Polymorphism in the epidermal growth factor gene is associated with birthweight in Sinhalese and white Western Europeans


1Human Genetics Unit; 2Department of Obstetrics and Gynaecology, Faculty of Medicine, University of Colombo, Kynsey Road, Colombo 00800, Sri Lanka; 3Division of Clinical Chemistry, Institute of Genetics, School of Molecular Medical Sciences and 4Division of Obstetrics, School of Human Development, University of Nottingham, Nottingham NG7 2UH, UK.

Birthweight predicts health later in life and is influenced by inherited factors. We investigated the association of the c.61G > A, and c.2566G > A polymorphisms in the epidermal growth factor (EGF) gene [GenBank NM_001963] with birthweight in three groups of healthy pregnant women, and in women with pregnancies affected by fetal growth restriction (FGR). Subjects comprised 171 Sinhalese women with normal pregnancies (Group A), 64 white Western European women with normal pregnancies (Group B), 101 white Western European women with normal pregnancies and their babies (Group C) and 107 women with pregnancies affected by FGR, their partners and their babies (Group D). Maternal EGF genotypes were associated with birthweight of healthy babies of women in Groups A (P = 0.03), B (P = 0.001) and C (P = 0.01). The association persisted following adjustment for confounding by gestational age, sex, maternal weight, parity and smoking habit. The trend from heaviest to lightest birthweights in all these groups was c.61AA > c.61GA > c.61GG and c.2566GG > c.2566GA > c.2566AA. The EGF haplotype associated with lower birthweight (c.61G, c.2566A) was transmitted at increased frequency from heterozygous parents to babies affected by FGR in Group D (P = 0.02). These findings support the hypothesis that growth factors expressed by the fetomaternal unit affect birthweight, and implicates polymorphism in the EGF gene in the aetiology of birthweight variability.

Keywords: birthweight/EGF/sinhalese/white Western European/genetic study

Introduction

Intrauterine growth is an important predictor of perinatal and adult health. Low birthweight is related to an increased risk of diseases in adult life, including coronary heart disease, hypertension and diabetes (McCarton et al., 1996). Fetal growth restriction (FGR) is associated with high perinatal mortality (Kramer et al., 1990) and intellectual impairment later in life (Lau and Rogers, 2004). Inherited factors influence birthweight (Klebanoff et al., 1989; Claussen et al., 2000; Ghezzi et al., 2003), but we know very little about the genes involved. Heritability of birthweight was estimated as 25–40% in one study of offspring of monzygotic and dizygotic twin sisters (Claussen et al., 2000). In another twin study, 30–40% of the variability in birthweight was attributed to maternal gene effects (Nance et al., 1983). In a genomewide screen for quantitative trait loci affecting birthweight, heritability was estimated to be 72% (Arya et al., 2006). There are no reports of studies specifically designed to address the question of heritability of FGR, but Claussen et al. (2000) analysed smallness for gestational age (SGA) as a dichotomous variable within their twin study and reported a heritability of 34%, similar to that for birthweight considered as a continuous variable.
exposed to hypoxia or cytokines in vitro (Levy et al., 2000; Johnstone et al., 2005).

EGF at the feto-maternal interface is reported to be of maternal origin (Haining et al., 1991; Bass et al., 1994), and in this location EGF could affect trophoblast proliferation, differentiation and invasion. EGF receptors (EGFR) are expressed by cytotrophoblast and are upregulated by EGF in these cells (DePalo and Das, 1988). EGF likely to be of fetal origin is present in amniotic fluid (Watanabe, 1990), and EGF mRNA transcripts and EGF protein have been identified in amnion and umbilical cord (Rao et al., 1995). Functional EGFR have also been demonstrated in umbilical cord tissue, including vascular endothelium and smooth muscle cells (Rao et al., 1995). An autocrine or paracrine role for EGF in modulating umbilical vascular tone has been proposed, which would in turn regulate the delivery of nutrients to the growing fetus.

