The Role of Bioinformatics in Gene Discovery

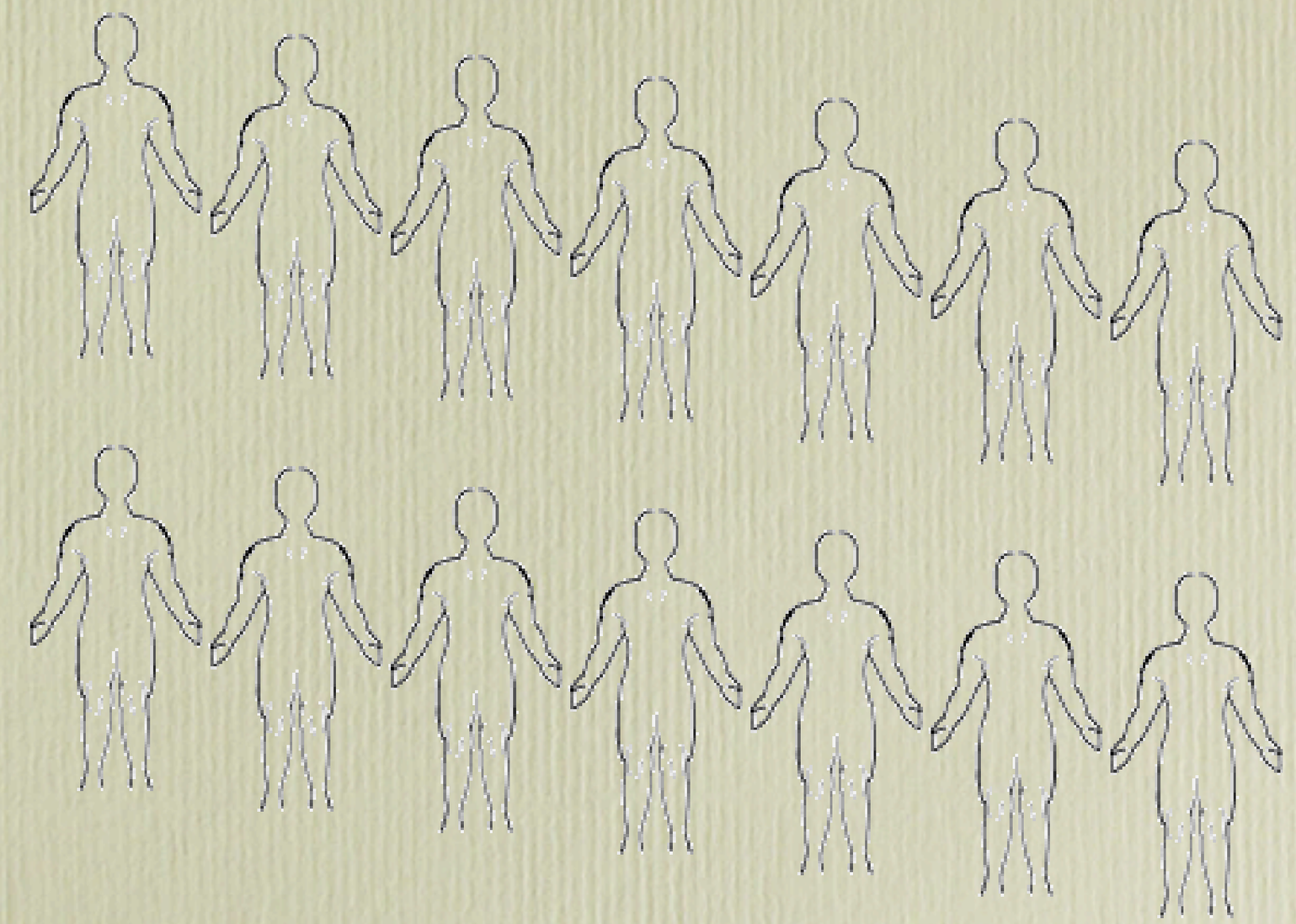
James J. Vincent

Laboratory of Molecular Biology
National Cancer Institute
National Institutes of Health

