Bioinformatics Studies of Genes that Cause Lysosomal Storage Diseases. A Comparison of *GLA*, *GBA* and *NPC1*

Olga Shaidakova and Farshad Tamari, Ph.D. Kingsborough Community College, Brooklyn, NY, USA.

Introduction

Lysosomal storage diseases fall under the umbrella of Inborn Errors of Metabolism, comprised of some 750 metabolic disorders (Saudubray and Garcia-Cazorla, 2018) and result from irregularities in metabolic pathways (Ferreira and van Karnebeek, 2019). They typically represent serious genetic diseases that usually result from genetic mutations from a single gene, and therefore, follow single-locus Mendelian inheritance. Examples of Lysosomal Storage Diseases include Niemann-Pick (NP) types A, B, and C1, as well as Fabry and Gaucher disease. While the diseases are caused by a mutation in a single gene, typically multiple mutations are mapped to the same gene locus. For example, for Niemann-Pick Type C1 disease, more than 180 mutations have been mapped to the NPC1 gene locus (Tamari et al., 2013c).

Niemann -pick type C1 and C2 result from mutations in the *sphingomyelin phosphodiesterase 1* (SMPD1) gene, while mutations for Fabry and Gaucher disease affect the *galactosidase alpha* (GLA), and *glucosylceramidase beta 1* (GBA1), respectively. As all three diseases fall under the category of Lysosomal Storage Diseases, it is plausible, and we hypothesize that the genes share some DNA identity.

Hypotheses and Goals

Goal: To characterize and compare sequences of three genes, GBA, GLA, and NPC-1 using bioinformatics software Jalview and Unipro UGENE

Hypothesis: Since mutations in the three genes cause Lysosomal Storage Diseases, there may exist DNA similarity between the three genes

Materials and Methods

To compare the DNA sequence for potential similarities, the DNA for the three genes were obtained from National Center for Biotechnology Information (NCBI). An input file, in FASTA format, bearing the sequences of the three genes was created after the removal of numbers and spaces from each of the three sequences. Removal of numbers and spaces between nucleotides in the sequences was completed by using: https://www.bioinformatics.org/sms2/filter_dna.html. For DNA alignment, the multiple alignment tool (CrustalW) function of the bioinformatics software Jalview was used with default settings. To characterize each sequence in terms of nucleotide composition, the bioinformatics software Unipro-UGENE was used.

Results

A sequence comparison among the genes indicated sequence similarity (Figure 1) between GBA and GLA of 61 %, GBA and NPC1 at 22% and GLA and NPC1 at 11%. (Table 1). Due to the size difference between the genes, with NPC1 being composed of 4760 nucleotides, GLA of 1260, and GBA of 2291 nucleotides, it is not surprising that the sequences do not align at the beginning and at the end of the genes (Figure 1, top left corner). There was a varying degree of similarity between the genes observed throughout the middle portions (Figure 1), with no alignment observed towards the end of the genes for the same reason as above. Surprisingly, the % nucleotide content when the three genes are compared indicated a % composition that was within a relatively narrow range (22-29% of each nucleotide, Table 2).

