University of Dundee

Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index

Hoggart, Clive J.; Venturini, Giulia; Mangino, Massimo; Gomez, Felicia; Ascari, Giulia; Zhao, Jing Hua

Published in:
PLoS Genetics

DOI:
10.1371/journal.pgen.1004508

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 16. Sep. 2020
Novel Approach Identifies SNPs in SLC2A10 and KCNK9 with Evidence for Parent-of-Origin Effect on Body Mass Index

Abstract

The phenotypic effect of some single nucleotide polymorphisms (SNPs) depends on their parental origin. We present a novel approach to detect parent-of-effect (POEs) in genome-wide genotype data of unrelated individuals. The method exploits increased phenotypic variance in the heterozygous genotype group relative to the homozygous groups. We applied the method to >56,000 unrelated individuals to search for POEs influencing body mass index (BMI). Six lead SNPs were carried forward for replication in five family-based studies (of ~4,000 trios). Two SNPs replicated: the paternal rs2471083-C allele (located near the imprinted KCNK9 gene) and the paternal rs3091869-T allele (located near the SLC2A10 gene) increased BMI equally (beta = 0.11 (SD), P < 0.0027) compared to the respective maternal alleles. Real-time PCR experiments of lymphoblastoid cell lines from the CEPH families showed that expression of both genes was dependent on parental origin of the SNPs alleles (P < 0.01). Our scheme opens new opportunities to exploit GWAS data of unrelated individuals to identify POEs and demonstrates that they play an important role in adult obesity.


Editor: Peter M. Visscher, The University of Queensland, Australia

Received February 5, 2014; Accepted May 14, 2014; Published July 31, 2014

Copyright: © 2014 Hoggart et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: ALSPAC: The UK Medical Research Council (MRC), the Wellcome Trust (combined grant ref. 0927310) and the University of Bristol provide core support for ALSPAC. We thank 23andMe for funding the genotyping of the ALSPAC children’s sample. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust. JPK was supported by a Wellcome Trust grant (088413MA). JPK, DME and GDS work in an MRC Unit funded by MC_UU_12013. This publication is the work of the authors and DME will serve as guarantor for the contents of this paper. ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HSNS268201100005C, HSNS2682610110009C, HSNS2682610110010C, HSNS2682610110011C, and HSNS2682610110012C), National Heart, Lung, and Blood Institute contract HHSN26826620062526C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. CoLaus: The CoLaus study was supported by research grants from the Swiss National Science Foundation (grant no: 33CSC0-122661) from GlaxoSmithKline and the Faculty of Biology and Medicine of Lausanne, Switzerland. ZK received financial support from the Leenaards Foundation and the Swiss National Science Foundation (31003A-139314). EPIC: The EPIC Norfolk diabetes case cohort study is nested within the EPIC Norfolk Study, which is supported by programme grants from the Medical Research Council, and Cancer Research UK and with additional support from the European Union, Stroke Association, British Heart Foundation, Research into Ageing, Department of Health, the Wellcome Trust and the Food Standards Agency. Genotyping was in part supported by the MRC-GSK pilot programme grant. We acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study, EPIC-Italy: Methylation analysis on EPIC-Italy samples was made possible by a grant of the Compagnia di San Paolo (Turin, Italy); and by Regione Piemonte (GM), EGCUT: Estonian Genome Center, University of Tartu (EGCUT) received targeted financing from Estonian Government SF0180142s08, Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVARENG). We acknowledge EGCUT technical personnel, especially Mr V. Soo and S. Smit. Data analyses were carried out in part in the High Performance Computing Center of University of Tartu, EU-NN: Support for the EU-nN research was in part provided by the French Ministry of Research and Higher Education, Project Agence Nationale de la Recherche-07-MRAR (France). FamHS: This work was partly supported by NIDDK grant R01DK0925601 (IBB, PI). Fenland: The Fenland Study is funded by the Wellcome Trust and the Medical Research Council (MC_U106179471). We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Coordination team and the Epidemiology Field, Data and Laboratory teams. HYPERGENES: This study was funded by HYPERGENES (FP7 - HEALTH-F4-2007-601150); INTEROMICS (MIUR - CNR Italian Flagship Project). Carlo Rivolta was supported by the Swiss National Science Foundation (Grant 310030_138346) and the Gebert Rüf Foundation, Switzerland (Rare Diseases - New Technologies grant). KORA S3/KORA S4: The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health). Ludwig-Maximilians-Universität, as part of a BMWi Innovative Part of this project was supported by BMWF grant number 01GSG023, by the German National Genome Research Network (NGFNPlus, project number 01GS0823 and project number 01GS0834) and through additional funds from the University of Ulm. LifeLines: The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNV), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. LOLIPOP: The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the NIHR Cardiovascular Biomedical Research Unit of Royal Brompton and Harefield NHS Foundation Trust (the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966,G0704931), the Wellcome Trust (084723/2/08Z) the NIHR (RP-PI-0407-10371),European Union FP7 (EpiMigrant, 279143)and Action on Hearing Loss (G51). We thank the participants and research staff who received the study material. NFBC66: NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114914, 24300796, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (756177, NHLBI grant SR01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (G0500539, G0600705, G0102319, PreVentSyn/SALVE). The DNA extractions, sample quality controls, biobank upkeep and aliquoting was performed in the National Public Health Institute, Biomedical Helsinki, Finland and supported financially by the Academy of Finland and Biocenter Helsinki. We thank the late Professor Paula Rautakallio (launch of NFBC1966), and Ms Ostu Tornwall and Ms Minttu Jussila (DNA biobanking). The authors would like to acknowledge the contribution of the late Academician of Science Leena Peltonen. QIMR: We acknowledge funding from the Australian National Health and Medical Research Council for Medical Research.
Introduction

