Gene expression regulation and the lactase gene

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size (Bases)</th>
<th>Estimated Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human (<em>Homo sapiens</em>)</td>
<td>3 billion</td>
<td>30,000</td>
</tr>
<tr>
<td>Laboratory mouse (<em>M. musculus</em>)</td>
<td>2.6 billion</td>
<td>30,000</td>
</tr>
<tr>
<td>Mustard weed (<em>A. thaliana</em>)</td>
<td>100 million</td>
<td>25,000</td>
</tr>
<tr>
<td>Roundworm (<em>C. elegans</em>)</td>
<td>97 million</td>
<td>19,000</td>
</tr>
<tr>
<td>Fruit fly (<em>D. melanogaster</em>)</td>
<td>137 million</td>
<td>13,000</td>
</tr>
<tr>
<td>Yeast (<em>S. cerevisiae</em>)</td>
<td>12.1 million</td>
<td>6,000</td>
</tr>
<tr>
<td>Bacterium (<em>E. coli</em>)</td>
<td>4.6 million</td>
<td>3,200</td>
</tr>
<tr>
<td>Human immunodeficiency virus (HIV)</td>
<td>9700</td>
<td>9</td>
</tr>
</tbody>
</table>

→ > 40% of the predicted human proteins are similar to fruit-fly or worm proteins
### Packing the information

1. DNA consists of two complementary nucleotid strains
2. DNA is coiled around proteins called histones building the chromatin
3. the human genome consists of $3.2 \times 10^9$ nucleotids which are arranged along chromosomes
4. DNA double strands are packed into chromosomes which are highly condensed

![Figure 1. Chromatin packing.](image)

### Human chromosomes

1. The human genome consists of 23 chromosome pairs.
2. Each cell contains two copies of a chromosome, one inherited by the mother and one by the father.
3. Those two copies are called homologous chromosomes
4. Sex chromosomes are not homologous.
5. Chromosome 1 (the largest human chromosome) has the most genes (2968), and the Y chromosome has the fewest (231).

![Figure 2. Banding pattern of the human chromosomes.](image)
The human genome

- functions unknown for > than 50% of discovered genes

- human genome sequence is almost (99.9%) exactly the same in all people

- ~ 2% of the genome encodes instructions for the synthesis of proteins

- Repeat sequences that do not code for proteins make up at least 50% of the human genome

- Repeat sequences have no direct functions, but they are creating entirely new genes or modifying and reshuffling existing genes.

The human genome

- Over 40% of the predicted human proteins share similarity with fruit-fly or worm proteins

- particular gene sequences are associated with numerous disorders (breast cancer, muscle disease, deafness, blindness…)

- ~ 3 million locations where single-base DNA differences occur in humans identified → search for DNA sequences associated with common diseases (cardiovascular disease, diabetes, arthritis, cancer…)

Gene expression regulation

Both of these cells contain the same genome, but they express different RNAs and proteins.

Figure 3. A mammalian neuron and a lymphocyte. Molecular Biology of the cell, 4th Edition

DNA – RNA - Protein

Figure 4. The pathway from DNA to protein. The flow of genetic information from DNA to RNA (transcription) and from RNA to protein (translation) occurs in all living cells. Molecular Biology of the cell, 4th Edition
Regulation of gene expression

Figure 5. Six steps at which eucaryotic gene expression can be controlled. Molecular Biology of the cell, 4th Edition

The 6 steps of gene expression regulation

- 1. transcriptional control
- 2. RNA processing control
- 3. RNA transport and localization control
- 4. translational control
- 5. mRNA degradation control
- 6. protein activity control
Transcriptional control

... controlling when and how often a given gene is transcribed

Figure 6. Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

Molecular Biology of the cell, 4th Edition

Transcriptional control – regulation by RNA polymerase

- Repressors
- Activators
- Enhancers
- General transcription factors

Figure 7. Transcription initiation by RNA polymerase II in a eucaryotic cell.

