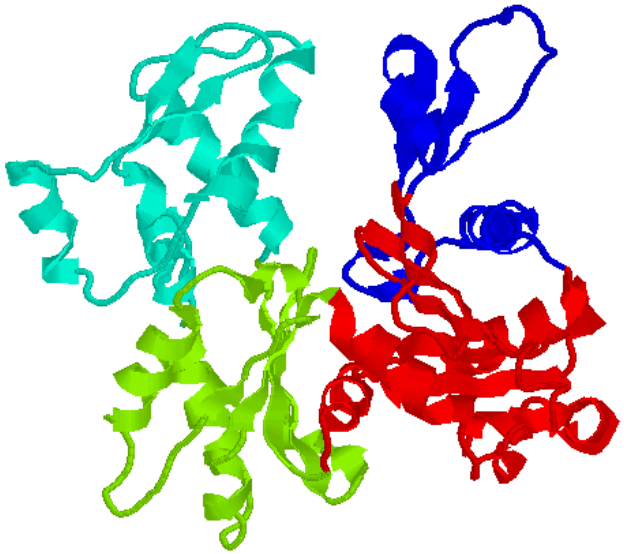
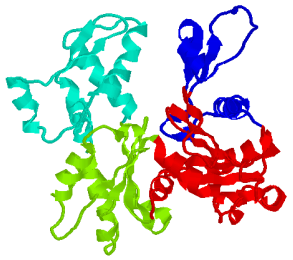
Genome Annotation



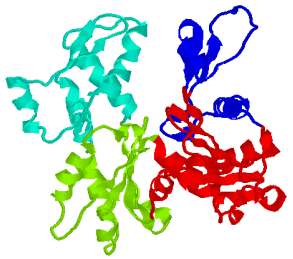
Dong Xu

Digital Biology Laboratory
Computer Science Department
Christopher S. Life Sciences Center
University of Missouri, Columbia
http://digbio.missouri.edu



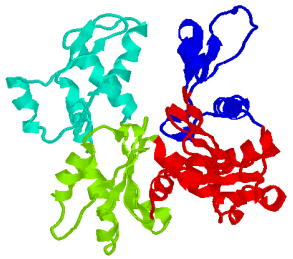
Lecture Outline

- Introduction
- Manual Curation
- Automatic Annotation
- Conclusions



What is Genome Annotation?

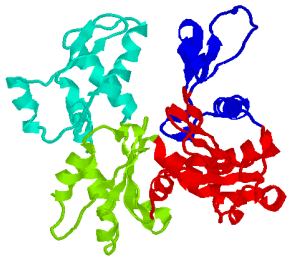
- Annotation is the process of interpreting raw sequence data into useful biological information by integrating computational analyses, other biological data and biological expertise
- It involves characterizing genomic features using computational and experimental methods
- Features could be repeats, genes, promoters, protein domains……..



What is Genome Annotation?

Questions:

- What genes does this genome contain?
- What proteins do they encode?
- How are they regulated?
- In what interactions or pathways do the proteins participate?



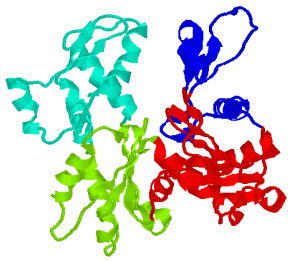
Types of Genome Annotation

Structural annotation

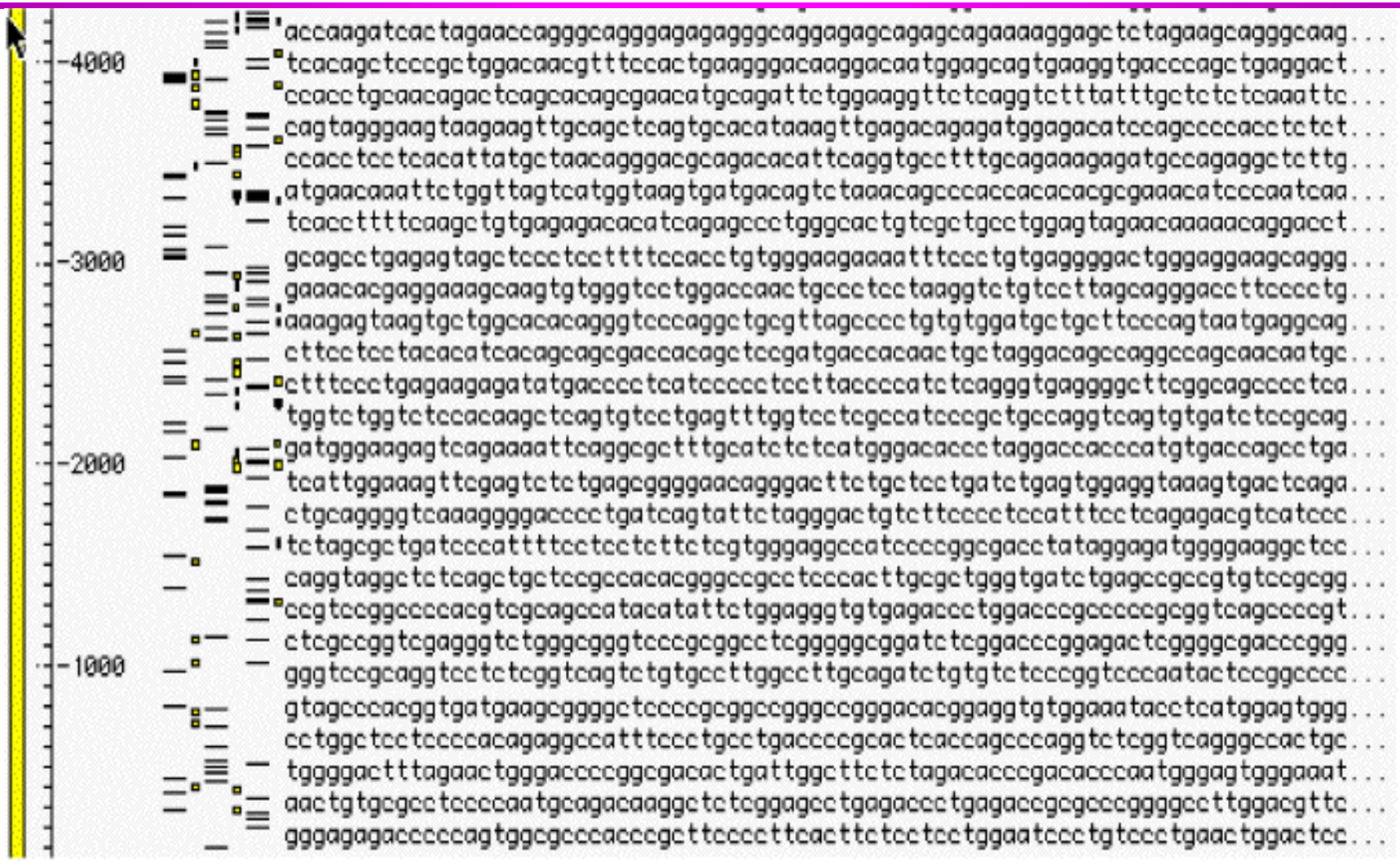
- Location of protein-coding genes
- Location of regions of homology with
  - other genomes
  - cDNA sequences
  - protein sequences
- Location and type of transcription regulatory elements

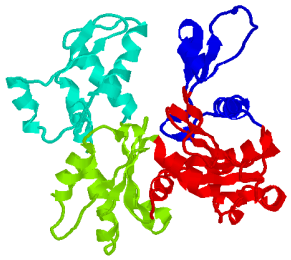
Functional annotation

- Molecular function of encoded proteins
- Membership in metabolic and regulatory networks

