

# UC Irvine

## UC Irvine Previously Published Works

### Title

Clinical and genetic heterogeneity in chromosome 9p associated hereditary inclusion body myopathy: exclusion of GNE and three other candidate genes

### Permalink

<https://escholarship.org/uc/item/42d529jr>

### Journal

Neuromuscular Disorders, 13(7-8)

### ISSN

0960-8966

### Authors

Watts, Giles DJ  
Thorne, M  
Kovach, MJ  
[et al.](#)

### Publication Date

2003-09-01

### DOI

10.1016/s0960-8966(03)00070-1

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



PERGAMON

Neuromuscular Disorders 13 (2003) 559–567



[www.elsevier.com/locate/nmd](http://www.elsevier.com/locate/nmd)

# Clinical and genetic heterogeneity in chromosome 9p associated hereditary inclusion body myopathy: exclusion of GNE and three other candidate genes

Giles D.J. Watts<sup>a</sup>, M. Thorne<sup>b</sup>, M.J. Kovach<sup>c</sup>, A. Pestronk<sup>d</sup>, Virginia E. Kimonis<sup>a,\*</sup>

<sup>a</sup>*Division of Genetics and Metabolism, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Fegan 5, Boston, MA 02115, USA*

<sup>b</sup>*HHMI at Children's Hospital Boston, Boston, MA, USA*

<sup>c</sup>*Biological and Environmental Sciences, University of Tennessee, Chattanooga, TN, USA*

<sup>d</sup>*Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA*

Received 6 November 2002; received in revised form 12 February 2003; accepted 21 March 2003

## Abstract

We have previously reported a new autosomal dominant inclusion body myopathy clinically resembling limb girdle muscular dystrophy, associated with Paget disease of bone in the majority and frontotemporal dementia in a third of individuals. The critical locus for this unique disorder now termed IBMPFD is 9p21.1–p12, spans 5.5 Mb and contains the gene responsible for the recessive quadriceps-sparing inclusion body myopathy (IBM2). Mutation analysis of the GNE gene associated with IBM2 in affected individuals from four IBMPFD families did not identify any mutations, indicating that the two disorders are not allelic. Expression studies indicate that GNE has a tissue-specific splice pattern, with four splice variants. Mutation analysis in three other candidate genes ( $\beta$ -tropomyosin, NDUFB6 and SMU1) did not identify any mutations.

© 2003 Elsevier B.V. All rights reserved.

**Keywords:** Inclusion body myopathy; Limb girdle muscular dystrophy; Paget disease of the bone; Frontotemporal dementia; GNE; IBM2

## 1. Introduction

Inclusion body myopathy (h-IBM) associated with Paget disease of the bone (PDB) and frontotemporal dementia (FTD), or IBMPFD, is a rare autosomal dominant disorder (MIM 605382). It is characterized by onset of proximal and distal muscle weakness, typically in adulthood and in the majority of cases of early-onset Paget disease and/or early-onset frontotemporal dementia in approximately 37% of cases. There also appears to be a mild asymmetry and variability in the patterns of muscle weakness [1,2]. EMG shows myopathic changes and in the latter stages of the disease, and muscle biopsies reveal non-specific myopathic changes or rimmed vacuoles in the cytoplasm. Serum creatine kinase (CK) levels are normal to slightly elevated, but are not significantly elevated enough for use as a

diagnostic method. Ultimately, the cause of death is through progressive muscle weakness and respiratory failure.

Paget disease of the bone (PDB) is characterized by abnormal, overactive osteoclasts leading to the thickening of bone. Histologically PDB is characterized by disorganized trabecular bone causing thickening of the cortex and coarse trabeculae. Within IBMPFD families the age of onset for Paget is earlier, typically before the age of 50 years. The most common complaint of affected individuals with Paget disease is of bone pain of the spine, pelvis and skull.

Frontotemporal dementia accounts for a substantial proportion of cases of primary degenerative dementia occurring before the age of 65 years in the general population. Disproportionately impaired executive or other frontal lobe functions, associated with changes in behavior and conduct, early in the course of the illness with relative sparing of memory and visuospatial abilities, provide strong support for the diagnosis of FTD. Frontotemporal dementia is seen at a mean onset age of 55 years (range 51–61 years) within IBMPFD families.

\* Corresponding author. Tel.: +1-617-355-6394; fax: +1-617-738-3574.

E-mail address: [virginia.kimonis@tch.harvard.edu](mailto:virginia.kimonis@tch.harvard.edu) (V.E. Kimonis).

The IBMPFD locus has been mapped to chromosome 9p21.1–p12 [1]. This region of chromosome 9 is gene rich and a multiple disease locus for many disorders including two other myopathies; IBM2 [3,4] and nemaline myopathy [5]. Because of the complex phenotype for IBMPFD, the selection of good functional candidate genes is not obvious and the expression of the causal genes is most likely not restricted to skeletal muscle.

There are several examples in the literature of the pleomorphic effects of gene mutations in which dominant and recessive forms of disease can be caused by genetically distinct mutations within the same gene. In particular, the IBM2 gene, which maps within the IBMPFD locus, was recently identified as the UDP-*N*-Acetylglucosamine 2-Epimerase/*N*-Acetylmannosamine Kinase (GNE) [4], which is also allelic for the rare disorder sialuria [6]. Sialuria is an autosomal dominant inborn error of metabolism associated with developmental delay, microcephaly, hepatosplenomegaly, coarse facial features with large tongue and massive urinary excretion of free sialic acid. Despite the absence of quadriceps sparing and the presence of PDB and FTD (Table 1) the IBM2 gene was considered a candidate gene for IBMPFD in view of the shoulder and hip girdle muscle weakness, mixed myopathic and neurogenic EMG findings, as well as myopathic changes and vacuoles in muscle biopsies within both disorders.

