ARTICLE

Assessment of transmission distortion on chromosome 6p in healthy individuals using tagSNPs

Pablo Sandro Carvalho Santos*,[1,4], Johannes Höhne[2,4], Peter Schlattmann[2], Inke R König[3], Andreas Ziegler[3], Barbara Uchanska-Ziegler[1] and Andreas Ziegler[1]

[1]Institut für Immungenetik, Charité-Universitätsmedizin Berlin, Berlin, Germany; [2]Institut für Biometrie und klinische Epidemiologie, Charité-Universitätsmedizin Berlin, Berlin, Germany; [3]Institut für medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany

The best-documented example for transmission distortion (TD) to normal offspring are the t haplotypes on mouse chromosome 17. In healthy humans, TD has been described for whole chromosomes and for particular loci, but multiple comparisons have presented a statistical obstacle in wide-ranging analyses. Here we provide six high-resolution TD maps of the short arm of human chromosome 6 (Hsa6p), based on single-nucleotide polymorphism (SNP) data from 60 trio families belonging to two ethnicities that are available through the International HapMap Project. We tested all approximately 70 000 previously genotyped SNPs within Hsa6p by the transmission disequilibrium test. TagSNP selection followed by permutation testing was performed to adjust for multiple testing. A statistically significant evidence for TD was observed among male parents of European ancestry, due to strong and wide-ranging skewed segregation in a 730 kb long region containing the transcription factor-encoding genes *SUPT3H* and *RUNX2*, as well as the microRNA locus *MIRN586*. We also observed that this chromosomal segment coincides with pronounced linkage disequilibrium (LD), suggesting a relationship between TD and LD. The fact that TD may be taking place in samples not selected for a genetic disease implies that linkage studies must be assessed with particular caution in chromosomal segments with evidence of TD.
European Journal of Human Genetics (2009) **17**, 1182–1189; doi:10.1038/ejhg.2009.16; published online 4 March 2009

Introduction

Transmission distortion (TD), also termed transmission ratio distortion or segregation distortion, is defined as a departure from Mendel's law of independent segregation of alleles. Causes of TD are numerous, and they include meiotic drive, gametic selection and postzygotic viability selection for or against particular genotypes.[1,2] TD has been observed in several species such as Neurospora, Drosophila and mouse,[3] and has been particularly well documented for mouse t haplotypes. These haplotypes are structural variants encompassing large segments of mouse chromosome 17 that are under extreme TD in wild mouse populations. Heterozygous males preferentially transmit the chromosome bearing t haplotypes, in up to 99% of the cases, to their offspring.[4] However, as homozygosity for t haplotypes is either lethal or causes sterility, and also because heterozygous males seem to suffer from a reproductive disadvantage when compared with wild-type males,[5] the frequency of t haplotypes is maintained constant (15–25%) in wild mouse populations.[4–6]

*Correspondence: PSC Santos, Institut für Immungenetik, Charité-Universitätsmedizin Berlin, Campus Benjamin Franklin, Freie Universität Berlin, Thielallee 73, 14195 Berlin, Germany. Tel: + 49 30 450 564738; Fax: + 49 30 450 564920;
E-mail: pablo.santos@charite.de
[4]These authors contributed equally to this work.*

Although the mechanisms underlying overtransmission of t haplotypes are now relatively well understood,[7] no such insight has been gained so far in humans.

There are several studies reporting the presence of TD in selected regions of the human genome,[8–16] yet most of them are based on families that had been selected for the presence of a particular disease. The question whether TD in these cases is related to the disease, or if it is a disease-independent phenomenon, cannot be answered when an investigation focuses on affected subjects or has its controls within affected families. As shown by at least two reports,[17,18] if a particular locus is normally subject to TD in the general population, and this fact is ignored in a linkage study investigating that locus (eg, an investigation of excess allele sharing among affected sibs), then this kind of 'background TD' might represent a substantial bias in such analyses. The reason for this possible bias is simple, as most linkage studies rely on the assumption that transmission of alleles of a given heterozygous locus takes place according to Mendelian inheritance, so that both male and female offspring have the same chances of inheriting both the alleles.

The HapMap project[19–21] offers a resource of data on genetic variation from four human populations that were not selected for a genetic disease. The data available for two ethnicities – CEU (North American residents with Northern and Western European ancestry) and YRI (Yoruba from Ibadan, Nigeria) – are based on mother–father–adult child family trios, making them adequate for testing TD. The issue of TD has so far been addressed in studies with these trios only very briefly in a supplement to one of the HapMap articles, and it was concluded that, although the inheritance of some single-nucleotide polymorphisms (SNPs) indicated the existence of TD, none achieved whole genomic statistical significance.[20] A principal obstacle in these studies concerns the problem of multiple comparisons, which results in a lack of substantial statistical power when a large number of markers are tested against a null hypothesis. Here, we addressed this issue by using tagSNPs, that is, SNPs that capture information of SNP haplotypes due to linkage disequilibrium (LD), thereby lowering the number of polymorphic markers under study. During the preparation of this paper, new populations were added to the current HapMap release. However, as data are still preliminary and not yet phased, we decided to keep analyses restricted to CEU and YRI.

