Identification of a low frequency missense mutation in MUC6 contributing to pulmonary artery hypertension by whole-exome sequencing

Lihui Zou1, Wenqing Li1, Jingli Han1, Yuanhua Yang2, Junhua Jin1, Fei Xiao1, Xiaomao Xu3 and Zhenguo Zhai4

1The MOH Key Laboratory of Geriatrics, Beijing Hospital, National Center of Gerontology, Beijing, PR China; 2Beijing Institute of Respiratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, PR China; 3Department of Respiratory and Critical Care Medicine, Beijing Hospital, National Center of Gerontology, Beijing, PR China; 4Department of Respiratory and Critical Care Medicine, The China-Japan Friendship Hospital, Beijing, PR China

Abstract
Pulmonary arterial hypertension (PAH) represents a progressive disease characterized by abnormally high blood pressure in the pulmonary artery. Although mutations in the bone morphogenetic receptor 2 (BMPR2) are found in 80% of heritable, their low penetrance suggests that other unidentified genetic modifiers are required for this disease. In this report, whole-exome sequencing (WES) and a linkage analysis were performed on genomic DNA isolated from four affected relatives and one non-affected relative in two PAH families. By focusing on meaningful variants which were presented in the four affected family members, but not presented in the non-affected individual, 49 SNP and eight indel variants in 39 genes were identified as candidates. Further high-throughput multiplex genotyping and Sanger sequencing were carried out to confirm the putative causal mutations in 150 individuals (30 idiopathic PAH [IPAH] patients, 30 chronic thromboembolic pulmonary hypertension [CTEPH] patients, and 90 normal controls). A heterozygous and deleterious mutation in the gene MUC6 (p.Pro1716Ser) was confirmed in the IPAH group (20/30, 67%) and CTEPH group (1/30, 3.33%); no variant was detected in the 90 normal controls.

MUC6, which is short for mucin 6, encodes high molecular weight glycoprotein produced by many epithelial tissues and forms an insoluble mucous barrier that protects the lumens. We re-confirmed this low frequency mutation with the 1000 Genomes database across all species; no population or frequency data of this allele were acquired. We also found that this mutation site was highly conserved in different species and predicted MUC6 has the protection function of the airway and pneumoangiogram based on genomic sequence data. The compound heterozygous MUC6 gene mutation (p.Pro1716Ser) suggests a novel disease mechanism leading to PAH.

Keywords
exome sequencing, mutation, MUC6 gene, pulmonary artery hypertension

Background
Pulmonary hypertension (PH) presents as a serious and life-threatening disease characterized by abnormally high blood pressure (hypertension) in the pulmonary circulation, usually leading to death with elevated pulmonary vascular resistance (PVR) and right ventricular failure. It is also very common in patients with collagen vascular disease, heart failure, chronic obstructive pulmonary disease (COPD), and sleep apnea. According to the most recent classification, PH was divided into five types, including
Mutation in MUC6 Contributing to PAH

Zou et al.

arterial, venous, hypoxic, thromboembolic, and miscellaneous. PAH and chronic thromboembolic pulmonary hypertension (CTEPH) are two main types of PH, share the same features of vasoconstriction, smooth-muscle cell and endothelial-cell proliferation and thrombosis.

Even if they receive treatment early, patients with PH have a very poor prognosis and high mortality. In particular, idiopathic PAH (IPAH) is the most severe type of PH, which has an untreated median survival of 2–3 years from the time of diagnosis. However, the pathogenesis and hereditary natures of PH have not been fully defined.

Despite intensive efforts in human genome scanning during the last decade, to date, only a few genes have been reported to be associated with PAH, and they are strikingly present in the same genetic and functional pathways. For example, mutations in BMPR2 associate with at least 50–70% of the familial cases,1,2 as well as approximately 15% of IPAH cases.3,4 The mutations of the genes encoding activin-like kinase-type 1 (ALK1)5 and endoglin (ENG), two other members of the transforming growth factor-beta superfamily, were also found among patients with hereditary hemorrhagic telangiectasia (HHT) and co-existent PAH. However, 25% of heritable PAH and 85% of IPAH cases cannot be accounted for by identified mutations in the known PAH genes.6 Therefore, it is crucial to analyze genetic profiles of patients for candidate genes and identify novel genes causing this complex disease.

