

Short Communication

For reprint orders, please contact: reprints@futuremedicine.com**Biomarkers
in Medicine**

Genome-wide analysis of nuclear magnetic resonance metabolites revealed parent-of-origin effect on triglycerides in medium very low-density lipoprotein in *PTPRD* gene

N Pervjakova^{*1,2,3}, V Kukushkina^{1,2}, T Haller¹, S Kasela¹, A Joensuu^{4,5}, K Kristiansson^{4,5}, T Annilo^{1,2}, M Perola^{1,4,5}, V Salomaa⁴, P Jousilahti⁴, A Metspalu^{1,2} & R Mägi¹¹Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia²Department of Biotechnology, Institute of Molecular & Cell Biology, University of Tartu, Tartu 51010, Estonia³Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, UK⁴National Institute for Health & Welfare (THL), Department of Public Health Solutions, Helsinki, Finland⁵Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland^{*}Author for correspondence: Tel.: +372 534 88 567; natalia.pervjakova@ut.ee

Aim: The aim of the study was to explore the parent-of-origin effects (POEs) on a range of human nuclear magnetic resonance metabolites. **Materials & methods:** We search for POEs in 14,815 unrelated individuals from Estonian and Finnish cohorts using POE method for the genotype data imputed with 1000 G reference panel and 82 nuclear magnetic resonance metabolites. **Results:** Meta-analysis revealed the evidence of POE for the variant rs1412727 in *PTPRD* gene for the metabolite: triglycerides in medium very low-density lipoprotein. No POEs were detected for genetic variants that were previously known to have main effect on circulating metabolites. **Conclusion:** We demonstrated possibility to detect POEs for human metabolites, but the POEs are weak, and therefore it is hard to detect those using currently available sample sizes.

First draft submitted: 15 January 2018; Accepted for publication: 22 February 2018; Published online: 14 March 2018

Keywords: EGCUT • FINRISK • NMR metabolites • parent-of-origin effects • *PTPRD* • VLDL

Circulating metabolites are intermediate products of metabolic processes that are known to participate in a wide spectrum of biological pathways and consequently associated with many diseases and pathological conditions [1]. In this context, metabolites could be used as a system-wide connection from genetic variants to disease- or phenotypic outcome. Therefore, identification of genetic variants associated with these molecules is an important task of genome-wide association studies (GWAS) [2–6]. The advances in proton nuclear magnetic resonance (¹H-NMR or shortened as NMR) spectroscopy provide a method for quantifying of a wide range of high and low molecular weight compounds without any prior selection [7]. To date, the NHGRI-EBI Catalog of published GWAS reports approximately 1529 genetic variants in 58 studies associated with different kind of metabolic phenotypes measured from diverse range of different tissue samples [8]. The most comprehensive GWAS of NMR metabolites to date was performed by Kettunen *et al.* where profiles of metabolic measures were analyzed in up to 25,000 individuals and 62 associations with various metabolites were described [9].

Although GWAS have been more successful discovering genetic variants associated with human metabolites, there is still a large proportion of ‘hidden heritability’. As a source of genetic variance, several analyses have been proposed, including those focusing on rare variants, structural variants, gene–gene and gene–environment interactions and parent-of-origin effects (POEs) [10,11]. POEs reflect a combination of genetic and epigenetic mechanisms, modulating different complex traits associated mainly with embryonic growth and development [12] and these effects have been described with many complex diseases such as diabetes mellitus, disglycemia and

Future
Medicine 

retinoblastoma [13–15]. In case of POEs, not only the presence of an allele, but also its parental origin affects the phenotypic outcome [16]. As an example, maternal allele-specific association of a common variant (rs1367117) located in the *APOB* gene was demonstrated in the analysis of adiposity [17] and borderline significant effect for fasting glucose and insulin levels [17]. This nonsynonymous missense variant that cause a Thr > Ile amino acid change was also observed in non-POE association with serum ApoB levels [18].

