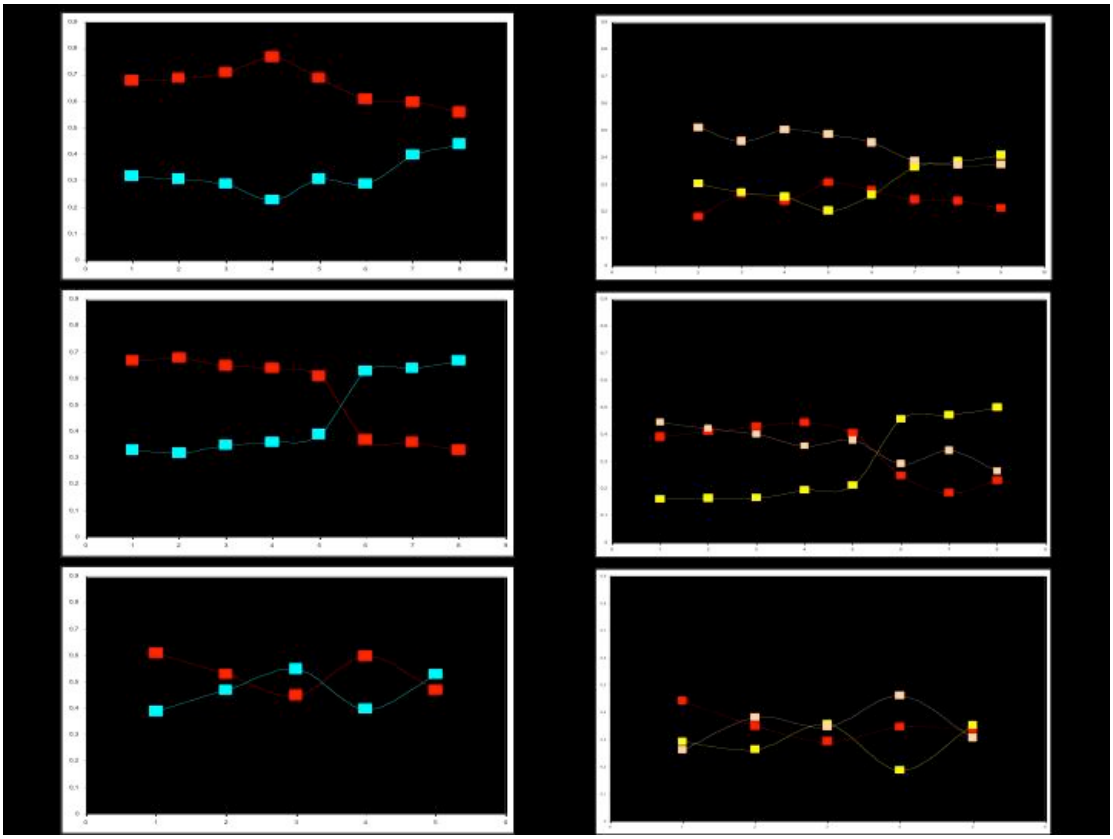
	123 SNPs			56 SNPs (tagSNPs in all pop)		
	Within population	Between pop Within groups	Among groups	Within population	Between pop Within groups	Among groups
Basque_Fr, Basque_Es, Catalan, Extremadura, Andalucia, North_Mor, Saharaui, South_Mor.	98.97 %	1.03 % P 0.00		98.97 %	1.03 % P 0.00	
North_Mor, Saharaui, South_Mor.	100.13 %	-0.13 % P 0.7		100.14 %	-0.14 % P 0.7	
Basque_Fr, Basque_Es, Catalan, Extremadura, Andalucia.	100.13 %	-0.13 % P 0.71		100.02 %	-0.02 % P 0.38	
Basque_Fr, Basque_Es, Catalan, Extremadura, Andalucia / **North_Mor, Saharaui, South_Mor.**	98.09 %	-0.13 % P 0.79	2.05 %	98.30 %	-0.07 % P 0.58	1.77 %
Basque_Fr, Basque_Es / **Catalan, Extremadura, Andalucia.**	100.13 %	-0.12 % P 0.63	-0.02 %	100.03 %	-0.00 % P 0.35	-0.03 %
Basque_Es, Catalan, Andalucia, Extremadura /Basque_Fr.	100.11 %	-0.14% P 0.67	0.02%	100.02 %	-0.01% P 0.41	-0.01%
Basque_Fr, Catalan, Andalucia, Extremadura /Basque_Es.	100.05 %	-0.17% P 0.78	0.12%	100.05 %	-0.04% P 0.43	0.06%