Recently, next-generation sequencing (NGS), particularly whole-exome sequencing (WES; also known as targeted exome capture) has become a highly efficient strategy to identify novel causative genes for complex disorders, especially sequencing and filtering multiple affected individuals within a family pedigree to identify the disease-causing mutations.7 It provides a powerful approach to identify new susceptibility genes of Mendelian disorders.

In the present study, we used WES to screen the genomes of four affected relatives and one healthy relative from two PAH families with the goal of discovering genetic variants predicted to contribute to PAH. Exome sequence data from PAH families were further confirmed in sporadic patients. We identified MUC6 gene mutation (p.Pro1716Ser) co-segregated with PAH. The mutation was completely new, which is not found in the single nucleotide polymorphism (SNP) database, the 1000 Genomes Project, and matching normal controls. This mutation site is also highly conserved in different species and MUC6 may have the protection function of the respiratory tract epithelium and pneumoniaiogram against infectious agents. These findings provide new insights into the molecular mechanism by which MUC6 acts as an important factor in PAH development. We therefore believe that WES can help to identify novel genetic variants in patients with PAH and can assist future efforts to select individualized therapeutic approaches of PAH.

Materials and methods

Individuals (patients)

The study was approved by the Ethics Committee of Beijing Hospital. The participants provided their written informed consent to participate in this study. PAH was diagnosed as an increase in mean pulmonary arterial pressure (mPAP) ≥ 25 mmHg at rest as assessed by right heart catheterization based on European Society of Cardiology guidelines.8 Blood samples from all available family members, sporadic idiopathic pulmonary hypertension (IPH) patients, and CTEPH patients were generously provided by Dr. Yang from the Chaoyang Hospital in Beijing, PR China. Non-PAH control DNA samples were conserved in our laboratory from the Beijing Hospital (90 age- and sex-matched controls). The blood samples were obtained after receiving written informed consent from all research participants. Total genomic DNA was extracted from all samples with extraction kit (Biochain, Beijing, PR China) following the protocol provided by the manufacturer, then quantified with NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA).

Whole-exome sequencing

The exome analysis was performed at Beijing Genomics Institution (BGI) in Shenzhen. The DNA (3 μg) of four affected relatives and one non-affected relative in the two PAH families was fragmented to approximately 200 bp following sonication for single- or paired-end library preparation. Exonic sequences were enriched via the NimbleGen (Madison, WI, USA) 2.1 M exon capture array, targeting 34 Mb of sequence from exons and flanking regions in nearly 20,000 known genes. Following exome capture, high-throughput sequencing was performed with the Genome Analyzer II platform (Illumina Inc., San Diego, CA, USA). The average depth of sequencing reads per base-pair (bp) was 70 ×. Paired sequencing reads were aligned to the reference human genome (UCSC, hg19) by using Short Oligonucleotide Analysis Package (SOAP)9. Nucleotides in exons and two flanking regions (10 bp flanking each exon) were further analyzed.

Candidate variant identification

Once all variants were obtained, variant filtration was performed. The detected variants were filtered according to the following criteria: DNA variants present in the Single Nucleotide Polymorphism database v132 (dbSNP132), the 1000 Genomes Project database, Hapmap database, and YH database were all removed with a minor allele frequency ≥ 0.005; variants within intergenic, intronic, and UTR regions and synonymous mutations were excluded from further analyses. Finally, functionally deleterious variants which have an impact on protein function were focused
in further study and predicted with protein prediction tools Gen2Phen.

