Published in final edited form as:

Hum Genet. 2009 December ; 126(6): 763–778. doi:10.1007/s00439-009-0724-8.

Modifier gene study of meconium ileus in cystic fibrosis: statistical considerations and gene mapping results

Ruslan Dorfman,

Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada
ruslan.dorfman@gmail.com

Weili Li,

Child Health Evaluative Sciences Program, Hospital for Sick Children, 555 University Avenue,
Toronto, ON M5G 1X8, Canada

Lei Sun,

Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada

Fan Lin,

Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada

Yongqian Wang,

Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada

Andrew Sandford,

The James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, University of
British Columbia, Vancouver, Canada

Peter D. Paré,

The James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, University of
British Columbia, Vancouver, Canada

Karen McKay,

Department of Respiratory Medicine, The Children's Hospital at Westmead, Westmead, NSW,
Australia

Hana Kayserova,

Institute of Molecular Physiology and Genetics, Slovak Academy of Sciences, Bratislava, Slovak
Republic

Tereza Piskackova,

Department of Biology and Medical Genetics, University Hospital Motol, 2nd School of Medicine,
Charles University Prague, Prague, Czech Republic

Milan Macek,

Department of Biology and Medical Genetics, University Hospital Motol, 2nd School of Medicine,
Charles University Prague, Prague, Czech Republic

© Springer-Verlag 2009

lisa.strug@utoronto.ca.

Web resources Mendelian inheritance in man listed in online <http://www.ncbi.nlm.gov/entrez>. The Haploview application can be downloaded at <http://www.broad.mit.edu/mpg/haploview>. Information about marker location can be found at UniSTS at the NCBI website above, and through the Ensembl genome browser at http://www.ensembl.org/Homo_sapiens/index.html. SNP frequencies were accessed at dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP>. TCAG Copy Number Variation database <http://projects.tcag.ca/variation/>.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-009-0724-8) contains supplementary material, which is available to authorized users.

Kamila Czerska,
Institute of Mother and Child, Warsaw, Poland

Dorota Sands,
Institute of Mother and Child, Warsaw, Poland

Harm Tiddens,
Erasmus Medical Centre Rotterdam, Sophia Children's Hospital, Rotterdam, The Netherlands

Sonia Margarit,
Human Genetics Center, Clínica Alemana, Universidad del Desarrollo de Santiago, Santiago, Chile

Gabriela Repetto,
Human Genetics Center, Clínica Alemana, Universidad del Desarrollo de Santiago, Santiago, Chile

Marci K. Sontag,
Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA

Frank J. Accurso,
Department of Pediatrics, Mike McMorris Cystic Fibrosis Care and Research Center, Children's Hospital, University of Colorado Denver, Aurora, CO 80045, USA

Scott Blackman,
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Garry R. Cutting,
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Lap-Chee Tsui,
University of Hong Kong, Hong Kong, China

Mary Corey,
Child Health Evaluative Sciences Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Peter Durie,
Program in Physiology and Experimental Medicine, Hospital for Sick Children, Toronto, ON, Canada

Julian Zielenski, and
Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada

Lisa J. Strug
Child Health Evaluative Sciences Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Abstract

Cystic fibrosis (CF) is a monogenic disease due to mutations in the *CFTR* gene. Yet, variability in CF disease presentation is presumed to be affected by modifier genes, such as those recently demonstrated for the pulmonary aspect. Here, we conduct a modifier gene study for meconium ileus (MI), an intestinal obstruction that occurs in 16–20% of CF newborns, providing linkage and association results from large family and case-control samples. Linkage analysis of modifier traits

is different than linkage analysis of primary traits on which a sample was ascertained. Here, we articulate a source of confounding unique to modifier gene studies and provide an example of how one might overcome the confounding in the context of linkage studies. Our linkage analysis provided evidence of a MI locus on chromosome 12p13.3, which was segregating in up to 80% of MI families with at least one affected offspring (HLOD = 2.9). Fine mapping of the 12p13.3 region in a large case-control sample of pancreatic insufficient Canadian CF patients with and without MI pointed to the involvement of *ADIPOR2* in MI ($p = 0.002$). This marker was substantially out of Hardy-Weinberg equilibrium in the cases only, and provided evidence of a cohort effect. The association with rs9300298 in the *ADIPOR2* gene at the 12p13.3 locus was replicated in an independent sample of CF families. A protective locus, using the phenotype of no-MI, mapped to 4q13.3 (HLOD = 3.19), with substantial heterogeneity. A candidate gene in the region, *SLC4A4*, provided preliminary evidence of association ($p = 0.002$), warranting further follow-up studies. Our linkage approach was used to direct our fine-mapping studies, which uncovered two potential modifier genes worthy of follow-up.

Introduction

Cystic fibrosis (CF) (MIM 219700) is an autosomal recessive monogenic disorder clinically characterized by pancreatic insufficiency (PI), progressive pulmonary disease, and elevated levels of chloride in sweat. In some CF newborn cases, typically with two severe mutations in the *CFTR* gene, intestinal obstruction due to meconium ileus (MI) is observed. MI develops in utero with intestinal obstruction in the distal ileum and cecum, and presents at birth with failure to pass meconium and vomiting. Immediate surgical intervention is required if there is evidence of intrauterine complications such as perforation, volvulus, or ileal atresia. In subjects with uncomplicated MI, administration of an enema may be sufficient to relieve the obstruction. If the enema fails to relieve the obstruction, surgical intervention is required. MI is an early clue to the diagnosis of CF because it occurs in about 15–20% of CF patients at birth, and is infrequent in newborns without CF (supplementary Scheme 1).

Similar intestinal obstruction is observed in *CFTR* knockout mice at birth, with the severity of obstruction depending on the genetic background (Rozmahel et al. 1996). Thus, it was established that MI is dependent on secondary genetic factors. Indeed, genetic and epidemio-logical studies in CF twins and siblings indicate high concordance rates of MI among monozygotic CF twins versus CF siblings and dizygotic twins (Blackman et al. 2006). Despite the high heritability of MI, the genetic basis of this complication has not yet been identified. This study aimed to (1) identify the genetic determinants of MI via linkage and fine-mapping studies, and (2) highlight statistical difficulties in the identification of modifier genes. Statistical approaches to the identification of modifier genes must be chosen with some care, as mapping of modifier genes typically involves selected samples ascertained previously based on the primary phenotype.

There have been a number of modifier gene studies of MI in CF. The first study in 1996, conducted on *CFTR* knockout mice, demonstrated that genetic background modifies both severity of intestinal disease and survival of CF knockout mice (Rozmahel et al. 1996). A genome-wide study on CD1 and 129/Sv congenic *CFTR*^{-/-} mice led to the identification of a highly significant locus (*CF Modifier1*, *Cfm1*) for severity of CF disease (neonatal survival to 12 weeks) on proximal chromosome 7. A follow-up linkage analysis of STR markers in the human orthologous region (19q13) used CF sibling pairs concordant or discordant for MI, and demonstrated that the *CFM1* locus is linked to MI in humans as well (Zielenski et al. 1999). A fine-mapping study of 98 SNPs genotyped in the *CFM1* region in Canadian CF patients detected an association with 4 markers in a zinc-finger (ZNF) gene cluster, but it

could not be replicated in an independent patient cohort (Zielenski et al. 2005). Two other studies of CF mouse models, analyzing survival of mice with intestinal obstruction, suggested several additional loci associated with a severe or mild intestinal phenotype (Haston and Tsui 2003; Norkina and De Lisle 2005).

A recent genome-wide linkage analysis of CF siblings, focusing on non-parametric approaches, did not find any evidence for linkage to 19q13 or other major loci for MI (Blackman et al. 2006). The study by Blackman et al. (2006) found suggestive loci for MI and no-MI, with multipoint non-parametric LOD scores 1.96, 2.14 and 2.18 at 4q35.1, 8p23.1 and 11q25, respectively, and they concluded from these results that multiple loci with small effects are the determinants of MI.

