Genetic regulation in eukaryotes

0. Introduction

Protists, Fungi, plants and animals belong to Eukaryotic organisms. Prokaryotic gene regulation will be discussed at the “lac operon” seminar. The definition of genetic regulation: in a strict sense: regulation of transcription; in a wider sense: regulation of expression and operation of gene products (RNAs and proteins).

Slide 0 Analysis of gene expression Gene expression has become a key issue in biology during the past couple of years. There are several reasons for this. Our curiosity to understand the processes controlling the formation of an adult human from a single cell; controlling the operation of adult body; and resulting in the amazing variability observable in human population and in other animals. Understanding genetic regulation is basically important in future medicine: in diagnostics and cure of illness, and in general, in personalized medicine. The technological development has enabled us by now to accurately analyze both the expression of individual genes and simultaneously a huge number of genes (functional genomics*: DNA chips* and protein chips*).

Slides 1-3 The importance of genetic regulation Slide 1 Why is it important to understand the principles of gene expression? As in most fields of science, two motivations drive scientific research: the joy of discovery or application of scientific results in practice. According to the newest scientific results gene expression and not gene structure determines phenotype. We have almost identical genomes to that of chimpanzee, but some important genes are expressed differently. Slide 2 Understanding genetic regulation will be essential in near future medicine. Nowadays, DNA and protein chips are widely used in diagnostics. Gene therapy will be the only remedy for various diseases. Information of gene expression will also be important in individual health care. Slide 3 The importance of genetic regulation is indicated by the awarded Nobel Prices in these disciplines. Jacob and Monod awarded the Nobel Price for the discovery of prokaryotic regulation at 1965. 2006 Chemistry: Roger Kornberg, studies of eukaryotic gene regulation; 2006 Medicine: Andrew Fire, Craig Mello: discovery of RNA interference.

Slides 4 Regulation from genes to proteins The expression of genes and gene products (RNAs and proteins) is regulated at several levels, which are as follows:

Transcription
- mRNA processing
- mRNA transport
- mRNA localization
- mRNA stability

\{ post-transcriptional regulation \}

Translation
- Protein degradation
- Protein processing and modification

\{ post-translational regulation \}

RNA polimerase-II transcribes pre-mRNAs. Introns are spliced out from the pre-mRNAs. For the sake of simplicity, only a single intron is depicted, but normally 20-50 introns occur within a gene. Together, only exons but not introns will be the part of mature mRNAs. The mRNAs are protected from nuclease attack by a cap at their 5’-terminals and by a stretch of adenine-containing nucleotides at their 3’-end. The export of RNAs from the nucleus, the translation, the modification of proteins is also regulated.
1. Regulation of transcription

Slide 5 Regulation of transcription The transcription of genes are regulated at two main levels. (1) At the chromatin level gene expression is controlled by histones. The density of histones bound to the DNA determines the strength of transcription. (2) cis and trans elements represent the other level of gene expression regulation.

1. Regulation of chromatin* regulation
   - Methylation
   - Acetylation

2. Interactions between regulatory elements
   - cis-elements: promoters, enhancers and silencers
   - trans-elements: RNA polymerase, transcription factors and co-factors

Cis-elements are localized on the same DNA strand as the gene
Trans-elements are located in the cytoplasm

Slide 6 Regulation of chromatin Regulation of histone–DNA binding allows the establishment of different chromatin states leading to distinct ‘readouts’ of the genetic information, such as gene activation or gene silencing. The acetylation of histone proteins removes positive charges, thereby reducing the affinity between histones and DNA, which makes it easier for RNA polymerase and transcription factors to find access to regulatory sequences and to activate transcription from them. Methylation of histones and DNA sequences leads to the opposite effect than that of acetylation. Methylation of histones results in a stronger binding to DNA. Methylation of DNA sequences has been shown to result in the silencing of gene expression in certain cells.

Slide 7 Structure of a gene A gene is composed of transcribed and regulatory regions. The transcribed region consists of non-coding (5'-UTR and 3'-UTR; untranslated regions) and coding region. Only those parts of the exons, which encode amino acids are called as coding regions. The parts of regulatory regions are the promoters, enhancers (facilitates transcription) and silencers (inhibit transcription). Introns are approximately twenty times longer than exons on average, while non-coding exons (5'-UTR and 3'-UTR) have the same size as the coding ones.