The effect of genetic variants on phenotypes may depend upon the parent from whom the variant was inherited [1,2]. Parent-of-origin effects (POEs) may arise through imprinting: mechanisms of which include cytosine methylation and histone deacetylation [2]. To date around 50 human genes are known to be imprinted and for most mammalian species less than 1% of the genome is confirmed to be imprinted [3]. One plausible explanation for this phenomenon is the parental conflict hypothesis, whereby both parents would like to maximize the influence of their genome on their offspring [4]. Current methods for detecting parent-of-origin effects rely on assigning parental ancestry to the inherited alleles. This is straightforward in linkage studies, which have identified potential POEs on type 2 diabetes, body mass index (BMI) [5,6], and alcohol intake [7–9]. However, only a very few of these findings have been replicated and the identified linkage peaks often span large chromosomal regions harbouring hundreds of genes, hence the causal gene or regulatory sequence is unknown. A notable exception is the work of Kong et al [1] who inferred parental origin through genealogy information and long-range phasing to subsequently test for POEs. This study identified six SNPs, four associated with risk of type 2 diabetes and the other two associated with each of breast cancer and basal-cell carcinoma.

Genome-wide association studies (GWASs) of unrelated individuals have very precisely identified a large number of genetic loci harbouring SNPs whose (alternative) allele counts associate with common traits. Since GWASs predominantly include unrelated individuals, the parental origin of the alleles cannot be determined, hence genetic effects influenced by the parental origin of the alleles are typically not considered. Here we present a novel approach that is able to detect POEs using genome-wide genotype data of unrelated individuals. We chose BMI as our target trait, due to previous findings [5,6] and the large available sample size. We report the discovery of two novel loci affecting BMI in a manner dependent on the parent-of-origin of the transmitted alleles.

Results

We applied our POE test, which compares the phenotypic variance of the heterozygous genotype group to the variance observed in the homozygous groups, to all SNPs genome-wide. The test, which is applicable to unrelated individuals, assumes that an increased variance in the heterozygous group arises because the heterozygous group consists of two subgroups (paternal reference allele/maternally alternative allele and maternal reference allele/paternal alternative allele) each with different means (see Figure 1). Differences in phenotypic variance were tested using the Brown-Forsythe test, modified to test the mean absolute deviations from the median in the heterozygous and homoygous groups (see Materials and Methods for details). We applied this test to BMI values (corrected for age and age-squared), separately in men and women, in 15 studies, totalling up to 56,092 individuals (detailed description of the cohorts can be found in Tables S1–S3, 13 of which participated in previous meta-analyses of the GIANT consortium [10]. In total 2,673,760 HapMap imputed and genotyped SNPs were tested. For each locus, a lead SNP (with the strongest POE association) was identified; other markers within 1 Mb or in LD ($r^2$>0.1) were excluded from further investigations. Sex-specific association summary statistics were then meta-analysed. No sex specific difference in effects were observed, therefore all reported results are sex-combined.

Our criteria to select SNPs to take forward to the replication stage resulted in the selection of six independent SNPs: four lead SNPs with POE P-value <3×10^{-6} and three SNPs in imprinted regions with P<5×10^{-4} (see Fig S1 for QQ-plot), one SNP fulfilled both criteria. See Table 1 for details of these results and Materials and Methods for details of the applied selection methods. These six SNPs were carried forward to the replication stage.

Replication in family-based studies

The replication stage utilised five family-based studies (see Tables S1–S5) to test for parent-of-origin effects at the six selected SNPs. Only heterozygous individuals are informative when testing for parent-of-origin effects; the number of heterozygous individuals for each of the tested SNPs ranged from 1,122 to 4,128 (see Table 2 and Table S4). A simplified parental asymmetry test (PAT, see Materials and Methods) was applied and SNPs successfully replicated if their PAT P-values were below 0.0003 (Bonferroni corrected significance threshold for family-wise error of 0.05 with six tests). Two of these SNPs, rs2471083 [T/C] (GIAS discovery BMI variance (het vs. hom): 1.058 vs 0.963, $P_{POE}$ = 9.34×10^{-7}; replication PAT $P = 0.00264$) and rs3091869 [T/C] (GIAS discovery BMI variance (het vs. hom): 1.046 vs 0.957, $P_{POE}$ = 4.7×10^{-10}; replication PAT $P = 0.00245$) successfully replicated. In particular, we found that heterozygous individuals who carry the rs2471083-C allele paternaly have 0.11 (SD unit) higher BMI on average than those carrying the C-allele maternally ($P = 0.00264$). Heterozygous carriers of the paternal rs3091869-C allele have 0.11 (SD unit) lower BMI on average than those carrying the maternal C-allele ($P = 0.00245$). Figure 2 shows the locuszoom plots of the POE association P-values for the two replicated loci (KCNK9 and SLC2A10).

Impact of the discovered variants

By combining the effect difference estimates ($\beta_M - \beta_P$) from the family-based studies and the marginal association effect sizes...
Large genetic association studies have revealed many genetic factors influencing common traits, such as body mass index (BMI). These studies assume that the effect of genetic variants is the same regardless of whether they are inherited from the mother or the father. In our study, we have developed a new approach that allows us to investigate variants whose impact depends on their parental origin (parent-of-origin effects), which can be observed (see Table S5). This could be due to insufficient power, (extreme) obese offspring in which parental origin of the alleles was not observed (see Table S5). Our novel approach revealed two SNPs, located near the genes KCNK9 and SLC2A10, influencing BMI in a parent-of-origin specific fashion. These loci were the first and fourth most significant genome-wide in our new POE test for unrelated individuals and both showed significant parent-of-origin effects in family studies. Some of these analyses showed significant association (see Supplemental Data S1, Figures S2, S3 and Table S8 for further details).

