

Published in final edited form as:

*Am J Hematol.* 2012 April ; 87(4): 377–383. doi:10.1002/ajh.23118.

## Genome-wide Association Study of N370S Homozygous Gaucher Disease Reveals the Candidacy of *CLN8* gene as a Genetic Modifier Contributing to Extreme Phenotypic Variation

Clarence K. Zhang<sup>1,\*</sup>, Philip B. Stein<sup>2,\*</sup>, Jun Liu<sup>2</sup>, Gregory M. Pastores<sup>3</sup>, Zuoheng Wang<sup>4</sup>, Ruhua Yang<sup>2</sup>, Judy H. Cho<sup>5</sup>, Peter K. Gregersen<sup>6</sup>, Johannes M. F. G. Aerts<sup>7</sup>, Hongyu Zhao<sup>1,4</sup>, and Pramod K. Mistry<sup>2,5</sup>

Clarence K. Zhang: clarence.k.zhang@yale.edu; Philip B. Stein: philip.stein@yale.edu; Jun Liu: jun.liu@yale.edu; Gregory M. Pastores: gregory.pastores@nyumc.org; Zuoheng Wang: zuoheng.wang@yaledu; Ruhua Yang: ruhua.yang@yale.edu; Judy H. Cho: judy.cho@yale.edu; Peter K. Gregersen: peterg@nshs.edu; Johannes M. F. G. Aerts: j.m.aerts@amc.uva.nl; Hongyu Zhao: hongyu.zhao@yale.edu

<sup>1</sup>Keck Biotechnology Laboratory Biostatistics Resource, Yale University School of Medicine 333 Cedar Street, LMP 4093, New Haven, Connecticut. 06562 <sup>2</sup>Pediatrics, Yale University School of Medicine 333 Cedar Street, LMP 4093, New Haven, Connecticut. 06562 <sup>3</sup>NYU Department of Neurology, Neurogenetics Unit, 403 East 34th Street New York, NY 10016 <sup>4</sup>Epidemiology and Public Health, Yale University School of Medicine 333 Cedar Street, LMP 4093, New Haven, Connecticut. 06562 <sup>5</sup>Department of Internal Medicine, Yale University School of Medicine 333 Cedar Street, LMP 4093, New Haven, Connecticut. 06562 <sup>6</sup>North Shore University Hospital, Center for Genomics and Human Genetics, Community Drive, Manhasset, NY 11030 <sup>7</sup>University of Amsterdam, Department of Biochemistry, Academic Medical Centre, Amsterdam, the Netherlands

### Abstract

Mutations in *GBA1* gene result in defective acid  $\beta$ -glucosidase and the complex phenotype of Gaucher disease (GD) related to the accumulation of glucosylceramide-laden macrophages. The phenotype is highly variable even among patients harboring identical *GBA1* mutations. We hypothesized that modifier gene(s) underlie phenotypic diversity in GD and performed a GWAS study in Ashkenazi Jewish patients with type 1 GD (GD1), homozygous for N370S mutation. Patients were assigned to mild, moderate or severe disease category using composite disease severity scoring systems. Whole-genome genotyping for >500,000 SNPs was performed to search for associations using OQLS algorithm in 139 eligible patients. Several SNPs in linkage disequilibrium within the *CLN8* gene locus were associated with the GD1 severity: SNP rs11986414 was associated with GD1 severity at p value  $1.26 \times 10^{-6}$ . Compared to mild disease, risk allele A at rs11986414 conferred an odds ratio of 3.72 for moderate/severe disease. Loss of function mutations in *CLN8* causes neuronal ceroid-lipofuscinosis but our results indicate that its *increased* expression may protect against severe GD1. In cultured skin fibroblasts, the relative expression of CLN8 was higher in mild GD compared to severely affected patients in whom CLN8 risk alleles were over-represented. In an *in vitro* cell model of GD, CLN8 expression was increased which was further enhanced in the presence of bioactive substrate, glucosylsphingosine. Taken together, CLN8 is a candidate modifier gene for GD1 that may function as a protective

Corresponding author: Pramod K. Mistry, MD, PhD, FRCP, Professor of Pediatrics and Internal Medicine, Chief, Pediatric Gastroenterology and Hepatology, Yale University School of Medicine, PO Box 208064, 333 Cedar Street; LMP 4093, New Haven, CT 06520, Telephone: 203-785-4649, Fax: 203-785-3365, pramod.mistry@yale.edu.

\*These authors contributed equally

sphingolipid sensor and/or in glycosphingolipid trafficking. Future studies should explore the role of CLN8 in pathophysiology of GD.

## Keywords

Gaucher disease; GWAS; genotype/phenotype correlations; phenotypic diversity; modifier genes; CLN8; N370S; GBA mutations

## Introduction

In Gaucher disease (GD), the defective activity of acid  $\beta$ -glucosidase due to mutations in the *GBA1* gene leads to progressive accumulation of the substrate, glucocerebroside in the lysosomes of mononuclear phagocytes [1]. The result is widespread accumulation of glucocerebroside-laden macrophages, systemic macrophage activation and a complex phenotype involving the liver, the spleen, the bones and the bone marrow [1]. Non-neuronopathic Type 1 Gaucher disease (GD1) represents the most common form accounting for >90% of currently known patients. Acute and chronic neuronopathic forms, type 2 and type 3 GD respectively, manifest neurodegenerative disease in addition to the systemic features [2]. Among lysosomal storage diseases, GD is the most common, with overall frequency of  $\sim 1$  in 40,000 individuals but in the Ashkenazi Jews, as many as  $\sim 1$  in 500 individuals are affected [1]. GD1 has served as an excellent prototype for genotype/phenotype delineation and therapeutic innovation for lysosomal storage diseases [3]. However, the pathways from lysosomal accumulation of glucocerebroside to clinical manifestations of GD are not known and this has hindered the delineation of basis of phenotypic variation [4-6].

The gene encoding acid  $\beta$ -glucosidase (*GBA1*) and its highly homologous pseudogene (96% identity) are closely linked on human chromosome 1q21 [7]. More than 200 mutations in the active *GBA1* gene have been catalogued, but only 4 account for >90% of disease alleles in patients of Ashkenazi Jewish ancestry and  $\sim 50\%$  of disease alleles in non-Jewish patients [1, 7]. In studies from North America and Europe, N370S comprise up to 70% of all disease alleles; N370S mutation occurs in homozygous form typically in patients of Ashkenazi Jewish ancestry and it is estimated  $\sim 1$  in 15 individuals in this population are carriers of this mutation [8]. A disease mutation comprising 55bp deletion in exon 9 spans the site of N370S mutation, which results in erroneous genotype assignment (N370S homozygous in N370S/55bp deletion) when allele-specific PCR assays are employed [7, 9].

There is some genotype/phenotype correlation so that the presence of at least one N370S allele absolutely precludes the development of neuronopathic disease [1]. As a group, N370S homozygous patients tend to have less severe disease compared to compound heterozygous N370S/L444P or N370S/84GG [7, 8, 10, 11]. However, within each *GBA1* genotype category and even among affected siblings there is extreme variation of disease severity [11, 15,16]. In fact, it has been suggested that homozygosity for N370S mutation results in such mild disease that a significant number never come to medical attention; yet others develop incapacitating disease at young age [7, 12].

