Published in final edited form as:

Twin Res Hum Genet. 2015 February ; 18(1): 61–72. doi:10.1017/thg.2014.74.

A twin study of breastfeeding with a preliminary genome wide association scan

L. Colodro-Conde^{1,2}, G. Zhu¹, R. A. Power³, A. Henders¹, A.C. Heath⁴, P.A.F. Madden⁴, G.W. Montgomery¹, S. E. Medland¹, J.R. Ordoñana², and N.G. Martin¹

¹QIMR Berghofer Medical Research Institute, 300 Herston Rd, Brisbane Queensland 4006, Australia

²Murcia Twin Registry, Department of Human Anatomy and Psychobiology, University of Murcia, Espinardo Campus, 30100 Murcia, Spain

³MRC Social Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, DeCrespigny Park, Denmark Hill, London, United Kingdom, SE5 8AF

⁴Department of Psychiatry, Washington University School of Medicine. 660 S. Euclid Ave., St. Louis, MO 63110, USA

Abstract

Breastfeeding has been an important survival trait during human history, though it has long been recognised that individuals differ in their exact breastfeeding behaviour. Here our aims were, first, to explore to what extent genetic and environmental influences contributed to the individual differences in breastfeeding behaviour; second, to detect possible genetic variants related to breastfeeding; and lastly, to test if the genetic variants associated with breastfeeding have been previously found to be related with breast size. Data were collected from a large community-based cohort of Australian twins, with 3,364 women for the twin modelling analyses and 1,521 of them included in the genome wide association study. Monozygotic twin correlations ($r_{MZ} = .52$, 95% *CI* .46 – .57) were larger than dizygotic twin correlations ($r_{DZ} = .35$, 95% *CI* .25 – .43) and the best-fitting model was the one composed by additive genetics and unique environmental factors, explaining 53% and 47% of the variance in breastfeeding behaviour, respectively. No breastfeeding-related genetic variants reached genome-wide significance. The polygenic risk score analyses showed no significant results, suggesting breast size does not influence breastfeeding. This study confers a replication of a previous one exploring the sources of variance of breastfeeding and, to our knowledge, is the first one to conduct a Genome-Wide Association Study on breastfeeding and look at the overlap with variants for breast size.

Keywords

breastfeeding; twin study; genome wide association study

Corresponding author: Lucia Colodro Conde. (T) +34 868 884262. (F) +34 868 884111. (E) lucia.c.c@um.es.

Conflict of interests

None

Breast milk, either from an infant's own mother or from another woman, can be assumed to have been the sole source of nourishment for infants during most of human evolution, having evolved as an adaptation to transfer immune factors to offspring and to space births (Sellen, 2007). It has been suggested that breastfeeding duration in prehistoric times was of two to three years and that, from the Middle Ages to the 19th century, most infants in Europe were typically weaned between the ages of one and two (Schön & Silvé, 2007). It was not until the end of the 19th century, that is, in our recent history, that artificial feeding became a safe alternative to breast milk, and breastfeeding became dependant on the mother's choice in most milieus. Consequently, nowadays a large variability in this behaviour can be observed both within and between countries, with global statistics showing that 37% of infants were exclusively breastfed until 6 months age in 2007 (World Health Organization, 2013), which is the current recommended minimum weaning age (World Health Organization & United Nations Children's Fund, 2003). This percentage varied considerably across countries, for example 18.8% in the U.S in 2011 (Centers for Disease Control and & Prevention, 2014), 17.6% in Australia in 2011–2012 (Australian Bureau of Statistics, 2013) and 28.5% in Spain in 2011–2012 (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2012). The instruments for data collection vary between countries in, for example, being explicit about the inclusion of expressed milk in exclusive breastfeeding, so some rates could be higher.

Given the positive impact of breastfeeding on infants and mothers' health (Ip et al., 2009; World Health Organization & United Nations Children's Fund, 2003), there is a current interest in determining which factors are related to women's decisions regarding the method of feeding their infants, i.e., the predispositions or situations that make them more likely to initiate breastfeeding or bottle feeding and the duration of breastfeeding. Breastfeeding is a biocultural behaviour (Holman & Grimes, 2003) and is influenced by multiple factors (Thulier & Mercer, 2009), such as the action of prolactin and oxytocin (Heinig & Dewey, 1997), pain from nipple trauma, mastitis, maternal exhaustion and perceived poor milk supply (Almqvist-Tangen et al., 2012)(Spencer, 2008), babies' sucking dynamics (Sakalidis et al., 2013), body mass index (Wojcicki, 2011), level of education (Colodro-Conde et al., 2011), working conditions (Calnen, 2010), psychological factors, personality, self-efficacy, anxiety (Brown, 2014; Colodro-Conde et al., in press (a); Colodro-Conde et al., in press (b); De Jager et al., 2012; Li et al., 2008; McFadden & Toole, 2006; Wagner et al., 2006), support from partner, family and peers (Thulier & Mercer, 2009), social norms (Swanson & Power, 2005), or advice of health professionals (Brown, Raynor, & Lee, 2011).