Using the CEU and YRI trios, the aim of our study was to apply tagSNPs in the course of an investigation of the presence of TD on the short arm of human chromosome 6 (Hsa6p). This region is largely syntenic to that segment of mouse chromosome 17 which harbors t haplotypes. Both chromosomal regions comprise also the major histocompatibility complex, with a multitude of genes participating in immune recognition,[22] female choice and reproductive success.[23–26] Preliminary evidence for TD on Hsa6p has been obtained in the low-resolution study of Zöllner et al.[27]

The results of our analysis show that the model approach of assessing TD chosen here has the potential to overcome limitations of earlier TD studies. It permitted us to identify a region on Hsa6p that is subject to TD in males of European ancestry.

Materials and methods
HapMap data
All data analyzed in this study were downloaded from the HapMap server on 9 February 2008 and correspond to HapMap phase II genotypings in public release 23, with genomic coordinates based on the NCBI Build 36. The haplotypes had been generated (phased) within the HapMap project, and phasing of haplotypes for ambiguous loci (eg, triple heterozygosity in a given locus within a family trio) was solved based on LD data. In total, 30 CEU family trios and 30 YRI family trios, comprising 180 Hsa6p haplotypes in each population, were analyzed. More information on subjects, samples, genotypings and phasing algorithm is given elsewhere.[20,21]

Statistical analysis
To quantify TD, we applied the transmission disequilibrium test (TDT)[28] to the phased data from 68 691 CEU and 74 820 YRI genotyped SNPs within Hsa6p. We calculated TDT statistics (McNemar-type TDT test, which is asymptotically χ_1^2 distributed[29]), and P-value and odds ratio (according to the McNemar's test) for each SNP on Hsa6p, and compiled three TD plots for each ethnicity: father specific; mother specific; and parent unspecific. For visual representation of results, minus $\log_{10}(P$-values$)$ are reported for each marker. To emphasize whole regions instead of a single loci, we also computed 'moving averages', using sliding windows with a width of 400 SNPs, stepping 10 SNPs further after each measurement. To overcome the problem of multiple comparisons, we reduced the number of SNPs by selecting different sets of tagSNPs and performed permutation tests on these markers. TagSNPs were selected from Hsa6p using a pair-wise tagging algorithm,[30] as implemented in HAPLOVIEW 4.[31] Tagging was performed with a maximum intermarker distance of 500 kb, a coefficient of determination of $r^2 = 0.8$ and considering a minor allele frequency (MAF) ≥ 0.15. The latter was chosen because all markers with a MAF less than 0.15 are not variable enough for TD to be detectable, being thus 'invisible' to the TDT tests. To assure that a tagSNP would represent a genomic region of a minimal size and not only itself (or a too low number of other markers), we selected a total of six different sets of tagSNPs, differing only in the minimum number of markers (k) that a tagSNP had to 'capture'. For example, for $k = 5$, only tagSNPs found able to 'capture' information of at least five SNPs (itself and four others) were selected. By lowering k, more tagSNPs are necessary to cover Hsa6p, but

independence between tagSNPs may decrease, and the genomic region captured by one given tagSNP is smaller. In contrast, by increasing k, the distance and independence between neighboring tagSNPs increase substantially, and tagSNPs represent broader genomic regions, having the disadvantage that the actual physical coverage (the percentage of chromosomal length covered by a given set of tagSNPs) decreases. We chose tagSNP assemblies considering $k = 1, 2, 5, 10, 20$ and 30. The parameters chosen for tagSNP selection were relatively strict and assured that the remaining LD between neighboring tagSNPs was very low (see Supplementary Figure S1).

For determining statistical significance, we permuted the transmitted and non-transmitted tagSNP alleles, and compared the distribution of TD test statistics over all tagSNPs. McIntyre *et al*[32] had shown that the Monte Carlo procedure used to obtain a global *P*-value for each set of SNPs guarantees the correct experiment-wise significance level, regardless of dependencies (LD) between the individual loci. For further Monte Carlo approaches addressing the same problem, see also the studies of Shi *et al*[33] and of Becker and Knapp.[34]

Results
Transmission distortion plots
TD plots for Hsa6p in the CEU population are shown in Figure 1. They are divided into father-specific, mother-specific and parent-unspecific components (Figure 1a–c,

respectively). Corresponding TD plots for the YRI population, in the same grouping, are given in Figure 2. The six plots are based on the uncorrected *P*-values of transmission/disequilibrium tests for each locus, and provide a visual reference for TD across Hsa6p in both populational groups. We also produced three TD plots regarding the SNPs that are common to YRI and CEU (Supplementary Figure S2). For dynamic visualization of the results in the genomic context of the HapMap genome browser (www.hapmap.org), we tabulated the results of TD tests in an electronic text file (Supplementary File S3) that can be integrated into the HapMap browser either locally or through the internet. In this manner, the results of our study can be visually and interactively compared with many other annotated features of the genome such as coding and regulatory regions, results of disease association studies, recombination rates and LD patterns.

