Published in final edited form as:
Ann Hum Genet. 2013 January ; 77(1): 31–46. doi:10.1111/j.1469-1809.2012.00734.x.

Anorectal atresia and variants at predicted regulatory sites in candidate genes

Tonia C. Carter[1], **Denise M. Kay**[2], **Marilyn L. Browne**[3,4], **Aiyi Liu**[1], **Paul A. Romitti**[5], **Devon Kuehn**[1], **Mary R. Conley**[1], **Michele Caggana**[2], **Charlotte M. Druschel**[3,4], **Lawrence C. Brody**[6], and **James L. Mills**[1]

[1]Division of Epidemiology, Statistics, and Prevention Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA

[2]Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, New York, USA

[3]Congenital Malformations Registry, New York State Department of Health, Albany, New York, USA

[4]Department of Epidemiology and Biostatistics, School of Public Health, University at Albany – State University of New York, Albany, New York, USA

[5]Department of Epidemiology, The University of Iowa College of Public Health, Iowa City, Iowa, USA

[6]Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

SUMMARY

Anorectal atresia is a serious birth defect of largely unknown etiology but candidate genes have been identified in animal studies and human syndromes. Because alterations in the activity of these genes might lead to anorectal atresia, we selected 71 common variants predicted to be in transcription factor binding sites, CpG windows, splice sites, and miRNA target sites of 25 candidate genes, and tested for their association with anorectal atresia. The study population comprised 150 anorectal atresia cases and 623 control infants without major malformations. Variants predicted to affect transcription factor binding, splicing, and DNA methylation in *WNT3A*, *PCSK5*, *TCF4*, *MKKS*, *GLI2*, *HOXD12*, and *BMP4* were associated with anorectal atresia based on a nominal *P* value <0.05. The *GLI2* and *BMP4* variants are reported to be moderately associated with gene expression changes (Spearman's rank correlation coefficients between −0.260 and 0.226). We did not find evidence for interaction between maternal pre-pregnancy obesity and variants in *MKKS*, a gene previously associated with obesity, on the risk of anorectal atresia. Our results for *MKKS* support previously suggested associations with anorectal malformations. Our findings suggest that more research is needed to determine whether altered *GLI2* and *BMP4* expression is important in anorectal atresia in humans.

Corresponding author: James L. Mills, MD, MS, 6100 Building, Room 7B03, DESPR, NICHD, NIH, Bethesda, MD 20892, Tel: 301-496-5394, Fax: 301-402-2084, jamesmills@nih.gov.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

Keywords

anorectal malformations; imperforate anus; hindgut; congenital abnormalities

INTRODUCTION

Anorectal atresia (imperforate anus) is a gastrointestinal birth defect that causes perinatal morbidity and requires surgical reconstruction. Prevalence is estimated to be 3–5 cases per 10,000 births (Spouge & Baird, 1986; Cuschieri & EUROCAT Working Group, 2001). Approximately 64–68% of cases have additional defects (Spouge &Baird, 1986; Cuschieri & EUROCAT Working Group, 2001) that can be categorized as non-syndromic multiple defects, chromosomal abnormalities, syndromes, sequences, and associations (Stoll et al. 2007). The etiology of anorectal atresia is uncertain; however, there is evidence that genetic factors are important contributors. First, although most cases are non-familial, case reports have described familial cases, some occurring over multiple generations (Weinstein, 1965; Schwoebel et al. 1984; Landau et al. 1997). Second, anorectal atresia is a component of recognized syndromes (e.g. Currarino, Pallister-Hall, and Townes-Brocks syndromes) resulting from mutations in specific genes (Hagan et al. 2000; Johnston et al. 2005; Botzenhart et al. 2007). Third, anorectal atresia is an inherited trait in certain lines of mice and pigs (Kluth et al. 1991; Hori et al. 2001).

To gain insight into the genetic factors involved in anorectal atresia etiology, we examined single nucleotide polymorphisms (SNPs) in selected candidate genes, specifically focusing on variants predicted to be in regulatory sites. In support of this approach, studies of other birth defects (non-syndromic oral clefts and Hirschsprung's disease) have identified SNPs located in transcription factor binding sites of candidate genes and have shown that these SNPs are strongly associated with these defects (Rahimov et al. 2008; Emison et al. 2010). This approach is relevant to the search for genetic risk factors for birth defects, including anorectal atresia, because variants in the regulatory sites of genes have the potential to alter gene activity and might be important for gene regulation during development. For our study of anorectal atresia, we examined SNPs from 25 candidate genes, chosen because reports of gene knockout in animals and mutation analysis in human syndromes that sometimes feature anorectal atresia provide evidence for their involvement in anorectal malformations (Mundt & Bates, 2010). Our objective was to determine whether SNPs with the potential to alter the activity of the 25 candidate genes are associated with anorectal atresia.

MATERIALS AND METHODS

Subjects

We conducted a nested case-control study based on the cohort of all live births in New York State for the birth years 1998–2005 (N=2,023,083). Live-born cases with anorectal atresia were identified from the New York State Congenital Malformations Registry. Physicians and hospitals are required by law to report to the registry all children under the age of two years who are diagnosed with one or more birth defects and who were born, or reside, in New York State. The study was restricted to cases that had anorectal atresia as the only major birth defect (N=155). Controls were live-born infants with no major birth defects. A random sample of controls was selected from the records of the New York State Newborn Screening Program after stratification by race/ethnicity. Controls (N=623) were frequency-matched to cases by race/ethnicity and the ratio of controls to cases was approximately 4:1.

New York State Congenital Malformations Registry records for cases were linked to records of the New York State Newborn Screening Program, and archived residual dried blood spots

were obtained for cases and controls. Five case records were not matched to the correct blood spot sample and were excluded. Therefore, 150 cases and 623 controls remained for analysis.

Birth certificates also provided data on maternal and infant characteristics. After the biological samples were processed, identifying information was removed from samples and data for all study subjects. This study was approved by the Institutional Review Board of the New York State Department of Health and was reviewed by the Office of Human Subjects Research at the National Institutes of Health.

SNPs

The bioinformatics tools SNPnexus (Chelala et al. 2009), SNPseek (Coassin et al. 2010), FastSNP (Yuan et al. 2006), miRBase (Griffiths-Jones et al. 2006), and the Genomatix suite (Werner, 2002) were used to identify SNPs predicted to alter transcription factor binding sites, CpG windows, splice sites, splicing enhancer/silencer sites, and miRNA target sites of the 25 candidate genes. The genes and their functions are presented in Table 1. The evidence and citations supporting their role in anorectal malformations are summarized in Supplementary Table 1. Initially, regions that encompassed the gene as well as 2 kb on either side were examined. SNPs with a minor allele frequency >0.1 in the 1000 Genomes Project or in any of the HapMap European, Han Chinese, Japanese, or Yoruban populations were selected. For three genes (*EFNB2*, *GDF11*, *SHH*) there were no relevant SNPs that matched these criteria, therefore the gene region was extended to 10 kb upstream and 5 kb downstream and the minor allele frequency threshold was lowered to 0.075. For one gene, *HOXD13*, a relevant SNP was identified only after the minor allele frequency threshold was reduced to 0.035. A total of 93 SNPs were identified, of which 11 were subsequently excluded because they were in strong linkage disequilibrium ($r^2 \geq 0.80$) with other selected SNPs, leaving 82 SNPs for which genotyping was attempted.

Laboratory analysis

Punches of 3 mm in diameter were made from each dried blood spot and sodium hydroxide precipitation was used to extract DNA from the punches. KBiosciences UK (Hoddesdon, Herts, UK) performed whole genome amplification and genotyping of ≥30 ng of the extracted DNA. Genotyping entailed the use of a competitive, allele-specific, primer extension pre-amplification method. Two separate rounds of whole genome amplification were performed for each study subject and the products of each round were genotyped. Three SNPs (*GDF11* rs7068:A>G, *PCSK5* rs7872060:G>A, *TCF4* rs2958162:T>C) were genotyped using genomic DNA because results from whole-genome amplified DNA did not pass the quality control criteria of the genotyping facility.

