10/31/05

RNA Structure & Function

Announcements

Seminar (Mon Oct 31)
12:10 PM IG Faculty Seminar in 101 Ind Ed II
Plant Steroid Hormone Signal Transduction
Yanhai Yin, GDCB

BCB link for Seminar Schedules (updated)
http://www.bcb.iastate.edu/seminars/index.html

BCB 544 Projects - Important Dates:
Nov 2 Wed noon - Project proposals due to David/Drena
Nov 4 Fri 10A - Approvals/responses to students
Dec 2 Fri noon - Written project reports due
Dec 5,7,8,9 class/lab - Oral Presentations (20')
(Dec 15 Thurs = Final Exam)

RNA Structure & Function Prediction

Mon
Review - promoter prediction
RNA structure & function

Wed
RNA structure prediction
2' & 3' structure prediction
miRNA & target prediction
RNA function prediction?

Reading Assignment (for Mon/Wed)

Mount Bioinformatics
- Chp 8 Prediction of RNA Secondary Structure
  pp. 327-355
- Ck Errata: http://www.bioinformaticsonline.org/help/errata2.html

Cates (Online) RNA Secondary Structure Prediction Module
- http://cse.rice.edu/content/m11065/latest/

Review last lecture:

Promoter Prediction
Promoter Prediction

- Overview of strategies
  - What sequence signals can be used?
  - What other types of information can be used?

- Algorithms ➔ a bit more about these in later lectures

- Promoter prediction software
  - 3 major types
  - many, many programs!

Promoter Prediction: Eukaryotes vs prokaryotes

Promoter prediction is easier in microbial genomes

Why?
- Highly conserved
- Simpler gene structures
- More sequenced genomes!
  (for comparative approaches)

Methods?
- Previously, again mostly HMM-based
- Now: similarity-based, comparative methods
  because so many genomes available

Promoter Prediction: Steps & Strategies

- Closely related to gene prediction!
- Obtain genomic sequence
- Use sequence-similarity based comparison
  (BLAST, MSA) to find related genes
  But: "regulatory" regions are much less well-conserved than coding regions
- Locate ORFs
- Identify TSS (Transcription Start Site)
- Use promoter prediction programs
- Analyze motifs, etc. in sequence (TRANSFAC)

Identify TSS -- if possible?
- One of biggest problems is determining exact TSS!
- Not very many full-length cDNAs!
- Good starting point? (human & vertebrate genes)
  Use FirstEF
  found within UCSC Genome Browser
  or submit to FirstEF web server

Promoter prediction strategies

1) Pattern-driven algorithms
2) Sequence-driven algorithms
3) Combined "evidence-based"

BEST RESULTS? Combined, sequential

Promoter Prediction: Pattern-driven algorithms

- Success depends on availability of collections of annotated binding sites (TRANSFAC & PROMO)
- Tend to produce huge numbers of FPs

Why?
- Binding sites (BS) for specific TFs often variable
- Binding sites are short (typically 5-15 bp)
- Interactions between TFs (& other proteins) influence affinity & specificity of TF binding
- One binding site often recognized by multiple BFs
- Biology is complex: promoters often specific to organism/cell/stage/environmental condition
Promoter Prediction: Pattern-driven algorithms

Solutions to problem of too many FP predictions?
• Take sequence context/biology into account
  • Eukaryotes: clusters of TFBSs are common
  • Prokaryotes: knowledge of σ factors helps
• Probability of “real” binding site increases if annotated transcription start site (TSS) nearby
  • But: What about enhancers? (no TSS nearby)
  
  & Only a small fraction of TSSs have been experimentally mapped
• Do the wet lab experiments!
  • But: Promoter-bashing is tedious

Promoter Prediction: Sequence-driven algorithms

Problems:
• Need sets of co-regulated genes
• For comparative (phylogenetic) methods
  • Must choose appropriate species
  • Different genomes evolve at different rates
  • Classical alignment methods have trouble with translocations, inversions in order of functional elements
  • If background conservation of entire region is highly conserved, comparison is useless
• Not enough data (Prokaryotes >>> Eukaryotes)
  • Biology is complex: many (most?) regulatory elements are not conserved across species!

Examples of promoter prediction/characterization software

Lab: used MATCH, MatInspector
  • TRANSFAC
  • MEME & MAST
  • BLAST, etc.

Others?
  • FIRST EF
  • Dragon Promoter Finder (these are links in PPTs)
  • also see Dragon Genome Explorer (has specialized promoter software for GC-rich DNA, finding CpG islands, etc)
  • JASPAR

Check out optional review & try associated tutorial:


Check this out: http://www.phylofoot.org/NRG_testcases/
Annotated lists of promoter databases & promoter prediction software

- URLs from Mount Chp 9, available online
  - Table 9.12 http://www.bioinformaticsonline.org/links/ch_09_1_2.html
- URLs for Baxevanis & Ouellette, Chp 5:

More lists:
- http://bioinformatics.ubc.ca/resources/links_directory?subcategory_id=104
- http://www3.oup.co.uk/nar/database/subcat/1/4/

New Today:

RNA Structure & Function

- RNA structure
  - Levels of organization
  - Bonds & energetics (more about this on Wed)
- RNA types & functions
  - Genomic information storage/transfer
  - Structural
  - Catalytic
  - Regulatory

