Pathogenesis of Lactose Intolerance: Expression and Mutation of LCT Gene

Xiaomi Ai[1, †], Aiyi Wang[2, †, *]

[1]RDFZ Xishan School, Beijing, China, axm_2005@163.com

[2]Beijing Huijia Private School, Beijing, China

*Corresponding author: 23wangaiyi@huijia.edu.cn

[†]These authors contributed equally

Keywords: Lactose intolerance, *LCT gene*, *MCM6 gene*, Pathogenesis.

Abstract: Lactose intolerance affects about two-thirds of the world's adult population. The evidence from the study of metabolic defects shows that the symptoms of specific inherited diseases reflected by some genes cannot dictate the synthesis of certain proteins or enzymes. It can be inferred that one of the types of lactose intolerance, congenital lactase deficiency (CLD), is closely related to the genetic disorder. *Lactase* gene (*LCT* gene) is the gene that directly associates with the pathology of CLD, its mutation will induce the expression of lactose intolerance. In addition to *LCT* gene, another gene that is near with it in the same region called *minichromosome maintenance complex component 6* gene (*MCM6* gene) is also responsible for the initiation of lactose intolerance because it is the enhancer of *LCT* gene, which controls whether the mutant genotype will express. Furthermore, the distinct magnitudes of single-nucleotide polymorphism of these two genes can trigger different phenotypes in the human body. Focus on one of the steps of the gene expression, transcription, whether the variation on the mRNA of *LCT* gene can influence the transcriptional gradient is also a noticeable factor of the pathogenesis. All these gene-related causes show that treating with genetic pathogenesis, like gene editing, is the critical breakthrough of healing lactose intolerance.

1. Introduction

Lactose intolerance is a digestive condition caused by a lack of ability to digest lactose, the primary carbohydrate found in dairy products. Severe lactose intolerance can cause a variety of symptoms, including diarrhea, vomiting, flatulence, and stomach discomfort, all of which can cause patients a variety of problems on a regular basis. Current research findings suggest that most of the symptoms of lactose intolerance are due to congenital lactase deficiency, which is a genetic cause [1]. It's also a problem that affects people all across the world, particularly Asians. Figure 1 shows that the rate in northern Europe ranges from 2% to 15%, while in Central Europe it ranges from 9% to 23% and in the United States it ranges from 6% to 22%. Lactose intolerance can be as high as 95% to 100% in Asia. At a minimum, it affects about two-thirds of the world's adult population [2].

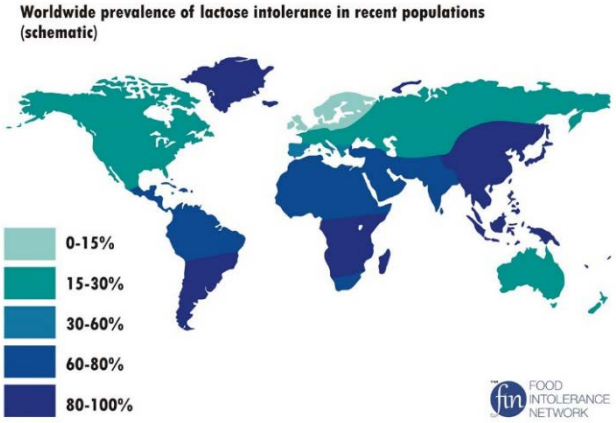

Figure 1. Worldwide prevalence of lactose intolerance in recent populations [3]

Lactose is the carbohydrate found in milk and dairy, which is composed of glucose and galactose, cells can also convert them into the storage carbohydrate glycogen, which provides a supply of sugar, serve as a source of energy for the human body as table sugar and starch do. Lactose, for example, is a disaccharide that cannot be absorbed by the small intestine and must be digested into monosaccharides. To accomplish this, the small intestine produces lactase, also known as lactase-phlorizin hydrolase (LPH), a type of enzyme found in the small intestine of mammals that catalyzes the breakdown of the -glycosidic bond in lactose. As a result, these small molecules are readily absorbed by the intestine and enter the bloodstream, where they serve as energy for the human body. In humans, lactase is particularly abundant during infancy, lactase non-persistence (LNP) are known as after weaning phase, people who can continue to digest lactose during adulthood, which capable of maintaining high levels of lactase content were called lactase persistence (LP). Why lactose intolerance can be classified as a genetic problem? Because lactase was encoded by a kind of gene called *lactase* gene (*LCT* gene) located on chromosome 2 at position 21 on the *LCT* gene region, including the adjacent *LCT* and *minichromosome maintenance complex component 6 (MCM6)* genes. Lactose intolerance was caused by insufficient lactase production that can't hydrolyze big molecules to digestible form, mutations of *LCT* gene will lead to interference in the function of the lactase enzyme. On the upstream approximately 39 kb 5′ of *LCT* gene, *MCM6* was the switch that modulates *LCT* gene's expression, *MCM6* gene's five single nucleotide polymorphisms are the main causes of LP [4].