The basis of such extraordinary phenotypic variation in this single gene/single enzyme disorder and widely regarded as involving a single cell type (the mononuclear phagocyte), is not understood. Studies have demonstrated lack of correlation between the residual acid  $\beta$ -glucosidase activity and disease severity, however these studies have been questioned due to failure to measure *in situ* lysosomal enzyme activity in intact cells [13, 14]. Sib-pairs with GD have been reported to exhibit different pattern of disease manifestations as well differ in

overall disease severity [15, 16]. Therefore, modifier gene(s) have been proposed to underlie phenotypic variation [13]. It should be kept in mind, that two pairs of identical twins have been reported with divergent phenotypes, suggesting that the basis of phenotypic diversity may be more complex [17,18].

A number of modulators of residual enzyme activity have been proposed as determinants of disease severity, i.e. lysosomal pH [19], saposin [20] and intracellular trafficking [21]. *A priori*, the determinants of phenotype variation in GD among patients with the same *GBA1* genotype could arise from environmental factors or through role of genetic modifier(s). Anecdotal evidence suggests that environmental factors through viral infections, i.e., EBV, may accelerate disease course [22]. With respect to potential genetic modifier(s) in the gene dense *GBA1* locus there are eleven widely distributed polymorphic sites that are in linkage disequilibrium. All N370S chromosomes appear to harbor an identical haplotype, regardless of geographic origin of the patient [7]. Therefore variation in disease severity cannot be explained on the basis of other mutations or polymorphic variations in the *GBA1* gene locus, the promoter or enhancer sequences—at least in this ethnic group [7, 23]. Studies have explored SNPs in genes encoding glucosylceramide synthase gene, *GBA2*, *GBA3*, vitamin D receptor, IL6 and TNF alpha, but there was no consistent correlation with disease severity [24-28]. No other plausible candidate genes have emerged based on current understanding of pathophysiology to explain phenotypic variation in GD...

We conducted a hypothesis-generating GWAS through unbiased exploration for association of SNPs with GD1 severity. The study was focused on N370S homozygous GD1 patients of Ashkenazi Jewish ancestry who exhibit extreme phenotypic variation with the rationale of conducting the study in genetically homogenous population to minimize population stratification.

## Methods

### Determination of Disease Severity

Patients were enrolled at the Gaucher Disease Treatment Centers of Yale School of Medicine and NYU School of Medicine. The study was approved by the Human Investigation Committee of the Yale University of Medicine and the Institutional Review Board of NYU School of Medicine.

Eligible patients represented consecutive N370S homozygous GD patients in whom comprehensive phenotype data prior to enzyme replacement therapy (ERT) was available. Patients had confirmed diagnosis of GD1 based on low (i.e., <10% of normal) acid  $\beta$ -glucosidase activity in peripheral blood leucocytes and *GBA1* mutation analysis. *GBA1* genotyping was performed as described previously to ensure absence of 55bp deletion in exon 9 [9]. Database included longitudinal phenotype data regarding bone, hematology, visceral involvement, age of onset, severity score index and whether prior splenectomy had been performed. Liver and spleen volumes were determined by volumetric MRI and depicted as fold-enlargement compared to normal ( $\times N$ ); normal liver volume is 2.5% body weight and spleen volume 0.2% body weight as described previously [10]. Hemoglobin, platelet count, MRI, Hermann Score [10], DXA scan, Bone Marrow Burden (BMB) score [29] were determined prior to ERT. Overall disease severity was determined according to Severity Score Index (SSI) [30] and the recently validated DS3 score [31]. Age of onset of symptoms, age at splenectomy and age at initiation of ERT were also recorded in the database. We aimed to assess association of SNPs based on ordinal ranking of disease severity, and therefore, the patients were assigned categories of mild, moderate and severe disease (Table I). After initial scoring via DS3 and SSI, patients were ranked by disease severity from most mild to severe. Additionally, experts in Gaucher disease (PKM and

GMP) who have followed patients for >10 yrs assigned a global assessment of disease severity of the patient. Using the DS3, SSI score and global assessments, patients were assigned to severity categories of mild, moderate and severe. Illustrative natural history of mildly affected GD1 patient compared to severely affected patient is depicted in Supplementary Figure 1. In addition to variation in overall disease severity, GD is recognized to exhibit different patterns of disease among affected patients, i.e., visceral vs skeletal disease. Therefore, in addition to overall disease severity, we assessed severity of individual disease compartments. A category designation was created for bone disease severity as well as severity of visceral disease. For bone disease, severity was based on pre-ERT Hermann score as described previously [10] with scores 1-2 classified as mild, 3 as moderate, and 4-5 as severe. Patients were also assessed based on visceral disease with spleen size:  $1-5 \times N$  as 0 point,  $5-10 \times N$  as 1 point,  $10-15 \times N$  as 2, and greater than 15 or splenectomy as 3 points. Severity scoring for hepatomegaly was as follows: less than  $1.25 \times N$  as 0 point,  $1.25-1.5 \times N$  as 1 point,  $1.5-1.75$  as 2 point and greater than 1.75 as 3 points; presence of cirrhosis/portal hypertension or hepatopulmonary syndrome as 4 points. Aggregate cores for liver and spleen involvement assigned severity of visceral disease: 0-1 classified as mild; 2-3 points, moderate and 4 points, severe.

Demographic information of the study population is depicted in Table I.

### Genotyping and Patient Exclusion Criteria

A total of 186 GD1 patients confirmed to be homozygous for N370S mutation were genotyped on Illumina 650 SNP array, which contains 620,901 SNPs. The GenTrain 2.0 calling algorithm implemented in GenomeStudio from Illumina was used to infer the genotype. Quality control measures included: excluding samples with <95% call rate, excluding SNPs with call rates <94%, MAF less than 0.5% and HWE p-value  $<1 \times 10^{-5}$ . The above quality control measurements were implemented in PLINK package [32].

Additionally, we excluded patients under age 18 yrs who were diagnosed through genetic screening because these patients tend to be asymptomatic and exhibit minimal or even no manifestations of the disease. Patients were also excluded if there were insufficient records for determination of SSI and DS3 scores prior to ERT. After these exclusions, we were left with 139 patients for the final analysis of SNP association with disease severity.

### Family Structure and Association Analysis

The Kinship program [33] was used to identify cryptic family relationship directly from genotypes. It estimates the IBD (identical by descent) coefficients and calculates kinship coefficients. These coefficients were used to infer duplicated samples, siblings and parent-offspring relationships.

After family relationship among individuals was established, a newly developed method, OQLS was used to compare the allele frequency of SNPs among different phenotype categories while accounting for the ordinal nature of trait information. OQLS (Wang, Z, unpublished) is an extension of the single-marker testing method for a binary trait, MQLS [34]. The advantage of OQLS/MQLS is that it is applicable to completely general combinations of family and case-control designs. OQLS/MQLS explicitly incorporates the phenotyped relatives which provide additional information on the enrichment of predisposing variants, and uses the dependence structure to derive optimal weights to increase the power of association analysis.