Population substructure was investigated by means of the Bayesian approach implemented in the STRUCTURE program. When we tested all 123 SNPs with K=2 or K=3 populations, no clear pattern could be seen (Fig. 4,5), which is a direct consequence of the low, albeit significant, level of differentiation between Iberian and North African populations. However, if we selected the 14 SNPs with overall $F_{ST} > 5\%$, then a slight but discernible structure emerged separating Iberia from N Africa. We can also plot the average probability per actual population of belonging to each theoretical ancestral population (Fig. 6). In this simplified representation, the patterns described above are clearer. In particular, we can observe that in graphs with K=3, the averages

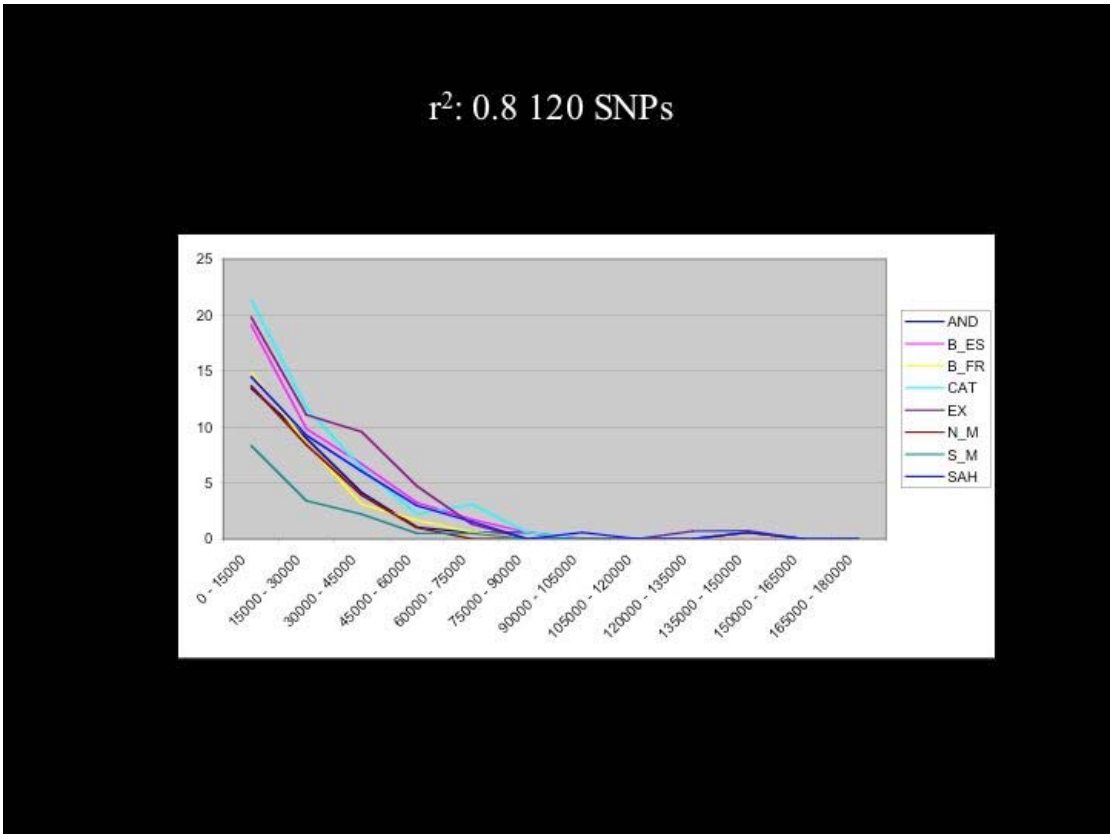
for two of the ancestral populations always overlap, indicating that K=2 fits best the structure present in our dataset.