Here, we discuss how the choice of statistical methodology for linkage analysis of modifier traits should be based on different criteria than those for linkage studies of complex or Mendelian diseases. Although various authors have recognized the importance of alternative study designs when searching for modifier genes (Daw et al. 2008; Houlston and Tomlinson 1998), the problem has not been clearly articulated, with different approaches discussed and contrasted. Here, we provide an example of how the analysis of modifier gene studies can be complicated by their ascertainment scheme, and thus advise investigators to be cognizant of the potential problem.

One central difference between complex disease linkage studies and modifier gene studies is related to the study design: in modifier gene studies, siblings are generally ascertained on the primary trait, e.g., CF or Type I diabetes, while linkage analysis is performed on a secondary or modifying trait such as MI, in the case of CF, or nephropathy in Type I diabetes (Rogus et al. 2008). Therefore, if analysis is conducted only in those CF patients affected for the trait of interest (MI) as in standard non-parametric linkage (NPL) analyses, observed linkage evidence could be due to a confounding phenotype because subjects are also affected for the primary trait (CF). See results of “Confounding in modifier gene studies of MI” for a more in-depth discussion.

Here, we report results from linkage analysis of the MI and no-MI phenotypes, searching for a protective locus in the latter case (“Parametric linkage analysis of MI and no-MI”); we contrast the results of parametric and NPL analysis for this MI modifier study, providing an approach to overcome the limitations of NPL in this situation (“Confounding in modifier gene studies of MI”); we report fine-mapping results from under an MI linkage peak (“Association studies of the MI linkage regions”), and present association results at a candidate gene near a no-MI linkage peak (“Association analysis of SLC4A4 in the no-MI linkage region”).

Materials and methods

Subjects

Cystic fibrosis patients and their family members were recruited from 37 specialized CF clinics across Canada and selected international clinics (Australia, Chile, the Netherlands, Poland, and USA). The protocols for this study were approved by ethical review boards at the Hospital for Sick Children and all participating institutions. Informed consent was obtained from each individual or his/her parent or guardian. The Canadian CF modifier study population consists of 2,441 patients diagnosed with CF between 1951 and 2006, and 3,092 parents. The linkage analyses were based on families with at least two affected CF siblings from the Canadian CF modifier study and selected international clinics. Patients were classified as having the phenotype of MI if the clinic from which they were recruited reported the presence of MI at birth from chart review, and no-MI otherwise. The linkage

analysis for the phenotype of no-MI included 226 families with PI CFTR mutations (pancreatic sufficiency (PS) is rare in individuals with MI; Ahmed et al. 2003; Kristidis et al. 1992) who had at least one child with no-MI at birth. For the purposes of this study, exocrine pancreatic function status was based on CFTR mutations (Ahmed et al. 2003). PI status was assigned if both CFTR mutations were known to be associated with PI. PS patients carried at least one known pancreatic sufficient mutation. Otherwise, those with CFTR mutations which were unidentified or not clearly associated with PS or PI were classified as PS or PI based on the reported clinical use of pancreatic enzymes. The MI analysis included 71 families with at least one MI child at birth. The number of families in each of the two scans, by country/site of origin, is provided in Table 1, while Table 2 provides information on the CFTR mutation distribution, age, gender, and ethnicity of the sample. Those from the US were recruited through CF centers at the University of Colorado, Denver and Johns Hopkins University (JHU), Baltimore (Blackman et al. 2006).

DNA samples

Individual blood samples were collected in glass tubes containing anticoagulant (Acid Citrate Dextrose; ACD-Vacutainer tubes, BD) and maintained at room temperature. Genomic DNA was extracted using the phenol/chloroform procedure (Cutting et al. 1989). The DNA stocks were stored at 4°C. Patients were recruited during the period 2003–2006 and the age in all analyses is based on the date the blood sample was taken.

Genotyping

Short tandem repeats (STRs)—For the purpose of genome-wide linkage analysis of multiple phenotypes, 1,178 individuals from 337 families were genotyped using the deCODE 4cM STR marker panel. This panel contains 1,054 highly polymorphic STR markers. Amplified fragments were subjected to electrophoresis using ABI 3700 and ABI 3730 DNA analyzers with CEPH family DNA used as controls. Allele calls were checked for consistency with Hardy–Weinberg equilibrium (HWE) and non-paternity. Errors were reconciled by resampling or exclusion of Mendelian inconsistent genotypes.

Single nucleotide polymorphisms (SNPs) in the 12p linkage region—We conducted a fine-mapping association study on 12p, genotyping 92 tagging SNPs using the Sequenom platform with the iPLEX Gold protocol at Genome Quebec and McGill University Innovation Centre in Montreal, Canada. The 92 tagging SNPs were chosen from HapMap with $r^2 > 0.9$ and minor allele frequency (MAF) above 0.1. These were placed in the 360.5 kb region surrounding the *ADIPOR2* gene (chr12: 1549476–1910002; Mar 2006 Assembly) encompassing the *WNT5B* and *CACNA2D4* genes. The genotyping was done on 1,104 individuals: 379 unrelated MI cases, 596 unrelated PI control patients who did not have MI themselves nor a family history of MI, and 43 informative Canadian nuclear MI families who contributed positively to the linkage signal. The 379 unrelated MI cases included the probands of the 43 nuclear families.

Three Mendelian errors were found at locus rs1029628 in the family data, and the genotypes of families with Mendelian errors were reset to unknown at this locus. In the case–control sample, six individuals were removed due to a low genotyping rate (proportion of missing genotypes > 10%), and more than 10% of individuals had missing genotypes at rs2190770, which was removed from the analysis. No SNP had MAF less than 1%, and none failed HWE tests ($p < 0.001$) in the MI control sample.

ADIPOR2 replication cohort—The 100K Affymetrix array (100K) was used to assay 29 MI concordant siblings and their parents recruited by JHU, as part of the CF Twin and Sibling Study (Blackman et al. 2006). The 100K array contained two SNPs, rs9300298 and

rs1044825, of four markers that were associated and out of HWE in our Canadian cases; neither of the two SNPs had low MAF in founders (MAF = 0.44 and 0.40, respectively) nor a high rate of missing genotypes.

Copy number variation (CNV) analyses—Two CNVs in the *ADIPOR2* gene were identified through the Database of Genomic Variants hosted by The Center for Applied Genomics (<http://projects.tcag.ca/variation/>), The Hospital for Sick Children, Toronto, ON, Canada. The CNV consisting of a 315 base deletion (variation #11592, also known as rs35417183) is at chr12: 1734273–1734587 in intron 2 of the *ADIPOR2* longest splice variant NM024551, and intron 1 of the shorter alternative splice variant AY424280. An insertion of 134 bases is present (variation #24961) at chr12: 1754352–1754485 in intron 3 of the long mRNA isoform. The genotyping for the 315 bp deletion and 134 bp insertion was performed in multiplexed reactions (supplementary Figure 1) using specific primers (supplementary Table 1) in multiplexed PCR at standard cycling conditions (60°C annealing for 45 s, 1 min extension at 72°C).

Re-sequencing of the *ADIPOR2* gene in MI discordant sibs—In order to identify potential mutations in the coding regions of the *ADIPOR2* gene, we sequenced the coding exons. These included flanking genomic regions and the eighth non-coding exon in 24 sibling pairs discordant for MI.

SNPs in *SLC4A4*—Tagging SNPs for *SLC4A4* (chr4: 72423681–72656663) (supplementary Table 2) were genotyped using Illumina Golden Gate technology in 601 nuclear families. Follow-up genotyping of rs1453458 was also performed on the Luminex platform in a sample of 278 cases and 1,315 controls. None of the genotyped SNPs in the chromosome 4 region failed QC or were out of HWE in founders.

