BMJ Open

Study protocol to investigate the environmental and genetic aetiology of atopic dermatitis: the Indonesian Prospective Study of Atopic Dermatitis in Infants (ISADI)

Conny Tanjung,1,2 Peter Rzehak,3 Muchtaruddin Mansyur,4 Zakiudin Munasir,5 Herawati Sudoyo,6,7 Suzanna Immanuel,8 Roedi Irawan,9 Eva Reischl,10 Hans Demmelmaier,3 Berthold Koletzko,3 Sri Rezeki Hadinegoro,5 Damayanti Rusli Sjarif5

ABSTRACT

Introduction: Atopic dermatitis (AD) is the most common skin disorder in young children worldwide, with a high impact on morbidity and quality of life. To date, no prospective study has been published on the incidence and potential predictors of AD in South East Asian populations. The Indonesian Prospective Study of Atopic Dermatitis in Infants (ISADI) will address the genetic, metabolic and dietary characteristics of mothers and their offspring, as well as potential determinants of AD within the first year of infant life.

Methods and analysis: This prospective study will be undertaken in about 400 infants to investigate the direct and indirect effects of filaggrin (FLG) gene mutations, the genetic variants of FADS1, FADS2 and FADS3 and the role of long-chain polyunsaturated fatty acids (LCPUFA) on the development of AD. We will use standardised protocols for subject recruitment, umbilical artery plasma analysis, buccal cell sampling for genotyping, fatty acid analysis, physical exams, 3-day food-intake recall of mothers and children, as well as comprehensive questionnaires on environmental, socioeconomic and AD-related factors, including family history. Monthly monitoring by telephone and physical exams every 3 months will be carried out to assess participants’ anthropometry, medical history and incidence of AD diagnosis during the first year of life. Hypotheses-driven analyses of quality-controlled dietary, genetic and metabolic data will be performed with state-of-the-art statistical methods (eg, AD-event history, haplotype, dietary or metabolic factor analysis). Direct and indirect effects of genetics and LCPUFA in buccal cell and cord plasma glycerophospholipids as potential mediators of inflammation on AD development will be evaluated by path analysis.

Strengthenes and limitations of this study

- This study is the first in Asia to evaluate the role of FADS genes on long-chain polyunsaturated fatty acids (LCPUFA) compositions in buccal cells and plasma.
- This study is the first in Asia to evaluate the roles of FADS genes and LCPUFA concentrations on the progression of AD.
- We hypothesise that in utero exposure to LCPUFA provides greater benefits to infants compared to exposure in infancy or childhood. As such, we will sample participants’ umbilical artery plasma in order to assess actual fetal conditions, rather than umbilical vein plasma, as performed in other studies.
- Diagnosis of AD will be based on Hanifin & Rajka criteria and confirmed by a dermatologist.
- We will assess for filaggrin mutations by single nucleotide polymorphism (SNP) analysis of five reported pathogenic SNPs. However, full gene sequencing would be more accurate, as the filaggrin (FLG) gene varies according to population.

Ethics and dissemination: The Permanent Medical Research Ethics Committee in Medicine and Health/Faculty of Medicine Universitas Indonesia/Dr Cipto Mangunkusumo Hospital (No. 47/H2.F1/ETIK/2014) approved the study protocol (extended by the letter no. 148/UN2.F1/ETIK/2015). We aim to disseminate our findings via publication in an international journal with high impact factor.

BACKGROUND

The incidence and prevalence of atopic dermatitis (AD) have increased markedly in recent years, possibly due to interactions between genetic and environmental factors. At least 20 genes are reportedly related to AD, but one genetic locus has been consistently linked with AD occurrence, the filaggrin (FLG) gene.1–2 Mutation of the gene causes loss of function and disturbances in
epidermal cytoskeleton aggregation, thereby increasing skin permeability to water and outside particles, such as allergens.\(^1\)\(^3\)

Much effort has been made to prevent AD, such as promotion of exclusively breastfeeding for at least 3 months, reduction of allergen exposure such as to dust mites and tobacco, using partially hydrolysed formula for infants unable to breastfeed, probiotic supplementation, restoration of the skin barrier and supplementation with long-chain polyunsaturated fatty acids (LCPUFA).\(^1\)\(^4\)–\(^7\)