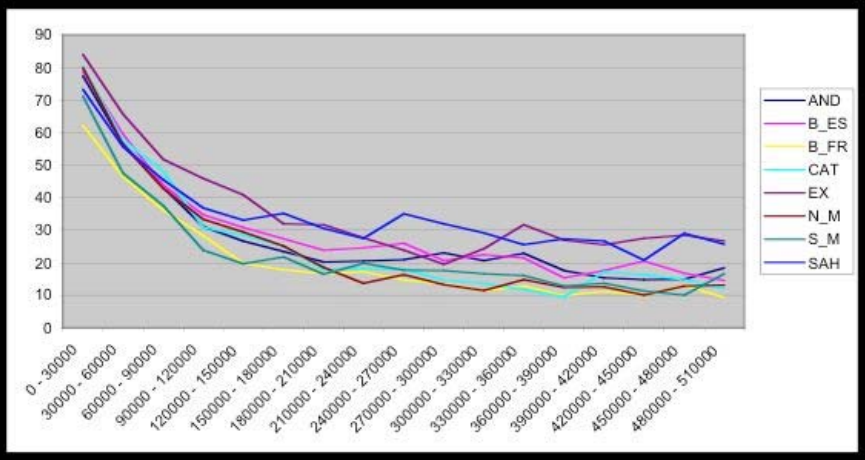

8pop 1 12 sps 1000000 3k
8pop 1 14SNPs 1000000 3k



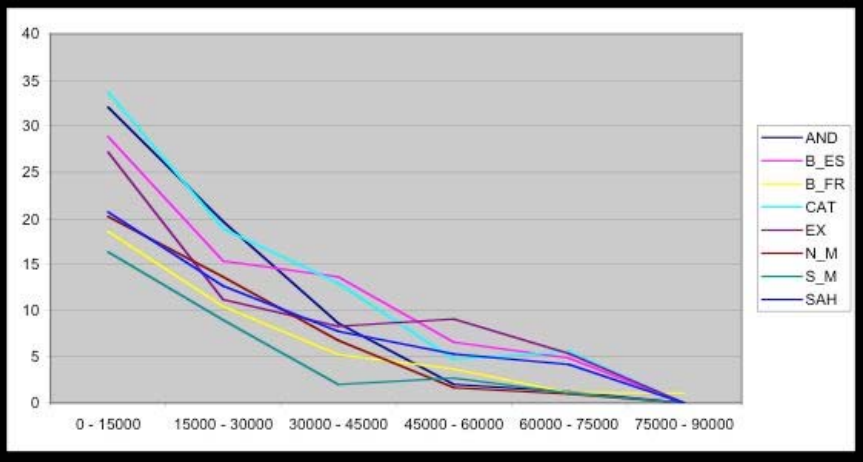
Finally, we analyzed linkage disequilibrium decay. In Figures 7 to 10, the proportion of SNP pairs in each physical distance bin with r2>0.8 (Figs. 7, 9) or |D'|>0.8 (Figs. 8, 10) are plotted. We considered separately all SNPs (Figs. 7, 8) and SNPs with minor allele frequencies above 0.05 (Figs. 9, 10). The four graphs show the same pattern, in which no single population consistently exhibits a slower LD decay.