**Variant validation using the MassARRAY™ system and Sanger sequencing**

Genotyping using genomic DNA was carried out with the MassARRAY™ System (Sequenom, Inc., San Diego, CA, USA), which was based on the matrix-assisted laser desorption ionization-time of flight mass spectrometry platform (MALDI-TOF MS). Specific primers for the respective polymorphic sites were designed for the PCR amplification. After purification of the amplification product, a single extension primer annealed to the amplification product and generated an allele-specific product. The allele-specific extension product was detected by mass spectrometry with distinct masses. Approximately 10% of the SNP mutations found in WES could not be validated by the Sequenom’s MassARRAY Designer software because of its special structure, including high GC content, copy number variation (CNV), and repeat region. These SNPs were confirmed by Sanger sequencing according to the standard protocol. Sequencing reads for each sample were assembled and analyzed with Variant Reporter software. Genotype frequencies for all the SNPs were in Hardy–Weinberg equilibrium, both in controls and in patients with \( P \) values > 0.05. The percentage of successful genotyping was > 90%.

**Statistical analyses**

Statistical analyses were performed with SPSS software (version 18, SPSS Inc., Chicago, IL, USA). The statistical significance of associations between categorical variables was based upon the one-tailed Fisher’s exact test; a \( P \) value < 0.05 was considered to be significant.

**Accession codes**

The following GenBank reference sequence was used for multiple sequence alignment: MUC6, NM_005961.

**Results**

**Clinical presentation of the two PAH families**

We analyzed two PAH families of three generations including four affected members and seven non-affected members (Fig. 1). We sequenced the exomes of five diagnosed members from these two families: a female patient and her sister diagnosed at 31 and 35 years, respectively, her non-affected mother, a male patient diagnosed at 46 years, and his affected daughter diagnosed at 15 years. Patient AI-2 had died because of cardiovascular disease; patient AI-1 died at a young age with a self-reported history of angina, syncope, and breathlessness, which were the symptoms of PAH. Physical examination of all the other members revealed completely normal results.

**Mutation detection and analysis by whole-exome sequencing**

We sequenced exomic regions of the patients AII-2, AII-3, BI-1, BII-2, and AI-3 and obtained 4.4–6.4 million bases of sequence per sample, achieving a mean depth of target region between 69.08 x and 77.20 x, respectively (Table 1). The mean percentage of covered bases was in the range of 99.08–99.19% from the five relatives, which is sufficient for analyzing the SNPs, short insertions, or deletions. The mean rates of nucleotide mismatch were 0.37%; more specifically, 93,600 SNPs and 6993 indels per sample were found. We focused only on variants on non-synonymous
and splice-site variants, gains of stop codon, or frame-shift mutations, which are possibly responsible for the protein sequences and functions. We also excluded variants that reported in the Single Nucleotide Polymorphism database v132 (dbSNP132), the 1000 Genomes Project database, Hapmap database, and YH database.

After filtration of variants, 600 SNPs and 88 indels still remained. Subsequently, we compared the variants presented in the four affected relatives (AII-2, AII-3, BI-1, and BII-2; Fig. 1) but not in the non-affected family member (AI-3; Fig. 1). Of these, 39 genes with 57 common homozygous or heterozygous variants in the four affected patients were identified. Therefore, it was probable that these 57 variants were predicted to be the causal mutations for PAH.

In order to exclude mutations in the genes associated with PAH, we focused on the variant result of PAH-related genes BMPR2, ALK1, and ENG in the patients. No mutation was found in these genes, further suggesting that the PAH development requires other factors.

Variants detected in IPH and CTEPH patients
To verify the candidate mutations, the genomes of four affected relatives (AII-2, AII-3, BI-1, and BII-2; Fig. 1) and the non-affected family member (AI-3; Fig. 1) of these, 39 genes with 57 common homozygous or heterozygous variants in the four affected patients were identified. Therefore, it was probable that these 57 variants were predicted to be the causal mutations for PAH.

In order to exclude mutations in the genes associated with PAH, we focused on the variant result of PAH-related genes BMPR2, ALK1, and ENG in the patients. No mutation was found in these genes, further suggesting that the PAH development requires other factors.

