

Linkage disequilibrium analysis identifies an *FGFR1* haplotype-tag SNP associated with normal variation in craniofacial shape

By Anna K. Coussens and Angela van Daal

Cooperative Research Centre for Diagnostics, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia
(Correspondence to Angela van Daal)

Abstract

Mutations in *FGFR1* and *TWIST1* have been reported to affect the timing of calvarial suture fusion resulting in craniosynostosis and facial abnormalities. We screened nonpathologic populations for genetic polymorphisms that may associate with normal craniofacial variation. We identified 17 single-nucleotide polymorphisms (SNPs) in *FGFR1*, 6 of which were novel (g.8591855G→A, g.8593685G→A, g.8602303C→T, g.8602475A→G (p.Ile293Val), g.8605849C→T, g.8607868G→A). No SNPs were found in *TWIST1*. *FGFR1* SNP haplotypes were reconstructed for Caucasian, Asian, Australian Aboriginal, and African American populations. All populations shared two linkage disequilibrium blocks, with one haplotype-tag SNP (htSNP) tagging each block. The htSNP g.8592931G→C was found to have a significant negative correlation with the cephalic index for all populations ($R = -0.187$, $p = 0.036$), with larger correlations in Asians and females. This finding is a starting point in the identification of a set of SNPs that can be genotyped to determine both normal and disease craniofacial phenotypes.

Human facial morphology is a complex physical trait controlled by genetic, environmental, mechanical, and epigenetic factors [1]. Numerous twin and familial studies have demonstrated that there is a clear genetic component to craniofacial measurements [2], [3] and [4]. The extent of craniofacial diversity that exists is testimony to a complex network of interactions, involving many genes. The small number of genes so far identified exhibit a significant amount of redundancy [5], suggesting the possibility of an accumulation of a large number of polymorphisms in genes involved in craniofacial development. It is suggested that most single-nucleotide polymorphisms (SNPs) arose after speciation, but before the separation of distinct populations [6]. This may explain why most facial types are seen in all populations [7]. Identifying the genetic variation that leads to specific craniofacial phenotypes would enable more effective diagnosis and treatments of cranial malformations and associated diseases, such as obstructive sleep apnea.

The development of specific facial types is governed by the shape of the head/skull [1]. At birth the cranial vault, which surrounds the brain, is made up of seven bones (calvaria), with adjoining bones separated by fibrous joints called calvarial sutures. Patients with craniosynostosis, the premature fusion of calvarial sutures, present with mild to severe facial abnormalities, including long or beaked noses; broad eyes; large brow ridges; extremely broad (brachycephalic), narrow (dolichocephalic), or asymmetrically shaped (plagiocephalic) skulls; and possible mental retardation (reviewed by Cohen [8] and Muenke and Wilkie [9]). Linkage studies have identified point mutations in six genes that cause one or more craniosynostosis syndromes (reviewed by Wilkie [10]). Identical mutations can result in the development of a range of different phenotypes, or even syndromes [11], [12], [13] and [14], suggesting that the genetic background influences phenotype expression [15]. Type 1 Pfeiffer syndrome (MIM 101600) and Saethre–

Chotzen syndrome (MIM 101400) patients have mutations in fibroblast growth factor receptor 1 (*FGFR1*) (MIM 136350) and *TWIST1* (MIM 601622), respectively. Patients with mutations in these genes may present with milder phenotypes compared to other syndromes, depending on the mutation. Mutations in *FGFR1* also result in Kallmann syndrome 2 (KAL2) (MIM 147950). Only *FGFR1* Kallmann patients present with occasional cleft lip or palate and dental agenesis [16]. The *FGFR1* gene, located at 8p11.2–p11.1, contains 19 exons spanning 55 kb and has at least nine known isoforms. It is a cell surface receptor that signals via two major transduction pathways: Ras/mitogen-activated protein kinase and PLC γ pathways (reviewed by Groth and Lardelli [17]). *TWIST1*, located at 7p21, contains 2 exons and spans 2.2 kb. It is a highly conserved basic helix-loop-helix (bHLH) transcription factor. Expression of these two genes in craniofacial primordia and mutations in both resulting in facial malformations indicate they are good candidate genes to screen for SNPs that regulate craniofacial morphogenesis.

We have identified SNP profiles of *FGFR1* and *TWIST1* in nondisease populations, screening individuals with cephalic index measurements in the extreme ends of normal variation. We reconstructed SNP haplotypes for a number of populations, identified haplotype-tag SNPs (htSNPs), and analyzed them for associations with the cephalic index and specific facial phenotypes. We found no significant SNP profiles in *TWIST1*, but identified two linkage disequilibrium (LD) blocks in *FGFR1* and two htSNPs, one of which shows association with a decrease in the cephalic index, the standard measure of head shape. The results from this study indicate that SNPs that regulate normal craniofacial variation as well as specific phenotypes of craniosynostosis and other cranial pathologies can potentially be identified.

Results

Twenty-two PCR fragments covering two genes were analyzed by denaturing high-performance liquid chromatography (dHPLC) in an attempt to identify SNPs that influence normal craniofacial variation. Eight base differences were identified in *TWIST1* (including rs2522205, rs2522206, rs2522201, rs2522202, rs2522203, and rs2522204), relative to the reference sequences, in all samples (Fig. 1). However, heterozygosity was not observed. A BLAST sequence search (BLAST 2.2.9, www.ncbi.nlm.nih.gov/BLAST) of the current human genome database (build 34, March 10, 2004), using sequence X91662, and subsequent sequence alignment with the resultant six *TWIST1* sequences, indicated that our two chosen *TWIST1* reference sequences (X91662 and Y10871) were the only ones that differed from the samples sequenced in this study and the other database sequences. The reference sequence of *TWIST1* should be AC003986.

Table 1 shows the 17 SNPs identified in *FGFR1* (Fig. 2). Fifteen of the 17 polymorphisms detected are transitions and 2 are transversions, 6 are novel, and the remainder are described in the NCBI dbSNP database. Seven exonic SNPs were identified, including 1 novel nonsynonymous change, p.Ile293Val; 4 synonymous changes; and 2 in the 5'UTR. p.Ile293Val was seen in only 1 of the 44 individuals screened and was not analyzed further. Eight SNPs had a heterozygote frequency greater than 5% and were genotyped in a further 30 individuals.