Slides 8, 9 Promoters Transcriptional regulation is by far the most important mode for the control of eukaryotic gene expression. The cis-regulatory sequences involve regulatory DNA motives, which are recognized by specific transcription factors. Basal promoter elements termed TATA boxes are located between 20 and 30 bases upstream of the transcriptional start site of eukaryotic genes. Proximal promoter elements, such as the CAAT box and GC box, reside within 40 to 250 bases upstream of the transcriptional start site. The various “boxes” are so called consensus sequences meaning that they do not or only slightly vary even across large evolutionary distances.

Slide 10 Enhancers and silencers The enhancer/silencer regulatory sequences are predominantly located upstream (5') of the genes, though some elements may occur downstream (3') or within the introns. Enhancer sequences can reside up to hundreds of thousands of base pairs from the coding region. The regulatory sequences of an average gene reside within 10,000 base pairs. The number and type of regulatory elements vary with each gene. The borders of genes are determined by the insulator sequences, whose function is to restrict the effect of regulatory sequences to the gene they control.
**Slide 11** Cell type-specific gene expression  Although, almost all of our cells comprise the same genetic content, there is a huge number of cell types, and each type of cells expresses different genes. The question is how it is possible. The various cell types developed by means of differentiation. The genetic basis of differentiation is the formation of different chromatin pattern (varying histone binding to the DNA) in different tissues. The histone binding pattern determines the type of transcription factors expressed in a cell, and the transcription factors decide whether the expression of a particular gene is ON or OFF. Normally, several transcription factors and co-factors are required for the control of a gene. If one of them is missing, there is no transcription. Further, even though, the appropriate transcription factors are present in a cell, no gene expression is induced if the regulatory region of the target gene is blocked by histones.

**4. Types of gene expression**

**Slide 12** Characteristics of gene expression  Three very important features of gene expression is under regulation: (1) when a gene product (RNA or protein) is produced during embryogenesis or after birth; (2) how much gene product is produced; (3) which tissues (where) it is expressed in.

**Slide 13** Types of gene expression

1. Constitutive (continuous) – e.g. housekeeping genes
2. Induced 
   - nutrition material-induced: glucose in liver cell
   - stress: heat shock, osmotic shock (salt)
   - cell-communication-induced: hormones, growths factors
   - developmentally regulated
3. cell type-specific

**Slide 14** Cell communication-induced gene expression  Cells more often communicate with each other by means of signal molecules. A signal molecule can bind to cell-membrane anchored a receptor inducing a strictly regulated cascade of biochemical events (this is far the most frequent situation), called signal transduction. Alternatively, signal molecules can enter the cell and exert their effects in the cytoplasm or in the nucleus. There are three basic types of them. They can be transcription factors, thus they directly influence gene expression by binding an enhancer sequence on he DNA (it is rare), or they can bind to a transcription factor, or to another factor, which exert its effect on a transcription factor through multiple steps. The intracellular binding partner of a signal molecule is also called receptor.

**Slide 15** Steroid hormone activation  Glucocorticoid receptors are located in the cytoplasm in an inactive state (hsp90 chaperons inhibit them). Steroid hormone binding dislocates hsp90, and results in the formation of a dimeric (two-subunit) molecule, which in turn, enter to the nucleus, bind to its response DNA element (GRE: glucocorticoid response element) and activate transcription from the linked gene.

**Slide 16** Interferon-γ activation  Interferon (IFN)-γ binding to its receptor induces JAK kinase activation, which in turn results in the phosphorylation of STAT-1α transcription factor. As a result, activated STAT-1α forms a dimeric molecule, enter to nucleus and induce transcription from genes harboring the appropriate response elements (alternative terms: recognition sequences, motives, consensus sequences).
Glossary

**Chromatin** is the complex of DNA and protein found inside the nuclei of eukaryotic cells. The major proteins involved in chromatin are histone proteins, although many other chromosomal proteins have prominent roles too. The functions of chromatin are to package DNA into a smaller volume to fit in the cell, to strengthen the DNA to allow mitosis and meiosis, and to serve as a mechanism to control expression. Changes in chromatin structure are affected mainly by methylation (DNA and proteins) and acetylation (proteins). Chromatin is easily visualized by staining, hence its name, which literally means *colored material*. Simplistically, there are three levels of chromatin organization: 1. DNA wrapping around nucleosomes - The "beads on a string" structure. 2. A 30 nm condensed chromatin fiber consisting of nucleosome arrays in their most compact form. 3. Higher level DNA packaging into the metaphase chromosome.