TagSNPs were selected from all six data sets, and permutation tests were applied to each of the six assemblies of tagSNPs (Table 1). For four of these *k*-values (from 2 to 20), the tests indicated statistically significant evidence for TD among CEU fathers, in a region corresponding to band 6p21.1, at a distance of 45 Mb from the telomere (Figure 1). The TD peak matches a 730 741-bp wide genomic region (715 CEU SNPs), comprising the whole reading frames of the loci *SUPT3H* and *MIRN586*, as well as the upstream region and four exons of *RUNX2*. This chromosomal segment is depicted in detail in Figure 3.
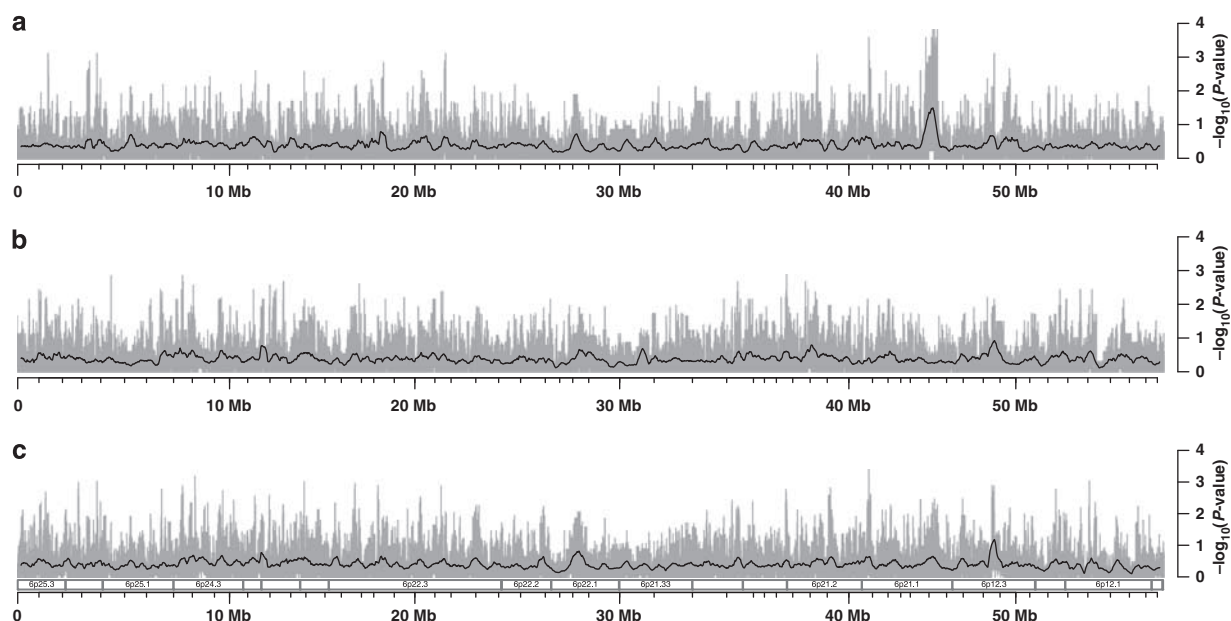

Figure 1 TD plots for chromosome 6p in the CEU panel. (**a**) Father specific, (**b**) mother specific and (**c**) parent unspecific. The gray bars represent $-\log_{10}$(*P*-values) of the transmission/disequilibrium tests in each SNP, whereas the smoothing black lines refer to averages of these same results in sliding windows (400 SNPs wide, stepping 10 SNPs after each measurement). Indications for selected cytogenetic bands are given under (**c**), but apply for all panels.
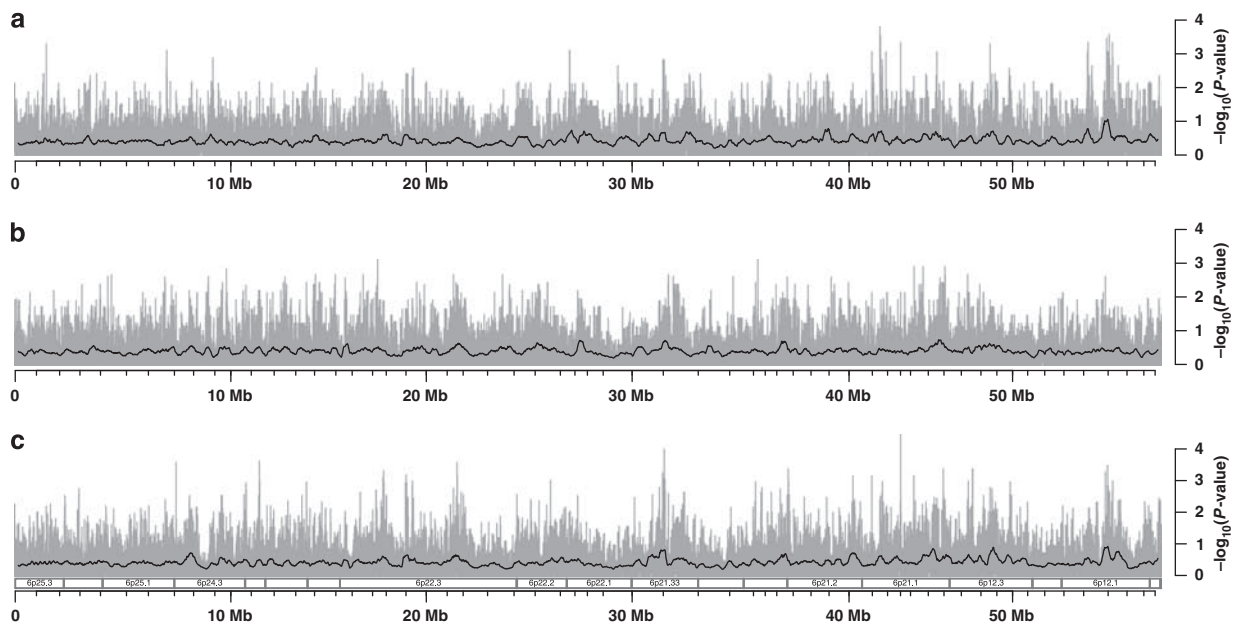
Figure 2 TD plots for chromosome 6p in the YRI panel. (**a**) Father specific, (**b**) mother specific and (**c**) parent unspecific. The gray bars represent $-\log_{10}(P\text{-values})$ of the transmission/disequilibrium tests in each SNP, whereas the smoothing black lines refer to averages of these same results in sliding windows (400 SNPs wide, stepping 10 SNPs after each measurement). Indications for selected cytogenetic bands are given under (**c**), but apply for all panels.