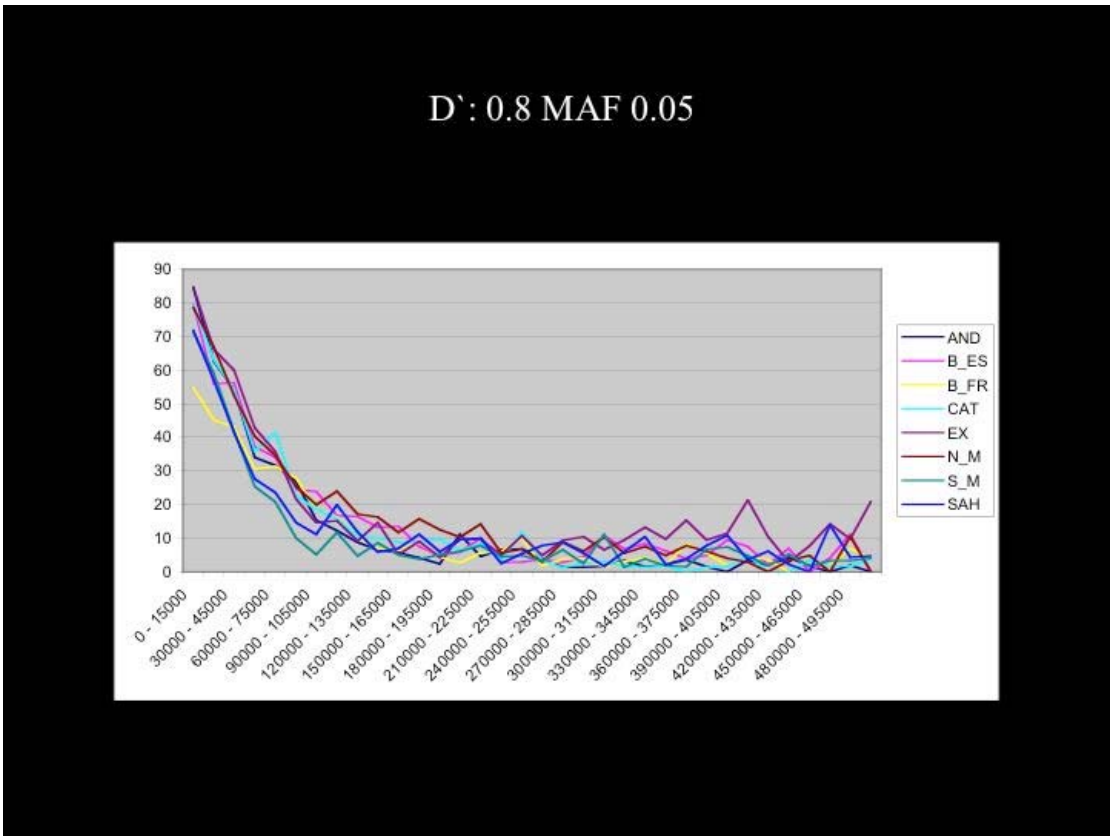


D`: 0.8 120 SNPs



r^2: 0.8 MAF 0.05




D`: 0.8 MAF 0.05
90
80
70
60
50
40
30
20
10
0
AND
B_ES
B_FR
CAT
EX
N_M
S_M
SAH
0 - 15000
30000 - 45000
60000 - 75000
90000 - 105000
120000 - 135000
150000 - 165000
180000 - 195000
210000 - 225000
240000 - 255000
270000 - 285000
300000 - 315000
330000 - 345000
360000 - 375000
390000 - 405000
420000 - 435000
450000 - 465000
480000 - 495000

DISCUSSION

We have sought to verify whether the culturally isolated Basque population is also genetically isolated to the point that would be especially appropriate for genetic association studies. To that effect, we have genotyped 123 SNPs in a ~1-Mb region in chromosome 22 without any known genes, in two Basque populations as well as in three other Iberian samples and in three population from North Africa. All populations showed similar levels of genetic heterogeneity as measured by average heterozigosity. The main source of differentiation within this sample set was Iberia vs. North Africa, while Iberia was genetically homogeneous, and LD decayed with physical distance at similar rates in all populations. Thus, we must conclude that, at least as gathered from this genomic region, Basques do not exhibit the genomic properties that would make them a particularly attractive population for association studies.

Our finding that Basques are not genetically differentiated contrasts sharply with previous reports that showed them to be outliers in the genetic variation of western Europe. However, the loci that with the largest allele frequency differences in previous studies (Calafell and Bertranpetit, 1994) were ABO, RH, and other blood groups, that is, glycoproteins of the red-cell membrane. It is tempting to speculate, then, that such differences were not due to founder effects and subsequent reduced gene flow, but rather that they were the result of microgeographical natural selection, possibly linked to pathogens.

The low genetic differentiation and lack of internal substructure within Iberian population actually facilitates the design of case-control studies, since patients and healthy controls from different regions in Spain can be freely pooled without population substructure producing spurious associations.

We have analyzed 66,240 genotypes from 1 Mb of the human genome in five samples from Iberia and three from North Africa, making it probably the largest genetic survey of this area. However, it is also limited to this particular genome segment, and results should be treated with caution. A broader study, with a wider geographical and genomic coverage would be needed to decide whether gene content has biased our results or to uncover other sources of genetic differentiation within Spain.