Failure to achieve intrauterine growth potential is termed FGR. Low birthweight is used as a surrogate marker for FGR, which has been variously defined as birthweight below the 10th, 5th or 3rd centile of a matched population. Computer software is available for the estimation of a corrected birthweight centile (CBC) which adjusts for gestational age at delivery, fetal sex, maternal parity, weight and height; all of these factors are significantly associated with birthweight (Gardosi et al., 1992). In addition, an association between maternal smoking and low birthweight has been repeatedly demonstrated (Lieberman et al., 1994).

We hypothesized that genetic polymorphism in the maternal EGF gene is associated with variation in birthweight.

Materials and Methods

Subjects

We tested our hypothesis initially in a group of normal pregnant Sinhalese women in Sri Lanka (Group A) and carried out replicate studies in two groups of healthy pregnant white Western European women (Groups B and C) from Nottingham, UK. The study was further extended to white Western European maternal–patriarchal–fetal triads from pregnancies affected by FGR (Group D).

Group A comprised 171 nulliparous Sinhalese women who had normal pregnancies recruited from maternity units in Colombo, Sri Lanka, between August 2001 and January 2003. Group B comprised 64 nulliparous white Western European women who had normal pregnancies recruited between 1993 and 1998 as controls in a study of pre-eclampsia (Morgan et al., 1995); Group C comprised 101 nulliparous and multiparous white Western European women who had normal pregnancies recruited between 2000 and 2003 as controls in a study of FGR and Group D included 107 women who had pregnancies complicated by FGR, their partners and babies recruited as cases for this study of FGR (Tower et al., 2006). All had singleton pregnancies.

FGR pregnancies were identified antenatally by ultrasound scan. Patients were approached if the fetal abdominal circumference, on locally used growth percentile charts, was <5th centile. Following delivery, the centile of the CBC was calculated using the GROW software, available at http://www.perinatal.nhs.uk. This software calculates the birthweight centile adjusted for maternal ethnicity, weight, height and parity, gestation at delivery and fetal sex. FGR pregnancy was defined as CBC ≤5; normally grown pregnancies were defined as CBC >5. Women with a history of chronic hypertension, diabetes, renal, cardiovascular, endocrine or autoimmune disease and pregnancies affected by chromosomal abnormalities were excluded from the study.

All subjects were recruited with the approval of the Ethics Committees of the host institutions, and volunteers provided written informed consent to participation.

Nomenclature


In order to test our hypothesis, we selected two genetic markers with potential functional effects in the EGF gene: the c.61G > A polymorphism in the 5’ untranslated region of the gene (Shahbazi et al., 2002), which has been reported to modify gene expression and the non-synonymous c.2566G > A polymorphism in exon 14 of the gene, which results in a substitution of the amino acid methionine by isoleucine in the LDL receptor homology domain of EGF (Semina et al., 1996).

Genotyping

DNA samples extracted from venous blood were available for all women and for partners of women in Group D. DNA samples extracted from umbilical cords were available for babies of women in Groups C and D.

The EGF c.61G > A polymorphism was genotyped using a PCR/RFLP assay described previously, using the restriction enzyme Alul (Shahbazi et al., 2002). The EGF c.2566G > A polymorphism was genotyped by mutation specific (MS)-PCR using two allele specific forward primers 5’-TTA TGT GTG GTT CTA TTC CGC TAT GCC ATC AGT AAG G-3’ and 5’-GGG CTA TCG CAT CAG GAA TA-3’. and a common reverse primer 5’-CAG ATT CCA GCC AAAG GAA AG-3’. The MS-PCR was conducted in a final volume of 15 μl containing genomic DNA, 0.08 μM EGF-G, 0.08 μM EGF-A, 0.1 μM EGF-F, 20 μM dNTP, 10 mM Tris-HCL, 50 mM KCl, 1.5 mM MgCl2, pH 8.3, 20°C and 1 U of Taq DNA Polymerase (Roche Diagnostics, Germany). The reactions were subjected to 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 1 min and extension at 72°C for 5 min. The 182-bp and 162-bp PCR products generated from the c.2566G allele and the c.2566A allele, respectively, were resolved by electrophoresis 3.5% agarose.