Eight SNPs failed either assay design or validation on test DNA and therefore were excluded. A ninth SNP could not be successfully genotyped on either genomic DNA or DNA that had undergone whole genome amplification and was also excluded. The minor allele of a tenth tri-allelic SNP (*UBR1* rs3917223; Table 1) was incorrectly specified and therefore the allele of interest was not genotyped. As a result, genotypes were available for 72 SNPs (Table 1).

Genotyping quality control measures included the use of blank wells and repeat genotyping of 5% of DNA samples. All SNPs were called successfully >98% of the time. When genotypes from the two rounds of whole genome amplification were compared, there were nine discordant calls in eight different SNPs among the 56,187 genotypes that were successfully called. In addition, there was one genotype error for *PQBP1* rs741932:T>C (on the X chromosome): a male subject had a heterozygous genotype for this SNP. This subject

was not included in analyses for *PQBP1* rs741932:T>C. No discordant genotypes were observed after repeat genotyping of 5% of samples. Genotypes that were discordant or thought to be due to error were set to missing for the statistical analyses. For one SNP (*HOXD13* rs35290213:A>C), all study subjects were homozygous for the major allele. However, its minor allele frequency was expected to be low (Table 1). After exclusion of this SNP, 71 SNPs in 23 genes remained for analysis.

Tests for deviation from Hardy-Weinberg equilibrium for the 71 SNPs were performed independently for cases and controls and separately by race/ethnicity (adjustment for 568 tests using the Bonferroni method; $P<8.8\text{x}10^{-5}$). None of the SNPs deviated from Hardy-Weinberg equilibrium.

Statistical analysis

Data on maternal and infant characteristics were compared between cases and controls using Fisher's exact test. Logistic regression was used to compare genotype distributions of the 71 SNPs between cases and controls. In regression analyses, two degree-of-freedom tests were used to generate *P* values for SNPs on autosomes and the X chromosome in females; a one degree-of-freedom test was used to generate *P* values for SNPs on the X chromosome in males. Analyses were performed for the study subjects overall, with race/ethnicity included in the regression model. Separate analyses were also performed for each race/ethnic group. The exception was for the group described as "other" because its sample size was too small to permit separate analyses. Potential confounders were selected from among the maternal and infant characteristics if the *P* values for their associations with anorectal atresia were <0.1. Maternal smoking during pregnancy (yes/no) was included as a covariate in logistic regression analyses because previous reports suggest that parental smoking is associated with anorectal malformations (Zwink et al. 2011). For SNPs in two genes on the X chromosome (*PQBP1* rs741932:T>C; *ZIC3* rs5931174:T>C), analyses were performed separately for males and females. Analyses were repeated after restriction to singleton births to determine whether birth plurality influenced the results. Genotype analyses were adjusted for multiple comparisons using the Bonferroni method (71 SNPs tested in entire study population and in each of four race/ethnic groups, and tests were repeated among singleton births resulting in total of 710 tests; $P<0.00007$). This adjustment was applied to analyses for the full study population as well as the subset of singleton births. SAS software, version 9.2 (SAS Institute, Cary, North Carolina) was used to conduct statistical analyses.

Measures of linkage disequilibrium were estimated using Haploview (Barrett et al. 2005), based on the genotypes of control subjects, or the HapMap and 1000 Genomes populations, as indicated in the text.

RESULTS

Mothers of case and control infants did not differ significantly by maternal age, education, smoking during pregnancy, pre-gestational diabetes, gestational diabetes, use of in vitro fertilization or other assisted reproductive technique, plurality, or birth year (Table 2). Case mothers were more likely to be nulliparous, a difference of borderline significance ($P=0.07$). There was a preponderance of males among case infants: the male-to-female ratio was 1.23 for case infants and 0.79 for control infants ($P=0.014$). We did not included infant sex as a covariate in logistic regression analyses because it was not considered to be a cause of birth defects.

Associations were observed between anorectal atresia and SNPs in some of the candidate genes at a nominal *P* value <0.05; these associations varied by race/ethnicity (Table 3). *PCSK5* rs7040769:T>C ($P=0.046$), *TCF4* rs8766:A>G ($P=0.044$), *MKKS* rs2013178:T>A

(*P*=0.0015), and *MKKS* rs1003994:G>A (*P*=0.0078) were associated with anorectal atresia in non-Hispanic whites. The *PCSK5* and *TCF4* SNPs are both predicted to be in exon splicing enhancer sites. The *MKKS* SNPs are predicted to be in transcription factor binding sites (Table 1) and are in moderately strong linkage disequilibrium with each other (r^2=0.79). In African-Americans, anorectal atresia was associated with *WNT3A* rs12401893:G>A (*P*=0.031), *GLI2* rs3738880:A>C (*P*=0.020), *PCSK5* rs872189:C>T (*P*=0.034), and *PCSK5* rs2279659:C>T (*P*=0.043). The *WNT3A* SNP is predicted to be in a CpG window; the *GLI2* and *PCSK5* SNPs are predicted to be in transcription factor binding sites (Table 1). The two *PCSK5* SNPs are not in linkage disequilibrium with each other (r^2=0). *HOXD12* rs35817516:G>A (*P*=0.020) and *MKKS* rs2013178:T>A (*P*=0.028) were associated with anorectal atresia in Hispanics; both SNPs are predicted to be in transcription factor binding sites. *BMP4* rs17563:T>C, predicted to be in an exon splicing silencer site, was associated with anorectal atresia in Asians (*P*=0.033). No associations were observed in the overall group of study subjects.

For each of the SNPs showing an association with anorectal atresia, we determined the risk genotype(s) by using logistic regression to calculate odds ratios and 95% confidence intervals (Supplementary Table 2). *PCSK5* rs7040769 CC, *MKKS* rs2013178 TT (in non-Hispanic whites), *MKKS* rs1003994 GG, *GLI2* rs3738880 AC and AA, *PCSK5* rs872189 CC, *PCSK5* rs2279659 TT, *HOXD12* rs35817516 AA, and *BMP4* rs17563 CC genotypes were associated with elevated odds ratios for anorectal atresia. Odds ratios were approximately 2.0 for *TCF4* rs8766 GG and *WNT3A* rs12401893 AA genotypes but these associations were not statistically significant.

Previous reports of associations between *MKKS* variants and obesity (Benzinou et al. 2006), and between obesity and anorectal atresia (Waller et al. 2007), prompted us to explore whether anorectal atresia was associated with an interaction between *MKKS* variants in the offspring and maternal obesity (body mass index ≥30 kg/m^2). Data on pre-pregnancy body mass index were available from the birth certificate for mothers of 79/150 (53%) cases and 340/623 (54%) controls. Although we observed *MKKS* variants to be associated with anorectal atresia in non-Hispanic whites and Hispanics (Table 3), the sample size for Hispanics was too small to test for interaction: only 12 Hispanic case mothers had data on maternal pre-pregnancy obesity. Therefore, we conducted the analysis for the overall group of study subjects and non-Hispanic whites (Table 4). We tested for additive interaction by calculating the interaction contrast ratio and its 95% confidence interval as described by Richardson & Kaufman (2009). The interaction contrast ratio represents the excess risk resulting from the interaction relative to the risk when exposure is absent (Kalilani & Atashili, 2006). When there is no interaction, the interaction contrast ratio has a value of zero. We also checked whether the magnitude of the odds ratio in the presence of both maternal pre-pregnancy obesity and the *MKKS* variant was greater than the sum of the odds ratios for the separate effects of maternal pre-pregnancy obesity and the *MKKS* variant. We found that the 95% confidence intervals for the interaction contrast ratios included zero and that there were no statistically significant elevations in the odds ratios when both maternal pre-pregnancy obesity and homozygosity for the *MKKS* variant in the offspring were present (Table 4). Based on these results, we did not find evidence for an interaction between *MKKS* variants and obesity in anorectal atresia.