ACKNOWLEDGMENTS

de la Chapelle, A. & Wright, F.A. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc. Natl. Acad. Sci. USA 95, 12416–12423 (1998).

Shifman, S. & Darvasi, A. The value of isolated populations. Nat. Genet. 28, 309–310 (2001).

Wright, A.F. , Carothers, A.D. & Pirastu, M. Population choice in mapping genes for complex diseases. Nat. Genet. 23, 397–404 (1999).

the International HapMap Consortium. The International HapMap Project. Nature 426, 789–796 (2003).

The International HapMap Consortium. A haplotype map of the human genome. Nature 437, 1299–1320 (2005).

Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. Science 273, 1516–1517 (1996).

Jorde, L.B. Linkage disequilibrium as a gene-mapping tool. Am. J. Hum. Genet. 56, 11–14 (1995).

Lee, N. et al. A genomewide linkage-disequilibrium scan localizes the Saguenay-Lac-Saint-Jean cytochrome oxidase deficiency to 2p16. Am. J. Hum. Genet. 68, 397–409 (2001).

Eaves, I.A. et al. The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat. Genet. 25, 320–323 (2000).

Shifman, S. , Kuypers, J. , Kokoris, M. , Yakir, B. & Darvasi, A. Linkage disequilibrium patterns of the human genome across populations. Hum. Mol. Genet. 12, 771–776 (2003).

Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. 22, 139–144 (1999).

Laan, M. & Paabo, S. Demographic history and linkage disequilibrium in human populations. Nat. Genet. 17, 435–438 (1997).

Angius, A. et al. Not all isolates are equal: linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations. Hum. Genet. 111, 9–15 (2002).

Shmulewitz, D. et al. Linkage analysis of quantitative traits related to obesity, type II diabetes, hypertentsion and dyslipidemia (metabolic syndrome) on the island of Kosrae, Federated States of Micronesia. Proc. Natl. Acad. Sci. USA (accepted).

Segal, H.G. Kosrae: The Sleeping Lady Awakens (Kosrae State Tourist Division, Department of Conservation and Development, Kosrae State Government, Federated States of Micronesia, 1995).

Reich, D.E. et al. Linkage disequilibrium in the human genome. Nature 411, 199–204 (2001).

Carlson, C.S. , Eberle, M.A. , Kruglyak, L. & Nickerson, D.A. Mapping complex disease loci in whole-genome association studies. Nature 429, 446–452 (2004).

Lin, S. , Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. Nat. Genet. 36, 1181–1188 (2004).

Patterson, N. et al. Methods for high-density admixture mapping of disease genes. Am. J. Hum. Genet. 74, 979–1000 (2004).

Spielman, R.S. , McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52, 506–516 (1993).

Churchill, G.A. & Doerge, R.W. Empirical threshold values for quantitative trait mapping. Genetics 138, 963–971 (1994).

Doerge, R.W. & Churchill, G.A. Permutation tests for multiple loci affecting a quantitative character. Genetics 142, 285–294 (1996).

Van Steen, K. et al. Genomic screening and replication using the same data set in family-based association testing. Nat. Genet. 37, 683–691 (2005).

Wilk, J.B. et al. Family-based association tests for qualitative and quantitative traits using single-nucleotide polymorphism and microsatellite data. Genet. Epidemiol. 21 (Suppl. 1), S364–S369 (2001).

Zhang, J. , Schneider, D. , Ober, C. & McPeek, M.S. Multilocus linkage disequilibrium mapping by the decay of haplotype sharing with samples of related individuals. Genet. Epidemiol. 29, 128–140 (2005).

Ober, C. , Abney, M. & McPeek, M.S. The genetic dissection of complex traits in a founder population. Am. J. Hum. Genet. 69, 1068–1079 (2001).

Barrett, J.C. , Fry, B. , Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).

Ardlie, K.G., Kruglyak, L., and Seielstad, M. 2002. Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. 3: 299–309.

Bertranpetit, J., Calafell, F., Comas, D., González-Neira, A., and Navarro, A. 2003. Structure of linkage disequilibrium in humans: Genome factors and population stratification. In Cold Spring Harb. Symp. Quant. Biol., pp. 79–88. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., et al. 2002. A human genome diversity cell line panel. Science 296: 261–262.