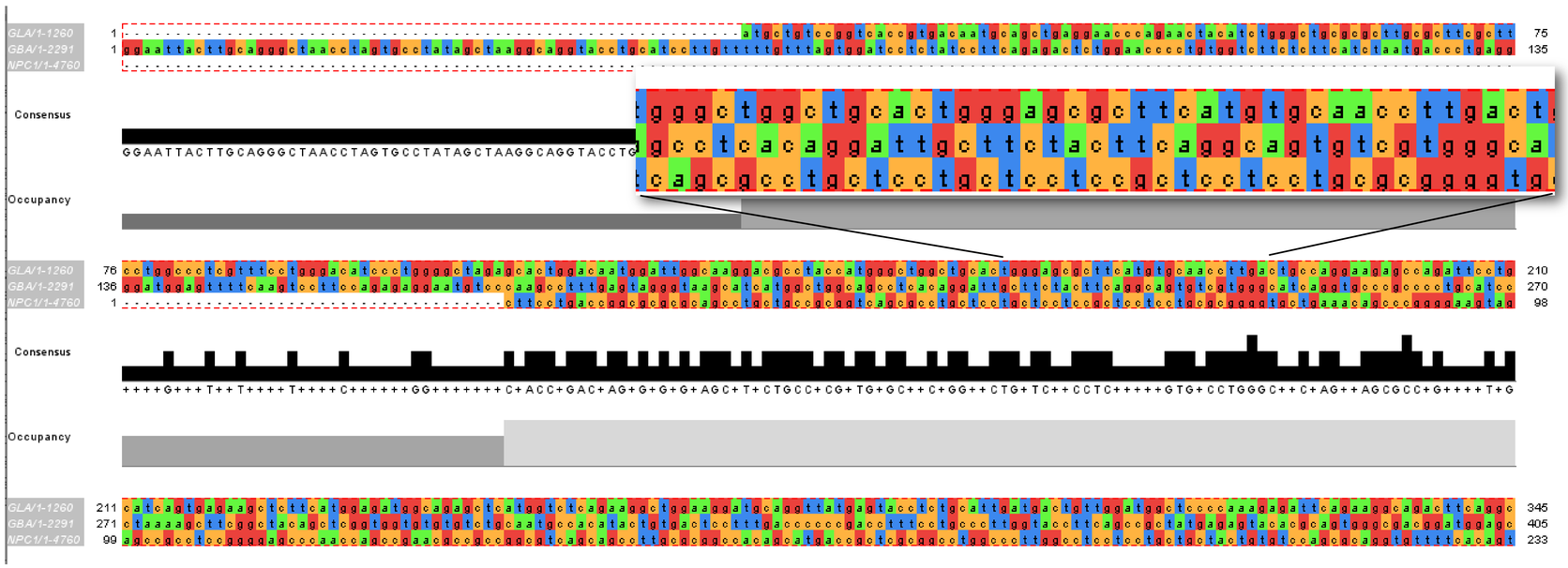

Figure 1: Alignment and comparisons of the three genes using Jalview

Discussion and Future Work

We set out to determine potential differences for three genes, mutations in which cause recessive genetic diseases. Because all three diseases fall under the umbrella of Lysosomal Storages Diseases, we hypothesized that there may be sequence similarity between the genes. The genes have significantly different sizes, so it is not surprising that the % similarity is relatively low. In addition, although all three genes contribute to a Lysosomal Storage Disease, they code for different proteins with different functions. For example, while NPC1 gene is responsible for the egress of unesterified cholesterol from the lysosomes, GLA is involved in the breakdown of globotriaosylceramide (GL) within the lysosomes and GBA helps breakdown glucocerebroside within the lysosomes. Given the similarity of the function between GLA and GBA, perhaps it is not surprising that they share the biggest sequence identity (61%, Table 1) among all the comparisons.

References

- Saudubray, Jean-Marie, and Ángels Garcia-Cazorla. "Inborn Errors of Metabolism Overview: Pathophysiology, Manifestations, Evaluation, and Management." Pediatric clinics of North America vol. 65,2 (2018): 179-208. doi:10.1016/j.pcl.2017.11.002
- Ferreira, Carlos R, and Clara D M van Karnebeek. "Inborn errors of metabolism." Handbook of clinical neurology vol. 162 (2019): 449-481. doi:10.1016/B978-0-444-64029-1.00022-9
- Tamari F, Chen FW, Li C, Chaudhari J, Ioannou YA (2013) PKC Activation in Niemann Pick C1 Cells Restores Subcellular Cholesterol Transport. PLoS ONE 8(8): e74169. https://doi.org/10.1371/journal.pone.0074169

This research was made possible through funds made available to FT by CUNY's "CUNY Research Scholars Program"

	GLA	GBA	NPC1
GLA	100%	61%	11%
GBA	61%	100%	22%
NPC1	11%	22%	100%

Table 1: % identity comparing GLA, GBA and NPC1

	GLA		GBA		NPC1	
N	*#N*	*%N*	*#N*	*%N*	*#N*	*%N*
A:	316	25	526	23	1 118	24
C:	293	23	653	29	1 179	25
G:	332	26	606	27	1 160	24
T:	319	25	506	22	1 303	27

Table 2: Nucleotide composition comparison of GLA, GBA and NPC1