There were no meaningful changes in the results after restricting the study population to singleton births. Also, none of the findings remained statistically significant after stringent adjustment for multiple comparisons using the Bonferroni method.

Because the observed associations were only statistically significant at a nominal *P* value <0.05 (Table 3), and because the tested SNPs could be a marker for other causative variants,

we used data from the HapMap and 1000 Genomes populations to check whether the 10 SNPs showing an association in Table 3 are in linkage disequilibrium with other rare coding variants. The results are shown in Supplementary Table 3. None of the 10 SNPs was in strong or moderate linkage disequilibrium (r^2>0.5) with rare coding variants. Five were within 1000 bp of at least one missense or frameshift variant in the same gene. Because data on the minor allele frequency of these nearby variants are limited, it is unclear whether they are rare in the general population. However, they are worth further investigation as risk factors for anorectal atresia.

We also explored whether SNPs that were statistically significant at a nominal *P* value <0.05 were associated with changes in gene expression. We used the Genevar database (Yang et al. 2010) which contains gene expression data from three different datasets: lymphoblastoid cell lines from HapMap individuals (Stranger et al. 2012); adipose tissue, lymphoblastoid cell lines, and skin samples from twins of Caucasian ancestry (Nica et al. 2011); and fibroblasts, lymphoblastoid cell lines, and T-cells from umbilical cord of newborns with Western European ancestry (Dimas et al. 2009). Data were available for five SNPs (*GLI2* rs3738880, *HOXD12* rs35817516, *PCSK5* rs7040769, *PCSK5* rs2279659, *BMP4* rs17563) and the results are presented in Supplementary Figures 1–14. Three different probes were used to measure *BMP4* gene expression in samples from HapMap individuals (Supplementary Figures 9–11) and twins (Supplementary Figures 12–14).

Associations were found between *GLI2* rs3738880 and *GLI2* gene expression in HapMap Luhya samples (Supplementary Figure 1, LWK, adjusted *P*=0.043), *HOXD12* rs35817516 and *HOXD12* gene expression in HapMap Gujarati samples (Supplementary Figure 3, GIH, adjusted *P*=0.033), *PCSK5* rs2279659 and *PCSK5* gene expression in umbilical cord T-cells (Supplementary Figure 7, GenCord-T, adjusted *P*=0.0089), and between *BMP4* rs17563 and *BMP4* gene expression in HapMap Gujarati samples (Supplementary Figure 9, GIH, adjusted *P*=0.018). However, the magnitude of the Spearman's rank correlation coefficients for these associations (rho between −0.260 and 0.297) was moderate. The positive value of rho (0.226) for *GLI2* rs3738880 in HapMap Luhya samples suggested that *GLI2* rs3738880 AA was associated with increased *GLI2* gene expression, and the negative value of rho (−0.260) for *BMP4* rs17563 in HapMap Gujarati samples suggested that *BMP4* rs17563 CC was associated with reduced *BMP4* gene expression. The number of samples with the *HOXD12* rs35817516 AA genotype in HapMap Gujarati samples (Supplementary Figure 3) and the *PCSK5* rs2279659 AA genotype in umbilical cord T-cells (Supplementary Figure 7) was too small to examine their effect on gene expression.

DISCUSSION

Mis-regulation of gene expression is a possible mechanism of birth defects. Therefore, we investigated whether SNPs predicted to affect gene function at transcriptional or post-transcriptional stages were associated with anorectal atresia, a birth defect of the hindgut. We observed that SNPs predicted to alter splicing, DNA methylation, or the binding of transcription factors in *WNT3A*, *PCSK5*, *TCF4*, *MKKS*, *GLI2*, *HOXD12*, and *BMP4* were associated with anorectal atresia, based on nominally significant results (*P*<0.05). The finding for *MKKS*, a gene involved in a human syndrome that sometimes includes anorectal atresia (Stone et al. 2000), supports its suggested association with anorectal atresia and indicates that the regulation of transcription of this gene could influence the occurrence of anorectal atresia. *PCSK5* SNPs showed associations in more than one race/ethnic group, as did a SNP in the *MKKS* gene. Associations with variants at predicted transcription factor binding sites also implicate the relevant transcription factors as contributors to anorectal atresia, and the genes encoding these transcription factors are a promising area for future research.

For most of the candidate genes in this study, evidence suggesting an involvement in anorectal atresia was obtained from animal studies (Mundt & Bates, 2010). Our study provides evidence that variants in *PCSK5*, *TCF4*, *GLI2*, *HOXD12*, and *BMP4* are also associated with anorectal atresia in humans. *MKKS* is a gene in which mutations have been detected among patients with a recognized syndrome (McKusick-Kaufman syndrome) that sometimes include anorectal malformations (Robinow & Shaw, 1979). Our data indicate that variants in this gene could also be involved in non-syndromic cases of anorectal atresia.

MKKS variants have been associated with obesity (Benzinou et al. 2006; Rouskas et al. 2008) and maternal pre-pregnancy obesity has been associated with anorectal atresia (Waller et al. 2007). Knockout of *Mkks* in mice alters leptin receptor signaling; this produces resistance to the effects of leptin to reduce body weight and food consumption, and leads to obesity in the affected animals (Seo et al. 2009). There are conflicting data on the association between obesity and two *MKKS* variants (rs1545:C>A and rs17852625:G>A) examined in this study. *MKKS* rs1545:C>A was associated with obesity in a Greek population (Rouskas et al. 2008) but no association with this SNP or rs17852625:G>A was observed in a Danish study (Andersen et al. 2005). We did not find evidence for association between these variants and anorectal atresia. The two SNPs associated with anorectal atresia in this study were not assessed in previous studies of *MKKS* and obesity. Further, maternal pre-pregnancy obesity did not influence the association between these SNPs and anorectal atresia. This requires additional investigation because the number of cases, even in our study, was limited. It is possible that *MKKS* has a role in both obesity and anorectal atresia; however, the interrelationships among the three factors are not yet clear.

Interactions among genes expressed in different embryonic cell layers are important in hindgut development. Sonic hedgehog signaling in the endoderm induces the expression of *Bmp4* and *Hox* genes in hindgut mesoderm during chick gut development (Roberts et al. 1995). Our results showing associations between anorectal atresia and SNPs in a number of genes (*GLI2*, *HOXD12*, and *BMP4*) that are downstream targets of sonic hedgehog suggest that similar interactions might be involved in anorectal morphogenesis in humans. Other evidence suggests there might be interactions between some sonic hedgehog targets and *PCSK5*, another gene associated with anorectal atresia in this study. *BMP4* is a substrate for cleavage by *PCSK5* during embryogenesis (Cui et al. 1998); in embryonic mice the C470R mutation in *Pcsk5* leads to abnormal expression of *Hox* genes (including *Hoxd12*) and a range of birth defects including anorectal malformations (Szumska et al. 2008); and NKX2-2, the transcription factor predicted to have a binding site in *PCSK5* that is affected by the SNP rs2279659:C>T, was observed to be a target of sonic hedgehog signaling (Vokes et al. 2007).

We observed that the *GLI2* rs3738880 AA genotype was associated with anorectal atresia in our African-American population and with increased *GLI2* expression in HapMap samples from the Luhya population in Kenya. It is possible that increased *GLI2* expression could lead to altered sonic hedgehog signaling. Future studies should examine what role this might play in anorectal atresia in African-Americans. The *BMP4* rs17563 CC genotype was associated with both decreased gene expression and elevated odds ratios for anorectal atresia. Another study investigating the functional effect of *BMP4* rs17563:T>C found that the quantity of *BMP4* mRNA in plasma was significantly greater among carriers of the T allele than carriers of the C allele (Capasso et al. 2009). Other evidence supports a role for *BMP4* rs17563:T>C in birth defects: this variant was associated with non-syndromic cleft lip with or without cleft palate in a Chinese study population (Lin et al. 2008). Oro-facial clefts represent another birth defect for which there is evidence from animal studies of the involvement of a network of genes, including sonic hedgehog signaling and *BMP4* (Lan & Jiang, 2009).