Carlson, C.S., Eberle, M.A., Rieder, M.J., Smith, J.D., Kruglyak, L., and Nickerson, D.A. 2003. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat. Genet. 33: 518–521.

Carlson, C.S., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. 2004a. Mapping complex disease loci in whole-genome association studies. Nature 429: 446–452.

Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., and Nickerson, D.A. 2004b. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J. Hum. Genet. 74: 106–120.

Comas, D., Plaza, S., Wells, R.S., Yuldaseva, N., Lao, O., Calafell, F., and Bertranpetit, J. 2004. Admixture, migrations, and dispersals in Central Asia: Evidence from maternal DNA lineages. Eur. J. Hum. Genet. 12: 495–504.

Crawford, D.C., Carlson, C.S., Rieder, M.J., Carrington, D.P., Yi, Q., Smith, J.D., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. 2004. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am. J. Hum. Genet. 74: 610–622.

Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D.M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., et al. 2002. A first-generation linkage disequilibrium map of human chromosome 22. Nature 418: 544–548.

de la Vega, F.M., Isaac, H., Collins, A., Scafe, C.R., Halldorsson, B.V., Su, X., Lippert, R.A., Wang, Y., Laig-Webster, M., Koehler, R.T., et al. 2005. The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. Genome Res. 15: 454–462.

Dunham, I., Hunt, R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., et al. 1999. The DNA sequence of human chromosome 22. Nature 402: 489–495.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. Science 296: 2225–2229.

The International HapMap Consortium. 2003. The International HapMap Project. Nature 426: 789–796.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437: 1299–1320.

Jobling, M.A., Hurles, M.E., and Tyler-Smith, C. 2004. Human evolutionary genetics: Origins, peoples, and disease. Garland Science, Taylor & Francis, New York.

Johnson, G.C., Esposito, L., Barratt, B.J., Smith, A.N., Heward, J., Di Genova, G., Ueda, H., Cordell, H.J., Eaves, I.A., Dudbridge, F., et al. 2001. Haplotype tagging for the identification of common disease genes. Nat. Genet. 29: 233–237.

Ke, X., Hunt, S., Tapper, W., Lawrence, R., Stavrides, G., Ghori, J., Whittaker, P., Collins, A., Morris, A.P., Bentley, D., et al. 2004. The impact of SNP density on fine-scale patterns of linkage disequilibrium. Hum. Mol. Genet. 13: 577–588.

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. 22: 139–144.

Kruglyak, L. and Nickerson, D.A. 2001. Variation is the spice of life. Nat. Genet. 27: 234–236.

Mateu, E., Pérez-Lezaún, A., Martínez-Arias, R., Andrés, A.M., Vallés, M., Bertranpetit, J., and Calafell, F. 2002. PKLR-GBA region shows almost complete linkage disequilibrium over 70 kb in a set of worldwide populations. Hum. Genet. 110: 532–544.

Mueller, J.C., Lohmussaar, E., Magi, R., Remm, M., Bettecken, T., Lichtner, P., Biskup, S., Illig, T., Pfeufer, A., Luedemann, J., et al. 2005. Linkage disequilibrium patterns and tagSNP transferability among European populations. Am. J. Hum. Genet. 76: 387–398.

Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321–324.

Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., et al. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294: 1719–1723.

Pritchard, J.K. and Przeworski, M. 2001. Linkage disequilibrium in humans: Models and data. Am. J. Hum. Genet. 69: 1–14.

Ramirez-Soriano, A., Lao, O., Soldevila, M., Calafell, F., Bertranpetit, J., and Comas, D. 2005. Haplotype tagging efficiency in worldwide populations in CTLA4 gene. Genes Immun. 6: 646–657.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. Science 273: 1516–1517.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. 2002. Genetic structure of human populations. Science 298: 2381–2385.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J., and Barbujani, G. 2000. Geographic patterns of mtDNA diversity in Europe. Am. J. Hum. Genet. 66: 262–278.

Soldevila, M., Calafell, F., Helgason, A., Stefansson, K., and Bertranpetit, J. 2005. Assessing the signatures of selection in PRNP from polymorphism data: Results support Kreitman and Di Rienzo's opinion. Trends Genet. 21: 389–391.

