5/5/03 Introduction to eukaryotic gene regulation

Differential gene expression in eukaryotes: some interesting phenomenology related to this topic

1) How do you know your head from your tail (the HOX gene complex)

Control of spatial (region and tissue specific expression) expression of genes as well as temporal expression

![Diagram of HOM-C gene expression](image)

![Diagram of Hox gene expression](image)
2) Why cloning vertebrates (using donor nuclei from adults)

doesn’t work very well -- the human cloning announced in 2001 involved the production of a handful of pathetic clumps of cells

Many scientists considered their cloning claim to be B--- S---

And Scientists are in general agreement that the Clonaid claims are bogus

http://www.clonaid.com/
3) Why we’re not chimps

It is well established that at least 98% of human DNA (coding and non-coding is identical to that of chimpanzees.

Which of our genes makes us human?

IN other words what genes have we acquired in the last ? years that the common ancestor of chimps and human did not have?
Maybe is this not the correct question?

Differential gene expression of our almost identical genomes may be the reason that you are a human and not a chimp

**PRIMATE EVOLUTION:**

**Gene Activity Clocks Brain's Fast Evolution**

Elizabeth Pennisi

A team of molecular biologists has taken a stab at defining what makes us human. Its answer: We’re set apart from other primates not so much by differences in the makeup of our genes but by relatively recent changes in how active those genes are. Such changes are most dramatic in the brain, where they’ve occurred at a faster rate in humans than in other primates, they report on page 340.

*Science 296: 233  April 12, 2002*
Intra- and Interspecific Variation in Primate Gene Expression Patterns

Wolfgang Enard,¹* Philipp Khaitovich,¹* Joachim Klose,² Sebastian Zöllner,¹ Florian Heissig,¹ Patrick Giavalisco,³ Kay Nieselt-Struwe,⁴ Elaine Muchmore,⁵,⁶ Ajit Varki,⁵ Rivka Ravid,⁷ Gaby M. Doxiadis,⁸ Ronald E. Bontrop,⁸ Svante Pääbo¹†

Although humans and their closest evolutionary relatives, the chimpanzees, are 98.7% identical in their genomic DNA sequences, they differ in many morphological, behavioral, and cognitive aspects. The underlying genetic basis of many of these differences may be altered gene expression. We have compared the transcriptome in blood leukocytes, liver, and brain of humans, chimpanzees, orangutans, and macaques using microarrays, as well as protein expression patterns of humans and chimpanzees using two-dimensional gel electrophoresis. We also studied three mouse species that are approximately as related to humans.

Science 296: 340  April 12, 2002

We will look at the approaches to assessing variation in gene expression at the whole genome level at the end of the quarter: microarrays and 2-D gel electrophoresis of proteins
General Themes seen in both eukaryotic and prokaryotic cells

Transcriptional regulatory circuitry in euk cells has the same basic components as prok cells

Cis-acting sites:
• gene specific -- often (but not necessarily) adjacent to the core promoter region
• known various names such as operators in proks and enhancers & silencers in euks
• these sites are inert until bound by their cognate regulatory proteins

Regulatory proteins
• trans-acting DNA-binding proteins that recognize and bind to specific cis-acting sites: call repressors, activators, negative regulators, positive regulators, transcription factors

A mechanism for regulating the activity of the regulatory protein
BE SURE TO LOOK AT FIGURE 7-58 IN ALBERTS
• level and/or activity of regulator is modulated in response to some signal
• signal could come from the external environmental, the internal environmental (ie. hormones in multi-cellular organisms), intracellular signals (developmental regulatory proteins)

Also involves
• core promoter sequences (not gene specific)
• generic TFs in eukaryotes (not gene specific) and sigma factor in proks
• RNA polymerase complex
Transcription Factors: DNA binding proteins that recognize specific promoter elements or other cis-acting sequences

These transcription factors have a variety of DNA-binding motifs
- A search of the human genome sequence has revealed more than 2000 genes that encode transcription factors which are involved in the regulation of gene expression
- In this figure they are categorized according their protein “family”
- Assignment to a family is based on specific protein motifs that relate to how the protein contacts DNA or other aspects of their function

Figure 1 Genome-wide comparison of transcriptional activator families in eukaryotes. The relative sizes of transcriptional activator families among Homo sapiens, D. melanogaster, C. elegans and S. cerevisiae are indicated, derived from an analysis of eukaryotic proteomes using the INTERPRO database, which incorporates Pfam, PRINTS and Prosite. The transcription factors families shown are the largest of their category out of the 1,502 human protein families listed by the IPI.
Nature 409:832 Feb. 15, 2001 Expressing the Human Genome
The homeodomain of the Engrailed protein binds to a particular site in the DNA. Helix 3 contacts the base pairs in the major groove, while the amino-terminal portion of the homeodomain enters the minor groove. (After Pabo and Sauer, 1992.)

