Transcription Regulation and Gene Expression in Eukaryotes  FS 08

Gene organization/localization

AIRE: a special (?) example of gene regulation

DNA Methylation

P. Matthias, May 7, 2008

Some cellular structures: interdependency of compartments?
Gene gating: A hypothesis

Gene gating: it’s all about location

ABSTRACT It is assumed that the genome of a higher eukaryotic organism is organized into a number of distinct three-dimensional (3-D) structures, each characteristic for a given differentiated state. These discrete 3-D structures are envisioned to develop in a hierarchical and largely irreversible manner from an omnipotent 3-D structure of the zygotic genome. The information for these processes is assumed to reside in the genome. The nuclear pore complexes, the peripheral nuclear lamina, and components of the nuclear core are proposed to be among the topologically most proximal organelles that interpret this information and thereby serve in the maintenance and the alteration of the 3-D structure of the genome during development, differentiation, and the cell cycle. The nuclear pore complexes are envisioned to serve as gating organelles capable of interacting specifically with expanded (transcribable) portions of the genome. Their nonrandom distribution on the nuclear surface would reflect the underlying periodic organization of the genome into expanded and compacted domains, alternating with each other. All transcripts of a given gated gene would leave the nucleus by way of that pore complex to which the gene is gated. Implications for cell asymmetry and polarity are discussed and evolutionary considerations are presented.
Chromosome territories and various nuclear bodies

Interchromosomal interactions
The nuclear periphery: lamins and others

**Lamins**

**A**

- pre-lamin
- CAAX-CD0H
- FTIs
- AAX endopeptidase
- C-CD0H
- C-CAAX
- farnesylated/carbonylmethylated pre-lamin A
- mature lamins A, B1, B2
- Znpsp2/FACE3

**B**

- NLS
- Ig-fold
- CAAX

**C**

- lamins A
- lamins C
- lamins B1
- lamins B2

**D**

- mature lamins A

**E**

- 389 417 436 544 646
- 389 417 436 544 572
- 390 415 438 545 583
- 404 435 470 577 617
Figure 4. Mutations in lamins that are known to cause human disease. The dark blue regions show the non-helical domains and the light blue shows the rod domain. Also indicated are the proposed chromatin, LAP2α, and emerin binding domains. The known protein mutations are indicated by squares, circles, triangles, and diamonds to indicate nonsense, point, deletion, and frameshift mutations, respectively. The red star or the red triangle indicates double mutations. The black circle indicates a recessive mutation. 27 homozygous mutations that may also be recessive. The specific amino acid modifications are shown in Table 1.
Lamin mutations disrupt heterochromatin organization

HGPS: Hutchinson-Gilford progeria syndrome
(Point mutation in LaminA gene)

LETTERS

Nuclear pore association confers optimal expression levels for an inducible yeast gene

LETTERS

Transcriptional repression mediated by repositioning of genes to the nuclear lamina
Choreography within the cell nucleus: Ig genes

**Stem cell**
- DNA
- V1, V2, V3, V4, V5, D1, D2, D3, D4, D5, J1, J2, J3, J4, J5, J6, C1, C2, C3, C4, C5, C6

**proB cell**
- DNA
- V1, V2, V3, V4, D1, D2, D3, J1, J2, J3, J4, J5, J6, C1, C2, C3, C4, C5, C6

**preB cell**
- DNA
- mRNA
- Polypeptide
- V1, V2, D1, D2, V3, C1, C2, C3, C4, C5, C6

**Translation**
- RNA splicing

---

**Choreography within the cell nucleus: Ig genes**

**Panel D**
- Lymphoid progenitors
- Pax5−/− pro-B cells
- Pax5+/− pro-B cells
- wt pro-B cells
- wt pro-B cells
- reconstituted Pax5− pro-B cells
- IgH
- Vh/I
- Vh:J
- Pro-B

**Panel A**
- Low complexity
- Locus contraction
- Locus expansion

**Panel B**
- CD19−T

**Panel C**
- Pro-B

**Panel D**
- Pro-B

**Panel E**
- E2A−
An interesting example of gene regulation: AIRE

Immune Tolerance

- State of immune system unresponsiveness to an antigen. Failure to respond to an antigen
- Tolerance to self antigens: Essential feature of the immune system
- Loss of tolerance to self: Destruction of self tissues, autoimmunity

- Central Tolerance: Tolerance established in lymphocytes developing in central lymphoid organs
- Peripheral Tolerance: Tolerance acquired by mature lymphocytes in the peripheral tissues.
APECED: Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy

OMIM Nr 240300

Destructive autoimmune reactions against e.g. parathyroid or thyroid glands, adrenal cortex, gonads, β cells (pancreas), gastric parietal cells, …

Gene responsible for disease was cloned: AIRE/aire
The AIRE protein

Fig. 1 Schematic presentation of the genomic structure of the AIRE gene, the AIRE protein, and the mutations described in AIRE. The mutations marked with black symbols have been identified by us.