Genomic imprinting is considered as the primary process that leads to POEs [19]. To date, approximately 100 genes are verified to be imprinted in humans and many more genes have been proposed for testing. Also, both imprinting and POEs can be tissue specific and depend on stage of the embryo development. Mott *et al.* demonstrated that the interaction between nonimprinted genes with imprinted ones can cause POEs on complex traits [20]. The role of genomic imprinting in human metabolic processes is relatively poorly understood, although convincing examples come from mice models where the process of imprinting is denoted as a key regulator of metabolism from infancy to adulthood [21]. Several imprinted genes in mice such as *Igf2r* and *Grb10* are known to influence energy homeostasis and have been proposed to coordinate glucose-regulated metabolism [22–24] while *Dlk1*, *Dio3* and *Gnas* are associated with thermogenesis in brown adipose tissue in mice progeny [25–29].

POE analysis are usually conducted using family trios and are often underpowered due to the lack of that kind of data. In contrast, huge amount of genotype data of unrelated individuals has become available in the era of biobanks. Hoggart *et al.* proposed a novel method for finding POEs in genotype data of unrelated individuals harboring SNPs at which such effects occur with increased variability to the phenotypic outcome [30]. This method is based on the assumption that POEs can be identified as increased variance of phenotype in heterozygotes group compared with homozygous individuals groups due to genetic effect differences among maternal and paternal mutation carriers. This approach was successfully tested on anthropometric phenotypes and two genes were identified having POE on BMI [30]. As POEs are often associated with metabolic diseases such as gestational diabetes [31] and metabolic syndrome [32], we aimed to investigate POEs of NMR metabolites measured in blood serum.

Materials & methods

Cohort description

The study sample contained 14,815 adult individuals of European descent from three population-based cohorts – one from Estonia (EGCUT) and two from Finland (DILGOM and FINRISK1997) – all being genotyped with a genome-wide genotyping array and imputed using 1000 Genomes Project multi-ethnic reference panel (Phase III, March 2012 release) (Supplementary Table 3).

The EGCUT (Estonian Biobank from Estonian Genome Center, University of Tartu) is a prospective, volunteer-based sample of the Estonian resident adult population (aged ≥ 18 years). The current number of participants is close to 52,000, which represents about 5% of the Estonian adult population. General practitioners and medical personnel in the special recruitment offices recruited adult volunteers throughout the country in the period of 2002–2010. At baseline, the GPs performed the standardized health examination of the participants, who also donated blood samples, where DNA, plasma, serum and white blood cells were immediately isolated [32].

The National FINRISK studies, held every 5 years, are population-based chronic disease risk factor monitoring surveys, conducted to monitor the health of the Finnish population among persons aged 25–74 at recruitment [33]. The participants are a random sample drawn from the population register, stratified according to geographical areas, sex and 10-year age groups to form a representative sample.

The FINRISK1997 (The National FINRISK 1997 Study) includes in total, 8444 individuals that were recruited to represent the population of five study areas across Finland. Standard clinical laboratory measurements were collected, and participants filled out questionnaires on physical activity and socioeconomic status [34].

The Dietary, Lifestyle and Genetic determinants of Obesity and the Metabolic syndrome (DILGOM, N = 5024) study was implemented within the framework of the National FINRISK 2007 Study. In 2007, all FINRISK health examinations participants were invited to a more detailed health examination (included in the DILGOM Study) concerning obesity and metabolic syndrome.

All participants from all three cohorts have given a written informed consent. The project have been approved by the Ethics Committee of Human Studies, University of Tartu, Estonia (approval 234/T-12), and the Ethics Committee of the National Public Health Institute, Helsinki, Finland and Ethics Committee of Helsinki and Uusimaa Hospital District (decision number 22/1/1997; 38/96; 558/E3/2001; 229/E0/06). All experiments in all cohorts were performed in accordance with relevant guidelines and regulations.