For top association regions, LD pruning was performed as described in the PLINK package. SNPs were pruned if their  $r^2$  were larger than 0.8. The HaploView [35] was used to calculate and visualize the fine scale LD structure and infer haplotype blocks. The SNAP tool [36]

was used to plot the regional association result and LD values. The eQTL browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>) was used to search for SNPs associated with the expression of CLN8 [37].

### Skin Fibroblast Cell Lines

Two millimeter punch skin biopsy was obtained from 6 patients (3 with severe GD and 3 with mild GD), and 2 healthy controls using the AcuPunch kit (Acuderm Inc, Lauderdale, FL). A sterile 12mm circular cover glass was placed on each of the skin biopsy fragments and 6ml of primary growth culture (DMEM supplemented with 15% fetal calf serum, 1% penicillin/streptomycin/gentamicin, 1% NEAA) was added. A fibroblast cell line was established once the outgrowth of fibroblasts skin biopsy covered 75% - 90% of the dish. The cells were digested with 0.05% trypsin/EDTA and split at a ratio of 1:3 to 1:4. Cell lines between 4 - 7 generations were used for the experiments.

### *In vitro* studies with fibroblast cell lines

The inhibitor of glucocerebrosidase, conduritol B epoxide (CBE) was purchased from Calbiochem (La Jolla, CA); Glucocerebroside (GL1) and Glucosylsphingosine (Lyso GL1) were purchased from Matreya (Pleasant Gap, PA). Stock solutions of the drug were prepared according to the manufacturer's instructions. Final concentrations of 50 $\mu$ M of CBE, 40 $\mu$ M of GL1, and 20 $\mu$ M of Lyso GL1 were used to treat the cells as described previously [37a]. CBE alone, CBE with GL1 or Lyso GL1 were separately added to the two normal (healthy) cell lines with the same amount of DMSO as control. After incubation for the specified periods, RNA was extracted for real time PCR.

Total RNA from each cell line was prepared using the TRIzol Plus RNA Purification kit (Invitrogen, Carlsbad CA). Briefly, adherent fibroblast cells were scratched and concentrated to an Eppendorf micro-centrifuge tube to which 1ml of TRIzol reagent was added. Following phase separation, the solutions containing RNA were transferred to a spin cartridge. After binding and washing, RNA was eluted into a fresh microtube and RNA concentration was determined.

First strand cDNA synthesis was performed using the AffinityScript Multiple Temperature cDNA Synthesis kit (Agilent Tech, Santa Clara, CA) according to the manufacturer's protocols. Oligo dT was used as primer for the reverse transcription. RNA-Oligo dT mixture was preheated at 70°C for 5min before gradual cooling at room temperature for 10min, then followed by addition of the reaction mixture on ice and at 42°C for 5min and 50°C for 55min.

Quantitative real-time PCR was performed using a 7500 Real-time PCR system (Applied Biosystems, Carlebad, CA). Primer and probe mixtures for CLN8 and Actin- $\beta$  (used as internal reference control) were ordered from the inventory of Applied Biosystems and pre-labeled with the reporter dye 6-carboxyfluorescein (6'-FAM) at the 5' end and with the quencher dye 6-carboxytetramethylrodamine (TAMRA) at the 3' end. Quantitative PCR was performed with 2 $\mu$ l of 1:3 diluted cDNA, 10 $\mu$ l of the FastStart universal probe master (Roche, Mannheim, Germany), 0.5 $\mu$ l of 900nM primers, and 0.5 $\mu$ l of 250nM probe in a final volume of 20 $\mu$ l. Each reaction was set up in triplicate and the mean values were used for analysis.

The study was approved by the Human Investigation Committee of the Yale University of Medicine and IRB of NYU School of Medicine.



## Results

### Genotype Quality Control and Family Structure

Five out of 186 samples had genotyping call rates 95% and these samples were removed from further analysis. Among the remaining 181 samples, 2 samples had call rates between 96 and 97%. The remaining 179 samples had call rates > 99%. After SNP cleaning using the quality control criteria described in methods section, 554,340 SNPs were left. The total genotyping rate in post-QC samples is 99.8%. There were 540,902 autosomal SNPs retained for final association analysis.

From pair-wise estimated distribution of IBD coefficients and kinship coefficients, we inferred 14 nuclear families. Among these 14 families, 3 were parent-offspring, 10 were sibling pairs and one family had 3 siblings (supplementary table IB). We found two categories of kinship coefficients, 31 pairs near 0.25 representing parental-child or sibling relationship (supplementary table IA). For a parent offspring pair, they are anticipated to share 1 allele IBD, whereas for a sib pair, the probabilities of sharing 2, 1 or 0 alleles IBD are 0.25, 0.5 and 0.25, respectively. Other sample pairs had kinship coefficients less than 0.05, and therefore less likely to be related with each other.

### Association Results

After applying the OQLS algorithm to our ordinal valued phenotype in a combined sample of family and unrelated cases/controls, the genetic association of 540,902 autosomal SNPs was identified in the three categories of disease severity (overall disease severity, severity of visceral disease and severity of bone disease). Figure 1 shows the Manhattan plot for the genome wide association results. No markers achieved genome wide significance based on the stringent Bonferroni correction ( $p < 9 \times 10^{-8}$ ). However, the Bonferroni correction assumes all markers behave independently. In practice, a significant proportion of markers do not segregate independently because they are in LD with each other. Therefore, the top signals displayed in the Manhattan plot are worthy of further investigation. Consequently, we analyzed SNPs with lowest P values for associations with disease severity. In addition, cluster of top associations displayed in the Manhattan plot (spikes) were examined in detail.

Supplementary Table II lists the top associated markers with each of the three disease severity categories. If not within the known gene region, their nearest gene neighborhoods and distances were displayed. The distance to the start of the gene (disstart) was calculated by the gene start position minus the SNP location; the distance to the end of the gene (distend) was calculated by the gene end position minus the SNP location.

When assessing SNP association with overall disease severity, the region harboring *CLN8* gene achieved the strongest association signal. In addition, the *CLN8* gene region also demonstrated strong association with the severity of visceral disease (Supplementary table II). Association signals for SNPs within *CLN8* and its 100 kb flanking regions with the overall disease severity are depicted in Figure 2a. LD structure within 100 kb flanking region is displayed in Figure 2b. Closer resolution of LD structure and haplotype block within the 25 kb flanking region is displayed in Figure 2c. Several SNPs within or near *CLN8* regions are in strong LD with each other and majority of these SNPs are strongly associated with the overall disease severity. Supplementary table S III shows the detailed association results for each SNP within the *CLN8* region for the overall disease category. Interestingly, among these SNPs two intronic SNPs (rs11136424 and rs4875958, pvalues  $1.59E^{-05}$  and  $2.69E^{-05}$ ) are strongly associated with *CLN8* expression as depicted in the eQTL browser. SNP rs11136424 is associated with *CLN8* expression in lymphoblastoid cell lines and SNP rs4875958 is associated with *CLN8* expression in liver tissue (Supplementary Figure 2). Both eQTL data are derived from people of European ancestry [37].