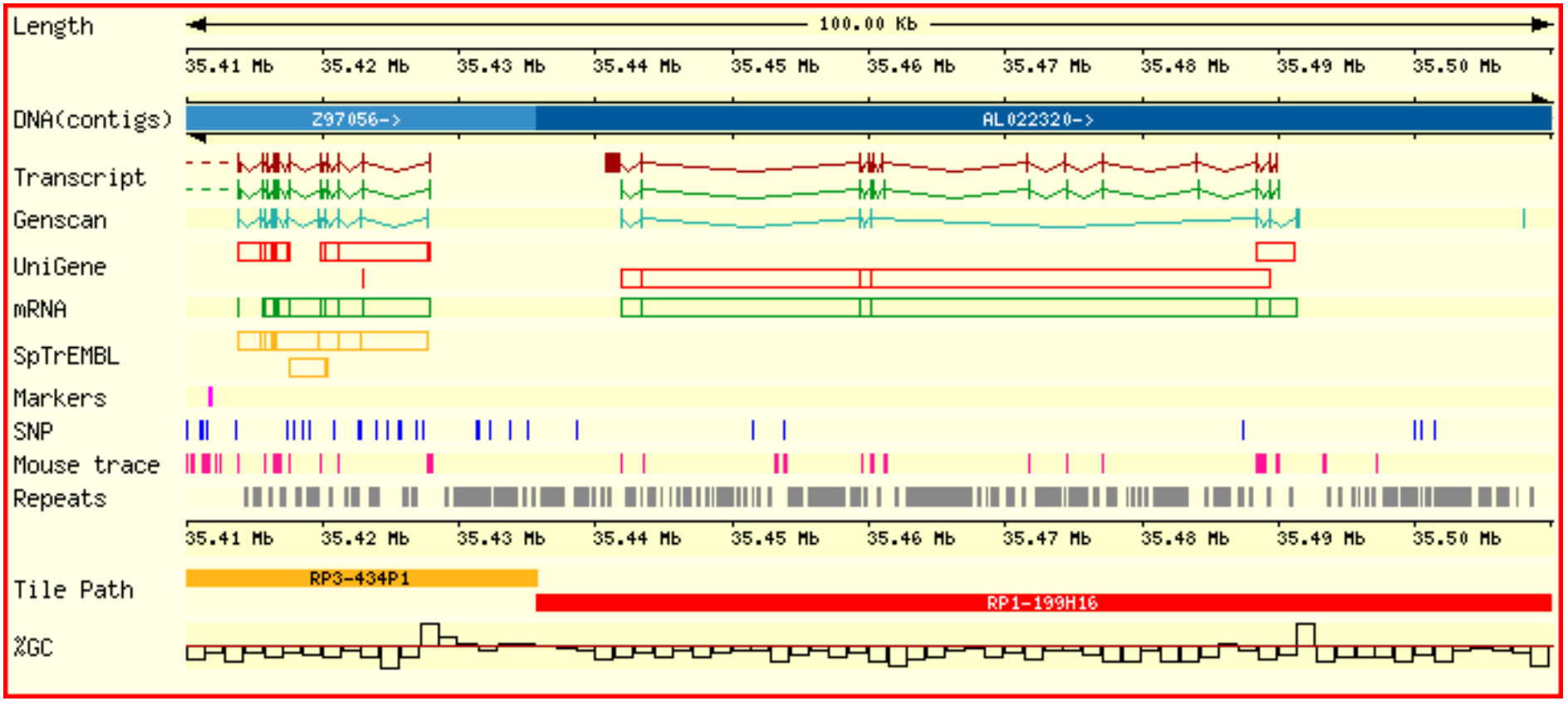


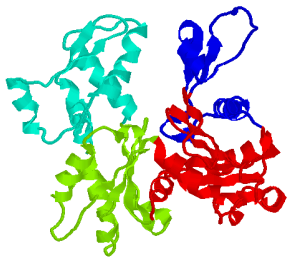
Aim: To get from here …





to here,

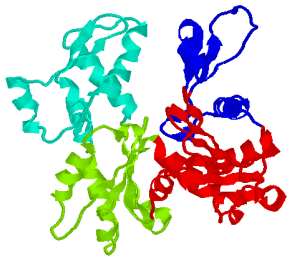

Length
100.00 Kb
35.41 Mb
35.42 Mb
35.43 Mb
35.44 Mb
35.45 Mb
35.46 Mb
35.47 Mb
35.48 Mb
35.49 Mb
35.50 Mb
DNA(contigs)
Z97056->
AL022320->
Transcript
Genscan
UniGene
mRNA
SpTrEMBL
Markers
SNP
Mouse trace
Repeats
Tile Path
RP3-434P1
RP1-199H16
%GC



What are genes? - 1

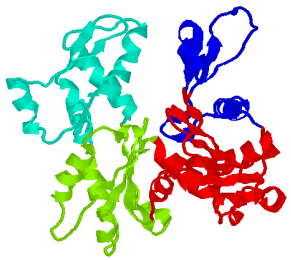
- Complete DNA segments responsible to make functional products
- Products
  - Proteins
  - Functional RNA molecules
    - RNAi (interfering RNA)
    - rRNA (ribosomal RNA)
    - snRNA (small nuclear)
    - snoRNA (small nucleolar)
    - tRNA (transfer RNA)

Non-coding RNA



What are genes? - 2

- Definition *vs. d*ynamic concept
- Consider
  - Prokaryotic *vs.* eukaryotic gene models
  - Introns/exons
  - Posttranscriptional modifications
  - Alternative splicing
  - Differential expression
  - Posttranslational modifications
  - Multi-subunit proteins



Prokaryotic Gene Structure

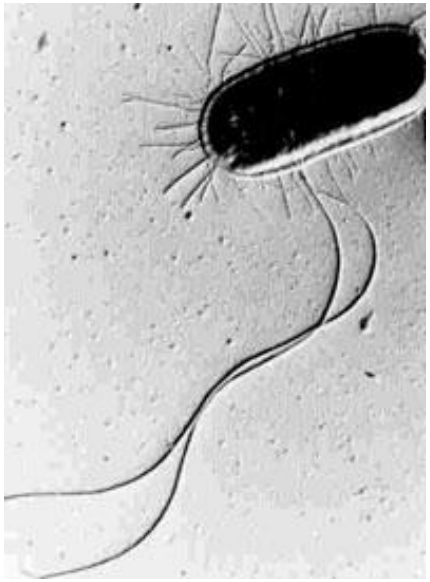
Coding region of Open Reading Frame

Promoter region (maybe)

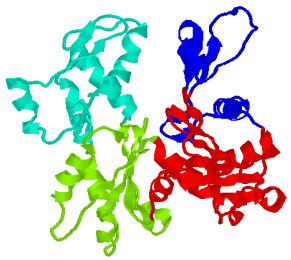
Ribosome binding site (maybe)

Termination sequence (maybe)

Start codon / Stop Codon

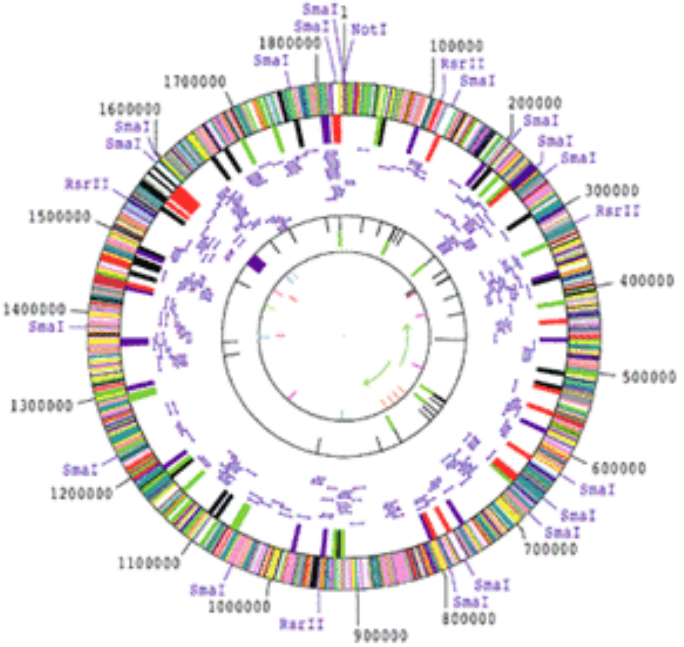


Open reading frame (ORF): a segment of DNA with two in-frame stop codons at the two ends and no in-frame stop codon in the middle



Prokaryotic gene model: ORF-genes

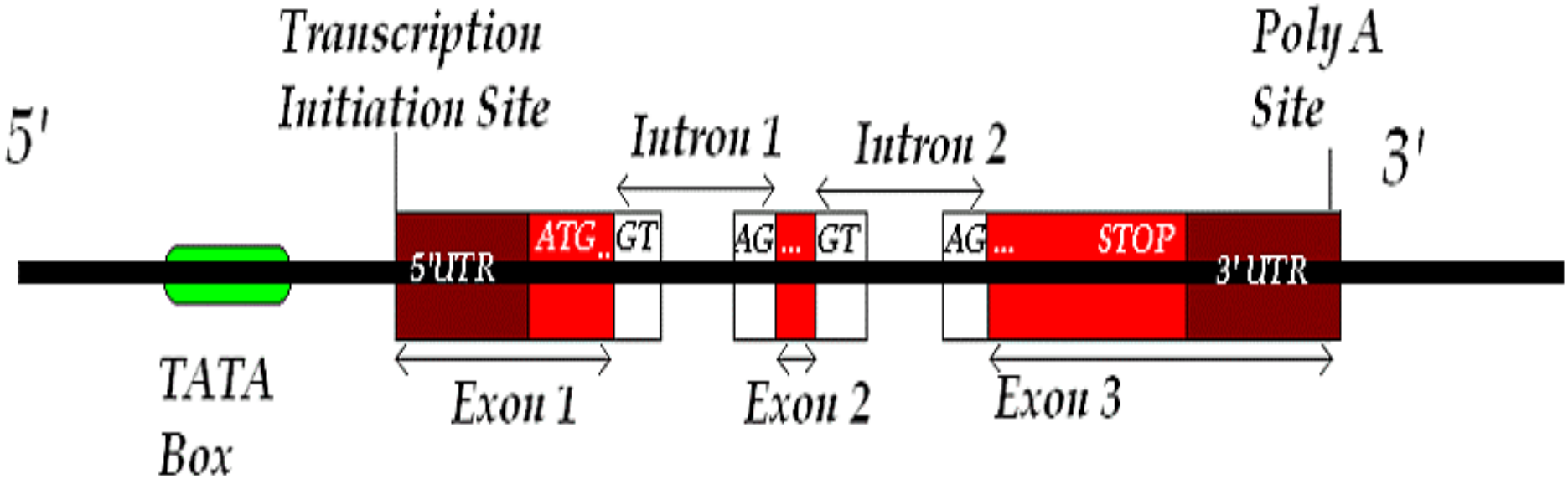
- "Small" genomes, high gene density
  - *Haemophilus influenza* genome 85% genic
- Operons
  - One transcript, many genes
- No introns
  - One gene, one protein
- Open reading frames
  - One ORF per gene
  - ORFs begin with start, end with stop codon (def.)