As quality assurance measures, sequenced positive control samples and DNA-free blanks were included in all batches, and all genotypes were confirmed by two independent observers.

Statistical analysis

Hardy–Weinberg equilibrium (HWE) was evaluated by χ² testing. Linear regression analysis was performed to study the association of EGF polymorphisms with birthweight in healthy pregnancies, with inclusion of potential confounding variables, including maternal weight, gestation at delivery, fetal sex, parity and smoking habit. The genotype was treated as a quantitative variable coded 0, 1 or 2 to represent the number of variant alleles, consistent with an additive model. Haplotype frequencies and hypothesis testing were performed using programs from the UNPHASED software suite (Dudbridge, 2003), which uses the expectation maximization algorithm to determine common haplotypes, and conducts a likelihood ratio test in a log-linear model. COCAPHASE was used to estimate haplotype frequencies from genotype data and to perform hypothesis testing using haplotypes in case–control association studies; TDTPHASE was used for hypothesis testing in case-parent trios.

Results

The characteristics of pregnancies of all four groups are shown in Table 1. In Group D, 52 women with FGR (48%) had abnormal umbilical artery Dopplers [either raised resistance (>95th centile) or absent/reversed end diastolic flow]. Of these, 15 women (14%) had absent or reversed end diastolic flow. In the FGR group, 49 women (46%) had an amniotic fluid index of less than the 5th centile.

EGF c.61G > A and c.2566G > A genotypes were in HWE in both white Western European and Sinhalese populations, although there were significant racial differences in allele frequencies (Table 2). The two polymorphisms were in linkage disequilibrium in both populations (D’ > 0.90, r² > 0.65); the two haplotypes c.61A, c.2566G and c.61G, c.2566A represented over 84% of haplotypes estimated by the maximum likelihood algorithm.

The results of linear regression analysis are shown in Table 2. There was a significant association between maternal EGF genotype and birthweight in healthy pregnancies in Sinhalese (Group A) and...
EGF gene polymorphism is associated with birthweight

EGF c.61G, c.2566A haplotype was observed in both father–offspring and mother–offspring pairs. Analysis of subgroups demonstrated that this trend was independent of maternal smoking habit during pregnancy, although the smaller number of informative transmissions resulted in a loss of statistical significance on subgroup analysis.

**Discussion**

The results reported here suggest that maternal EGF genotypes contribute to the variation in birthweight in healthy pregnancies. Demonstration of an association between maternal EGF genotypes and birthweight in three independent sample collections, representing two distinct ethnic groups, and careful correction for possible confounding factors, provides strong confirmation of the validity of these findings. The proportion of variation in birthweight which could be attributed to EGF genotype in this study was between 1% (Group C) and 10% (Group B). Compared with babies born to mothers who were homozygous for EGF c.61G or c.2566A, those born to mothers homozygous for EGF c.61A or c.2566G were on average between 250 and 390 g heavier in white Western European pregnancies, and 170–200 g heavier in Sinhalese pregnancies.

The observation that the EGF haplotype associated with the lowest birthweight, c.61G, c.2566A, is also transmitted preferentially to babies affected by FGR lends further support to the concept of an EGF susceptibility haplotype, in this case, acting in the fetus. It is important to recognize that the term 'FGR' does not merely refer to the lower extreme of the birthweight distribution, but implies a pathological failure to achieve birthweight potential due to a failure of trophoblast invasion (Khong et al., 1986). The molecular mechanisms underlying birthweight variability in otherwise healthy pregnancies, and those which result in the pathological features of FGR are therefore likely to differ. SGA is used as a convenient surrogate for FGR, but must inevitably include some babies who are constitutionally small. In this study, sequential antenatal measures of impaired fetal growth, the observation of reduced liquor volume (oligohydramnios) on ultrasound (Lin et al., 1990) and/or the presence of abnormal umbilical artery Doppler waveforms were utilized to improve the