Among SNPs within the CLN8 region, the topmost SNP rs11986414 achieved a p value of  $1.26 \times 10^{-6}$ . The two alleles of this SNP are A and G; A is the risk allele associated with more severe disease. As shown in Supplementary Table IVa, there was over-representation of A allele in patients with moderate and severe disease compared to the patients with mild disease. Compared to patients with mild disease, A allele conferred odds ratio (OR) of 3.62 (95% CI 1.96 - 6.70) for moderate disease and 3.85 (95% CI 1.96 - 7.55) for severe disease (Table S IVb). The OR between moderate and severe disease categories is 1.06 (95% CI 0.51 - 2.22). Therefore genetic risk conferred by rs11986414 appears to be similar for both moderate and severe disease. The OR with combined severe and moderate vs mild disease category is 3.72 (95% CI 2.19 - 6.31). To understand the genetic inheritance model, the genotypic odds ratios for combined severe and moderate category vs mild disease were further evaluated. Supplementary table IVc shows the OR for AA vs AG, AG vs GG and AA vs GG: 2.38 (95% CI 0.76 - 9.04), 4.52 (95% CI 2.05 - 9.97) and 11.84 (95% CI 3.33 - 42.0), respectively. The large confidence interval for odds ratios with genotype AA are due to the fact that there are only 4 patients with genotype AA in the mild category. These results are consistent with the notion that there is a gene dosage effect, where each addition of allele A increases the risk of severe disease.

For SNPs association with the severity of bone disease, the top SNP is rs510933 on chromosome 6 (p value  $7.87E^{-07}$ ) (Supplementary table IIc). There are 4 other SNPs from this region. However, this region is devoid of genes within 200kb. The most proximal genes near this region are around 500kb away and include STXBP5, *SAMD5*, *SASH1*. A noteworthy candidate modifier gene for severity of bone disease is indicated by SNP rs7920091, a SNP in *ADAM12* (p= $1.11E^{-05}$ ). *ADAM12* is member of the ADAM (a disintegrin and metalloprotease) protein family. Polymorphisms in this gene have been associated with osteoarthritis and its progression [38]. The ADAM family is also thought to regulate human osteoclast differentiation and activity [39]. Other noteworthy associations were with SNPs rs6129532 and rs6124225 on chromosome 20 linked with HSPE pseudogene [HSPE1P1] (p value  $1.39E^{-05}$  and  $2.12E^{-05}$  respectively) and rs11087800, intronic SNP in *ANGPT4* (p value  $1.45E^{-05}$ ). *ANGPT4* is involved in angiogenesis. This may be relevant since avascular osteonecrosis in GD is considered a primarily vascular complication reflecting endothelial cell involvement in GD and may involve impaired angiogenic support due to osteoblastic dysfunction [6, 40, 41].

### Expression of CLN8 in skin fibroblasts and macrophages

Using Real Time PCR, CLN8 relative expression was higher in GD patients compared to controls (Figure 3b). There was also a trend of GD patients with mild disease having higher expression of CLN8 compared to GD patients with severe disease. To assess the relationship between relative expression of CLN8 and the number of risk alleles in *CLN8* gene region, two SNPs (rs4595147 and rs11986414) with the lowest LD were selected from the associated SNPs within the CLN8 region. Of the 6 samples analyzed (3 severe and 3 mild patients), the total number of risk alleles appear to be marginally correlated with relative expression (correlation coefficient -0.796, p value 0.058). In control fibroblasts cultured with CBE, an *in vitro* chemically-induced cellular model of GD, there was slight increase of CLN8 expression compared to untreated cells (Figure 3a). There was more striking increase (2 - 2.5 fold) of CLN8 expression in CBE-treated cells exposed to potent bioactive lipid substrate of acid  $\beta$ -glucosidase, Lyso GL1. Of the two lipid substrates of acid  $\beta$ -glucosidase, Lyso-GL1 was the more potent inducer of CLN8 expression compared to GL1 (Figure 3a). We performed RT PCR for CLN8 expression in primary monocyte-derived macrophages of GD1 patients, and the results were similar to the findings in skin fibroblasts (data not shown). To further assess the expression pattern of CLN8 in different tissues, we interrogated Novartis expression atlas ([www.biogps.org](http://www.biogps.org)) to determine whether it is

expressed preferentially in immune cells [42, 43]. Supplementary Figure 3 depicts the tissue specific expression pattern of CLN8. Interestingly, highest expression is observed in immune cells including macrophages and monocytes. Curiously, lowest expression was observed in brain tissues, which is most vulnerable to loss of function mutations in this gene.

## Discussion

The basis of remarkable variation of severity of GD1 and its heterogeneity among patients with identical *GBA1* gene mutations is not understood [1]. Therefore, it is not possible to identify patients who will benefit from pre-emptive intervention with expensive ERT or small molecule therapy before onset of incapacitating disease manifestations. Identification of genetic modifier(s) will enhance the ability to accurately predict natural history in individual patients in order to develop a personalized monitoring and therapeutic strategy as well as refine genetic counseling.

There are significant inherent obstacles to conduct robust GWAS studies for discovery of modifier gene(s) in a Mendelian disorder such as GD. The chronicity and the heterogeneity of the disease as well as the lack of sufficiently well-characterized patients limit the ability to conduct such studies. The greatest hurdle is that the new generation of patients tends to be treated early before the full natural history unfolds [52,53]. Therefore the accuracy of ascertainment of true disease phenotype with respect to severity is severely compromised. Moreover, the instruments to depict disease severity have limitations to adequately capture the full scope of heterogeneity of disease manifestations. The latter is exemplified most strikingly by some GD patients who predominantly exhibit bone manifestations with only mild visceral or hematological disease and vice versa; these patterns are in contrast to occasional patients who manifest uniformly severe disease in all disease compartments [10]. These patterns of disease, may in fact be driven by modifier gene(s) that underlie monocyte trafficking and/or macrophage-mediated determinants of individual organ injury. Therefore, in our study, we elected to depict disease severity by multiple approaches, including SSI, validated DS3 as well as global physician assessment by two investigators with extensive experience of GD. Moreover, we also examined individual disease compartments, including the viscera and the bone. Thus the final dataset of mild, moderate and severe disease are remarkably consistent with these classifications in the context of overall disease severity as well as severity of individual disease compartments (Table I).

Our study revealed several candidate modifier genes, not previously considered on the basis of current understanding of GD pathophysiology: CLN8, ADAM12 and ANGPT4. The lead candidate modifier gene was *CLN8*. The physiological function of *CLN8* is not known but a loss of function mutation is responsible for a subtype of neuronal ceroid-lipofuscinosis (NCL), the Northern epilepsy mental retardation (EPMR) syndrome, characterized by seizures in childhood and progressive mental retardation [44]. The *CLN8* gene encodes a predicted 286-amino acid transmembrane protein containing a C-terminal ER-retrieval signal (KKRP) [45]. Consistent with this structure, *CLN8* protein, a nonglycosylated membrane protein, is located in the ER and shuttles between the ER and the ER-Golgi intermediate compartment. Highest expression of CLN8 is surprisingly in the liver and the spleen (compared to the brain) and it has been suggested that it plays an important role as a sensor of sphingolipids and sphingolipid trafficking in all cell types [46]. *CLN8* is a member of a family of homologues that includes ER-resident proteins: yeast Lag1p and human TRAM (translocating chain-associating membrane protein) [47]. A C-terminal motif of 200 amino acid residues designated the Tram- Lag-CLN8 (TLC) domain in these proteins has been proposed to play a role in ceramide synthesis and glycosphingolipid trafficking [47].