Table 1.Seventeen SNPs identified in *FGFR1* by dHPLC analysis

SNP ID	Site	Description ^a	Predicted protein ^b	dbSNP ID No. (NCBI)	Heterozygosity (<i>n</i> = 44)	Genotyping method
1	Exon 18	g.8591855G→A		Novel	0.091	<i>AvaII</i> ^c
2	Intron 17	g.8591936C→T		rs4647904	0.091	<i>TauI</i> ^d
3	Intron 14	g.8592931G→C		rs4647905	0.273	<i>Cac8I</i> ^c
4	Intron 14	g.8593685G→A		Novel	0.023	
5	Intron 8	g.8599907C→T		rs4647910	0.023	
6	Intron 7	g.8602047C→T		rs3925	0.045	
7	Exon 7	g.8602303C→T		Novel	0.023	
8	Exon 7	g.8602475A→G	p.Ile293Val	Novel	0.023	
9	Intron 5	g.8604400G→A		rs2293971	0.273	<i>Kpn2I</i> ^d
10	Exon 5	g.8605849C→T		Novel	0.023	
11	Intron 3	g.8607200G→C		rs2304000	0.273	<i>BanII</i> ^c
12	Exon 3	g.8607602C→T		rs2915665	0.018	
13 ^f	Intron 2	g.8607868G→A		Novel	0.018	
14	Intron 2	g.8607944G→A		rs4647907	0.273	dHPLC mixing
15	5'UTR	g.8646085C→T	5'UTR	rs4647909	0.045	
16	5'UTR	g.8646435C→T	5'UTR	rs3213849	0.409	<i>SmaI</i> ^c
17	Promoter	g.8646793A→G	Promoter	rs930828	0.432	Kinetic PCR

^a Nucleotide numbering according to GenBank Accession No. NT_007995.14. *Note.* SNPs with heterozygosity greater than 0.05, used for LD and haplotype analyses, are in bold.

^b Amino acid numbering according to GenBank Accession No. NP_000595.

^cNEB, Beverly, MA, USA.

^d Fermentas, Hanover, MD, USA.

^e Roche Diagnostics, Castle Hill, NSW, Australia.

^f g.8607868G→A was identified during the dHPLC mixing experiment to genotype g.8607944G→A (*n* = 55)

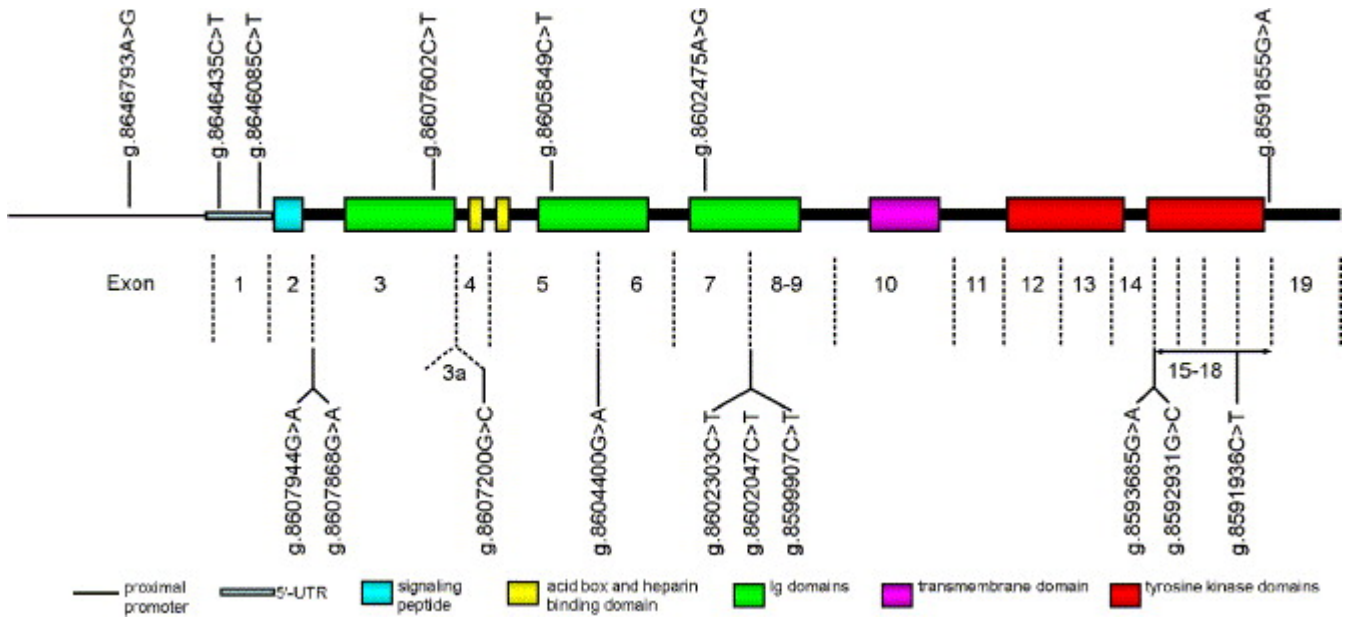


Fig. 2.

FGFR1 protein with the position of exon boundaries relative to encoded protein sequence and identified SNP locations (modified from Groth and Lardelli [17], used by permission of the publisher). Exonic SNPs are above and intronic SNPs are below. Exons 8 and 9 encode alternatively spliced C-terminal halves of the third Ig domain. Isoforms 5 and 6 have an extended exon 7, which prematurely truncates in intron 7; g.8602303C→T is in the extended exon. g.8599907C→T is in intron 8. Isoform 8 uses the additional exon, exon 3a, within intron 3, resulting in a truncated protein. g.8607200G→C is 3 bp downstream of the exon 3a stop codon. Domain sizes are not to scale.

Haplotype and LD analysis

Haplotypes were reconstructed for the eight common ($f > 0.05$) *FGFR1* SNPs that span 55 kb, all of which are in HWE. The 74 genotypes were analyzed as one population for haplotype reconstruction. The PHASE program identified 13 best reconstruction haplotypes, 9 present within the sample analyzed, with 3 major haplotypes representing >87% of the sample (Table 2). These three major haplotypes are also the three major haplotypes in each population analyzed (African American, Australian Aboriginal, Asian, and Caucasian), supporting the use of all samples as one population for initial haplotype reconstruction. SNPTagger identified two htSNPs, representing >90% of all haplotypes within the sample: g.8592931G→C (rs4647905) and g.8646435C→T (rs3213849). Haplotypes from the four populations were also analyzed separately, resulting in the selection of the same two htSNPs representing >80% Caucasian, >85% African American, and 100% of Asian and Australian Aboriginal haplotypes. These two htSNPs distinguish between the three major haplotypes. Exact population differentiation based on haplotype frequencies found African Americans to be significantly different ($p < 0.05$) from Caucasians and Australian Aboriginals, while Asians and African American haplotype frequencies did not quite reach a significant difference ($p = 0.054$) (Table 3). Population differentiation based on the frequency of the two htSNPs within each population was similarly examined and resulted in nearly identical p values (Table 3), indicating that the pair-wise LD pattern and haplotype structure seen in each population are similar and represented well by the same two htSNPs.

Table 2.