Molecular Biology of the cell, 4th Edition
Transcriptional control – gene regulatory proteins

Figure 8. Summary of the mechanisms by which specific gene regulatory proteins control gene transcription in procaryotes. (A) Negative regulation; (B) positive regulation.
Molecular Biology of the cell, 4th Edition

Transcriptional control – Epigenetic Modifications

"Epi" – above, over, outside or beside

a) Methylation
b) Histone modifications (Biotinylation, Poly(ADP-ribosylation))
c) X-chromosome inactivation
d) Genomic Imprinting
DNA methylation

is the addition or removal of a methyl group predominantly where cytosine bases occur consecutively

Figure 9. www.mcb.ucdavis.edu

Methylation of cytosines is a heritable modification of DNA!

Histone modifications

… modifications at the amino acids that constitute the N-terminal tails of histones

Figure 11. https://www.univie.ac.at/chromosomes/studentenpage/skripten/Jantsch/AMG%20Jantsch%202.pdf

→ Histones are small proteins that mediate the folding of DNA into chromatin
→ DNA is wrapped around octamers of core histones
X-chromosome inactivation

Sex is determined by the X and Y – chromosome. To balance the unequal X-chromosome dosage between the XX female and XY male, mammals have adopted a unique form of dosage compensation:

The X-chromosome inactivation (one of the two X chromosomes is transcriptionally silenced through epigenetic mechanisms)

The silencing involves only those genes that are on the same X chromosome. The inactive state of those X chromosomes is maintained during cell divisions.

Genomic imprinting

Two copies of every autosomal gene are inherited. Both copies are functional for the majority of these genes.

Imprinted genes are those genes in which one copy is turned off in a parent-of-origin dependent manner.

Examples:
- Paternally expressed imprinted genes tend to promote growth while it is suppressed by those genes which are maternally expressed.
- Paternally expressed imprinted genes enhance the extraction of nutrients from the mother during pregnancy, whereas the maternal genome limits it.
Consequences of epigenetic modifications

- **DNA methylation** is generally associated with silencing of genes, directly by inhibiting the binding of specific transcription factors and indirectly by recruiting methyl-CpG binding proteins and their associated repressive chromatin-remodelling activities.
- **Histone modifications** leads to in-/activation of genes
- **X-chromosome inactivation** balances the unequal X-chromosome dosage between the XX female and XY male.
- **Genomic imprinting** inactivates one copy of the father or the mother -> the gene is inherited just by one parent.

The 6 steps of gene expression regulation

- 1. transcriptional control
- 2. RNA processing control
- 3. RNA transport and localization control
- 4. translational control
- 5. mRNA degradation control
- 6. protein activity control
RNA-processing control

… controlling how the RNA transcript is spliced or otherwise processed

- Capping
- Addition of poly(A)tail
- Splicing

Figure 12. Alternative splicing of the α-tropomyosin gene from rat. Molecular Biology of the cell, 4th Edition

RNA-processing control

Capping, Splicing, Polyadenylation

Figure 13. Summary of the steps leading from gene to protein in eucaryotes and bacteria. Molecular Biology of the cell, 4th Edition
The 6 steps of gene expression regulation

• 1. transcriptional control
• 2. RNA processing control
• 3. RNA transport and localization control
• 4. translational control
• 5. mRNA degradation control
• 6. protein activity control

RNA transport and localization control

… selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized

Figure 14. Schematic illustration of an "export-ready" mRNA molecule and its transport through the nuclear pore. As indicated, some proteins travel with the mRNA as it moves through the pore, whereas others remain in the nucleus. Once in the cytoplasm, the mRNA continues to shed previously bound proteins and acquire new ones; these substitutions affect the subsequent translation of the message. Because some are transported with the RNA, the proteins that become bound to an mRNA in the nucleus can influence its subsequent stability and translation in the cytosol. RNA export factors, shown in the nucleus, play an active role in transporting the mRNA to the cytosol.