This review aims to summarize how do the current *LCT* and *MCM6* genes affect lactose intolerance, discussing the contribution of current genetic findings to lactose intolerance as a disease, and describing genetic testing methods for early prevention [5].

2. Relationship between LCT gene and lactose intolerance

The arrangement of nitrogenous bases on the strand of the DNA is decisive for the encoding of genes. When the lactase is needed for digesting lactose in small intestine, the region that contains the *LCT* gene (*LCT* gene region) will split and come through transcription. The transcription will produce a strand of RNA that contains complementary bases with *LCT* gene. The RNA strand that is created at this process is called the messenger RNA (mRNA) of the *LCT* gene. The mRNA will be passed on to a specific organelle called ribosome, the place of translation, and the transfer RNA (tRNA) matches nucleotides on mRNA in a pattern of three nucleotides per group encoding one amino acid. Then, a polypeptide chain will be formed during the series of amino acids built with the sequence of nitrogenous bases. This polypeptide chain provided by the transcription and translation of *LCT* gene is the basis for the formation of lactase. Finally, one or more linking polypeptides will constitute the lactase [6].

LCT gene encodes for the generation of lactase, which controls whether the infant will appear the symptoms of lactose intolerance. This inborn lactose intolerance is called congenital lactase deficiency (CLD). It is a rare autosomal recessive inheritance, which means the pathogenic individual's parents both have one copy of mutated gene. Its main phenomenon is infants' inability to digest lactose, whether by breast milk or formula. After ingesting the food that contains lactose, the infants will develop severe watery diarrhea. *LCT* gene contains 17 exons which have a length of 13352 bp. It not only determines the lactase production and persistence of infants but also adults. Lactose intolerance that happened in adulthood is generally caused by the decreasing expression of *LCT* gene after infancy. It is also the most common cause of lactose intolerance. The regulatory element in *LCT* gene region is a DNA sequence that keeps the *LCT* gene's activity. This key substance sites in another gene that is nearby *LCT* gene, called *MCM6* gene. The changes in this element will maintain the production of lactase in the human's small intestine. Only if people inherit these significant changes, they can avoid getting lactose intolerance [7].

3. Various mutants of the LCT gene cause CLD

The mutations of *LCT* gene will occur when the number or order of bases in the gene is disrupted. Nucleotides in the DNA molecule of *LCT* gene can be deleted, doubled, rearranged, or replaced, each alteration having a particular effect. The abnormally short enzyme and the change of a single protein building block (amino acid) in lactase are caused by *LCT* gene mutations. These mutations will interfere with the function of lactase, resulting in the non digestion of lactose in the small intestine.

The studies show that there are five kinds of mutation in *LCT* gene, which will cause CLD. The previous studies present that the mutation that the sequence analyses of regional transcripts in *LCT* gene do not discover the mutation that caused CLD (figure 2). So, the former researchers genotype some microsatellite markers for Finish 32 children affected by CLD, which cover chromosome 2, position 2q21 (the position of *LCT* gene) to 2q22 and construct some haplotypes in chromosomes. According to the sequence analysis of *LCT* gene, one major haplotype, cen-5-T−13910-7-5-12-9-6-11-4-15-12-9-7-5-4-4-tel, is emerge disease chromosomes and indicates that the founder haplotype is derived. Three of five distinct mutations cause the premature truncation of lactase, the other two are missense mutations with Single Nucleotide Variation (SNV). Chromosomes with nonsense mutations all have the major pathogenic haplotype. The first mutation is c.4170T→A (Finmajor). Finmajor is observed in 27 of 32 patients. The form that they preserve the founder haplotype is homozygous. Besides, the codon of this mutation is prematurely terminated in exon 9, which causes Y1390X. The second one is the disappearance of nucleotides in exon 14, c.4998_5001delTGAG. This mutation happens in two patients' paternal disease chromosomes, not only brings a frameshift (S1666fsX1722) and a premature stop codon and leads to the production of heterozygous for Finmajor and S1666fsX1722. The third mutation revealed in one patient is the deletion of two nucleotides in exon 2, c.653_654delCT, which codon 218 and protein truncation at codon 224, S218fsX224 will be frameshift changed, and this mutation is also carried in the paternal disease chromosome. The two Single-nucleotide polymorphisms (SNPs), *rs121908937* and *rs386833833*, are the likely pathogenic or pathogenic variants in CLD. They are also related to the last two mutations discovered in the patients: The c.804G→C transversion indicates the amino acid substitution of SNPs, *rs121908937*, changes from glutamine (Q) to histidine (H) at codon 268 in exon 3 (Q268H); the c.4087G→A transition leading the changing of SNPs, rs386833833, from glycine (G) to serine (S) at codon 1363 in exon 9 (G1363S). The variant in *rs121908936* and *rs121908937* are very crucial determinants for the initiation of CLD [8].