Discussion

Our novel approach revealed two SNPs, located near the genes KCNK9 and SLC2A10, influencing BMI in a parent-of-origin specific fashion. These loci were the first and fourth most significant genome-wide in our new POE test for unrelated individuals and both showed significant parent-of-origin effects in family studies. Both SNPs exhibit polar overdominance, where homozygous individuals have equal (baseline) phenotypes and heterozygous genotypes confer relative risk/protection, depending on the parental origin. Polar overdominance, has been observed in humans for type2 diabetes [1] and BMI [20], however it is very rare and its molecular mechanism is unknown.

RT-PCR experiments revealed that gene expression levels of KCNK9 and SLC2A10 in LCLs were also influenced in a parent-of-origin manner. The expression of these genes is highest in the brain (although it is also expressed in testis, liver, colon, adrenal gland and kidney; see http://www.genecards.org/) indicating a potential neuronal involvement. Expression levels of KCNK9 and SLC2A10 in living brain cells might have been more informative and robust, however such information is not available. The applied qPCR method was optimised to ensure that the expression levels measured in LCLs were representative only of the target transcript and amplification efficiency was assessed to be sensitive enough to allow the detection of even small changes in gene expression. Interestingly, rs2471083 alleles, regardless of their

...
parental origin, show marginally significant association (P = 0.03) with KCNK9 expression levels in the hippocampus (http://www.broadinstitute.org/gtex/). Our methylation analyses did not reveal any evidence that the POEs were driven by differences in inherited paternal and maternal methylation. Neither of our two SNPs tag common copy number variants (CNVs) (based on the CNV reference data used in Heid et al. [21]) and we found only one sample (out of 14,315 available in-house, whose BMI Z-score was +1.18) with a 76 kb deletion overlapping rs2471083. Hence, the effect of the two discovered SNPs are unlikely to be driven by CNVs. To check whether the two confirmed SNPs (rs3091869, rs2471083), or SNPs in LD (10 with \( r^2 > 0.8 \) in 1000 Genomes EUR population), show regulatory activity, we queried RegulomeDB (http://regulome.stanford.edu). None of these SNPs were annotated to have more than minimal binding evidence (RegulomeDB score below 4).

A previous study proposed to detect POE in inbred F₂ mice by a two-component mixture distribution fitting of the heterozygous genotype group and further two components for the homozygous groups [22]. This method requires a parametric distribution of the phenotype to be assumed, small violations of this assumption can result in heavily biased parameter estimates. The method we chose is more robust to a wide range of phenotype distributions (due to the underlying Brown-Forsythe test employed), computationally faster (making it attractive for testing millions of SNPs) and applicable to probabilistic genotype calls. Our POE test for unrelated GWAS samples is similar to a test proposed to detect gene-environment interactions [23] in that it exploits differences in phenotypic variance to detect a phenomenon not directly measured. Inflated phenotypic variance in the heterozygous group might also be the result of other phenomena: (i) a phenotype altering effect (be it genetic or environmental) acting only on the

---

**Figure 1.** Explanation of the POE test. Top panel illustrates the phenotype distributions in the four genotype groups that would be observed if the parent-of-origin of the alleles were known. Bottom panel shows how these distributions change if the parent-of-origin is unobserved. The resulting heterozygous group will have increased variance due to its heterogeneity. This example describes a scenario we observe for the two replicated hits, namely that the paternal- and maternal effects are of the same size, but opposite in direction (\( \beta_M = -\beta_P \)). Therefore the average phenotype in the B/B group is the same as in the A/A group, as the paternal and maternal B allele effects cancel each other out. In the A/B group there are two subpopulations: the A-pat/B-mat group with phenotypic mean of \( \beta_M \) and the A-mat/B-pat group with \( \beta_P \) mean. Thus, the two subpopulations combined also have zero mean, but increased variance.

doi:10.1371/journal.pgen.1004508.g001
Table 1. Discovery POE association results for the six SNPs selected for replication.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>position (B36)</th>
<th>Gene (distance in kb)</th>
<th>MAF</th>
<th>N (AA+BB)</th>
<th>N (AB)</th>
<th>P-value (POE)</th>
<th>P-value (main)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genome-wide (P&lt;5e-6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs155570</td>
<td>2</td>
<td>59778667</td>
<td>AC007131.1(419)</td>
<td>0.2</td>
<td>32643</td>
<td>15232</td>
<td>2.12E-06</td>
<td>1.70E-03</td>
</tr>
<tr>
<td>rs2471083</td>
<td>8</td>
<td>140588522</td>
<td>KCNK9(105)</td>
<td>0.27</td>
<td>33885</td>
<td>22132</td>
<td>9.34E-07</td>
<td>8.60E-01</td>
</tr>
<tr>
<td>rs1345919</td>
<td>15</td>
<td>31961311</td>
<td>AVEN(0)</td>
<td>0.22</td>
<td>36850</td>
<td>19142</td>
<td>4.90E-06</td>
<td>5.15E-01</td>
</tr>
<tr>
<td>rs3091869</td>
<td>20</td>
<td>44859324</td>
<td>SL2CA10(61)</td>
<td>0.44</td>
<td>23891</td>
<td>23143</td>
<td>4.70E-06</td>
<td>3.93E-02</td>
</tr>
<tr>
<td>Imprinted regions (P&lt;5e-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1398940</td>
<td>4</td>
<td>90156206</td>
<td>FAM13A(0)</td>
<td>0.39</td>
<td>25173</td>
<td>22737</td>
<td>4.50E-04</td>
<td>5.86E-01</td>
</tr>
<tr>
<td>rs2471083</td>
<td>8</td>
<td>140588522</td>
<td>KCNK9(105)</td>
<td>0.27</td>
<td>33885</td>
<td>22132</td>
<td>9.34E-07</td>
<td>8.60E-01</td>
</tr>
<tr>
<td>rs6500550</td>
<td>16</td>
<td>3686241</td>
<td>TRAP1(0)</td>
<td>0.31</td>
<td>27302</td>
<td>20188</td>
<td>1.69E-04</td>
<td>8.44E-03</td>
</tr>
</tbody>
</table>