AIRE localization

Fig. 2 Subcellular localization of the mouse and human Aire/AIRE proteins in NHEJ3 and BHK cells, respectively. NIH cells were transiently transfected with wild-type mouse Aire cDNA (A, B). BHK cells were infected with recombinant Semliki Forest virus encoding the wild-type human AIRE (C). (A) Mouse Aire showed a distinct speckled distribution in the nucleus. (B) Some cytoplasmic immunostaining of cells expressing mouse Aire was observed. (C) Human AIRE was detected both as small dots in the nucleus, excluding the nucleolus, and as larger granules in the cytoplasm.
How does AIRE work?

Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self

Jens Derbinski, Antje Schulte, Bruno Kyewski and Ludger Klein
"Promiscuous" gene expression

Fig. 1. Schematic view of the architecture of the thymus. The major cell types and the sequential cell-cell interactions along the migratory route of developing T cells are depicted. The different antigen-presenting cells are shown in different colors. Medullary epithelial cells, the role of which in self-tolerance is the focus of this review, are highlighted in red.

| Table 1. Cell type-specific patterns of promiscuous gene expression in C57BL/6 mice

<table>
<thead>
<tr>
<th>Gene expression</th>
<th>Macrophage</th>
<th>DC</th>
<th>mTEC</th>
<th>cTEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CRP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Complement C5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>α- Fetoprotein</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Albumin</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Haptoglobin</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>eN-Antibody</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PLP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MOG</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5-HT</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GAD65</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GAD67</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Somatostatin</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Trypsin</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EGF</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Inhibin</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>eN- Crystalin</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B-Refined S-antigen</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RRIP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Abbreviations: rACL1641, receptor of calcitonin-like 1; xTEC, cortical TEC; cTEC, cortical TEC reactive protein; GAD, glutamic acid decarboxylase; Gp100, melanoma antigen (protein silver); PrP, Prion; I gH, HLA-A; protein pump of the stomach; eN-Ab, endothelial fatty acid binding protein; RRIP, intrathymic cell-specific binding protein; MAGE, melanoma antigen gene; OPG, osteoprotegerin; TEC, thymic epithelial cell.

*No signal, +/+, no or weak signal in duplicate analysis; +, reproducible signal.

*These genes are also expressed in mTECs of N00 mice.
"Promiscuous" gene expression

Thymic microenvironment

- Aire

Die

Yes

Match?

No

Immature thymocytes

Mature T cell repertoire

Fewer self-reactive T cells

Match?

Die

Yes

No

Gene expression in thymic microenvironment: Aire expression affects T cell development and repertoire formation.
Fig. 2. Genomic locations of the top 200 aire-activated (red) and the top 200 aire-repressed (blue) loci. Aire-regulated genes appear to distribute without preference among the chromosomes.

Fig. 3. Aire-regulated genes are clustered. (A) Histogram of intergenic distances <1 Mb for the top 200 aire-activated genes (filled bars) compared with the mean of 1,088 randomly drawn sets of 200 genes. (B) Histogram for the top 200 aire-repressed genes. (C) Two hundred genes were randomly drawn 10,000 times, and the number of genes <200 kb apart were calculated.
Effects of DNA methylation

- The presence of the methyl group alters the appearance of the major groove of DNA and thus can affect the binding of transcription factors
- The positions of 5-methyl cytosine can be passed on during DNA replication (by 'maintenance methylases') and thus act as a long-term molecular markers
Effects of DNA methylation

- Methylation causes:
  - Promoters to be switched off by changing the interaction between DNA and transcription factors and/or chromatin structure
  - X-chromosome inactivation
  - Lack of expression from repetitive and parasite DNA sequences
  - Imprinting
  - Cancer
Effects of DNA methylation

- In mammalian genomes the dinucleotide sequence CpG occurs at only 25% of the expected frequency
  - >98% of C residues in CpG motifs are methylated
  - Methylated C residues mutate to T (deamination) at a very high rate
- CpG deficiency thus corresponds precisely to expectations based on C -> T mutation rates!

Significance of CpG dinucleotides
Analysis of DNA methylation

- Two (non-exclusive) models:
  - Binding of transcription factors prevents methylases from accessing particular sequences
    - Removal of Sp1 target sites flanking CGIs results in methylation of CGI
  - A methylation-targeting mechanism steered by transcription factors
    - Cytosine methyltransferases DNMT1 and DNMT3a associated with Rb, E2F1, HDACs etc.

Establishment of DNA methylation patterns
Distribution of DNA methylation in the gene(ome)

Establishment of DNA methylation patterns
CpG island methylation in disease

Growth regulation and signal transduction - 14-3-3-sigma, Cyclin A1, IGF2, IGFBP7, ABL1, IRF7, LKB1, NF-L, RASSF1, FHIT,

Tumor suppressor - P15 INK4B, P16 INK4A, WT1, VHL, Rh, HIC1, MDGI, APC, N33, DBCCR1, CAV1 (Caveolin), P27KIP1, P57KIP2

Enhancer and transcription factor - GATA-3, DCIS-1, Oct-6, POU3F1, H19, PAX6, MYF3 (MYOD1)