Molecular Biology of the cell, 4th Edition
RNA transport and localization control

- The main advantages of this mechanism over transcriptional control in the nucleus is that it endows local sites with independent decision-making authority.

- Localized RNAs typically contain codes, expressed within cis-acting elements, that specify subcellular targeting. Such codes are recognized by trans-acting factors, adaptors that mediate translocation along cytoskeletal elements by molecular motors.

- Most transported mRNAs are assumed translationally dormant while en route.

- In some cell types (e.g. in neurons) it is considered that translation remains repressed after arrival at the destination site (e.g., a postsynaptic microdomain) until an appropriate activation signal is received.

The 6 steps of gene expression regulation

- 1. transcriptional control
- 2. RNA processing control
- 3. RNA transport and localization control
- 4. translational control
- 5. mRNA degradation control
- 6. protein activity control
Translational control

... selecting which mRNAs in the cytoplasm are translated by ribosomes

Mechanisms of translation control:
- transport control (only the mRNA which is transported to the cytoplasm can be translated)
- number of ribosomes
- translation factors (if the concentration of these factors is too low in the cell, the translation start or the elongation process can be decelerated or inhibited)
- mRNA localisation (a specific place in the cytoplasm leads to the production of the protein at a specific position in the cell)
- Regulation by untranslated regions (UTRs)

The 6 steps of gene expression regulation

- 1. transcriptional control
- 2. RNA processing control
- 3. RNA transport and localization control
- 4. translational control
- 5. mRNA degradation control
- 6. protein activity control
mRNA degradation control

...selectively destabilizing certain mRNA molecules in the cytoplasm

- the stability of different mRNAs in the cytoplasm varies widely.
- many eukaryotic mRNAs are quite stable, some have unusually short half-lives.
- the stability is determined by the cap-structure and the length of the poly-A tail of the mRNA.
- mRNA degradation is carried out by ribonucleases (deadenylation, degradation of the poly-A tail).
- mRNA stability is also dependent on base pair structure of the transcript.

New investigated field:
Regulation of mRNA stability by miRNA and siRNA which are building ribonucleoprotein complexes with recognize the mRNA in a sequent specific way -> blocking the translation complex or degradation of the transcript

Figure 15. Two mechanisms of eukaryotic mRNA decay.
(A) Deadenylation-dependent decay. Most eukaryotic mRNAs are degraded by this pathway. The critical threshold of poly-A tail length that induces decay may correspond to the loss of the poly-A binding proteins. The deadenylation enzyme associates with both the 3' poly-A tail and the 5' cap, and this arrangement may coordinate decapping with poly-A shortening. Although 5' to 3' and 3' to 5' degradation are shown on separate RNA molecules, these two processes can occur together on the same molecule. (B) Deadenylation-independent decay. It is not yet known with certainty whether decapping follows endonucleolytic cleavage of the mRNA.

Molecular Biology of the cell, 4th Edition
The 6 steps of gene expression regulation

- 1. transcriptional control
- 2. RNA processing control
- 3. RNA transport and localization control
- 4. translational control
- 5. mRNA degradation control
- 6. protein activity control

Protein activity control

… selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made.

Proteins built after translation can be
- functional or
- have to undergo a maturation process (exo/-endopeptidasen)
- or functional groups (phosphorylation, acetylation, methylation … ) have to be added
Example – regulation of transcription

In bacteria, the lac repressor protein blocks the synthesis of enzymes that digest lactose when there is no lactose to feed on. When lactose is present, it binds to the repressor, causing it to detach from the DNA strand.

Lactase gene - E. coli

Figure 16. (A) A diagram of a small portion of the genome of the bacterium Escherichia coli containing genes (called lacI, lacZ, lacY, and lacA) coding for four different proteins.