Three haplotypes account for >80% of population variation for *FGFR1* SNPs

Haplotype ^a	SNP1	SNP2	SNP3	SNP9	SNP11	SNP14	SNP16	SNP17	Frequency ^b	Cumulative frequency	Caucasian (n = 45)	Asian (n = 10)	Australian Aboriginal (n = 10)	African American (n = 8)
H1	0	0	0	0	0	0	0	0	0.385	0.385	0.367	0.300	0.350	0.625
H2	0	0	0	0	0	0	1	1	0.318	0.703	0.356	0.250	0.400	0.063
H3	0	0	1	1	1	1	0	0	0.169	0.872	0.111	0.450	0.150	0.188
H4	0	0	1	1	1	1	1	1	0.034	0.905	0.033		0.100	
H5	0	1	1	1	1	1	0	0	0.034	0.939	0.056			
H6	1	0	0	0	0	0	0	0	0.027	0.966	0.044			
H7	0	0	0	0	0	0	0	1	0.014	0.980	0.011			0.063
H8	0	0	0	1	1	1	0	0	0.014	0.993	0.022			
H9	0	0	0	1	0	0	1	1	0.007	1.000				0.063

^aMajor alleles are denoted as 0 and minor alleles as 1. htSNPs are in bold. SNP numbering from Table 1.

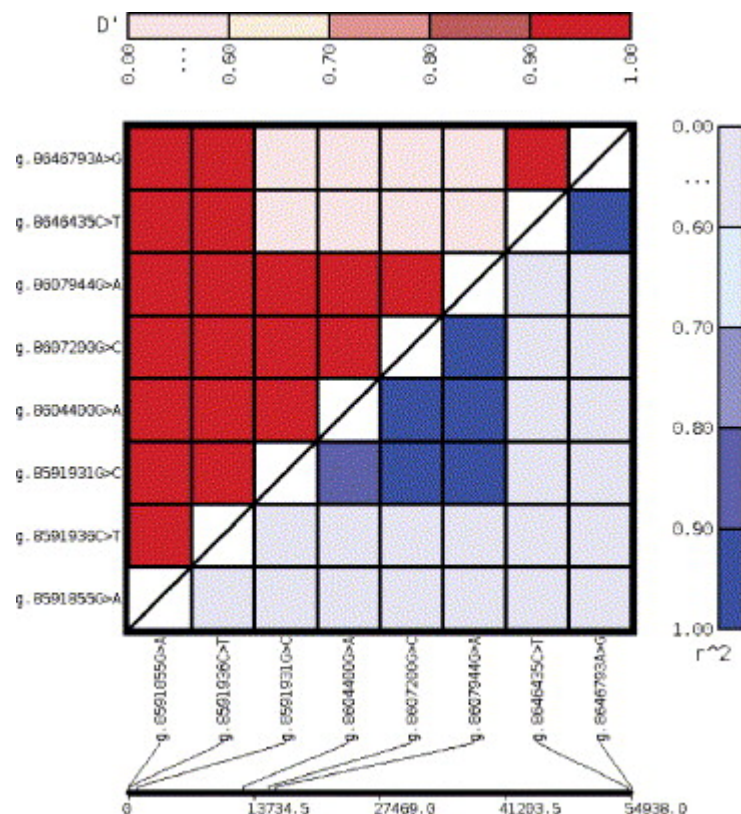
^bFrequency of each haplotype within the sample of 74 individuals (148 haplotypes), listed in decreasing order. The three major haplotypes accounting for >87% of the haplotypes in the sample are the three major haplotypes in the four populations analyzed. One Indian was also present in the 74 individuals and had haplotypes H1 and H2.

Table 3.

Population differentiation and standard neutral model measures

Population	Caucasian	Asian	Australian Aboriginal	African American
Caucasian		0.0869	0.8051	0.0357
Asian	0.0754		0.1390	0.0523
Australian Aboriginal	0.8083	0.1304		0.0465
African American	0.0334	0.0554	0.0442	
Neutral models				
Tajima's <i>D</i>	0.8733	0.9739	0.9400	0.5991
Chakraborty's	0.6250	0.9241	0.8545	0.2458

htSNP *p* values above diagonal and eight loci haplotype *p* values below diagonal, for population differentiation based on haplotype frequencies. Neutral model measures' *p* values are also indicated. Significant *p* values (<0.05) are in bold.

**Fig. 3.**

Pair-wise linkage disequilibrium between the eight common *FGFR1* SNPs. Samples were analyzed as one population.

The pattern of pair-wise LD between the eight common SNPs is shown in Fig. 3. The LD pattern was calculated for all populations as one sample, since they have near-identical pair-wise LD strengths. Two distinct LD blocks exist. LD block 1 encompasses g.8592931G→C, g.8604400G→A, g.8607200G→C, and g.8607944G→A, which are in complete LD ($D' = 1$ and $r^2 \geq 0.89$, $p < 0.0001$), with g.8607200G→C and g.8607944G→A in absolute LD ($r^2 = 1$). Two SNPs being in absolute LD when four populations are combined emphasizes the fact that the LD patterns are highly conserved between populations. g.8591936C→T is also in LD with SNPs in this block, but is much weaker than the rest, with $D' = 1$ and $r^2 \leq 0.113$ ($p \leq 0.0009$; significant at the Bonferroni-adjusted level, 0.0018). The second LD block comprises g.8646435C→T and g.8646793A→G ($D' = 1$, $r^2 = 0.943$, $p < 0.0001$). Limited LD exists between SNPs of the two blocks ($D' < 0.60$, $r^2 < 0.081$).

Two measures to detect departure from the standard neutral model were calculated for each population (Table 3). No population gave significant values for Tajima's D ($p > 0.599$), a model of infinite loci, or for Chakraborty's test for infinite alleles ($p > 0.245$) (Table 3). These measures indicate that the SNPs within these loci are acting under neutral selection.

Association analyses

The two htSNPs that distinguish the three major haplotypes of each population were analyzed for association with the cephalic index measure and various facial variations. Initial analysis using χ^2 tests for cephalic index, head shape, and face shape indicated no associations with g.8646435C→T. Exploratory analysis of a further 25 facial variations also found no associations. Conversely, g.8592931G→C appeared to be associated with the cephalic index, face shape, and mandible shape. A further 100 people were genotyped for g.8592931G→C, increasing the sample size to 133 for cephalic index (CI) measurements and 154 for photographs. The means of the cephalic index, grouped by genotypes, were distributed differently for each population (Fig. 4). Consequently, stepwise linear regression was performed on CI values represented by ranked normal scores to analyze CI association with g.8592931G→C. A significant coefficient, $R = -0.182$ ($p = 0.036$), was produced, indicating that the g.8592931G→C variant allele is associated with a decrease in CI. Separating males ($n = 54$) and females ($n = 79$) resulted in significance for females only, $R = -0.242$, $p = 0.031$ compared to $R = -0.083$, $p = 0.548$ for males. g.8592931G→C also showed a greater correlation coefficient for Asian samples ($R = -0.770$, $p = 0.043$). Asians normally have a CI > 80 ; however, both homozygous C individuals had CI < 80 (Fig. 4). Linear regression was also carried out for dominant and recessive inheritance effects of g.8592931G→C, with dominant inheritance significantly associated with CI ($R = -0.175$, $p = 0.045$). For exploratory interest, the head shape categories were analyzed using multinomial regression and ANOVA by population groups, but no significance was found. Face shape was similarly analyzed for each population. Caucasians had a significant association between face shape and the dominant model of g.8592931G→C ($p = 0.024$), but not the additive model. This result indicated that Caucasians with at least one copy of g.8592931G→C are more likely to have a long and narrow or oval face. These post hoc analyses provide an initial indication that the g.8592931G→C-associated reduction in CI has an affect on face shape and should be confirmed using larger population groups.