Overview

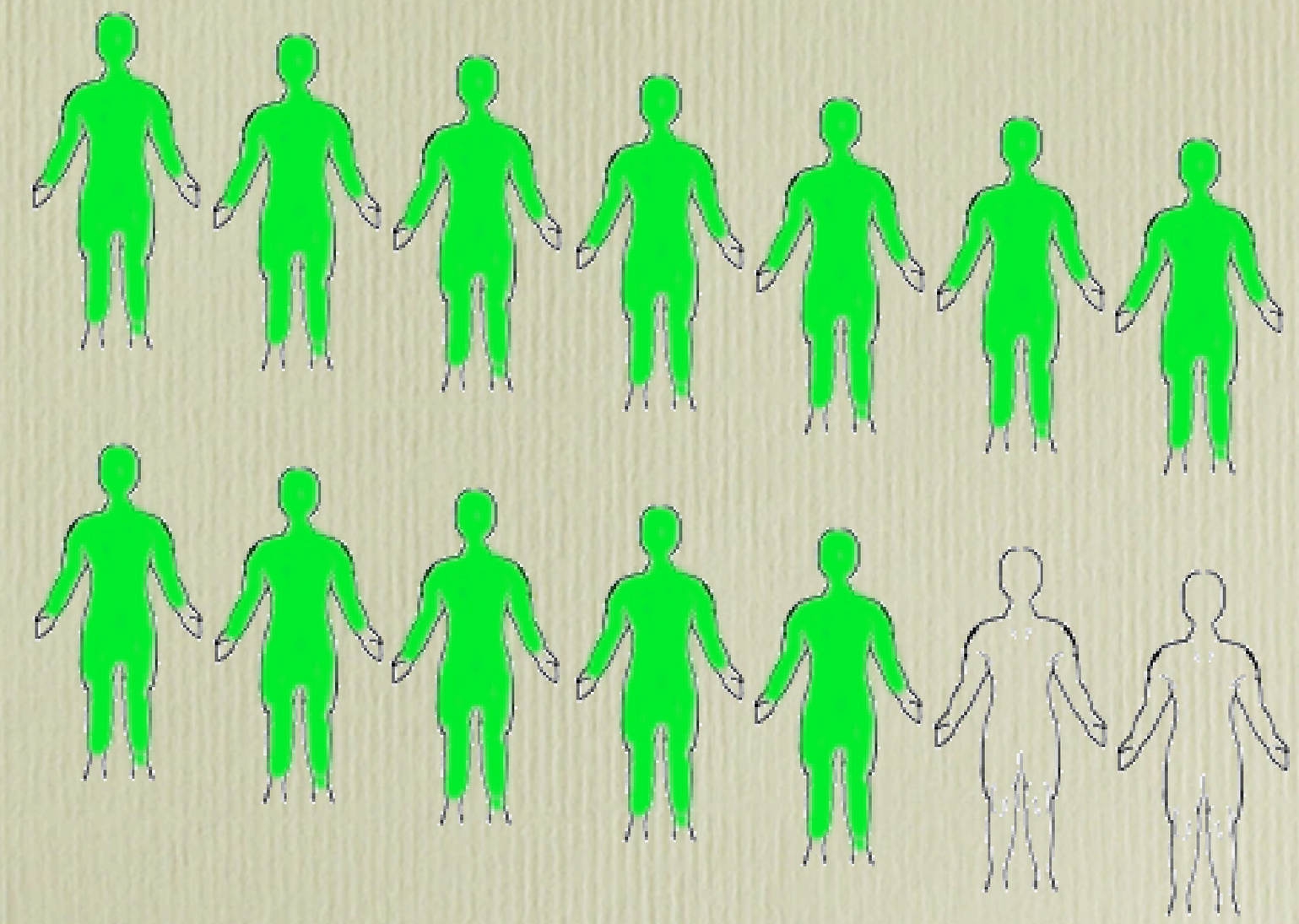
- Why we look for new genes
- How we find them with a computer
- We found what we were looking for

We make immunotoxins ...



14 Patients with Hairy Cell Leukemia
Phase 1 Clinical Trial

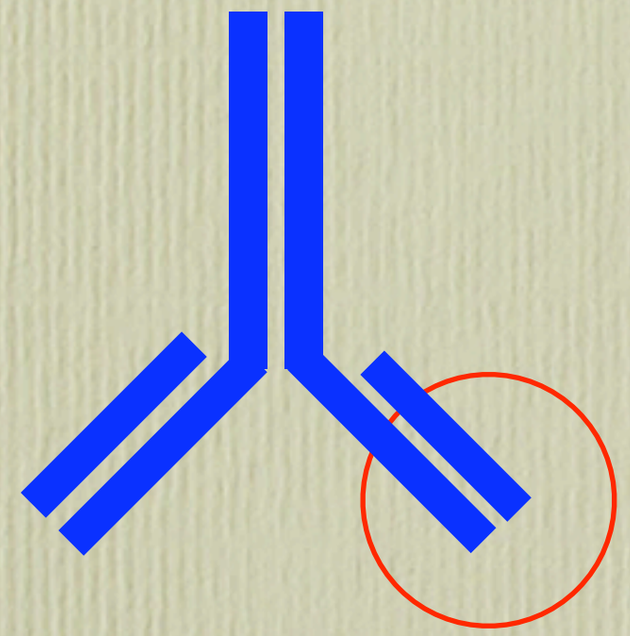
… and they work well.