Statistical methodology

Parametric linkage analysis—We used parametric or LOD score analysis for the genome-wide linkage scan. The MLINK program of the FAST-LINK software package (Cottingham et al. 1993; Schaffer et al. 1994) was used for two-point analysis, while GENEHUNTER (Kruglyak et al. 1996) was used for multipoint. Heterogeneity LOD scores were computed in the multipoint analyses. All non-CF individuals were coded as unknown for the MI trait.

LOD score analysis requires specification of disease model parameters such as mode of inheritance, penetrance, gene frequency, and sporadic rate, despite the model generally being unknown. However, a large body of research has shown that in fact, apart from the inheritance model at the locus being studied, LOD score analysis is robust to misspecification of other parameters (Clerget-Darpoux et al. 1986; Elston 1989; Greenberg and Hodge 1989; Pal et al. 2001; Vieland et al. 1992a, b, 1993).

We performed the LOD score analysis twice, once under a dominant model and once under a recessive model. We specified a low sporadic rate of 0.0002, an arbitrary penetrance of 0.50, and arbitrary gene frequencies of 0.1 for recessive and 0.01 for dominant models. We did so following the procedure discussed by Hodge and Elston (1994), Abreu et al. (2002), Greenberg and Abreu (2001), and Hodge et al. (1997)) referred to as the maximized maximum LOD score (MMLS). MMLS specifies that one performs the analysis twice, once under dominant and once under a recessive assumption (Greenberg et al. 1998), with a low sporadic rate and arbitrary values for penetrance and gene frequencies. The MMLS approach is supported by a large body of research indicating that MMLS is an estimate of maximization over the whole parameter space, with a known approximate significance

criterion. The consequent increase in type I error is approximately a half a degree of freedom under a Chi-squared distribution (Hodge et al. 1997), a 0.3 penalty on the LOD score being a conservative penalty to assign before declaring genome-wide linkage. MMLS is a reasonable estimate of a LOD score which maximizes over the whole parameter space (i.e., a MOD score; Hodge and Elston 1994) since the values chosen for penetrance and gene frequency have minimal effect on power and type I error assuming the inheritance model at the locus is correctly specified (Greenberg 1990; Greenberg and Hodge 1989; Hodge and Elston 1994; Pal et al. 2001); this holds for both simple and complex models. Of course, the actual value of the LOD score may vary with different parameter values. The complete genetic model for a disease/trait (i.e., all the susceptibility loci and their interactions) does not need to be specified in the LOD score analysis (Dizier et al. 1996; Durner et al. 1999). Only the parameters for the locus under analysis are required, because the contribution of all other loci can be subsumed as reduced penetrance or phenocopy frequency for analysis purposes (Greenberg and Hodge 1989). However, misspecifying the disease inheritance model does lead to a significant penalty in the LOD score (Clerget-Darpoux et al. 1986; Hodge et al. 1997). This means that linkage to a dominant gene can be missed if the data are analyzed under a recessive model, and vice versa, thus explaining why in MMLS one must analyze under both a dominant and recessive model. Consequently, this method will give information about possible inheritance at that locus (Elston 1989). We used $\text{LOD} > 2$ as a criterion reflecting a locus of interest to follow-up on. This criterion was based on simulation results indicating low type I error rates for $\text{LOD} > 2$ in our 71 MI families (<http://onion.cpmc.columbia.edu/shimshon/>).

NPL analysis—We also performed non-parametric two-point and multi-point linkage analysis using GENEHUNTER. We did so as proof of principle to illustrate the difficulty in interpretation for modifier gene studies. Since standard NPL implementations include affected individuals by default, we ran a separate NPL analysis including families with siblings discordant for MI by coding the unaffected as affected, and evaluating linkage evidence via significant under-sharing. The inclusion of discordant pairs is well documented for QTL analysis [e.g., “Statistical analysis in discordant pair studies is generally done by testing for reduced identity by descent (IBD) sharing in the pairs.” (Forrest and Feingold 2000)] and the reduced IBD sharing is a direct result of Mendel’s law of segregation (Blackwelder and Elston 1985), but it has not been used for binary traits due to design issues. For example, the amount of reduced IBD sharing by discordant sibpairs is less than the amount of increased sharing by affected sibpairs under most disease models. Thus, affected sibling pairs would be chosen for collection over discordant sibling pairs (DSPs) for power considerations in a study where confounding, as we have highlighted for modifier gene studies, was not an issue.

Association analysis—For the single SNP and CNV analyses, we used PLINK (Purcell et al. 2007) and analyzed all the SNPs for association using the Cochran-Armitage trend and 2df genotypic tests; we analyzed some SNPs under recessive or dominant models when warranted. We also evaluated HWE in the cases and controls separately, and for those SNPs out of HWE in the cases, we tested for association under a recessive model. We used the method of Wang and Shete (2008) to combine evidence for association from the recessive model and HWE analysis. In the replication sample from JHU, we used FBAT Version 1.7.3 (Laird et al. 2000), assuming a genotypic 2df model and an offset = 0 to be consistent with the case-control analysis in the Canadian sample.

Multi-SNP analysis—We used multivariate logistic regression to determine whether multiple associated SNPs in a region were pointing to a single causal variant or were representing multiple independent mutations (Clayton et al. 2004). Multivariate logistic

regression was also used for interaction analysis. We used Phase 2.1.1 (Stephens and Donnelly 2003; Stephens et al. 2001) to construct haplotypes in regions where frequencies differed between cases and controls. Haploview (Barrett et al. 2005) was used to determine the pair-wise LD in the region, in our sample of controls.

Descriptive statistics were calculated using the R statistical package. Regression analysis was performed to evaluate the relationship between MI and lung function on a population-based sample of CF patients collected across Canada (Li et al. 2009), which included 285 PI MI cases and 1,225 PI MI controls. Lung function was measured as a derivative of FEV1 percent-predicted adjusted for mortality with age (Taylor et al. 2009).

Results

Results from linkage and association analyses for the MI and no-MI phenotypes are summarized in Fig. 1, and the details are provided in the following sections.

Parametric linkage analysis of MI and no-MI

In the MI scan, we included 71 nuclear families ascertained to have two CF children, 17 families were concordant for MI and 54 were discordant. In the no-MI scan, the data consisted of 226 informative nuclear families, among which 176 were concordant for no-MI, and 50 were discordant. Four of the discordant families included in the MI analysis had CFTR mutations that may be associated with pancreatic sufficiency, and thus were not eligible to be included in the no-MI analysis.

The two-point parametric linkage analysis identified two loci with LOD scores over 2.0 for the MI phenotype. For the no-MI phenotype, five markers provided linkage evidence with LOD scores greater than 2. These markers together with their LOD scores, cM positions, and flanking marker information are presented in Table 3 for both phenotypes. We followed-up these regions with multipoint analysis, calculating multipoint heterogeneity LOD scores (HLOD). The largest multipoint HLOD for MI occurred at D12S1656 (5.21 cM) with an HLOD = 2.935 (Fig. 2) under a dominant model, despite the limited linkage information available in the 71 nuclear families. There was no evidence of heterogeneity at D12S1656 ($\hat{\alpha}=1$); 54 of the 71 families (76%) contributed positive LOD scores to the overall linkage signal. In multipoint analysis for no-MI, we observed an HLOD of 3.19 at D4S2389 (cM position = 84.72) under a recessive mode of inheritance; however, there was a great deal of heterogeneity ($\hat{\alpha}=0.26$) (Fig. 3); 115 of the 226 (51%) families contributed positive linkage evidence to the overall linkage signal. 78% of all the discordant families contributed positively to this locus, while only 43% of the concordant families contributed positively, possibly pointing to a source of the heterogeneity. D4S2389 is in the GC locus; the GC protein is a reported nutritional biomarker in CF (Speeckaert et al. 2008). The two-point parametric LOD scores we obtained at 4q35.1, 8p23.1, and 11q25, the loci reported by Blackman et al. (2006) for the MI phenotype, were 0.07, 0.16, and 0, respectively. At 20p11.22 and 21q22.3, the loci reported by Blackman et al. (2006) for the no-MI phenotype, we observed parametric LODs of 0.02 and 0, respectively.