From the omega-3 LCPUFA eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) and the omega-6 LCPUFA arachidonic acid (ARA) prostaglandins (PG), leukotrienes (LT) and thromboxanes (TX) are synthesized.\(^8\)–\(^15\) To date, clinical trials and meta-analyses have shown inconsistent results. Studies in the past 5–10 years showed that polymorphisms in the gene encoding fatty acid desaturases (FADS genes) influence the contribution of polyunsaturated fatty acids (PUFA) and LCPUFA, which are derived by desaturation and chain elongation from PUFA, to total lipids.\(^11\)–\(^13\)\(^16\)–\(^19\)

To date, there have been no studies in Indonesia on FLG gene mutations, the composition of LCPUFAs in infants, FADS1 and FADS2 gene polymorphisms, nor on possible associations of these gene polymorphisms. These data are needed, considering the large size of the Indonesian population (fourth largest in the world with a population of 237,641,326 people), consisting of 1,128 tribes.\(^20\) The results of this study are expected to provide information about the interaction of genetic variation, nutrition and the progression of AD in infants.

**AIMS AND OBJECTIVES**

The general objective of this study is to characterise the impact of genetic variation in the FLG gene and the FADS1, FADS2 and FADS3 gene cluster on LCPUFA in plasma and buccal cell lipids, as well as the occurrence and severity of AD in Indonesian infants. Specific objectives include the characterisation of the frequency of FLG and FADS1, FADS2 and FADS3 gene single nucleotide polymorphisms (SNPs) assessed in umbilical artery leucocytes, the fatty acid composition of umbilical artery plasma and buccal cell lipids and the impact of FLG gene and FADS1, FADS2 and FADS3 gene SNPs, LCPUFA status, maternal diet and breastfeeding on AD in the first year of life in Indonesian infants.

**ATOPIC DERMATITIS**

AD (also commonly known as atopic eczema) is a chronic skin disorder characterised by inflammation and itching. It is one of the most common disorders found in children with a high impact on morbidity and quality of life. It often precedes the occurrences of allergic rhinitis and asthma, which has been referred to as the ‘atopic march’.\(^21\)–\(^23\) The worldwide prevalence of AD has increased during the past three decades, and currently 10–20% of children are affected.\(^1\)\(^24\)–\(^26\) AD often begins in very early childhood, with as many as 45% of all cases reported to be manifested in the first 6 months of life.\(^1\)\(^21\)–\(^23\)

Several approaches have been explored for AD prevention, such as reducing exposure to common environmental allergens such as house dust mites, tobacco and cows’ milk protein, probiotic supplementation, restoration of skin barrier and LCPUFA supplementation.\(^26\)–\(^28\) However, no prospective study on the incidence and potential predictors of AD in Southeast Asian populations has been published.\(^25\)

**FILAGRIN GENE**

SNPs of the filaggrin gene (FLG, access number in GenBank NM_0002016) have been described as a major etiologic factor of keratinisation disorders, including ichthyosis vulgaris and AD, in Western populations.\(^26\) The filaggrin gene is located on the 1q21 chromosome in an area called the epidermal differentiation complex (EDC). The EDC is involved in the formation of the stratum corneum, the outermost layer of skin which acts as a barrier.\(^26\)

Sequencing of exon 3 is challenging because of its size and the repetitions between units 10 and 12. So far, some 49 mutations in the FLG gene have been reported, all of which are missense and frameshift mutations.\(^27\)\(^28\) Genetic analyses in Asian populations have shown very different results from European populations. The R501X (Arg501Stop) and 2282del4 mutations are most commonly found in European populations. In Asian populations, more commonly found mutations are 3321delA in China and Korea; 441delA, 1249insG, 7945delA, Q2147X, E2422X and R4307X in Chinese-Singaporeans; R501X and 3321delA mutations in Japan; R501X and 2282del4 in Pakistan and E1795X in Taiwan.\(^29\)–\(^36\) Hence, FLG gene mutation screening is challenging in Asian AD patients. More knowledge is needed on the distribution of FLG mutations in different Asian populations, including people from Indonesia.