We have screened four genes within the candidate region for mutations by direct sequence analysis of genomic DNA from two affected individuals from each of three unrelated IBMPFD families and one unrelated, unaffected control individual; however, since no disease causing mutations have been identified, we have excluded these four positional candidate genes for the IBMPFD disease. Details of these genes are summarized in Table 2. We have further characterized the expression profile of the GNE gene.

## 2. Materials and methods

Consent was obtained from each subject prior to participation. Research studies were approved by, the Springfield Committee for Research Involving Human Subjects, Springfield, IL and later from the Children's Hospital, Boston, MA. Individuals included in the analysis were over age 18 years. Members of these families participated in clinical, biochemical and radiological studies. A diagnosis of myopathy was based on the presence of muscular weakness on physical examination and by creatine kinase measurements, and in several patients by EMG changes or biopsy findings suggestive of myopathy. Paget disease of bone was diagnosed by skeletal surveys including skull, spine, hips, long bones, hands and feet, measurements of alkaline phosphatase and urine for pyridinolinium studies. The diagnosis of dementia was made by comprehensive neuropsychological assessments or was confirmed from the medical records in deceased individuals.

### 2.1. Mutation analysis

Mutation analysis was carried out on two affected individuals from each pedigree. Non-affected individuals and unrelated married-in individuals were selected as controls. Primers for genomic DNA were designed at least 50 bp up or down from exon/intron boundaries for all exons of the genes: GNE,  $\beta$ -tropomyosin, NDUFB6 and SMU1. Sequences longer than 1000 base pairs were divided into multiple overlapping segments for amplification. Polymerase chain reaction (PCR) products were gel-purified using the Gel Extraction Kit (Qiagen, Valencia, CA) and DNA sequence was determined by an ABI 377 sequencer, using a dRhodamine terminator cycle sequencing kit (Applied Biosystems Inc., Foster City, CA). Sequence and trace file comparisons were carried out using Lasergene 99 (DNASar Inc., Madison, WI) software.

### 2.2. Expression and splicing pattern analysis

Expression of the known GNE transcript was also investigated in normal tissues using reverse transcription (RT)–PCR. Using a panel of total RNA from 20 human tissues (Clontech Laboratories) and quadriceps, total RNA (isolated by standard techniques [7]) and cDNA was synthesized by using a poly(dT) primer with SuperScript II reverse transcriptase (Gibco BRL Life Technologies, Grand Island, NY). Using combinations of forward and reverse primers designed within the exon boundaries of the cDNA sequence, PCR was carried out on cDNA from all 21 tissues. Splice variants were confirmed by sequencing of the PCR product.

### 2.3. Real-time quantitative PCR analysis

Total RNA was extracted from post mortem muscle samples (deltoid, gastrocnemius, quadriceps and tibialis anterior) from a 3-week-old infant and cDNA synthesized and described above. The real-time quantitative PCR analysis was performed using ABI PRISM 7700 Sequence Detection System (Applied Biosystems, Branchburg, NJ). The TaqMan PCR reactions were a multiplex of the Assays-on-Demand (Applied Biosystems) gene expression assay for GNE (assay ID: Hs00178556\_m1) and the TaqMan GAPDH control assay (Applied Biosystems). TaqMan amplification reactions were set up in a reaction volume of 20  $\mu$ l using the TaqMan Universal PCR Master Mix (Applied Biosystems) and a total of 75 ng of cDNA pre-reaction. DNA amplifications were carried out in eight-well reaction optical tubes/strips (Applied Biosystems). The TaqMan PCR conditions were used as described in TaqMan guidelines using 40 cycles of 95 °C for 15 s and 60 °C for 1 min, with initial steps of 2 min at 50 °C and 10 min at 95 °C. Each sample was analyzed in duplicate and a negative control of the cDNA reaction minus reverse transcriptase for each sample. The Relative Quantification

Table 1  
Comparison of clinical features of chromosome 9 linked inclusion body myopathies.

	Clinical features	Ethnic background	Inheritance	Myopathy studies				Paget disease of bone	Frontotemporal dementia
				Gait	CPK	EMG	Muscle biopsy		
IBMPFD	Onset at mean age of 42.7 (range 3–66) years  Proximal muscle weakness Occasional progressive distal weakness No quadriceps sparing	Caucasian: English, German, Romanian, Jewish	Autosomal dominant	Marked lordosis  Difficulty with stairs	Normal to mildly elevated	Myopathic abnormalities mainly with irritative changes	Myopathic	Onset mean age 42.2 years involving: Pelvis, spine, hip, scapula and skull	Onset mean age 54 years
IBM2	Mild asymmetry Onset 20–40 years  Initial distal muscle weakness (foot drop) progressing to proximal muscles Pronounced quadriceps sparing Weakness symmetrical Reflexes normal in early stages, later diminished	All ethnic groups, predominantly Iranian Jewish	Autosomal recessive	Walking preserved until late stages	Normal to mildly elevated	Myopathic > neuropathic pattern	Rimmed vacuolated muscle fibers	–	–
Nonaka myopathy	Onset in early adulthood  Pronounced distal muscle weakness Quadriceps sparing	Japanese	Autosomal recessive		CPK elevated	Myopathic pattern	Rimmed vacuolated muscle fibers	–	–