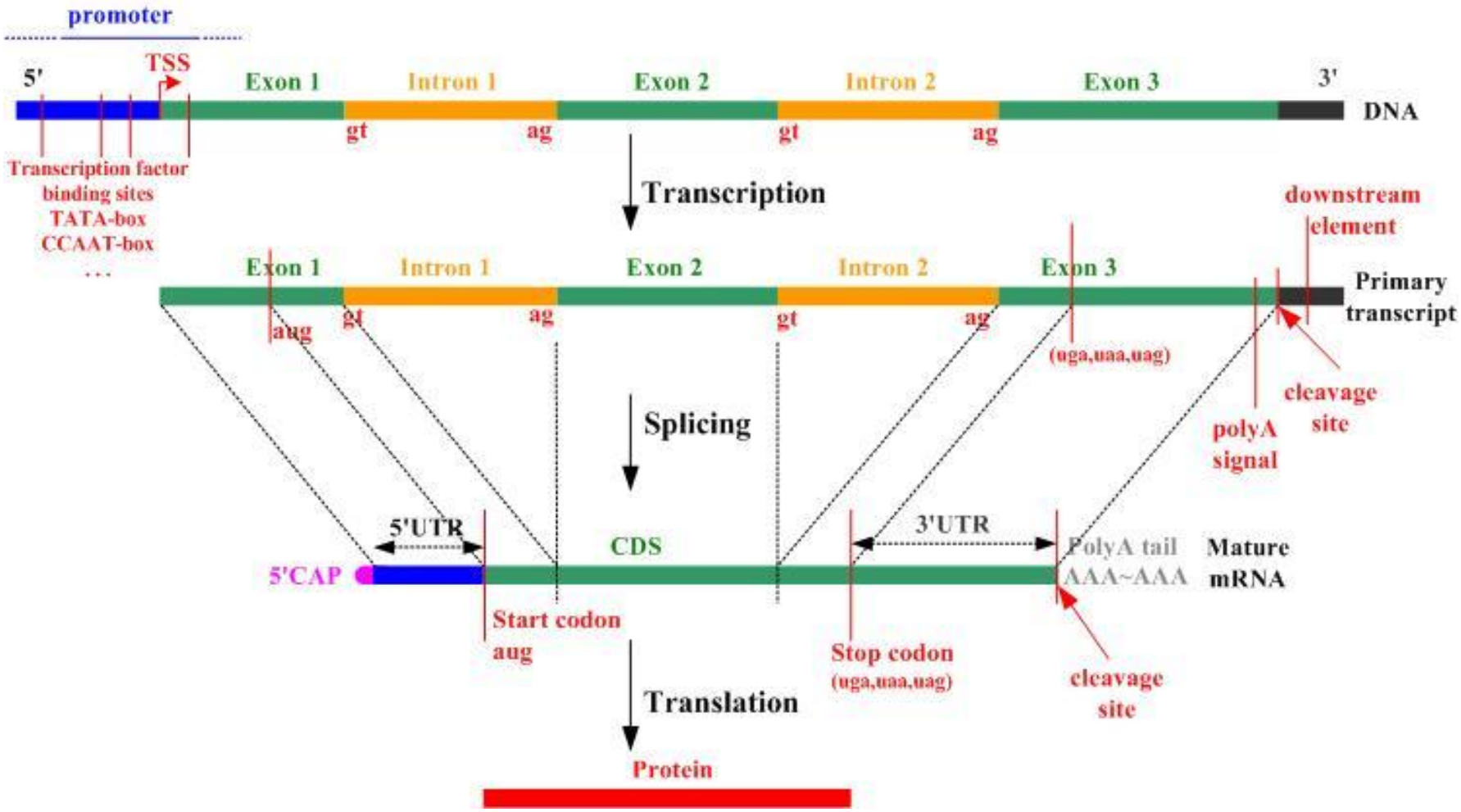
TIGR: http://www.tigr.org/tigr-scripts/CMR2/CMRGenomes.spl

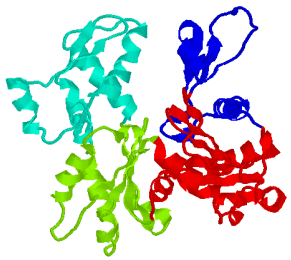
NCBI: http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html

Eukaryotic gene model: spliced genes



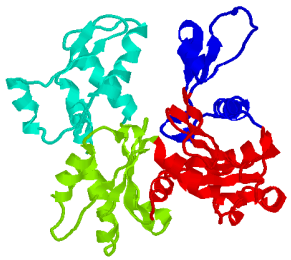
Eukaryotic Gene Structure





Genetic Code

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC: Phe	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr	UGU, UGC: Cys	U C A G
	UUA, UUG: Leu		UAA: Stop; UAG: Stop	UGA: Stop; UGG: Trp	
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His	CGU, CGC, CGA, CGG: Arg	U C A G
			CAA, CAG: Gln		
A	AUU, AUC, AUA: Ile	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn	AGU, AGC: Ser	U C A G
	AUG: Met		AAA, AAG: Lys	AGA, AGG: Arg	
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp	GGU, GGC, GGA, GGG: Gly	U C A G
			GAA, GAG: Glu		



Reading Frame

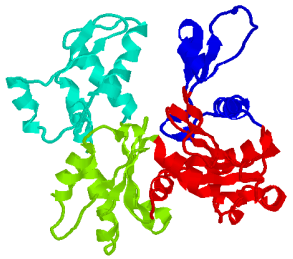
- Reading (or translation) frame: each DNA segment has six possible reading frames

Forward strand: 

Reading frame #1	Reading frame #2	Reading frame #3
ATG	TGG	GGC
GCT	CTT	TTA
TAC	ACG	CGC
GCT	CTT	TTG
TGC	GA.	A..

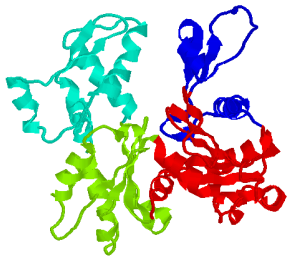
Reverse strand: 

Reading frame #4	Reading frame #5	Reading frame #6
TCA	CAA	AAG
AGC	GCG	CGT
GTA	TAA	AAG
AGC	GCC	CCA
CAT	AT.	T..



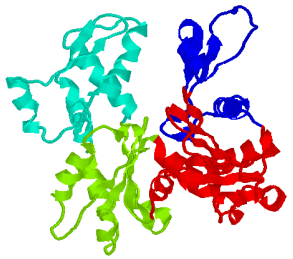
Lecture Outline

- Introduction
- Manual Curation
- Automatic Annotation
- Challenges or Pitfalls
- Conclusions



Manual Curation

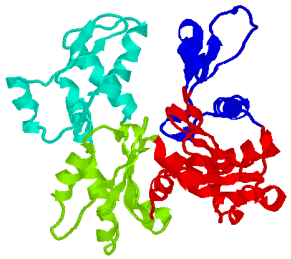
- Annotation for all genes in the genome has been manually reviewed by a curator and should be regarded as accurate as possible given the data available at the time of curation.
- Annotation data assigned has been based on all evidence available to the curator.
- In addition, gene models are curated to remove overlapping genes, resolve frameshifted genes, and determine the initiation codon of each gene.



Manual Curation

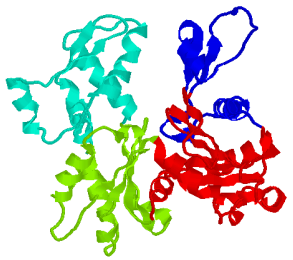
Task involves identifying Genes

- Known
- Novel
- Novel transcript
- Putative
- Pseudogene



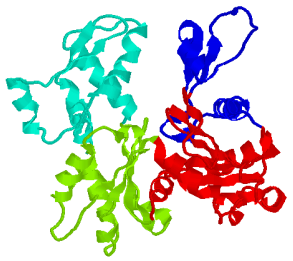
Annotation nomenclature

- **Known Gene** – Predicted gene matches the entire length of a known gene.
- **Putative Gene** – Predicted gene contains region conserved with known gene. Also referred to as “like” or “similar to”.
- **Unknown Gene** – Predicted gene matches a gene or EST of which the function is not known.
- **Hypothetical Gene** – Predicted gene that does not contain significant similarity to any known gene or EST.



Things curators are looking to annotate?

- CDS
- mRNA
- Alternative RNA splicing
- Promoter and Poly-A Signal
- Pseudogenes
- ncRNA

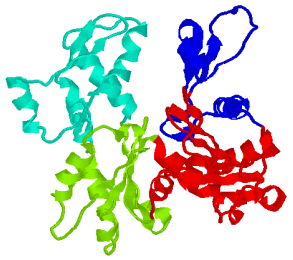


Pseudogenes

- Could be as high as 20-30% of all Genomic sequence predictions could be pseudogene
- Non-functional copy of a gene
  - Processed pseudogene
    - Retro-transposon derived
    - No 5’ promoters
    - No introns
    - Often includes polyA tail
  - Non-processed pseudogene
    - Gene duplication derived
  - Both include events that make the gene non-functional
    - Frameshift
    - Stop codons
- We assume pseudogenes have no function, but we really don’t know!