---

**Table 1: Characteristics of study participants**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Group A (n = 171)</th>
<th>Group B (n = 64)</th>
<th>Group C (n = 101)</th>
<th>Group D (n = 107)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>27.1 (5.1)</td>
<td>27.5 (3.7)</td>
<td>29.3 (5.7)</td>
<td>26.5 (5.8)</td>
</tr>
<tr>
<td>Nulliparous</td>
<td>171 (100%)</td>
<td>64 (100%)</td>
<td>58 (57%)</td>
<td>65 (61%)</td>
</tr>
<tr>
<td>Maternal weight at booking (kg)</td>
<td>48.8 (8.0)</td>
<td>65.1 (9.9)</td>
<td>67.5 (13.1)</td>
<td>61.2 (10.6)</td>
</tr>
<tr>
<td>Gestation at delivery (weeks)</td>
<td>39.4 (1.1)</td>
<td>39.9 (1.4)</td>
<td>38.8 (1.5)</td>
<td>35.1 (4.4)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>0</td>
<td>7 (11%)</td>
<td>14 (14%)</td>
<td>52 (49%)</td>
</tr>
<tr>
<td>Male infant sex</td>
<td>105 (61%)</td>
<td>60 (25%)</td>
<td>52 (52%)</td>
<td>44 (41%)</td>
</tr>
</tbody>
</table>

Data are presented as mean (SD) or number (%), with the exception of BWC, which are shown as median (inter-quartile range)

---

**Table 2: Infants’ birthweights in kilograms tabulated according to maternal or fetal EGF genotype**

<table>
<thead>
<tr>
<th>EGF c.61G &gt; A</th>
<th>G allele frequency</th>
<th>Genotypes</th>
<th>Adjusted P</th>
<th>EGF c.2566G &gt; A</th>
<th>A allele frequency</th>
<th>Genotypes</th>
<th>Adjusted P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA</td>
<td>AG</td>
<td>GG</td>
<td></td>
<td>GG</td>
<td>GA</td>
<td>AA</td>
</tr>
<tr>
<td>N</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean birthweight (SD in kg)</td>
<td>37 (0.44)</td>
<td>78 (0.42)</td>
<td>56 (0.38)</td>
<td>0.209</td>
<td>41 (0.41)</td>
<td>73 (0.40)</td>
<td>57 (0.41)</td>
</tr>
<tr>
<td>Group B mothers</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>25</td>
<td>9</td>
<td>0.001</td>
<td>32 (0.49)</td>
<td>24 (0.45)</td>
<td>8 (0.17)</td>
</tr>
<tr>
<td>Mean birthweight (SD in kg)</td>
<td>3.63 (0.50)</td>
<td>3.50 (0.45)</td>
<td>3.27 (0.19)</td>
<td></td>
<td>3.66 (0.49)</td>
<td>3.44 (0.45)</td>
<td>3.27 (0.17)</td>
</tr>
<tr>
<td>Group C mothers</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>35</td>
<td>51</td>
<td>15</td>
<td>0.013</td>
<td>43 (0.41)</td>
<td>46 (0.40)</td>
<td>12 (0.59)</td>
</tr>
<tr>
<td>Mean birthweight (SD in kg)</td>
<td>3.47 (0.42)</td>
<td>3.38 (0.41)</td>
<td>3.22 (0.54)</td>
<td></td>
<td>3.48 (0.41)</td>
<td>3.35 (0.40)</td>
<td>3.18 (0.59)</td>
</tr>
<tr>
<td>Group C babies</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>32</td>
<td>53</td>
<td>15</td>
<td>0.427</td>
<td>39 (0.36)</td>
<td>49 (0.46)</td>
<td>12 (0.57)</td>
</tr>
<tr>
<td>Mean birthweight (SD in kg)</td>
<td>3.39 (0.39)</td>
<td>3.42 (0.44)</td>
<td>3.22 (0.51)</td>
<td></td>
<td>3.43 (0.36)</td>
<td>3.37 (0.46)</td>
<td>3.24 (0.57)</td>
</tr>
</tbody>
</table>