SNPs have multiple genotypes: because *rs121908936* has two alleles A and T, there are three different genotypes (A; A), (A; T), and (T; T) [9]. Each genotype has a specific magnitude: *rs121908936* (A; A) is the normal genotype with max magnitude, 5; *rs121908936* (A; T) is heterozygous with magnitude 2.5; *rs121908936* (T; T) has magnitude 0. Similarly, *rs121908937* has two alleles C and G, and three genotypes: (C; C) with max magnitude 5, (C; G) with magnitude 2.5, (G; G) with magnitude 0 [10]. The changing of magnitude in SNPs decides the expression of *LCT* gene. Both in *rs121908936* and *rs121908937*, when the magnitude is 5, it will cause the CLD; when the magnitude is 2.5, the carrier of CLD alleles will appear; magnitude=0 is the most common condition that arises in ClinVar. Basically, the mutation of LCT gene is diverse, they all can cause the CLD in different ways.

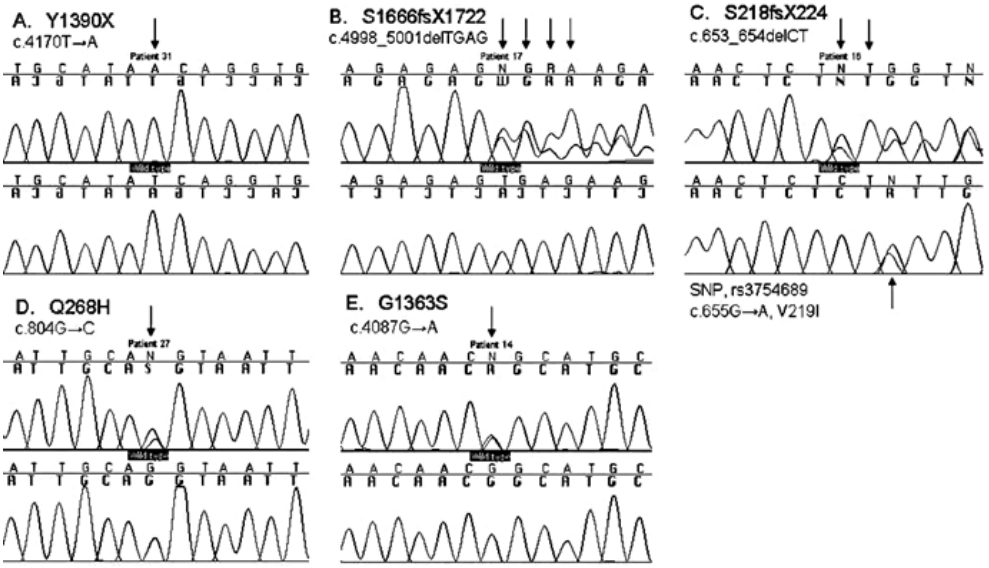

Figure 2. Five distinct CLD mutation in LCT gene [8]

4. MCM6 gene mutation underlies the genetic etiology of LP

About two-thirds of people around the world will have LNP, left one-third of people will have LP. The reasons that cause LP were five or more SNPs in a region called *MCM6* by a kind of protein, on the same haplotype and in an intron of it, one of the highly conserved micro chromosome maintenance proteins (MCM) necessary to initiate eukaryotic genome replication is protein. So *MCM6* gene is being described as a regulatory enhancer that modulates *LCT* expression. These two genes are located close together, the transcription start site of the MCM6 gene is located at 5′ of the *LCT* transcription start site, approximately 39 kb, and it has been scanned by fluorescence in situ hybridization in chromosome 2, position 2q21 (figure 3), showing that it is being expressed in a variety of adult and fetal human tissues.