SNP matching any of the following two selection criteria were retained: (1) All lead SNPs with POE P-value below 5 x 10^{-6}. (2) All lead SNPs falling in previously reported imprinted regions with POE P-value <5 x 10^{-4}. Note that one SNP is listed twice as it met both criteria. Chr: chromosome, Gene: nearest gene, MAF: minor allele frequency, N (AA+BB): homozygous sample size, N (AB): heterozygous sample size, P-value (main): main effect association P-value in Speliotes et al. [10]. The parent of origin P-value, “P-value (POE)”, is the Brown-Forsythe test P-value. SNPs marked in bold are those which replicated in the family studies, see Table 2.

doi:10.1371/journal.pgen.1004508.t001

Table 2. Replication of the 6 discovery SNPs in trios (or parent-offspring pairs).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Coded allele</th>
<th>Other allele</th>
<th>beta (coded paternal) - beta (coded maternal)</th>
<th>P-value for beta difference</th>
<th>beta meta</th>
<th>P meta</th>
<th>N meta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QMR</td>
<td>FAMHS</td>
<td>FHS</td>
<td>ALSPAC</td>
<td>GS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QMR</td>
<td>FAMHS</td>
<td>FHS</td>
<td>ALSPAC</td>
<td>GS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs155570</td>
<td>G</td>
<td>A</td>
<td>0.218 -0.033 -0.098 0.23 0.136</td>
<td>5.82E-03 7.97E-01 1.72E-01 2.43E-02 2.32E-01</td>
<td>0.079 5.57E-02</td>
<td>1646:1093</td>
<td></td>
</tr>
<tr>
<td>rs1345919</td>
<td>A</td>
<td>T</td>
<td>-0.17 -0.167 0.098 -0.032 -0.006</td>
<td>3.72E-02 2.06E-01 1.56E-01 7.15E-01 9.54E-01</td>
<td>-0.03 4.46E-01</td>
<td>1231:1696</td>
<td></td>
</tr>
<tr>
<td>rs1398940</td>
<td>G</td>
<td>A</td>
<td>0.104 0.033 0.098 -0.07 0.008</td>
<td>1.41E-01 7.57E-01 9.50E-02 2.70E-01 9.30E-01</td>
<td>0.038 2.45E-01</td>
<td>1984:2144</td>
<td></td>
</tr>
<tr>
<td>rs2471083</td>
<td>C</td>
<td>T</td>
<td>0.012 0.064 0.24 0.007 0.135</td>
<td>8.81E-01 5.62E-01 1.48E-04 9.31E-01 1.89E-01</td>
<td>0.112 2.64E-03</td>
<td>1453:1959</td>
<td></td>
</tr>
<tr>
<td>rs3091869</td>
<td>C</td>
<td>T</td>
<td>-0.107 -0.152 -0.117 -0.062 -0.199</td>
<td>1.19E-01 1.29E-01 2.75E-01 3.35E-01 5.05E-02</td>
<td>-0.112 2.45E-03</td>
<td>1477:1696</td>
<td></td>
</tr>
<tr>
<td>rs6500550</td>
<td>T</td>
<td>C</td>
<td>0.053 -0.306 0.039 -0.169 -0.03</td>
<td>4.64E-01 3.14E-03 5.79E-01 2.22E-02 7.29E-01</td>
<td>-0.057 1.08E-01</td>
<td>1952:1844</td>
<td></td>
</tr>
</tbody>
</table>

For each target SNP, in each family we searched for trios (or parent-offspring pairs) with heterozygous offspring and determined the parent of origin of the alleles (whenever possible). From each family at most one heterozygous offspring with known parental origin was then collected and grouped according to the parental origin of the alleles. The equality of phenotypic means in the two groups was tested using a Student t-test (“P-value for beta difference column”). The difference between the phenotypic means were meta-analysed using inverse-variance weighting. Table notations: “beta meta” and “P meta”: meta-analysis estimate of beta (coded paternal) - beta (coded maternal) effect size differences and the corresponding P-value, “N meta”: total number of heterozygous offspring with (coded-maternal/other-paternal), (coded-paternal/other-maternal) genotypes, respectively.

doi:10.1371/journal.pgen.1004508.t002
heterozygous group; (ii) an overdominant effect combined with a genetic or environmental interaction or non-linear, monotonic phenotype transformation that has different derivatives for low and high trait values; (iii) a large marginal additive effect combined with a (monotonic) transformation for which the second derivative is maximised at the mean phenotype value of the heterozygous group (see Materials and Methods for details). More generally, the combination of the scale on which the phenotype is measured and a strong marginal association with an allelic dosage may give rise to spurious associations using variance tests [24]. Recently some evidence has emerged about loci which effect the variance of phenotypes (through impacting environmental plasticity, canalization, developmental stability, etc.) that can be detected via association with phenotypic variability [25]. Therefore, the top hits obtained by our POE test may need further prioritisation before proceeding to trio-based confirmation. We recommend the following checks: (a) Exclude SNPs with overdominant effects; (b) For SNPs with low POE P-value, test gene-environment (GxE) interaction (as done in [23]) via modelling phenotypic variance as a function of the genotype dosage (coded in additive, recessive or dominant fashion). If this test is more significant than the POE test, it is probably a GxE that is driving the POE association and also as a side effect we will observe significant difference in the variance between the two homozygous groups. (c) If a SNP with low POE P-value has marginal effect on the trait, repeat the POE test for various transformed versions of the phenotype such as log and inverse-normal quantile. If the resulting POE P-values are not robust, give lower priority to the examined SNP.