Table 1 Relation between the minimum number of captured SNPs (k), the number of tagSNPs found in Hsa6p with a given k (#tagSNPs), the coverage of Hsa6p (proportion of markers captured by each set of tagSNPs, Cov), the maximum TD test statistics observed considering a given k (maxT$_{TDT}$) and the corresponding Monte Carlo P-values, regarding CEU fathers

k	#tagSNPs	Cov	maxT$_{TDT}$	P-value
1	9781	0.75[a]	14.22	0.085
2	5146	0.61	14.22	0.041
5	2461	0.47	14.22	0.022
10	1091	0.31	12.25	0.046
20	326	0.16	12.25	0.014
30	133	0.09	9.94	0.088

[a]25% of the SNPs have MAF < 0.15.

After correction for multiple comparisons, no statistically significant evidence for TD was found for any other samples than CEU fathers (other results not shown). In fact, analysis of the SUPT3H–MIRN586–RUNX2 ($S–M–R$) region of CEU mothers revealed a tendency for a compensating under transmission of alleles over transmitted by fathers in 92 loci (eg, SNP *rs6899845*: out of 17 heterozygous fathers, 16 transmit the A allele and only 1 transmits the C allele, whereas 7 out of the 11 heterozygous mothers transmit the C and 4 transmit the A allele). Moreover, isolated SNPs in the $S–M–R$ region do show evidence of TD among YRI fathers (see Discussion), and TD

remains suggestive in the YRI/CEU combined father-specific analysis, even though TD does not 'resist' correction for multiple comparisons (Supplementary Figure S2).

Figure 3 provides information on SNP density (Figure 3b) and further loci contained in the chromosomal segment around *SUPT3H* and *RUNX2* (Figure 3c), a close-up view of the results from the CEU father-specific TDT (Figure 3d) and an LD plot based on the same SNPs tested here (Figure 3e). None of the above-mentioned 715 CEU SNPs encode a non-synonymous amino-acid exchange.

Our observations can be illustrated through the following example: tagSNP *rs2677101* is located in the above-mentioned region, and it represents (tags) the information of six SNPs. Regarding this marker, 17 out of the 18 heterozygous fathers transmit the G allele, whereas only one heterozygous father transmits the A allele. This corresponds to a (father-specific) TDT statistics of 14.22 (odds ratio = 17; $P = 0.0002$). To check for possible phasing bias on the original data, we tested TD again considering only the SNPs for which transmission was absolutely indisputable, and re-phased the segment comprising the mentioned 715 CEU SNPs. As results (not shown) were fully compatible with the original data, but implicated the removal of some families, we performed further analyses considering the original HapMap phasing.