The magnitude of the genetic and environmental sources of variance in breastfeeding behaviour were explored in a previous study in a cohort of Spanish twins from the Murcia Twin Registry (Ordoñana et al., 2006, 2013). The best fitting models explained the observed variance through additive genetic and non-shared environmental factors for initiation and duration in the first born child and the average for the complete offspring, with heritability ranging between 44% and 54% (Colodro-Conde et al., 2013). It was hypothesised that among the genetic factors influencing breastfeeding behaviour, those related to the regulation of hormone production could play a part in breastfeeding outcomes. In support of this hypothesis, Jonas et al. (2013) found that the single nucleotide polymorphism (SNP) rs2740210 in the oxytocin peptide gene was associated with exclusive breastfeeding at 3 and

6 months postpartum. Other authors, however, did not find a relationship between selected oxytocin receptor SNPs and breastfeeding (Tharner et al., 2012). As breastfeeding has been suggested to be significantly influenced by genetic factors, it would be interesting to explore the genetic variants underlying this trait.

Several studies have concluded that body mass index is associated with breastfeeding behaviour, so mothers who are overweight or obese are less likely to initiate lactation, and are prone to early cessation of breastfeeding (Jevitt et al., 2007; Wojcicki, 2011). A study with health professionals showed that from their experience, initiating was more difficult than continuing breastfeeding for women with large breasts and/or obesity (Katz et al., 2010). The link between overweight, elevated serum testosterone concentration and pathologies such as the polycystic ovary syndrome could be underlying this relationship (Balen et al., 1995; Barber, McCarthy, Wass, & Franks, 2006). Heritability of breast cup size has been estimated to be 56%, and one third of this variance was shared with body mass index (Wade et al., 2010). Some genetic variants associated with breast size also influence breast cancer risk (Eriksson et al., 2012).

This paper has three main objectives. First, it aims to replicate previous findings related to heritability of breastfeeding by exploring the magnitude of the genetic and environmental influences in the variation of breastfeeding behaviour in a sample of Australian twins. For that aim, we used a twin design, calculating twin correlations for breastfeeding and fitting an ACE variance components model. Second, in order to detect possible genetic variants related to breastfeeding behaviour, we conducted a Genome-Wide Association Study (GWAS) of this phenotype, using data collected from this large community-based cohort of Australian twins. Lastly, we report the results of a polygenic risk score analysis examining whether an individual's number of genetic variants predisposing to breast size are associated with breastfeeding.

Materials and methods

Participants

Participants in this study were 3,364 female twin mothers from the QIMR health and life style studies Cohorts I and II with data on breastfeeding. Women in Cohort I were born between 1892 and 1963 ($n = 3,205$) and women in Cohort II were born between 1964 and 1971 ($n = 419$). At the time of the survey (1988-93), the mean age was 43.41 years ($SD = 12.32$, range = 25–86 years), for women in Cohort I and 24.81 ($SD = 1.87$, range = 19–29 years) for women in Cohort II. Further details of the sample, data collection and zygosity determination are described elsewhere for cohort I (Heath et al., 1997) and for cohort II (Knopik et al., 2004).

The sample comprised 992 complete twin pairs: 629 monozygotic (MZ) and 363 dizygotic (DZ) and 1,380 individual twins from incomplete pairs (411 MZ and 347 DZ from female-female pairs, 621 from female-male pairs and 5 of unknown zygosity). Genome wide genotypic data were available for 1,521 of the 3,364 individuals.

Procedure

In 1988–1993, participants of both cohorts completed a mailed questionnaire including information about their childbearing. Some of these who did not complete the mailed questionnaire were interviewed over the telephone in 1989–1992. A small subsample of participants of Cohort I ($n = 341$) was re-surveyed two years after their initial contact to establish the test-retest reliability of the interview measures.

As part of follow up studies, blood samples were collected from the majority of participants in Cohorts I and II. DNA samples were genotyped using the Illumina 317, 370 and 610 SNP chips. Following quality control, the data were imputed using the subset of $\sim 281,000$ markers which were available across all chips. The genotypes were phased using MACH, then imputed (including the X chromosome), using Minimac based on the 283 European reference individuals in the March 2013 release of phased haplotype from the 1000 Genomes project. In the analyses presented here, we use data from 6.59 M imputed SNPs that passed quality control ($R^2 > 0.3$ and minor allele frequency $> 1\%$) for association analysis. Details of SNP typing, quality control, data cleaning (including ancestry exclusion) and analysis are given in detail elsewhere (Medland et al., 2009).

This study was approved by the Queensland Institute of Medical Research Human Research Ethics Committee and the storage of the data follows national regulations regarding personal data protection. All of the participants provided informed consent.

Measures

Data were based on retrospective self-reports. All the participants were asked about details of the birth of up to eight children, including birth complications, time of birth, birth weight, hours of labour, feelings of depression and type of feeding in the first months (see Appendix 1). The variable we analysed was the mean number of months they breastfed each child, averaged across all live births and standardised to a Z-score.

Data analysis

Twin modelling—Data preparation and descriptive and preliminary analyses were performed in SPSS v.19 (SPSS, 2010). Assumptions of the twin design were checked, including the homogeneity of the means and variances of first and second born twins and across zygosity groups. Further details of the twin design, including checking at assumptions, can be found elsewhere (Neale & Cardon, 1992; Posthuma et al., 2003).