Data were analysed by linear regression of birth weight on maternal EGF genotype, adjusting for the following confounders: amaternal weight at booking; bperiod of gestation; csex of baby; dsmoking status; eparity. All Sinhalese women were non-smokers. The GenBank reference sequence for EGF is NM_001963.
The measure of linkage disequilibrium, \( r^c \), between c.2566G and this region with minor allele frequencies acting as markers for functional polymorphisms elsewhere in the EGF gene which is critical in pathological pregnancies. It should be noted, however, that the lower statistical power in analysis of dichotomous (FGR versus healthy) compared with continuous (birthweight) data may have resulted in failure to detect an effect of maternal genes in FGR. The results of the TDT analysis of fetal EGF in FGR are of marginal significance, and will need confirmation in a larger study. It is nevertheless intriguing that they implicate the identical susceptibility haplotype to the maternal EGF haplotype associated with lower birthweights in healthy pregnancies.

Identification of the genetic mechanisms underlying the association of EGF genotypes with intrauterine growth will require extensive investigation. It is clear from the 2-SNP haplotyping undertaken in this study that linkage disequilibrium extends over a region exceeding 67 kb at this locus and the polymorphisms studied may therefore be acting as markers for functional polymorphisms elsewhere in the gene. The strong linkage disequilibrium across the EGF gene is confirmed by the latest HapMap release, which tested 44 common SNPs in this region with minor allele frequencies >0.1, including the c.2566G > A polymorphism in exon 14 (http://www.hapmap.org/). The measure of linkage disequilibrium, \( r^2 \), between c.2566G > A and over 40 of the other common SNPs in EGF exceeds 0.50, suggesting that both c.61G > A and c.2566G > A may be acting as proxies for other functional SNPs.

**Table 3:** Haplotype frequency estimates of the EGF gene in white Western European women who delivered growth restricted infants (FGR) (Group D) and who delivered normal infants (Controls) (Group C)

<table>
<thead>
<tr>
<th>Maternal EGF haplotype</th>
<th>Group D mothers [FGR] ((n = 107))</th>
<th>Group C mothers [Controls] ((n = 101))</th>
</tr>
</thead>
<tbody>
<tr>
<td>G G</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>G A</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td>A G</td>
<td>0.62</td>
<td>0.60</td>
</tr>
<tr>
<td>A A</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Likelihood ratio statistic 0.69, Degrees of freedom 2, \( P = 0.71 \).

Haplotypes with frequencies <0.03 (representing <6 chromosomes) have been dropped from hypothesis testing. The GenBank reference sequence for EGF is NM_001963.

These two SNPs were nevertheless selected for their possible functional effects. The non-synonymous c.2566G > A polymorphism does not lie within the region encoding mature EGF, but causes an iso- leucine for methionine substitution in the EGF precursor, a 160 kDa protein which is known to be biologically active (Breyer and Cohen, 1990). The c.61G > A SNP in the 5′ untranslated region of EGF does not coincide with known transcription factor binding sites, but is adjacent to a putative nuclear factor-κB binding site (Shahbazi et al., 2002). EGF production by cultured peripheral blood mononuclear cells from 34 healthy individuals was reported to be significantly higher in cells homozygous for c.61G (Shahbazi et al., 2002). A subsequent study of 42 patients with glioblastoma multiforme reported higher EGF expression in tumour tissue associated with possession of one or two copies of the c.61G allele (Bhowmick et al., 2004). The association of the c.61G allele, which was associated with higher EGF expression in monocytes and glioblastoma cells, with low birthweight is consistent with observations in transgenic mice over-expressing EGF, which had birthweights only half that of their normal littermates (Chan and Wong, 2000).

There have been relatively few reports of attempts to identify genes affecting birthweight. In a recent genomewide screen, Arya et al. (2006) identified eight quantitative trait loci linked to birthweight with LOD scores >1.2. The strongest linkage was on chromosome 6, but interestingly a suggestive locus was also detected on chromosome 4q. The maximum LOD score at this locus was at marker D4S1625, about 30 megabase pairs from the EGF gene.