**DNA chips**: Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously.

**Epigenetic inheritance**: the same genetic content can determine more than one phenotype as a result of, for example, maternal effects.

**Exon shuffling**: gaining novel domains of proteins by acquiring a new exon from another gene located at other part of the genome during evolution.

**Forward genetics**: The experimental procedure that begins with a random mutation and a subsequent search for the altered phenotype and the mutant gene responsible for this phenotype.

**Functional genomics** uses high-throughput techniques like DNA microarrays* and proteomics* to describe the function and interactions of genes. These techniques allow the analysis of the expression level of a huge number of genes at the same time.

**Gene expression** is a multi-step process that begins with transcription, followed by post transcriptional modification and translation.

**Gene networks** are genetic modules composed of functionally-linked genes, which determine the development and operation of particular traits, physiological processes or behaviors.

The **genome** of an organism is the whole hereditary information of an organism that is encoded in the DNA (or, for some viruses, RNA). This includes both the genes and the non-coding sequences.

**Genomics** is the study of an organism's/species’ genome.

A **knockout animal** is a genetically engineered animal (usually mouse) one or more of whose genes have been made inoperable. Knockout is a route to learning about a gene that has been sequenced (revealing the order of bases) but has an unknown or incompletely known function.

**Phenotype**: anything that is part of the observable structure, function or behavior of an organism

**Protein chips** (= protein microarrays) are measurement devices used in biomedical applications to determine the presence and/or amount (referred to as quantitation) of proteins in biological samples.

**Proteome** The entirety of proteins in existence in an organism. Most importantly, while the genome* is a rather constant entity, the proteome differs from cell to cell and is constantly changing through its biochemical interactions with the genome and the environment.

**Proteomics** is the large-scale study of proteins (simultaneous analysis of a large number of proteins). This term was coined to make an analogy with genomics*, and while it is often viewed as the "next step", proteomics is much more complicated than genomics.
**Regulatory RNAs** A commonly used synonym is non-coding RNA, RNA molecules that function without being translated into proteins. Their functions include regulation of gene expression at the levels of transcription (chromatin modification) and translation.

**Reverse genetics** The experimental approach that begins with a cloned segment of DNA, followed by the insertion of this DNA to the host genome. The foreign DNA can serve both a transgene, which is over-expressed in the host animal, or it can serve to knock out an endogenous gene of the host. The aim of reverse genetics to find altered phenotype resulted by the genetic manipulation.

**Ribonuclease (RNase)** is an enzyme that catalyzes the breakdown of RNA into smaller components.

**Signal transduction**: is any process by which a cell converts one kind of signal or stimulus into another. Processes referred to as signal transduction often involves a sequence of biochemical reactions inside the cell, which are carried out by enzymes and other proteins linked through second messengers.

**Silent codon positions**. The genetic code is redundant, which means that in most cases more than one codon determines single amino acids. Those base replacements, which do not result in the change of amino acids are called as silent changes. Those positions of codons (generally third ones), which contain the replaceable bases are called silent codon positions.

**Small interfering RNA (siRNA)**, are a class of 20-25 nucleotide-long RNA molecules that interfere with the expression of genes. They are naturally produced as part of the RNA interference (RNAi) pathway by the enzyme Dicer. They can also be exogenously (artificially) introduced by investigators to bring about downregulation of a particular gene. siRNA's have a well defined structure. Briefly, this is a short (usually 21-nucleotide) double-strand of RNA (dsRNA) with 2-nucleotide overhangs on either end, including a 5’ phosphate group and a 3’ hydroxy (-OH) group.

**Spliceosome** is a complex of RNA and many protein subunits that remove the non-coding introns from unprocessed mRNA. The RNAs that spliceosomes consist of are named U1, U2, U4, U5, and U6, and participate in several RNA-RNA and RNA-protein interactions.

**Transcription factor**: a protein that binds DNA at a specific promoter and enhancer or other transcription factors, and thereby directly controls transcription. Transcription factors can be selectively activated or deactivated by other proteins, often as the final step in signal transduction.

**Transgenic organism**: An organism that has integrated foreign DNA into its germ line as a result of the experimental introduction of DNA. Recombinant DNA techniques are commonly used to produce a transgenic organism.

**Ubiquitin** is a 76 residue polypeptide that can be conjugated to specific proteins by members of a complex family of enzyme cascade systems, whereby signaling that the particular protein is destined for degradation.