In an induced *in vitro* cellular model of GD, we found there was an induction of CLN8 expression and the level of expression was further increased in the presence of potent bioactive lipid substrate of acid  $\beta$ -glucosidase, Lyso-GL1. In cells derived from, GD1 patients, there was increased basal expression of CLN8 relative to control cells; moreover there was a trend towards increased expression in patients with mild disease compared to patients with moderate/severe disease. We found the SNP alleles in CLN8 region associated with severe GD correlated with relatively lower expression of CLN8 compared to the lower risk alleles (Figure 3). Interestingly, CLN8 and CLN6, two integral ER trans-membrane proteins and cathepsin D, are increased in expression (3-4-fold) in the spleens of mice with conditional deletion of *GBA1* [6]; loss of function mutations in these three genes result in NCLs [48]. Moreover, MYC, a regulator of *CLN8* was marginally associated with GD1 severity (top SNP p value  $4.96 \times 10^{-5}$ ). Finally interrogation of Novartis gene expression pattern highlighted prominent relative expression of CLN8 in immune cells compared to other cell types and organs. Taken together, CLN8 is a plausible candidate modifier gene of GD and its induction likely protects cells from the damaging effect of accumulating lipids in GD1.

Interestingly loss of function mutation in CLN8 results in lysosomal accumulation of autofluorescent material but with lower levels of plasmalogen [47] and the latter has been suggested to promote lipid peroxidation. In fact, in GD1, serum plasmalogen levels are reduced, the baseline levels correlate negatively with chitotriosidase and furthermore increase after ERT [50]. This observation is also consistent with our notion that induction of CLN8 may be protective towards damaging cellular effects of glycosphingolipid accumulation in GD1.

*CLN8* and the other potential candidate modifier genes (*ADAMI2* and *ANGPT4*) identified in our GWAS do not achieve the threshold for genome-wide significance after stringent Bonferroni correction [51]. Given the potential impediments to conduct a replication study, an alternative approach to define the potential role of CLN8 as a modifier of GD will derive from studies in transgenic mice since CLN8 KO as well as conditional KO mouse models of GD are available [44, 54].

In conclusion, *CLN8* is a plausible candidate modifier gene for Gaucher disease based on our GWAS results and *in vitro* studies in chemically-induced cellular model of GD. Further studies are warranted to define the physiologic role of CLN8 in health and its role in GD pathophysiology. The availability of mouse models should facilitate these studies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the patients at Yale and NYU Gaucher Disease Treatment Centers for participation in this study. The study was supported by the National Gaucher Foundation and by Lysosomal Fellowship Training Grant from Genzyme-Sanofi to PS. PKM thanks Dr Jeffrey Gruen and Karen K. Weisbord for advice.

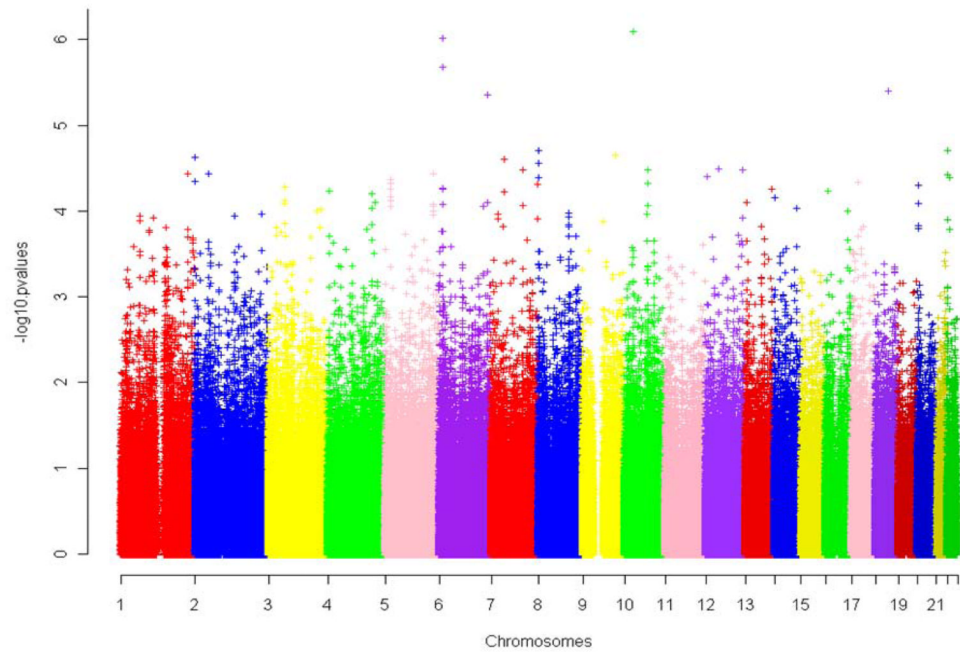
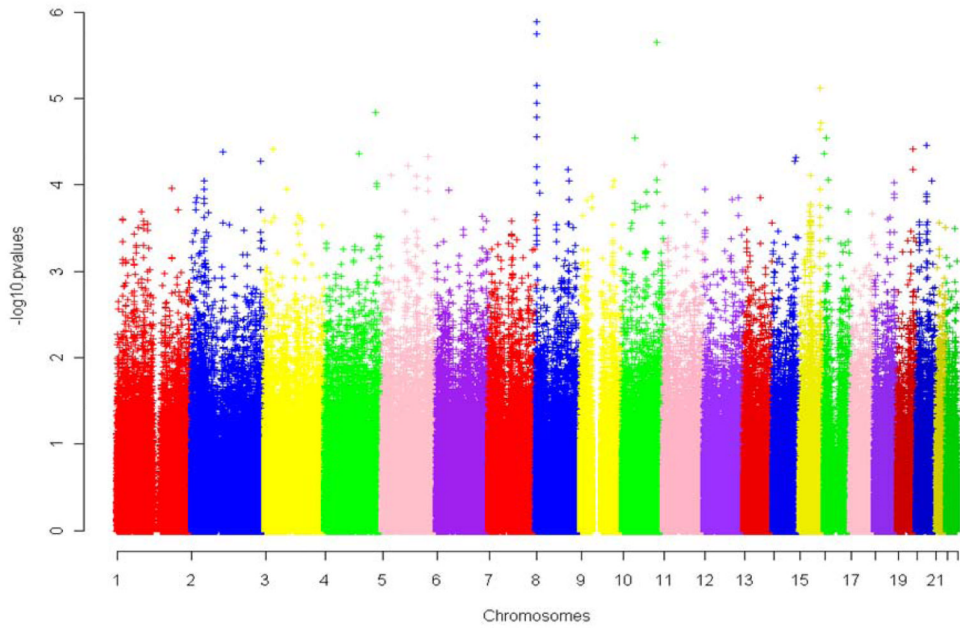
Contract grant sponsor: PKM is supported by NIDDK K24DK066306 mid-career clinical investigator award, the National Gaucher Foundation and a grant from Genzyme Corporation for investigator-initiated study on biomarker discovery and validation. PS was supported by a Lysosomal Storage Disease Fellowship Training Grant from Genzyme-Sanofi.