Confounding in modifier gene studies of MI

For a modifier linkage study of MI, every MI subject has CF and all CF subjects have lung disease, with quantitative measures of lung disease known to be correlated in CF siblings (Vanscoy et al. 2007). Standard NPL analysis uses affected individuals only. Thus, if NPL analysis of MI demonstrates linkage, is the marker linked to MI, to CF, to the pulmonary phenotype or another CF phenotype that is correlated between siblings? If MI and lung disease are uncorrelated, this confounding can be eliminated or greatly reduced by including

subjects unaffected with MI. This is because lung disease is correlated between CF siblings, thus reduced IBD sharing by CF siblings discordant for MI would provide evidence in favor of linkage to MI, but against linkage to lung disease.

In a large Canadian population-based sample of 1,510 PI CF patients (see “Materials and methods”), there is no evidence that lung function and MI are correlated in all PI patients ($p = 0.44$, with 285 MI and 1225 no-MI patients) or when restricted to just $\Delta F508$ homozygotes ($p = 0.90$, with 191 MI and 731 no-MI patients). Thus, MI and lung disease are uncorrelated and the discordants can provide linkage evidence for MI.

The use of discordant pairs in linkage analysis is well documented for QTL analysis (Forrest and Feingold 2000; Guo and Elston 2000; Risch and Zhang 1995) and the same principle applies to binary traits (Forrest and Feingold 2000). Thus, although the use of discordant relative pairs in NPL studies is not a novel idea, in this type of modifier gene study, it plays a critical role and informs the linkage method choice. In parametric analysis, unaffected individuals are always included, thus parametric analysis is generally robust to this confounding.

Contrasting parametric and NPL analysis of MI—Multipoint results from a genome-wide scan using NPL analysis, incorporating discordant siblings’ information into the analysis, are shown in Fig. 4 for MI and Fig. 5 for no-MI. This analysis provides proof of principle that the addition of discordants can clarify the NPL linkage information. On chromosome 7, the closest marker to CFTR, D7S522, shows significant over-sharing in both concordant MI and discordant families (where MI and no-MI have both been labeled as “affected”), indicative of linkage evidence for CF not MI. This is not surprising, because these discordant MI siblings are concordant for CF and are expected to have over-sharing in the CFTR region. All DSPs in the parametric analysis provide evidence against linkage at D7S522. At the MI-linked D12S1656 locus identified via the parametric linkage method, we observed over-sharing from concordant MI families, and under-sharing from the discordant MI families, providing corroborating evidence that D12S1656 is a susceptibility locus for MI, not CF or lung disease (Fig. 4). Likewise, at D4S2389 for no-MI we see evidence of under-sharing from the discordant MI families, although given the degree of heterogeneity observed from the parametric analysis, the power to detect this locus using NPL is reduced (Greenberg and Abreu 2001). It should be noted that there are some loci in the NPL analysis which appear to provide evidence of linkage, and these loci are not corroborated by the parametric analysis, e.g., chromosome 6 at 41.08 cM for the MI phenotype. The two-point parametric LOD score at this locus is 1.85 under the recessive model. Thus, the two methods could serve as complimentary, and this locus should also be followed-up on in future work. Here, we were able to disentangle linkage evidence from NPL for MI, versus other traits correlated in CF siblings but uncorrelated with MI, by including discordant families in the analysis. A combined NPL statistic from the two sets of families can then be calculated.

Association studies of the MI linkage regions

Fine mapping of the chromosome 12p13.3 MI linkage region—Given the novelty and size of the LOD score at D12S1656 in the 71 nuclear families for MI and the high proportion of families contributing to that LOD score, we elected to perform fine-mapping studies on chromosome 12. We targeted candidate genes under the linkage peak using a case-control analysis of unrelated PI patients with and without MI (379 MI cases, 596 controls).

Of the 92 SNPs genotyped, there were four markers out of HWE in cases only ($p < 0.001$): rs6489326 and rs9300298 in ADIPOR2, rs7955261 in the putative promoter region of ADIPOR2 and rs1044825 in the neighboring gene, CACNA2D4, with increased linkage

disequilibrium (LD) (Fig. 6). Table 4 provides the HWE statistics calculated in the independent samples of cases and controls for the four SNPs that were out of HWE in cases. In the case–control analysis, assuming an additive model, rs1003939, which is located between *WNT5B* and *ADIPOR2*, was most significantly associated with MI in the 92 tested SNPs [$p = 0.003$, OR = 0.72, 95% CI (0.58, 0.89)] (Fig. 7). However, when analyzed under a two degrees of freedom (df) genotypic model, the most significantly associated SNPs were the four SNPs out of HWE in the cases, i.e., rs1044825, rs6489326, rs9300298, and rs7955261, with $p = 0.002$, 0.005, 0.01, and 0.01, respectively (Fig. 7), driven by significance under a recessive model. The estimated odds ratio from the 2df genotypic model at rs1044825 suggests that those which are homozygous for the minor allele (C) have 1.7 times (95% CI 1.23–2.35) the odds of having MI than the heterozygotes. CFTR mutation status was incorporated into the logistic regression model to assess confounding and interaction; CFTR, whether coded as number of $\Delta F508$ mutations or whether a person was homozygous $\Delta F508$ or not, had little effect on the observed MI association with rs1044825, rs6489326, rs9300298, or rs7955261, and there was no evidence of statistical interaction (Table 5).

We used the Wang et al. method (Wang and Shete 2008) to combine the association evidence for *ADIPOR2* at rs1044825 from the recessive model and the HWE test. This method resulted in a significant association ($p = 0.0001$) using both the mean-based and median-based tail-strength measures; this result remains significant at the 5% level after Bonferroni correction for the number of SNPs ($N = 92$) tested in the 12p13.3 region, and the number of different statistical tests conducted on rs1044825.

Neither multi-SNP analysis using logistic regression, nor haplotype analysis in PHASE provided additional location information beyond that gleaned from the single SNP analyses.

Further analysis of Hardy–Weinberg disequilibrium in the chromosome

12p13.3 region—We further analyzed the departure from HWE in the 379 unrelated MI cases using exact tests to determine whether the association was cohort-dependent. Since MI treatment practices changed significantly from 1970 to 1980, with less mortality associated with MI as a result, this could be one explanation for HWE departures in the cases. When we stratified the MI-affected subjects by whether they were born before 1980 or born after 1980, only the younger group of cases was significantly out of HWE (rs1044825, $p = 0.0003$). We incorporated this result into a logistic regression model predicting the odds of MI, in which we included as predictors an indicator function for whether the patient was born before or after 1980, the genotype at the significant SNP (e.g., rs6489326), and the genotype-by-cohort indicator interaction. From the significant observed interaction term (rs9300298, $p = 0.004$), we are able to conclude that the genotype–MI association differs by whether an individual was born before or after 1980. Specifically, with the heterozygotes as the reference group, the observed odds ratio for those which are homozygous for the minor allele is 2.2 times greater in individuals born before 1980 than in individuals born after 1980 [95% CI (1.02, 4.72)], with OR = 2.767 (95% CI 1.42, 5.41) for those born before 1980 and OR = 1.260 (95% CI 0.45, 3.48) for those born after 1980.