Molecular Biology of the cell, 4th Edition
Lactase gene – E. coli

The lac operon consists of 3 structural genes (promoter, terminator and operator):

- \textit{lacZ} encodes β–galactosidase (cleaves lactose into glucose and galactose)
- \textit{lacY} encodes β–galactoside permease, a membrane bound transport protein that pumps lactose into the cell.
- \textit{lacA} encodes β–galactoside transacetylase, an enzyme that transfers an acetyl group from acetyl-CoA to β–galactosides.

Lactase:
The lac operon (E. coli)

Figure 17. Dual control of the lac operon.
Molecular Biology of the cell, 4th Edition

Glucose and lactose levels control the initiation of transcription of the lac operon

- through their effects on the lac repressor protein and CAP
- Lactose addition increases the concentration of allolactose, which binds to the repressor protein and removes it from the DNA.
- Glucose addition decreases the concentration of cyclic AMP; because cyclic AMP no longer binds to CAP, this gene activator protein dissociates from the DNA, turning off the operon.
- (in reality the situation is more complex. In addition, expression of the lac operon is never completely shut down.)
Lac operon (E. coli)

http://www.hschickor.de/genregu1.html

http://www.youtube.com/watch?v=EsWhrJ51Y6U

Lactase gene (human)

The small intestinal enzyme lactase or lactase-phlorizin hydrolase (LPH) is encoded by the gene LCT

Figure 18. http://genecards.ccbb.re.kr
**Lactase gene**

The gene for human LPH:

- is located on chromosome 2q21
- comprises 17 exons and covers approximately 49 kb
- has a messenger RNA (mRNA) of slightly more than 6 kb
- from initiation codon to stop codon, human LPH mRNA encodes 1927 amino acids forming the complete translation product

---

**Lactase or lactase-phlorizin hydrolase**

![Figure 19. Lactase; PDB - An Information Portal to Biological Macromolecular Structures](http://www.pdb.org/pdb/explore.do?structureId=1TG7)
Biochemistry of lactase

The lactase gene is transcribed and translated to make a glycosylated enzyme precursor, which is then transported to the cell membrane and cleaved to produce the active enzyme. The enzyme dangles into the small intestine, bound to the membranes of cells of the intestinal wall.

Figure 20.

Lactase-phlorizin hydrolase (LPH) -1

Fig. 21. Structure and biosynthesis of the lactase protein. Overview of the domain structure of the LPH protein. The translation product of the lactase gene is produced as a pre-pro-LPH form. A signal peptide is cleaved during translocation revealing the pro-LPH form. The pro-region is later cleaved off during processing revealing the mature functional LPH protein. The region necessary for apical sorting and dimerization is indicated. Sequence analysis of the lactase sequence has revealed four internal repeats as indicated. Two of the repeats (I–II) are located in the pro-region, whereas two repeats (III– IV) are found in the mature LPH. Modified; J.T. Troelsen / Biochimica et Biophysica Acta 201723 (2005) 19–32.
Lactase expression

In mammals milk is the main nutrient consumed after birth.

- essential for newborn mammals to be able to hydrolyse the large amounts of lactose found in milk

- high lactase expression continues until the weaning period at which time mammals change their carbohydrate intake from milk to an intake mainly based on plants /and or meat
Regulation of lactase expression

The lactase gene has a complicated pattern of regulation:

– the cellular level (differentiation-dependent expression in the enterocytes)
– the organ level (tissue-specific and differential expression along the longitudinal axis of the small intestine)
– During development (postweaning decline)

Regulation of lactase expression

Lactase expression (the cellular level):

• … is a commonly used marker for differentiated enterocytes
• … differs between rodents and humans around birth:
  – mouse: lactase expression rises from a relatively low level just after birth to a maximum level 3 days after birth (rodent intestine is not fully mature at birth)
  – human: are born with a high expression
Regulation of lactase expression

Lactase expression (the organ level):

• … is exclusively restricted to the small intestine of all investigated mammals
• ... is highest in the mid-jejunum with a lower level in the proximal jejunum and the duodenum (in adult rat and rabbit)

A minor lactase mRNA expression can be detected in the colon during the postnatal period.