Statistical analyses employed full information maximum-likelihood modelling (FIML) procedures using the statistical package Mx (Neale et al., 2006). In FIML, both complete and incomplete pairs of twins can be used in the analyses and the goodness-of-fit of a model to the observed data is distributed as chi-square (χ^2). By testing the change in chi-square (χ^2) against the change in degrees of freedom (df), we can test whether dropping or equating specific model parameters significantly worsens the model fit. The best-fitting model was chosen in each case by deducting the residual deviance of the compared models and by comparing Akaike's information criterion (AIC).

In a first step, we determined twin pair correlations per zygosity group for breastfeeding. We tested whether MZ correlations were higher than those of DZ twin pairs, which would suggest a genetic influence on individual differences in this trait. Then, variance component model-fitting was conducted to partition the variation in breastfeeding into genetic, shared environmental and non-shared environmental influences. Observed MZ and DZ twin correlations generally reflect a combination of A, C, D, and E influences, and structural equation modelling determines the combination that best matches the observed data (Posthuma et al., 2003). Age at the survey time was modelled as covariate in all the analyses.

SNP-based genome-wide association study—1000 Genome imputed dosage data were tested for association with breastfeeding using the additive family-based association test in Merlin-off line (Chen & Abecasis, 2007). Correction for age at the survey time was performed by fitting the covariate in the regression model. We adopted a genome wide significance level for the association between SNP and phenotype of 5×10^{-8} or smaller to correct for the total number of independent tests (Dudbridge & Gusnanto, 2008).

Gene-based analysis—A gene-based test Versatile Gene-based Association Study (VEGAS), designed for use with GWAS data with related individuals (Liu et al., 2010), was conducted to determine the level of association across the gene while correcting for linkage disequilibrium (LD) and gene size. Details of this procedure are summarized elsewhere (Verweij et al., 2010). In brief, this test explores association of each gene (including 50Kb up and down of the coding region) taking into account the p -values of all SNPs (after first pruning out those in high LD) ($r^2 > .20$), and the LD between them. A p -value below $\alpha = 2.8 \times 10^{-6}$ was considered to be genome wide significant as the gene-based association test included 17,585 genes (0.05/17,585 genes).

Polygenic study—To examine a potential shared genetic a etiology between breastfeeding and breast size, we tested the effects of SNPs previously associated with breast size through a polygenic risk scoring approach. This took the top 2,079 SNPs associated with breast size in a previous GWAS in an independent sample (Eriksson et al., 2012) at a p -value of .0001. These were then clumped within PLINK (Purcell et al., 2007) to correct for LD within the QIMR sample ($R^2 < .2$ in 250kb windows), identifying 66 overlapping independent SNPs associated with breast size. A polygenic score was generated using PLINK weighting on the beta for each SNP. This score was then tested for association with breastfeeding phenotypes, correcting for covariates of five principal components. Variance explained by this score was derived as that of a model including the polygenic score and covariates minus the variance explained by a model including only covariates as predictors of breastfeeding.

Results

Descriptive Results

As stated before, our sample is entirely comprised of mothers who reported data about breastfeeding. The mean number of children was 2.52 ($SD = 1.26$, Range = 1–12). Three out

of four (75.2%) women breastfed for at least one month in each one of their births. The mean duration of breastfeeding was 5.31 months ($SD = 4.69$, range: 0–48), with slight changes according to child order (Table 1). Table 2 presents the correlation matrix for the duration of breastfeeding for the first five children. Breastfeeding reports showed a high test-retest reliability ($r = .96$, $p < .001$) and internal consistency among breastfeeding durations (Cronbach's $\alpha = .98$; $ICC = .85$).

Twin Correlations and Variance component model fitting

The MZ twin correlation for the mean breastfeeding duration ($r_{MZ} = .52$, 95% $CI .46 - .57$) was larger than for DZ twins ($r_{DZ} = .35$, 95% $CI .25 - .43$), which suggests the presence of genetic influences (see Figure 1). Lower correlations were found for the breastfeeding duration of the first-born child ($r_{MZ} = .41$, 95% $CI .34 - .47$; $r_{DZ} = .27$, 95% $CI .17 - .37$).

A univariate model was fit to disentangle the sources of variance of the mean breastfeeding duration (see Table 3). Since the DZ twin correlation was more than half the MZ twin correlation, C was estimated instead of D. The best-fitting model was the one that included additive genetic and non-shared environmental sources of variation –shared environmental factors did not account for a significant portion of variation. Additive genetic factors accounted for 53% in breastfeeding (95% $CI 47\% - 58\%$) and the remaining 47% was due to unique environmental factors (95% $CI 42\% - 53\%$). The reduced model fit well, as it did not differ significantly from the full ACE model. The model fitting analysis revealed a similar structure of the underlying variance for the breastfeeding duration in the first-born child, though with lower heritability (.42 vs. .53 for average duration).

Genome-wide association study

We conducted a genome-wide association study of breastfeeding in 1,521 individuals from 1,073 Australian families. The average age of the genotyped sample was 46.46 years ($SD = 11.38$). Breastfeeding mean duration in this subsample was 5.52 months ($SD = 4.75$). We tested 6.59 M SNPs for association with breastfeeding, correcting for age at the time of survey.