Haplotypes of the $S–M–R$ region
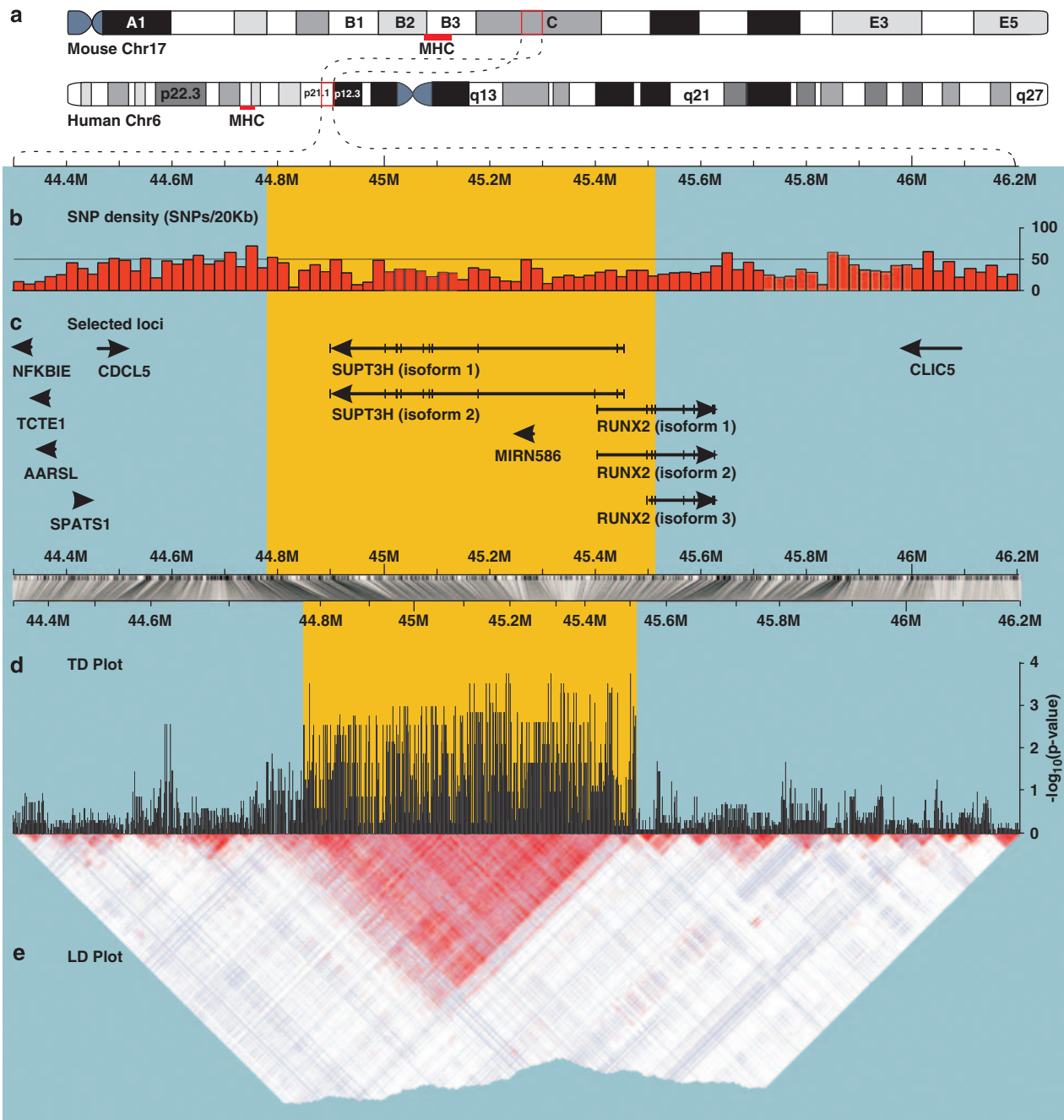Aiming to test whether the transmission pattern on the $S–M–R$ region corresponded to specific haplotypes of

Figure 3 1.9 Mb region of Hsa6p including the *S−M−R* region. The region under evidence of TD in CEU fathers (yellow background) comprises 715 SNPs, spanning 730 741 base pairs. (**a**) Representation of the synteny between mouse Chr17 and human Chr6 with regard to the region between *NFKBIE* and *CLIC5*. For reference, the position of the MHC is indicated in both chromosomes with a red bar. (**b**) SNP density, measured in 20 kb wide bins. (**c**) Localization of selected loci from the region. The arrows indicate transcriptional orientation, and exons are depicted (only for *SUPT3H* and *RUNX2*) through vertical ticks on the lines representing isoforms. (**d**) Plot of the CEU, father-specific *P*-values of the transmission/disequilibrium tests. The peak at the left of the picture, around coordinate 44.62M, corresponds to three isolated SNPs that were not considered further. (**e**) LD plot based on the SNPs from the TD plot, CEU panel. The strong LD spanning the area under the TD peak is evident. Whereas panels b and c are plotted having all base pairs in the region as reference, panels d and e are plotted based only on the HapMap SNPs available for this region. The horizontal bar between panels c and d represents the correspondence between these two different reference systems (scaling). This is why the yellow background is wider above the bar than below it, although it highlights the same genomic region.

the whole region, we performed phylogenetic analyses on its markers, considering the 60 CEU father haplotypes. After removal of S–M–R region-homozygous subjects (fathers having both their haplotypes clustering together in the phylogeny), the resulting phylogenetic tree revealed that the remaining 24 haplotypes occur in two major groups (Supplementary Figure S4), one consisting typically (10 out of 12 times) of transmitted and the other of non-transmitted haplotypes.