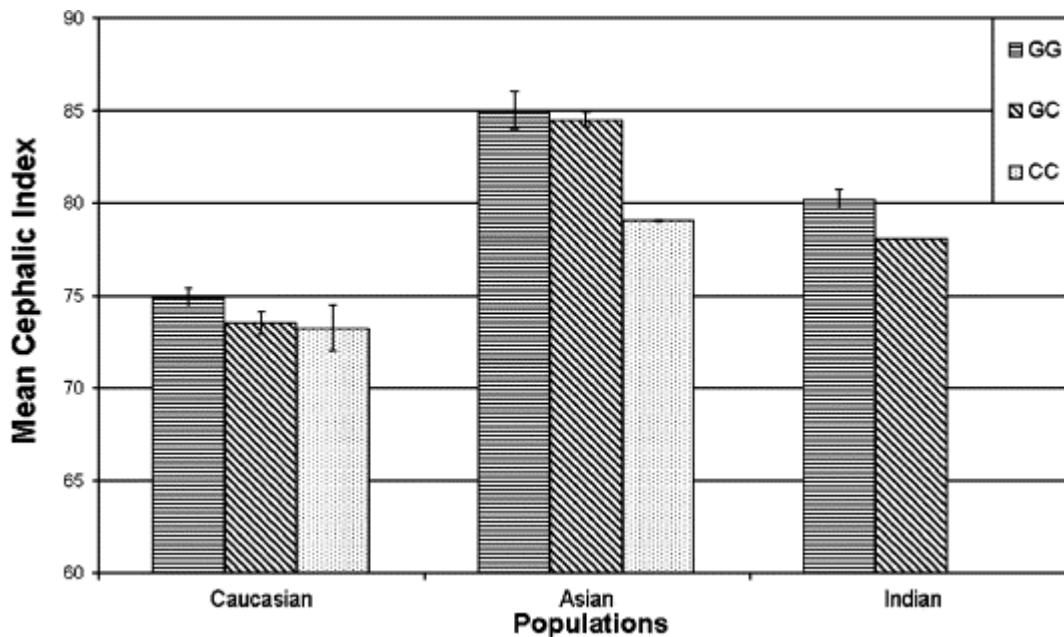


Fig. 4.

The distribution of mean cephalic index (CI) for g.8592931G→C genotypes within three populations. The mean CI for each population is different and therefore CI values must be normalized and ranked for combined population data to be analyzed. The reducing effect of the C allele on mean CI can be observed for each population, acting in a linear additive state. There were no homozygous C Indians and Australian Aboriginals and African Americans were not included due to the small number that had measurements taken.

Discussion

The genetic control of craniofacial development has been the focus of a large number of studies in recent years (reviewed by Francis-West and colleagues [18]). The majority of these studies analyzed embryonic development of facial primordia. Few published studies have focused on the identification of polymorphisms that regulate normal craniofacial variation. One such study compared the sequence of the TCOF1 gene in different dog breeds, identifying a nonsynonymous base change in breeds with brachycephalic head shapes [19]. This supports the likelihood of genetic polymorphisms in humans that also affect head shape. A multitude of studies have analyzed patients with syndromic and nonsyndromic craniosynostosis in an attempt to identify causative mutations in *FGFR1* and *TWIST1*, as well as multiple other genes (reviewed by Wilkie [10]). There are no such published reports describing nucleotide polymorphisms in normal human populations and their relationship to facial phenotypes. Only two coding SNPs exist on dbSNP for *FGFR1*; one was identified from in silico clone fragment overlap analysis and the other (rs2915665) was verified in this study. The nonsynonymous SNP g.8602475A→G identified in this study is not present. Similarly, all but two SNPs listed for *TWIST1* were identified from clone overlap analysis. Currently, there are no htSNPs on the HAPMAP database in *TWIST1* and although there are multiple htSNPs in *FGFR1*, frequency data are available only for the Caucasian population.

We used dHPLC analysis to identify common polymorphisms within coding, flanking intronic, and putative proximal promoter regions of *TWIST1* and *FGFR1* in normal populations. No SNPs were identified in *TWIST1* and sequence alignment indicated that six of the dbSNPs are not commonly polymorphic. More than 50 *TWIST1* coding mutations that result in the development of craniosynostosis have been described [11], [20], [21], [22] and [23]. While this suggests that the *TWIST1* locus may be a good candidate for SNP detection, it appears that the highly conserved

bHLH domain does not allow the accumulation of nonpathologic SNPs. Common polymorphisms affecting normal craniofacial variation were therefore not found in 44 samples from a range of geographic populations.

FGFR1 has only one recurrent gain-of-function mutation (Arg252Pro) identified, resulting in craniosynostosis [24]. Recent studies have identified 22 nonrecurring loss-of-function point mutations in *FGFR1* resulting in KAL2 [16], [25] and [26]. Similar to *TWIST1*, due to the number of mutations already identified in this gene we predicted that *FGFR1* may accumulate SNPs that do not result in a pathologic deformation, but may influence normal variation. In support of this, we have identified 17 SNPs within the *FGFR1* genomic sequence, 6 of which were novel (Table 1). All, but one, were identified during the initial screening stage. g.8607868G→A was the only SNP to be identified in the additional 30 genotyped samples, many of which were sequenced. This suggests that the initial screening set of 44 individuals was a representative sample for common *FGFR1* polymorphisms. One of the novel SNPs was a nonsynonymous Ile293Val substitution. It is in the IgIII domain downstream of the Pfeiffer syndrome mutation Pro252Arg, which occurs in the IgII–III linker region [24]. The individual with this SNP has not been diagnosed with any form of craniosynostosis or KAL2. The individual is unaffected presumably because the conservative amino acid change does not alter *FGFR1* function.

The specific promoter region of human *FGFR1* has yet to be fully identified. Recent studies in chickens suggest multiple Sp1 and Sp3 elements between –69 and –14 and two Sp1 binding sites more than 1 kb upstream of the transcription start site activate the *FGFR1* promoter [27] and [28]; g.8646793A→G at –265 is potentially within the human *FGFR1* promoter and g.8646435C→T and g.8646085C→T are in the 5'UTR. These SNPs may therefore affect *FGFR1* expression. g.8607200G→C is found in intron 3, 3 bases downstream from the exon 3a stop codon (Fig. 2). Isoform 8 is the only transcript that uses this alternate exon, producing a prematurely truncated protein that is thought to act as a dominant negative repressor of *FGFR1* function (reviewed by Groth and Lardelli [17]). Current work within our laboratory suggests that an additional isoform, which also uses exon 3a, but does not truncate at the stop codon, exists (unpublished results). g.8607200G→C therefore lies in the extended exon 3a sequence, which possibly encodes an additional dominant negative isoform. Whether g.8607200G→C affects the function of this isoform is unknown and further work is needed to clarify this.