BMP4 rs17563:T>C is in a predicted exon splicing silencer motif and could possibly regulate *BMP4* expression by effects on mRNA splicing. Exon splicing enhancers and silencers are important in regulating gene expression as illustrated by the *INSR* gene encoding the human insulin receptor (Sen et al. 2009). The binding of splicing regulatory proteins to exon splicing enhancer and silencer sites in *INSR* leads to expression of two insulin receptor isoforms as a result of alternative splicing of exon 11 of *INSR*. During embryogenesis, there is increased expression of insulin receptor lacking exon 11, whereas in the adult, the insulin receptor containing exon 11 is expressed predominantly. Further investigation is required to explore the mechanism by which the rs17563:T>C SNP could influence *BMP4* expression.

Because the associations between changes in gene expression and the *GLI2* rs3738880 and *BMP4* rs17563 SNPs were only moderate in magnitude, we need to be cautious in considering these variants as genetic risk factors for anorectal atresia. More evidence supporting a role for these variants in anorectal atresia is needed. Also, because the gene expression changes were observed in tissues other than hindgut, the functional effects of these SNPs in the hindgut need to be investigated.

Our comparison of selected demographic and other non-genetic factors between cases and controls did not find any associations with anorectal atresia. A review and meta-analysis of parental risk factors for anorectal malformations (Zwink et al. 2011) also found no association with maternal smoking. However, in contrast to our observations, the results of the meta-analysis showed that pre-gestational diabetes and gestational diabetes were risk factors. Our study probably had low power to detect associations with pre-gestational diabetes and gestational diabetes because these diseases were rare in our study population. The lack of an association with parity and assisted reproductive technology in our study also contradicts other reports (Reefhuis et al. 2009; van Rooij et al. 2010). This could be partly due to differences in the case groups between our study and these two reports. Our cases had anorectal atresia as their only major malformation; in the other two studies some cases had multiple major malformations involving more than one organ system. Because there is evidence that both genetic and non-genetic factors influence the risk of anorectal malformations, their interaction should be investigated further.

One of the strengths of this study was the relatively large, population-based sample of anorectal atresia cases and controls. Because of the rarity of this defect, many previous reports have had smaller sample sizes and included mostly clinic-based cases. Also, because cases and controls were drawn from the general New York State population, we were able to examine associations in the four major race/ethnic groups that make up this population. Limitations of our study include the possibility that our adjustment for multiple comparisons using the Bonferroni method was too conservative because we examined genes for which there is strong prior evidence for an involvement in anorectal malformations. However, none of the observed associations remained statistically significant after the adjustment, and we cannot exclude the possibility that the associations were due to chance. We were also uncertain about the accuracy of data on maternal height and pre-pregnancy weight obtained from the birth certificate. These data could have been based on measurements or on maternal self-report, and misclassification of maternal obesity was possible.

We were also limited by the lack of medical record data on cases; therefore, we could not distinguish cases that had a fistula or determine whether the defect was above or below the level of the levator ani muscle. Consequently, we could not investigate whether associations with genetic factors varied by these characteristics. Findings from a study that examined *SHH, GLI2* and *BMP4* expression in tissue from the terminal rectum of cases with anorectal malformations and controls suggest that genetic factors could differ according to the level of

the defect (Zhang et al. 2009). Compared with controls, expression of all three genes was lower among cases whose malformation occurred above the pubococcygeal line. Only *GLI2* expression was lower among cases whose malformation was below the pubococcygeal line.

We demonstrated that a gene (*MKKS*) responsible for a human syndrome that sometimes includes anorectal atresia is likely to play a role in non-syndromic cases of anorectal atresia. Our results, which require confirmation, also lead us to conclude that a number of genes (*WNT3A*, *PCSK5*, *TCF4*, *GLI2*, *HOXD12*, and *BMP4)* identified as being involved in anorectal malformations in animals might contribute to anorectal atresia in humans. One of the genes (*GLI2*) mediates sonic hedgehog signaling and others (*HOXD12* and *BMP4*) are known targets of the sonic hedgehog signaling pathway; this suggests that normal functioning of this pathway could be critical to human embryonic hindgut development. Our findings indicate that sonic hedgehog pathway signalling is a promising area for future research into the etiology of anorectal atresia. Our observations that SNPs in *GLI2* and *BMP4* were also associated with changes in gene expression suggest a mechanism by which these SNPs could play a role in anorectal atresia. However, the associations with gene expression were moderate and more evidence is needed to clarify these associations. Further investigation into the regulation of expression of these genes in the hindgut could be informative for determining the mechanisms leading to anorectal atresia.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank April J. Atkins, Emily C. McGrath, and Robert J. Sicko for laboratory and technical assistance. We are also grateful for the data management skills of Sandra D. Richardson. This work was supported by the Intramural Research Program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (Contract # HHSN267200703431C; NICHD # N01-DK-7-3431).

REFERENCES

Andersen KL, Echwald SM, Larsen LH, Hamid YH, Glumer C, Jorgensen T, Borch-Johnsen K, Andersen T, Sorensen TI, Hansen T, Pedersen O. Variation of the McKusick-Kaufman gene and studies of relationships with common forms of obesity. J Clin Endocrinol Metab. 2005; 90:225–230. [PubMed: 15483080]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

Benzinou M, Walley A, Lobbens S, Charles MA, Jouret B, Fumeron F, Balkau B, Meyre D, Froguel P. Bardet-Biedl syndrome gene variants are associated with both childhood and adult common obesity in French caucasians. Diabetes. 2006; 55:2876–2882. [PubMed: 17003356]

Botzenhart EM, Bartalini G, Blair E, Brady AF, Elmslie F, Chong KL, Christy K, Torres-Martinez W, Danesino C, Deardorff MA, Fryns JP, Marlin S, Garcia-Minaur S, Hellenbroich Y, Hay BN, Penttinen M, Shashi V, Terhal P, Van Maldergem L, Whiteford ML, Zackai E, Kohlhase J. Townes-Brocks syndrome: twenty novel SALL1 mutations in sporadic and familial cases and refinement of the SALL1 hot spot region. Hum Mutat. 2007; 28:204–205. [PubMed: 17221874]

Capasso M, Ayala F, Russo R, Avvisati RA, Asci R, Iolascon A. A predicted functional single-nucleotide polymorphism of bone morphogenetic protein-4 gene affects mRNA expression and shows a significant association with cutaneous melanoma in Southern Italian population. J Cancer Res Clin Oncol. 2009; 135:1799–1807. [PubMed: 19557432]

Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics. 2009; 25:655–661. [PubMed: 19098027]

Coassin S, Brandstatter A, Kronenberg F. Lost in the space of bioinformatic tools: a constantly updated survival guide for genetic epidemiology. The GenEpi Toolbox. Atherosclerosis. 2010; 209:321–335. [PubMed: 19963217]

Cui Y, Jean F, Thomas G, Christian JL. BMP-4 is proteolytically activated by furin and/or PC6 during vertebrate embryonic development. EMBO J. 1998; 17:4735–4743. [PubMed: 9707432]

Cuschieri A. EUROCAT Working Group. Descriptive epidemiology of isolated anal anomalies: a survey of 4.6 million births in Europe. Am J Med Genet. 2001; 103:207–215. [PubMed: 11745992]

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez-Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–1250. [PubMed: 19644074]

Emison ES, Garcia-Barcelo M, Grice EA, Lantieri F, Amiel J, Burzynski G, Fernandez RM, Hao L, Kashuk C, West K, Miao X, Tam PK, Griseri P, Ceccherini I, Pelet A, Jannot AS, de Pontual L, Henrion-Claude A, Lyonnet S, Verheij JB, Hofstra RM, Antinolo G, Borrego S, McCallion AS, Chakravarti A. Differential contributions of rare and common, coding and noncoding RET mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet. 2010; 87:60–74. [PubMed: 20598273]