For our confirmed SNPs multiple lines of evidence show that the parent-of-origin effects are real, most convincingly clear replication in independent family data of parent-of-origin associations of the hit SNPs with both BMI and gene expression levels. Further, the GWAS discovery associations are very unlikely to be artefacts of the factors discussed above: (i) there is no evidence of overdominant, additive, recessive or dominant effects (the mean BMI values are near identical in the three genotype groups), hence the signals cannot be driven by gene-environment interactions or be an artefact of the scale on which the phenotype is measured (ii) no SNP within 500 kb has any detectable marginal effect on BMI thus the association cannot be driven by haplotype-specific marginal effects [26]; (iii) the phenotypic variances in the two homozygous groups, are almost identical (rs2471083: $\sigma^2_C = 0.96$, $\sigma^2_T = 1.06$, $\sigma^2_{CT} = 0.98$ and rs3091869: $\sigma^2_C = 0.97$, $\sigma^2_T = 1.05$, $\sigma^2_{CT} = 0.95$); (iv) POE test with log- and inverse-normal quantile transformed BMI values resulted in similar results (Table S6), further reducing the likelihood of an artefact resulting from the scale on which the phenotype is measured [24].

Some of the negative results of the other SNPs carried forward to the replication phase in the family data could be explained by lack of power. The power to replicate POE associations in family-based studies is dependent on the available number of heterozygous individuals (for details see Supplemental Data S1) and thus increases with minor allele frequency (MAF). Therefore, it is unsurprising that the two SNPs which replicated had relatively high MAF (>27%).

Linkage studies have identified four regions exerting POE on BMI (10p12, 12q24, 13q32 [5] and 2q31 [27]). We looked up SNPs in these regions in our genome-wide discovery POE association results. The reported linkage regions showed enrichment for lower than expected POE P-values (see Figure S4 for regional QQ-plots), however, no SNPs survived Bonferroni correction. We also tried to replicate a SNP in exon 5 of DLK1 (rs1802710) because this SNP showed polar overdominance for obesity in children [20], but only a very slight trend ($P = 0.32$) was visible in our study. Previously reported BMI-associated loci [10] show some enrichment for lower POE P-values (Supplemental Data S1, Tables S9, S10 and Figure S5), however these need to be replicated in family studies.

Previous work comparing strength of associations of mother-offspring BMI with father-offspring BMI did not reveal intrauterine influence on obesity in children [28]. A similar conclusion was reached in a systematic review of seven studies [29], while stronger maternal influence was observed in a recent longitudinal study [30]. The difference in conclusions may be due to that fact that the former studies included predominantly older children than the longitudinal study (0–3.5 years). At early age the diet of the offspring may be more similar to that of the mother than the father (e.g. due to breastfeeding), which might have contributed to the higher mother-offspring BMI similarity found by Linabery et al. [30].

In summary, our findings indicate that POEs may play a role in adult obesity. The two identified SNPs have strong parent-of-origin effect on BMI, close to that of the FTO, contributing substantially to the heritability of BMI. Our follow-up experiments demonstrated parent-of-origin specific gene expression modulation, but failed to link methylation activity of these loci to BMI values. Inevitably for newly discovered loci, further studies are warranted to determine how these variations functionally influence obesity in humans. The reliance of our approach on difference in phenotypic variance means that it cannot be extended to binary
outcomes. Since there are other phenomena which can give rise to significant POE association, we recommend that top hits from our method are followed up in family studies, where parental origin of alleles can be inferred. In addition, our variance based POE test for GWAS data is naturally much less powerful than actually testing the mean values of the two heterozygous subgroups in trios. However, GWASs of unrelated individuals are several-fold more numerous and typically much larger than studies with a trio design; hence our methodology provides a great advance in parent-of-origin research by providing means to exploit all available GWAS data of unrelated individuals in order to identify parent-of-origin effects on continuous phenotypes.

Materials and Methods

Ethics statement

All participating studies were approved by the respective institutional Ethics Committees. All study participants gave written consent including for genetic studies.

Detecting parent of origin effects

If we denote the alleles of a bi-allelic SNP by “A” (reference) and “B” (alternative) the possible genotypes are A/A, A/B and B/B. Standard GWASs estimate the effect of the alternative allele dosage on the phenotype in question. In this work we are interested in associations in which a phenotype \(y\) is influenced by the alleles of a particular SNP and the effect depends on the parental origin of these alleles. In the presence of a parent-of-origin effect the heterozygous genotype group is split into two subgroups, depending on the parental origin of the A and B alleles.

We assume that the phenotype of any individual in the A/A genotype group is modelled by \(y_{AA} = \mu_{AA} + \varepsilon\), where \(\mu_{AA}\) is the mean and \(\varepsilon\) is an individual level error with mean zero and variance \(\sigma^2\). If the maternal and paternal effects of the B allele are \(\beta_M\) and \(\beta_P\), it follows that the phenotype of an individual in the B/B group is \(y_{BB} = (\mu_{AA} + \beta_M + \beta_P) + \varepsilon\) and its variance is \(\sigma^2\). (Note that as a consequence the maternal and paternal effects of the A allele are \(-\beta_P\) and \(-\beta_M\).) Here we assume \(\sigma^2\) is constant across genotype groups (A/A, A/B and B/B) and \(\beta_M\) and \(\beta_P\) are fixed effects. The effects of violations of these assumptions are covered in the discussion. The phenotype in the heterozygous group is a 50%-50% mixture of two distributions (Fig 1a):

\[ y_{AB} = \mu_{AA} + \pi\beta_M + (1-\pi)\beta_P + \varepsilon, \]

where \(\pi\) is a Bernoulli random variable (with parameter \(1/2\)), taking values \(\pi = 1\) if the B allele is inherited from the mother and \(\pi = 0\) if inherited from the father. The heterozygous phenotype distribution can be simplified to \(y_{AB} = (\mu_{AA} + \beta_P) + \pi(\beta_M - \beta_P) + \varepsilon\). Since \(\pi\) and \(\varepsilon\) are independent random variables, the phenotypic variance of the heterozygous genotype group \(\sigma_{AB}^2\) is

\[ \text{Var}(y_{AB}) = \text{Var}(\pi)(\beta_M - \beta_P)^2 + \text{Var}(\varepsilon) = \frac{1}{4}(\beta_M - \beta_P)^2 + \sigma^2. \]