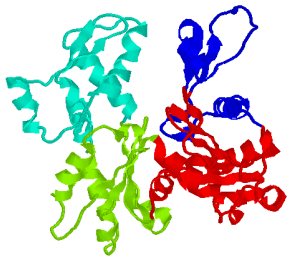
Example Pseudogene

```
LOCUS       NG_005487               1850 bp    DNA     linear   ROD 14-FEB-2006
DEFINITION  Mus musculus ubiquitin-conjugating enzyme E2 variant 2 pseudogene
            (LOC625221) on chromosome 6.
ACCESSION   NG_005487
VERSION     NG_005487.1  GI:87239965
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1850)
  AUTHORS   Wilson,R.
  TITLE     Mus musculus BAC clone RP24-201D17 from 6
  JOURNAL   Unpublished (2003)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AC121925.2.
FEATURES             Location/Qualifiers
     source          1..1850
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /note="AC121925.2 32277..34126"
     gene            101..1750
                     /gene="LOC625221"
                     /pseudo
                     /db_xref="GeneID:625221"
     repeat_region   1792..1827
                     /rpt_family="ID"
ORIGIN
        1 tcttctgcct caattcctca agtgctagta tcatatgccc atgccattat ttttaactcc
       61 ccttttttcat gctaagaatt gaacacacgg ccctgcgtgc ggtggtgcgt ctggtagcag
      121 gagaagatgg cggtctccac aggagttaaa gttcctcgta attttcgctt gttggaagaa
```



Noncoding RNA (ncRNA)

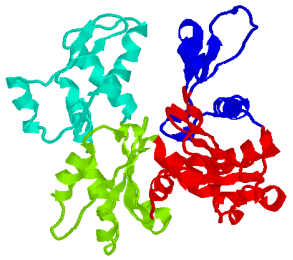
- ncRNA represent 98% of all transcripts in a mammalian cell
- ncRNA have not been taken into account in gene counts
  - cDNA
  - ORF computational prediction
  - Comparative genomics looking at ORF
- ncRNA can be:
  - Structural
  - Catalytic
  - Regulatory



Example ncRNA

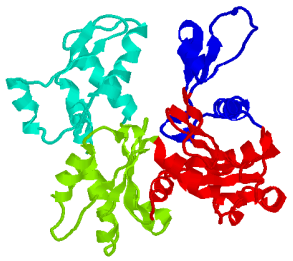
From NW_632744.1

```
gene            complement(55100..55691)
                /locus_tag="CR40465"
                /note="synonym: CR_tc_AT13310"
                /db_xref="GeneID:3354945"
misc_RNA        complement(55100..55691)
                /locus_tag="CR40465"
                /note="This annotation is identical to the ncRNA
                CR_tc_AT13310 annotation, also mapped identically to 2L
                [20224138,20223553]
                last curated on Thu Jan 15 13:37:02 PST 2004"
                /db_xref="FlyBase:FBgn0058465"
                /db_xref="GeneID:3354945"
```



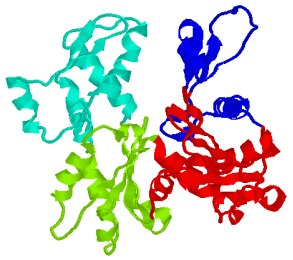
Noncoding RNA (ncRNA)

- tRNA – transfer RNA: involved in translation
- rRNA – ribosomal RNA: structural component of ribosome, where translation takes place
- snoRNA – small nucleolar RNA: functional/catalytic in RNA maturation
- Antisense RNA: gene regulation / silencing?



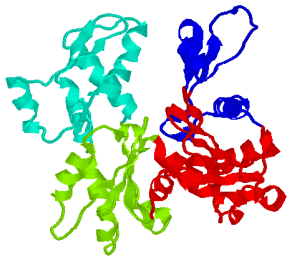
Repetitive Sequence

- Definition
  - DNA sequences that made up of copies of the same or nearly the same nucleotide sequence
  - Present in many copies per chromosome set



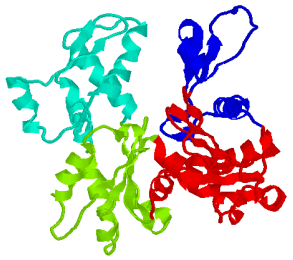
Repeat Filtering

- RepeatMasker
  - Uses precompiled representative sequence libraries to find homologous copies of known repeat families
  - Use Blast
  - http://www.repeatmasker.org/



Lecture Outline

- Introduction
- Manual Curation
- Automatic Annotation
- Conclusions

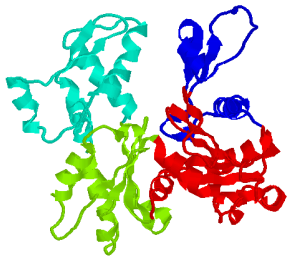


Automatic Annotation

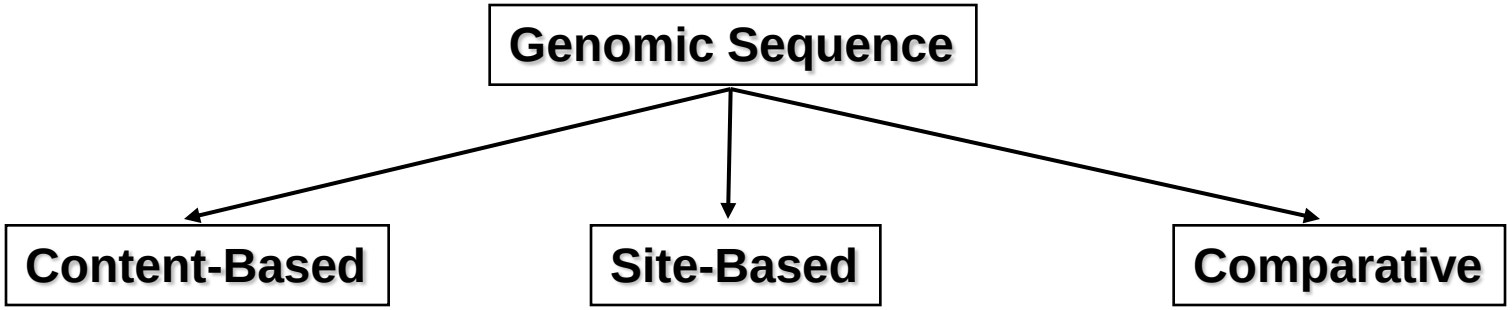
- Automated annotation describes annotation which has been generated by an computational algorithm without being further curated.

Who does automatic annotation?

- EnsEMBL
- NCBI
- UCSC



Gene-Finding Strategies



Bulk properties of sequence:

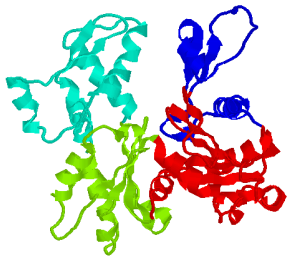
- Open reading frames
- Codon usage
- Repeat periodicity
- Compositional complexity

Absolute properties of sequence:

- Consensus sequences
- Donor and acceptor splice sites
- Transcription factor binding sites
- Polyadenylation signals
- “Right” ATG start
- Stop codons out-of-context

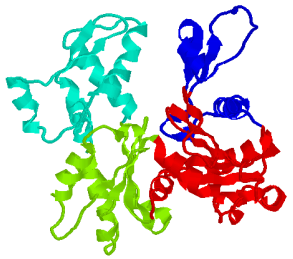
Inferences based on sequence homology:

- Protein sequence with similarity to translated product of query
- Modular structure of proteins usually precludes finding complete gene



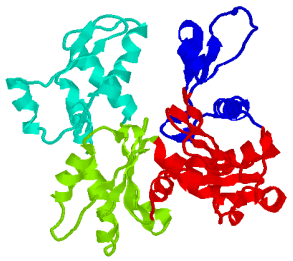
Gene-Finding Strategies

- **Homology-based gene prediction**
  - Similarity Searches (*e.g.* BLAST)
  - Genome Browsers
  - RNA evidence (ESTs)
- ***Ab initio gene prediction***
  - Gene prediction programs
  - Prokaryotes
    - ORF identification
  - Eukaryotes
    - Promoter prediction
    - PolyA-signal prediction
    - Splice site, start/stop-codon predictions



Homology based approaches

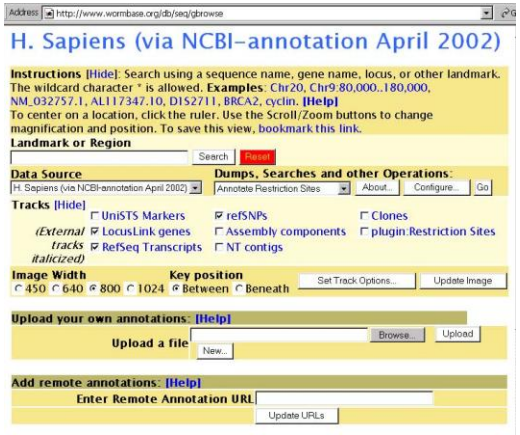
- Idea is new species are not produced from scratch, they are evolutionary related to extant species
- Search by local alignment programs
  - EST/cDNA to genome : BlastN, FASTA
  - Protein to genome : TBlastN