RNA structure: 3 levels of organization

Rob Knight
Univ Colorado

Covalent & non-covalent bonds in RNA

Primary: Covalent bonds
Secondary/Tertiary Non-covalent bonds
  - H-bonds (base-pairing)
  - Base stacking

G-C, A-U, G-U ("wobble") & variants

See: IMB Image Library of Biological Molecules
http://www.imb.jena.de/ImgLibDoc/nana/IMAGE_NANA.html#sec_element
Common structural motifs in RNA

- Helices
- Loops
  - Hairpin
  - Interior
  - Bulge
  - Multibranch
- Pseudoknots

Fig 6.2
Baxevanis & Ouellette 2005

RNA functions

- Storage/transfer of genetic information
  - Genomes
    - Many viruses have RNA genomes
      - Single-stranded (ssRNA)
        - E.g., retroviruses (HIV)
      - Double-stranded (dsRNA)
  - Transfer of genetic information
    - mRNA = "coding RNA" - encodes proteins

Structural
  - E.g., rRNA, which is major structural component of ribosomes (Gloria Culver, ISU)
  - But - its role is not just structural, also:

Catalytic
  - RNA in ribosome has peptidyltransferase activity
    - Enzymatic activity responsible for peptide bond formation between amino acids in growing peptide chain
    - Also, many small RNAs are enzymes "ribozymes" (W Allen Miller, ISU)

Regulatory
  - Recently discovered important new roles for RNAs in normal cells:
    - In "defense" - esp. in plants
    - In normal development
      - E.g., siRNAs, miRNA
  - As tools:
    - For gene therapy or to modify gene expression
      - RNAi (used by many at ISU: Diane Bassham, Thomas Baum, Jeff Essner, Kristen Johansen, Jo Anne Powell-Coffman, Roger Wise, etc.)
      - RNA aptamers (Marit Nilsen-Hamilton, ISU)

RNA types & functions

<table>
<thead>
<tr>
<th>Types of RNAs</th>
<th>Primary Function(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA - messenger</td>
<td>translation (protein synthesis) regulatory</td>
</tr>
<tr>
<td>rRNA - ribosomal</td>
<td>translation (protein synthesis) catalytic</td>
</tr>
<tr>
<td>t-RNA - transfer</td>
<td>translation (protein synthesis)</td>
</tr>
<tr>
<td>hnRNA - heterogeneous nuclear</td>
<td>precursors &amp; intermediates of mature mRNAs &amp; other RNAs</td>
</tr>
<tr>
<td>snRNA - small cytoplasmic</td>
<td>signal recognition particle (SRP) RNA processing catalytic</td>
</tr>
<tr>
<td>siRNA - small nuclear</td>
<td>mRNA processing, poly A addition catalytic</td>
</tr>
<tr>
<td>snoRNA - small nucleolar</td>
<td>RNA processing/methylation</td>
</tr>
<tr>
<td>regulatory RNAs (siRNA, miRNA, etc.)</td>
<td>regulation of transcription and translation, other?</td>
</tr>
</tbody>
</table>
Expression of a Typical Eukaryotic Gene

Gene Expression Challenges for Computational Biology

• Understand the ‘code’ for each step in gene expression (set of default recognition rules), e.g., the ‘splicing code’

• Understand the rules for sequence-specific recognition of nucleic acids by protein and ribonucleoprotein (RNP) factors

• Understand the regulatory events that occur at each step and the biological consequences of regulation

Lots of data
Genomes, structures, transcripts, microarrays, ChIP-Chip, etc.

Sequence-specific Transcription Factors

• typically bind in clusters

  • Regulatory modules

Yan (ISU) A computational method to identify amino acid residues involved in protein-DNA interactions

Kadonaga Cell 2004
**Integration of transcription & RNA processing**

Maniatis & Reed Nature 2002

---

**Early Steps in Pre-mRNA Splicing**

- Formation of exon-spanning complex
- Subsequent rearrangement to form intron-spanning spliceosomes which catalyze intron excision and exon ligation

Matlin, Clark & Smith Nature Mol Cell Biol 2005

---

**Alternative Splicing**

> 50% of human genes undergo alternative splicing

Matlin, Clark & Smith Nature Mol Cell Biol 2005

---

**Splicing Regulation**

ESE/ESS = Exonic Splicing Enhancers/Silencers
ISE/ISS = Intronic Splicing Enhancers/Silencers

Matlin, Clark & Smith Nature Mol Cell Biol 2005

---

**Coupling of Splicing & Nonsense-mediated mRNA Decay (NMD)**

Maniatis & Reed Nature 2002

---

**C. elegans lin-4 Small Regulatory RNA**

V. Ambros lab

We now know that there are hundreds of miRNA genes
(Ambros, Bartel, Carrington, Ruvkun, Tuschl, others)
MicroRNA Biogenesis

MicroRNA and RNAi pathways

miRNA and RNAi pathways

miRNA Challenges for Computational Biology

• Find the genes encoding microRNAs
• Predict their regulatory targets
• Integrate miRNAs into gene regulatory pathways & networks

Need to modify traditional paradigm of "transcriptional control" by protein-DNA interactions to include miRNA regulatory mechanisms