Eight of the 17 SNPs identified had a heterozygote frequency >0.05 in the 44 individuals screened, including 3 of the SNPs located in potential regulatory or coding regions. Haplotype reconstruction of the 8 SNPs identified three common haplotypes among all populations and 2 htSNPs (g.8592931G→C and g.8646435C→T) describing more than 80% of the haplotypes in each population (Table 2). Overall, the representation of each population by the same 2 htSNPs makes these SNPs extremely useful for global association studies.

Pair-wise linkage disequilibrium analysis identified two distinct LD blocks within *FGFR1*, with one htSNP each (Fig. 3). The first LD block between g.8592931 and g.8607944 encompasses 15 kb, with g.8592931G→C tagging it. The second LD block includes g.8646435C→T (htSNP) and g.8646793A→G, which are separated by 357 bp. There are weak pair-wise LD values between SNPs in the two blocks, suggesting the possibility of a recombination hot spot between exon 1 and exon 3, the location of the two innermost SNPs of each block (Fig. 2). Intron 1 spans 10.4 kb and intron 2 spans 27.4 kb, with the 178-bp exon 2 between them. The likely low selection pressure acting on this large intronic region makes it a good candidate for a recombination hot spot. LD information exists in the HAPMAP database (www.hapmap.org) for g.8592931G→C, g.8604400G→A, and g.8602047C→T. LD values verify for g.8592931G→C and

g.8604400G→A ($r^2 = 0.95$), but disagree for g.8602047C→T ($r^2 = 0.91$ with g.8592931G→C), which we did not genotype further due to low heterozygote frequency. LD information is present only for the Utah-derived Caucasian population (CEU), and our second htSNP g.8646435C→T is not included.

A limitation of this study is the small population sizes for the non-Caucasian populations. However, the fact that they all share the same three common haplotypes and the same two htSNPs suggests that the samples are a good representation of variation seen among the populations. Also, since near-identical pair-wise LD exists in all populations, samples can be grouped for haplotype reconstruction, even though it has been shown that different haplotype structures often exist between populations [29] and [30]. Exact population differentiation also determined the African American haplotype frequencies to be significantly different from the Caucasian and Australian Aboriginal haplotype frequencies. This is the commonly observed trend in craniometric and other genetic marker population studies [31], [32], [33], [34] and [35], again supporting the idea that the population sizes were not significantly affecting the results.

This is the first study to analyze high-frequency htSNPs for an association with craniofacial variation in nonpathologic human populations. The eight common SNPs in *FGFR1* potentially make up part of the genetic background that regulates craniofacial shape. Initial phenotype association analysis of htSNP g.8646435C→T was not able to detect an association with any facial characteristics, but further studies may find a correlation with disease phenotypes.

Haplotype-tag SNP g.8592931G→C was found to have a significant correlation with the cephalic index, the primary measure of head shape. Analyzing ranked normal scores for CI values produced a correlation coefficient, $R = -0.187$ ($p = 0.036$), indicating a significant negative association of g.8592931G→C with the cephalic index. Dominant and recessive inheritance models were similarly analyzed, but were less significant, suggesting it acts in an additive manner. This correlation can be seen in Fig. 4, with the mean CI of each genotype gradually decreasing for heterozygous and then homozygous variant (C) individuals. The largest effect was seen in Asians ($R = -0.770$, $p = 0.043$), although the sample size was small and further analysis is required. Additionally, the finding of a significant association in females only ($p = 0.031$) is interesting given the findings of a recent twin study, which found that sons show a higher genetic component for head size measures than daughters [4]. This difference could be due to the smaller number of males sampled in our study compared to females or uneven population distribution of sexes, or it could reflect a true sexual dimorphism.

No association of g.8592931G→C with the three head shape categories was found, which is likely to be due to the large groupings of this variable not being significantly affected by the small g.8592931G→C-associated reduction in CI. Six face shape categories were analyzed for association: long and narrow, oval, diamond, egg-shaped, round and short, and square. Caucasians were found to have a significant correlation between narrow and oval faces and the dominant form of g.8592931G→C. These are only exploratory analyses as multiple testing was not taken into account and greater numbers, especially for other populations, are needed to carry out meaningful analysis of these six face shape categories. Considering that craniometric traits undergo neutral selection [7] and [36] it was expected that our SNPs related to craniometric variation would also undergo neutral selection; this was the case.

The four SNPs in LD block 1, tagged by g.8592931G→C, are all associated with a reduction in CI. The other members of the LD block were not analyzed for phenotype association as their frequencies are not significantly different from that of g.8592931G→C. Since they are all intronic it is unclear how they would affect *FGFR1* function. Debate exists regarding the

functionality of intronic regions and it has been suggested that introns may indeed play a role in regulation of gene expression (reviewed by Makalowski [37] and Wickelgren [38]). It can therefore not be discounted that these intronic SNPs are playing a role in regulating *FGFR1*. There is also the possibility that further isoforms exist that use the intronic regions containing these SNP as coding regions. Our analysis clearly identified *FGFR1* htSNPs useful for other such association studies, adding to the vast amount of SNP information being compiled by the HAPMAP project. The significant correlation between the htSNP of *FGFR1* and the measure of cranial shape suggests that the genetic background of *FGFR1* affects normal craniofacial variation. This work emphasizes that whole gene scans are likely to identify an array of SNPs that regulate craniofacial shape.

Material and methods

Samples

DNA was extracted from blood from healthy Caucasian (137), Asian (13), Australian Aboriginal (11), African American (8), and Indian (5) individuals by standard methods. Forty-four samples (Caucasian (35), Asian (7), Australian Aboriginal (1), African American (1)) from 22 dolichocephalic and 22 brachycephalic individuals were used for initial polymorphism screening. The individuals were chosen to represent the extreme ends of normal variation in the cephalic index measurement in an attempt to identify common SNPs regulating craniofacial shape. A further 30 individuals (Caucasian (10), Asian (3), Australian Aboriginal (9), African American (7), Indian (1)) were genotyped for haplotype reconstruction, while an additional 100 were genotyped for one htSNP for association analyses. Caucasian samples were obtained from the Australian public, a large proportion of which are of British Isles ancestry. Asian samples were also obtained from the Australian public and included six of South-East Asian and four of East Asian ancestry. African American samples were obtained from Tamyra Moetti (Federal Bureau of Investigation, Quantico, VA, USA). Australian Aboriginal samples were obtained from Leo Freney (Queensland Health, John Tonge Centre, Brisbane, QLD, Australia). Admixture data for the latter two populations are unavailable. This work was approved by the Queensland University of Technology Ethics Committee (1743/1H).

Phenotype analysis

Cranial measurements euryon–euryon (eu-eu) and glabella–opisthocranium (g-op) were taken by standard anthropometric techniques using spreading calipers. Frontal and profile photographs were taken. The CI was calculated as $(eu-eu \times 100)/g-op$. CI measurements were grouped into three categories describing head shape: dolichocephalic (74.9 or less), mesocephalic (75.0–79.9), and brachycephalic (80.0 or greater) [39]. Twenty-seven topographical facial variations, segregated into multiple variants, were analyzed from the photographs [1] and [40]. Facial variants analyzed included face shape, forehead angle, nose shape, nasal septum slope, mandible protrusion, mandible shape, and eye shape.