CNV analysis in ADIPOR2—In the associated region of the *ADIPOR2* gene, we identified two putative CNV variants in the TCAG-CNV database: a deletion of 315 bases (rs35417183), which is located in intron 2, and an insertion of 134 bases in the middle of intron 3. Since the deletion and insertion regions contain enhancers (supplementary Table 1), which may affect the expression levels of *ADIPOR2*, we genotyped all cases and controls for these CNVs. The 315 bp deletion and 134 bp insertion occur on different haplotypes. The 315 bp deletion showed some association evidence in the 2df genotypic test ($p = 0.04$). The 134 bp insertion was not associated with MI ($p = 0.83$).

Family-based association analysis of ADIPOR2 in an independent cohort—In an attempt to replicate our findings, we used the genotype data from the 100K Affymetrix array that was assayed on 29 MI families recruited by JHU. The 100K array contained two SNPs (rs9300298 and rs1044825) of the four markers that were associated under the 2df model and out of HWE in our study cases. We used FBAT to analyze the two SNPs in common. In FBAT, we assumed the presence of linkage, selected the offset value so that only MI-affected offspring contributed to the test statistics, and we then analyzed under additive and genotypic models. The association of rs9300298 in *ADIPOR2* replicated for the same alleles in this independently collected cohort, using the 2df genotypic model ($p = 0.02$). rs1044825 was not associated with MI ($p = 0.35$) in this sample.

Re-sequencing of ADIPOR2 in MI discordant siblings—Sequencing of the *ADIPOR2* coding exons in 24 sibling pairs discordant for MI led to identification of 15 sequence variants (supplementary Table 3). None of these polymorphisms identified were significantly different between MI cases and their no-MI siblings.

Association analysis of SLC4A4 in the no-MI linkage region

Multipoint linkage analysis for the no-MI phenotype provided an HLOD of 3.19 at D4S2389 (cM position = 84.72) under a recessive mode of inheritance, with evidence of heterogeneity ($\hat{\alpha}=0.26$). Since 78% of the no-MI discordant families contributed positively to the HLOD at D4S2389, we used FBAT to perform a multi-allelic test of association with D4S2389 in the 50 no-MI discordant families, under the additive and genotypic models. D4S2389 provided some evidence of association under the genotypic model ($p = 0.046$). The region of D4S2389 contained a plausible candidate gene, *SLC4A4*, which encodes for a bicarbonate transporter. In a previous study of lung function, 26 tagging SNPs in the *SLC4A4* gene (supplementary Table 2) were genotyped in 601 families of varying size (e.g., sibling pairs, nuclear families, etc.). We were thus able to analyze these data for association with our no-MI phenotype. Fifteen families were removed from analysis due to a high Mendelian error rate (due to incorrect family structure), and one SNP was excluded from the analysis for low MAF = 0.009. No SNP was out of HWE in founders. Family-based association analyses using the phenotype of no-MI indicated some association evidence in this gene when analyzed under a recessive model ($p = 0.01$ at rs1453458; bp = 69782442). However, in this analysis, only six families were informative for association leading us to question the applicability of the asymptotic theory of the test statistic. Because rs1453458 originally had a relatively low MAF = 0.04, rs1453458 had been genotyped in additional samples of Canadian CF patients. Family-based association analysis was performed on this SNP in the additional sample after removing 5 families with Mendelian errors, and a significant association was again found between this SNP and the no-MI phenotype under a recessive model ($p = 0.002$), this time with 11 informative nuclear families.

Discussion

Here we present results of linkage analysis for the modifier trait of MI in CF patients; we present association results from loci identified via linkage for both MI and no-MI (a protective locus); and we articulate a type of confounding that is often a result of the design of modifier studies, providing details of the method we used to disentangle linkage evidence for MI.

Standard implementations of NPL might not be the best analytical tool for use in modifier gene studies due to the study design. From NPL results, it is difficult to disentangle whether a locus is linked to the modifying trait, the original disease (e.g., CF), or other phenotypic outcomes that correlate in affected sibling pairs. Through this MI example, we are able to

illustrate the importance of carefully considering the choice of analytical tools in modifier gene studies.

Others have also recognized the difficulty imposed by not including unaffected individuals in the analysis for linkage to modifier genes (Daw et al. 2008; Houlston and Tomlinson 1998; Rogus et al. 2008). Our approach incorporating DSPs into our NPL analysis did, indeed, provide a way to use NPL for our modifier study. However, DSPs are always included in parametric linkage analysis, and parametric analysis can also explicitly account for locus heterogeneity (Ott 1999).

Using parametric linkage, we were able to identify a locus for MI on chromosome 12p13.3 and a protective locus on chromosome 4q13.3. The 12p13.3 locus was corroborated using the NPL approach including discordants. There was also evidence in the NPL approach corroborating the 4q13.3 locus; however, in the presence of heterogeneity, NPL has reduced power (Greenberg and Abreu 2001). Moreover, through fine mapping in a larger cohort of PI patients, we were able to find some evidence of association to *ADIPOR2* and *SLC4A4*.

Fine-mapping results in the region of *SLC4A4* require further scrutiny. The association in *ADIPOR2* was replicated in an independently collected cohort from the US CF Twin and Sibling Study, providing further evidence that *ADIPOR2* plays a role in MI susceptibility.

Interestingly, the *ADIPOR2* association with MI differed by birth cohort. Out of the four associated SNPs in patients born after 1980, two were in strong deviation from HWE even after Bonferroni adjustment for the number of SNPs tested, and the other two were close to the threshold. There was a significant excess of minor allele homozygotes (Table 4) whose OR for developing MI differed from those born before 1980. The deviation from HWE in the MI cases is unlikely to be explained by a lower survival of CF newborns with MI born prior to 1980. This is because in such a case one would expect that the deviation from HWE would be evident in the early birth cohort (prior to 1980), rather than the observed deviation in patients born after 1980. An alternative hypothesis is based on an unexpected side effect of prenatal genetic testing implemented after the discovery of the *CFTR* gene. This prenatal genetic testing could have led to preferential abortion of fetuses homozygous for the deltaF508 mutation, thus reducing the risk of developing MI at birth, and resulting in HWE deviations.

The *ADIPOR2* gene encodes one of two adiponectin receptors. Adiponectin is involved in a broad range of metabolic processes including lipid and glucose metabolism (Kadowaki and Yamauchi 2005; Oh et al. 2007) while playing an anti-inflammatory role (Ouchi and Walsh 2007). Due to the broad expression of *ADIPOR2* in various tissues and the presence of adiponectin circulating in blood, it is not clear how the adiponectin signaling pathway may affect MI. One possibility is that the disruption of adiponectin's anti-inflammatory activity could contribute to an increased risk of MI (Ouchi and Walsh 2007).

In an attempt to identify potential variants that might explain the association between *ADIPOR2* and MI, we re-sequenced the exonic and flanking regions of the gene in the discordant MI sibling pairs. While several changes were identified in the coding and non-coding regions, none of these could be deemed as functional or were associated with MI. Thus, we conclude that unidentified intronic or promoter variants may affect levels of *ADIPOR2* protein. Future work will focus on genetic and expression analyses of *ADIPOR2* in relevant tissues.

The linkage analysis for no-MI detected a broad linkage region on chromosome 4q13. The largest HLOD was observed at D4S2389, located in intron 1 of the *GC* gene. *GC* encodes for the vitamin D-binding protein (DBP/GC). The *GC* protein is a major carrier of vitamin D

and its metabolites in plasma (Speeckaert et al. 2008). It has been recently demonstrated that GC/DBP concentration in the serum of CF patients is significantly decreased (Speeckaert et al. 2008), which may explain severe vitamin D deficiency observed in CF patients due to its reduced absorption in the gut (Rovner et al. 2007). Another likely candidate gene in this region, *SLC4A4/NBC1*, encodes a sodium bicarbonate co-transporter that regulates bicarbonate secretion in multiple organs, including pancreatic duct cells and intestine (Soleimani and Burnham 2001). Levels of expression and rate of endocytosis for a number of transporters, including *SLC4A4*, depend on the levels of the CFTR protein (Lee et al. 2001; Shumaker et al. 1999; Shumaker and Soleimani 1999; Soleimani 2001). It has been recently proposed that defective bicarbonate secretion mucins are less soluble and have a tendency to aggregate in organs affected by CF (Quinton 2008). Hence, a gain-of-function mutation or higher expression levels of *SLC4A4* could ameliorate bicarbonate secretion, thus having a protective effect on MI.