NMR biomarkers quantification, quality control & imputation

In total, 135 circulating metabolites were set to be quantified using proton NMR spectroscopy of native blood plasma in the Estonian cohort and serum in the Finnish cohorts. All cohorts were analyzed using the same high-throughput NMR metabolomics platform in the same analysis laboratory using a method described by Soininen *et al.* [35]. This methodology provided information on different NMR metabolites including lipoproteins, proteins, low molecular weight metabolites, such as amino acids, glycolysis precursors and other small molecules. All measured metabolites were already described here [9,36]. As a part of our NMR data quality control (QC), 23 phenotypes were discarded due to the missings more than 10% of measurements. NMR measurement outliers were defined as measurement values, which were deviating more than three standard deviations from the mean and these values were removed from the data. As the second step, imputation was performed to fill the missing NMR values (call rate = 0.9) for the remaining metabolites, using Multivariate Imputation by Chained Equations (MICE) method implemented into the R package MICE [37]. MICE imputation method required two data modifications performed before imputation: the individual values needed to be scaled to the same mean; in case of highly correlated phenotypes, one of them needed to be removed. The pairwise correlation threshold was set to 0.995, and 30 phenotypes were removed as a result, leaving us 82 phenotypes in total for the further analyses. As the final step, residuals for all 82 metabolites were calculated adjusting for sex, age, age², BMI and 10 principal components, calculated based on the genetic data, to correct for population stratification (Supplementary Table 4).

Genotyping, quality control & imputation

All cohorts were genotyped using Illumina genotyping arrays: Illumina Human370CNV and Illumina OMNI-Express for the EGCUT data and Illumina Infinium HumanHap 610K and Illumina OMNI-Express for the DILGOM and FINRISK1997 cohorts, respectively. Sample quality control covered exclusion on the basis of genome-wide call rate, minor allele frequency, extreme heterozygosity, sex discordance, cryptic relatedness and ethnical outliers (Supplementary Table 3). SNP quality control was done on the basis of call rate across samples and deviation from Hardy–Weinberg equilibrium. Before the imputation, nonautosomal SNPs, SNPs with minor allele frequency <1% and palindromic SNPs were removed. All studies were imputed using 1000 Genomes Project multi-ethnic reference panel (Phase III, March 2012 release) using IMPUTEv2 software [38].

POE analysis

The POE analyses in all three population-based cohorts were performed using QUICKTEST software [30]. This software implements a method for POE analysis using genetic data of unrelated individuals, based on comparing the phenotypic variance of the heterozygous genotype group to the variance observed in the homozygous groups [30]. The method is based on the assumption that increased variance in the heterozygous group in unrelated individuals arises because that group contains two subgroups (paternal reference allele/maternal alternative allele and maternal reference allele/paternal alternative allele) each with different means. The POE test utilize the Brown–Forsythe test, modified to test the mean absolute deviation from the median in the heterozygous and homozygous groups [30] (Supplementary Figure 2). The POE test was performed on 5861 individuals in the EGCUT, 516 in the DILGOM and 8438 individuals in the FINRISK1997 cohort.

Meta-analysis

Meta-analysis for each analyzed 82 NMR metabolites were performed in up to 14,815 participants from 3 studies. Before meta-analysis, SNPs with minor allele frequency <1% and poorly imputed variants (IMPUTE info <0.8) were excluded from further analysis. Summary statistics for autosomal SNPs that passed quality control for all studies were combined in inverse-variance weighted fixed effect meta-analysis for each trait, as implemented in GWAMA [39].

Principal component analysis

The significance level threshold was calculated using principal component analysis. As NMR data are highly intercorrelated, simple Bonferroni correction would be too conservative, and therefore we used principal component analysis to estimate the number of independent phenotypes [40] using the R package Prcomp [41]. In total, we found 26 principal components describing more than 99% of the variability in the NMR data. Therefore, we used genome-wide threshold divided by the number of independent phenotypes ($5 \times 10^{-8}/26 = 1.92 \times 10^{-9}$) as the significance threshold corrected for the number of independent tests.