SNP identification by dHPLC analysis

DNA samples were analyzed for mutations in *TWIST1* and *FGFR1* using a Prostar Helix dHPLC system (Varian, Palo Alto, CA, USA). Twenty-three primer pairs (Table 4) were designed for amplification of 20 *FGFR1* exons (including all isoform variants) and the 2 *TWIST1* exons, as well as putative promoter and flanking intronic sequences. Reference sequences with GenBank Accession Nos. X91662.1 and Y10871.1 were used for the *TWIST1* promoter and exon 1 and for exon 2, respectively; NT_007995.14 was used for *FGFR1*. PCR amplicons were designed to be smaller than 750 bp for optimal dHPLC analysis. As a result of the high GC content of *TWIST1*, nested PCR using 5% (v/v) DMSO for internal PCR was

implemented for *TWIST1* fragments, using 1/500 dilutions of the external 2230-bp amplicon. Expand Long Template polymerase (1.75 units; Roche Diagnostics, Castle Hill, NSW, Australia), 1× PCR buffer 3 (2.75 mM MgCl₂), 0.4 mM each dNTP, 10% (v/v) DMSO, 0.3 μM each primer, and 75–300 ng DNA in a 50-μl reaction were used for the external product, following recommended cycling conditions (Table 4). Standard PCRs for other fragments were performed in 50-μl reactions including 1 unit Platinum Taq DNA polymerase (Invitrogen, Melbourne, VIC, Australia), 1× PCR buffer (20 mM Tris–HCl (pH 8.4), 50 mM KCl), 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.4 μM each primer, and 50–200 ng DNA. Cycling conditions were 5 min at 95°C, cycling 40 s at 95°C, 30 s at the determined T_m (Table 4), 40 s at 72°C, and final extension of 5 min at 72°C. Prior to dHPLC analysis, PCR products underwent an initial denaturation followed by a reannealing phase from 95 to 65°C over 30 min, facilitating the formation of heteroduplexes. A reference sample was analyzed at 1°C increments within a range 2°C above and below the range of temperatures given by the DNA melt program (<http://insertion.stanford.edu/melt.html>), to determine the optimal denaturing temperature(s) (T_d) at which each amplicon was analyzed (Table 4). Fragments were analyzed at a number of temperatures if no heteroduplexes were detected. The stationary phase consisted of 3.5-μm alkylated silica matrix, 50 × 4.6-mm i.d. (Eclipse dsDNA analysis column; Agilent Technologies, Melbourne, VIC, Australia). The mobile phase was 0.1 M triethylammonium acetate buffer (pH 7.0) containing 0.2 mM EDTA. Five microliters of PCR product was eluted with a linear acetonitrile gradient that varied depending on the size of product (Table 4). Varian Star Reviewer 2.0 software (Varian) was used to normalize chromatographic data and identify heteroduplex samples. Three samples per unique elution profile were sequenced for each PCR product and compared to the reference sequence to identify polymorphisms. PCR products were purified with ExoSAP-IT (USB, Cleveland, OH, USA) and bidirectionally sequenced, using Big Dye Terminator 3.1 chemistry, and analyzed on an ABI 310 genetic analyzer (Applied Biosystems, Foster City, CA, USA). Sequence data analyses were performed using SeqMan II 4.03 software (DNASTAR, Madison, WI, USA).

Genotyping

The 44 samples screened by dHPLC were analyzed by restriction fragment length polymorphism (RFLP) analysis for two known *TWIST1* SNPs (g.1231A→G and g.1387G→A [22] and [23], reference sequence X91662), using 1 unit of NciI (NEB, Beverly, MA, USA) and MspAII (NEB), respectively, and 1× recommended restriction buffer in 20-μl reactions. The 44 screening samples plus an additional 30 were genotyped for commonly identified (heterozygote frequency >0.05) *FGFR1* SNPs (Table 1). g.8591855G→A, g.8591936C→T, g.8592931G→C, g.8604400G→A, g.8607200G→C, and g.8646435C→T were genotyped by RFLP analysis using 1 unit of enzyme (Table 1) and 1× recommended restriction buffer in 20-μl reactions. All RFLP products were analyzed on 3% agarose gels. g.8646793A→G was genotyped by kinetic PCR using primers with a mismatch 3 bp from the 3' end and either SNP base at the 3' base (Table 4) and analyzed on 1.5% agarose gels. g.8607944G→A was genotyped by performing homoduplex mixing and subsequent dHPLC analysis. Each sample PCR product for g.8607944G→A was mixed with an equal concentration of PCR product from an identified homozygous variant sample. Mixed samples were denatured as described previously and analyzed by dHPLC at the optimal T_d. Known heterozygous and homozygous samples were run in parallel as controls. With this analysis, heterozygotes, common homozygotes, and variant homozygotes gave different peak shapes. A single fourth peak represented an additional SNP (g.8607868G→A), identified by sequencing.

Table 4. Primers used for PCR and sequencing and dHPLC conditions used to analyze generated PCR products