If a parent-of-origin effect is present \(\beta_M\) and \(\beta_P\) are different, thus \(\sigma_{AB}^2\) is larger than the variance observed in the homozygous groups (\(\sigma^2\)) (Figure 1). Therefore, although in regular GWAS data we cannot identify the two subgroups within A/B genotypes, we can detect POE via increased phenotypic variance in the heterozygous group relative to the homozygous groups.

In the presence of a marginal association a phenotype transformation could alter the genotype group variances and introduce bias into the test [24]. For this reason we analysed untransformed age-, age²-corrected BMI values (normalised to have zero mean and unit variance) separately for men and women. Standard variance tests (such as the F-test) are, however, sensitive to deviations from the Gaussian distribution.

Therefore, we used a robust version of the Brown-Forsythe test. Briefly, we first centred the phenotype values (at zero) in each genotype group to avoid inflated variance in the presence of marginal effects in the group of all homozygote individuals. We denote these centred phenotypes by \(z\), where

\[ z_i = \begin{cases} y_i - \mu_{AA} & \text{if } g_i = AA \\ y_i - \mu_{AB} & \text{if } g_i = AB \\ y_i - \mu_{BB} & \text{if } g_i = BB \end{cases} \]

Here \(g_i\) stands for the genotype of individual \(i\), and \(\mu_j\) represents the median phenotype value in genotype group \(j\), where \(j\) can take the values of AA, AB or BB. We then regress the absolute deviations from the median onto a 0–1 coded genotype group identifier (1 for heterozygous and 0 for homozygous individuals) in order to estimate the POE effect size [31]. This regression result in a slope estimate

\[ \hat{b} = z_{het} - z_{hom}, \]

where \(z_{het} = \sum_{i\in AA}(z_i)/n_{AA}\) and \(z_{hom} = (\sum_{i\in A}(z_i) + \sum_{i\in B}(z_i))/n_{AA} + n_{AB}\). The corresponding standard error is

\[ \text{SE}_{\hat{b}}^2 = \frac{\text{RSS}}{n-1} \frac{n_{AA} + n_{BB}}{n}, \]

where

\[ \text{RSS} = \sum_{i}(z_i - \hat{y}_i)^2 - \hat{b}^2n_{AB}(n_{AA} + n_{BB})/n \]

Finally, the POE P-value is assigned based on the test statistic \(\hat{b}/\text{SE}_{\hat{b}} \sim \mathcal{N}(0,1)\). The test was extended to imputed genotype probabilities and implemented in the latest version (v0.98) of the Quicken software (http://www3.unil.ch/wpmu/sgg/quicktest/). The robustness of this test to deviations from normality has been studied in [32] and its power in [31].

SNP selection strategy

We applied our POE test genome-wide to all HapMap imputed markers in a set of cohorts and results were combined across cohorts using fixed-effect inverse-variance weighting meta-analysis. SNPs were selected for replication if they met at least one of the following two criteria: (1) POE P-value <5×10⁻⁶ or (2) POE P-value <5×10⁻⁴ and within 500 kb of previously reported imprinted regions according to the Catalogue of Parent of Origin Effects database (http://igc.otago.ac.nz/home.html). At loci which met either criteria, a lead SNP (with the strongest POE association) was identified; other markers within 1 Mb or in LD (\(r^2 > 0.1\)) were excluded from further investigations. In total 2,673,768 HapMap imputed and genotyped SNPs were analysed,
of which 29,457 were considered as lying in imprinted regions, criterion [2]. Using the procedure of Gao et al. [33] we estimated the effective number of tests considered by each criterion to be \( \sim 1,000,000 \) and 6,100 respectively, justifying the \( \sim 100 \) fold drop in the \( P \)-threshold applied to the second criterion.

Testing in family-based studies

We tested our findings in family-based studies using a simplified parental asymmetry test [34] (PAT). For each target SNP, in each family we searched for trios (or parent-offspring pairs) with heterozygous offspring and determined the parent of origin of the alleles (whenever possible, i.e. at least one homozygous parent). From each family at most one heterozygous offspring with known parental origin was then collected and grouped according to the parental origin of the alleles. Note that although POE is acting in every genotype group, it can only be detected in the heterozygous group.

As at the discovery phase, we used sex-, age- and age\(^2\)-corrected BMI residuals as phenotype. The equality of phenotypic means in the two groups was tested using a Student t-test. When significant differences were detected we also estimated the difference between paternal and maternal effect sizes, which is simply the difference between the phenotype averages in the paternal- and maternal-groups.