Gene prediction through comparative genomics

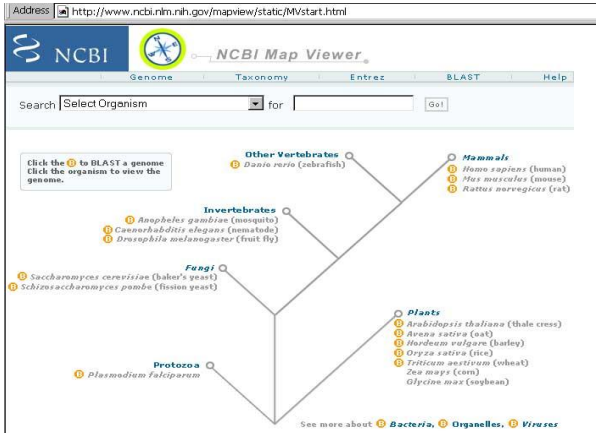
- Highly similar (Conserved) regions between two genomes are useful or else they would have diverged
- If genomes are too closely related all regions are similar, not just genes
- If genomes are too far apart, analogous regions may be too dissimilar to be found

Genome Browsers



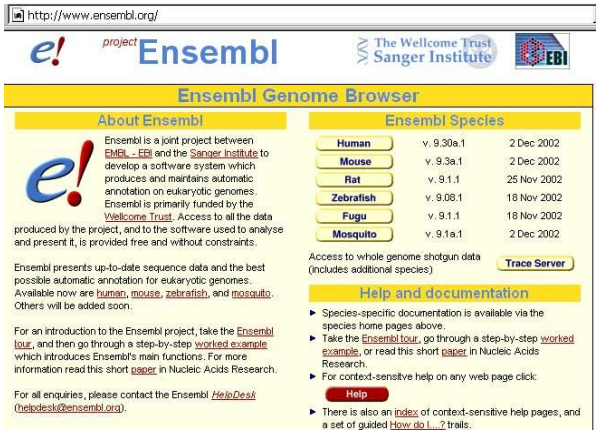
Generic Genome Browser (CSHL)

www.wormbase.org/db/seq/gbrowse



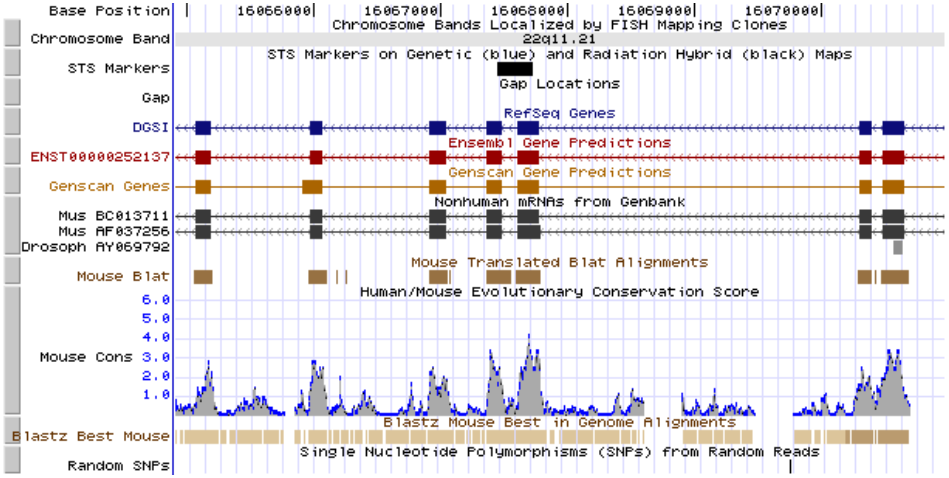
NCBI Map Viewer

www.ncbi.nlm.nih.gov/mapview/



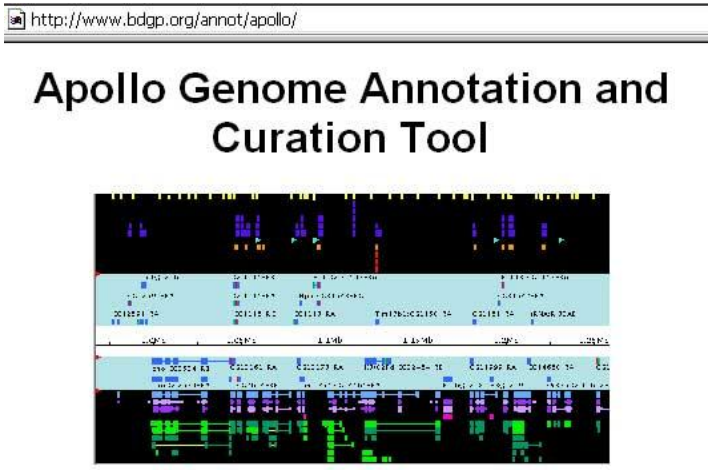
Ensembl Genome Browser

www.ensembl.org/



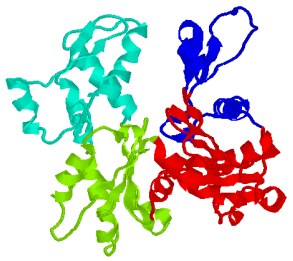
UCSC Genome Browser

genome.ucsc.edu/cgi-bin/hgGateway?org=human



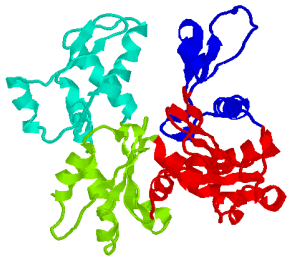
Apollo Genome Browser

www.bdgp.org/annot/apollo/



Gene discovery using ESTs

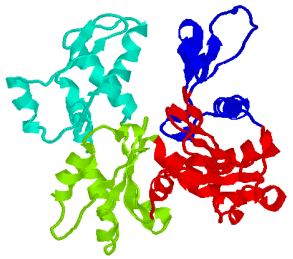
- Expressed Sequence Tags (ESTs) represent sequences from expressed genes.
- If region matches EST with high stringency then region is probably a gene or pseudo gene.
  - EST overlapping exon boundary gives an accurate prediction of exon boundary.



Ab initio approach -1

Rely on

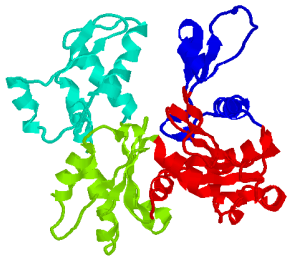
- Identification of specific signals : start codon, stop codon, ribosomal binding site
  - The **Shine-Dalgarno Sequence** (AGGAGG) is the signal for initiation of **protein biosynthesis** in **bacterial mRNA**.
  - It is located 5' of the first coding AUG, and consists primarily, but not exclusively, of purines.
  - It is the ribosomal binding site.



Ab initio approach -2

Rely on

- Differences in nucleotide-motif composition between
  - protein coding and non-coding sequences
  - Correct reading frame of a gene and other reading frames



Coding Signal Detection

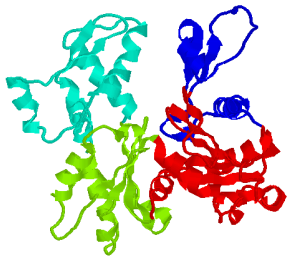
- Frequency distribution of dimers in protein sequence (shewanella)