**LONG-CHAIN POLYUNSATURATED FATTY ACIDS**

EPA, DHA and ARA have a major impact on human health outcomes, such as motor and cognitive development, mental health and psychiatric disorders, cardiovascular disorders and immunologic and inflammatory responses.\(^8\)\(^10\) LCPUFA are essential components of all plasma membranes, modulate transcription and act in cellular signalling pathways through their eicosanoid, docosanoid and resolvin metabolites.\(^37\)

Inflammatory cells usually contain relatively large amounts of ARA and several other LCPUFA. ARA is a substrate for the synthesis of PG, TX, LT and other oxidative derivatives formed by cyclooxygenase (COX) and lipoxygenase (LOX) activity in inflammatory and epithelial cells.\(^38\)\(^39\)
FATTY ACID DESATURASE (FADS) GENE AND ITS RELATIONSHIP TO AD

The FADS1, FADS2 and FADS3 gene cluster is located on chromosome 11q12–q13.1 and has a size of 91.9 kb.\textsuperscript{46} A first candidate gene study published in 2006 in Germany demonstrated that SNPs in the FADS1 and FADS2 genes are predictive of the conversion of precursor PUFA to LCPUFA, as well as the risk of allergic response.\textsuperscript{11} Participants with minor alleles had higher plasma PUFA precursor concentrations and a lower product/substrate ratio, indicative of reduced desaturase activity. These findings were subsequently replicated in many other populations.\textsuperscript{12–14,41–44}

In several studies, the FADS gene location in chromosome 11q12–11q13.1 has been linked with atopy. In a German birth cohort, a significant association between FADS SNPs and the risk of AD was found, while in a Dutch birth cohort study no such association was found, potentially related to different assessment methods for AD.\textsuperscript{42} Further insights in the association of FADS SNPs and PUFA status with AD might be gained by replication studies in populations with different living conditions, such as in Indonesian infants.

**METHODS**

**Study area and population**

The study will be performed in a Primary Healthcare facility in Kemayoran District, Central Jakarta, for 18 months. Participants are apparently healthy newborn infants born full term (37–42 weeks of gestation), whose parents agreed to be included in the study by providing written informed consent.

**Inclusion and exclusion criteria**

Inclusion criteria:
1. Newborns: full term, healthy, birth weight more than 2500 g and no major congenital anomaly.
2. Mothers: agree to take part in the study and sign the informed consent.
3. Mothers with normal gestational history and without complications such as gestational hypertension and gestational diabetes mellitus, and not vegetarian.
4. Mothers without severe illness during or after labour.
5. Mothers who did not take omega-3 and omega-6 supplementation during pregnancy and breastfeeding.

Exclusion criteria:
Newborns who received omega-3 or omega-6 supplementation in syrup/caplet form.

**Study design**

In this prospective birth cohort, we use two research designs, cross-sectional and longitudinal (survival analysis). The cross-sectional design will be used to look for associations between FADS1, FADS2 and FADS3 SNPs and PUFA composition at birth in umbilical artery plasma glycerophospholipids and in buccal swab cell glycerophospholipids.\textsuperscript{45–46} For the first time, data on genetic variants in the FLG gene and the FADS gene cluster, as well as PUFA levels, will be available for Indonesian neonates. Moreover, the relationship between genetic make-up and PUFA levels will be explored by regression methods. In addition, the association of the FLG gene and FADS1, FADS2 and FADS3 SNPs with the emergence of AD in the first year of life will be assessed. We will consider dietary PUFA intake and neonatal PUFA status as covariates by standard logistic regression methods,\textsuperscript{47} as well as by survival analysis approaches to account for the differences in observation times and censoring due to incident AD, disease-free time over the study period or lost to follow-up.\textsuperscript{48}

The longitudinal design will be used to assess average time and differences in incident AD development among groups of infants with different genotypes, PUFA cord plasma glycerophospholipid levels. These analyses will account for covariates such as family history of atopy and infants’ diet and will be assessed over the first year of life by a survival analysis approach.\textsuperscript{48} Thus, this prospective design will also allow us to disentangle the relative contributions of genetic and nutritional aspects, as will statistical path analyses to test for a potential mediating effect of PUFA levels among genotypes and AD development.\textsuperscript{50}

**Ethical aspects**

Participants in this study will be treated in accordance with the Declaration of Helsinki. Ethical review was performed by The Permanent Medical Research Ethics Committee in Medicine and Health/Faculty of Medicine Universitas Indonesia/Dr Cipto Mangunkusumo Hospital no. 47/H2.F1/ETIK/2014 and extended by the letter no. 148/UN2.F1/ETIK/2015.