Discussion

The preliminary evidence for an excess of allele sharing among siblings regarding chromosome 6,[27] and the extensive synteny between its short arm and mouse chromosome 17[35] with its collection of t haplotypes, suggests the presence of TD encompassing Hsa6p in healthy humans as well. The TD plots shown in this study provide an overview of TD, pinpointing loci with evidence of skewed segregation rates within the CEU (Figure 1) and YRI (Figure 2) populations. Although many loci in all six plots appear to be under TD with $-\log_{10}(P\text{-value}) \geq 2.0$, statistical significance in all cases but one is lost after correction for multiple comparisons. By choosing representative collections of tagSNPs, the problem of multiple comparisons could be overcome, as shown by the simulations that yield statistical significance for CEU father-specific TD at the S–M–R region. It is difficult to quantify the observed results with one single, global P-value, due to the internal structure of the investigated sample (high number of markers related to each other by LD, giving rise to haplotype blocks of variable sizes). We believe, however, that the optimal set of tagSNPs to test for TD corresponds to a k size lying somewhere between the two extremes 1 (good coverage of Hsa6p, but too many single SNPs and short-ranging haplotypes are included in the analysis) and 30 (long-ranging haplotypes, but too few SNPs are included, resulting in low coverage of Hsa6p). This is precisely the span in which significant P-values were observed (Table 1). At this interval of k, our approach leads to plausible significance levels that provide an impression of how unexpected our findings are.

The important advantage of performing permutation tests on tagSNPs is that we can reasonably reduce the data. Moreover, in close analogy to the Moving Block Bootstrap for time series, tagSNPs can be considered independent. This assumption leads to a more powerful method, because it enables to permute haplotypes instead of whole chromosomal set of markers, as proposed in the Monte Carlo procedure by McIntyre.[32]

TD is expected to exert a strong selective force, leaving drastic footprints in the genomic LD landscape. As shown for t haplotypes,[4] TD implies suppression of recombination, and the emergence of a strong and well-delimited LD block. This is in accordance with the strong LD that we observed within the haplotype block determined by the S–M–R region. To test the correspondence between TD and LD in other contexts, we systematically tested LD in the region around eight markers previously reported to be under TD[10] dependent on grandparental origin of alleles. Even though we could not test TD in a similar way as the original study,[10] and did not detect TD for these eight markers in the HapMap trios, in fact, with the exception of one marker (which was not tested due to dubious location in genomic coordinates), all others were found to be located within strong and well-delimited haplotype blocks (see Supplementary Figure S5 for the LD plots). TD may thus be actively shaping the human LD landscape, as its presence within a given genomic region will drive the non-random inheritance of a combination of alleles, thereby giving rise to the emergence of LD. The only plausible explanation for the TD difference found between CEU and YRI is the strong discrepancy regarding allelic frequencies, which leads to the observation that many SNPs under TD in CEU lacked statistical power in YRI due to low MAF. This is illustrated by the fact that 26% of the CEU SNPs with an MAF $\geq 5\%$ are monomorphic in the YRI panel.

Apart from the MIRN586 locus, which specifies a microRNA precursor[36] and might thus play a regulatory role, the S–M–R region contains two genes that encode transcription factors. The human SUPT3H locus gives rise to the transcription factor Spt3, which is a subunit of two GCN5 histone acetyltransferase-containing coactivator complexes: STAGA (SPT3-TAF9-GCN5 acetylase) and TFTC (TBP-free-TAF complex),[37] both involved in histone acetylation, chromatin remodeling and transcriptional regulation.[37,38] In contrast, much more is known about RUNX2. Its product, Runx2, is essential for chondrogenesis, osteogenesis, tooth formation and hematopoiesis.[39–42] By contributing to the retention of gene expression patterns after mitosis, Runx2 participates also in the epigenetic regulation of specific genes during cell division.[43,44] In addition, a recent study found expression of Runx2 in testis and sperm of mice, suggesting a role in the male reproductive system.[45] The promoter region of SUPT3H and both known promoters of RUNX2 are responsible for alternative isoform expression.[46] These regulatory regions are located within the chromosomal segment covered by the 715 SNPs that are subject to CEU male-specific TD. The possibility that the two different haplotype groups identified through the phylogenetic analysis (Supplementary Figure S4) might correspond to an alternative isoform expression of these two genes (Figure 3c) is intriguing, and would, if confirmed, deliver a hint to the biological role of TD in the S–M–R region.

Two association studies addressing the transmission of SNPs of RUNX2 within families have recently been

published. The first[16] focused on affected trios from four populations, and found evidence of linkage of many SNPs in *RUNX2* with cleft lip, but only when parent-of-origin effects were considered (which is indicative of genomic imprinting). We obtained evidence that most of these 11 SNPs can be subject to skewed transmission ratios as well, but only among YRI fathers (Supplementary File S3). In the second study,[15] genotyping of four markers (two linked to each promoter of *RUNX2*, P1 and P2) in Caucasian families showed them to be associated with hand bone phenotypes. In our study, the two P1-linked markers show a tendency toward TD both among CEU fathers ($P = 0.0075$ and 0.0023) and among YRI parents (parent unspecific, $P = 0.0736$ and 0.0164). In case of the two P2-linked markers, only one was present in the HapMap, and we found it to be under TD among CEU fathers ($P = 0.0016$). It remains currently undetermined whether the TD in CEU fathers, which we observed in our study, has the potential to influence the outcome of the two above-mentioned studies.[15,16] If this holds true, one would expect that the excess of maternal transmission observed by Sull *et al*[16] might be even more pronounced than already found to lead to the 'disappearance' of the CEU paternal 'background TD' (the generally present TD, independent of phenotype) in the *RUNX2* region. The absence of coding, non-synonymous polymorphisms among the markers found to be under TD points to the importance of regulatory regions (such as P1 and P2),[15,46] or to a role for the product of the *MIRN586* gene. SNP variants within introns, potential regulatory regions or open reading frames of unknown function have already been associated with phenotypic traits in other studies.[47]