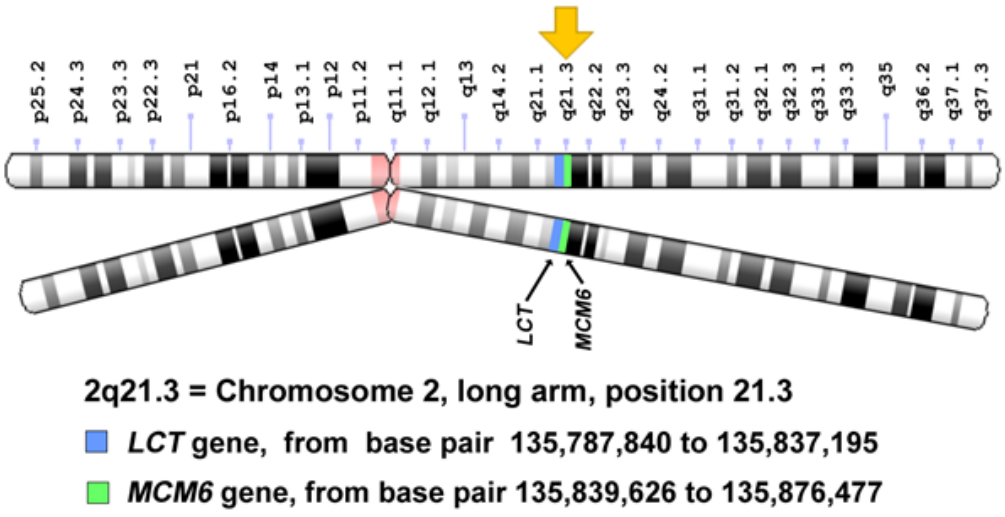

Figure 3. The location of MCM6 gene on LCT gene [11]

In an epigenetic map analysis related to the in vivo transcription gradient of the mouse lactase gene, it was found that a large cluster of modified cytosine was associated with LCT mRNA variation in the whole intestine at *MCM6* exon 13 – intron 13. The researchers looked at the coding data of adult mice's small intestine to see if sites associated with *LCT* transcriptional variation could act as genomic regulatory elements. They discovered a stable element at *MCM6* exon 13 – intron 13 (h3kme1) and an active enhancer at *LCT* intron 2 (h3k4me1 and h3k27ac), as well as an active promoter at *LCT* exon 1

(h3k4me1 and h3k27 (h3k4me3 and h3k27ac). The binding of CTCF, a protein that enhances the connection between transcriptional regulatory regions by changing chromatin structure, can be shown in *MCM6* exon 13 – intron 13, *LCT* exon 1 and intron 2. *MCM6* exon 13 – intron 13 locus was found to be a key regulator for the age-dependent establishment and maintenance of *LCT* transcriptional gradient. These findings are consistent with reports based on genetics and evidence of its ability to regulate *LCT* transcription, that is, this upstream region is an enhancer of *LCT* transcription. The new observation found that the epigenetic control of *MCM6* exon 13 – intron 13 is important for *LCT* regulation in both mice and humans, earlier findings by adding an epigenetic dimension.

Furthermore, the *MCM6* exon 13 – intron 13 region is not sensitive to environmental signals (i.e. milk), implying that the *MCM6* exon 13–intron 13 epigenetic programming is not scalable. Other regulatory factors that control the *LCT* gradient, such as *LCT* intron 2 and intron 8, on the other hand, can adjust epigenetically to environmental inputs and partially restore *LCT* expression in the adult gut after weaning. Therefore, local *DNA* modification changes may promote intestinal *LCT* transcription gradient with age [12].

It has been described the genetic basis of population diversity in lactase production as a dominant trait. Although it isn't finished yet. As we mentioned in the last paragraph: The reasons that cause LP were five or more SNPs in *MCM6* gene's region, these types of SNPs are *rs4988235*, *rs41525747*, *rs41380347*, *rs869051967* and *rs145946811*, in addition to the most extensive and closely related LP variants, these five SNPs have also been reported as functional markers, including in vitro transfection tests and in vivo studies. Except these five genetic markers, so far, in a specific population, 18 new SNPs mapping *MCM6* gene are also associated with LP. Therefore, a total of 23 known SNPs is considered as the basis of genetic etiology of LP.