Hagan DM, Ross AJ, Strachan T, Lynch SA, Ruiz-Perez V, Wang YM, Scambler P, Custard E, Reardon W, Hassan S, Nixon P, Papapetrou C, Winter RM, Edwards Y, Morrison K, Barrow M, Cordier-Alex MP, Correia P, Galvin-Parton PA, Gaskill S, Gaskin KJ, Garcia-Minaur S, Gereige R, Hayward R, Homfray T. Mutation analysis and embryonic expression of the HLXB9 Currarino syndrome gene. Am J Hum Genet. 2000; 66:1504–1515. [PubMed: 10749657]

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34:D140–D144. [PubMed: 16381832]

Hori T, Giuffra E, Andersson L, Ohkawa H. Mapping loci causing susceptibility to anal atresia in pigs, using a resource pedigree. J Pediatr Surg. 2001; 36:1370–1374. [PubMed: 11528608]

Johnston JJ, Olivos-Glander I, Killoran C, Elson E, Turner JT, Peters KF, Abbott MH, Aughton DJ, Aylsworth AS, Bamshad MJ, Booth C, Curry CJ, David A, Dinulos MB, Flannery DB, Fox MA, Graham JM, Grange DK, Guttmacher AE, Hannibal MC, Henn W, Hennekam RC, Holmes LB, Hoyme HE, Leppig KA, Lin AE, Macleod P, Manchester DK, Marcelis C, Mazzanti L, McCann E, McDonald MT, Mendelsohn NJ, Moeschler JB, Moghaddam B, Neri G, Newbury-Ecob R, Pagon RA, Phillips JA, Sadler LS, Stoler JM, Tilstra D, Walsh Vockley CM, Zackai EH, Zadeh TM, Brueton L, Black GC, Biesecker LG. Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of GLI3 mutations. Am J Hum Genet. 2005; 76:609–622. [PubMed: 15739154]

Kalilani L, Atashili J. Measuring additive interaction using odds ratios. Epidemiol Perspect Innov. 2006; 3:5. [PubMed: 16620385]

Kluth D, Lambrecht W, Reich P, Buhrer C. SD-mice – an animal model for complex anorectal malformations. Eur J Pediatr Surg. 1991; 1:183–188. [PubMed: 1892807]

Lan Y, Jiang R. Sonic hedgehog signaling regulates reciprocal epithelial-mesenchymal interactions controlling palatal outgrowth. Development. 2009; 136:1387–1396. [PubMed: 19304890]

Landau D, Mordechai J, Karplus M, Carmi R. Inheritance of familial congenital isolated anorectal malformations: case report and review. Am J Med Genet. 1997; 71:280–282. [PubMed: 9268096]

Lin JY, Chen YJ, Huang YL, Tang GP, Zhang L, Deng B, Li M, Ma H, Luan RS. Association of bone morphogenetic protein 4 gene polymorphisms with nonsyndromic cleft lip with or without cleft palate in Chinese children. DNA Cell Biol. 2008; 27:601–605. [PubMed: 18771417]

Mundt E, Bates MD. Genetics of Hirschsprung disease and anorectal malformations. Semin Pediatr Surg. 2010; 19:107–117. [PubMed: 20307847]

Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, Ingle C, Nestle FO, di Meglio P, Min JL, Wilk A, Hammond CJ, Hassanali N, Yang TP, Montgomery SB, O'Rahilly S, Lindgren CM, Zondervan KT, Soranzo N, Barroso I, Durbin R, Ahmadi K, Deloukas P, McCarthy MI, Dermitzakis ET, Spector TD. MuTHER Consortium. The architecture of gene regulatory

variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011; 7:e1002003. [PubMed: 21304890]

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, Pennacchio LA, Schutte BC, Murray JC. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet. 2008; 40:1341–1347. [PubMed: 18836445]

Reefhuis J, Honein MA, Schieve LA, Correa A, Hobbs CA, Rasmussen SA. The National Birth Defects Prevention Study. Assisted reproductive technology and major structural birth defects in the United States. Hum Reprod. 2009; 24:360–366. [PubMed: 19010807]

Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am J Epdemiol. 2009; 169:756–760.

Roberts DJ, Johnson RL, Burke AC, Nelson CE, Morgan BA, Tabin C. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. Development. 1995; 121:3163–3174. [PubMed: 7588051]

Robinow M, Shaw A. The McKusick-Kaufman syndrome: recessively inherited vaginal atresia, hydrometrocolpos, uterovaginal duplications, anorectal anomalies, postaxial polydactyly, and congenital heart disease. J Pediatr. 1979; 94:776–778. [PubMed: 448491]

Rouskas K, Paletas K, Kalogeridis A, Sarigianni M, Ioannidou-Papagiannaki E, Tsapas A, Kouvatsi A. Association between BBS6/MKKS gene polymorphisms, obesity and metabolic syndrome in the Greek population. Int J Obes (Lond). 2008; 32:1618–1625. [PubMed: 18813213]

Schwoebel MG, Hirsig J, Schinzel A, Stauffer UG. Familial incidence of congenital anorectal anomalies. J Pediatr Surg. 1984; 19:179–182. [PubMed: 6726575]

Sen S, Talukdar I, Webster NJ. SRp20 and CUG-BP1 modulate insulin receptor exon 11 alternative splicing. Mol Cell Biol. 2009; 29:871–880. [PubMed: 19047369]

Seo S, Guo DF, Bugge K, Morgan DA, Rahmouni K, Sheffield VC. Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling. Hum Mol Genet. 2009; 18:1323–1331. [PubMed: 19150989]

Spouge D, Baird PA. Imperforate anus in 700,000 consecutive liveborn infants. Am J Med Genet. 1986; Suppl 2:151–161.

Stoll C, Alembik Y, Dott B, Roth MP. Associated malformations in patients with anorectal anomalies. Eur J Med Genet. 2007; 50:281–290. [PubMed: 17572165]

Stone DL, Slavotinek A, Bouffard GG, Banerjee-Basu S, Baxevanis AD, Barr M, Biesecker LG. Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. Nat Genet. 2000; 25:79–82. [PubMed: 10802661]

Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET. Patterns of cis regulatory variation in diverse human populations. PLoS Genet. 2012; 8:e1002639. [PubMed: 22532805]

Szumska D, Pieles G, Essalmani R, Bilski M, Mesnard D, Kaur K, Franklyn A, El Omari K, Jefferis J, Bentham J, Taylor JM, Schneider JE, Arnold SJ, Johnson P, Tymowska-Lalanne Z, Stammers D, Clarke K, Neubauer S, Morris A, Brown SD, Shaw-Smith C, Cama A, Capra V, Ragoussis J, Constam D, Seidah NG, Prat A, Bhattacharya S. VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5. Genes Dev. 2008; 22:1465–1477. [PubMed: 18519639]

van Rooij IA, Wijers CH, Rieu PN, Hendriks HS, Brouwers MM, Knoers NV, de Blaauw I, Roeleveld N. Maternal and paternal risk factors for anorectal malformations: a Dutch case-control study. Birth Defects Res A Clin Mol Teratol. 2010; 88:152–158. [PubMed: 20073076]

Vokes SA, Ji H, McCuine S, Tenzen T, Giles S, Zhong S, Longabaugh WJ, Davidson EH, Wong WH, McMahon AP. Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development. 2007; 134:1977–1989. [PubMed: 17442700]

Waller DK, Shaw GM, Rasmussen SA, Hobbs CA, Canfield MA, Siega-Riz AM, Gallaway MS, Correa A. National Birth Defects Prevention Study. Prepregnancy obesity as a risk factor for structural birth defects. Arch Pediatr Adolesc Med. 2007; 161:745–750. [PubMed: 17679655]

Weinstein ED. Sex-linked imperforate anus. Pediatrics. 1965; 35:715–718. [PubMed: 14271507]