Table 1. Parent-of-origin meta-analysis results for circulating metabolites identified from three cohorts.

SNP	Chr	Position (b37)	Reference allele	Other allele	MAF	N (AA + BB)	N (AB)	POE p-value	Gene	Metabolite
rs1412727	9	9423523	A	G	0.17	9392	5423	8.01×10^{-11}	<i>PTPRD</i>	M-VLDL-TG
rs10273950	7	70664155	G	T	0.02	13,945	870	7.37×10^{-9}	<i>WBSCR17</i>	Ile
rs189028844	10	54062144	A	T	0.01	13,302	390	9.11×10^{-9}	<i>PRKG1-AS1</i>	IDL-TG
rs56406361	8	30371938	T	A	0.24	9942	4873	9.62×10^{-9}	<i>RBPMS</i>	AcAce
rs662136	7	54621526	T	C	0.34	8442	6371	1.27×10^{-8}	<i>VSTM2A</i>	IDL-P
rs554447888	4	129769586	G	A	0.01	12,856	314	1.53×10^{-8}	<i>PHF17</i>	Pyr
rs196581	7	36315219	T	C	0.15	9997	4818	2.76×10^{-8}	<i>EEPD1</i>	VLDL-D
rs10010912	4	25011501	C	A	0.03	13,563	1252	3.03×10^{-8}	<i>LG12</i>	S-HDL-L
rs196581	7	36315219	T	C	0.15	9997	4818	3.16×10^{-8}	<i>EEPD1</i>	L-VLDL-FC
rs12244405	10	52643179	C	T	0.28	9333	2083	3.59×10^{-8}	<i>A1CF</i>	L-LDL-FC
rs4742595	9	9420915	T	C	0.17	9547	5268	4.15×10^{-8}	<i>PTPRD</i>	M-VLDL-FC

AcAce: Acetoacetate; Chr: Chromosome; Gene: Nearest gene; HDL: High-density lipoprotein; IDL: Intermediate-density lipoprotein; IDL-TG: Triglyceride in IDL; IDL-P: Concentration of IDL particle; Ile: Isoleucine; LDL: Low-density lipoprotein; L-LDL-FC: Free cholesterol in large LDL; L-VLDL-FC: Free cholesterol in large VLDL; MAF: Minor allele frequency; M-VLDL-FC: Free cholesterol in medium VLDL; M-VLDL-TG: Triglycerides in medium size VLDL; N (AA + BB): Homozygous sample size; N (AB): Heterozygous sample size; POE p-value: The Brown-Forsythe test p-value; Pyr: Pyruvate; S-HDL-L: Total amount of lipids in small HDL; VLDL-D: Mean diameter for VLDL particle; VLDL: Very low-density lipoprotein.

Results

We performed a genome-wide analysis to determine plausible POEs for the NMR metabolites. Genotype data of 9,362,347 SNPs was available in up to 14,815 individuals of European descent from three cohorts: one from Estonia (EGCUT) and two from Finland (DILGOM and FINRISK1997). We found 9 loci, with possible POEs, affecting levels of 11 metabolites (Table 1). Nine associations identified in the initial POE analysis reached the genome-wide significant threshold ($p \leq 5 \times 10^{-8}$). Principal components analysis was used to calculate the genome-wide p-value threshold for the number of independent phenotypes analyzed, resulting in the new threshold $p \leq 1.92 \times 10^{-9}$ (see methods for the multiple correction calculation). After multiple testing correction only the variant rs1412727 remained significant (MAF = 0.17; $p = 8.29 \times 10^{-11}$) located in the intron of the *PTPRD* (protein tyrosine phosphatase, receptor-type, D) gene and associated with metabolites called triglycerides in medium-size very low-density lipoprotein (VLDL) (Supplementary Figure 1). Also no heterogeneity was observed for this variant among all cohorts ($i^2 = 0$; HetP = 0.836304; Cochran's Q statistic = 0.854866).