The Manhattan plot of association p -values for 6.59 M SNPs is shown in Figure 2. While no SNP achieved genome-wide significance ($p < 6.6 \times 10^{-8}$), regions of suggestive association signals, with the smallest p -value of 1.2×10^{-7} obtained for a SNP (rs6950451) were observed on chromosome 7. Suggestive association signals were also detected in chromosomes 2 (SNP rs930421) and 18 (rs9807759), with p -values of 1.2×10^{-6} in both cases.

The quantile–quantile plot ($Q-Q$ plot) of the observed versus expected (under the null-hypothesis of no association) $\log_{10}$ (p -value) from the association analysis is presented in Figure 3. The genomic control λ (1.007) was close to 1.0, indicating that there was no evidence for inflation of the test statistics or a bias because of possible population stratification in the results (Bacanu et al., 2000) and that the family-based association model had correctly accounted for relatedness.

The top SNPs and gene regions from GWAS analysis with the strongest association with breastfeeding are shown in Table 4.

Figure 4 shows the locus zoom plots of the unconditioned analyses for the top five SNPs, rs6950451, rs930451, rs9807759, rs17381960, rs7446359, and also for rs498793, since there are genes at or nearby this latter SNP which have been emphasized for their role in the composition of fatty acids in mothers' milk (Standl et al., 2012; Glaser et al., 2011). Notably, although LD structure surrounding the chromosome 7 locus is rather sparse, rs6950451 was well imputed ($R^2 = .84$). Conditional analyses on the top SNP were also performed, and there were no associated SNPs with $p < .001$.

Gene-based analysis

Gene-based test results are shown in Table 5. Although none of the genes reached genome wide significance, we list the top 20 genes. Two of these (MRAs and OXER1) have already been noted in the top SNP analysis (Table 4), but the other genes listed are all novel.

Polygenic study

Next, we tested an association between the polygenic score for breast size and we found no significant associations between a polygenic score of genetic predisposition to breast size and breastfeeding among the 66 SNPs with $p < 10^{-4}$ available ($\beta = 0.011$, $r^2 = .003$, $p = .67$).

Discussion

In this investigation we used a broad behavioural genetics approach to analyse individual differences in breastfeeding. The first objective was to explore the proportion of genetic and environmental sources in the variation of breastfeeding. Correlations were higher for MZ twins than the ones for DZ twins, suggesting the presence of genetic factors implicated in this behaviour. The results of the twin analysis showed that the best-fitting model was the one including additive genetics effects and unique environmental factors, explaining 53% (CI 95%: .47 – .58) and 47% (CI 95%: .42 – .53) of the variance in breastfeeding behaviour, respectively. The results are also compatible with the full model, in which 17% (CI 95%: 0 – .35) of variance accounted by common environmental factors. The values obtained here are very similar to the ones found recently in a Spanish sample (Colodro-Conde et al., 2013), where 54% of the variance of breastfeeding duration to all the offspring was due to additive genetic factors and the remaining 46%, to unique environmental factors. Despite the samples coming from two distinct cultural backgrounds (Australia vs. Spain), and measures being different (quantitative vs. ordinal), the similar results obtained confirm that genetic factors are an important source of variation between women in breastfeeding behaviour. While we have not been able to detect any significant effect of shared environmental factors, this maybe because of a lack of power due to sample size and our results are compatible with a C^2 as high as 35%.

Confirmation that genetic factors influence breastfeeding behaviour justified our exploring the genetic variants underlying this trait by GWAS, as is discussed below.

The present study is, to our knowledge, the first to perform a GWAS on breastfeeding behaviours. No genome-wide significant SNPs were identified. However, though not significant, we found a promising genetic region on chromosome 7, around rs6950451. Likewise, the VEGAS gene-based analysis did not find any significant results. In addition, we detected suggestive association of some genetic regions with plausible links to breastfeeding. Among them, the oxoeicosanoid (OXE) receptor 1 gene ($p = 1.2 \times 10^{-3}$), in chromosome 2, is a receptor for eicosanoids and polyunsaturated fatty acids (Rebhan et al., 1997), which are present in the breast milk (Koletzko et al., 2008).

It is well known that breast milk provides a unique supply of long chain polyunsaturated fatty acids, and that their synthesis is controlled by key enzymes encoded by the FADS gene cluster (Marquardt et al., 2000; Martin et al., 2011). In this work we found a suggestive association of the FADS1 ($p = 2.3 \times 10^{-2}$), FADS2 ($p = 3.9 \times 10^{-2}$) and FADS3 ($p = 4 \times 10^{-2}$) genes, on chromosome 11 (Figure 4, panel F), with breastfeeding. Several studies have shown strong associations between the FADS gene cluster and fatty acid levels in breast milk (Standl et al., 2012; Glaser et al., 2011). Maternal genetic variants in the FADS gene cluster have previously been associated not only with higher colostrum levels of long chain polyunsaturated fatty acids but also with higher cognitive scores in their children (Morales et al., 2011).

Accordingly, changes in these gene regions may have an effect on breastfeeding duration. As no other study has explored the molecular genetic basis of breastfeeding we cannot yet replicate our findings. The fact that we did not find genome-wide significant results is to be expected considering the low power.

This paper has some limitations that need to be taken into consideration. First, measures of breastfeeding are based on self report and much of the data used was reported a long time after the breastfeeding took place. A larger sample is also required for the GWAS as we expect that breastfeeding is influenced by a large number of SNPs with very small effect size. Ideally, we seek to conduct a meta-analysis combining different samples.