Werner T. Finding and decrypting of promoters contributes to the elucidation of gene function. In Silico Biol. 2002; 2:249–255. [PubMed: 12542410]

Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010; 26:2474–2476. [PubMed: 20702402]

Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ, Wang HH, Yao A, Chen YT, Hsu CN. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. Nucleic Acids Res. 2006; 34:W635–W641. [PubMed: 16845089]

Zhang J, Zhang ZB, Gao H, Zhang D, Wang WL. Down-regulation of SHH/BMP4 signalling in human anorectal malformations. J Int Med Res. 2009; 37:1842–1850. [PubMed: 20146882]

Zwink N, Jenetzky E, Brenner H. Parental risk factors and anorectal malformations: systematic review and meta-analysis. Orphanet J Rare Dis. 2011; 6:25. [PubMed: 21586115]

Table 1

SNPs in candidate genes for anorectal atresia

Gene symbol; name; location	Gene function	SNP	Minor allele frequency[a]	SNP location	Predicted functional element at SNP location
EPHB2; EPH receptor B2; 1p36.1-p35	Ephrin ligand-receptor interactions regulate diverse processes including axon guidance and vascular system morphogenesis	rs12723359:G>A	0.467	intron	TF binding site for NR6A1
		rs4654817:C>T	0.190	intron	TF binding site for SRF
		rs2675494:C>T	0.479	intron	TF binding site for MEF2A
WNT3A; Wingless-type MMTV integration site family, member 3A; 1q42	Signaling protein involved in regulation of cell fate and patterning during embryogenesis	rs12401893:G>A	0.197	intron	CpG window (DNA methylation)
GLI2; GLI family zinc finger 2; 2q14	Transcription factor and mediator of sonic hedgehog signaling	rs3738880:A>C	0.439	exon (p.A1156S)[b]	TF binding site for PAX5
HOXD12; Homeobox D12; 2q31.1	Transcription factor	rs35817516:G>A	0.173[c]	exon (p.R186Q)[b]	TF binding site for PAX5
HOXD13; Homeobox D13; 2q31.1	Transcription factor	rs35290213:A>C	0.039[d]	exon (p.S252A)[b]	TF binding site for MEF2A
WNT5A; Wingless-type MMTV integration site family, member 5A; 3p21-p14	Signaling protein involved in regulation of many developmental processes that rely on cellular migration	rs1047898:A>G	0.494	3'UTR	TF binding site
		rs815541:C>G	0.115	intron	CpG window (DNA methylation)
EPHB3; EPH receptor B3; 3q21-qter	Ephrin ligand-receptor interactions regulate diverse processes including axon guidance and vascular system morphogenesis	rs6797724:G>T	0.136[e]	intron	TF binding site for FOXC1, LHX3
		rs9862375:G>A	0.108[f]	exon (synonymous)	Exonic splice enhancer/silencer
		rs7652597:T>C	0.349	exon (synonymous)	Splice site disruption
		rs11719912:G>A	0.117[f]	exon (synonymous)	TF binding site for ELK1
		rs9881589:G>A	0.167	exon (synonymous)	TF binding site for NF1

Gene symbol; name; location	Gene function	SNP	Minor allele frequency[a]	SNP location	Predicted functional element at SNP location
FGF10; Fibroblast growth factor 10; 5p13-p12	Signaling protein involved in regulation of cell division	rs2330544:T>G	0.492	intron	TF binding site for GATA1
		rs12523512:C>G	0.237	intron	TF binding site for GATA1, LMO2
GLI3; GLI family zinc finger 3; 7p13	Transcription factor and mediator of sonic hedgehog signaling	rs4364531:A>G	0.440	intron	TF binding site for IRF1
		rs846265:G>T	0.386	intron	TF binding site for HLF
		rs7793034:A>G	0.479	intron	TF binding site
		rs10951666:C>T	0.319	intron	TF binding site for PBX1, ARNT
		rs1125413:T>G	0.282	intron	TF binding site for FOXL1
		rs3801189:C>T	0.440	intron	TF binding site for POU3F2
		rs17810462:A>G	0.134	intron	TF binding site for FOXJ2
		rs3801228:A>G	0.238	intron	TF binding site for TBP
		rs3801232:C>T	0.384	intron	TF binding site for YY1, TLX2
HOXA13; Homeobox A13; 7p15.2	Transcription factor	rs2189239:C>T	0.144	3’UTR	miRNA target site for miR-488, miR-520a, miR-525
SHH; Sonic hedgehog; 7q36	Signaling protein involved in regulation of embryonic patterning and morphogenesis	rs45611433:G>A	0.123	5’ near gene	TF binding site for GATA family
		rs7782709:C>G	0.076[e]	5’ near gene	TF binding site for EP300
		rs7782892:C>G	0.126	5’ near gene	TF binding site for MZF1
MNX1; Motor neuron and pancreas homeobox 1; 7q36	Transcription factor	rs10262191:C>A	0.184	5’ near gene	TF binding site for PBX1, SRY
PCSK5; Proprotein convertase subtilisin/kexin type 5; 9q21.3	Member of subtilisin-like proprotein convertase family involved in cleavage of precursor proteins that include precursors of growth factors, receptors, polypeptide hormones, adhesion molecules, proteases, as well as cell surface proteins of infectious viruses and bacteria	rs7020560:G>A	0.148	exon (synonymous)	TF binding site for AHR
		rs7040769:T>C	0.243	exon (synonymous)	Exonic splice enhancer
		rs872189:C>T	0.374	intron	TF binding site for POU2F1
		rs12005917:T>C	0.228	intron	TF binding site for MEF2A

Gene symbol; name; location	Gene function	SNP	Minor allele frequency[a]	SNP location	Predicted functional element at SNP location
		rs1571790:T>A	0.436	intron	TF binding site for NR3C1
		rs7872060:G>A	0.449	intron	TF binding site for ALX1
		rs3824474:A>G	0.185	intron	TF binding site for E2F1
		rs10781342:C>T	0.153	intron	TF binding site for MAX
		rs2279659:C>T	0.242	intron	TF binding site for NKX2-2
		rs10521468:C>T	0.133	exon (synonymous)	Exonic splice enhancer
		rs2643325:C>T	0.239	exon (p.G1090S)[b]	TF binding site for TFAP4
		rs1110223:A>G	0.323	exon (p.K1320E)[b]	TF binding site for MYC, MAX
		rs3001772:C>T	0.490	exon (p.T1343M)[b]	TF binding site for PAX6
		rs2495207:C>T	0.322	exon (p.R1366H)[b]	TF binding site for NR3C1
FGFR2; Fibroblast growth factor receptor 2; 10q26	Part of a signaling cascade that influences cell division and differentiation	rs1047100:G>A	0.177	exon (synonymous)	Exonic splice enhancer
GDF11; Growth differentiation factor 11; 12q13.2	Regulation of cell growth and differentiation	rs2462936:C>T	0.126	5’ near gene	TF binding site for ZBTB6
		rs7068:A>G	0.247	3’ near gene	TF binding site for HSF1, MYT1, HLTF
HNF1A (also known as *TCF1*); HNF1 homeobox A; 12q24.2	Transcription factor	rs1169289:G>C	0.412	exon (synonymous)	Exonic splice enhancer
		rs2464196:G>A	0.318	exon (p.S487N)[b]	TF binding site for SRF
		rs1169310:G>A	0.362	3’UTR	miRNA target site for miR-640
EFNB2; Ephrin-B2; 13q33	Ephrin ligand-receptor interactions regulate diverse processes including axon guidance and vascular system morphogenesis	rs9301143:C>T	0.335	5’ near gene	TF binding site for POU3F1