Fragment ID	Forward primer 5'-3'	Reverse primer 5'-3'	T _m (°C)	Cycle	Size (bp)	dHPLC gradient (°C)	DHPLC T _d (°C)	dHPLC method ^a
<i>FGFR1</i>								
Ex1-5'	CCTCAACTCTATGTGAAAAGCAC	GAGAAAAGTCCTTGGGTTCC	60	32	654	56–70	65, 66	Large
Ex1-3'	GACCGAGACCCCTCGTAG	AGGGAGCCAGGAGGTGAAAGG	66	35	668	64–71	70	Large
Ex2	TCTGGACGCCTTAGTAAGTCCAC	TCACCTGCAACCATATCACCTC	66	30	387	60–64	61 ^b –63 ^b	Medium
Ex3	CTCAGTAGCCTCCAGTAAGTGATG	CAGCAGTTTCTGAAGCAGAGTGG	63	37	538	61–65	63	Large
Int3	CCAGCCCTCTCTGCCAGTTTC	GCCCTCGCACAGCTCCCTTG	63	30	417	61–65	63	Large
Ex4–5	TGTCCGTGTTTCATCTGGAAGTCTG	TGAAAAGCATGTAATCAGGACTTC	57	29	628	56–67	61	Large
Ex6	TGGTGTATTTGGGAGTGTTC	CTATCCTGACTCTGCCCTAAG	63	31	514	59–63	61, 63 ^b	Large
Ex7	GCAAGGTCCCATGACAAGTG	ATGCACCCCATTTGCCAGAAAG	66	31	641	59–63	62 ^b , 63	Large
Ex8	AATCTCACTTAACATCTCCCTTCG	GGTGGCAAGGACAGTCCAGTAC	63	33	368	55–57, 60–63	62 ^b , 63 ^b	Large
Ex9	TGGTTTTCGTTTGTTCCTTGTTG	CAGGATGTGGCACCAGGCAG	66	31	431	59–63	62	Medium
Ex10	AGGGGACAGGAGACAGGTG	GCTACACTAGCTTGAAACAGACTC	63	31	475	59–64	63 ^b	Large
Ex11–12	TGACTAAGAATGGGAAGGAGTCAC	GGCAGCAAAGGCACCAGAGA	66	31	675	58–63	61 ^b , 62 ^b	Large
Ex13	CCCACTCCCTTAGCCTTTATCC	CTCTTAACCCCTTCCCTAGC	66	31	288	58–63	61 ^b , 62 ^b	Medium
Ex14	AGGCAGGAGATGGGAGGTTG	CCTAAGACAACACACAGGGCAC	66	31	448	60–64	63	Large
Ex15–16	CGCTTGCTGTGATGAGAAGCCTG	GCCTTTCAACATCTGGAGCAGAG	66	30	691	60–65	63	Large
Ex17–18	TCCAGATGTTGAAAGGCTGATCTG	GGCGAGAGGAAGCAGCGATG	66	31	603	61–65	64	Large
Ex19-5'	GCACGCACCTCCCTGAGCAG	GCAAAACAGACCAAACCGACAGG	66	29	614	61–67	66 ^b , 65 ^b	Large
Ex19-3'	GGGAGTGGGAGCCAATGAAC	CACCTTTTCATACAGCCCATAG	63	31	629	56–63	62 ^b , 58 ^b	Large
Ex1-5' A ^c	CCTCAACTCTATGTGAAAAGCAC	CCCTAGACATAGCAGCGATT	57	33	183	—	—	—
Ex1-5' G ^c	CCTCAACTCTATGTGAAAAGCAC	CCCTAGACATAGCAGCGATC	57	35	183	—	—	—
<i>TWIST</i>								
External	CTGCCTCTTTCGAGCACCTTC	TTTCTAAGACTAGTGAGCACTGTTCTTATC	50	11 + 16	2230	—	—	—
Promoter	TTGAATGGTTTGGGAGGACGAA	CGGTCTGGCTTCTCTCGCTGTT	57	29	588	60–68	67	Large
Ex1-5'	AAGCTGGCGGGCTGAGGCG	CAAGAAGTCTGCGGGCTGTG ^d	64	27	391	64–72	70	Large
Ex1-3'	AGCTCCTCGTAAGACTGCGGAC ^d	AATCGAGGTGGACTGGGAACCG ^d	57	26	512	59–68	66	Universal
Ex2	GACCTAAACAATAACCGACTC	TTAAAAATATAGACCAAACCTAAGG	57	33	737	57–63	58	Universal

^a Medium method elution gradient was 46–60% over 6 min. Large method was 52–66% over 6 min. Universal method was 45–72% over 9 min. All had a flow rate of 0.5 ml/min.

^b No heteroduplexes identified at this T_d.

^c Primers were used for allele-specific genotyping of g.8646793A→G.

^d Primers taken from Howard and colleagues [11].

Linkage disequilibrium analysis

LD and haplotype frequencies were analyzed for the eight genotyped *FGFR1* SNPs. Hardy–Weinberg equilibrium (HWE) was analyzed using the HWE 1.20 program (<http://linkage.rockefeller.edu/ott/linkutil.htm>). Haplotypes were reconstructed for the 74 genotyped individuals using the PHASE 2.0.1 program [41] and [42] and the no-recombination model. Haplotype phase acceptance was set at 95% confidence, two- to four-loci intervals, and five separate runs with 100 iterations each, from which the best reconstruction haplotypes were chosen. The output 148 sample haplotypes were analyzed for pair-wise linkage disequilibrium measures (D' and r^2) using the HaploXT 1.1 program with verbose option [43] (<http://archimedes.well.ox.ac.uk/pise/haploxt.html>). The resultant 2×2 contingency tables were entered into the 2BY2 program [44] to calculate Fisher's exact test two-tailed p values. To obtain overall significance of 0.05, the Bonferroni procedure was applied to correct for multiple testing: 28 pair-wise LD values gave significance of each individual testing to be 0.0018. LD blocks were analyzed using the JLIN 1.11 program [45]. htSNPs were identified from the PHASE reconstructed 148 haplotypes using SNPtagger (<http://www.well.ox.ac.uk/~xiayi/haplotype/index.html>), analyzing haplotypes as one group and for each population. Population haplotype comparisons were performed using the Arlequin 2.0 package [46]. Exact sample differentiation based on haplotype frequencies, as well as two tests to detect departure from neutral selection models, Tajima's D and Chakraborty's test for infinite alleles, was calculated. Significant values for all these tests were taken for 105 and 104 (neutral models) coalescent model simulations, respectively.

Phenotype association

Association between common htSNPs and cranial and facial phenotypes was initially performed using χ^2 tests separating population data. g.8592931G→C was chosen for more extensive analyses. Population mean CI values, grouped according to genotype, were calculated and compared graphically. Population mean CI values were not comparative; consequently, CI measurements were normalized (by blom) and ranked within populations to obtain comparative CI data. Stepwise linear regression analysis was performed for ranked CI data, while multinomial regression and ANOVA, with Monte Carlo significance set at 0.95, were performed for face shape and head shape categories. All statistical analyses were performed using SPSS 11.5 software (SPSS, Chicago, IL, USA). All p values given are two-sided.

Acknowledgments

We thank Dale Nyholt for his assistance with the association analysis statistics. We also thank all the people who volunteered to participate in this study.

References

- [1] D.H. Enlow, Facial growth, Saunders, Philadelphia (1990).
- [2] E.J. Devor, Transmission of human craniofacial dimensions, *J. Craniofac. Genet. Dev. Biol.* 7 (1987), pp. 95–106.