Table 2  
Overview of genes screened, including gene name, gene symbol, accession number, number of exons, length of cDNA in base pairs, number of amino acids and the BAC clone to which the gene maps

Gene name	Symbol	Accession no.	No. of exons	cDNA (bp)	No. of amino acids	Assigned BAC clone
UDP- <i>N</i> -acetylglucosamine-2-epimerase/ <i>N</i> -acetylmannosamine kinase	GNE	NM_005476 (c)	12	2210	722	RP11-421H8
Tropomyosin isoform 2	TPM2	NM_003289 (c)	9	1044	284	RP11-112J3
NADH dehydrogenase (ubiquinone) 1 beta subcomplex 6	NDUFB6	NM_002493 (c)	4	594	128	RP11-205M20
SMU-1	SMU-1	BC002876 (c)	12	1542	513	RP11-54K16

values were calculated as described by the manufacturer (Applied Biosystems).

### 3. Results

#### 3.1. *GNE* gene

We have determined the genomic structure for the *GNE* gene, with a particular emphasis on trying to identify any potential splice variances, by aligning the cDNA and expressed sequence tag (EST) sequences (human and non-human) to the working draft sequences of the BAC RP11-421H8 (AL158830). This identified a new 5' exon (89 bp) in addition to the 13 exons present in the database (Fig. 1A). Since there is no unified numbering system for the exons of *GNE* and for the sake of clarity, the new exon will be referred to as exon A1 (Fig. 2A). Exon A1 contains one open reading frame (ORF) present in each of two different splice

variants; splice forms II and III. This forms different 5' amino acid sequences without producing a frame shift in the original ORF (Figs. 1B,C and 2B).

It was also found that exon 1, which is a 5' untranslated region (UTR), only constitutes 238 bp of the published sequence of 423 bp [4] (Fig. 1D). This does not change the coding of the gene, however, as the start codon is within the second exon.

While characterizing the expression profile of *GNE* in 21 different tissues, we identified four splice variants that were expressed in a tissue-specific pattern (Fig. 2B and Table 3). A majority of tissues only expressed one of the four splice variants, for example liver and thyroid gland only expressed variant III (Fig. 2B). Splice variants III and IV were not expressed together in the same tissue, whereas variants I, II and IV were jointly expressed in kidney, prostate and trachea, although colon expressed only variants I and II. Both splice variants II and III were jointly expressed in placenta and salivary gland. All other tissues expressed only

#### A: Exon A1 :

```
atttcaaacacactagggtacagagctcgtgcttcgggatggaacctatggtatctgcagagggagtcgctttcaagga 84
cctca 89
```

#### B: Splice variant II :

```
atttcaaac 9
cacactagggtacagagctcgtgcttcgggatggaacctATGGTTATCTGCAGAGGGAGTCATGCTTTCAGGACCTCAAACT 93
M V I C R G S H A F K D L K L
CTATTTAAGAACCTCTCAAACGAAACAAGCAAATCATGGAGAAGAATGGAATAACCGAAAGCTGCCGGTTTGTGTTGCT 175
Y F K N L S K R N K Q I M E K N G N N R K L R V C V A
```

#### C: Splice variant III :

```
atttcaaac 10
acactagggtacagagctcgtgcttcgggatggaacctATGGTTATCTGCAGAGGGAGTCATGCTTTCAGGACCTCAAAAT 94
M V I C R G S H A F K D L R N
ACATATCGAATGATTGAACAAGATGACTTTGACATTAAACACAGGCTACACACAATTGTGAGGGAGAAGATGAGGCAGCCATG 178
T Y R M I E Q D D F D I N T R L H T I V R G E D E A A M
```

#### D: Exon 1 corrected sequence :

```
caatcaecgagctctctcgggtggcgtcagggcgagtcgctcgggatcggcactccgggtggcgtggctgggagcagcaggag 84
tggggacaaggctcagagcagcagctcctaccgccgcgccgagcgggaaaccgagggcagcctcgcctctgcctgcttcg 168
tggcgttgggtcgtccctcggcagggagcgggtggcggcgtgggagggagcctctgacggcgtctggtagagcccgccgg 252
cggcgtgggactccggagagatcctcccacccccagggcggcctccccggcctcctcccagcctcccctcccggcc 336
gctttcaccaccacggcccgagggcccaactggaccggcctcctcccacatgtcctgcaggcctcaccataggccacctct 420
ttg 423
```

Fig. 1. (A) Sequence for the new alternatively spliced coding first exon. (B,C) Exon A1 with the two 5' amino acid sequences shown, for splice variants II and III. The original ORF is shown in the gray box. (D) The sequence in bold represents the amended exon sequence previously reported. Of the original 423 bp only the first 238 bp is present in the type I splice variant. The splice donor site is underlined. (E) Nucleotide sequence and predicted amino acid sequence for the *GNE* gene, from the first coding exon. Exons are numbered as the first coding exon being number 1. The domain structure of *GNE* is shown as grey for the Epimerase 2, UDP-*N*-acetylglucosamine 2-epimerase domain and light grey for the ROK family domain, with the domain giving rise to Sialuria indicated as the darker grey box within the epimerase domain. Also, the splice site for exons 11 and 12 are indicated showing the *AluY* and *AluSx* sequences.