## References

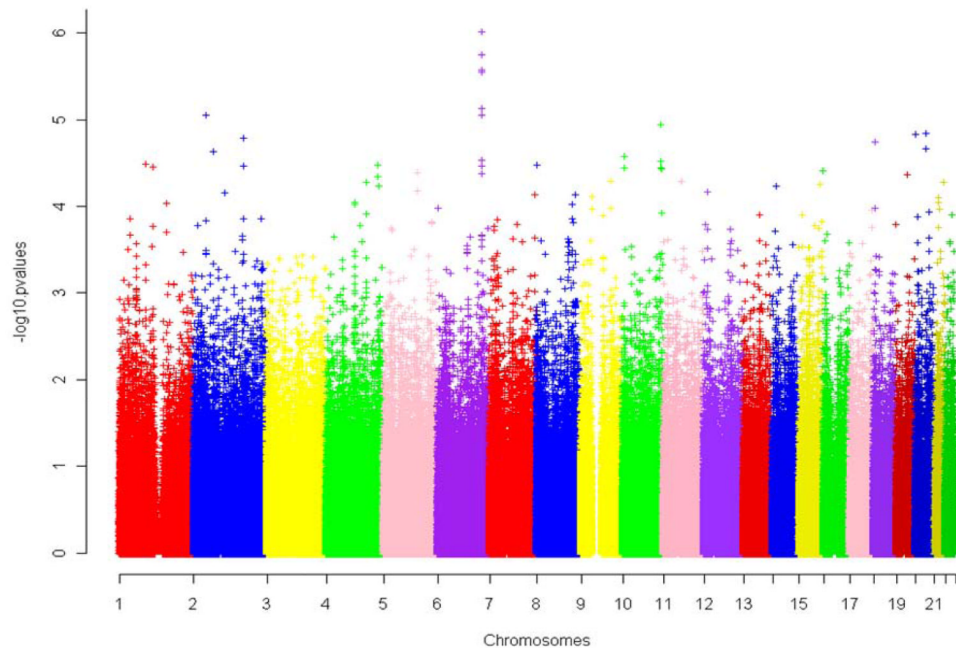
1. Grabowski GA. Phenotype, diagnosis, and treatment of Gaucher's disease. *Lancet*. 2008; 372(9645): 1263–71. [PubMed: 19094956]
2. Brady RO, Barton NW, Grabowski GA. The role of neurogenetics in Gaucher disease. *Arch Neurol*. 1993; 50(11):1212–24. [PubMed: 8215980]
3. Cox TM. Gaucher disease: clinical profile and therapeutic developments. *Biologics*. 4:299–313. [PubMed: 21209725]
4. Jmoudiak M, Futerman AH. Gaucher disease: pathological mechanisms and modern management. *Br J Haematol*. 2005; 129(2):178–88. [PubMed: 15813845]
5. Kitatani K, et al. Acid beta-glucosidase 1 counteracts p38delta-dependent induction of interleukin-6: possible role for ceramide as an anti-inflammatory lipid. *J Biol Chem*. 2009; 284(19):12979–88. [PubMed: 19279008]
6. Mistry PK, et al. Glucocerebrosidase gene-deficient mouse recapitulates Gaucher disease displaying cellular and molecular dysregulation beyond the macrophage. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:19473–19478. [PubMed: 20962279]
7. Hruska, KS.; LaMarca, ME.; Sidransky, E. Gaucher disease: Molecular biology and genotype/phenotype correlations In Gaucher disease Eds AH Futerman and A Zimran. CRC Press;
8. Grabowski GA. Gaucher disease: gene frequencies and genotype/phenotype correlations. *Genet Test*. 1997; 1(1):5–12. [PubMed: 10464619]
9. Torralba MA, et al. High prevalence of the 55-bp deletion (c.1263del55) in exon 9 of the glucocerebrosidase gene causing misdiagnosis (for homozygous N370S (c.1226A > G) mutation) in Spanish Gaucher disease patients. *Blood Cells Mol Dis*. 2002; 29(1):35–40. [PubMed: 12482401]
10. Taddei TH, et al. The underrecognized progressive nature of N370S Gaucher disease and assessment of cancer risk in 403 patients. *Am J Hematol*. 2009; 84(4):208. [PubMed: 19260119]
11. Mistry PK, et al. A reappraisal of Gaucher disease-diagnosis and disease management algorithms. *Am J Hematol*. 86(1):110–5. [PubMed: 21080341]
12. Beutler E, et al. Gaucher disease: gene frequencies in the Ashkenazi Jewish population. *Am J Hum Genet*. 1993; 52(1):85–8. [PubMed: 8434610]
13. Goker-Alpan O, et al. Divergent phenotypes in Gaucher disease implicate the role of modifiers. *J Med Genet*. 2005; 42(6):e37. [PubMed: 15937077]
14. Aerts JM, et al. Biomarkers in the diagnosis of lysosomal storage disorders: proteins, lipids, and inhibobodies. *J Inherit Metab Dis*. 34(3):605–19. [PubMed: 21445610]
15. Amato D, et al. Gaucher disease: variability in phenotype among siblings. *J Inherit Metab Dis*. 2004; 27(5):659–69. [PubMed: 15669682]
16. Elstein D, et al. Disease severity in sibling pairs with type 1 Gaucher disease. *J Inherit Metab Dis*. 33(1):79–83. [PubMed: 20049528]
17. Lachmann RH, et al. Twin pairs showing discordance of phenotype in adult Gaucher's disease. *QJM*. 2004; 97(4):199–204. [PubMed: 15028849]
18. Biegstraaten M, et al. A monozygotic twin pair with highly discordant Gaucher phenotypes. *Blood Cells Mol Dis*. 46(1):39–41. [PubMed: 21056933]
19. van Weely S, et al. Role of pH in determining the cell-type-specific residual activity of glucocerebrosidase in type 1 Gaucher disease. *J Clin Invest*. 1993; 91(3):1167–75. [PubMed: 8450045]
20. Grabowski GA, Horowitz M. Gaucher's disease: molecular, genetic and enzymological aspects. *Baillieres Clin Haematol*. 1997 Dec; 10(4):635–56. [PubMed: 9497856]
21. Ron I, Horowitz M. ER retention and degradation as the molecular basis underlying Gaucher disease heterogeneity. *Hum Mol Genet*. 2005 Aug 15; 14(16):2387–98. [PubMed: 16000318]
22. Kolodny EH, et al. Phenotypic manifestations of Gaucher disease: clinical features in 48 biochemically verified type 1 patients and comment on type 2 patients. *Prog Clin Biol Res*. 1982; 95:33–65. [PubMed: 6289358]
23. Diaz GA, et al. Gaucher disease: the origins of the Ashkenazi Jewish N370S and 84GG acid beta-glucosidase mutations. *Am J Hum Genet*. 2000; 66(6):1821–32. [PubMed: 10777718]