Of the possible candidate genes affecting birthweight, insulin-like growth factor IGF-1 (IGF) has attracted particular attention since publication of a case report of severe intrauterine growth restriction in a patient with a homozygous partial deletion of the IGF-1 gene (Woods et al., 1996). Fetal homozygosity for functional IGF-1 mutations appears to cause both FGR and post-natal growth failure, whereas heterozygosity may be associated with more subtle manifestations of intrauterine growth restriction (Walenkamp et al., 2005). Interaction between EGF and IGF and IGF-binding protein-3 is well documented (Hembree et al., 1994; Edmondson et al., 1999), and illustrates the complex networks within which maternal and fetal genes must act to determine birthweight.

**Acknowledgements**

We wish to thank the research participants for having consented to take part in these investigations and the staff of the recruiting hospitals in Sri Lanka and the UK for their co-operation. We also wish to thank H.D. Samarasinghe for assistance with subject recruitment in Sri Lanka. V.H.W.D. held a PhD studentship funded jointly by the President’s Fund Sri Lanka, University of Colombo, Sri Lanka and the University of Nottingham, UK. C.T. held a clinical research training fellowship funded by the Wellcome Trust.

**References**


These two SNPs were nevertheless selected for their possible functional effects. The non-synonymous c.2566G > A polymorphism does not lie within the region encoding mature EGF, but causes an isoleucine for methionine substitution in the EGF precursor, a 160 kDa protein which is known to be biologically active (Breyer and Cohen, 1990). The c.61G > A SNP in the 5′ untranslated region of EGF does not coincide with known transcription factor binding sites, but is adjacent to a putative nuclear factor-κB binding site (Shahbazi et al., 2002). EGF production by cultured peripheral blood mononuclear cells from 34 healthy individuals was reported to be significantly higher in cells homozygous for c.61G (Shahbazi et al., 2002). A subsequent study of 42 patients with glioblastoma multiforme reported higher EGF expression in tumour tissue associated with possession of one or two copies of the c.61G allele (Bhowmick et al., 2004). The association of the c.61G allele, which was associated with higher EGF expression in monocytes and glioblastoma cells, with low birthweight is consistent with observations in transgenic mice over-expressing EGF, which had birthweights only half that of their normal littermates (Chan and Wong, 2000).

There have been relatively few reports of attempts to identify genes affecting birthweight. In a recent genomewide screen, Arya et al. (2006) identified eight quantitative trait loci linked to birthweight with LOD scores >1.2. The strongest linkage was on chromosome 6, but interestingly a suggestive locus was also detected on chromosome 4q. The maximum LOD score at this locus was at marker D4S1625, about 30 megabase pairs from the EGF gene.

Of the possible candidate genes affecting birthweight, insulin-like growth factor IGF-1 (IGF) has attracted particular attention since publication of a case report of severe intrauterine growth restriction in a patient with a homozygous partial deletion of the IGF-1 gene (Woods et al., 1996). Fetal homozygosity for functional IGF-1 mutations appears to cause both FGR and post-natal growth failure, whereas heterozygosity may be associated with more subtle manifestations of intrauterine growth restriction (Walenkamp et al., 2005). Interaction between EGF and IGF and IGF-binding protein-3 is well documented (Hembree et al., 1994; Edmondson et al., 1999), and illustrates the complex networks within which maternal and fetal genes must act to determine birthweight.

**Acknowledgements**

We wish to thank the research participants for having consented to take part in these investigations and the staff of the recruiting hospitals in Sri Lanka and the UK for their co-operation. We also wish to thank H.D. Samarasinghe for assistance with subject recruitment in Sri Lanka. V.H.W.D. held a PhD studentship funded jointly by the President’s Fund Sri Lanka, University of Colombo, Sri Lanka and the University of Nottingham, UK. C.T. held a clinical research training fellowship funded by the Wellcome Trust.

**References**


Submitted on October 8, 2006; resubmitted on January 26, 2007; accepted on January 28, 2007