The involvement of Spt3 and Runx2 in chromatin remodeling and/or epigenetic regulation is suggestive of imprinting, but is still poorly explored. The presence of TD in loci subject to genomic imprinting has already been discussed and reported for many human genes.[1,8,10] Naumova *et al*[10] discuss their findings on TD of loci known to be imprinted from the point of view that TD (grandparental origin dependent) results possibly from the loss of embryos carrying genes for which imprinting marks were incorrectly reset. De Villena *et al*[1] propose an alternative way of understanding genomic imprinting, discussing a tight relationship between epigenetics, imprinting and TD. In this study, we suggest adding LD to this group of phenomena that are apparently in constant interaction.

Finally, the presence of two attributes reinforces the evidence that TD is relevant for the *S–M-R* region. First, both the high number of analyzed markers and the limited number of subjects make wide-ranging searches like the one carried out here unlikely to identify loci in which the level of TD is low, although regions worthy of follow-up may be pinpointed.[48] Second, the fact that the HapMap families have been selected based on large sibship might represent a bias against the presence of TD, as large families

are expected to have fewer spontaneous abortions as compared with the general population.[18] The fact that we did observe TD, despite these two forces acting against its detection, provides further support for its existence, at least in CEU fathers. However, as we did not find sufficient evidence for compensating TD among CEU mothers, it remains enigmatic how heterozygosity for the loci within the *S–M–R* region is maintained in the population. A possibly comparable situation exists for t haplotypes on mouse chromosome 17.[4] An explanation for the co-existence of wild-type and t haplotypes has only recently been provided, in the territorial behavior of male mice: heterozygous males show impaired ability to maintain their territories, leading to a significant reproductive disadvantage when compared with males homozygous for the wild-type haplotype.[5] Although several questions remain unanswered,[49] the results of our study contribute to an understanding of the dynamics of LD in the human genome, and may be seen as a first step toward a genome-wide study of TD in healthy individuals. In particular, our approach reveals that the statistical problem of multiple comparisons can be circumvented when carefully chosen tagSNPs are considered instead.

Acknowledgements

Conflict of interest

The authors declare that no competing interests exist.

References

1 de Villena FPM, de la Casa-Esperón E, Sapienza C: Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends Genet* 2000; **16**: 573–579.
2 de Villena FPM, Sapienza C: Nonrandom segregation during meiosis: the unfairness of females. *Mamm Genome* 2001; **12**: 331–339.
3 Lyttle TW: Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet* 1993; **9**: 205–210.
4 Lyon MF: Transmission ratio distortion in mice. *Annu Rev Genet* 2003; **37**: 393–408.
5 Carroll LS, Meagher S, Morrison L *et al*: Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. *Evolution* 2004; **58**: 1318–1328.
6 Ardlie KG, Silver LM: Low frequency of mouse t haplotypes in wild populations is not explained by modifiers of meiotic drive. *Genetics* 1996; **144**: 1787–1797.
7 Bauer H, Véron N, Willert J *et al*: The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev* 2007; **21**: 143–147.
8 Naumova AK, Leppert M, Barker DF *et al*: Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. *Am J Hum Genet* 1998; **62**: 1493–1499.
9 Eaves IA, Bennett ST, Forster P *et al*: Transmission ratio distortion at the INS-IGF2 VNTR. *Nat Genet* 1999; **22**: 324–325.

10 Naumova AK, Greenwood CMT, Morgan K: Imprinting and deviation from Mendelian transmission ratios. *Genome* 2001; **44**: 311–320.

11 Lemire M, Roslin NM, Laprise C et al: Transmission-ratio distortion and allele sharing in affected sib pairs: a new linkage statistic with reduced bias, with application to chromosome 6q25.3. *Am J Hum Genet* 2004; **75**: 571–586.

12 Hanchard N, Rockett K, Udalova I et al: An investigation of transmission ratio distortion in the central region of the human MHC. *Genes Immun* 2006; **7**: 51–58.

13 Dean NL, Loredo-Osti JC, Fujiwara TM et al: Transmission ratio distortion in the myotonic dystrophy locus in human preimplantation embryos. *Eur J Hum Genet* 2006; **14**: 299–306.

14 Alias L, Barceló MJ, Gich I et al: Evidence of a segregation ratio distortion of SMN1 alleles in spinal muscular atrophy. *Eur J Hum Genet* 2007; **15**: 1090–1093.

15 Ermakov S, Malkin I, Keter M et al: Family-based association study of polymorphisms in the RUNX2 locus with hand bone length and hand BMD. *Ann Hum Genet* 2008; **72**: 510–518.