Effect size estimation

In order to estimate paternal \( \beta_P \) and maternal \( \beta_M \) effect sizes it is sufficient to know their mean \( (\hat{\beta}_M + \hat{\beta}_P) / 2 \) and their difference \( (\hat{\beta}_M - \hat{\beta}_P) \). The difference between paternal and maternal effect alleles can also be derived from GWAS of unrelated individuals. It is easy to see that the test statistic defined as

\[
T = 4 \left( \sigma^2_{AB} \frac{(n_{AA} - 1) \sigma_{AA}^2 + (n_{BB} - 1) \sigma_{BB}^2}{n_{AA} + n_{BB} - 2} \right)
\]

gives an unbiased estimate of \( (\hat{\beta}_M - \hat{\beta}_P)^2 \). Since \( \text{Var}(\sigma_{a}^2) = \sigma^2_{a0} / n_{a0} - 1 \), the variance of \( T \) is \( \approx \frac{1}{n_{AB}} + \frac{1}{n_{AA} + n_{BB} - 2} \). Therefore, the absolute difference in paternal and maternal effects \( |\hat{\beta}_M - \hat{\beta}_P| \) can be estimated if the phenotypic variances in the three genotype groups are known. However, these estimates will be strongly subject to the winner’s curse [33], thus we used the family studies to derive more reliable estimates of \( (\hat{\beta}_M - \hat{\beta}_P) \). To reduce the effect of differences in the distribution of BMI between the family-based studies, we meta-analysed the difference estimates \( (\hat{\beta}_M - \hat{\beta}_P) \) from each study in order to obtain a combined estimate of \( (\hat{\beta}_M - \hat{\beta}_P) \). The average of the maternal- and paternal effects, \( (\hat{\beta}_M + \hat{\beta}_P) / 2 \), is the association effect size using a simple additive genetic model, which can be most accurately estimated from the largest-to-date meta-analytic study on BMI [10] (including \( \sim 250,000 \) individuals).

Effect of phenotype transformation in case of marginal association

If there is an additive marginal genetic effect influencing the trait certain transformations may inflate the phenotypic variance of the heterozygous group. Let \( \mu \) be the phenotypic mean in the heterozygous group and \( \delta \) the marginal effect of the SNP (on the original scale). Let \( g(t) \) denote an S-shaped transformation function of the form

\[
g(t) = \frac{1}{1 + e^{at - \mu}}
\]

that is applied to the trait.

In the following we show that for any \( \delta \) value arbitrarily large phenotypic variance inflation can be achieved in the heterozygous group, compared to the two homozygous groups by an appropriate parameter choice for \( a \). Using a second order Taylor expansion the variance of the transformed phenotype in the heterozygous group can be estimated by

\[
\text{Var}(g(y)|G = AB) \approx g'(\mu)^2 \sigma^2 + \frac{1}{4} g''(\mu) [\text{Var}(y)|G = AB)^2] - 4g'(\mu)^2 \sigma^2
\]

If we assume the phenotype follows a Gaussian distribution \( y|G \) then, \( \text{Var}(y|G = AB)^2 \) simplifies to \( 4g'(\mu)^2 + 2\sigma^2 \) and thus

\[
\text{Var}(g(y)|G = AB) \approx g'(\mu)^2 \sigma^2 + \frac{1}{2} g''(\mu)^2 \sigma^2
\]

Without loss of generality one can assume that \( \sigma^2 = 1 \). The variance in AA genotype group can be estimated similarly and thus

\[
\text{Var}(g(y)|G = AB) - \text{Var}(g(y)|G = AA)
\]

\[
\approx [g'(\mu)^2 - g'(\mu - \delta)^2] + \frac{1}{2} g''(\mu)^2 - g''(\mu - \delta)^2
\]

Using the special form of \( g(t) \), the variance difference can be expressed as

\[
\text{Var}(g(y)|G = AB) - \text{Var}(g(y)|G = AA)
\]

\[
\approx \frac{1}{16} - \frac{e^{2a\delta}}{(1 + e^{a\delta})^4} - \frac{a^2}{2} \frac{(e^{a\delta} - e^{-a\delta})^2}{(1 + e^{a\delta})^4}
\]

\[
= \frac{1}{16} - \frac{a^2}{2} \frac{e^{2a\delta} - e^{-2a\delta}}{(1 + e^{a\delta})^4} - \frac{a^2}{2} \frac{(e^{a\delta} - e^{-a\delta})^2}{(1 + e^{a\delta})^4}
\]

and since \( g'(\mu - \delta) = g'(\mu + \delta) \) and \( g''(\mu - \delta) = -g''(\mu + \delta) \)

\[
\text{Var}(g(y)|G = AB) - \text{Var}(g(y)|G = BB)
\]

\[
\approx \text{Var}(g(y)|G = AB) - \text{Var}(g(y)|G = AA).
\]

It is easy to see that for a fixed \( \delta \)

\[
\lim_{a \to \infty} f(a, \delta) = 0
\]

As \( a \to \infty \), \( f(a, \delta) \to 0 \) faster than \( a^2 \to \infty \), thus for any effect size \( \delta \) we can find a transformation function \( g \) such that the variance inflation of the heterozygous group exceeds any arbitrary threshold.