**E: Coding region of GNE showing alternate start codon and non-coding *Alu* repeats:**

```

actctat 8
tttaagaacctctcaaaacgaacaagcaaatcATGGAGAAGAATGGAAATAACCGAAAGCTGCGGGTTTGTGTGCTACTTGT 92
      M E K N G N N R K L R V C V A T C
      ▶Start codon
AACCGTGCAGATTATTCTAAACTTCCCGGATCATGTTGGCATTAAAAACCGAACCTGAGTCTCTTGAACCTGATGTTGTGTA 176
N R A D Y S K L A P I M F G I K T E P E F F E L D V V V
      Exon 2 Exon 3
CTTGGCTCTCACCTGATAGATGACTATGAAAATACATATCGAATGATTGAAACAAGATGACTTTGACATTAACACCAGGCTACAC 260
L G S H L I D D Y G N T Y R M I E Q D D F D I N T R L H
ACAATTGTGAGGGGAGAAGATGAGGCACCCATGGTGAAGTCAGTAGGCCCTGGCCCTAGTGAAGCTGCCAGATGTCCTTAATGCG 344
T I V R G E D E A A M V E S V G L A L V K L P D V L N R
CTGAAGCCTGATATCATGATTTGTCATGGAGACAGGTTTGTAGCCCTGGCTCTGGCCACATCTGCTGCCCTGATGAACATCCGA 428
L K P D I M I V H G D R F D A L A L A T S A A L M N I R
ATCCTTCAATTTGAAGTGGGGAGTCAAGTGGACCATTTGATGACTCTATCAGACATGCCATAA CAAAATGGCTCAATATCAT 512
I L H I E G G E V S G T I D D S I R H A I T K L A H Y H
GTGTGCTGCACCCGACGTGAGAGCAGCACCTGATATCCATGTGTGAGGACCATGATCGCATCCGTTTGGCAGGCTGCCCTTCC 596
V C C T R S A E Q H L I S M C E D H D R I L L A G C P S
TATGACAACTTCTCTCAGCCRAAGACAAGACTACATGAGCATCATTCGCATGTGGCTAGGTGATGATGTAATACTAAGAT 680
Y D K L L S A K N K D Y M S I I R M W L G D D V K S K D
      Exon 4
TACATTGTTGCACTACAGCACCCCTGTGACCACCTGACATTAAGCATTCCATAAAAAATGTTGAATTAACATTGGATGCACTTATC 764
Y I V A L Q H P V T T D I K H S T K M F E L T L D A L I
TCATTTAAACAGCGGACCCCTAGTCTCTGTTCCAAATATTGACGCAAGGAGCAAAAGAGATGGTTCAGTGTGCGGAAGAAGGGC 848
S F N K R T L V L F P N I D A G S K E M V R V M R K K G
ATTGAGCATCATCCCAACTTCTCGTCAGTTAAACACGTCCTCAATTTGACAGTTTATACAGTTGGTCCCATGCTGGCTGTATG 932
I E H H P N F R A V K H V P F D Q F I Q L V A H A G C M
ATTGGAAACAGCAGCTGTGGGGTTCGAGAAGTTGGAGCTTTTGGAAACCTGTGATCAACCTGGAAACAGTCAAGATTGGAAGA 1016
I G N S S C G V R E V G A F G T P V I N L G T R Q I G R
      Exon 6
GAAAAGGGGAGAATGTTCTCATGTCGGGATGCTGACACCCAAAGACAATAATTGCAAGCACTGCACCTTCAGTTTGGTAAA 1100
E T G E N V L H V R D A D T Q D K I L Q A L H L Q F G K
      Exon 7
CAGTACCCCTGTTCAAAGATATATGGGGATGGAAATGCTGTTCCAAAGATTTTGAAGTTTCTCAAATCTATCGATCTTCAAAG 1184
Q Y P C S K I Y G D G N A V P R I L K F L K S I D L Q E
CCACTGCAAAAGAAATTTCTGCTTCTCCTGCTGGAAGGAGAATATCTCTCAAGATATTGACCATAATCTTGAACACTTAAAGTGC 1268
P L Q K K F C F P P V K E N I S Q D I D H I L E T L S A
      Exon 8
TTGGCCGTTGATCTTGGCGGGACCAACCTCCGAGTTGCAATAGTCAGCATGAGGGTGAATAGTTAAGAAGTATACTCAGTTC 1352
L A V D L G G T N L R V A I V S M K G E I V K K Y T Q F
AATCCTAAACCTATGAAGAGAGGATTAATTTAATCCTACAGATGTGTGGAAGCTGCAGCAGAAGCTGTAACCACTGAACTGC 1436
N P K T Y E E R I N L I L Q M C V E A A A E A V K L N C
AGAATTTGGGAGTAGGCATTTCCACAGTGGCGGTGTAATCCTCGGGAAGGAATTTGCTGCTATTCAACCAACTGATCCAA 1520
R I L G V G I S T G G R V N P R E G I V L H S T K L I Q
      Exon 9
GATGGAACTCTGAGACTTTAGACCCCTTCTGACACTTGCATCTCCCTGTGTTGGGTAGACATGATGGCACTGTGCT 1604
E W N S V D L R T P L S D T L H L P V W V D N D G N C A
      Exon 10
GCCCTGGCGAAAGGAATTTGGCCAAAGAAAGGACTGGAACACTTTGTTACACTTATCACAGCAAGGAATCGGTGGTGA 1688
A L A E R K F G Q G K G L E N F V T L I T G T G I G G G
ATTATCCATCAGCATGAATTTGATCCAGGAAGCTCTTCTGTGCTGCAGAAGTGGCCACCTGTTGTGCTCTGGATGGCCCT 1772
I I H O H E L I H G S S F C A A E L G H L V V S L D G P
GATTGTTCTGTGGAAGCCATGGTGCATTAAGCATAAGCCTCTGGAATGGCTTGCAGAGGGAGGCAAAAAAGCTCCATGAT 1856
D C S C G S H G C I E A Y A S G M A L Q R E A K K L H D