24. Beutler E, West C. Polymorphisms in glucosylceramide (glucocerebroside) synthase and the Gaucher disease phenotype. *Isr Med Assoc J.* 2002; 4(11):986–8. [PubMed: 12489486]
25. Dekker N, et al. The cytosolic beta-glucosidase GBA3 does not influence type 1 Gaucher disease manifestation. *Blood Cells Mol Dis.* 46(1):19–26. [PubMed: 20728381]
26. Altarescu G, et al. The interleukin-6 promoter polymorphism in Gaucher disease: a new modifier gene? *QJM.* 2003; 96(8):575–8. [PubMed: 12897342]
27. Altarescu G, et al. TNF-alpha levels and TNF-alpha gene polymorphism in type I Gaucher disease. *Cytokine.* 2005; 31(2):149–52. [PubMed: 15919211]
28. Greenwood A, et al. Vitamin D receptor (VDR) polymorphisms in the cardiac variant of Gaucher disease. *Pediatr Cardiol.* 31(1):30–2. [PubMed: 19784695]
29. Maas M, et al. Quantification of bone involvement in Gaucher disease: MR imaging bone marrow burden score as an alternative to Dixon quantitative chemical shift MR imaging--initial experience. *Radiology.* 2003; 229(2):554–61. [PubMed: 14526090]
30. Zimran A, et al. Gaucher disease. Clinical, laboratory, radiologic, and genetic features of 53 patients. *Medicine (Baltimore).* 1992; 71(6):337–53. [PubMed: 1435229]
31. Weinreb NJ, et al. A validated disease severity scoring system for adults with type 1 Gaucher disease. *Genet Med.* 12(1):44–51. [PubMed: 20027115]
32. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–75. [PubMed: 17701901]
33. Choi Y, Wijsman EM, Weir BS. Case-control association testing in the presence of unknown relationships. *Genet Epidemiol.* 2009; 33(8):668–78. [PubMed: 19333967]
34. Thornton T, McPeck MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet.* 2007; 81(2):321–37. [PubMed: 17668381]
35. Barrett JC, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21(2):263–5. [PubMed: 15297300]
36. Johnson AD, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008; 24(24):2938–9. [PubMed: 18974171]
37. Veyrieras JB, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genetics.* 2008 Oct.4(10):e1000214. [PubMed: 18846210]
- 37a. Sillence DJ, et al. Glucosylceramide modulates membrane traffic along the endocytic pathway. *J Lipid Res.* 2002; 43(11):1837–45. [PubMed: 12401882]
38. Kerna I, et al. Missense single nucleotide polymorphism of the ADAM12 gene is associated with radiographic knee osteoarthritis in middle-aged Estonian cohort. *Osteoarthritis Cartilage.* 2009; 17(8):1093–8. [PubMed: 19268722]
39. Bartholin L, et al. FLRG, a new ADAM12-associated protein, modulates osteoclast differentiation. *Biol Cell.* 2005; 97(7):577–88. [PubMed: 15574124]
40. Elleder M. Glucosylceramide transfer from lysosomes--the missing link in molecular pathology of glucosylceramidase deficiency: a hypothesis based on existing data. *J Inher Metab Dis.* 2006; 29(6):707–15. [PubMed: 17080304]
41. Zaidi M, et al. ACTH protects against glucocorticoid-induced osteonecrosis of bone. *Proc Natl Acad Sci U S A.* 2010 May 11; 107(19):8782–7. [PubMed: 20421485]
42. Su AI, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A.* 2004 Apr 20; 101(16):6062–7. [PubMed: 15075390]
43. Wu C, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology.* 2009; 10:R130. [PubMed: 19919682]
44. Ranta S, et al. The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nat Genet.* 1999 Oct; 23(2):233–6. [PubMed: 10508524]
45. Lonka L, et al. The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. *Hum Mol Genet.* 2000; 9(11):1691–7. [PubMed: 10861296]
46. Lonka L, et al. Localization of wild-type and mutant neuronal ceroid lipofuscinosis CLN8 proteins in non-neuronal and neuronal cells. *J Neurosci Res.* 2004; 76(6):862–71. [PubMed: 15160397]
47. Schorling S, et al. Lag1p and Lac1p are essential for the Acyl-CoA-dependent ceramide synthase reaction in *Saccharomyces cerevisiae*. *Mol Biol Cell.* 2001; 12(11):3417–27. [PubMed: 11694577]

48. Kousi M, Lehesjoki AE, Mole SE. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum Mutat.* 2012 Jan; 33(1):42–63. [PubMed: 21990111]
49. Hermansson M, et al. Mass spectrometric analysis reveals changes in phospholipid, neutral sphingolipid and sulfatide molecular species in progressive epilepsy with mental retardation, EPMR, brain: a case study. *J Neurochem.* 2005; 95(3):609–17. [PubMed: 16086686]
50. Moraitou M, et al. Plasmalogen levels in Gaucher disease. *Blood Cells Mol Dis.* 2008; 41(2):196–9. [PubMed: 18501647]
51. Hunter DJ, Kraft P. Drinking from the fire hose--statistical issues in genomewide association studies. *N Engl J Med.* 2007; 357(5):436–9. [PubMed: 17634446]
52. Weinreb N, et al. A benchmark analysis of the achievement of therapeutic goals for type 1 Gaucher disease patients treated with imiglucerase. *Am J Hematol.* 2008; 83(12):890–5. [PubMed: 18819093]
53. Mistry PK, et al. Timing of initiation of enzyme replacement therapy after diagnosis of type 1 Gaucher disease: effect on incidence of avascular necrosis. *Br J Haematol.* 2009; 147(4):561–70. [PubMed: 19732054]
54. Enquist IB, et al. Effective cell and gene therapy in a murine model of Gaucher disease. *Proc Natl Acad Sci U S A.* 2006 Sep 12; 103(37):13819–24. [PubMed: 16954197]

