Molecular Studies on Complex Biological Systems:

Sources for Transcriptome and Proteome Complexity

October 17, 2007       Dr. Stefan Maas, BioS Lehigh U.
What’s the molecular difference?
What accounts for the often massive and seemingly arbitrary differences in genome size observed among eukaryotic organisms?

The fruit fly  
*Drosophila melanogaster*  
180 Mb

The mountain grasshopper  
*Podisma pedestris*  
18,000 Mb
Is an Expansion in Gene Number driving Evolution of Higher Organisms?

- **Vertebrata**: 30,000
- **Urochordata**: 16,000
- **Arthropoda**: 14,000
- **Nematoda**: 21,000
- **Fungi**: 2,000 – 13,000
- **Vascular plants**: 25,000 – 60,000
- **Unicellular sps.**: 5,000 – 10,000
- **Prokaryotes**: 500 – 7,000
From genes to proteins

DNA

RNA

Protein
Gene splicing: Removal of non-coding introns
Alternative splicing: One gene, several proteins!

Mature splice variant I

Alternative Splicing

Mature splice variant II
Principles of Alternative Splicing
RNA editing

transcription

A-to-I editing

splicing + translation

genomic DNA

pre-mRNA

RNA editing
Effects of Editing

- Adenosine converted to Inosine
- Interpreted as Guanosine
- Expand the proteome
Consequences of Editing

- ADAR1 (p150 + p110)
- ADAR2
- ADAR3

- 5' UTR
- 3' UTR
- Nucleus
- Splice site creation
- Codon change
- Amino acid change
- Intron editing
- 3'UTR editing
- Tankyrase, NADH-Reductase (SU 14.5B)
  - Modulation of mRNA stability and transport?
- Neuronal Cell Adhesion Molecule (NrCAM), Phosphodiesterase (PDE8A)
  - Influence on splicing?

- AMPA-type Glutamate Receptors
  - Regulation of gating behavior
  - Modification of channel kinetics
  - Control of receptor trafficking
- 5HT2C Serotonin Receptor
  - Regulation of G-protein coupling efficiency

J. Biol. Chem. (2003), Vol. 278, pp 1392
# Mammalian substrates of A-to-I pre-mRNA editing

<table>
<thead>
<tr>
<th>Gene</th>
<th>codon</th>
<th>amino acid</th>
<th>editing [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GluR-B</td>
<td>CAG/CIG</td>
<td>Q/R</td>
<td>100</td>
</tr>
<tr>
<td>GluR-B,C,-D</td>
<td>AAG/AIG</td>
<td>R/G</td>
<td>60-80</td>
</tr>
<tr>
<td>GluR-5,-6</td>
<td>CAG/CIG</td>
<td>Q/R</td>
<td>40-80</td>
</tr>
<tr>
<td>GluR-6</td>
<td>AUU/IUU</td>
<td>I/V</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>UAC/UIC</td>
<td>Y/C</td>
<td>80</td>
</tr>
<tr>
<td>5-HT&lt;sub&gt;2C&lt;/sub&gt;</td>
<td>AUA/IUA</td>
<td>I/V</td>
<td>40-90</td>
</tr>
<tr>
<td>Serotonin-receptor</td>
<td>AAU/AIU</td>
<td>N/S</td>
<td>35-40</td>
</tr>
<tr>
<td></td>
<td>AUU/IUU</td>
<td>I/V</td>
<td>45-75</td>
</tr>
</tbody>
</table>
Diversity through RNA editing

GluR-6 pre-mRNA

M1 M2 M3

AUU UAC IUU UIC CAG CIG

I / V Y / C Q / R

unedited edited

5' 3'

GluR-6

N M1 M2 C
Diversity through RNA editing

GluR-6 pre-mRNA

M1 M2 M3

AUU UAC CAG
IUU UIC CIG

I / V Y / C Q / R

GluR-6

N M1 M2 C

I Y I C V Y I Y I C V Y V C I C V Y V C

unedited 10 %

unedited 5 %
unedited 5 %
fully edited 10 %

unedited 5 %

unedited 65 %

fully edited
Even more diversity

paralytic pre-mRNA

1,536 variants

alternative splicing
Even more diversity

paralytic pre-mRNA

1,032,192 variants

alternative splicing

RNA editing

constitutive exon

alternative exon

editing site
Q/R-site editing of glutamate receptor subunit GluR-B

mRNA
-UUU AUG CAG CAA GGA-
F M Q/R Q G

GluR-B

cytosol

N

M1 M2 M3 M4
Q/R-site editing of glutamate receptor subunit GluR-B

mRNA

-UUU AUG CAG CAA GGA-
F M Q/R Q G

postsynaptic neuron

GluR-B

cytosol

>99.9% R
RNA editing enzyme deficient mice

Survival (%) vs. Postnatal day

ADAR2^-/-/GluR-B^+/+

RNA editing enzyme deficient mice: Rescue by GluR-B point mutation

(Higuchi, Maas, Single, Hartner et al., 2000, Nature 406, 78-81)
Question:

- Could too much or too little RNA editing cause disease or alter the progression of known diseases?
Glioblastoma multiforme (GBM)

http://www.thejohnphilpthompsonfoundation.org/GlioblastomaMultiforme_1_.jpg
### Q/R site editing in normal human brain and gliomas

(Maas et al., 2001, *PNAS* 98, 14687-92)

