Goals:
1. Relate miRNA structure to their function as post-transcriptional regulators.
2. Generalize approaches we have previously discussed to predict miRNA genes and targets.

miRNA overview
- Short RNA molecules (~22 nt)
- Post-transcriptional regulation: silencing/repression
- Form RNA-induced Silencing Complex (RISC) which:
  1. Catalyzes mRNA cleavage (open question: how?)
  2. Destabilizes mRNA, competes with ribosomes
- Evidence they evolved independently in plants and animals
- Seem to be important for complex, multicellular life: play roles in development, differentiation (but also disease, cancer, etc.)
- Have subtle effects, precise timing
- RISC opens new field of RNA interference (RNAi): models (knockouts), drugs

miRNA biogenesis

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>Cytoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA</td>
<td>Helicase</td>
</tr>
<tr>
<td>~1000 nt</td>
<td>miRNA*</td>
</tr>
<tr>
<td>~10 nt</td>
<td>~22 nt</td>
</tr>
<tr>
<td>~50 nt</td>
<td></td>
</tr>
<tr>
<td>mismatch</td>
<td></td>
</tr>
<tr>
<td>Drosha</td>
<td></td>
</tr>
<tr>
<td>Pasha</td>
<td></td>
</tr>
<tr>
<td>Dicer</td>
<td></td>
</tr>
<tr>
<td>miRNA</td>
<td></td>
</tr>
<tr>
<td>RISC</td>
<td></td>
</tr>
<tr>
<td>(Ago2)</td>
<td></td>
</tr>
</tbody>
</table>
miRNA Targeting
- Ago2 only exposes positions 2-5 (seed region)
- Complementary base pairing with target causes conformation change, exposing 6-8.
- Non-canonical binding sites: only 6 nt, supplementary region 13-16, compensatory binding

miRNA gene finding
Filtering: look for 22 nt fragments in RNA-seq. Look for small loop, stacking (Zuker algorithm). Look for conserved hairpin sequence.

Problems: need to distinguish miRNA from degraded fragments; can't detect recently evolved miRNAs; no easily interpretable rules, insight into the biology.

Homology: align putative miRNA transcripts to known examples, cluster miRNAs to learn common patterns

Classification:
- HMM over double-stranded structures (stacking)
- SVM/NBC/random forest to distinguish miRNAs from others

Problems: what features are important? How to pick a negative set? How to pick a positive set?
- Some features: folding free energy, length, AU, UG, GG pairs, stem loop properties

miRNA target prediction
- Look for perfect match in nt. 2-8, extend to larger target
- All of the previous issues still apply
- Should we consider folding energy of the target?