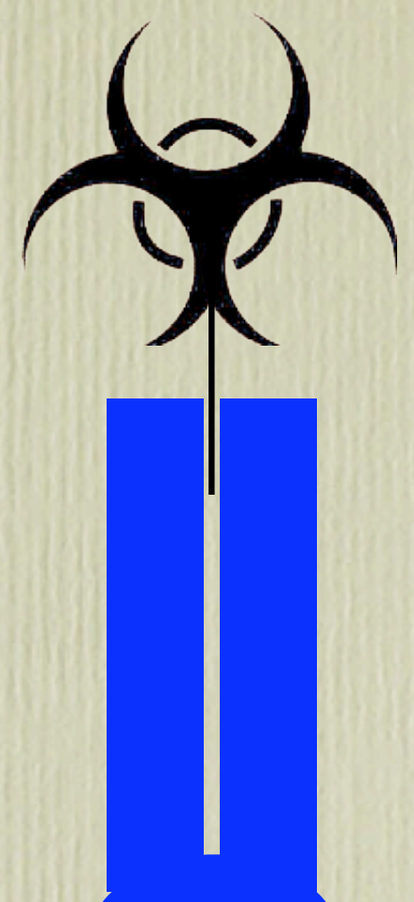
86% Complete Remission

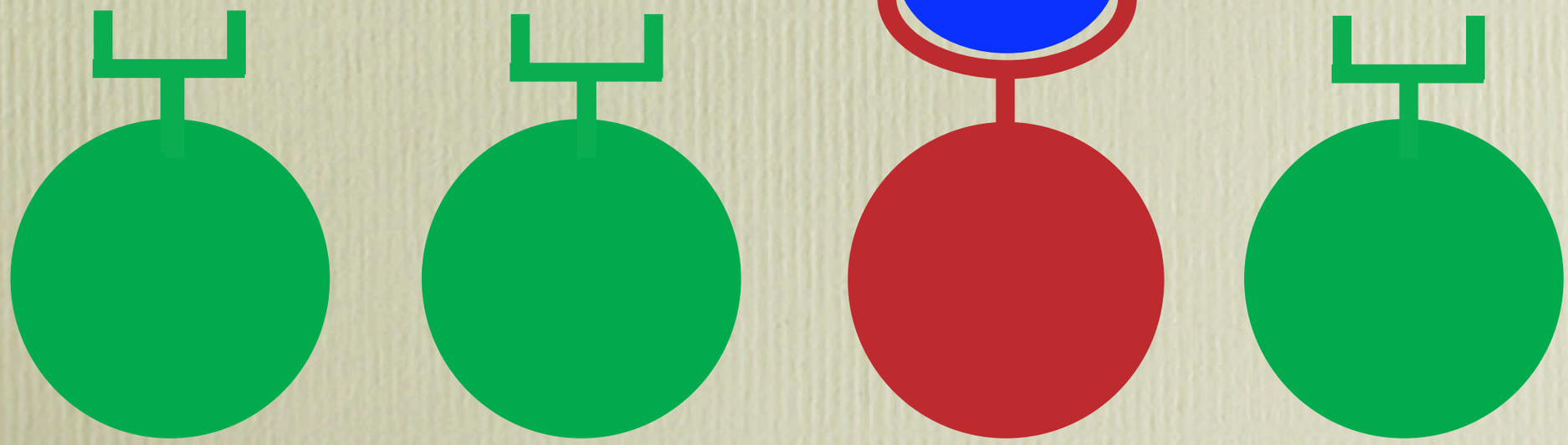
How We Make Immunotoxins

antibody

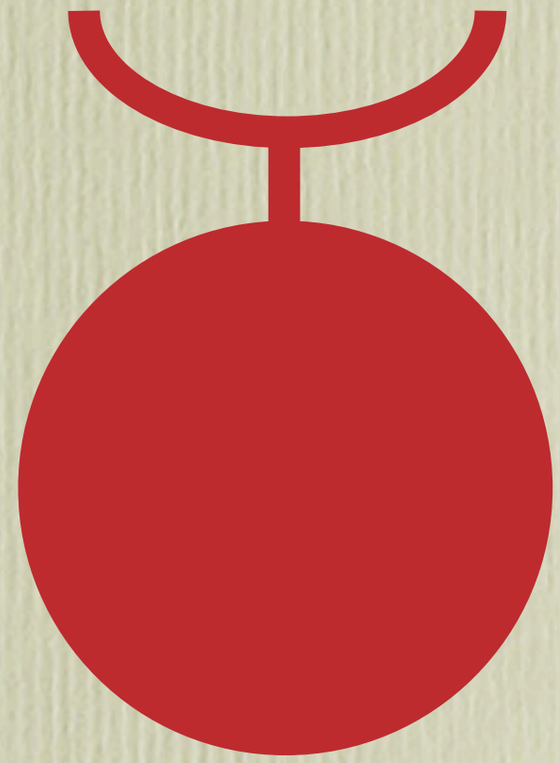


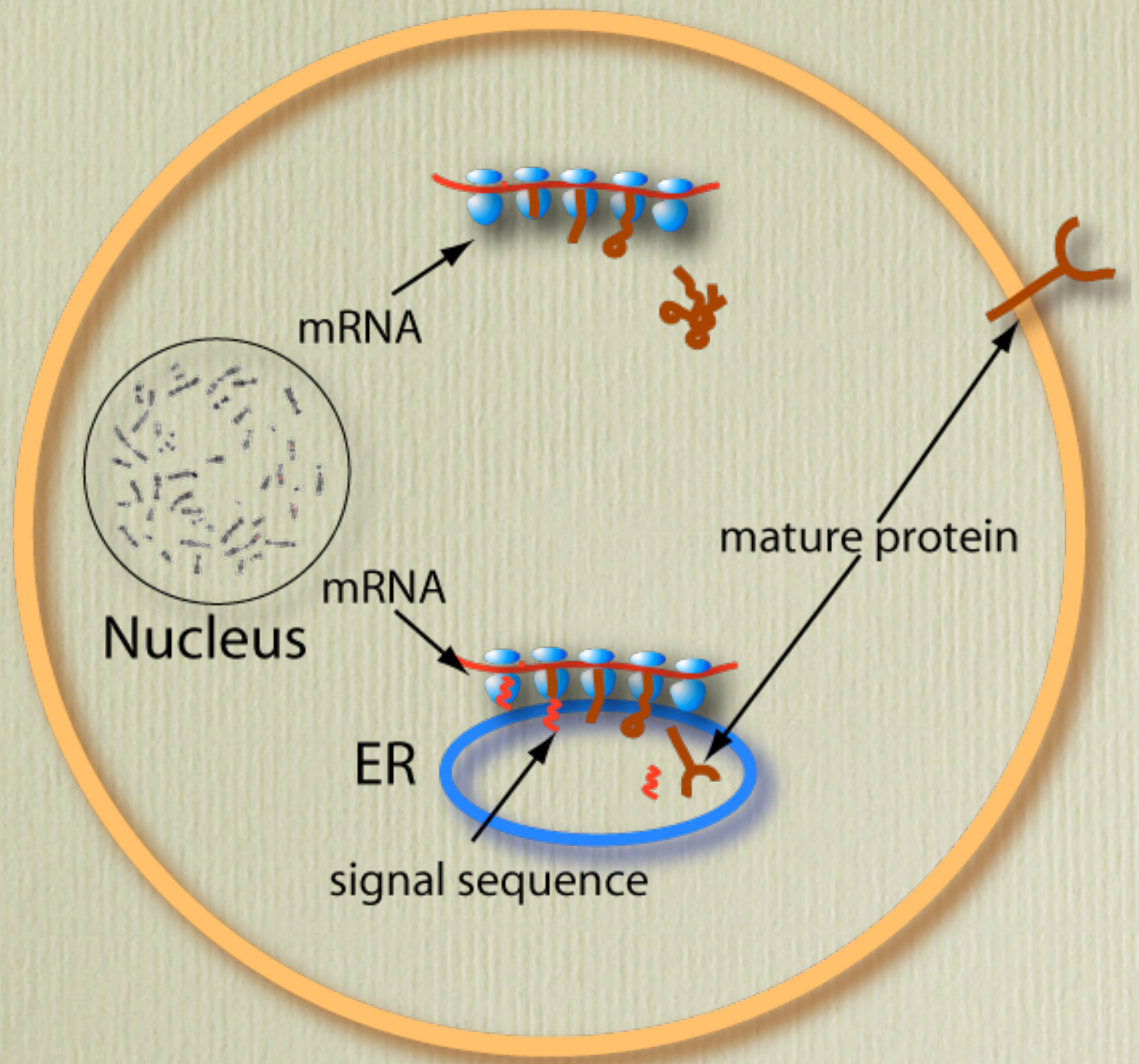
toxin