16 Sull JW, Liang KY, Hetmanski JB et al: Differential parental transmission of markers in RUNX2 among cleft case-parent trios from four populations. *Genet Epidemiol* 2008; **32**: 505–512.

17 Greenwood CM, Morgan K: The impact of transmission-ratio distortion on allele sharing in affected sibling pairs. *Am J Hum Genet* 2000; **66**: 2001–2004.

18 Paterson AD, Sun L, Liu XQ et al: Transmission ratio distortion in families from the Framingham Heart Study. *BMC Genet* 2003; **4**: S48.

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

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

21 International HapMap Consortium: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**: 851–861.

22 Horton R, Wilming L, Rand V et al: Gene map of the extended human MHC. *Nat Rev Genet* 2004; **5**: 889–899.

23 Ober C, Hyslop T, Elias S et al: Human leukocyte antigen matching and fetal loss: results of a 10 year prospective study. *Hum Reprod* 1998; **13**: 33–38.

24 Kruse C, Steffensen R, Varming K et al: A study of HLA-DR and -DQ alleles in 588 patients and 562 controls confirms that HLA-DRB1*03 is associated with recurrent miscarriage. *Hum Reprod* 2004; **19**: 1215–1221.

25 Ziegler A, Kentenich H, Uchanska-Ziegler B: Female choice and the MHC. *Trends Immunol* 2005; **26**: 496–502.

26 Ilmonen P, Penn DJ, Damjanovich K et al: Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics* 2007; **176**: 2501–2508.

27 Zöllner S, Wen X, Hanchard NA et al: Evidence for extensive transmission distortion in the human genome. *Am J Hum Genet* 2004; **74**: 62–72.

28 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.

29 Ziegler A, König IR: A Statistical Approach to Genetic Epidemiology: Concepts and Applications. Weinheim: Wiley-VCH, 361p.

30 de Bakker PIW, Yelensky R, Pe'er I et al: Efficiency and power in genetic association studies. *Nat Genet* 2005; **37**: 1217–1223.

31 Barrett JC, Fry B, Maller J et al: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

32 McIntyre LM, Martin ER, Simonsen KL et al: Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. *Genet Epidemiol* 2000; **19**: 18–29.

33 Shi M, Umbach DM, Weinberg CR: Identification of risk-related haplotypes with the use of multiple SNPs from nuclear families. *Am J Hum Genet* 2007; **81**: 53–66.

34 Becker T, Knapp M: A powerful strategy to account for multiple testing in the context of haplotype analysis. *Am J Hum Genet* 2004; **75**: 561–570.

35 Graves JAM, Wakefield MJ, Peters J et al: *Report of the committee on comparative gene mapping; in: Human Gene Mapping*. Johns Hopkins Press: Baltimore, 1996, pp 1351–1408.

36 Cummins JM, He Y, Leary RJ et al: The colorectal microRNAome. *Proc Natl Acad Sci USA* 2006; **103**: 3687–3692.

37 Nagy Z, Tora L: Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* 2007; **26**: 5341–5357.

38 Martinez E, Palhan VB, Tjernberg A et al: Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors *in vivo*. *Mol Cell Biol* 2001; **21**: 6782–6795.

39 Aberg T, Wang XP, Kim JH et al: Runx2 mediates FGF signaling from epithelium to mesenchyme during tooth morphogenesis. *Dev Biol* 2004; **270**: 76–93.

40 Stein GS, Lian JB, van Wijnen AJ et al: Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. *Oncogene* 2004; **23**: 4315–4329.

41 de Bruijn MFTR, Speck NA: Core-binding factors in hematopoiesis and immune function. *Oncogene* 2004; **23**: 4238–4248.

42 Okumura AJ, Peterson LF, Lo MC et al: Expression of AML/Runx and ETO/MTG family members during hematopoietic differentiation of embryonic stem cells. *Exp Hematol* 2007; **35**: 978–988.

43 Young DW, Hassan MQ, Pratap J et al: Mitotic occupancy and lineage-specific transcriptional control of rRNA genes by Runx2. *Nature* 2007; **445**: 442–446.

44 Young DW, Hassan MQ, Yang XQ et al: Mitotic retention of gene expression patterns by the cell fate-determining transcription factor Runx2. *Proc Natl Acad Sci USA* 2007; **104**: 3189–3194.

45 Jeong JH, Jin JS, Kim HN et al: Expression of Runx2 transcription factor in non-skeletal tissues, sperm and brain. *J Cell Physiol* 2008; **217**: 511–517.

46 Stock M, Otto F: Control of RUNX2 isoform expression: the role of promoters and enhancers. *J Cell Biochem* 2005; **95**: 506–517.

47 Styrkarsdottir U, Halldorsson BV, Gretarsdottir S et al: Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 2008; **358**: 2355–2365.

48 Evans DM, Morris AP, Cardon LR et al: A note on the power to detect transmission distortion in parent-child trios via the transmission disequilibrium test. *Behav Genet* 2006; **36**: 947–950.

49 Baker AEM: Mendelian inheritance of t haplotypes in house mouse (*Mus musculus domesticus*) field populations. *Genet Res* 2008; **90**: 331–339.