In addition to the genome-wide analysis, we selected previously established associations from the study by Kettunen *et al.* [9], where meta-analysis of 123 metabolites in 24,925 individuals revealed 62 associations with NMR metabolite levels, and tested if any of these associations show POEs. In total, five loci had p-values less than 0.05 (Supplementary Table 1), but none of them reached the threshold considered for significance after multiple testing correction ($p_{\text{corrected}} < 0.001$).

We also analyzed a variant rs1367117, which has been reported to have POE on glucose level [6,42]. No significant association between the level of glucose from NMR metabolite panel and this variant was detected in our data.

Discussion

We performed a genome-wide analysis to find POEs of common variants for NMR metabolites in humans. We identified 11 signals that reached genome-wide significance threshold, but only 1 variant remained significant after the multiple testing correction considering the number of phenotypes analyzed. We assume that even though NMR metabolites have a very strong genetic component, POEs in NMR metabolites are small and require larger sample sizes to be detected.

We report a novel plausible POE association with the genetic variant rs1412727 that affects the level of the metabolite triglycerides in medium-size VLDL particle. This variant is a nonsynonymous intronic variant located in *PTPRD* gene, and thus we report for this gene as our main candidate. Measurement of associated molecule represents the number of an endogenously secreted triglycerides into the core of VLDL particles [35]. After being produced in liver, the VLDLs then circulate in the bloodstream to go to the peripheral adipose and muscle tissue, where they are further broken down by lipoprotein lipases releasing triglycerides that can be used to supply fatty acids and glycerol to the cell as a source of energy [43,44]. Several studies have provided evidence that the level of lipoproteins associated with triglycerides is causally related to the development of coronary heart disease [45–47].

PTPRD gene itself is already established gene from GWAS and known to be associated with several traits and diseases, such as Type 2 diabetes [48], age at menarche [49], epilepsy [50] and obsessive-compulsive disorder [51]. To date, it is known that the protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family that has a phosphorylation of tyrosine residues activity. This is the key regulatory mechanism important for cellular signaling that affects cellular events including cell metabolism, proliferation, adhesion, differentiation, migration and development [52–54]. Although changing the phosphorylation activity of proteins is among the most well-studied mechanisms and well-used modification methods for studying the regulation of protein structure and function in cells and organisms, the function of *PTPRD* gene is yet poorly understood. The mechanism of this gene could be similar to another gene that belongs to the same PTP family – *PTPIB* gene is implicated as a negative regulation factor for fat triglycerides [55]. Briefly, expression reduction of *PTPIB* results in a decrease of fat deposition and triglyceride levels by downregulating *SREBP1*, followed by a further downregulation of *SPOT14*, *FAS*, *LPL* and *PPAR* – important genes known to be involved in lipogenesis and fatty acid synthesis [55]. Moreover, *PTPRD* gene has been demonstrated in association with weight gain caused by antipsychotic medications, which are used for patients with bipolar disorder and schizophrenia. The antipsychotic-induced weight gain phenotype may be involved in multiple pathways related to metabolic process. Thus, we believe that identification of the POE in the *PTPRD* gene could help in uncovering etiology of related diseases.

Discovering of POEs can help to reveal some of the hidden heritability. It is worth noting that it is not only important to find novel associations on the genotype–phenotype level, but also to consider already established variants for having more complex effects on phenotypes, which could be missed while testing association with using only additive genetic models. Therefore, we also investigated whether POEs can also be found for the genetic loci that have been previously associated with NMR metabolites by Kettunen *et al.* [9]. Using the genetic additive model and a smaller dataset, we demonstrated the same associations with circulating metabolites while we did not detect any POEs. Also, we were not able to confirm POE of a previously established signal on the level of fasting glucose demonstrated by Hochner *et al.* located in the *APOB* gene [17]. Although our study has revealed the POE on human circulating metabolites, future studies are required to provide replication and deeper understanding about this regulation in brain tissue of adult organisms.