A Finnish team described a recent finding that two polymorphisms were found in introns 9 and 13 of *MCM6*. The -13910C/ T intron 13 is 14 Kb upstream of the *LCT* gene when it was numbered from the initiation codon – ATG – from the *LCT* gene. The other polymorphism -22018G/A in intron 9, is located 22 Kb upstream of the *LCT* gene, 100% and 97% correlated to the lactose intolerance respectively. Homozygous genotypes CC and GG have low lactase levels, which is not persistent to lactase, so it will cause lactose intolerance. Homozygous genotypes TT and AA have high lactase levels, they are persistent to lactose, therefore no lactose intolerance being triggered. While heterozygous genotypes CT and GA have medium lactase levels, both of which are lactose tolerant. All the T allele exists as lactase persistence, which is in people who are tolerant with lactose and is missing in individuals without persistence, which is the people who had lactose intolerance [13].

Therefore, lactase persistence or non-persistence is related to the non-coding variation in *MCM6* gene, because the single nucleotide polymorphism in the intron region of this gene is associated with differential transcriptional activation of the adjacent lactase gene promoter [13].

5. Feasibility of LCT gene and MCM6 gene detection in diagnosing lactose intolerance

Lactose intolerance is widespread in some countries and sometimes it is fatal, like CLD, which will greatly affect life. The routine test for lactose intolerance is a breath test. The hydrogen/methane breath test (HBT) measures hydrogen in the breath to detect gastrointestinal symptoms in a variety of illnesses. When there is a problem with digestion or absorption of food in the small intestine, large volumes of hydrogen gas might be created. Because lactase was not enough to hydrolyze lactose and make it used by the human body, the non-digested lactose will go into colon and ferment by colonic microbiota, which will turn to gas and short-chain fatty acid to cause diarrhea and fart. The hydrogen generated by bacteria is then taken into the bloodstream as it passes through the small intestine and colon. The hydrogen-rich blood travels to the lungs, where it is released and exhaled in the breath, allowing it to be measured. [14].

However, HBT is a test that can only be done if the patient is consuming lactose. As mentioned above, the relationship between *LCT* gene and *MCM6* gene and lactose intolerance is direct, whether it is clinically possible for some patients with a genetic family history or repeated diarrhea to seek the cause, and actively carry out gene identification to make a clear diagnosis. And for the detection of

newborns, from birth, a positive and reasonable arrangement of diet can be carried out to promote the growth and development of newborns. In addition to physically avoiding foods with high lactose content, carrying out genetic testing can also in advance prevent it.

The -13910C/ T and -22018G/A mutations are the very crucial factors that control the expression of *LCT* gene. In order to prove these variations affect the absorption of lactose, researchers study on twenty Brazilian volunteers above 18 years old, half of whom present lactose tolerance and the other half present lactose intolerant. At the beginning of the experiment, the researchers phlebotomize the participants for DNA analysis. The DNA fragment which contains the mutation -13910C/ T are be shown by the primers C/T-for 5'-AAGACGTAA GTTACCATTTAATAC-3' and C/T-rev 5'-CGTTAATACCCACTGACCTATCCT-3' through polymerase chain reaction (PCR). The variation -22018G/A also be found by PCR with primers G/A-for 5'-TAAGA ACATTTTACACTCTTC-3' and G/A-rev 5'-AGAAAATGGGTTTTCGCCATG-3'. This DNA test proposes that the CC/GG genotype indicates the LNP, but the CT/TT and GA/AA genotypes present the LP for individuals. Then, the researchers use HBT to determine again whether the participants have the symptoms of LNP. All 10 samples (CT/GA, CT/AA, and TT/AA genotypes) have negative HBT. The proportion of their detailed genotype is 6 people with CTGA, 3 people with TTAA, and 1 person with CTAA. Other 10 people with the symptom of LNP after the test, which contains genotype CC/GG, 9 of them have positive HBT result and 1 has a negative result. The specific one of the ten CCGG genotype individuals with negative HBT results without any symptoms can be explained by the general decrease of lactase or the intestinal flora not generating hydrogen. Overall, the results of the testing procedure show an association between mutations in the *MCM6* gene fragment and LP: mutations in -13910C/ T and -22018G/A corresponded to the negative result of HBT, which mean individuals are LP; the absence of mutations in the gene also corresponded to positive HBT results, meaning that participants are LNP [15].