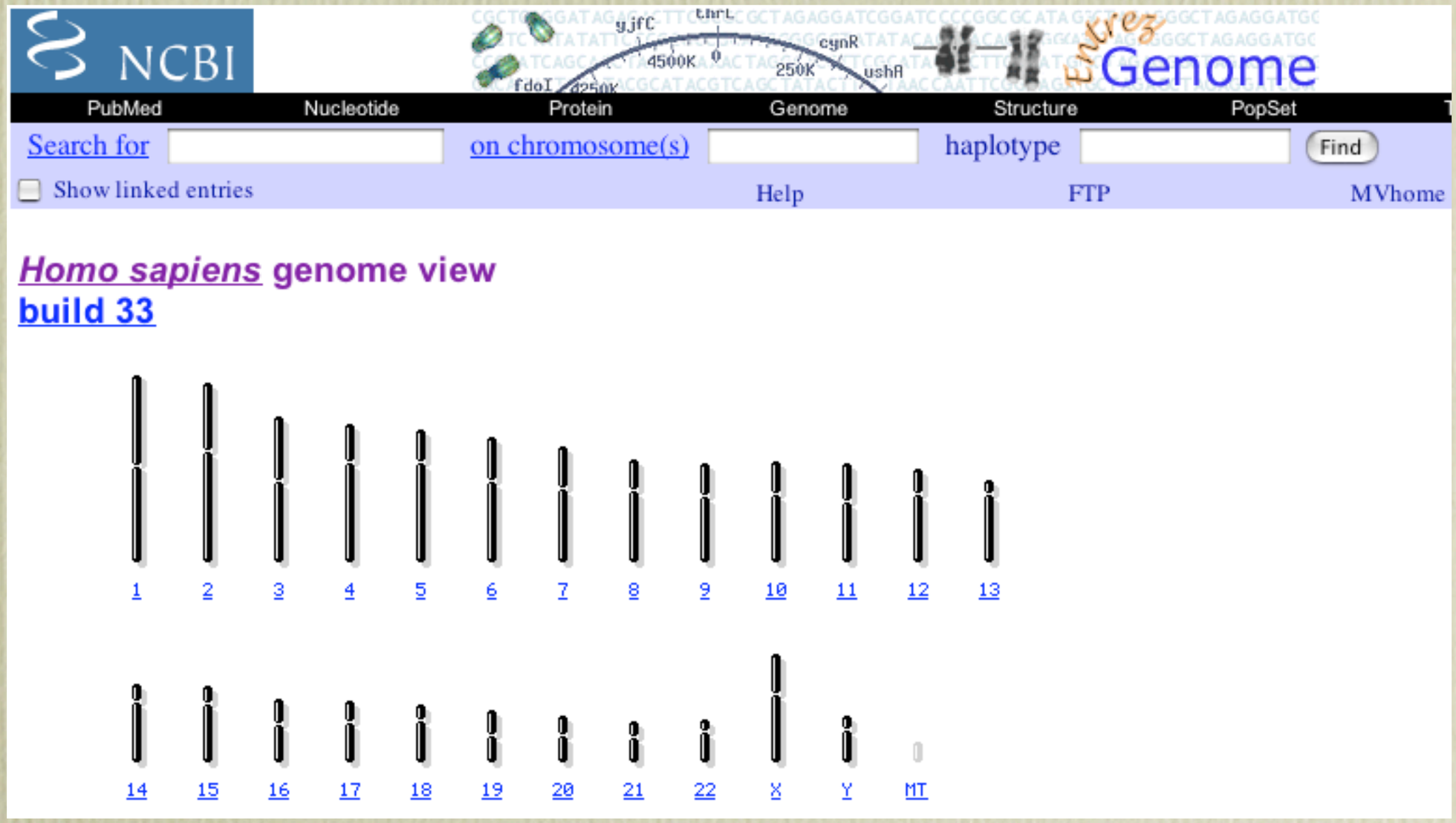
Find the Receptor

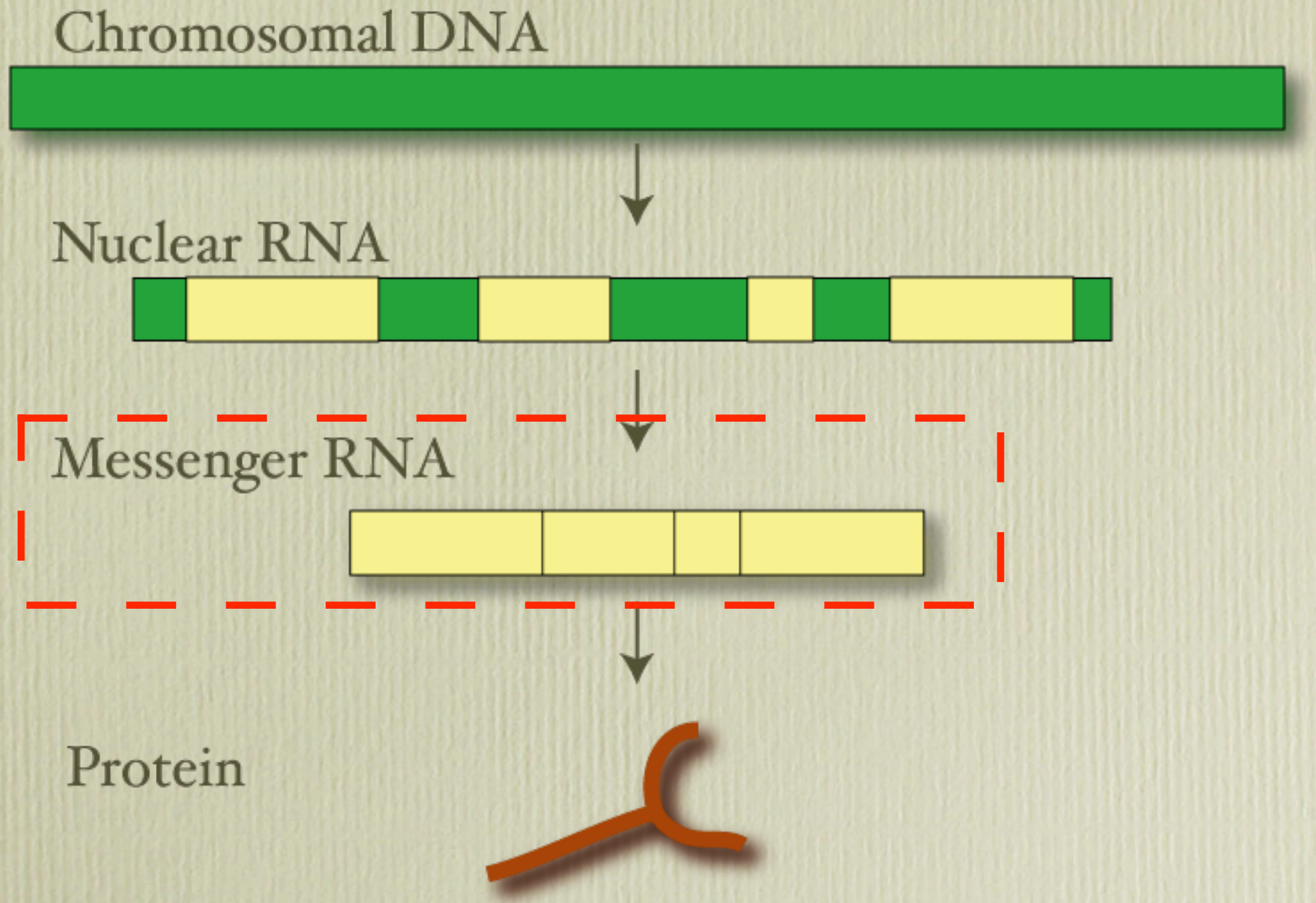



mRNA
Nucleus
mRNA
mature protein
ER
signal sequence

The Human Genome

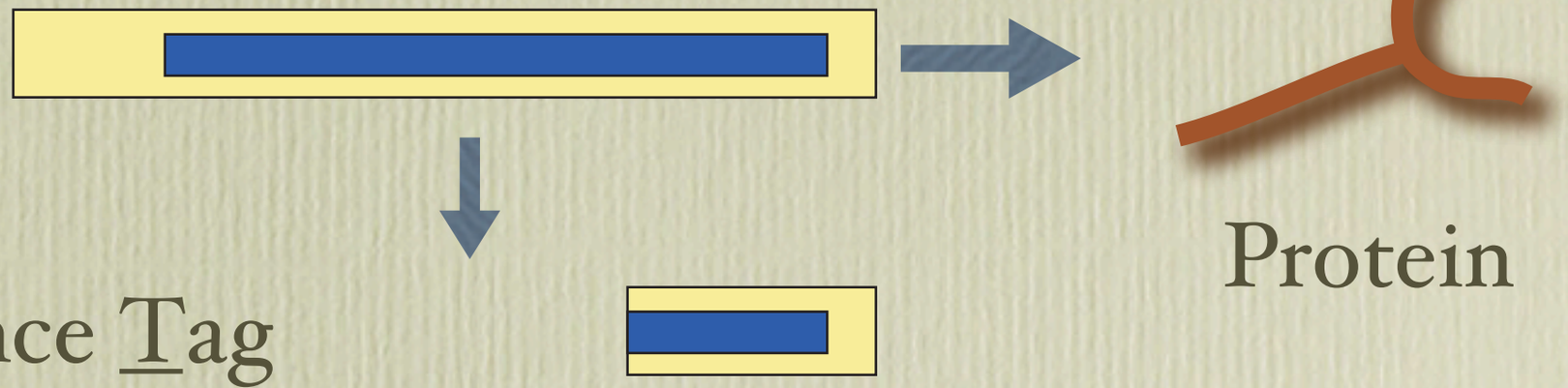
Mostly Finished