Name	ala	arg	asn	asp	cys	glu	gln	gly	his	ile	leu	lys	met	phe	pro	ser	thr	trp	tyr	val
ala	9.5	4.1	4.3	5.3	1.2	6	4.8	6.5	2	6.5	11.5	6	2.6	3.7	3.5	6.2	5	1.1	2.7	6.5
arg	7.9	5.5	3.9	5.3	1.1	6	5.5	5.9	2.6	6.5	11.4	5	2.2	4.7	3.6	5.5	4.4	1.4	4	6.6
asn	9.6	4.9	4.2	4.9	1	5.3	5.6	7.4	2.3	6	10	4.9	2	3.5	5.1	6.1	5.5	1.5	3.1	6.1
asp	9.3	4	4.7	5.1	1	6.7	2.9	7	1.8	7.1	9.6	6.3	2.3	4.3	3.9	5.9	5.1	1.6	3.6	6.6
cys	8.4	4.8	3.3	5.4	1.7	5.6	5.2	8.1	4.3	5.4	10.2	3.8	1.8	4.1	4.5	6.3	4.3	1.6	3.4	6.8
glu	9.4	5.8	3.6	4.5	0.8	4.9	7	5.8	2.6	5.9	12.7	5	2.4	4	3.5	5.4	5	1.1	2.8	6.8
gln	10.3	4.9	3	4.4	0.9	4.5	6.8	7	2.7	5.5	12.8	4.1	2	3.9	3.8	5.8	5.3	1.4	3	6.9
gly	8.1	4.8	3.9	5.1	1.2	6	4.6	6.4	2.4	6.8	10.5	5.8	2.7	4.8	2.4	5.8	5.1	1.4	3.7	7.5
his	7.3	4.7	4	4.8	1.5	4.9	5.6	6.9	3	6.2	10.8	4.8	1.6	5	5.2	6.8	4.9	1.7	4.2	5.1
ile	11	4.7	4.9	6.5	1.1	6.9	3.6	7.2	2.1	5.3	8.6	5.3	1.8	3.2	4.2	7	5.6	0.9	2.9	6.1
leu	10.4	4.2	4.3	5.2	1.1	5.2	3.7	6.8	2	5.6	10.6	5.3	2.3	3.8	4.5	7.4	6.2	1	2.6	6.6
lys	10.6	5.2	3.8	5.2	0.5	5.3	5.9	6.6	2.6	5.2	11.3	4.7	1.9	2.8	4.6	6	5.5	1.2	2.6	7.6
met	10.8	4.8	3.8	4.6	0.7	4.6	4.9	7	1.7	4.7	11.4	5.2	2.8	3.3	5.1	7.4	6.3	0.9	2	6.8
phe	9.6	3.7	5.2	6.5	1.2	6.4	2.7	7.9	1.9	6.7	7.4	5	2.5	3.9	3.6	8	5.8	1.3	3.3	6.3
pro	8.4	3.6	4.6	5.4	0.7	7.6	5.2	5.4	2.3	6.1	11.2	5.5	2.4	4.2	2.8	6.5	5.4	1.4	2.9	7.5
ser	9.1	4.6	3.7	5	1	5.4	5.2	7.2	2.6	6	11.6	4.5	2.2	4.1	4.1	6.5	5	1.2	3.2	6.8
thr	9.1	4.2	3.7	5.6	0.9	5.7	5.7	7.5	2.2	5.5	12	4.2	2	3.5	5.5	6.2	5.3	1.1	2.6	6.7
trp	7.1	6.3	3.2	4.8	1.3	3.9	8.5	6.6	3.6	5	14.2	3.2	2.4	4.6	3.9	5.8	4.3	1.3	3	6.1
tyr	7.9	6.5	3.6	4.9	1.2	4.5	7	7.1	2.6	5	11.7	4	1.6	4.7	4.9	6.4	4.6	1.5	3.4	5.7
val	9.6	4.1	4.4	5.9	1	6.2	3.4	6.4	1.8	6.5	10.2	5.2	2.5	3.7	3.8	7.2	6.1	1.1	2.7	7.1

The average frequency is 5%

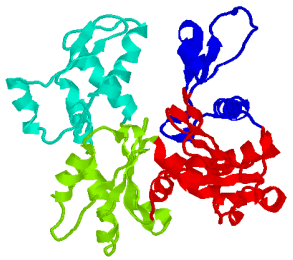
Some amino acids prefer to be next to each other

Some other amino acids prefer to be not next to each other



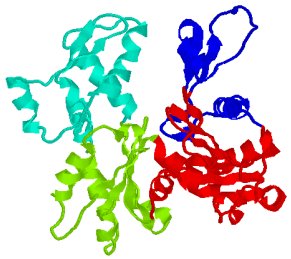
Ab initio approach -3

- Compositional differences
  - Nucleotides in coding and non-coding regions evolve under different constraints
  - First and second codon position are constrained by the encoded amino acid
  - Third codon position is subject to mutational and translation efficiency constraints
  - Nucleotide in non-coding regions can evolve independently



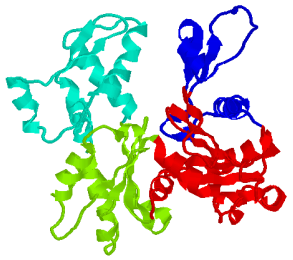
Ab initio gene prediction

- Prokaryotes
  - ORF-Detectors
- Eukaryotes
  - Position, extent & direction: through promoter and polyA-signal predictors
  - Structure: through splice site predictors
  - Exact location of coding sequences: through determination of relationships between potential start codons, splice sites, ORFs, and stop codons



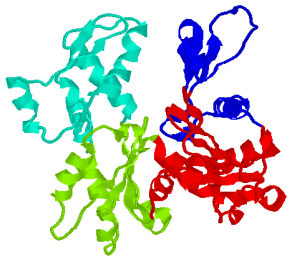
Gene prediction programs

- Rule-based programs
  - Use explicit set of rules to make decisions.
  - Example: GeneFinder
- Neural Network-based programs
  - Use data set to build rules.
  - Examples: Grail, GrailEXP
- Hidden Markov Model-based programs
  - Use probabilities of states and transitions between these states to predict features.
  - Examples: Genscan, GenomeScan



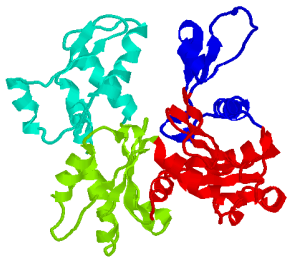
Tools for Annotation

- EnsEMBL (EBI)
- Sequin (NCBI)
- PseudoCAP (SFU)
- GMOD (CSHL)
- Pegasys (UBiC)
- Apollo (EBI/Berkeley)

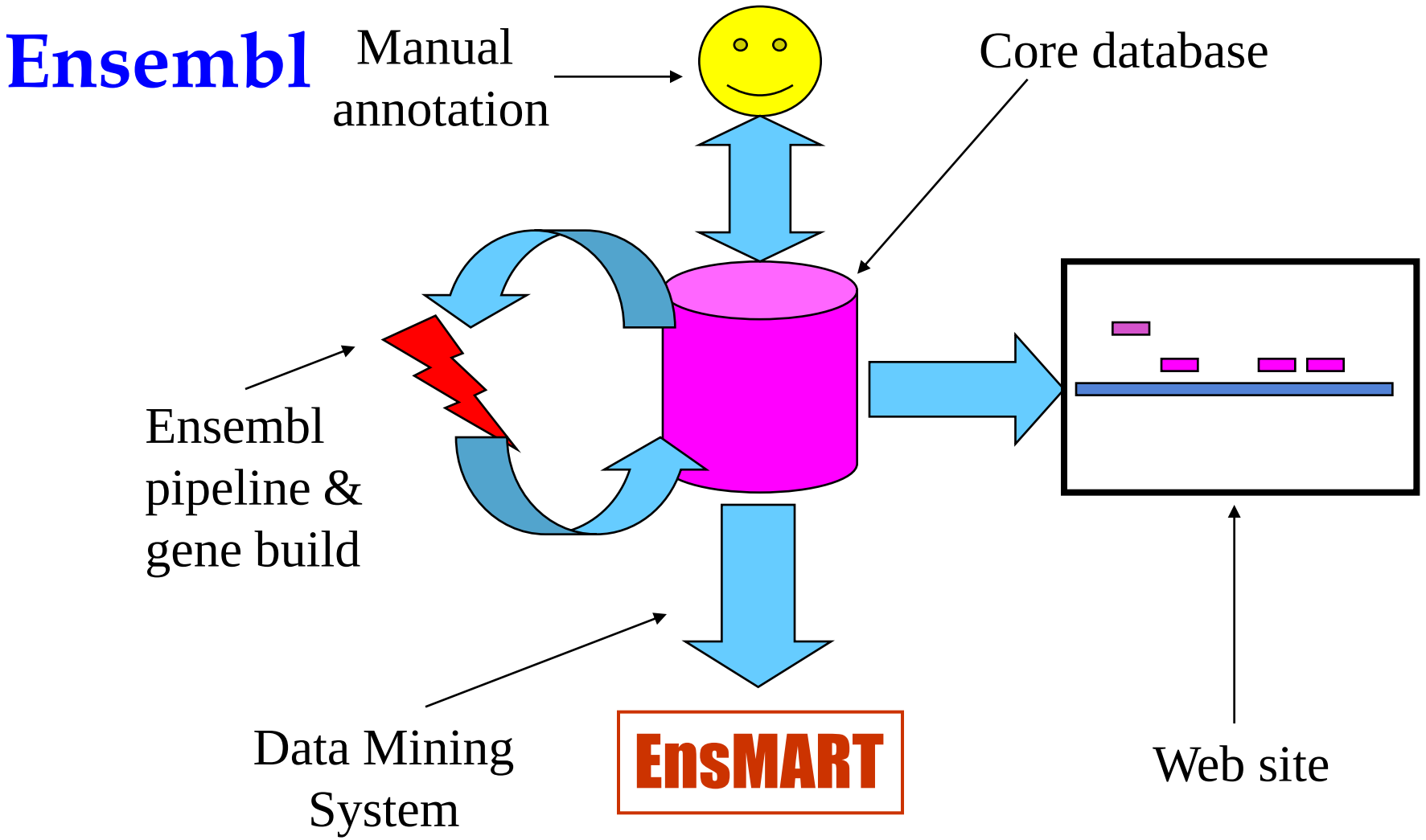


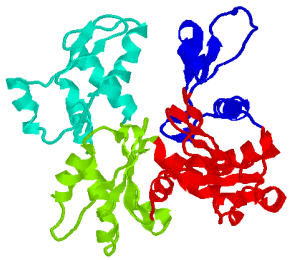
Tools for Annotation

- ORF detectors
  - NCBI: http://www.ncbi.nih.gov/gorf/gorf.html
- Promoter predictors
  - CSHL: http://rulai.cshl.org/software/index1.htm
  - BDGP: fruitfly.org/seq_tools/promoter.html
  - ICG: TATA-Box predictor
- PolyA signal predictors
  - CSHL: argon.cshl.org/tabaska/polyadq_form.html
- Splice site predictors
  - BDGP: http://www.fruitfly.org/seq_tools/splice.html
- Start-/stop-codon identifiers
  - DNALC: Translator/ORF-Finder
  - BCM: Searchlauncher