Table 1. Total variants identified through the WES analysis in patients with PAH.

| Members  | A-I3 (control) | A-II2 | A-II3 | B-II | B-II |
|----------|----------------|-------|-------|------|------|
| Target region (bp) | 44,131,461 | 44,234,141 | 44,131,461 | 44,234,141 | 44,131,461 |
| Reads mapped to genome  | 61,341,026 | 59,875,249 | 64,170,438 | 58,639,948 | 62,215,793 |
| Mean depth of target region (X) | 71.80 | 71.74 | 77.20 | 69.08 | 73.80 |
| Coverage of target region (%) | 99.19 | 99.08 | 99.12 | 99.13 | 99.11 |
| Total SNPs (n) | 94,598 | 92,141 | 93,742 | 93,123 | 94,400 |
| Missense | 12,022 | 12,110 | 12,283 | 11,829 | 12,088 |
| Total number of indels | 7085 | 6999 | 7050 | 6915 | 6928 |
| Frameshift variants | 245 | 233 | 228 | 241 | 233 |

Importance of the MUC6 mutation
The 1000 Genomes Project describes the variants and genotypes of the samples from 26 different populations around the world. In order to reconfirm this rare mutation in all the population or some particular population, we selected the data from 1000 genomes and found that no population or frequency data of this allele was acquired. In the meantime, we summarized all the variation consequences of MUC6 gene (Table 3).

We then compared the human sequence for interspecies homology and found that the mutation site of the MUC6 gene is highly conserved among different species (Fig. 3a). We also used the Atlas of Genetics and Cytogenetics in Oncology and Haematology database (http://atlasgeneticsoncology.org/index.html) to analyze the peptide structure of MUC6. The schematic diagram indicated that MUC6 contains D1, D2, D’, and D3 domains at the N-terminal region, the TR domain at the central region, the STP domain, and the CK domain at the C-terminal region. The MUC6 gene has a highly similar 5’ region with other mucin genes; however, it shows no similarity in the 3’ region which contains long stretches of tandem repeats that encode serine- and threonine-rich domains (Fig. 3b).
Pro1716Ser is located in the tandem repeat domain, which exhibits a high level of polymorphism that can result in differences in the length of the protein backbone. Therefore, this polymorphism may affect the structural rearrangement of the VNTR domain that will lead to reduction of the protection of the respiratory tract epithelium against infectious agents.

Discussion

WES is widely performed to identify causal variants in rare heritable disorders. Accompanied by genetic linkage studies, it is usually used in large sick families with multiple affected and non-affected individuals to identify the genes that carry the disease-causing mutations. Thus, other populations including sporadic patients and normal individuals can be screened for mutations in those candidate genes. Finally, case-control studies can be performed when a gene was deemed to play an important role in the pathogenesis of the disease.13

In this study, we reported two PAH families and our WES result of novel causative gene identification in PAH. Family A lived in Beijing, which is a northern city in China. Family B lived in the southern Chinese provinces. We found that the compound heterozygous MUC6 gene mutation (p.Pro1716Ser) on chromosome 11p15 at position 1,017,655 (GRCh37/hg19) was significantly associated with PAH. The presence of MUC6 variants in PAH family members and nearly 70% of IPHs suggests that MUC6 plays an important role in the genesis of this disease in Chinese people. Furthermore, the increase in PVR among PAH patients can be explained by vasoconstriction, proliferative and obstructive remodeling of the pulmonary vessel wall, inflammation, and thrombosis. However, CTEPH is the consequence of mechanical obstruction of the pulmonary arteries, which is the most important pathophysiological process caused by non-resolution of acute thrombotic masses that later undergo fibrosis. Our study demonstrated the MUC6 gene mutation (p.Pro1716Ser) was not the causal mutation of CTEPH due to the different pathophysiological features.
The MUC6 gene is a member of the mucins family, which has a high molecular weight, with heavily glycosylated proteins (glycoconjugates) produced by epithelial tissues. Members of the mucins family show very little sequence homology, but all of them contain long stretches of tandem repeats of sequence that encode serine- and threonine-rich domains. These repeated regions in MUC6 are composed of 507 bp and comprise 50% of the polypeptide.

The human MUC6 gene is located at chromosome position 11p15, a recombination-rich region which also contains other mucin genes (MUC2, MUC5AC, and MUC5B). The MUC6 gene extends approximately 24 kb and the mRNA length is approximately 8 kb. It encodes secretory-type gastric mucin, a high molecular weight secreted glycoprotein with a molecular mass of about 105 kDa, highly expressed in the gastric mucosa, duodenal Brunner’s glands, pancreas, gall bladder, endocervix, seminal vesicle, and the common bile duct. It plays an essential role in the cytoprotection of epithelial surfaces from acid, proteases, mechanical trauma, and pathogenic microorganisms in a variety of epithelial tissues, especially in the gastrointestinal tract. Airway mucins are major components of the soluble layer that protect the lungs from any potentially pathogens and environmental toxins.