alberts figure 7-16  7-23
How is a gene catalogued as having a zinc finger binding motif?
overhead of conserved residues in Zinc Finger

Conserved bases confer a structural framework but do not confer binding specificity. Be sure to look at Figure 7-28 in Alberts
• DNA binding domain
• Transactivation domains -- role?
• Hormone binding? role?
• dimerization

see also zinc finger figures in Chapter 7 of alberts
CIS-ACTING SITES IN EUKARYOTIC PROMOTERS

• CORE PROMOTER (includes TATA box)

• PROMOTER-PROXIMAL ELEMENTS: found within 100-200 bases of the transcription initiation site

• DISTANCE-INDEPENDENT SITES: cis-acting elements that can exert their effects at considerable distance either upstream or downstream from the promoter (includes enhancers and silencers)

DISTANCE-INDEPENDENT SITES:
• enhancers or silencers (in higher eucaryotes)
• upstream activator sequences (UAS) in yeast

♦ gene-specific (and in multicellular eukaryotes may control tissue specificity of expression -- see below)
♦ can be located 5' of, 3' of, or within an intron of a gene. The orientation of these sequences relative to the transcription start site is relatively unimportant.
♦ Furthermore, such sites may be located at quite a distance (thousands of base pairs) away from the gene that they are regulating.
♦ Enhancers generally contain multiple clustered binding sites for transcription factors, which interact cooperatively

hormone response elements (HRE’s): cis-acting sites that bind to steroid hormone receptors
How can a cis-acting site affect RNA polymerase at a promoter 1000’s of bases distant from the site?

Core RNA polymerases do not recognize promoters on their own
- TFII A, B, D, E, F & H are basal or general transcription factors used at every eukaryotic promoter
- The action of these factors is required to assist pol II in recognizing promoter sites and initiation transcription

The general transcription factors are essentially the same for each gene transcribed by pol II

The regulatory transcription factors and their cis-acting sites will vary from gene to gene

What feature of this promoter complex are not characteristic of prokaryotic cells?
**Activators**
These proteins bind to genes at sites known as *enhancers* and speed the rate of transcription.

**Repressors**
These proteins bind to selected sets of genes at sites known as *silencers* and thus slow transcription.

**Coactivators**
These "adapter" molecules integrate signals from activators and perhaps repressors.

**Basal transcription factors**
In response to injunctions from activators, these factors position RNA polymerase at the start of transcription and initiate the transcription process.
Tissue specific expression of a gene in multicellular eukaryotes is often controlled by enhancer elements

A cis-acting site is effectively inert until it is contacted by its cognate transcription factor: availability of active transcription factors controls the *temporal* and *spatial* expression of many genes

Each enhancer and any other type of cis-acting site will contact a specific transcription factor(s) -- analogous to the specificity of interaction between a prokaryotic operator and repressor

E= enhancer  P = promoter  TF = transcription factor
The activity or availability of a transcription factor can be controlled in many different ways:

- transcriptional control
- splicing
- translational control
- post-translational control (allostery, phosphorylation, dimer or heterodimer formation, sequestration in a cellular compartment)

Look at the alberts figure 7-58
But what controls the regulators of the regulators and so forth and so on back to the beginning?

Facing up to eukaryotic organisms:

What does molecular biology tell us about this process?

The final outcome: variety of tissues producing a variety of specific gene products located in the proper orientation to each other.

Gene Expression in Eukaryotes
Gene --------------------------> Active protein product

1. Rate of transcription initiation **
2. Alternative splicing patterns
3. Transcript stability
4. Translational regulation
5. Post-translational processing or activation

Focus on #1  but some additional complications

↓ cis-acting sites  ↓ trans acting proteins

RNA pol II complex with promoter:
rate of formation

↑ CpG methylation       ↑ chromatin structure
(complex euks)

♦ Normal development depends on a precise sequence of changes in the
  configuration of the chromatin and the methylation state of the genomic DNA

♦ These so-called epigenetic alterations are involved in tissue-specific expression
  of genes
Targeted Chromatin Accessibility by Remodeling Complexes: Making genes accessible to transcription factors

- Eukaryotic genes are packaged into chromatin, which greatly impedes the binding of many proteins to their target sequences
- A fundamental mechanism controlling the selectivity of gene expression is the limited ability of many transcription factors to access the genome