Gene symbol; name; location	Gene function	SNP	Minor allele frequency[a]	SNP location	Predicted functional element at SNP location
BMP4; Bone morphogenetic protein 4; 14q22-q23	Cell-cell signaling molecule required for numerous developmental processes	rs17563:T>C	0.347	exon (p.V152A)[b]	Exonic splice silencer
UBR1; Ubiquitin protein ligase E3 component n-recognin 1; 15q13	Recognition of proteins targeted for degradation though the ubiquitin system	rs3917223:A>G/T	0.119[c]	exon (p.T1548A or p.T1548S)[b]	Exonic splice silencer
SALL1; Sal-like 1 (Drosophila); 16q12.1	Transcription factor	rs1465338:A>G	0.457	intron	TF binding site for ARNT
		rs11645288:G>A	0.141	exon (synonymous)	Exonic splice enhancer
TCF4; Transcription factor 4; 18q21.1	Transcription factor	rs8766:A>G	0.372	exon (synonymous)	Exonic splice enhancer
		rs1261076:G>A	0.433	intron	TF binding site for SREBF1
		rs8094490:C>T	0.138	intron	TF binding site
		rs3794894:G>T	0.170	intron	TF binding site for ZEB1
		rs1660237:T>C	0.374	intron	TF binding site for GATA1
		rs2958162:T>C	0.484	intron	TF binding site for MYC, MAX, USF1
		rs12956276:G>A	0.303	intron	TF binding site for FOXF2, FOXD1
MKKS; McKusick-Kaufman syndrome; 20p12	Protein processing in embryogenesis	rs1545:C>A	0.180	exon (p.G532V)[b]	Exonic splice enhancer
		rs17852625:G>A	0.180	exon (synonymous)	Exonic splice silencer
		rs2013178:T>A	0.324	5’ near gene	TF binding site for ATF1
		rs1003994:G>A	0.243	5’ near gene	TF binding site for TP53
SALL4; Sal-like 4 (Drosophila); 20q13.2	Transcription factor	rs17802735:C>G	0.184[c]	exon (synonymous)	Exonic splice enhancer
		rs6021437:T>C	0.335	exon (synonymous)	Exonic splice enhancer
		rs6126344:A>C	0.316	exon (p.L507R)[b]	Exonic splice enhancer

Gene symbol; name; location	Gene function	SNP	Minor allele frequency[a]	SNP location	Predicted functional element at SNP location
		rs13038893:C>T	0.246	exon (synonymous)	Exonic splice enhancer
		rs6096585:G>A	0.226	intron	CpG window (DNA methylation)
PQBP1; Polyglutamine binding protein 1; Xp11.23	Nuclear polyglutamine-binding protein that regulates transcription	rs741932:T>C	0.469[c]	intron	Splice site disruption
ZIC3; Zic family member 3; Xq26.2	Transcription factor	rs5931174:T>C	0.323[c]	3' near gene	TF binding site for CDX1

[a]Based on 1000 Genomes project, unless otherwise noted

[b]GenBank reference sequences for encoded proteins were NP_005261.2 for *GLI2*, NP_067016.3 for *HOXD12*, NP_001177411.1 for *PCSK5*, NP_000536.5 for *HNF1A*, NP_001193.2 for *BMP4*, , NP_061336.1 for *MKKS*, NP_065169.1 for *SALL4*, NP_000514.2 for *HOXD13*, and NP_777576.1 for *UBR1*

[c]Minor allele frequency based on HapMap European (CEU) population

[d]Minor allele frequency based on population that includes individuals of European and African ancestry

[e]Minor allele frequency based on pilot data for the Yoruban (YRI) population in the 1000 Genomes project

[f]Minor allele frequency based on pilot data for the European (CEU) population in the 1000 Genomes project TF, transcription factor; UTR, un-translated region

Table 2

Comparison of characteristics between anorectal atresia cases and controls

Characteristic	Cases (N = 150)	Controls (N = 623)	*P* value[a]
	N (%)	N (%)	
Maternal age (years)			0.41
<20	8 (5.3)	50 (8.0)	
20-34	108 (72.0)	453 (72.7)	
≥35	34 (22.7)	120 (19.3)	
Maternal race/ethnicity			0.99
White, non-Hispanic	81 (54.0)	340 (54.6)	
African-American	17 (11.3)	68 (10.9)	
Hispanic	39 (26.0)	163 (26.2)	
Asian	12 (8.0)	48 (7.7)	
Other	1 (0.7)	4 (0.6)	
Maternal education (years)			0.75
<12	28 (18.7)	114 (18.3)	
12	44 (29.3)	173 (27.8)	
>12	72 (48.0)	328 (52.6)	
Missing	6 (4.0)	8 (1.3)	
Parity			0.066
Nulliparous	73 (48.7)	251 (40.3)	
Multiparous	77 (51.3)	372 (59.7)	
Maternal smoking during pregnancy			0.19
Yes	21 (14.0)	63 (10.1)	
No	128 (85.3)	560 (89.9)	
Missing	1 (0.7)	0 (0.0)	
Maternal pre-pregnancy diabetes			0.66
Yes	2 (1.3)	6 (1.0)	
No	148 (98.7)	617 (99.0)	
Gestational diabetes			0.64
Yes	7 (4.7)	24 (3.8)	
No	143 (95.3)	599 (96.2)	
In vitro fertilization or other assisted reproductive technique			1.00
Yes	2 (1.3)	10 (1.6)	
No	148 (98.7)	613 (98.4)	
Plurality			0.57
Singleton	145 (96.7)	608 (97.6)	
Multiple birth	5 (3.3)	15 (2.4)	
Infant sex			0.014

Characteristic	Cases (N = 150)	Controls (N = 623)	*P* value[a]
	N (%)	N (%)	
Male	89 (59.3)	300 (48.2)	
Female	61 (40.7)	323 (51.8)	
Birth year			0.82
1998	12 (8.0)	75 (12.0)	
1999	21 (14.0)	71 (11.4)	
2000	22 (14.7)	78 (12.5)	
2001	18 (12.0)	71 (11.4)	
2002	21 (14.0)	85 (13.6)	
2003	18 (12.0)	87 (14.0)	
2004	23 (15.3)	87 (14.0)	
2005	15 (10.0)	69 (11.1)	

[a]Fisher's exact test used to compare characteristics between cases and controls

Table 3

P values for associations between SNPs in candidate genes and anorectal atresia[a]

SNP	All subjects (N = 773)	Non-Hispanic white (N = 421)	African-American (N = 85)	Hispanic (N = 202)	Asian (N = 60)
EPHB2 rs12723359:G>A	0.98	0.61	0.47	0.91	0.61
EPHB2 rs4654817:C>T	0.28	0.47	0.12	0.82	0.89
EPHB2 rs2675494:C>T	0.38	0.37	0.95	0.96	0.29
WNT3A rs12401893:G>A	0.61	0.67	0.031	0.88	0.82
GLI2 rs3738880:A>C	0.31	0.92	0.020	0.074	0.67
HOXD12 rs35817516:G>A	0.25	0.96	0.61	0.020	0.99
WNT5A rs1047898:A>G	0.52	0.78	0.55	0.64	0.87
WNT5A rs815541:C>G	0.97	0.99	0.31	0.35	0.98
EPHB3 rs6797724:G>T	0.92	0.23	0.56	0.65	0.97
EPHB3 rs9862375:G>A	0.10	0.96	0.65	0.22	0.99
EPHB3 rs7652597:T>C	0.89	0.53	0.87	0.99	0.77
EPHB3 rs11719912:G>A	0.36	0.99	0.42	0.35	0.99
EPHB3 rs9881589:G>A	0.30	0.29	0.94	0.85	0.90
FGF10 rs2330544:T>G	0.62	0.54	0.68	0.47	0.72
FGF10 rs12523512:C>G	0.60	0.47	0.82	0.22	0.55
GLI3 rs4364531:A>G	0.56	0.37	0.61	0.47	0.94
GLI3 rs846265:G>T	0.80	0.99	0.92	0.10	0.29
GLI3 rs7793034:A>G	0.93	0.81	0.16	0.71	0.66
GLI3 rs10951666:C>T	0.33	0.42	0.48	0.38	0.47
GLI3 rs1125413:T>G	0.77	0.98	0.29	0.60	0.24
GLI3 rs3801189:C>T	0.93	0.48	0.51	0.96	0.16
GLI3 rs17810462:A>G	0.39	0.47	0.51	0.82	0.37
GLI3 rs3801228:A>G	0.15	0.093	0.66	0.51	0.45
GLI3 rs3801232:C>T	0.13	0.40	0.30	0.11	0.42
HOXA13 rs2189239:C>T	0.31	0.21	0.81	0.93	0.98