In summary, we have confirmed that genetic factors are significant in the explanation of breastfeeding behaviour variability. However, we could identify neither any breastfeeding-related SNP reaching conventional levels of genome-wide significance nor any overlap with an individual's genetic predisposition towards breast size. This study provides replication of a previous one exploring the sources of variance of breastfeeding in a Spanish sample and, to our knowledge, is the first one to conduct a preliminary GWAS on breastfeeding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to the twins for their participation and to the data collectors and data managers Anthony Conciatore, Richard Parker and Scott Gordon. We also thank Liz Rietschel, Jana Strohmaier, Brendan Zietsch, Ada Johansson, Patrick Jern, and Miguel Renteria, for their comments on earlier drafts of this paper.

Financial support:

Genome-wide genotyping in the Australian sample was funded by NIH Grants AA013320, AA013321, AA013326, AA011998 and AA017688. LCC was supported for a research visit at Genetic Epidemiology Laboratory, QIMR, and for a postdoctoral fellowship by Seneca Foundation -Regional Agency for Science and Technology, Murcia, Spain (12431/FPI/09 and 19151/PD/13).

References

- Almqvist-Tangen G, Bergman S, Dahlgren J, Roswall J, Alm B. Factors associated with discontinuation of breastfeeding before 1 month of age. *Acta Paediatrica*. 2012; 101(1):55–60. [PubMed: 21767302]
- Australian Bureau of Statistics. Australian Health Survey: Health Service Usage and Health Related Actions; p. 2011-2012. <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/6664B939E49FD9C1CA257B39000F2E4B>
- Bacanu S-A, Devlin B, Roeder K. The power of genomic control. *The American Journal of Human Genetics*. 2000; 66(6):1933–1944.
- Balen AH, Conway GS, Kaltsas G, Techatrasak K, Manning PJ, West C, Jacobs HS. Andrology: Polycystic ovary syndrome: the spectrum of the disorder in 1741 patients. *Human Reproduction*. 1995; 10(8):2107–2111. [PubMed: 8567849]
- Barber T, McCarthy M, Wass J, Franks S. Obesity and polycystic ovary syndrome. *Clinical endocrinology*. 2006; 65(2):137–145. [PubMed: 16886951]
- Brown A. Maternal trait personality and breastfeeding duration: the importance of confidence and social support. *Journal of Advanced Nursing*. 2013
- Brown A, Jordan S. Impact of birth complications on breastfeeding duration: an internet survey. *Journal of Advanced Nursing*. 2012
- Brown A, Raynor P, Lee M. Healthcare professionals' and mothers' perceptions of factors that influence decisions to breastfeed or formula feed infants: a comparative study. *Journal of advanced nursing*. 2011; 67(9):1993–2003. [PubMed: 21507050]
- Calnen G. A Strategy for the future. *Breastfeeding Medicine*. 2010; 5(2):95–95.
- Centers for Disease Control and, & Prevention. Breastfeeding among US children born 2001–2011. CDC National Immunization Survey. 2014. http://www.cdc.gov/breastfeeding/data/nis_data/
- Chen W-M, Abecasis GR. Family-based association tests for genome wide association scans. *The American Journal of Human Genetics*. 2007; 81(5):913–926.
- Colodro-Conde L, Sánchez-Romera JF, Ordoñana JR. Heritability of Initiation and Duration of Breastfeeding Behavior. *Twin Research and Human Genetics*. 2013; 16(2):575–580. [PubMed: 23398685]
- Colodro-Conde L, Sánchez-Romera JF, Tornero-Gómez MJ, Pérez-Riquelme F, Polo-Tomás M, Ordoñana JR. Relationship Between Level of Education and Breastfeeding Duration Depends on Social Context Breastfeeding Trends Over a 40-Year Period in Spain. *Journal of Human Lactation*. 2011; 27(3):272–278. [PubMed: 21788656]
- Colodro-Conde L, Limiñana Gras RM, Ordoñana JR. Estilos de Personalidad y Adaptación Psicológica en la toma de decisiones sobre Lactancia en mujeres primíparas [Personality Styles and Psychological Adjustment when making decisions about Breastfeeding in primiparous women]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*. (in press (a)).
- Colodro-Conde L, Limiñana Gras RM, Sánchez-López MP, Ordoñana JR. Gender, Health and Initiation of Breastfeeding. *Women & Health*. (in press (b)).
- De Jager M, Hartley K, Terrazas J, Merrill J. Barriers to breastfeeding – A global survey on why women start and stop breastfeeding. *European Obstetrics & Gynaecology*. 2012; 7(Suppl 1):25–30.
- Dudbridge F, Gusnanto A. Estimation of significance thresholds for genome wide association scans. *Genetic epidemiology*. 2008; 32(3):227–234. [PubMed: 18300295]
- Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, Tung JY. Genetic variants associated with breast size also influence breast cancer risk. *BMC medical genetics*. 2012; 1353(1)