<table>
<thead>
<tr>
<th>Tissue/Mutation Site</th>
<th>Controls</th>
<th>Glioblastoma Tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>cortex</td>
<td>90%</td>
<td>69%</td>
</tr>
<tr>
<td>WM1</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>WM2</td>
<td>100%</td>
<td>79%</td>
</tr>
<tr>
<td>AC1 premalignant tumor</td>
<td>83%</td>
<td>76%</td>
</tr>
<tr>
<td>1</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>2</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>3</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>4</td>
<td>76%</td>
<td>76%</td>
</tr>
<tr>
<td>5</td>
<td>79%</td>
<td>79%</td>
</tr>
<tr>
<td>6</td>
<td>85%</td>
<td>85%</td>
</tr>
<tr>
<td>7</td>
<td>69%</td>
<td>69%</td>
</tr>
</tbody>
</table>

The graph shows the percentage of Q/R site editing in normal human brain and gliomas tissue samples. Q/R site editing in normal human brain samples is lower compared to glioblastoma tissues, indicating a possible link between Q/R site editing and glioblastoma development.
RNA editing of ion channel GluR-B

mRNA

- UUU AUG CAG CAA GGA-
F    M    Q/R    Q    G

ADAR2

I

unedited Glu-R

postsynaptic neuron

Glu Ca²⁺ Glu

Q R

>99.9% R

synapse
RNA editing and cancer

**Too little editing:**

- Malignant growth,
- Activation of oncogenes/repression of tumor suppressor genes

[Glioblastoma (GluR-B)]

**Too much editing:**

- activation of oncogenes or
- inhibition of tumor suppressor genes

[BC10, leukemia (PTPN6), pancreatic cancer (prox1); hepatoma (APOBEC1), neurofibromatosis (NF1), Wilm’s tumor (WT1)]
A-to-I RNA editing and human diseases

- Glioblastoma multiforme => GluR-B underediting due to ADAR2 deficiency (Maas et al. PNAS 2001)
- Amyotrophic Lateral Sclerosis (ALS) => GluR-B Q/R-site underediting in motor neurons (Kawahara et al. Nature 2004)
- Suicidal Depression => Change in Serotonin receptor editing (Gurevich et al., Neuron 2002)

Hyperediting phenotypes: Lupus erythematosis, inflammatory lung disease
Question:

How many genes are subject to RNA editing in the human genome?

Investigate using the complete human genome sequence as an important tool.
The “Smoking Gun” of A-to-I RNA editing

- TTT ATG CAG CAA GGA-  
  genomic DNA

- UUU AUG CAG CAA GGA-  
  unedited mRNA

- UUU AUG CGG CAA GGA-  
  edited

Reverse transcription

- TTT ATG CAG CAA GGA-  
  cDNA

- TTT ATG CGG CAA GGA-
In silico evidence of RNA editing

Athanasiadis et al., *PLoS Biology* 2004
Experimental validation

Athanasiadis et al., *PLoS Biology* 2004

*Number of edited sites exceeds predicted numbers*
Analysis of 103,723 human cDNA sequences

17,406 contain one or more Alu

Athanasiadis et al., *PLoS Biology* 2004
Transposable elements in the human genome

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of copies</th>
<th>Percentage of total genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINEs</td>
<td>1,558,000</td>
<td>13.1</td>
</tr>
<tr>
<td>\textit{Alu}</td>
<td>1,090,000</td>
<td>10.6</td>
</tr>
<tr>
<td>LINEs</td>
<td>868,000</td>
<td>20.4</td>
</tr>
<tr>
<td>\textit{LINE1}</td>
<td>516,000</td>
<td>16.9</td>
</tr>
<tr>
<td>LTR elements</td>
<td>443,000</td>
<td>8.3</td>
</tr>
<tr>
<td>DNA elements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{mariner}</td>
<td>294,000</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>14,000</td>
<td>0.1</td>
</tr>
<tr>
<td>Unclassified</td>
<td>3,000</td>
<td>0.1</td>
</tr>
<tr>
<td>Total of all types</td>
<td></td>
<td>44.7</td>
</tr>
</tbody>
</table>

Alu-mediated RNA foldback structures

Levanon et al. EMBO Rep. 2005
Experimental validation

Athanasiadis et al., PLoS Biology 2004
Tissue origin of Alus determines editing extent

Athanasiadis et al., *PLoS Biology* 2004
Distance between inverted Alu pairs

Athanasiadis et al., PLoS Biology 2004
Functions (?)

- None
- Accelerated evolution
- Alu biology
- Prevent RNAi
- RNA transport, stability, translation
Spectrum of edited human RNAs

- GluRs
- $5\text{HT}_{2c}$
- etc

- ADAR2 int
- simple repeats
- MIR pairs
- LINE pairs
- Alu pairs

- ds character
- site selectivity
- inosine content
- Prevalence?

Average for each class
Summary - Conclusions

**We do know:**

- Complexity is generated by processes of alternative splicing and RNA editing
- RNA editing regulates the function of genes through recoding and probably through other mechanisms
- We can use genomic and transcriptomic sequence information to search for RNA variations

**We don’t know:**

- What is the total impact of RNA editing on complexity? How is it regulated in vivo?
- Are there other molecular phenomena that contribute to complexity?
- What is all the non-coding RNA (and DNA) doing?