**Sample size calculation**

We will use a consecutive sampling procedure to collect participants.

The number of participants was calculated using the formula of proportion comparison sample size between two unpaired groups.\textsuperscript{31}

\[
N1 = N2 = \frac{Z_\alpha \sqrt{2PQ} + Z_\beta \sqrt{(P_1 Q_2 + P_2 Q_1)}}{(P_1 - P_2)^2}
\]

where

\[N\] is the sample size;
\[Z_\alpha\] is the alpha raw deviate of 5%, \(Z_\alpha\) value is 1.96;
\[Z_\beta\] is the beta raw deviate=0.84;
\[P_2\] is the atopic dermatitis proportion in minor allele (MiA)=0.06;
\[P_1\] is the atopic dermatitis proportion in major allele (MaA)=0.16;
\[(P_1-P_2)=0.1\text{ and }P=(P_1+P_2)/2=0.11; Q=(1-p)=0.89.

On the basis of the calculations using the above formula, the minimum required number of participants
for each group is 152 newborns. Taking into consideration a 10% loss to follow-up, a minimum of 335 participants are needed. A good power of the study will be achieved if the minimum major allele to minor allele ratio is 3:1, which would result in at least 90 participants in the minor allele group.

**Procedures**

Briefly, FLG gene mutation and FADS1, FADS2 and FADS3 gene polymorphisms will be examined from the buccal coat of umbilical artery and LCPUFA level will be measured in plasma of umbilical artery and in buccal cells in about 400 newborns from a Primary Healthcare Centre in Kemayoran District.

Umbilical artery blood specimens will be collected directly after birth and put into EDTA tubes. Buccal cells will be collected from infants within 1 hour of birth by brushing the surfaces of the inner mouth mucosa 20–25 times with gentle pressure, using a Gynobrush (Herenz Centre in Kemayoran District.

The brush is then put into a Sarstedt tube (62.554.502—15 mL). The brush is held in place by the tube cap, so that the cells stuck to the brush will sediment on centrifugation. Centrifugation will be performed at 1400×g for 10 min at 4°C, the supernatant removed and the tube stored at −80°C. A second buccal specimen will be obtained from the infants at the age of 12 months or at the time when AD is diagnosed. Specimens will be immediately frozen at −80°C and transported by air to Ludwig-Maximilians-Universität München on dry ice where they will be stored at −80°C until analysis.

Questionnaire data consisting of parental age, parity, address, educational level, family income, ethnicity, atopic history of parents and blood siblings, number of siblings, problems in pregnancy, smoking history in pregnancy and maternal exposure to nicotine during pregnancy will be recorded by investigators. Monthly monitoring by telephone will be performed to collect information on breastfeeding, duration of exclusive/predominant or any breastfeeding, formula feeding including type and amount of formula given and any other food provided. At the time when complementary feeding starts, 3-day food-intake recalls will be obtained from mothers and infants and evaluated by NUTRISURVEY (http://www.nutrisurvey.de). Infants will undergo examinations at 3, 6, 9 and 12 months of age at the study centre, including medical history, physical examination and possible diagnosis of AD based on Hanifin & Rajka criteria.

In addition, at any time parents report the presence of skin disorders or a suspicion of AD, the infant will be invited to Pantai Indah Kapuk Hospital for assessment by a skilled dermatologist. If the infant is unable to visit the hospital, a home visit will be performed and photographs of the skin lesion will be subsequently evaluated by an experienced dermatologist.

LCPUFA analysis from buccal cell lipids will be performed after methanol-based ultrasound facilitated lipid extraction. Specimen preparation and gas chromatography (Model 7890 gas chromatograph; Agilent, Darmstadt, Germany) will be performed as described previously. Fatty acid results will be reported in percentage of total fatty acids analysed (mol%). In this analysis, nine PUFA values will be included, namely linoleic acid (LA/C18:2n−6), γ-linolenic acid (GLA/C18:3n−6), dihomo-γ-linolenic acid (DGLA/C20:3n−6), arachidonic acid (ARA/C20:4n−6), adrenic acid (A/C22:4n−6), α-linolenic acid (ALA/C18:3n−3), eicosapentaenoic acid (EPA/C20:5n−3), docosapentaenoic acid (DPA/C22:5n−3) and docosahexaenoic acid (DHA/C22:6n−3).