- Glaser C, Lattka E, Rzehak P, Steer C, Koletzko B. Genetic variation in polyunsaturated fatty acid metabolism and its potential relevance for human development and health. *Maternal & child nutrition*. 2011; 7(s2):27–40. [PubMed: 21366865]
- Heath AC, Bucholz K, Madden P, Dinwiddie S, Slutske W, Bierut L, Martin N. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychological medicine*. 1997; 27(6):1381–1396. [PubMed: 9403910]
- Heinig MJ, Dewey KG. Health effects of breastfeeding for mothers: a critical. *Nutrition Research Reviews*. 1997; 10:35–56. [PubMed: 19094257]
- Holman DJ, Grimes MA. Patterns for the initiation of breastfeeding in humans. *American journal of human biology*. 2003; 15(6):765–780. [PubMed: 14595868]
- Ip, S.; Chung, M.; Raman, G.; Chew, P.; Magula, N.; DeVine, D.; Lau, J. Breastfeeding and maternal and infant health outcomes in developed countries. Rockville, MD: Agency for Healthcare Research and Quality; 2009. (Evidence Report/Technology Assessment No. 153)
- Jevitt C, Hernandez I, Groër M. Lactation complicated by overweight and obesity: supporting the mother and newborn. *Journal of Midwifery & Women's Health*. 2007; 52(6):606–613.
- Jonas W, Mileva-Seitz V, Girard A, Bisceglia R, Kennedy J, Sokolowski M, Steiner M. Genetic variation in oxytocin rs2740210 and early adversity associated with postpartum depression and breastfeeding duration. *Genes, Brain and Behavior*. 2013; 12(7):681–694.
- Katz KA, Nilsson I, Rasmussen KM. Danish health care providers' perception of breastfeeding difficulty experienced by women who are obese, have large breasts, or both. *Journal of Human Lactation*. 2010; 26(2):138–147. [PubMed: 19910520]
- Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, Martin NG. Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. *Psychological Medicine*. 2004; 34(8):1519–1530. [PubMed: 15724882]
- Koletzko B, Lien E, Agostoni C, Böhles H, Campoy C, Cetin I, Forsyth S. The roles of long-chain polyunsaturated fatty acids in pregnancy, lactation and infancy: review of current knowledge and consensus recommendations. *Journal of perinatal medicine*. 2008; 36(1):5–14. [PubMed: 18184094]
- Li J, Kendall G, Henderson S, Downie J, Landsborough L, Oddy W. Maternal psychosocial well-being in pregnancy and breastfeeding duration. *Acta Paediatrica*. 2008; 97(2):221–225. [PubMed: 18254911]
- Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Martin NG. A versatile gene-based test for genome-wide association studies. *The American Journal of Human Genetics*. 2010; 87(1):139–145.
- Marquardt A, Stöhr H, White K, Weber BH. cDNA cloning, genomic structure, and chromosomal localization of three members of the human fatty acid desaturase family. *Genomics*. 2000; 66(2):175–183. [PubMed: 10860662]
- Martin NW, Benyamin B, Hansell NK, Montgomery GW, Martin NG, Wright MJ, Bates TC. Cognitive Function in Adolescence: Testing for Interactions Between Breast-Feeding and < i> FADS2 Polymorphisms. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2011; 50(1):55–62. [PubMed: 21156270]
- McFadden A, Toole G. Exploring women's views of breastfeeding: a focus group study within an area with high levels of socio-economic deprivation. *Maternal & child nutrition*. 2006; 2(3):156–168. [PubMed: 16881928]
- Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, Henders AK. Common variants in the trichohyalin gene are associated with straight hair in Europeans. *The American Journal of Human Genetics*. 2009; 85(5):750–755.
- Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI). Encuesta Nacional de Salud de España [Spanish National Health Survey]. 2012. <http://www.msssi.gob.es/en/estadEstudios/estadisticas/sisInfSanSNS/home.htm>
- Morales E, Bustamante M, Gonzalez JR, Guxens M, Torrent M, Mendez M, Vrijheid M. Genetic variants of the FADS gene cluster and ELOVL gene family, colostrums LC-PUFA levels, breastfeeding, and child cognition. *PLoS One*. 2011; 6(2):e17181. [PubMed: 21383846]