Invasion/metastasis suppressor - E-cadherin, *TIMP-3, Mts-1, CD-44, H-cadherin, THBS1, MUC2, CSPG2, COL9A1, CASP8 (CASPASE 8), GPC3, MT1a, CX26 (Connexin 26), NEP, TLS3 (T-Plasmin), UPA (Urokinase), ZO2 (Zona Occludens 2)

DNA repair/detoxify carcinogens - MGMT, HMLH1, GST Pi, BRAC1, RPA2, HOXA5, TERT, COX2, SIM2

Hormone and kinase receptor - Androgen receptor, ER, PgA, BLT1, Calcitonin, EPHA3, EPO, GALNR2, Teasin, LRP-2 (Megalnin), TGFBR1, RAR-Beta 2

Angiogenesis inhibitor - TSP-1, *TIMP3

Tumor antigen - MAGE-1, ABO

Transporter - CFTR, MDR1, NIS

GTP-protein - Endothelin receptor B, GNAL (GTP-binding protein - Olfactory subunit), P14/ARF

DNA methyltransferases

![DNMT1 (193.5 kDa)](image)

![DNMT3 family](image)

Figure 2: Schematic structure of the three catalytically active DNA methyltransferases in mammals showing the N-terminal regulatory and C-terminal catalytic domains, and other regions with known or proposed functions. The catalytic domains of the three enzymes are conserved, but there is little similarity between their N-terminal regulatory domains. PON1, domain that interacts with phosphorylated cell cycle antigen; NLS, nuclear localization signal; KGM, lysine-glycine repeat hinge region; HDAC, histone deacetylase interaction domain; PHD, plant homeodomain motif that shows homology to the ATR-X (x-thalassemia, mental retardation, X-linked) gene.
Methylated DNA binding proteins

MeCP2

- MeCP2 is an abundantly expressed DNA-binding protein, located in the nucleus and associated with 5-methylcytosine-rich heterochromatin
- Two known functional domains
  - an 84 aa methyl-CpG-binding domain (MBD)
  - a 104 aa transcriptional repression domain (TRD)
- The MBD binds to symmetrically methylated CpG dinucleotides, and the TRD recruits the co-repressor Sin3A
MeCP1 is a complex

Diseases caused by mutations in the methylation machinery

Figure 5 | Summary of mutations in DNA methyltransferase 3B (DNMT3B) identified in ICF syndrome patients. Many individuals were compound heterozygotes for two independent mutations (denoted by ‘+’). Nearly all mutations affect the catalytic domain. Loss of catalytic activity was confirmed for the D809G mutation in cell line overexpression studies. The conserved methyltransferase motifs within the catalytic domain are indicated with roman numerals. The cysteine-rich PHD (plant homeodomain) region in the N-terminus is highly homologous to a PHD motif within the e-thalassemia, mental retardation, X-linked (ATR-X) protein. The S-adenosyl—methionine (SAM) binding domain is indicated, as well as motifs involved in catalysis (IVX) and DNA binding (IVX). Several mutations (Ins.) result in altered splicing patterns (Alu) and the insertion (Ins.) or deletion (Δ) of coding sequence.
Diseases caused by mutations in the methylation machinery

### Box 1. Immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Cytogenetic features</th>
<th>Genetic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertelorism</td>
<td>Chromosomal fusions</td>
<td>Mutations of ICFNAB</td>
</tr>
<tr>
<td>Low-set ears</td>
<td>Extension of juxtacentromeric chromatin</td>
<td></td>
</tr>
<tr>
<td>Epicanthal folds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macroglossia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced immunoglobulins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Succumb to infectious disease before adulthood</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Box 2. Rett syndrome

- Incidence 1 in 10,000–15,000
- X-linked mutations of MECP2
- Males die at birth
- Progressive neurodevelopmental disorder
- Common cause of mental retardation in females
- Normal development until 6–19 months of age
- Gradual loss of speech/purposeful hand use
- Microcephaly, seizures, autism and ataxia

![Diagram showing the influence of Rett syndromes on the function of MeCP2](image)
Diseases caused by mutations in the methylation machinery

Figure 1

Diseases caused by mutations in DNMT3b and MeCP2. Schematic representations of the two proteins are shown with various protein motifs indicated. (a) Schematics of the DNMT3b in EF syndrome by Hanaoka et al. They are indicated in blue, those identified by Osano et al. in green, and those identified by Xu et al. in red. (b) One of the families studied by Hanaoka et al. was also studied by Chen et al. The mutants G595S codifies a proline to arginine transition that occurs in a trinucleotide codon GCT. L1500P refers to a leucine to proline transition that results in a C3T substitution and a C3T substitution in a trinucleotide codon CCT. Mutations in MeCP2 that alter the basic core domain are shown in orange, while those identified by Wang et al. (b) are indicated in red, while those identified by Wang et al. (b) are indicated in red, while those identified by Wang et al. (b) are indicated in red.

TRANSDUCTION OF THE METHYLATION SIGNAL

KAISO

MeCP1 = MBD2 + M2NURD

MBD3

MBD2

MBD1

Transcription Factor

Sin3/MBD

+ MeCP2