```

Fig. 1 (continued)

splice variant IV, except in the lung where expression of any GNE splice form was not detectable by PCR.

It is observed that exons A1 and exon 1 are mutually exclusive 5' exons since they were not found together as part of the same transcript. Splice variant I consists of exons 1 through 13 being spliced together, using the start codon in exon 2, to form a 2169 bp ORF that codes for 722 amino acids (Figs. 1E and 2B). In splice variant II, the ORF seen in variants I and IV is maintained but has an additional unique 28-amino-acid sequence since the start codon is in exon A1 (Figs. 1B and 2B). Splice variant III is absent for the first 59 amino acids of the protein sequence seen in variants I and IV (Fig. 1E), since exon 2 is spliced out and exon A1 is spliced in phase to exon 3, with the start codon being within exon A1 (Fig. 1C). Splice variant IV uses the same ORF as splice

variant I, although the transcript is lacking the non-coding exon 1 (Figs. 1E and 2B).

### 3.2. Mutation analysis of *GNE* gene

Using primers for genomic DNA, the entire length for all 14 exons of GNE were screened for mutations by direct sequencing from two affected individuals from each of the four families. Mutation analysis of the 14 exons identified two non-coding single nucleotide polymorphisms (SNPs), one heterozygous SNP within intron 11 (IVS11 – 80 G → C) and one homozygous within the 3' UTR of exon 13 (320 T → C). Since both were also present in normal individuals they were not considered as mutations.

**Exon 11**  
 GAGGACCTGCTCTTGGTGAAGGGATGTCAGTGCCAAAAGATGAGGCTGTGGGTGOGCTCCATCTCATCCAAGCTGCGAAACTT 1940  
 E D L L L V E G M S V P K D E A V G A L H L I Q A A K L

GGCAATGCGAAGGCCAGAGCATCTAAGAACAGCTGGAACAGCTTTGGTCTTGGGGTTGTGAACATCTCCATACCATGAAT 2024  
 G N A K A Q S I L R T A G T A L G L G V V N I L H T M N

CCCTCCCTGTGATCTCTCCGGAGTCTGGCCAGTCACTATATCCACATGTCAAAGACGTCAATCGCCAGCAGGCTTGTC 2108  
 P S L V I L S G V L A S H Y I H I V K D V I R Q Q A L S

TCCGTGAGGACGTGGATGTGTGGTTTCGGATTGGTTGACCCCGCCCTGCTGGGTGCTGCCAGCATGGTCTGGACTACACA 2192  
 S V Q D V D V V V S D L V D P A L L G A A S M V L D Y T

ACACGACGGATCTACTAGacctccaggaacagacatggaccttctccagagctcctgagtggaatcaagttctgtcttttag 2276  
 T R R I Y \*

gatgaccttttctaacaatcaaatctggtattgaactgcaggtgactttggcagagaaatgthttcaacttttggctcctctt 2360

ccagagtcacctttcccactcctattttttagatgctattctttctgatgtctcttactaggggctatttttagctcaaac 2444

ctgtaagttacagtcacaattttctgtgccaagcagctacaataatagagaggaaagccttcttagaactctgcttactaatgt 2528

attaataccactgagaccttcaggccttgcctggatataccttcatcctgaagtttgcattaat.aatccttccagccggcagc 2612  
*AluY*

actggtcagcctctaaatcccagcacttttggagcccgagccagccagctcaacagctcagga.gatcagaccctcctgcta 2696  
*AluY*

acatggtgaaacatggtgaaacccctctctactaaaaatacaaaaaattagctggtgtggtgagcctccagctactcggga 2780  
*AluY*

**Exon 12** | **Exon 13**  
 gctgagcagcaggaatgcatgaaacccagctgagctcaatgcaacctctgcccccaagtccagtattctc 2864  
*AluX*

ctgctcagcctctgagtagctggattacaggcacatgaccacgcccggctgattttttttagatttttagtagacaggg 2948  
*AluX*

tttaccatgltggccagcctggtctlgaactcctgacctcaggtgatccgcccgcctctgcttcccaggctgctgggattaca 3032  
*AluX*

gattgaaacatcacacctgacctagtgacaggttttttgggtacttttagatgatcaagaatcatgcatatatacttt 3116

cagatttctattttgggaaatgaaggtttctacaacatattgtttcagtggttcaataaaactgaaggactcaacattacatt 3200

gaactatatacttctcctagtggttagtgtaaaagaggttggctgatttctaaactctgcccagcctgcagtaactcaggg 3284

gcctggttattgttcagacattccatggtgattcctgggaaggaagcttggctgctcagtttctgagtgctgggtgagataatg 3368

ttctgggaagggacatctgttctttggtgtaactctcatggtgaaatctgctctgtacatcagacaattgcattgctaccaag 3452