SNP	All subjects (N = 773)	Non-Hispanic white (N = 421)	African-American (N = 85)	Hispanic (N = 202)	Asian (N = 60)
SHH rs45611433:G>A	0.22	0.99	0.99	0.11	0.45
SHH rs7782709:C>G	0.32	0.13	0.81	0.58	0.99
SHH rs7782892:C>G	0.61	0.51	0.73	0.70	0.99
MNX1 rs10262191:C>A	0.99	0.91	0.18	0.91	0.28
PCSK5 rs7020560:G>A	0.97	0.96	0.55	0.69	0.99
PCSK5 rs7040769:T>C	0.080	0.046	0.68	0.24	0.99
PCSK5 rs872189:C>T	0.22	0.52	0.034	0.71	0.43
PCSK5 rs12005917:T>C	0.32	0.56	0.99	0.56	0.85
PCSK5 rs1571790:T>A	0.28	0.095	0.30	0.60	0.32
PCSK5 rs7872060:G>A	0.20	0.31	0.31	0.42	0.56
PCSK5 rs3824474:A>G	0.29	0.18	0.81	0.92	0.72
PCSK5 rs10781342:C>T	0.99	0.99	0.76	0.89	0.61
PCSK5 rs2279659:C>T	0.56	0.82	0.043	0.77	0.79
PCSK5 rs10521468:C>T	0.95	0.49	0.30	0.74	0.84
PCSK5 rs2643325:C>T	0.98	0.12	0.36	0.37	0.43
PCSK5 rs1110223:A>G	0.54	0.30	0.095	0.67	0.73
PCSK5 rs3001772:C>T	0.83	0.70	0.30	0.70	0.93
PCSK5 rs2495207:C>T	0.70	0.23	0.20	0.97	0.78
FGFR2 rs1047100:G>A	0.59	0.88	0.47	0.68	0.99
GDF11 rs2462936:C>T	0.34	0.29	0.31	0.24	0.99
GDF11 rs7068:A>G	0.30	0.15	0.71	0.77	0.99
HNF1A rs1169289:G>C	0.49	0.35	0.66	0.10	0.47
HNF1A rs2464196:G>A	0.85	0.73	0.61	0.33	0.32
HNF1A rs1169310:G>A	0.96	0.97	0.58	0.19	0.65
EFNB2 rs9301143:C>T	0.57	0.24	0.91	0.55	0.70
BMP4 rs17563:T>C	0.47	0.48	0.59	0.58	0.033
SALL1 rs1465338:A>G	0.59	0.71	0.74	0.89	0.41
SALL1 rs11645288:G>A	0.89	0.12	0.76	0.58	0.74

SNP	All subjects (N = 773)	Non-Hispanic white (N = 421)	African-American (N = 85)	Hispanic (N = 202)	Asian (N = 60)
TCF4 rs8766:A>G	0.66	0.044	0.80	0.30	0.46
TCF4 rs1261076:G>A	0.91	0.092	0.98	0.061	0.43
TCF4 rs8094490:C>T	0.94	0.29	0.94	0.42	0.99
TCF4 rs3794894:G>T	0.71	0.054	0.60	0.43	0.62
TCF4 rs1660237:T>C	0.43	0.33	0.69	0.33	0.60
TCF4 rs2958162:T>C	0.11	0.84	0.19	0.20	0.69
TCF4 rs12956276:G>A	0.74	0.84	0.97	0.89	0.31
MKKS rs1545:C>A	0.85	0.86	0.38	0.61	0.92
MKKS rs17852625:G>A	0.89	0.86	0.55	0.58	0.95
MKKS rs2013178:T>A	0.095	0.0015	0.15	0.028	0.70
MKKS rs1003994:G>A	0.14	0.0078	0.62	0.31	0.73
SALL4 rs17802735:C>G	0.53	0.76	0.73	0.75	0.53
SALL4 rs6021437:T>C	0.33	0.62	0.98	0.31	0.75
SALL4 rs6126344:A>C	0.55	0.65	0.60	0.46	0.87
SALL4 rs13038893:C>T	0.55	0.65	0.56	0.77	0.34
SALL4 rs6096585:G>A	0.55	0.74	0.89	0.46	0.49
PQBP1 rs741932:T>C					
Male	0.82	0.49	0.68	0.11	0.11
Female	0.44	0.10	0.10	0.93	0.32
ZIC3 rs5931174:T>C					
Male	0.25	0.88	0.42	0.72	0.083
Female	0.67	0.90	0.16	0.21	0.95

[a]Logistic regression used to calculate P values from two degree-of-freedom tests (for variants on autosomes and the × chromosome in females) and one degree-of-freedom tests (for variants on the × chromosome in males); all models adjusted for parity and maternal smoking; models that include all study subjects also adjusted for race/ethnicity

Table 4

Anorectal atresia and the interaction between *MKKS* SNPs in the offspring and maternal pre-pregnancy obesity[a]

Cases (N)	Controls (N)	*MKKS* rs2013178 TT genotype in offspring	Maternal pre-pregnancy obesity[b]	Odds ratio (95% confidence interval)[c]	Interaction contrast ratio (95% confidence interval)[de]
All subjects					
42	182	No	No	Ref	0.59 (−0.27, 1.90)
25	81	Yes	No	1.29 (0.73, 2.29) *P*=0.39	
4	49	No	Yes	0.38 (0.12, 1.18) *P*=0.094	
8	26	Yes	Yes	1.69 (0.68, 4.16) *P*=0.26	
Non-Hispanic white					
29	151	No	No	Ref	0.49 (−1.14, 3.05)
20	49	Yes	No	2.15 (1.11, 4.15) *P*=0.023	
3	39	No	Yes	0.43 (0.13, 1.52) *P*=0.15	
7	16	Yes	Yes	2.37 (0.89, 6.33) *P*=0.085	
		***MKKS* rs1003994 GG genotype in offspring**			
All subjects					
38	162	No	No	Ref	0.60 (−0.20, 1.80)
29	103	Yes	No	1.20 (0.68, 2.11) *P*=0.54	
2	43	No	Yes	0.26 (0.06, 1.16) *P*=0.078	
10	32	Yes	Yes	1.49 (0.64, 3.50) *P*=0.36	
Non-Hispanic white					
27	140	No	No	Ref	0.66 (−0.79, 3.21)

Cases (N)	Controls (N)	*MKKS* rs2013178 TT genotype in offspring	Maternal pre-pregnancy obesity[b]	Odds ratio (95% confidence interval)[c]	Interaction contrast ratio (95% confidence interval)[de]
22	62	Yes	No	1.80 (0.94, 3.43) *P*=0.076	
2	37	No	Yes	0.30 (0.07, 1.35) *P*=0.10	
8	18	Yes	Yes	2.45 (0.95, 6.31) *P*=0.063	

[a]Analyses included 79 cases and 340 controls that had data available on maternal pre-pregnancy body mass index and other covariates used in the regression analyses

[b]Maternal pre-pregnancy body mass index ≥30 kg/m^2

[c]Logistic regression models adjusted for parity and maternal smoking; models that included all subjects were also adjusted for maternal race/ethnicity

[d]As described by Richardson & Kaufman (2009), the interaction contrast ratio was calculated from the product term for interaction in a linear odds ratio model; confidence intervals were based on the likelihood ratio

[e]Models for calculation of the interaction contrast ratio and 95% confidence interval were adjusted for parity and maternal smoking; models that included all subjects were also adjusted for maternal race/ethnicity