- [3] R.C. Hauspie, C. Susanne and E. Defrise-Gussenhoven, Testing for the presence of genetic variance in factors of face measurements of Belgian twins, *Ann. Hum. Biol.* 12 (1985), pp. 429–440.
- [4] K. Sharma, Sex differences in genetic determinants of craniofacial variations—a study based on twin kinships, *Acta Genet. Med. Gemellol. (Rome)* 47 (1998), pp. 31–41.
- [5] H.J. Kim, D.P. Rice, P.J. Kettunen and I. Thesleff, FGF-, BMP- and Shh-mediated signalling pathways in the regulation of cranial suture morphogenesis and calvarial bone development, *Development* 125 (1998), pp. 1241–1251.
- [6] J.L. Mountain, A.A. Lin, A.M. Bowcock and L.L. Cavalli-Sforza, Evolution of modern humans: evidence from nuclear DNA polymorphisms, *Philos. Trans. R. Soc. London B Biol. Sci.* 337 (1992), pp. 159–165.
- [7] J.H. Relethford, Craniometric variation among modern human populations, *Am. J. Phys. Anthropol.* 95 (1994), pp. 53–62.
- [8] M.M. Cohen Jr., Sutural biology and the correlates of craniosynostosis, *Am. J. Med. Genet.* 47 (1993), pp. 581–616.
- [9] M. Muenke and A.O. Wilkie, Craniosynostosis syndromes. In: C. Scriver, A. Beaudet, W. Sly, D. Valle, B. Childs and B. Vogelstein, Editors, *The metabolic and molecular bases of inherited disease*, McGraw–Hill, New York (2000), pp. 6117–6146.
- [10] A.O. Wilkie, Craniosynostosis: genes and mechanisms, *Hum. Mol. Genet.* 6 (1997), pp. 1647–1656.
- [11] T.D. Howard et al., Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre–Chotzen syndrome, *Nat. Genet.* 15 (1997), pp. 36–41.
- [12] M.R. Passos-Bueno et al., Clinical spectrum of fibroblast growth factor receptor mutations, *Hum. Mutat.* 14 (1999), pp. 115–125.
- [13] T. Roscioli et al., Clinical findings in a patient with *FGFR1* P252R mutation and comparison with the literature, *Am. J. Med. Genet.* 93 (2000), pp. 22–28.
- [14] P. Rutl et al., Identical mutations in the *FGFR2* gene cause both Pfeiffer and Crouzon syndrome phenotypes, *Nat. Genet.* 9 (1995), pp. 173–176.
- [15] F. Carinci et al., Expression profiles of craniosynostosis-derived fibroblasts, *Mol. Med.* 8 (2002), pp. 638–644.
- [16] J. Albuissou et al., Kallmann syndrome: 14 novel mutations in *KAL1* and *FGFR1* (*KAL2*), *Hum. Mutat.* 25 (2005), pp. 98–99.
- [17] C. Groth and M. Lardelli, The structure and function of vertebrate fibroblast growth factor receptor 1, *Int. J. Dev. Biol.* 46 (2002), pp. 393–400.

- [18] P.H. Francis-West, L. Robson and D.J. Evans, Craniofacial development: the tissue and molecular interactions that control development of the head, *Adv. Anat. Embryol. Cell Biol.* 169 (2003) (III-VI), pp. 1–138.
- [19] K.E. Haworth et al., Canine TCOF1; cloning, chromosome assignment and genetic analysis in dogs with different head types, *Mamm. Genome* 12 (2001), pp. 622–629.
- [20] V. El Ghouzzi et al., Mutations of the TWIST gene in the Saethre–Chotzen syndrome, *Nat. Genet.* 15 (1997), pp. 42–46.
- [21] D. Johnson et al., A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1, *Am. J. Hum. Genet.* 63 (1998), pp. 1282–1293.
- [22] C.F. Yang, J.Y. Wu, F.J. Tsai, C.C. Lee and W.D. Lin, Identification of a polymorphism (G83S) in the TWIST gene in Taiwanese, *Hum. Mutat.* 16 (2000), p. 448.
- [23] V. Kasparcova et al., Molecular analysis of patients with Saethre–Chotzen syndrome: novel mutations and polymorphisms in the TWIST gene, *Am. J. Hum. Genet.* 63 (1998), p. abs 2126 Suppl..
- [24] M. Muenke et al., A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome, *Nat. Genet.* 8 (1994), pp. 269–274.
- [25] N. Sato et al., Clinical assessment and mutation analysis of Kallmann syndrome 1 (KAL1) and fibroblast growth factor receptor 1 (*FGFR1*, or KAL2) in five families and 18 sporadic patients, *J. Clin. Endocrinol. Metab.* 89 (2004), pp. 1079–1088.
- [26] C. Dode et al., Loss-of-function mutations in *FGFR1* cause autosomal dominant Kallmann syndrome, *Nat. Genet.* 33 (2003), pp. 463–465.
- [27] S.G. Patel and J.X. DiMario, Two distal Sp1-binding cis-elements regulate fibroblast growth factor receptor 1 (*FGFR1*) gene expression in myoblasts, *Gene* 270 (2001), pp. 171–180.
- [28] R. Parakati and J.X. DiMario, Sp1- and Sp3-mediated transcriptional regulation of the fibroblast growth factor receptor 1 gene in chicken skeletal muscle cells, *J. Biol. Chem.* 277 (2002), pp. 9278–9285.
- [29] L.B. Jorde, W.S. Watkins, J. Kere, D. Nyman and A.W. Eriksson, Gene mapping in isolated populations: new roles for old friends?, *Hum. Hered.* 50 (2000), pp. 57–65.
- [30] S.B. Gabriel et al., The structure of haplotype blocks in the human genome, *Science* 296 (2002), pp. 2225–2229.
- [31] B. Latter, Genetic differences within and between populations of the major human subgroups, *Am. Nat.* 116 (1980), pp. 220–237.
- [32] R.C. Lewontin, The apportionment of human diversity, *Evol. Biol.* 6 (1972), pp. 381–398.

- [33] L.B. Jorde et al., The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data, *Am. J. Hum. Genet.* 66 (2000), pp. 979–988.
- [34] C. Romualdi et al., Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms, *Genome Res.* 12 (2002), pp. 602–612.
- [35] R.J. Hennessy and C.B. Stringer, Geometric morphometric study of the regional variation of modern human craniofacial form, *Am. J. Phys. Anthropol.* 117 (2002), pp. 37–48.
- [36] J.H. Relethford, Apportionment of global human genetic diversity based on craniometrics and skin color, *Am. J. Phys. Anthropol.* 118 (2002), pp. 393–398.
- [37] W. Makalowski, Genomics: not junk after all, *Science* 300 (2003), pp. 1246–1247.
- [38] I. Wickelgren, Molecular biology: spinning junk into gold, *Science* 300 (2003), pp. 1646–1649.
- [39] W.M. Bass, *Human osteology: a laboratory and field manual*, Missouri Archaeological Society, Columbia, MO (1995).
- [40] A. Malinowski and W. Bozilow, *Podstawy antropometrii: metody, techniki, normy*, Wydawnictwo Naukowe PWN, Warsaw (1997).
- [41] M. Stephens, N.J. Smith and P. Donnelly, A new statistical method for haplotype reconstruction from population data, *Am. J. Hum. Genet.* 68 (2001), pp. 978–989.
- [42] M. Stephens and P. Donnelly, A comparison of Bayesian methods for haplotype reconstruction from population genotype data, *Am. J. Hum. Genet.* 73 (2003), pp. 1162–1169.
- [43] G.R. Abecasis and W.O. Cookson, GOLD—graphical overview of linkage disequilibrium, *Bioinformatics* 16 (2000), pp. 182–183.
- [44] B. Weir, (1996). *Genetic Data Analysis II*, Sinauer, <http://linkage.rockefeller.edu/ott/linkutil.htm>, Sunderland, MA.
- [45] K. Carter, P. McCaskie, L. Palmer (2004). *Java Linkage Disequilibrium Plotter*, WAIMR, <http://www.genepi.com.au/projects/jlin/>, Perth, Western Australia.
- [46] S. Schneider, D. Roessli, L. Excoffier (2000). *Arlequin version 2.000: software for population genetics data analysis*, Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, <http://lgb.unige.ch/arlequin/>, Geneva.