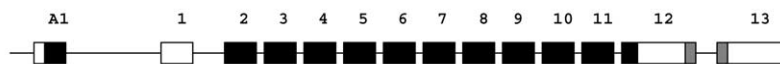
tttcatccaataattgaaaggatggtatgaaatcaaaacaaatattagtttttattaaactcatgggaaggtcaatatat 3536

tccacgtaaattattacatattggttaagtaattgcatgttaatttttaattgtaaatattttggttactgtctgagcoaa 3620

atttcaagaaaaataaataacatttcttggtaaaaaaaa 3664

Fig. 1 (continued)

**A: GNE exons with unified numbers**



**B: GNE Tissue Specific Splice Variants**

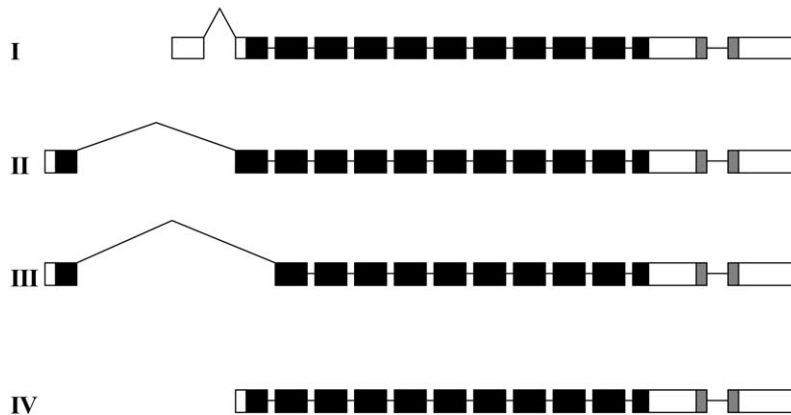


Fig. 2. Illustration of the four splice variants of GNE. (A) Sequence diagram of GNE exon/intron structure. Numbered boxes represent exons (not to scale). Black boxes are coding regions. Connecting lines between the boxes represent introns (not to scale). (B) A depiction of the four splice variants found for GNE.



Table 3  
Tissue distribution of the four GNE transcripts

Tissue	ISplice variant			
	I	II	III	IV
Adrenal gland				+
Bone marrow				+
Brain, cerebellum				+
Brain (whole)				+
Heart				+
Kidney	+	+		+
Liver			+	
Lung	–	–	–	–
Placenta		+	+	
Prostate	+	+		+
Salivary gland		+	+	
Skeletal muscle				+
Spleen				+
Testis				+
Thymus				+
Thyroid gland			+	
Trachea	+	+		+
Uterus				+
Colon	+	+		
Small intestine				+
Quad				+

### 3.3. Real-time quantitative PCR analysis

While we were characterizing the expression patterns of the various splice forms, it seemed that although the cDNA was synthesized from the same amount of RNA, there was variation in the intensity of the resulting PCR products when visualized on a gel. In particular, GNE expression in quadriceps appeared reduced when compared to other tissues such as heart and mixed skeletal muscle samples. To investigate this observation in a quantitative manner, we performed real-time PCR on deltoid, gastrocnemius, quadriceps and tibialis anterior samples from the same individual. The  $\Delta C_T$  values for each sample were: deltoid = 8.2, gastrocnemius = 8.035, quadriceps = 8.36 and tibialis anterior = 8.375. We then calculated the Relative Quantification values (relative to quadriceps). Tibialis anterior was closest to quadriceps with 1% less expression of GNE, while deltoid and gastrocnemius showed an increase in GNE expression of 10 and 20%, respectively. Since the different expression levels observed for the GNE mRNA from these muscles was small, the mechanism by which the quadriceps, and in some cases the deltoids, escape muscle wasting in IBM2 is probably not due to differing expression levels.

### 3.4. Mutation analysis of $\beta$ -tropomyosin, *NDUFB6* and *SMU1* genes

Due to the dominant nature of IBMPFD it is possible that the causal gene is involved in the structural integrity of

the laminal membrane. For this reason we screened these three genes:  $\beta$ -tropomyosin, *NDUFB6* and *SMU1* (Table 2).  $\beta$ -Tropomyosin (*TPM2*) [5] is the gene responsible for a rare form of nemaline myopathy. NADH-Ubiquinone Oxidoreductase 1 Beta Subcomplex 6 (*NDUFB6*), is part of the first multisubunit enzyme complex of the mitochondrial respiratory chain [8]. *SMU1* is involved in the splicing control of several extracellular matrix proteins [9], in particular perlecan, since mutations of the perlecan gene are known to cause Schwartz–Jampel syndrome (chondrodystrophic myotonia) [10] and dyssegmental dysplasia (Silverman–Handmaker type) [11,12]. Both of these disorders are associated with neuromuscular and skeletal components.

We determined the genomic structure for the  $\beta$ -tropomyosin, *NDUFB6* and *SMU1* genes by aligning the cDNA and EST sequences to the working draft sequences (Table 2). All coding and non-coding exons were sequenced for all three genes. *TPM2* and *NDUFB6* were not found to contain any mutations or coding/non-coding SNPs although *SMU1* had two non-coding SNPs, one heterozygous SNP within intron 6 (IVS6 – 47 A → G) and one heterozygous within intron 11 (IVS11 – 9 A → G). Since both were also present in normal individuals, they were not considered mutations.