Regulation of lactase expression

Lactase expression (during development):

• … is down-regulated during the weaning period (postweaning decline) leading to a low level of lactase expression thereafter

postweaning decline:
- rats: at day 20 after gestation
- human: lactase production usually drops about 90% during the first four years of life (varies widely), but there are human populations which tolerate fresh milk and other dairy products throughout their lives (lactase persistence)
Transcriptional regulation of lactase expression

The presence or absence of lactase is regulated at the level of gene transcription.

- "promoter regions" of nearby DNA which control whether or not they are transcribed into RNA to make proteins.
- promoters permit the transcription of the associated gene when they bind to specific "transcription factors".
- promoters of intestinal genes, are activated when they bind to specific transcription agents that are expressed in the cells of the intestinal epithelium.

Lactase promoter

A 150 bp region of the lactase promoter upstream of the transcriptional initiation site is well-conserved in rat, mouse, rabbit pig and human indicating that important regulatory cis-elements are located in this region.

-> Transgenic analyses of the lactase promoter (transgenic mice)
-> Analyses of the lactase promoter in intestinal cell lines (Caco-2 cells)
Lactase promotor

![Sequence alignment of the conserved approx. 150 bp lactase promoter. The cis-elements in this region are boxed. The transcription factors binding to the cis-elements are indicated in parentheses. J.T. Troelsen / Biochimica et Biophysica Acta 2017 23 (2005) 19–32](image)

Transcription factors regulating lactase expression - I

3 cis elements have been identified in transfection experiments:

- CE1a
- CE2c
- GATA –site

Cdx-2 and HNF1α, and GATA-4 (investigated in Caco-2 cells.)
Transcription factors regulating lactase expression -II

Figure 24. Schematic comparison of the lactase expression and the expression of transcription factors known to bind to the proximal lactase promoter. Expression pattern of lactase [85], Cdx-2, HNF-1α, GATA-4 and Pdx-1 along the longitudinal axis of the mouse small intestine. J.T. Troelsen / Biochimica et Biophysica Acta 20 1723 (2005) 19–32

Transcription factors regulating lactase expression -III

Figure 25. Upstream gene regulatory regions influencing pig, rat, and human lactase promoter activity. Regulatory cis-elements are indicated. The identity of transcription factors binding the cis-elements is indicated ("?" indicates unknown factors). The identified regulatory roles of the cis-elements found by functional promoter analyses are indicated by arrows as well as their gene regulatory role (activation=+, repression=-). J.T. Troelsen / Biochimica et Biophysica Acta 20 1723 (2005) 19–32
Lactase expression – summary - I

• Lactase expression in humans and animals is regulated in a complex pattern in the small intestine

• A 150bp proximal promoter is important for the regulation of lactase expression

• This promoter region binds transcription factors (Cdx-2, HNF-α and GATA factors)

• These transcription factors are very important for not only lactase expression but also for the differentiation and maintenance of the intestinal epithelium

Lactase expression – summary - II

• The activity of proximal lactase promoter is modulated by transcriptional enhancers and silencers located upstream the lactase gene

• No sequence homology has been found between these regulatory regions in the human, pig and rat
Functional role of SNPs

Some polymorphisms have been identified, which may have a functional role in regulating lactase gene activity:

… are located in introns of MCM6 (a gene involved in cell cycle regulation) – 3kb upstream of the initiation site of the human lactase gene

… possible functional role of SNPs associated with adult-type hypolactasia (lactase persistence)

2 most important polymorphisms so far:

• T/C – 13910
• A/G - 22018
Adult-type hypolactasia

- Autosomal recessive gastrointestinal condition that is the result of a decline in the activity of lactase in the intestinal lumen after weaning
- Down-regulation of lactase is considered as a normal phenomenon among mammals
- However, some humans have the ability to maintain lactase activity and digest lactose throughout their lives (lactase persistence – maybe influence of polymorphisms in the MCM6 gene)