6. Conclusion

Lactose intolerance is a widespread and ethnically unique disease. The study of its genetic pathology can help people understand and prevent the disease in their lives through early genetic testing. Current studies have shown that lactose intolerance that happened in adulthood is generally caused by the decreasing expression of *LCT* gene after infancy. The regulatory element in *LCT* gene region is *MCM6* gene that keeps the *LCT* gene's activity. The few SNPs (rs4988235, rs41525747, rs41380347, rs869051967, and rs145946811) mutation on *MCM6* gene are related to the lactose persistence of people, homozygous genotypes CC and GG will cause lactose intolerance, homozygous genotypes TT and AA can persist with lactose, and heterozygous genotypes CT and GA have medium lactase levels, both of which are lactose tolerant. And according to the pathogenesis of lactose intolerance, people can test whether they are digestible to lactose by hydrogen/ methane breath test or gene sequencing. What additional SNPs are associated with lactose intolerance disorders? Do their mutations have a greater impact on the ability to digest lactose in humans? More importantly, of course, with the development of modern medical technology, can we target the invention and study of treatments based on our understanding of the genetic pathology of lactose intolerance? In the future, personalized gene therapy or gene-editing technology may be a potential treatment for lactose intolerance.

References

[1] Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Orr, R. B. (2021). Campbell Biology (12th ed.). Pearson Education Limited.

[2] Lactose Intolerance by Country 2022. (n.d.). Worldpopulationreview.com. Retrieved March 12, 2022, from https://worldpopulationreview.com

[3] FIN - Food Intolerance Network. (n.d.). Www.food-Intolerance-Network.com. Retrieved March 13, 2022, from https://www.food-intolerance-network.com

[4] MCM6 minichromosome maintenance complex component 6 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Www.ncbi.nlm.nih.gov.
[5] Anguita-Ruiz, A., Aguilera, C. M., & Gil, Á. (2020). Genetics of Lactose Intolerance: An Updated Review and Online Interactive World Maps of Phenotype and Genotype Frequencies. Nutrients, 12 (9). https://doi.org/10.3390/nu12092689
[6] Lactose intolerance: MedlinePlus Genetics. (2020, August 18). Medlineplus.gov. https://medlineplus.gov/genetics/condition/lactose-intolerance/
[7] The Editors of Encyclopedia Britannica. (2019). gene | Definition, Structure, Expression, & Facts. In Encyclopædia Britannica. https://www.britannica.com/science/gene
[8] Kuokkanen, M., Kokkonen, J., Enattah, N. S., Ylisaukko-oja, T., Komu, H., Varilo, T., Peltonen, L., Savilahti, E., & Järvelä, I. (2006). Mutations in the Translated Region of the Lactase Gene (LCT) Underlie Congenital Lactase Deficiency. American Journal of Human Genetics, 78 (2), 339–344. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1380240/
[9] rs121908936 - SNPedia. (n.d.). Www.snpedia.com. https://www.snpedia.com
[10] rs121908937 - SNPedia. (n.d.). Www.snpedia.com. Retrieved February 25, 2022, from https://www.snpedia.com/index.php/Rs121908937
[11] Shiels, R. (n.d.). LibGuides: Enzyme Deficient Diseases: Lactose Intolerance. Libguides.msben.nsw.edu.au. https://libguides.msben.nsw.edu.au/enzymedeficientdisease/lactose
[12] Oh, E., Jeremian, R., Oh, G., Groot, D., Susic, M., Lee, K., Foy, K., Laird, P. W., Petronis, A., & Labrie, V. (2017). Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome. Scientific Reports, 7 (1). https://doi.org/10.1038/srep41843
[13] A.S. Arroyo, M., P. Lopes, A. C., B. Piatto, V., & Victor Maniglia, J. (2010). Perspectives for Early Genetic Screening of Lactose Intolerance: - 13910C/T Polymorphism Tracking in the MCM6 Gene. The Open Biology Journal, 3 (1), 66–71. https://doi.org/10.2174/1874196701003010006 6
[14] Hydrogen/Methane Breath Test. (n.d.). National Jewish Health. https://www.nationaljewish.org
[15] Bulhões, A. C., Goldani, H. A. S., Oliveira, F. S., Matte, U. S., Mazzuca, R. B., & Silveira, T. R. (2007). Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (LCT) gene in adult-type hypolactasia. Brazilian Journal of Medical and Biological Research, 40 (11), 1441–1446.