Example : Ensembl Automatic Annotation Process -1





Ensembl Automatic Annotation Process -2

Raw Compute

Sequence data arrives in contigs

Repeat masking

Ab initio predictions (Genscan)

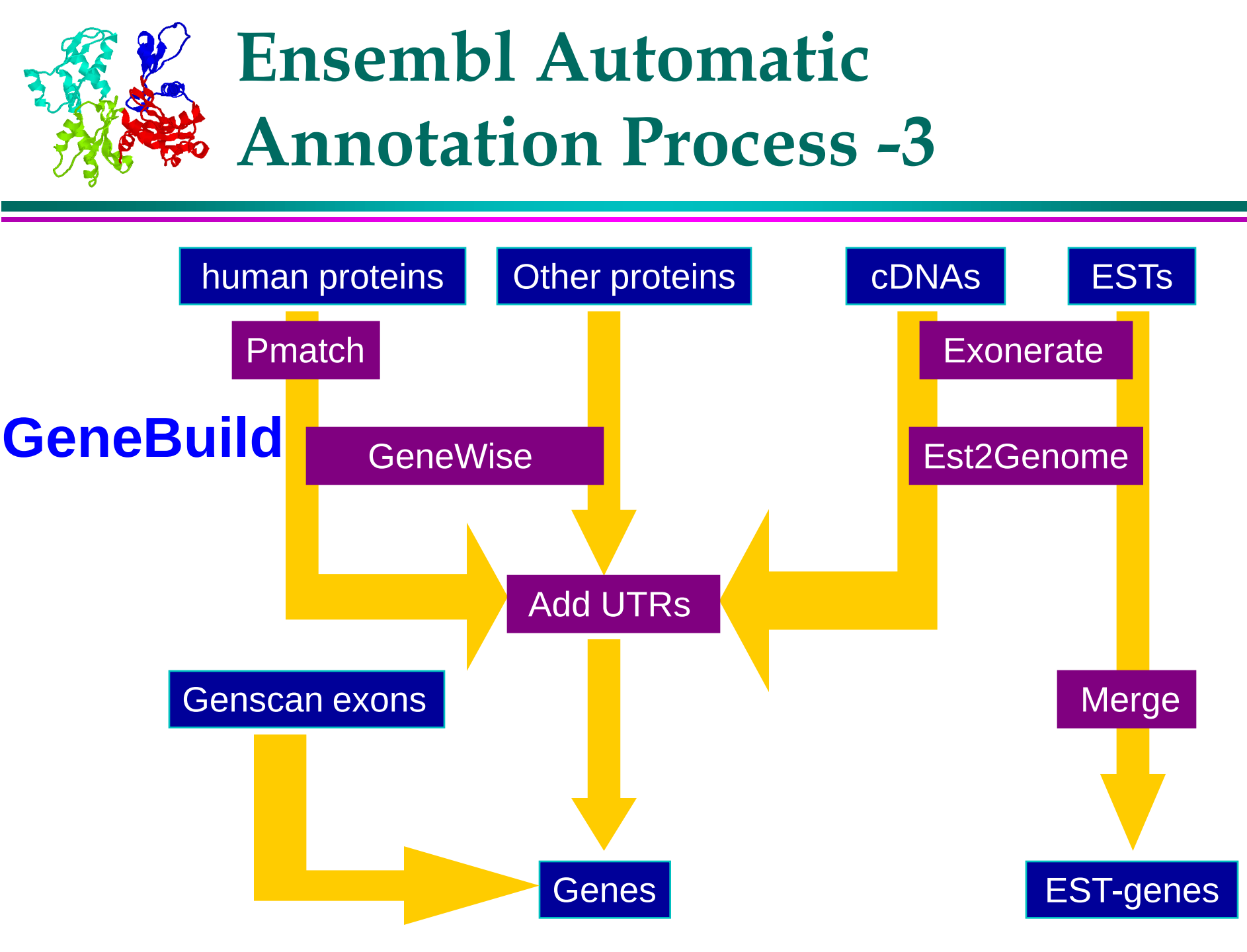
Blast the predictions against: swall, vertebrate RNA, unigene

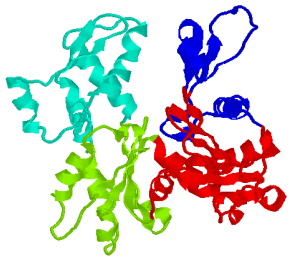
ePCR places markers on the sequence

Assembly information is used to position contigs on a “golden path”

EnsEMBL core

Ensembl Automatic Annotation Process -3





Ensembl Automatic Annotation Process -4

Genewise

Protein Sequences

⇩

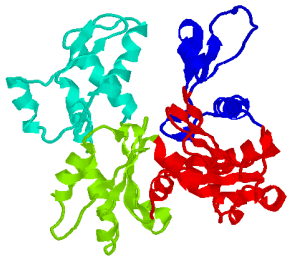
Aligned to the Genome

⇩

Blast and MiniSeq

⇩

Genewise



Ensembl Automatic Annotation Process -5

ESTs and cDNA

Map cDNAs and ESTs using Exonerate
(determine coverage, % identity and location in genome)

↓

Store hits and filter on percentage identity and length coverage

↓

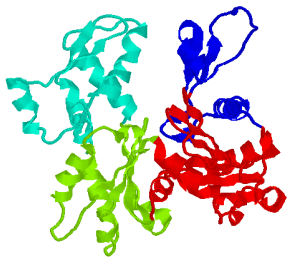
blast sequence and create a miniseq

↓

Run est2genome on miniseq
(determine strand, splicing)

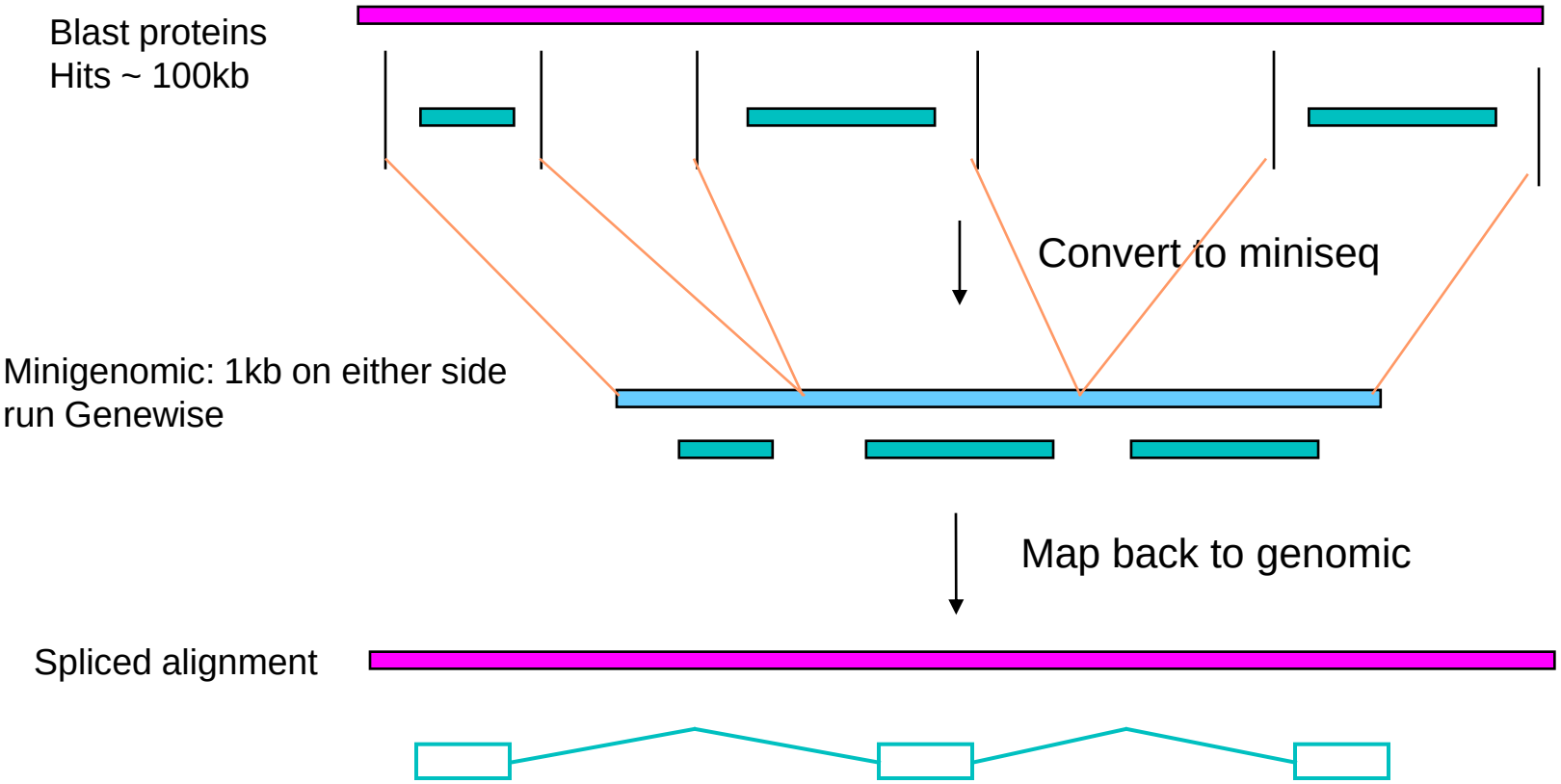
↓

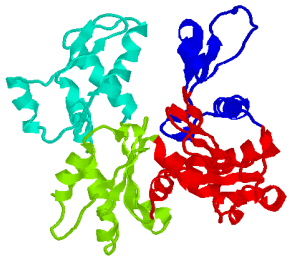
Map transcripts back into genome-assembly