Chromosomal DNA
Nuclear RNA
Messenger RNA
Protein

Expressed Sequence Tags

Messenger RNA



Expressed Sequence Tag

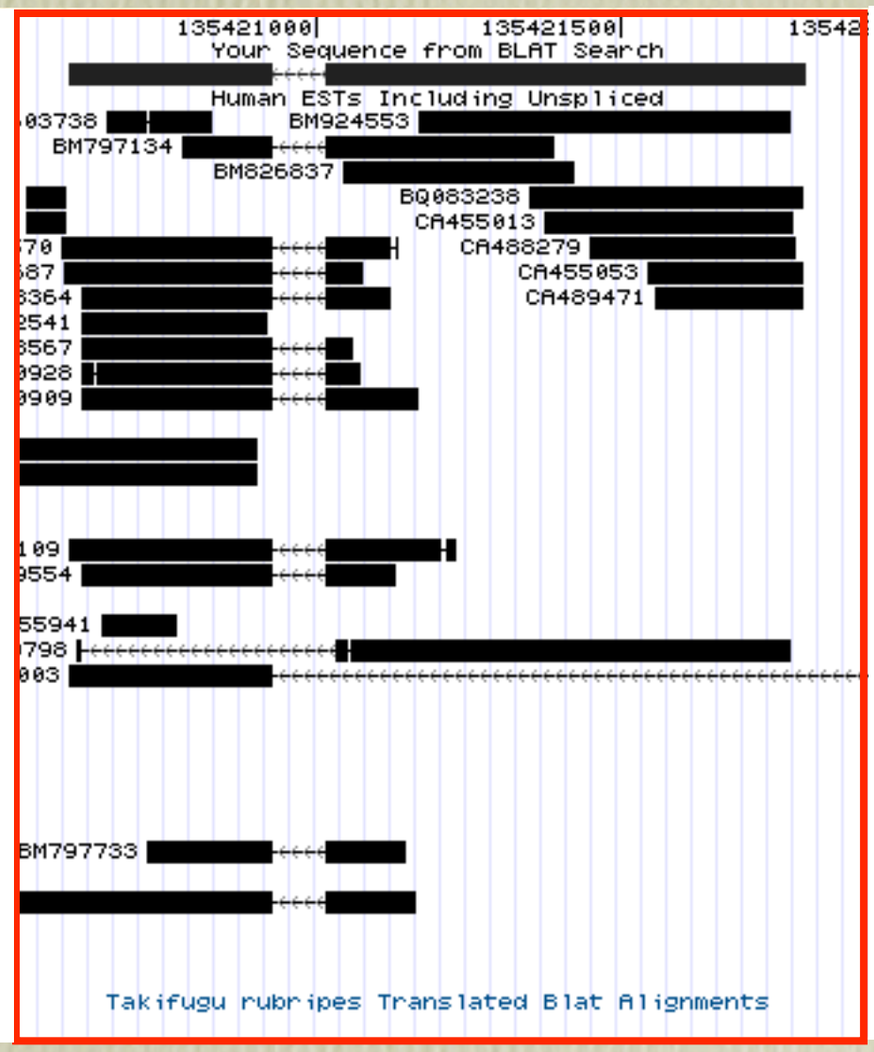
EST

dbEST

- 5 million ESTs
- 8000 libraries
- 1500 tissue types