While determining the genomic structure of *SMU1* a BLAST search identified a potential *SMU1* pseudogene located on chromosome 1q23.2–24.3. A BLAST search of the *SMU1* mRNA sequence against the AL157713 (RP11-110J1) showed a 96% identical match to the *SMU1* sequence from chromosome 1 which did not contain any introns. Two ORFs of 95 and 185 amino acids were identified which both showed 95% identity to the *SMU1* sequence on chromosome 9p21.1. The fragmented ORF and the absence of introns would suggest that this sequence on chromosome 1q23.1–24.3 is a pseudogene of *SMU1*.

## 4. Discussion

Hereditary inclusion body myopathy with early-onset Paget disease of bone and frontotemporal dementia is distinct among the clinically and genetically heterogeneous family of hereditary myopathies. The clinical features distinguish it from the recessive types of h-IBM linked to chromosome 9p (Table 1). The weakness primarily involves the proximal muscles clinically being indistinguishable from the limb girdle muscular dystrophies. Striking features of this syndrome consist of a marked proximal weakness later occasionally involving the distal musculature with individuals walking with an exaggerated lordotic gait, having difficulty raising their arms up, scapular winging. There is a moderate variability in the patterns of muscle weakness, noted both within and among families, as well as a common pattern of weakness involving proximal and with mild asymmetry. Onset of the myopathy is typically in



adulthood; however, earlier involvement is documented in some patients.

Paget disease of bone has been reported previously presenting with neuromuscular disorders, such as non-specific muscular dystrophy [13], dystrophia myotonia [14], amyotrophic lateral sclerosis [15] and scapuloperoneal muscular dystrophy [16], although there have been no reports of a molecular basis for these disorders to date.

We previously identified linkage to a 5.5 Mb region on chromosome 9p21.1–p12 [1] which encompasses the IBM2 locus. There are several examples in the literature of the pleiomorphic effects of gene mutations in which dominant and recessive forms of disease can be caused by genetically distinct mutations within the same gene. The *GNE* gene is an example of this as it is the casual gene for IBM2 and sialuria [4,6]. While no mutations in *GNE* were identified for IBM2, we were able to further characterize the expression of the gene in a number of tissues. This showed that different splice forms of *GNE* are expressed in all tissues, with kidney, liver, placenta, prostate, salivary gland, thyroid gland, trachea and colon splice variants (II and III, Fig. 2B) possessing different 5' amino acid sequences to those of all other tissues. The full ORF of *GNE* (variant IV) was expressed in quadriceps.

When using real-time PCR as a method to quantify the amount of gene expression, we minimized the number of variables between the samples being tested. For this reason, in this analysis we used muscle samples from the same individual and did not use the mixed skeletal muscle or quadriceps samples used in the expression panel, since the latter were not age- or sex-matched. We found the expression levels of the *GNE* mRNA did not seem to differ greatly between deltoid, gastrocnemius, quadriceps and tibialis anterior, for the samples tested. Since IBM2 is a progressive disorder it is possible that the expression of *GNE* could vary with age.

It is interesting to note that in the mouse and rat, *GNE* gene contains only 12 exons although the coding region is very well conserved when compared to human. In addition to the two non-coding 5' exons, the evolution of the human gene exon 12 has included an *AluY* repeat, which has spliced with an *AluSx* repeat, which is at the start of exon 13. Neither of the *Alu* repeats are coding, but they are transcribed as part of the 3' UTR.

Due to the complex phenotype of IBM2, the selection of candidate genes, based on their function, is difficult. The most common attribute of the disease is the myopathy and given the dominant nature of the disorder, it is possible that the causal gene is a structural one, although the added combination of bone and brain have not previously been reported for any of the structural genes known to cause myopathies. Therefore, all genes placed in the IBM2 interval are considered as candidate genes.

We have constructed a complete sequence contig across the IBM2 candidate interval, using BAC sequence data from the public database. This has allowed us to determine

the genomic structures of *GNE*, *NDUFB6*, *SMU1* and *TPM2*. We have screened these four genes for mutations in two affected individuals from each of the four IBM2 families and one unrelated, unaffected control individual. Our studies identified four SNPs within the non-coding and intronic sequences of *GNE* and *SMU1*. Two criteria were used to exclude the candidate genes: (1) no co-segregation of sequence changes with the IBM2 phenotype and (2) the identification of the same sequence change within unrelated, unaffected control individuals (i.e. *GNE* and *SMU1*).

Additional candidate genes that map within the critical region of the syndrome of IBM2 on chromosome 9, including talin (*TLN*) [17,18], cAMP response element-binding protein 3 (*CREB3*) [19], clathrin light chain (*CLTA*) [20], interleukin-11 receptor alpha chain (*IL11RA*) [21], serine protease inhibitor, Kazal type 4 (*SPINK4*) and ubiquitin associated protein (*UBAP*) [22], express proteins that may be involved in cellular and structural functions of muscle and/or bone cells.