Accessibility of DNA to protein interaction is regulated by diverse enzymatic complexes that modulate nucleosomal structure by “remodelling” or histone modification
Transcriptional state of eukaryotic chromatin. The ground state is generally restrictive, and the enhancer (ENH) and promoter (TATA) elements are covered by nucleosomes. Repressors (R) can bind to the chromatin (either to the DNA or nucleosome) and bind a set of proteins that silence the DNA. Alternatively, the DNA can bind activators (A) that bind to other chromatin modifying proteins. These allow TFIID and RNA polymerase II to replace nucleosomes and initiate transcription.
In vertebrates:

Methylation of cytosines provides a gene silencing mechanism

![Methylation of cytosines](image)

Figure 7–80. Molecular Biology of the Cell, 4th Edition.

*Formation of 5-methyl cytosine*

In vertebrates this is confined to selected cytosines located in a CG sequence

This methylation pattern can be stably inherited:

![Methylation pattern](image)

Figure 7–81. Molecular Biology of the Cell, 4th Edition.

Methyl-directed maintenance methyltransferases maintain the methylation pattern
DNA methylation may help turn genes off:

1. *de novo* methyltransferases may take advantage of a “bare” promoter to methylate CG regions that are not protected by TF’s or the RNA pol II complex
2. the methylation may stimulate binding by other proteins which shut down transcription permanently by altering the chromatin configuration
Meet (left to right): Rainbow, Allie and cc (carbon copy) who is a genetic clone of Rainbow. Allie is cc’s surrogate mom.

Animal cloning is inefficient
- It took 188 unsuccessful attempts at producing a cloned kitten before CC was “born”
- cloning results in gestational and neonatal developmental failures
- at BEST, a few percent of the nuclear transfer embryos survive to birth and, of those, many die within the perinatal period
- newborn clones often display respiratory distress and circulatory problems and even apparently healthy survivors may suffer from immune dysfunction, or kidney or brain malfunction

The fetal abnormalities and abnormalities in those few clones that are born alive probably result from failures in genomic reprogramming.
Adult sheep

Skin fibroblast cells

Egg chromosomes removed

Electrically induced egg and attached skin cell

Skin cell nucleus inside egg

Embryo culture

Embryo transfer to foster mother
Gene/genomic Reprogramming:

- Normal development depends on a precise sequence of changes in the pattern of gene expression: transcription and translation in adult cells is very different from that in embryonic cells.
- Some genes are expressed in only embryonic cells; other genes are expressed only in adult cells.
- Epigenetic programming: during development some genes are more or less “permanently” turned off: they are chemically modified (methylation of purine/pyrimidine bases) and complexed with specific proteins that prevent access by the transcription machinery.
- These epigenetic modifications control tissue-specific expression of genes.
- During normal reproduction, genomic reprogramming takes place during oogenesis and spermatogenesis (formation of the egg and sperm), processes that take years and months (respectively) in humans.
- During nuclear cloning, the reprogramming of the somatic nucleus must occur within minutes or, at most, hours between the time that nuclear transfer is completed and the onset of cell division in the activated egg.

---

Figure 1. Reprogramming in normal development and nuclear cloning.

(A) The genome of primordial germ cells is hypomethylated ("reset," white boxes). Reprogramming and establishment of parent-specific epigenetic marks occur over the course of gametogenesis so that the genome of sperm and egg is competent to express the genes that need to be activated in early embryonic (red hatched box) and later (green hatched box) development. During cleavage and early postimplantation development, "embryonic" genes, such as Oct 3/4, become activated (solid red box) and are repressed at later stages (black boxes) when tissue-specific genes (green boxes) are activated in adult tissues (labeled A, B, and C).

Adult stem cells are thought to be less differentiated and may be more effective NT donors because they may require less reprogramming (see text). Epigenetic reprogramming of imprinted and nonimprinted genes occurs during gametogenesis in contrast to X inactivation and the readjustment of telomere length, which take place postzygotically.

(B) Reprogramming of a somatic nucleus after nuclear transfer may result in
(i) no activation of "embryonic" genes and early lethality,
(ii) faulty activation of embryonic genes and an abnormal phenotype, or (iii) faithful activation of "embryonic" and "adult" genes and normal development of the clone. The latter outcome is the exception, and the percentage in each category is estimated from data on cumulus cell NT animals.
In normal reproduction, the sperm and egg genomes are competent to express the genes that need to be activated in early development.

Black = genes turned off  red = early embryonic genes  green = tissue specific genes  

PGC = primordial germ cell