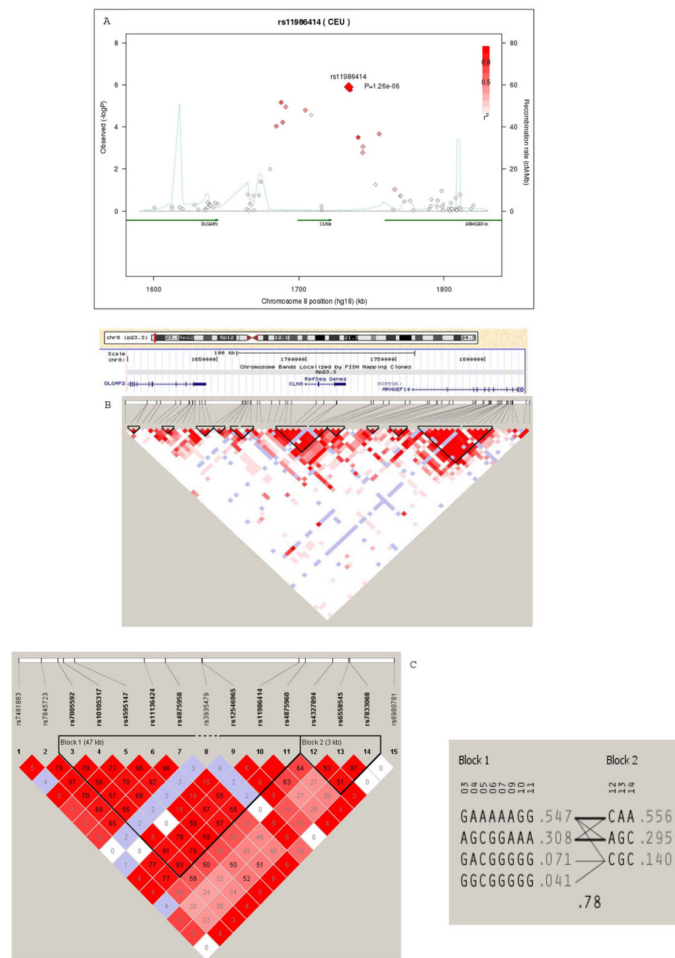






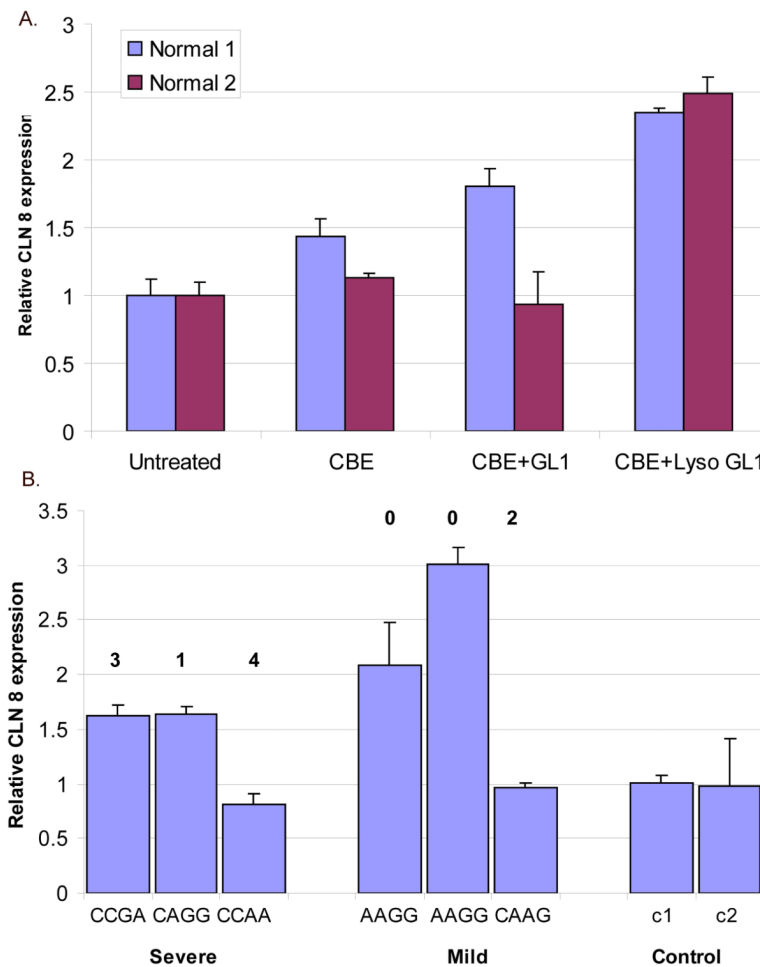
**Figure 1. Manhattan plots for genome wide association results in three disease severity categories**

The x-axis shows the chromosomes arranged according to their genome location. The y-axis shows the  $-\log p$  value of the association strength. Each chromosome is in different colors. A. Overall disease severity. B. Severity of visceral disease. C. Severity of skeletal disease.



**Figure 2. Association plots for the CLN8 and its 100kb flanking region with the overall disease severity**

A. Regional Manhattan plot and LD structure with the top SNP rs11986414. The x-axis shows the chromosome position, the left y-axis shows the  $-\log(p)$  for the association strength, and the right y-axis shows the recombination rate calculated from Hapmap3 CEU reference population. The color for each marker represents the LD  $r^2$  value with the topmost snp rs11986414. B. LD structure. The pair-wise LD pattern is shown for all markers within the 100kb flanking region. The chromosomal location is shown above, the red, blue and white colors represent different LD measured by  $D'$ . The blocks bordered with thick black box represent haplotype blocks inferred. C. LD structure and haplotype block for the 25kb flanking region. The value inside each cell displays the LD measurement  $r^2$ , the color scheme is based on  $D'$ . The inferred haplotype block and haplotype frequency are shown too.



**Figure 3. CLN8 expression in a) healthy control fibroblasts at baseline and after addition of GL1 and Lyso-GL1 b) in GD fibroblasts stratified according to disease severity**

**A.** CBE alone, CBE with GL1 or Lyso GL1 were separately added to the two normal (healthy) cell lines with replacement of medium and drugs every 5-7days. RNA was then extracted and real time PCR was performed for CLN8 expression. Actin- $\beta$  was used as internal reference control. Concentration of culture solutions were: 50  $\mu$ M CBE, 40  $\mu$ M GL1, 20  $\mu$ M Lyso GL1. **B.** CLN8 expression at baseline in mild and severe GD patient fibroblasts compared to control cells by real time PCR. Results are expressed as times control. The cycle threshold for actin was 19.8 and the average control cycle threshold was 27.8. Each patient's risk alleles are displayed on top for the 2 SNPs (rs11986414 and rs4875960) associated with CLN8. Allele C for the first SNP and allele A for second SNP are associated with a higher risk of disease severity.

**Table 1**  
**Demographics of Study Population of N370S Homozygous Gaucher disease**

Severity	Age	SSI	DS3	Liver Size	Spleen	% Female	% Splenectomy
<b>Mild</b>	47.7±1.9 (n=80)	4.2±0.2 (n=80)	1.7±0.1 (n=80)	1.1±0.0 (n=80)	4.9±0.3 (n=80)	55%	0%
<b>Moderate</b>	51.8±3.3 (n=35)	7.4±0.3 (n=35)	4.3±0.1 (n=35)	1.3±0.1 (n=35)	7.7±1.1 (n=32)	55%	9.6%
<b>Severe</b>	57.5±3.5 (n=24)	12.1±0.4 (n=24)	7.0±0.3 (n=24)	1.6±0.2 (n=24)	13.5±2.4 (n=17)	50%	29.2%
<b>Total</b>	50.4±1.5 (n=139)	6.4±0.3 (n=139)	3.2±0.2 (n=139)	1.2±0.0 (n=139)	6.8±0.5 (n=129)	55%	7.2%