**FLG gene mutation and FADS1, FADS2 and FADS3 gene polymorphism analysis**

DNA will be extracted from the buccal coat of the umbilical artery by the Puregene DNA isolation kit (Gentra Systems, http://www.gentra.com). Genotyping will be performed using iPLEX Gold Chemistry (Sequenom) and matrix-assisted laser desorption ionisation-ionisation-time of flight (MALDI-TOF) mass spectrometry, with methods to detect allelic differences, as previously described. In brief, locations containing certain SNPs will be amplified by PCR using specific primers. After deactivation by alkaline phosphatase, single base elongation will be performed. In this reaction, primary elongation is performed in accordance to the print order. After salt ion removal by ion switch and elongation reaction, the specimen will be transferred to silicone chip and covered with 3-hydroxypicolinic acid. The differences from specific alleles will be measured by MALDI-TOF. Allele recognition from SNPs will be performed by Mass ARRAY Typer V4.0.5 (Sequenom). We plan to assess 18 SNPs for the FLG gene.

SNPs for FADS genes were selected based on three criteria: (1) the SNP has been studied in previous publications; (2) the SNP candidates in consideration are SNPs that have already been shown to be associated with LC-PUFA status or AD and (3) minor allele frequency is >10%.

Since FLG gene SNPs in Indonesian populations have not been reported and this gene varies among populations, the SNPs were selected based on the National Centre of Biotechnology Information (NCBI) reporting of pathogenic SNPs, which were rs61816761 (for R501X), rs797045090 (Q715X), rs121909626 (S2554X), rs1032929. The brush is then put into a Sarstedt tube (62.554.502—15 mL). The brush is held in place by the tube cap, so that the cells stuck to the brush will sediment on centrifugation. Centrifugation will be performed at 1400×g for 10 min at 4°C, the supernatant removed and the tube stored at −80°C. A second buccal specimen will be obtained from the infants at the age of 12 months or at the time when AD is diagnosed. Specimens will be immediately frozen at −80°C and transported by air to Ludwig-Maximilians-Universität München on dry ice where they will be stored at −80°C until analysis.

Questionnaire data consisting of parental age, parity, address, educational level, family income, ethnicity, atopic history of parents and blood siblings, number of siblings, problems in pregnancy, smoking history in pregnancy and maternal exposure to nicotine during pregnancy will be recorded by investigators. Monthly monitoring by telephone will be performed to collect information on breastfeeding, duration of exclusive/predominant or any breastfeeding, formula feeding including type and amount of formula given and any other food provided. At the time when complementary feeding starts, 3-day food-intake recalls will be obtained from mothers and infants and evaluated by NUTRISURVEY (http://www.nutrisurvey.de). Infants will undergo examinations at 3, 6, 9 and 12 months of age at the study centre, including medical history, physical examination and possible diagnosis of AD based on Hanifin & Rajka criteria.

In addition, at any time parents report the presence of skin disorders or a suspicion of AD, the infant will be invited to Pantai Indah Kapuk Hospital for assessment by a skilled dermatologist. If the infant is unable to visit the hospital, a home visit will be performed and photographs of the skin lesion will be subsequently evaluated by an experienced dermatologist.

LCPUFA analysis from buccal cell lipids will be performed after methanol-based ultrasound facilitated lipid extraction. Specimen preparation and gas chromatography (Model 7890 gas chromatograph; Agilent, Darmstadt, Germany) will be performed as described previously. Fatty acid results will be reported in percentage of total fatty acids analysed (mol%). In this analysis, nine PUFA values will be included, namely linoleic acid (LA/C18:2n−6), γ-linolenic acid (GLA/C18:3n−6), dihomo-γ-linolenic acid (DGLA/C20:3n−6), arachidonic acid (ARA/C20:4n−6), adrenic acid (A/C22:4n−6), α-linolenic acid (ALA/C18:3n−3), eicosapentaenoic acid (EPA/C20:5n−3), docosapentaenoic acid (DPA/C22:5n−3) and docosahexaenoic acid (DHA/C22:6n−3).