Cluster together into genes

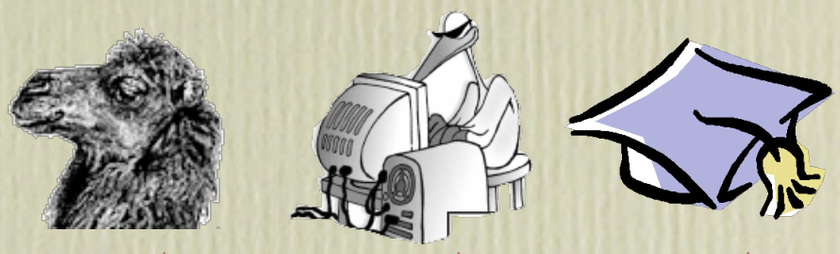
Clustering Results



Sequence Analysis Pipeline

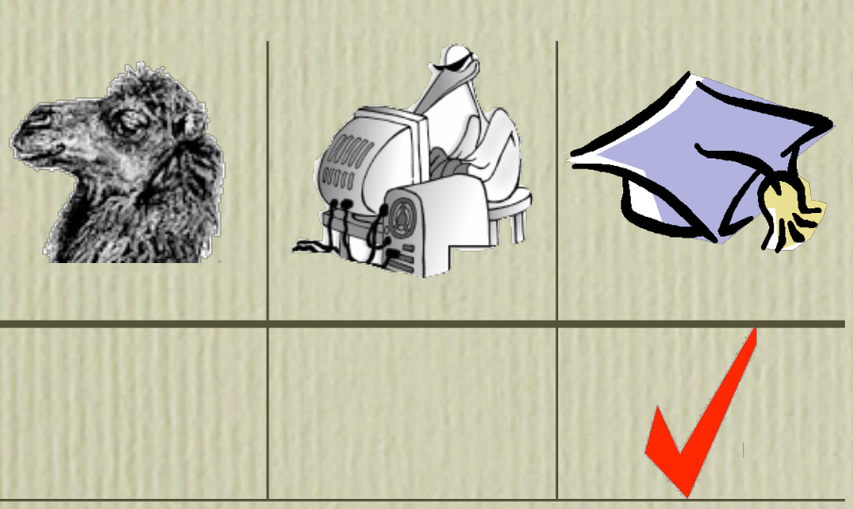
Build Clone Contig → Cluster → Align to Genome → Find ORF TM region → Display in Browser