Conclusion

POEs that include the process of genomic imprinting are important phenomena that are found for genes associated with the range of metabolic traits, such as Type 2 diabetes, menarche, metabolic syndrome and others. We, therefore, aimed to identify the POE on human circulating metabolites that serve as an intermediate trait between genetic variants and pathogenic outcomes. We performed genome-wide scan for POEs and also investigated them in already established genetic loci associated with NMR metabolites. We were not able to detect POEs for variants described in previously published comprehensive GWAS. On the other hand, the genome-wide POE analysis revealed the association of *PTPRD* gene with metabolite called triglycerides in medium VLDL molecule. In conclusion, we believe that POEs can be found for metabolic measurements, although expected effect is probably small.

Summary points

- Parent-of-origin effects (POEs) are present for several metabolic traits, for example, Type 2 diabetes, metabolic syndrome, BMI, menarche, gestational diabetes and many others.
- We used the modified Brown–Forsythe test to detect POEs in population-based data.
- We conducted genome-wide association POE analysis for 82 metabolites in 14,815 individuals of European ancestry.
- The analysis did not reveal POE in genetic variants previously associated with nuclear magnetic resonance metabolites.
- Genome-wide association POE analysis demonstrated association of *PTPRD* gene with metabolite called triglycerides in medium very low-density lipoprotein molecule.
- Variants in *PTPRD* gene region has been previously associated with epilepsy, age at menarche, Type 2 diabetes and obsessive-compulsive disorder.
- The function of *PTPRD* gene is yet poorly understood, although it was suggested to be implicated as a negative regulation factor for fat triglycerides.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at:
www.futuremedicine.com/doi/suppl/10.2217/bmm-2018-0020

Acknowledgements

The authors would like to acknowledge High Performance Computing Center of University of Tartu.

Conflict of interest

The authors declare no conflict of interest.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Authors' contributions

N Pervjakova and R Mägi conceived and designed the experiments. N Pervjakova, V Kukushkina, T Haller and S Kasela performed the experiments. M Perola, K Kristiansson, A Joensuu, A Metspalu, T Annilo, P Jousilahti and V Salomaa contributed data. N Pervjakova wrote the paper. R Mägi supervised the work.

Financial & competing interests disclosure

This study was supported by EU H2020 grant numbers 692145, 676550, 654248, Estonian Research Council Grant IUT20–60 and PerMed I, NIASC, EIT – Health and European Union through the European Regional Development Fund (Project Number: 2014–2020.4.01.15–0012 GENTRANSMED). K Kristiansson was supported by Academy of Finland grant number 250207. DILGOM NMR metabolomics were funded by the Yrjö Johansson Foundation. DILGOM GWAS data was funded by the Finnish Academy grant number 269517. V Kukushkina was supported by the Finnish Foundation for Cardiovascular Research. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Holmes E, Loo RL, Stampler J *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453(7193), 396–400 (2008).
2. Gieger C, Geistlinger L, Altmaier E *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 4(11), e1000282 (2008).
3. Hicks AA, Pramstaller PP, Johansson *et al.* Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* 5(10), e1000672 (2009).
4. Illig T, Gieger C, Zhai G *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* 42(2), 137–141 (2010).
5. Demirkan A, van Duijn CM, Ugozai P *et al.* Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet.* 8(2), e1002490 (2012).
6. Teslovich TM, Musunuru K, Smith AV *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707–713 (2010).
7. Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat. Rev. Genet.* 13(11), 759–769 (2012).
- **Discusses the study design considerations and applications to unravel the genetic contribution to human metabolic phenotypes.**
8. Welter D, MacArthur J, Morales J *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006 (2014).
9. Kettunen J, Demirkan A, Würtz P *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* 7, 11122 (2016).