**IDENTIFICATION OF THE VARIABLE**

For cross-sectional study design:

The dependent variables are PUFA percentage contents in buccal cells and plasma glycerophospholipids, namely LA (C18:2n−6), GLA (C18:3n−6), DGLA (C20:3n−6), ARA (C20:4n−6), A (C22:4n−6), ALA (C18:3n−3), EPA (C20:5n−3), DPA (C22:5n−3) and DHA (C22:6n−3).
The independent variables are 18 SNPs of the FADS gene, namely rs174548, rs174556, rs174561, rs8834458, rs968567, rs174570, rs174574, rs2727271, rs174576, rs174578, rs174579, rs174602, rs498793, rs526126, rs174575, rs174448, rs174449 and rs174455.

For prospective study design:

The dependent variable is AD, while independent variables are:

▸ 18 SNPs of the FADS1, FADS2 and FADS3 genes, namely rs174548, rs174556, rs174561, rs8834458, rs968567, rs174570, rs174574, rs2727271, rs174576, rs174578, rs174579, rs174602, rs498793, rs526126, rs174575, rs174448, rs174449 and rs174455.

▸ PUFA contents: LA (C18:2n−6), GLA (C18:3n−6), DGLA (C20:3n−6), AA (C20:4n−6), A (C22:4n−6), ALA (C18:3n−3), EPA (C20:5n−3), DPA (C22:5n−3) and DHA (C22:6n−3) in plasma and buccal cell glycerophospholipids.

▸ 5 SNPs of the FLG gene mutation, namely rs61816761, rs797045090, rs121909626, rs74129447 and rs558269137.

DISSEMINATION
We aim at dissemination of findings via publication in an international journal with high impact factor.

PLAN OF DATA PROCESSING, ANALYSIS AND PRESENTATION
All data processing will be performed with SAS 9.4 or R 3.2.4 (http://www.r-project.org). Allele frequency, Hardy–Weinberg equilibrium (HWE), linkage disequilibrium (LD) and further measures will be assessed with R-package Genetics (https://cran.r-project.org/web/packages/genetics/). The latter will be calculated by Fisher’s exact test. The LD test for each pair of SNP locus will be tested with Lewontin’s D. In addition, paired correlation $r^2$ in the study population will be calculated. Normal distribution of fatty acid will be assessed by the Kolmogorov–Smirnov test, box plots and QQ plots. Haplotype analysis will be performed to evaluate for association of the genetic profile and more than one SNP at a time with AD (https://cran.r-project.org/web/packages/haplo.stats/index.html). The results of this study will be disseminated in a high impact journal.

Logistic regression will be performed to evaluate the effect of each SNP and concentration of each fatty acid on the development of AD that arises in the first year of life. Incidence of AD will be estimated by survival analysis, using Kaplan–Meier product limit estimates to assess mean survival times for various groups and to depict disease-free and event times within the first year of life. Differences in survival curves among groups will be tested by log-rank tests. The specific survival analysis model of Cox-proportional hazard regression will be applied to allow multiple adjustments, covariates and potential predictors. Direct and indirect effects of genetic and nutrition (including maternal diet) on AD with PUFA in buccal cells and plasma glycerophospholipids as a potential mediator will be evaluated by moderated mediation path analysis.

REFERENCES
12. Xie L, Innis SM. Genetic variants of the FADS1 FADS2 gene cluster are associated with altered (n–6) and (n–3) essential fatty acids in plasma and erythrocyte phospholipids in women during pregnancy and in breast milk during lactation. *J Nutr* 2008;138:2222–8.
Study protocol to investigate the environmental and genetic aetiology of atopic dermatitis: the Indonesian Prospective Study of Atopic Dermatitis in Infants (ISADI)

Conny Tanjung, Peter Rzehak, Muchtaruddin Mansyur, Zakiudin Munasir, Herawati Sudoyo, Suzanna Immanuel, Roedi Irawan, Eva Reischl, Hans Demmelmaier, Berthold Koletzko, Sri Rezeki Hadinegoro and Damayanti Rusli Sjarif

*BMJ Open* 2017 7: doi: 10.1136/bmjopen-2016-012475

Updated information and services can be found at:
http://bmjopen.bmj.com/content/7/3/e012475

**References**

This article cites 54 articles, 10 of which you can access for free at:
http://bmjopen.bmj.com/content/7/3/e012475#BIBL

**Open Access**

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**

Articles on similar topics can be found in the following collections

- Genetics and genomics (109)
- Paediatrics (613)

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/