FBAT suggested that the minor allele of rs1453458 in *SLC4A4* is significantly overrepresented in CF families whose proband does not have MI. Intriguingly, the loss of function alleles in *SLC4A4* are associated with renal acidosis (Romero 2005) and no gain-of-function alleles in the general population have been identified to date. Future functional studies will determine whether the minor allele of rs1453458 is associated with increased protein expression. In addition, the whole GC locus should be fine mapped to determine the presence of another gain-of-function allele.

The loci identified in the present study have not been previously reported in human studies. However, the chromosome 4 locus is an orthologous locus to the mouse *Cfm2*. The *Cfm2* region on chromosome 5 was mapped using congenic mice to map loci for severe intestinal obstruction in CFTR KO mice in the 129/B6 intercross (L-C Tsui, personal communication). In the present study, we have been less successful replicating MI loci identified in human studies: we were unable to replicate the loci identified in Blackman et al. (2006), or the *CFMI* locus on chromosome 19q13 (Rozmahel et al. 1996; Zielenski et al. 1999). One possible explanation for the 19q13 locus could be locus heterogeneity. For example, the original 1999 study was conducted in a cohort with a high representation of families recruited from France and Quebec. Therefore, it is possible that the 19q13 locus could harbor a modifier gene that is responsible for MI in CF families of French origin.

Genetic heterogeneity is but one limitation, complicating the localization of MI genes. Another is the MI phenotype itself, which may be recorded with variable reliability. Although MI is a dramatic presentation at birth, the details may be inconsistently documented in the medical charts of older patients (Wood et al. 2007). For the present study, the presence of MI at birth was accepted as reported by the clinic where the patient was recruited, and thus some degree of phenotypic misclassification is expected.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported by Genome Canada through the Ontario Genomics Institute per research agreement 2004-OGI-3-05, by the Canadian Cystic Fibrosis Foundation, by the Ontario Research Fund—Research Excellence Program. L.J. Strug is supported by the NIH (HG-0004314) and the Natural Sciences and Engineering Research Council. R. Dorfman was supported by the joint Fellowship of Canadian Institutes of Health Research and Ontario Women's Health Council. Support by VZFN00064203 to Milan Macek. The authors express gratitude to all CF patients and their families for participating in this study. The authors thank Nicole Anderson, Jennifer Breaton, Mary Cristofi, Roxanne Rousseau, and Michael Van Spall for their exceptional effort in recruiting and ascertaining Canadian CF families for the study, as well Drs. Miroslava Balasckova, Vera Vavrova and Dana Zemkova for

provision of patients with MI. The authors are indebted to the following individuals from member institutions of the Canadian Consortium for CF Genetic Studies for ascertaining patient data and blood samples from CF patients and their families: S. Aaron, P. Barrett, B. Beaurivage, Y. Berthiaume, P. Bigonnesse, M. Boland, L. Boucher, J. Boucher, S. Bourgh, F. Brosseau, N.E. Brown, C. Brunoro, N. Bureau, A. Cantin, L. Charette, G. Cote, A. Dale, G. Davidson, K. Devesceri, R. Dicaire, V. Fauvel, A. Freitag, D.N. Garey, M. Gaul, W. Gervais, J. Gjevre, F. Gosse, A. Gravelle, B. Habbick, R. Hennessey, S.B. Holmes, J. Hopkins, D. Hughes, M. Jackson, J. Jacob, A. Jeanneret, P. Kean, W. Kepron, T. Kovesi, V.J. Kumar, L. Lands, M. LaPerriere, J. Leong, R. Levesque, D. Lougheed, M. Lowe, B. Lyttle, K. Malhotra, J.E. Marcotte, S. Marsolais, C. Martineau, E. Matouk, D. McCulloch, R.T. Michael, M. Montgomery, R. Morris, E.M. Nakielna, F. Paquet, H. Pasterkamp, N. Patterson, L. Pedder, L. Peterson, N. Petit, C. Piche, M. Plante, H.R. Rabin, K. Ramlall, F. Raymong, L. Rivard, G. Rivard, M. Roussin, M. Ruel, J. Salgado, L. Semple, E. Sheppard, F. Simard, A. Smith, M. Solomon, R. Stackhouse, J. Tabak, L. Taylor, A. Tsang, E. Tullis, C. Turtle, K. Vandamheen, M. van Spall, R. van Wylick, I. Waters, T. Wells, S. Wiltse, and P. Zuberbuhler.

References

- Abreu PC, Hodge SE, Greenberg DA. Quantification of type I error probabilities for heterogeneity LOD scores. *Genet Epidemiol.* 2002; 22:156–169. [PubMed: 11788961]
- Ahmed N, Corey M, Forstner G, Zielenski J, Tsui LC, Ellis L, Tullis E, Durie P. Molecular consequences of cystic fibrosis transmembrane regulator (CFTR) gene mutations in the exocrine pancreas. *Gut.* 2003; 52:1159–1164. [PubMed: 12865275]
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–265. [PubMed: 15297300]
- Blackman SM, Deering-Brose R, McWilliams R, Naughton K, Coleman B, Lai T, Algire M, Beck S, Hoover-Fong J, Hamosh A, Fallin MD, West K, Arking DE, Chakravarti A, Cutler DJ, Cutting GR, the CF Twin and Sibling Study. Relative contribution of genetic and nongenetic modifiers to intestinal obstruction in cystic fibrosis. *Gastroenterology.* 2006; 131:1030–1039. [PubMed: 17030173]
- Blackwelder WC, Elston RC. A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol.* 1985; 2:85–97. [PubMed: 3863778]
- Clayton D, Chapman J, Cooper J. Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol.* 2004; 27:415–428. [PubMed: 15481099]
- Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J. Effects of misspecifying genetic parameters in LOD score analysis. *Biometrics.* 1986; 42:393–399. [PubMed: 3741977]
- Cottingham RW Jr, Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet.* 1993; 53:252–263. [PubMed: 8317490]
- Cutting GR, Antonarakis SE, Buetow KH, Kasch LM, Rosenstein BJ, Kazazian HH Jr. Analysis of DNA polymorphism haplotypes linked to the cystic fibrosis locus in North American black and Caucasian families supports the existence of multiple mutations of the cystic fibrosis gene. *Am J Hum Genet.* 1989; 44:307–318. [PubMed: 2563631]
- Daw EW, Lu Y, Marian AJ, Shete S. Identifying modifier loci in existing genome scan data. *Ann Hum Genet.* 2008; 72:670–675. [PubMed: 18494837]
- Dizier MH, Babron M-C, Clerget-Darpoux F. Conclusion of LOD-score analysis for family data generated under two-locus models. *Am J Hum Genet.* 1996; 58:1338–1346. [PubMed: 8651311]
- Durner M, Vieland VJ, Greenberg DA. Further evidence for the increased power of LOD scores compared with nonparametric methods. *Am J Hum Genet.* 1999; 64:281–289. [PubMed: 9915967]
- Elston RC. Man bites dog? The validity of maximizing LOD scores to determine mode of inheritance. *Am J Med Genet.* 1989; 34:487–488. [PubMed: 2624257]
- Forrest WF, Feingold E. Composite statistics for QTL mapping with moderately discordant sibling pairs. *Am J Hum Genet.* 2000; 66:1642–1660. [PubMed: 10762549]
- Greenberg DA. Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: inferring mode of inheritance and estimating penetrance. *Genet Epidemiol.* 1990; 7:467–479. [PubMed: 2292371]
- Greenberg DA, Abreu PC. Determining trait locus position from multipoint analysis: accuracy and power of three different statistics. *Genet Epidemiol.* 2001; 21:299–314. [PubMed: 11754466]
- Greenberg DA, Hodge SE. Linkage analysis under “random” and “genetic” reduced penetrance. *Genet Epidemiol.* 1989; 6:259–264. [PubMed: 2786485]

- Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet.* 1998; 63:870–879. [PubMed: 9718328]
- Guo X, Elston RC. Two-stage global search designs for linkage analysis II: including discordant relative pairs in the study. *Genet Epidemiol.* 2000; 18:111–127. [PubMed: 10642425]
- Haston CK, Tsui LC. Loci of intestinal distress in cystic fibrosis knockout mice. *Physiol Genom.* 2003; 12:79–84.
- Hodge SE, Elston RC. Lods, wrods, and mods: the interpretation of LOD scores calculated under different models. *Genet Epidemiol.* 1994; 11:329–342. [PubMed: 7813895]
- Hodge SE, Abreu PC, Greenberg DA. Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. *Am J Hum Genet.* 1997; 60:217–227. [PubMed: 8981965]
- Houlston RS, Tomlinson IPM. Modifier genes in humans: strategies for identification. *Eur J Hum Genet.* 1998; 6:80–88. [PubMed: 9781018]
- Kadowaki T, Yamauchi T. Adiponectin and adiponectin receptors. *Endocr Rev.* 2005; 26:439–451. [PubMed: 15897298]
- Kristidis P, Bozon D, Corey M, Markiewicz D, Rommens J, Tsui LC, Durie P. Genetic determination of exocrine pancreatic function in cystic fibrosis. *Am J Hum Genet.* 1992; 50:1178–1184. [PubMed: 1376016]
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet.* 1996; 58:1347–1363. [PubMed: 8651312]
- Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol.* 2000; 19(Suppl 1):S36–S42. [PubMed: 11055368]
- Lee MG, Ahn W, Lee JA, Kim JY, Choi JY, Moe OW, Milgram SL, Muallem S, Kim KH. Coordination of pancreatic HCO₃-secretion by protein–protein interaction between membrane transporters. *JOP.* 2001; 2:203–206. [PubMed: 11875260]
- Li, W.; Zou, F.; Lee, S.; Corey, M.; Sun, L.; Wright, F.; Strug, LJ. Analysis of population stratification in North American CF patients. North American CF conference; Minneapolis. 2009.
- Norkina O, De Lisle RC. Potential genetic modifiers of the cystic fibrosis intestinal inflammatory phenotype on mouse chromosomes 1, 9, and 10. *BMC Genet.* 2005; 6:29. [PubMed: 15921521]
- Oh DK, Ciaraldi T, Henry RR. Adiponectin in health and disease. *Diabetes Obes Metab.* 2007; 9:282–289. [PubMed: 17391153]
- Ott, J. Analysis of human genetic linkage. 3rd edn. The Johns Hopkins University Press; Baltimore: 1999.
- Ouchi N, Walsh K. Adiponectin as an anti-inflammatory factor. *Clin Chim Acta.* 2007; 380:24–30. [PubMed: 17343838]
- Pal DK, Durner M, Greenberg DA. Effect of misspecification of gene frequency on the two-point LOD score. *Eur J Hum Genet.* 2001; 9:855–859. [PubMed: 11781703]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
- Quinton PM. Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis. *Lancet.* 2008; 372:415–417. [PubMed: 18675692]
- Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science.* 1995; 268:1584–1589. [PubMed: 7777857]
- Rogus JJ, Poznik GD, Pezzolesi MG, Smiles AM, Dunn J, Walker W, Wanic K, Moczulski D, Canani L, Araki S, Makita Y, Warram JH, Krolewski AS. High-density single nucleotide polymorphism genome-wide linkage scan for susceptibility genes for diabetic nephropathy in type 1 diabetes: discordant sibpair approach. *Diabetes.* 2008; 57:2519–2526. [PubMed: 18559660]
- Romero MF. Molecular pathophysiology of SLC4 bicarbonate transporters. *Curr Opin Nephrol Hypertens.* 2005; 14:495–501. [PubMed: 16046910]
- Rovner AJ, Stallings VA, Schall JJ, Leonard MB, Zemel BS. Vitamin D insufficiency in children, adolescents, and young adults with cystic fibrosis despite routine oral supplementation. *Am J Clin Nutr.* 2007; 86:1694–1699. [PubMed: 18065588]

- Rozmahel R, Wilschanski M, Matin A, Plyte S, Oliver M, Auerbach W, Moore A, Forstner J, Durie P, Nadeau J, Bear C, Tsui LC. Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nat Genet.* 1996; 12:280–287. [PubMed: 8589719]
- Schaffer AA, Gupta SK, Shriram K, Cottingham RW Jr. Avoiding recomputation in linkage analysis. *Hum Hered.* 1994; 44:225–237. [PubMed: 8056435]
- Shumaker H, Soleimani M. CFTR upregulates the expression of the basolateral $\text{Na}^{(+)}\text{-K}^{(+)}\text{-2Cl}^{(-)}$ cotransporter in cultured pancreatic duct cells. *Am J Physiol.* 1999; 277:C1100–C1110. [PubMed: 10600761]
- Shumaker H, Amlal H, Frizzell R, Ulrich CD II, Soleimani M. CFTR drives $\text{Na}^{+}\text{-nHCO}_3^{-}$ cotransport in pancreatic duct cells: a basis for defective HCO_3^{-} secretion in CF. *Am J Physiol.* 1999; 276:C16–C25. [PubMed: 9886916]
- Soleimani M. Impaired pancreatic ductal bicarbonate secretion in cystic fibrosis. *JOP.* 2001; 2:237–242. [PubMed: 11875265]
- Soleimani M, Burnham CE. $\text{Na}^{+}\text{:HCO}_3^{-}$ cotransporters (NBC): cloning and characterization. *J Membr Biol.* 2001; 183:71–84. [PubMed: 11562789]
- Speeckaert MM, Wehlou C, Vandewalle S, Taes YE, Robberecht E, Delanghe JR. Vitamin D binding protein, a new nutritional marker in cystic fibrosis patients. *Clin Chem Lab Med.* 2008; 46:365–370. [PubMed: 18303991]
- Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003; 73:1162–1169. [PubMed: 14574645]
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001; 68:978–989. [PubMed: 11254454]
- Taylor, C.; Webel, A.; Cutting, GR.; Knowles, MR.; Corey, M. A new cf lung disease severity phenotype for genetic modifier studies. North American CF conference; Minneapolis. 2009.
- Vanscoy LL, Blackman SM, Collaco JM, Bowers A, Lai T, Naughton K, Algire M, McWilliams R, Beck S, Hoover-Fong J, Hamosh A, Cutler D, Cutting GR. Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med.* 2007; 175:1036–1043. [PubMed: 17332481]
- Vieland V, Greenberg DA, Hodge SE, Ott J. Linkage analysis of two-locus diseases under single-locus and two-locus analysis models. *Cytogenet Cell Genet.* 1992a; 59:145–146. [PubMed: 1737484]
- Vieland VJ, Hodge SE, Greenberg DA. Adequacy of single-locus approximations for linkage analyses of oligogenic traits. *Genet Epidemiol.* 1992b; 9:45–59. [PubMed: 1634106]
- Vieland VJ, Greenberg DA, Hodge SE. Adequacy of single-locus approximations for linkage analyses of oligogenic traits: extension to multigenerational pedigree structures. *Hum Hered.* 1993; 43:329–336. [PubMed: 8288263]
- Wang J, Shete S. A test for genetic association that incorporates information about deviation from Hardy–Weinberg proportions in cases. *Am J Hum Genet.* 2008; 83:53–63. [PubMed: 18589394]
- Wood SD, Adams S, Dunn J, Baler J, Xu A, Pace RG, Kohl A, Yeatts J, Stonebreaker JR, Konstan MW, Drumm ML, Knowles MR. Inaccuracy of reporting of meconium ileus on case report forms of the gene modifier study. *Pediatr Pulmonol.* 2007; S30:272.
- Zielenski J, Corey M, Rozmahel R, Markiewicz D, Aznarez I, Casals T, Larriba S, Mercier B, Cutting GR, Krebsova A, Macek M Jr, Langfelder-Schwind E, Marshall BC, DeCelle-Germana J, Claustres M, Palacio A, Bal J, Nowakowska A, Ferec C, Estivill X, Durie P, Tsui LC. Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13. *Nat Genet.* 1999; 22:128–129. [PubMed: 10369249]
- Zielenski J, Dorfman R, Markiewicz D, Corey M, Ng P, Mak W, Durie P, Tsui LC. Tagging SNP analyses of the CFM1 locus in CF patients with and without meconium ileus. *Pediatric Pulmonol.* 2005; S28