In summary, *GNE* mutations were not associated with IBM2, indicating genetic heterogeneity with IBM2. We have also excluded three other candidate genes for IBM2. This study identified several non-coding SNPs of which none lead to a dominant mutation causing IBM2. Discovery of the gene associated with IBM2 will provide insight into the pathogenesis of muscle, bone and neurological disease. Perhaps an overlap exists between the biological processes that control bone and muscle maintenance and regeneration. Elucidation of the gene defect responsible for IBM2 should contribute to our understanding of these systems and possibly result in a treatment for this complex disorder.

## Acknowledgements

We thank the families and their physicians for their enthusiastic participation and contribution in our research studies. Funding of this study is from the National Institutes of Health NIAMS R03 AR 46869, NINDS K02 NS02157 award, and previously from the Excellence in Academic Medicine Program at Southern Illinois University School of Medicine, the Muscular Dystrophy Association and the Paget Foundation. Electronic-Database Information: GenBank at the National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>

## References

- [1] Kovach MJ, Waggoner B, Leal SM, et al. Clinical delineation and localization to chromosome 9p13.3–p12 of a unique dominant disorder in four families: hereditary inclusion body myopathy, Paget disease of bone, and frontotemporal dementia. *Mol Genet Metab* 2001;74:458–75.

- [2] Kimonis VE, Kovach MJ, Waggoner B, et al. Clinical and molecular studies in a unique family with autosomal dominant limb-girdle muscular dystrophy and Paget disease of bone. *Genet Med* 2000;2: 232–41.
- [3] Eisenberg I, Avidan N, Potikhan T, et al. The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat Genet* 2001;29: 83–7.
- [4] Kayashima T, Matsuo H, Satoh A, et al. Nonaka myopathy is caused by mutations in the UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase gene (GNE). *J Hum Genet* 2002;47:77–9.
- [5] Donner K, Ollikainen M, Ridanpaa M, et al. Mutations in the beta-tropomyosin (TPM2) gene – a rare cause of nemaline myopathy. *Neuromuscul Disord* 2002;12:151–8.
- [6] Seppala R, Lehto VP, Gahl WA. Mutations in the human UDP-N-acetylglucosamine 2-epimerase gene define the disease sialuria and the allosteric site of the enzyme. *Am J Hum Genet* 1999;64:1563–9.
- [7] Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Press; 1989.
- [8] Smeitink J, Loeffen J, Smeets R, et al. Molecular characterization and mutational analysis of the human B17 subunit of the mitochondrial respiratory chain complex I. *Hum Genet* 1998;103:245–50.
- [9] Spike CA, Shaw JE, Herman RK. Analysis of smu-1, a gene that regulates the alternative splicing of unc-52 pre-mRNA in *Caenorhabditis elegans*. *Mol Cell Biol* 2001;21:4985–95.
- [10] Nicole S, Davoine CS, Topaloglu H, et al. Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia). *Nat Genet* 2000;26:480–3.
- [11] Arikawa-Hirasawa E, Le AH, Nishino I, et al. Structural and functional mutations of the perlecan gene cause Schwartz–Jampel syndrome, with myotonic myopathy and chondrodysplasia. *Am J Hum Genet* 2002;70:1368–75.
- [12] Arikawa-Hirasawa E, Wilcox WR, Le AH, et al. Dyssegmental dysplasia, Silverman–Handmaker type, is caused by functional null mutations of the perlecan gene. *Nat Genet* 2001;27:431–4.
- [13] McBride TI. Paget’s disease and muscular dystrophy. Report of an unusual association in one family. *Scott Med J* 1966;11:238–43.
- [14] Caughey JE, Gwynne JF, Jefferson NR. Dystrophia myotonica associated with familial Paget’s disease (osteitis deformans) with sarcomata. *J Bone Joint Surg* 1957;38B:316–25.
- [15] Tucker WSJ, Hubbard WH, Stryker TD, et al. A new familial disorder of combined lower motor neuron degeneration and skeletal disorganization. *Trans Assoc Am Physicians* 1982;95:126–34.
- [16] Waggoner B, Kovach MJ, Winkelman M, et al. Heterogeneity in familial dominant Paget disease of bone and muscular dystrophy. *Am J Med Genet* 2002;108:187–91.
- [17] Gilmore AP, Ohanian V, Spurr NK, Critchley DR. Localisation of the human gene encoding the cytoskeletal protein talin to chromosome 9p. *Hum Genet* 1995;96:221–4.
- [18] Ben-Yosef T, Francomano CA. Characterization of the human talin (TLN) gene: genomic structure, chromosomal localization, and expression pattern. *Genomics* 1999;62:316–9.
- [19] Lu R, Yang P, O’Hare P, Misra V. Luman, a new member of the CREB/ATF family, binds to herpes simplex virus VP16-associated host cellular factor. *Mol Cell Biol* 1997;17:117–26.
- [20] Ponnambalam S, Jackson AP, LeBeau MM, et al. Chromosomal location and some structural features of human clathrin light-chain genes (CLTA and CLTB). *Genomics* 1994;24:440–4.
- [21] Van Leuven F, Stas L, Hilliker C, et al. Molecular cloning and characterization of the human interleukin-11 receptor alpha-chain gene, IL11RA, located on chromosome 9p13. *Genomics* 1996;31: 65–70.
- [22] Qian J, Yang J, Zhang X, et al. Isolation and characterization of a novel cDNA, UBAP1, derived from the tumor suppressor locus in human chromosome 9p21–22. *J Cancer Res Clin Oncol* 2001;127: 613–8.