Cell lines, nucleic acids isolation, sequencing and qPCR

Lymphoblastoid cell lines were derived from peripheral blood leukocytes of 95 members of 11 CEPH families [36] (#102, #884, #1335, #1340, #1341, #1345, #1346, #1347, #1362, #1408, #13292). They were purchased from the Coriell Cell Repository
We then loaded 10 primers for qPCR, we first performed a series of test amplifications facilitate the detection of poorly expressed genes. To validate instructions and a mix of random hexamers and oligo-dT that (Invitrogen/Life Technologies) according to the manufacturer’s instructions. Primer sequences were designed to amplify a 328-bp region on chromosome 8 that spans the rs2471083 polymorphism (forward primer: 5'--ACGACGAGTGCAGTGAG-G--3'; reverse primer: 5'--GTGACATTGGGAGACATGGGA--3'); and a 146-bp region on chromosome 20 that spans the rs9092611 polymorphism (forward primer: 5'--GCCACACGTGGTGATAGT--3'; reverse primer: 5'--TAACTGTCATTGTGCCCTGG--3'). PCR amplification was performed in a 25 μl reaction using GoTaq polymerase (Promega). After purification of PCR products (ExoSAP-IT, USB), sequencing reactions were carried out by using 1 μl of each of the 3.2 μM sequencing primers and 0.5 μl of BigDye Terminator v1.1 (Applied Biosystems). Following on-column purification (EdgeBio), sequencing products were run on an ABI-3130 XLS sequencer (Applied Biosystem). To synthesize cDNA, 2 μg of total RNA was retrotranscribed using the Superscript III reverse transcriptase (Invitrogen/Life Technologies) according to the manufacturer’s instructions and a mix of random hexamers and oligo-dT that facilitate the detection of poorly expressed genes. To validate primers for qPCR, we first performed a series of test amplifications by using a defined range of primer concentrations (50–200 nM). We then loaded 10 μl of each qPCR product on 1% agarose gels to check the specificity of the amplification product, which should correspond to a 113-bp (KCNK9) and 148-bp (SLC2A10) fragment. To test KCNK9 and SLC2A10 PCR efficiency a standard curve made of five serial dilutions of brain and lung cDNA were used, respectively, since the two genes are known to be highly expressed in these organs. We obtained a standard curve slope of −3.49 for KCNK9 and of −3.37 for SLC2A10, corresponding to 94% and 98% PCR efficiency. For more details see Supplemental Data S1.

**Comparing Ct values**

The output of the analysis was threshold cycles (Ct), i.e. the number of cycles at which the fluorescent signal of the reaction crosses a pre-determined threshold value. Since standard quantification methods (including normalization by housekeeping genes) introduce a considerable amount of experimental noise for very lowly expressed genes, raw Ct values were used to perform an absolute quantification of KCNK9 and SLC2A10 transcripts. As negative controls, housekeeping genes (HPRT1, GAPDH) were also tested for parent-origin-effect to exclude the possibility that the observed difference in KCNK9 and SLC2A10 expression levels was due to the sample preparation process. Raw Ct values were inverse-normal quantile transformed and a linear mixed model was fitted (using the R function lmer) modelling the technical replicates as random effects and parental origin as a fixed effect.

**Supporting Information**

**Figure S1** QQ-plot of the POE test P-values for SNPs in imprinted regions (left) and for the whole genome (right). (PNG)

**Figure S2** Left hand side plots describe SNP-methylation associations (mQTLs), where each point is a methylation probe. X-axis represents their physical position and y-axis the −log10 association P-value with the target SNP, whose location is indicated by the dashed line. Note that rs3092611 was used as a proxy for rs3091869 ($r^2 = 0.998$). The corresponding QQ-plots appear on the right hand side. Neighbouring methylation probes are strongly correlated therefore expected P-values were computed by estimating the effective number of tests. For expected P-values we computed the effective number of tests [33]. (PNG)

**Figure S3** Left hand side plots describe methylation associations with BMI in MZ twins. Each point is a methylation probe, X-axis represents their physical position and y-axis the −log10 association P-value with BMI. Location of the target SNP is indicated by the dashed line. Note that rs3092611 was used as a proxy for rs3091869 ($r^2 = 0.998$). The corresponding QQ-plots appear on the right hand side. Neighbouring methylation probes are strongly correlated therefore expected P-values were computed by estimating the effective number of tests. For expected P-values we computed the effective number of tests [33]. (PNG)

**Figure S4** POE association results for previously reported imprinted BMI linkage regions. Second dashed line corresponds to the Benjamini-Hochberg 5% FDR threshold. (PNG)

**Figure S5** POE association P-value QQ-plot for the top 58 independent SNPs with marginal BMI-association P-value $<10^{-5}$ in Speliotes et al. [10]. (PNG)

**Table S1** Description of the BMI distribution in the participating cohorts. (XLSX)

**Table S2** Brief summary of the participating cohorts. (XLSX)

**Table S3** Information on the genotyping methods in the participating cohorts. (XLSX)

**Table S4** Replication of the 6 discovery SNPs in family studies. Extended version of Table 2. (XLSX)

**Table S5** Parental asymmetry test results for our 6 candidate SNPs in adolescent population. (XLSX)

**Table S6** Analysis of the effect of gender- and age-correction and phenotype transformations on the POE association results. (XLSX)

**Table S7** Genotyping results and parental allele determination for members of the CEPH families. (XLSX)

**Table S8** Association results between methylation- and BMI-differences among MZ twins of the TwinsUK study. (XLSX)

**Table S9** POE association P-values for the 32 SNPs associated with BMI in the largest-to-date meta-analysis [10]. (XLSX)

**Table S10** POE association P-values for the top 58 independent SNPs with BMI-association P-value $<10^{-5}$ in the largest-to-date meta-analysis [10]. (XLSX)

**Data S1** Supporting information contain text on the expression- and methylation analysis, the effective sample size derivation in mother-offspring vs trio studies, POE look-ups for SNPs with marginal BMI association, study descriptions and acknowledgements. (DOCX)
Acknowledgments

A full list of acknowledgments can be found in the Supporting Information. We would like to thank Rosanna Pescini-Gobert for help with the qPCR analysis. A full list of acknowledgments can be found in the Supporting Information.

Members of the Scientific Board of the LifeLines Cohort Study

P J Bakker, M Kuivenhoven, M van Duijn, J Posthuma, E Bijl, M van Boxtel, M Ingelbrecht, R van Duijn, S Polderman, M Nyholt, R Grotz, P Hofman, T Sijbrands, L Kromhout, J Schouten, M Holling, M Hofman, J van Duijn, L van den Beucken.

Members of the Scientific Committee of the Generation Scotland Study


References