Weiss, K.M. and Clark, A.G. 2002. Linkage disequilibrium and the mapping of complex human traits. Trends Genet. 18: 19–24

FINAL REMARKS

METABOLIC NETWORK AND DISEASE SUSCEPTIBILITY

Hepcidin, the long awaited iron hormone, is the principal downregulator of the transport of iron across the small intestine and the placenta, and its release from macrophages. The present view is that hepcidin downregulates iron efflux from the intestine and macrophages by binding FPN and, following complex internalisation, leads to FPN degradation. In HFE, TfR2, and HJV related HC, hepatic expression or serum/urine levels of HAMP are inappropriately low.42 43 53 54 Its expression in the liver is also significantly impaired in HFE, TfR2, and HJV knockout mice24 55–57 and hepatic deposition of iron in HFEKO animals can be prevented by hepcidin overexpression.58 These findings suggests a unifying pathogenic model for all forms of HC in which HFE, TfR2, and HJV are all independent but complimentary regulators of hepcidin synthesis in the liver (fig 1). Hypothetically, these three proteins may be important for sensing circulatory iron and turning on hepcidin gene transcription, albeit with different mechanisms and functional impact. As the central role in iron homeostasis and in hemochromatosis disease development of this small peptide was established, we decided to analyze the variability in the promoter and in the coding region in 84 Italian subjects that presented symptoms of iron overload and resulted negative for the three reported HFE mutations. In the screening study a total of 420 amplicons were analyzed and 5 different variants were detected. Of these 5 variants 3 were in non-coding region and two in were found in exons, but resulted synonymous mutations. These results reflect a picture of high degree of conservations, especially if we consider that the subject that entered in the study share some defects in iron metabolism pathways and are not representative of general population. In three different studies several rare variants in HAMP gene region are reported as causative for hemochromatosis disease. In our case we can not rule out a possible involvement of the variant detected in iron overload phenotype, but it is difficult to demonstrate. From our results we can conclude that even though hepcidin plays a pivotal role in iron homeostasis regulations the molecular variability of the HAMP gene give only a marginal contribution in iron overload phenotypes.

NEW APPROACH ON CANDIDATE GENE STUDIES

The present data suggest that HP promoter polymorphism can be used as a marker in human population genetic studies. We determined the frequency distribution of HP alleles and promoter polymorphism using PCR and direct DNA sequencing of PCR products. We sequenced a 300-bp genomic DNA fragment of the HP promoter region and found population specificity of the distribution of not only the HP alleles but also the base substitutions at the promoter region and the haplotypes.

The 61C and 101G substitutions appear to be specific to individuals of African origin, because they were not detected in Europeans or Asians. Interestingly, the 61C substitution, located in one of three interleukin 6 responsive elements (Oliviero and Cortese 1989), was also found in African Americans, with the HP 2,1mod phenotype having a lower serum Hp level (Maeda 1991).

We have also seen this substitution in South African Xhosas at frequencies that are similar to those observed in Ghanaians (results not shown). We previously reported that the 61AVC and 101CVG substitutions were strongly associated with ahaptoglobinemia and hypohaptoglobinemia, respectively, which are frequent in African populations (Teye et al. 2003). The uniqueness of the 61C and 101G substitutions in the African population is highlighted by the fact that they are associated with the HP*2 and HP*1S alleles, respectively. The 104A substitution also appeared to be specific to African populations because we detected only one (1%) allele out of 100 European alleles, whereas the Ghanaian population had as many as 92 (37.4%) out of 246 alleles and the Chinese population had none. The South African white population is estimated to be 93% European and 7% non-European (a mix of people of southern African and Asian descent) (Botha et al. 1975). Therefore the probability of the one 104A substitution coming from an African gene cannot be excluded. The European population showed a higher frequency of the 242T substitution, which was the only major marker for this population. The 55G and 191G substitutions occurred at high and low frequencies, respectively, in all study populations, so they may not be suitable for population genetic studies. However, the 61C, 101G, 104A, and 242T substitutions could be useful for population genetic studies in African and European populations.

The Chinese population appears to be less polymorphic with respect to the SNPs than the African or European populations. However, the presence of the HP*del allele and the absence of the HP*1F allele serve as unique markers within the Chinese population.