- Neale, M.; Boker, S.; Xie, G.; Maes, H. Mx: Statistical Modeling Virginia Commonwealth University. 7th. Richmond, VA: Virginia Institute for Psychiatric and Behavioral Genetics. Department of Psychiatry; 2006.
- Neale, MC.; Cardon, LR. Methodology for genetic studies of twins and families. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1992.
- O'Brien B, Newton N. Psyche versus soma: Historical evolution of beliefs about nausea and vomiting during pregnancy. *J Psychosom Obstet Gynaecol*. 1991; 12:91–120.
- Ordoñana JR, Perez-Riquelme F, Gonzalez-Javier F, Carrillo E, Gomez-Amor J, Martinez-Selva JM. An initiative in Spain for the study of women's health: The Murcia Twin Registry. *Twin Research and Human Genetics*. 2006; 9(6):865–867. [PubMed: 17254422]
- Ordoñana JR, Rebollo-Mesa I, Carrillo E, Colodro-Conde L, García-Palomo FJ, González-Javier F, Pérez-Riquelme F. The Murcia Twin Registry: A Population-Based Registry of Adult Multiples in Spain. *Twin Research and Human Genetics*. 2013; 16(01):302–306. [PubMed: 23046559]
- Posthuma D, Beem AL, de Geus EJC, van Baal GCM, von Hjelmberg JB, Iachine I, Boomsma DI. Theory and practice in quantitative genetics. *Twin Research and Human Genetics*. 2003; 6(5):361–376.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007; 81(3):559–575.
- Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. GeneCards: encyclopedia for genes, proteins and diseases; 1997. <http://www.genecards.org/>
- Sakalidis VS, Williams TM, Garbin CP, Hepworth AR, Hartmann PE, Paech MJ, Geddes DT. Ultrasound imaging of infant sucking dynamics during the establishment of lactation. *Journal of Human Lactation*. 2013; 29(2):205–213. [PubMed: 22965645]
- Schön RA, Silvén M. Natural parenting: Back to basics in infant care. *Evolutionary Psychology*. 2007; 5(1):102–183.
- Sellen DW. Evolution of infant and young child feeding: implications for contemporary public health. *Annu. Rev. Nutr*. 2007; 27:123–148. [PubMed: 17666009]
- Spencer JP. Management of mastitis in breastfeeding women. *Am Fam Physician*. 2008; 78(6):727–731. [PubMed: 18819238]
- SPSS. IBM SPSS Statistics Base. Vol. 19. Chicago, IL: SPSS Inc; 2010.
- Standl M, Sausenthaler S, Lattka E, Koletzko S, Bauer C, Wichmann H, Schaaf B. FADS gene cluster modulates the effect of breastfeeding on asthma. Results from the GINIplus and LISAplus studies. *Allergy*. 2012; 67(1):83–90. [PubMed: 21933193]
- Swanson V, Power KG. Initiation and continuation of breastfeeding: theory of planned behaviour. *Journal of advanced nursing*. 2005; 50(3):272–282. [PubMed: 15811106]
- Tharner A, Luijk MP, Raat H, IJzendoorn MH, Bakermans-Kranenburg MJ, Moll HA, Tiemeier H. Breastfeeding and its relation to maternal sensitivity and infant attachment. *Journal of Developmental & Behavioral Pediatrics*. 2012; 33(5):396–404. [PubMed: 22580735]
- Thulier D, Mercer J. Variables associated with breastfeeding duration. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*. 2009; 38(3):259–268.
- Verweij KJ, Zietsch BP, Medland SE, Gordon SD, Benyamin B, Nyholt DR, Madden PA. A genome-wide association study of Cloninger's temperament scales: implications for the evolutionary genetics of personality. *Biological psychology*. 2010; 85(2):306–317. [PubMed: 20691247]
- Wade TD, Zhu G, Martin NG. Body mass index and breast size in women: same or different genes? *Twin Research and Human Genetics*. 2010; 13(5):450.
- Wagner CL, Wagner MT, Ebeling M, Chatman KG, Cohen M, Hulsey TC. The role of personality and other factors in a mother's decision to initiate breastfeeding. *Journal of Human Lactation*. 2006; 22(1):16–26. [PubMed: 16467284]
- Wojcicki JM. Maternal prepregnancy body mass index and initiation and duration of breastfeeding: a review of the literature. *Journal of women's health*. 2011; 20(3):341–347.
- World Health Organization. Global health observatory data repository [database]. 2013

World Health Organization, & United Nations Children's Fund. Global strategy for infant and young child feeding. Geneva, Switzerland: World Health Organization; 2003.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Scatter plot of twin correlations with 95% confidence intervals (CI) for breastfeeding duration. Note. t1: twin 1, t2: twin 2.

Figure 2.

Manhattan plot showing the results of the genome-wide association analyses for breastfeeding. Genes at or nearby best SNPs are indicated. The vertical axis shows the $-\log_{10}$ of the association p -values and the horizontal axis shows the chromosome numbers divided into 22 autosomes and the X chromosome.

Figure 3.

Quantile–quantile plot for breastfeeding mean duration. The horizontal axis shows the $-\log_{10}$ of expected p -values of association from a 1 degrees of freedom chi-square distribution and the vertical axis shows the $-\log_{10}$ of p -values from the observed chi-square distribution. The coloured dots represent the top hit SNPs. Genes at or nearby best SNPs are indicated.

Figure 4.

Regional association plots for breastfeeding. (A), showing the top associated SNP rs6950451 ($p = 1.2 \times 10^{-7}$) on chromosome 7; (B - F), showing the possible associated regions; on chromosomes 2, 5, 11, 13 and 18.

Table 1

Breastfeeding duration (months) according to childbirth order.

Child order	<i>n</i>	Mean	<i>SD</i>	Range
1 st child	3295	5.32	4.74	0–48
2 nd child	2568	5.19	5.17	0–48
3 rd child	1362	5.24	5.51	0–36
4 th child	547	5.67	6.63	0–60
5 th child	209	4.93	5.99	0–36
6 th child	76	3.93	4.44	0–20
7 th child	25	4.76	8.07	0–40
8 th child	8	4.12	3.44	0–9

Table 2

Correlations (N) for breastfeeding duration (months) in the first five children.

Child order	1 st child	2 nd child	3 rd child	4 th child	5 th child
1 st child	1 (3,295)				
2 nd child	.67* (2,508)	1 (2,568)			
3 rd child	.51* (1,317)	.71* (1,336)	1 (1,362)		
4 th child	.47* (525)	.68* (531)	.77* (531)	1 (547)	
5 th child	.49* (199)	.61* (205)	.68* (203)	.82* (206)	1 (209)

Note: Spearman's rho (N).