In the present study, the overexpression of MUC6 was an early event in the carcinogenesis of the pancreas, bile duct, and endocervix. Alterations in the expression of MUC6 were also correlated significantly with dedifferentiation of bronchial epithelium. Epigenetic modification in the promoter regions, the upstream transcription factors, and signaling pathways is involved in the regulation of expression of MUC6 gene.

The other researchers addressed the association of genetic variability in MUC2 and evolution of gastric cancer precursor lesions (GCPLs), especially in H. pylori-infected patients, suggesting an important role of the secreted mucin in the predisposition to gastric carcinogenesis. MUC6 gene polymorphism contributes to an elevated risk

| Type                      | Number | Description                                                                 |
|---------------------------|--------|-----------------------------------------------------------------------------|
| Downstream gene variant   | 1699   | A sequence variant located ’3’ of a gene                                   |
| Upstream gene variant     | 1621   | A sequence variant located ’5’ of a gene                                   |
| Intron variant            | 1102   | A transcript variant occurring within an intron                            |
| 3 prime UTR variant       | 23     | A UTR variant of the ’3’ UTR                                                |
| 5 prime UTR variant       | 2      | A UTR variant of the ’5’ UTR                                                |
| Coding sequence variant   | 2      | A sequence variant that changes the coding sequence                        |
| Synonymous variant        | 415    | A sequence variant where there is no resulting change to the encoded amino acid |
| Splice region variant     | 44     | A sequence variant in which a change has occurred within the region of the splice site, either within 1–3 bases of the exon or 3–8 bases of the intron |
| Missense variant          | 850    | A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved |
| Inframe deletion          | 11     | An inframe non-synonymous variant that deletes bases from the coding sequence |
| Inframe insertion         | 4      | An inframe non-synonymous variant that inserts bases into the coding sequence |
| Frameshift variant        | 51     | A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three |
| Stop gained               | 23     | A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript |
| Splice donor variant      | 6      | A splice variant that changes the 2 base region at the ’5’ end of an intron |
| Splice acceptor variant   | 3      | A splice variant that changes the 2 base region at the ’3’ end of an intron |
| All                       | 5812   | All variations                                                               |

Table 3. Summary of variation consequences of MUC6 gene.
in gastric cancer patients from northern Portugal.\textsuperscript{15} Some data suggest that short \textit{MUC6} alleles are associated with \textit{H. pylori} infection.\textsuperscript{22} However, the biological significance of mucin gene polymorphism is less well understood. A previous study suggests the polymorphisms of these variable number tandem repeat of \textit{MUC6}, \textit{MUC2}, and \textit{MUC5AC} are predicted to induce the difference in the length of the encoded mucin polypeptides.\textsuperscript{23}

As we know, mucins are overproduced in the patients with airway obstruction, which greatly contribute to lung pathophysiology in asthma, COPD, or cystic fibrosis.\textsuperscript{24–27} Based on the information, we can deduce a conclusion that the compound heterozygous mutations of \textit{MUC6} gene (p.Pro1716Ser) may impair the protective function of airways and pneumoangiograms. Finally, the decrease of pulmonary vasodilation can lead to PH.

**Conclusions**

In summary, in our current study we identified a PAH-related novel \textit{MUC6} gene mutation (p.Pro1716Ser) by WES. The major challenge for mutation analysis is to determine whether the variants are potentially causative factors for the disease. Since the role(s) of the candidate gene is not well understood in this paper, further studies are needed to elucidate the molecular mechanisms of the phenotype and genotype relationships in the coming years.

**Conflict of interest**

The author(s) declare that there is no conflict of interest.

---

**Fig. 3.** Identification of MUC6 gene mutation. (a) Heterozygous mutation site of MUC6 gene is highly conserved in different species. (b) Schematic representation of MUC6 peptide structure.