The Value of a C.S. Degree



Script web tools	✓	✓	✓
Large relational databases	✓	✓	✓
Reusable software objects		✓	✓
Software for other people			✓
Software for other software			✓
Robust, documented software			✓

System Administration
Backup
Networking
Queueing System



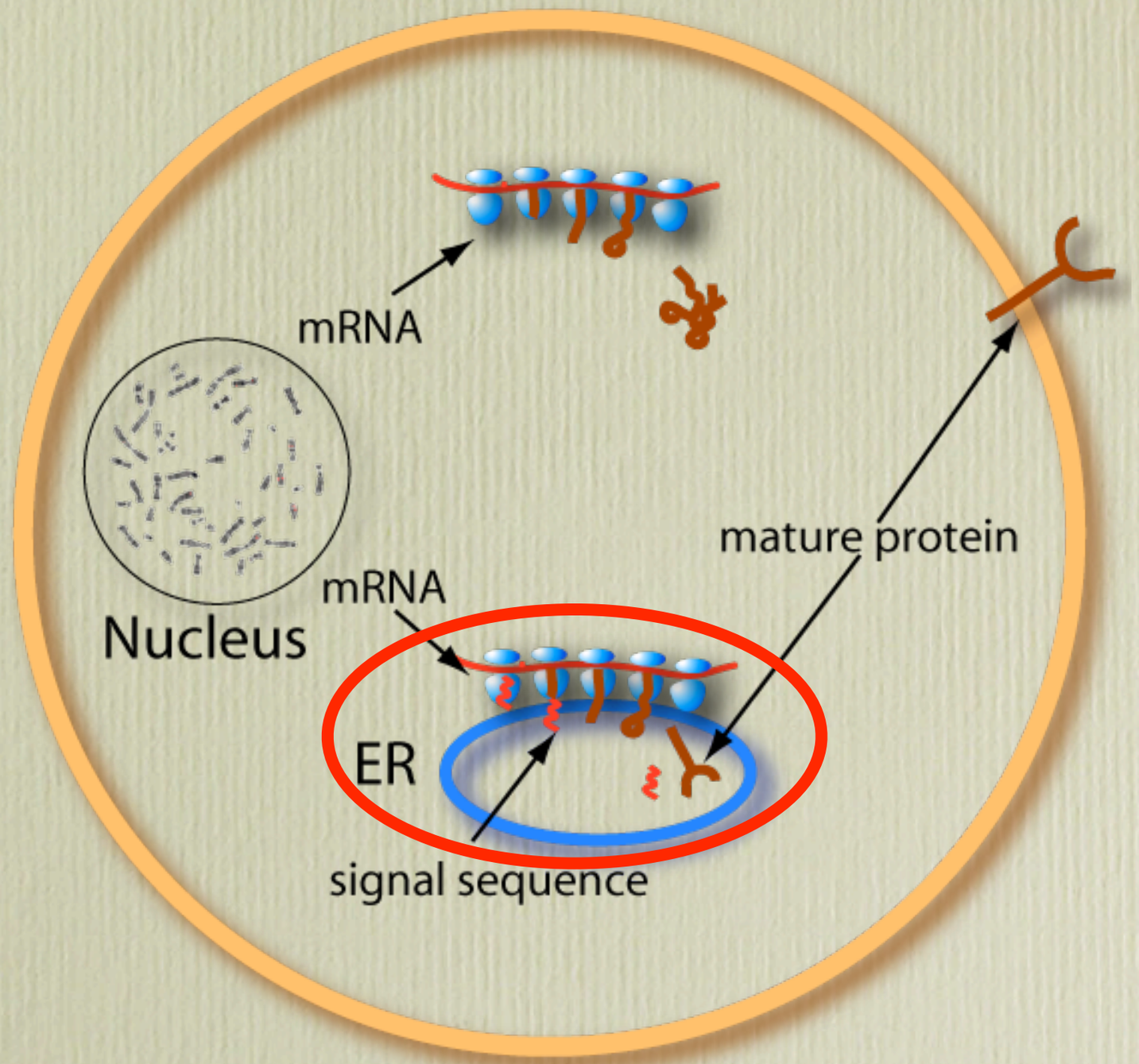