* $p < .01$

Model-fitting results for univariate models for breastfeeding mean duration and proportions of variance explained by additive genetic influences (A), common environment (C) and unique environment (E).

Table 3

Model	Parameter estimates (CI = 95%)			Goodness-of-fit index				
	A	C	E	-2LL	df	AIC	X ²	p
ACE	0.34 (0.14,0.55)	0.17 (0.0,0.35)	0.48 (0.43,0.54)	8862.88	3359	2144.88	-	-
AE	0.53 (0.47,0.58)	-	0.47 (0.42,0.53)	8866.20	3360	2146.20	3.11	1 0.08
CE	-	0.45 (0.40,0.50)	0.55 (0.50,0.60)	8874.71	3360	2154.71	11.83	1 0.001
E	-	-	1 (1,1)	9080.60	3361	2358.60	217.72	2 <0.001

A: additive genetic factors, AIC: Akaike's information criterion, BF: breastfeeding mean duration across all births, C: common environmental factors, CI: confidence interval, df: degrees of freedom, E: unique environmental factors, -2LL: twice negative log-likelihood, X²: difference in X² to saturated model, df: difference in degrees of freedom to saturated model. Bold values indicates best fitting model.

Table 4

Top ten SNPs and potential candidates and their gene regions from GWAS analysis showing the strongest associations with breastfeeding.

SNP	CHR	bp	Alleles	AL1%	B	SEB	p-value	At or Near Gene(s)
rs6950451	7	22979808	A/G	77.1	-0.259	0.049	1.2×10^{-7}	SNORD93, FAM126A
rs930451	2	42981239	A/G	68.4	-0.204	0.042	1.2×10^{-6}	OXER1, HAAO, MTA3
rs9807759	18	60269356	C/T	84.4	-0.401	0.083	1.2×10^{-6}	ZCCHC2
rs17381960	13	50095682	G/A	96.0	-0.778	0.164	2.2×10^{-6}	PHF11, RCBTB1
rs7446359	5	127890720	G/A	98.3	-1.185	0.256	3.6×10^{-6}	FBN2
rs7635879	3	193736373	T/A	55.2	0.176	0.039	5.8×10^{-6}	DPPA2P3, LOC647323
rs7699884	4	181295238	T/C	60.2	0.179	0.040	7.7×10^{-6}	(No Gene in ± 400 kb)
rs2277212	10	11299735	T/A	74.7	-0.298	0.067	8.7×10^{-6}	CELF2, CELF-AS2
rs218271	4	55411591	C/G	79.4	-0.238	0.054	8.9×10^{-6}	KIT
rs4724103	7	42339894	C/T	55.3	-0.167	0.038	9.8×10^{-6}	GLI3
rs498793	11	61624705	C/T	61.3	-0.185	0.043	1.7×10^{-5}	FADS2, FADS3, FADS1

Table 5

Genes ($d < 5 \times 10^{-3}$) from VEGAS gene-based analysis showing the strongest associations with breastfeeding.

Gene	Chr	Start position	End position	N SNPs	p-value	Best SNP	SNP p-value
MRAS	3	138066489	138124377	100	1.1×10^{-5}	rs4678408	1.8×10^{-5}
EED	11	85955805	85989785	129	4.1×10^{-5}	rs7951030	2.0×10^{-5}
PAPD4	5	78908242	78982471	135	1.9×10^{-4}	rs4704567	5.8×10^{-5}
DMAPI	1	44679124	44686351	43	3.2×10^{-4}	rs1291169	1.9×10^{-4}
INP5B	1	38326368	38412729	69	3.2×10^{-4}	rs12120737	2.8×10^{-4}
SEL1L2	20	13830049	13971262	244	4.1×10^{-4}	rs6042425	2.1×10^{-5}
PAPOLB	7	4897368	4901625	68	6.5×10^{-4}	rs6966725	3.7×10^{-5}
TPRG1	3	188889762	189041271	303	7.0×10^{-4}	rs1562758	7.9×10^{-5}
MTF1	1	38275238	38325292	53	7.0×10^{-4}	rs4329476	4.1×10^{-4}
CMYA5	5	78985658	79096049	245	1.1×10^{-3}	rs6859704	5.8×10^{-5}
OXER1	2	42989638	42991401	121	1.2×10^{-3}	rs930421	1.2×10^{-6}
ARID3B	15	74833547	74890472	70	1.3×10^{-3}	rs11072492	1.5×10^{-4}
MYCT1	6	153019029	153045715	193	1.7×10^{-3}	rs6934838	1.2×10^{-4}
C20 or 7	20	13765671	13799067	149	2.0×10^{-3}	rs6033833	1.2×10^{-4}
CHRNA4	15	78916635	78933587	95	2.3×10^{-3}	rs8043123	6.2×10^{-5}
ATP9B	18	76829396	77138282	358	2.9×10^{-3}	rs9954562	1.0×10^{-4}
MMP7	11	102391238	102401478	197	3.4×10^{-3}	rs7951520	3.5×10^{-5}
CPNE2	16	57126454	57181878	137	3.7×10^{-3}	rs2216758	7.8×10^{-5}
ADAMTS7	15	79051544	79103773	63	4.9×10^{-3}	rs12899940	3.2×10^{-5}