**Funding**

National Key Research and Development Program of China (2016YFC0905600); National Natural Science Foundation of China (81571384, 81570049); Beijing Natural Science Foundation (7172193); National Science and Technology Major Project for Significant New Drugs Creation (2017ZX09304026).

**References**

1. Newman JH, Wheeler L, Lane KB, et al. Mutation in the gene for bone morphogenetic protein receptor II as a cause of primary pulmonary hypertension in a large kindred. \textit{N Engl J Med} 2001; 345: 319–324.

2. Cogan JD, Pauciulo MW, Batchman AP, et al. High frequency of BMPR2 exonic deletions/duplications in familial pulmonary arterial hypertension. \textit{Am J Respir Crit Care Med} 2006; 174: 590–598.

3. Humbert M, Deng Z, Simonneau G, et al. BMPR2 germline mutations in pulmonary hypertension associated with fenfluramine derivatives. \textit{Eur Respir J} 2002; 20: 518–523.

4. Koehler R, Grünig E, Pauciulo MW, et al. Low frequency of BMPR2 mutations in a German cohort of patients with sporadic idiopathic pulmonary arterial hypertension. \textit{J Med Genet} 2004; 41(12): e127.

5. Fujiwara M, Yagi H, Matsuoka R, et al. Implications of mutations of activin receptor-like kinase 1 gene (ALK1) in addition to bone morphogenetic protein receptor II gene (BMPR2) in children with pulmonary arterial hypertension. \textit{Circ J} 2008; 72(1): 127–133.

6. Desai AA. Novel inheritable caveolin-1 mutations in pulmonary arterial hypertension. \textit{Circ Cardiovasc Genet} 2012; 5(6): 706–707.
7. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011; 12(11): 745–755.
8. Gaïa N, Hoeper MM, Humbert M, et al. Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Respir J 2009; 34(6): 1219–1263.
9. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009; 25(15): 1966–1967.
10. Gabriel S, Ziaugra L and Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr Protoc Hum Genet 2009; Chapter 2: Unit 2.12.
11. Humbert M, Morrell NW, Archer SL, et al. Cellular and molecular pathobiology of pulmonary arterial hypertension. J Am Coll Cardiol 2004; 43(12 Suppl S): 13s–24s.
12. Bras J, Guerreiro R and Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. Nat Rev Neurosci 2012; 13(7): 453–464.
13. Bartman AE, Buisine MP, Aubert JP, et al. The MUC6 secretory mucin gene is expressed in a wide variety of epithelial tissues. J Pathol 1998; 186(4): 398–405.
14. Garcia E, Carvalho F, Amorim A, et al. MUC6 gene polymorphism in healthy individuals and in gastric cancer patients from northern Portugal. Cancer Epidemiol Biomarkers Pre 1997; 6(12): 1071–1074.
15. Toribara NW, Roberton AM, Ho SB, et al. Human gastric mucin. Identification of a unique species by expression cloning. J Biol Chem 1993; 268(8): 5879–5885.
16. Rose MC. Mucins: structure, function, and role in pulmonary diseases. Am J Physiol 1992; 263(4 Pt 1): L413–429.
17. Yamada N, Kitamoto S, Yokoyama S, et al. Epigenetic regulation of mucin genes in human cancers. Clin Epigenetics 2011; 2(2): 85–96.
18. Carvalho F, Seruca R, David L, et al. MUC1 gene polymorphism and gastric cancer—an epidemiological study. Gastroenterol J 1997; 14(1): 107–111.
19. Jia Y1, Persson C, Hou L, et al. A comprehensive analysis of common genetic variation in MUC1, MUC5AC, MUC6 genes and risk of stomach cancer. Cancer Causes Control 2010; 21(2): 313–321.
20. Marín F, Bonet C, Muñoz X, et al. Genetic variation in MUC1, MUC2 and MUC6 genes and evolution of gastric cancer precursor lesions in a long-term follow-up in a high-risk area in Spain. Carcinogenesis 2012; 33(5): 1072–1080.
21. Vinall LE, Hill AS, Pigny P, et al. Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. Hum Genet 1998; 102(3): 357–366.