Fig. 1.
Summary of linkage and association analyses conducted in this study for the MI and no-MI phenotypes

Fig. 2.

Parametric multipoint heterogeneity linkage analysis of chromosome 12, under a dominant model, for the MI phenotype in 71 families with at least one MI-affected child. The largest multipoint HLOD for MI occurred at D12S1656 (cM position = 5.21, HLOD = 2.935). There was no evidence of heterogeneity at D12S1656 ($\hat{\alpha}=1$). 54 of the 71 families (76%) contributed positive LOD scores to the overall linkage signal

Fig. 3. Parametric multipoint heterogeneity linkage analysis of chromosome 4, under a recessive model, for the no-MI phenotype in 226 families with at least one no-MI child. At D4S2389 (cM position = 84.72), HLOD = 3.19. There is evidence of heterogeneity ($\hat{\alpha}=0.26$). 115 of the 226 (51%) families contributed positive linkage evidence to the overall linkage signal

Fig. 4.

Non-parametric multipoint linkage for the MI phenotype in 17 MI concordant families and 54 MI discordant families at all autosomal chromosomes. Analysis with MI concordants is plotted as a *solid line* and analysis with MI discordants is plotted as a *dashed line*. At chromosome 12, families concordant for MI exhibit over-sharing of alleles at D12S1656, while those discordant for MI exhibit under-sharing of alleles at the same locus. This provides corroborating evidence that D12S1656 is a susceptibility locus for MI

Fig. 5.

Non-parametric multipoint linkage for the no-MI phenotype in 176 no-MI concordant families and 50 no-MI discordant families at all autosomal chromosomes. Analysis with no-MI concordants is plotted as a *solid line* and analysis with no-MI discordants is plotted as a *dashed line*. At D4S2389 on chromosome 4, evidence of over-sharing was observed from the no-MI concordant families, and evidence of under-sharing was observed from the no-MI discordant families

Fig. 6. Linkage disequilibrium in the 12p13.3 linkage region on chromosome 12 spanning *ADIPOR2*, *WNT5B* and *CACNA2D4*; shades of *white* indicate $r^2 = 0$, shades of *grey* indicate $0 < r^2 < 1$, and shades of *black* indicate $r^2 = 1$

Fig. 7.

Case-control association for the MI phenotype in the 12p13.3 linkage region on chromosome 12: three candidate genes under the linkage peak were genotyped in 379 unrelated PI patients who presented with MI at birth and 596 unrelated PI patients without MI. *Solid circles* represent $-\log_{10}$ scaled p values obtained under the additive model; *solid diamonds* represent $-\log_{10}$ scaled p values under the genotypic model; *solid horizontal line* represents the significance threshold of 0.05. The size of the *diamond* decreases and the *shade* fades as the pair-wise LD between the top SNP rs1044825 and the rest of the SNPs decreases. Among the 92 SNPs tested for association with MI, rs1003939 provided the smallest p value under an additive model [$p = 0.003$, OR = 0.72, 95% CI (0.58, 0.89)], and rs1044825 resulted in the smallest p value under the genotypic model [$p = 0.002$, OR = 1.7, 95% CI (1.23, 2.35)]

Table 1

Number of families included in the parametric linkage analysis for the MI and no-MI phenotypes by country of origin

	MI	No-MI
Australia	4	4
Canada	42	127
Chile	0	9
Netherlands	1	6
Poland	4	23
USA	20	57

Table 2

Distribution of CFTR mutation genotype, gender, mean age and ethnic origin for the MI concordants, MI discordants and no-MI concordants

	MI/MI (N = 34)	MI/no-MI (N = 114)	No-MI/no-MI (N = 358)
CFTR mutation (% ΔF508/ ΔF508, % ΔF508/other, % other/other)	65/29/6	52/42/6	53/35/12
Gender (% female)	62	37	44
Average Age	14.63	14.89	16.09
% Caucasian	61.76	62.28	61.73

Table 3

STR markers with two-point LOD scores greater than 2 for the MI and no-MI phenotypes

Phenotype	Chromosome	Marker name (flanking markers)	Mol	Position (cM)	Two-point LOD Score
MI	12	D12S1656 (D12S352, D12S37)	Dominant	4.23	2.07 (1.45, 1.76)
	13	D13S1315 (D13S778, D13S26)	Dominant	118.99	2.16 (0.66, 0.13)
No-MI	4	D4S2912 (D4S391, D4S1587)	Dominant	48.97	2.28 (0.61, 0.65)
	4	D4S398 (D4S3000, D4S3004)	Dominant	77.91	2.68 (0.25, 0.54)
	15	D15S1507 (D15S1036, D15S98)	Dominant	68.15	2.21 (0.48, 0.45)
	18	D18S979 (D18S465, D18S485)	Dominant	96.96	2.08 (0, 0.36)
	18	D18S554 (D18S1009, - ^d)	Recessive	117.94	2.10 (0.62, - ^d)

Mode of inheritance (Mol) under which the LOD score was maximized is listed

^dFlanking marker was not genotyped; D18S554 was the last marker on chromosome 18 that was genotyped in our analysis

Hardy–Weinberg equilibrium (HWE) *p* values obtained from exact tests for the four SNPs in the 12p13.3 linkage region that are out of HWE in MI cases only, together with their odds ratios and confidence intervals from two degrees of freedom genotypic test

Table 4

SNP	Base pair position	HWE test (MI)	Genotype count (MI)	HWE test (no-MI)	Genotype count (no-MI)	MA	OR	95% CI
rs7955261	1659632	0.0007	110/155/112	0.80	136/296/154	C	1.54	(1.12, 2.12)
rs6489326	1712498	0.0006	104/155/119	0.41	125/301/156	C	1.62	(1.17, 2.23)
rs9300298	1736465	0.0004	106/153/116	0.87	137/297/153	C	1.50	(1.09, 2.98)
rs1044825	1771746	0.000009	106/146/127	1	126/293/170	C	1.70	(1.23, 2.35)

MA minor allele

Supplementary to Fig. 7: detailed information for the most significantly associated SNPs in the case-control association analysis for the MI phenotype, assuming a genotypic model

Table 5

SNP	Base pair position	Minor allele	MAF	Allele distribution (MI)	Genotype count (MI)	Allele distribution (no-MI)	Genotype count (no-MI)
rs7955261	1659632	C	0.49	375/379	110/155/112	568/604	136/296/154
rs6489326	1712498	C	0.48	363/393	104/155/119	551/613	125/301/156
rs9300298	1736465	C	0.49	365/385	106/153/116	571/603	137/297/153
rs1044825	1771746	C	0.46	358/400	106/146/127	545/633	126/293/170