The data revealed a dense SNP map within the HP promoter region, containing as many as six polymorphic sites within a short fragment of 300 bp. Thus, together with other markers, HP polymorphism at the allele and promoter levels may provide new pieces to the anthropogenetic puzzle of human migration and population differentiation. All the substitutions can be detected in a single PCR and sequencing reaction because the DNA fragment is less than 400 bp, making it simple and convenient to analyze.

POPUALTIONS IN ASSOCIATION STUDIES

Isolate populations have demonstrate to be effective for mutation discovery involved in rare Mendelian disorders, now is largely accepted that isolated populations of varied types will proof to be as effective in elucidating the genetic basis of complex disease. Although accepted the model behind the genetic of complex diseases it is completely different to the one behind the genetic of rare diseases, so it is not possible to transfer the good results obtained by the use of isolated populations in rare disease study, in faithful expectation from their utilization in complex disease research. Two main qualities are addressed to isolated populations to support their candidacy: increased levels of LD and lower indices of heterozigosity. These promising qualities were extensively investigated in only one study on a Micronesian isolated population. Except this single study there aren't other available consistent data, especially from European country, where several putative isolated populations are recently proposed as ideal source of subject for successful case control study. Aim of our study is to verify if on the basis of cultural evaluations and from fragmentary molecular data is possible to elect a hypothetical isolated population better than other general populations in gene finding in association study. To do that we genotyped a final number of 123 SNPs selected in a geneless region of 1 Mb in the chromosome 22 in eight populations: two Basque populations that are considered as isolated populations, three general Iberian population: from Catalonia region, Extremadura region and Andalusia region and finally three north African populations. From the analysis of genotype data a picture of extended homogeneity is emerged. The two Basque populations present neither the larger LD signals nor the lowest levels of heterozigosity. From the analysis of populations diversity a pattern of genetic homogeneity arose between the 5 Iberian populations, with no significative values of Fst. Only considering the three African populations the genetic divergence became appreciable and statistically significative. Although significative the level of differentiation between Iberian and North African populations is less deep then those reported in previous study based on the analysis of the y chromosome and mitochondrial DNA. In particular when we perform the structure analysis considering all 123 SNPs we can observe that the high degree of genetic homogeneity of Iberian populations is shared also with north African populations. When we use the 14 SNPs with the higher Fst values we became able to distinguish populations at continental levels, but we must remember that the real picture is the one where the contribution of all the different loci is considered. Backward to Basque issue

our results clearly indicate that Basque do not present the expected enhanced quality required for better results in gene finding association studies. Also clear is that the lack of those enhanced qualities is due to the fact that the two Basque populations for the variants we tested results absolutely not isolated. Although we must be careful in stating general considerations from the analysis of a single genomic region of only 1 Mb, our results supply several interesting new indications and suggest interesting questions. Basque are considered isolated on the basis of cultural issues and from results from several genetic studies based on analysis of molecular variability of y chromosome and mt-DNA and from analysis of variants in HLA region. Results from y chromosome and from mt-DNA have proved to be useful tools in defining populations genetic identity and in reconstructing population history, but is difficult to assess to which extent those molecular marker are representative of the whole genomic variability. Moreover results from the analysis of HLA region were largely used to describe molecular diversity in populations and from the analysis of such variants derived the results that with more strength indicates Basques as a genetic isolate. About this results it is possible to risk a consideration, as HLA region is the genomic region mainly involved in the immunity response to pathogens it is possible that the strong signal of isolation in Basques is increased by the effect of natural selection.

Although our data have not a genome wide scale, they are related to a genomic region where the impact of natural selection is supposed to be minimal and the net of marker considered is dense enough for a good evaluation of linkage disequilibrium decay. From our data appear clear that from cultural issues and from molecular studies that stress population diversity it is not possible elaborate reliable consideration on populations genomic structure, concluding that genetic isolation has to be deeply empirically investigated and tested to be proposed as powerful tool in association studies. Moreover the picture of low levels heterozigosity and the high degree of homogeneity emerged in our results could suggest that a possible alternative to isolated populations in association study is to use populations selected with light demographic rules in order to widen the subject recruitment basin and gain the possibility to collect wider and phenotypically more homogenous cohorts of patients.

INDEX