- **An extended genome-wide association study of genetic influences on 123 circulating metabolic traits from almost 25,000 individuals of European ancestry.**
- 10. Manolio TA, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. *Nature* 461(7265), 747–753 (2009).
- 11. Eichler EE, Flint J, Gibson G *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11(6), 446–450 (2010).
- 12. Hager R, Johnstone RA. The influence of phenotypic and genetic effects on maternal provisioning and offspring weight gain in mice. *Biol. Lett.* 2(1), 81–84 (2006).
- 13. Morris AP, Voight BF, Teslovich TM *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of Type 2 diabetes. *Nat. Genet.* 44(9), 981–990 (2012).
- **Classical example of genome-wide association studies performed on continuous phenotype.**
- 14. Kobaly K, Vellanki P, Sisk RK *et al.* Parent-of-origin effects on glucose homeostasis in polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* 99(8), 2961–2966 (2014).
- 15. Eloy P, Dehainault C, Sefta M *et al.* A parent-of-origin effect impacts the phenotype in low penetrance retinoblastoma families segregating the c.1981C>T/p.Arg661Trp mutation of RB1. *PLoS Genet.* 12(2), 1–14 (2016).
- 16. Wolf JB, Hager R, Cheverud JM. Genomic imprinting effects on complex traits. *Epigenetics* 3, 295–299 (2008).
- 17. Hochner H, Allard C, Granot-Hershkovitz E *et al.* Parent-of-origin effects of the *APOB* gene on adiposity in young adults. *PLoS Genet.* 11(10), 1–23 (2015).
- **Demonstrates a maternal-specific association between a common *APOB* gene variant and adiposity.**
- 18. Haas BE, Weissglas-Volkov D, Aguilar-Salinas CA *et al.* Evidence of how rs7575840 influences apolipoprotein B-containing lipid particles. *Arterioscler. Thromb. Vasc. Biol.* 31(5), 1201–1207 (2011).
- 19. Surani MAH, Barton SC. Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science* 222(4627), 1034–1036 (1983).
- 20. Mott R, Yuan W, Kaisaki P *et al.* The architecture of parent-of-origin effects in mice. *Cell* 156(1–2), 332–342 (2014).
- 21. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. *Nat. Rev. Genet.* 14(9), 609–617 (2013).
- **Describes different parent-of-origin effects in humans and mice.**
- 22. Woods SC, Seeley RJ, Porte D Jr, Schwartz MW. Signals that regulate food intake and energy homeostasis. *Science* 280(5368), 1378–1383 (1998).
- 23. Chanprasertyothin S, Jongjaroenprasert W, Ongphiphadhanakul B. The association of soluble *IGF2R* and *IGF2R* gene polymorphism with Type 2 diabetes. *J. Diabetes Res.* 2015, 216383 (2015).
- 24. Liu B, Liu F. Feedback regulation of mTORC1 by Grb10 in metabolism and beyond. *Cell Cycle* 13(17), 2643–2644 (2014).
- 25. Ishida M, Moore GE. The role of imprinted genes in humans. *Mol. Aspects Med.* 34(4), 826–840 (2013).
- 26. Xie T, Chen M, Gavrilova O, Lai EW, Liu J, Weinstein LS. Severe obesity and insulin resistance due to deletion of the maternal *Gsα* allele is reversed by paternal deletion of the *Gsα* imprint control region. *Endocrinology* 149(5), 2443–2450 (2008).
- 27. Yu S, Gavrilova O, Chen H *et al.* Paternal versus maternal transmission of a stimulatory G-protein α subunit knockout produces opposite effects on energy metabolism. *J. Clin. Invest.* 105(5), 615–623 (2000).
- 28. Haig D. Huddling: brown fat, genomic imprinting and the warm inner glow. *Curr. Biol.* 18(4), R172–R174 (2008).
- 29. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat. Rev. Genet.* 15(8), 517–530 (2014).
- 30. Hoggart CJ, Venturini G, Mangino M *et al.* Novel approach identifies SNPs in *SLC2A10* and *KCNK9* with evidence for parent-of-origin effect on body mass index. *PLoS Genet.* 10(7), 1–12 (2014).
- 31. Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* 115(3), e290–e296 (2005).
- 32. Leitsalu L, Haller T, Esko T *et al.* Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* 44(4), 1137–1147 (2015).
- 33. Borodulin K, Vartiainen E, Peltonen M *et al.* Forty-year trends in cardiovascular risk factors in Finland. *Eur. J. Public Health* 25(3), 539–546 (2014).
- 34. Vartiainen E, Laatikainen T, Peltonen M *et al.* Thirty-five-year trends in cardiovascular risk factors in Finland. *Int. J. Epidemiol.* 39(2), 504–518 (2010).
- 35. Soininen P, Kangas AJ, Wurtz P *et al.* High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism. *Analyst* 134(9), 1781–1785 (2009).
- **Demonstrates an update protocol for human metabolite measurements using nuclear magnetic resonance.**
- 36. Auro K, Joensuu A, Fischer K *et al.* A metabolic view on menopause and ageing. *Nat. Commun.* 5, 4708 (2014).

37. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* 20(1), 40–49 (2011).
38. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5(6), e1000529 (2009).
39. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11, 288 (2010).
40. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32(4), 361–369 (2008).
41. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* 9(4), e93766 (2014).
42. Chasman DI, Pare G, Mora S *et al.* Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet.* 5(11), e1000730 (2009).
43. Chapman MJ, Ginsberg HN, Amarenco P *et al.* Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: evidence and guidance for management. *Eur. Heart J.* 32(11), 1345–1361 (2011).
44. Toth PP. Triglyceride-rich lipoproteins as a causal factor for cardiovascular disease. *Vasc. Health Risk Manag.* 12, 171–183 (2016).
45. Do R, Willer CJ, Schmidt EM, Al E. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* 45(11), 1345–1352 (2013).
46. Abbasi F, McLaughlin T, Lamendola C *et al.* High carbohydrate diets, triglyceride-rich lipoproteins, and coronary heart disease risk. *Am. J. Cardiol.* 85(1), 45–48 (2000).
47. Rosenson RS, Davidson MH, Hirsh BJ, Kathiresan S, Gaudet D. Genetics and causality of triglyceride-rich lipoproteins in atherosclerotic cardiovascular disease. *J. Am. Coll. Cardiol.* 64(23), 2525–2540 (2014).
48. Tsai FJ, Yang CF, Chen CC *et al.* A genome-wide association study identifies susceptibility variants for Type 2 diabetes in Han Chinese. *PLoS Genet.* 6(2), e1000847 (2010).
49. Perry J, Day F, Elks CE *et al.* Parent-of-origin specific allelic associations among 106 genomic loci for age at menarche. *Nature* 514(7520), 92–97 (2014).
50. Speed D, Hoggart C, Petrovski S *et al.* A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. *Hum. Mol. Genet.* 23(1), 247–258 (2014).
51. Mattheisen M, Samuels JF, Wang Y *et al.* Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. *Mol. Psychiatry* 20(3), 337–344 (2015).
52. Hunter T. A thousand and one protein kinases. *Cell* 50(6), 823–829 (1987).
53. Mustelin T, Taskén K. Positive and negative regulation of T-cell activation through kinases and phosphatases. *Biochem. J.* 371(Pt 1), 15–27 (2003).
54. Xu Y, Fisher GJ. Receptor type protein tyrosine phosphatases (RPTPs) – roles in signal transduction and human disease. *J. Cell Commun. Signal.* 6(3), 125–138 (2012).
55. Rondinone CM, Trevillyan JM, Clampit J *et al.* Protein tyrosine phosphatase 1B reduction regulates adiposity and expression of genes involved in lipogenesis. *Diabetes* 51(8), 2405–2411 (2002).