Ensembl Automatic Annotation Process -6

Miniseq - the need for speed





NCBI GenBank Features

-10_signal
-35_signal
3'clip
3'UTR
5'clip
5'UTR
attenuator
CAAT_signal
CDS
conflict
C_region
D-loop
D_segment
enhancer
exon
GC_signal
gene
iDNA
intron
J_segment
LTR
mat_peptide
misc_binding
misc_difference
misc_feature
misc_recomb
misc_RNA
misc_signal
misc_structure
modified_base
mRNA
N_region
old_sequence
polyA_signal
polyA_site
precursor_RNA
primer_bind
prim_transcript
promoter
protein_bind
RBS
repeat_region
repeat_unit
rep_origin
rRNA
satellite
scRNA
sig_peptide
snoRNA
snRNA
S_region
stem_loop
STS
TATA_signal
terminator
transit_peptide
tRNA
unsure
variation
V_region
V_segment

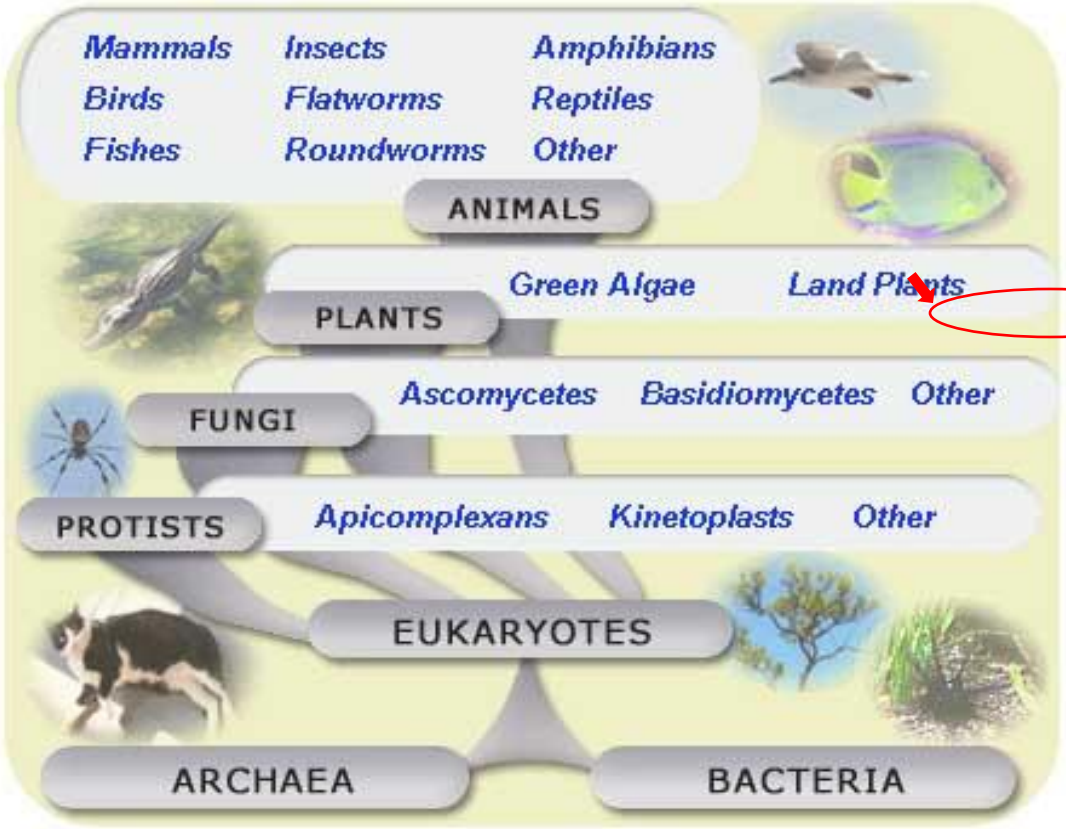
About Entrez

Entrez Genome Project
Home
Overview
Help
Statistics
Sequencing Centers

Submitting
Project Submissions
Project Instructions
General Genome Submissions
Feature Tables
Bacterial Genome Submissions
Whole Genome Shotgun Sequences

Related Resources
DOE Projects
DOE SAI Survey
Genome News Network
Genomes OnLine Database
IntlGenome
NHGRI Projects
NIAID Projects
TIGR Projects

Welcome to the NCBI Entrez Genome Project database.
This searchable database is a collection of complete and incomplete large-scale sequencing, assembly, annotation, and mapping projects for cellular organisms. The database is organized into organism-specific overviews that function as portals from which all projects in the database pertaining to that organism can be browsed and retrieved. Read more...



Updates

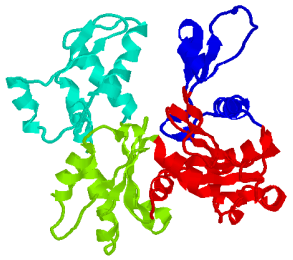
Proposal to Improve the use of Locus Tags in Microbial Genomes
NCBI and ASM have a joint proposal to improve the use of locus tags in microbial genomes. An updated search interface is available here.

NCBI Resources

Entrez Gene
gene-related information
Entrez Genome
sequence and map data from whole genomes
Eukaryotic Projects
eukaryotic-specific genome projects
Genomic Biology
organism-specific links
Prokaryotic Projects
prokaryotic-specific genome projects
Organellar Genomes
organellar reference sequences and tools
Plant Genomes
major plant genome projects
RefSeq
the reference sequence project
Viral Genomes
viral reference sequences and tools
WGS Sequences
whole genome shotgun sequences

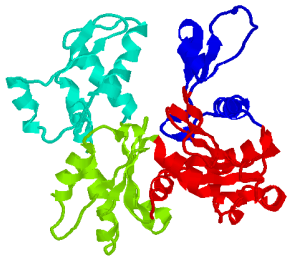
NCBI Tools

COGs
clusters of orthologous groups
GenePlot
pairwise comparison of protein homologs
Genomic BLAST
with complete and unfinished genomes



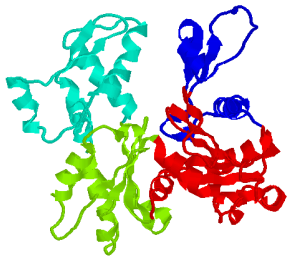
Assigning function to ORF

- in order to assign function, all predicted ORF's are translated to amino acid sequence and analysed by homology searches against sequence databases (usually Genbank)
- for each ORF there are three possible results -
- i) clear sequence homology indicating function
- ii) blocks of homology to defined functional motifs
- - these should be confirmed experimentally
- iii) no significant homology or homology to proteins of unknown function



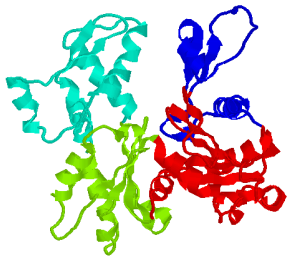
Lecture Outline

- Introduction
- Manual Curation
- Automatic Annotation
- Conclusions



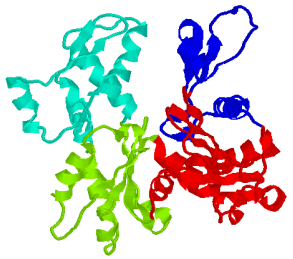
Challenges or Pitfalls

- First and last exons difficult to annotate because they contain UTRs.
- Smaller genes are not statistically significant so they are thrown out.
- Algorithms are trained with sequences from known genes which biases them against genes about which nothing is known.



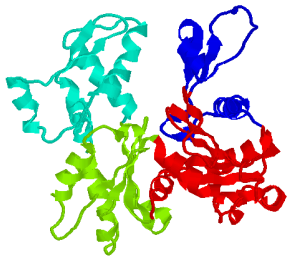
Conclusions

- Trust but verify
- Beware of gene prediction tools!
- Always use more than one gene prediction tool and more than one genome when possible.
- Active area of bioinformatics research, so be mindful of the new literature in this .



Readings

- http://www.genome.org/cgi/content/full/15/12/1777
- Play with **ORF Finder** http://www.ncbi.nlm.nih.gov/gorf/gorf.html
- Study **Microbial Genomes Resources** http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html

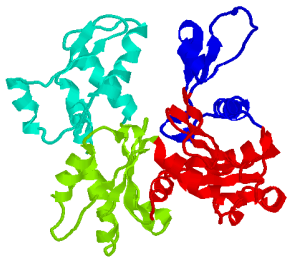


Seminar

Teaching through Undergraduate Research in Microbial Genome Annotation

Cheryl Kerfeld, Ph.D.
Seth Axen
Education Program
DOE Joint Genome Institute
Walnut Creek, California

Thursday, Sept 24, 2009
1:00—2:30 pm
Monsanto Auditorium
Bond Life Sciences Center



Acknowledgments

This file is for the educational purpose only. Some materials (including pictures and text) were taken from the Internet at the public domain.