Breast Cancer Genes from dbEST

0

Kristi Egland, PhD

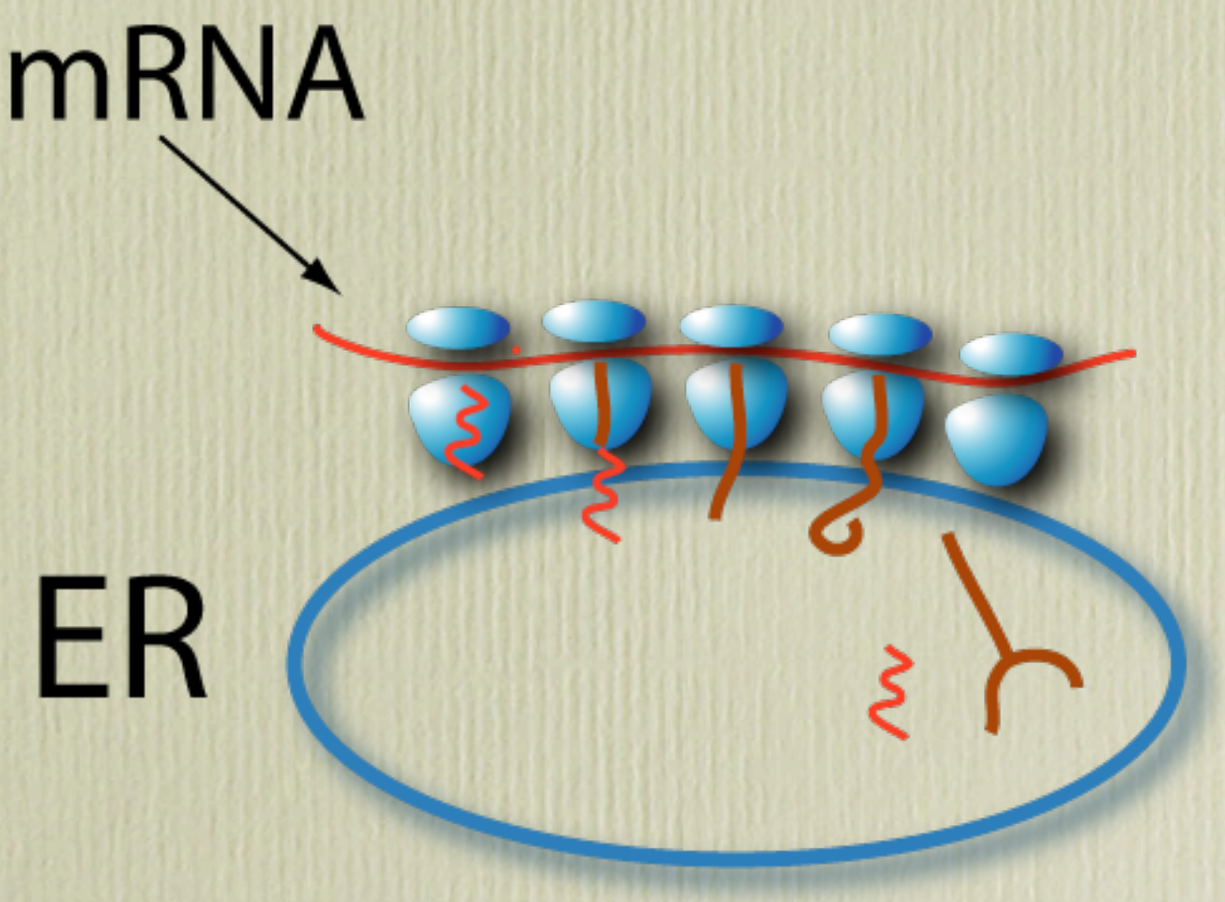
Molecular Biologist

We Make Our Own ESTs



The MAPcL

Membrane Associated Polyribosomal cDNA Library

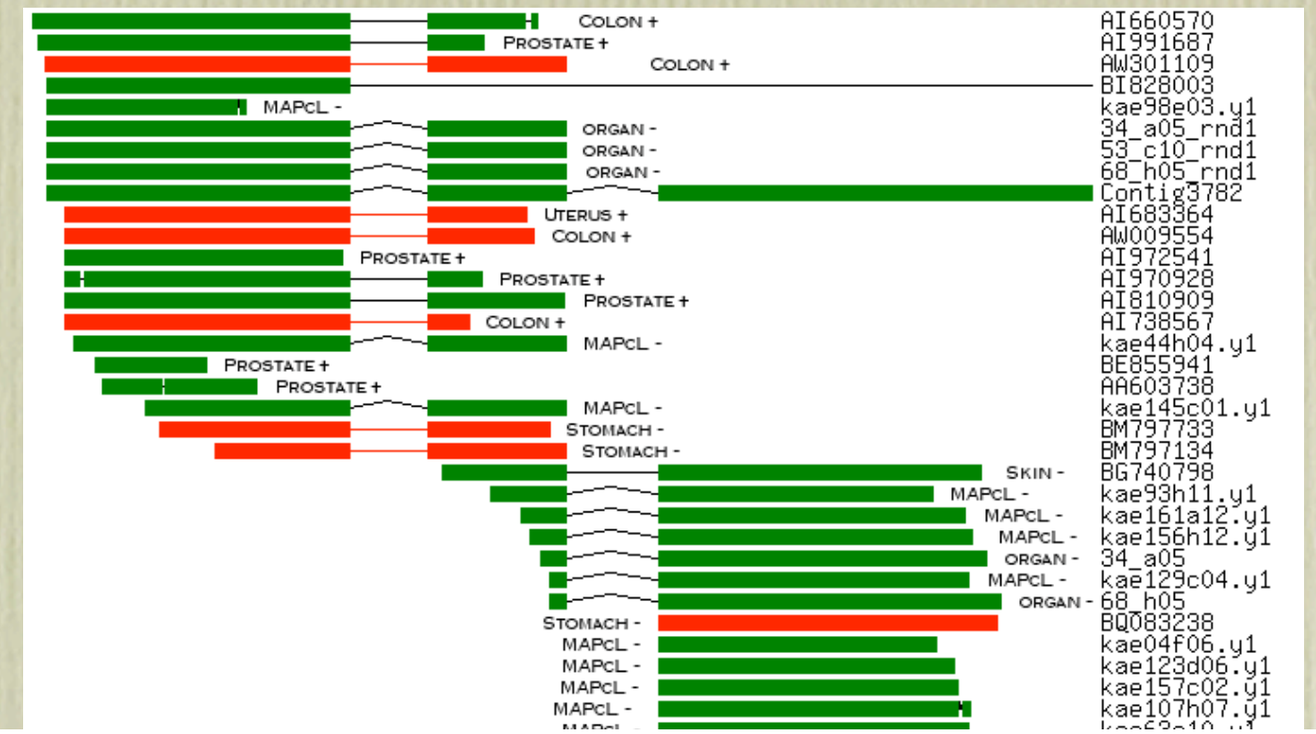


Two years + $$$$$ → 25,000 sequences

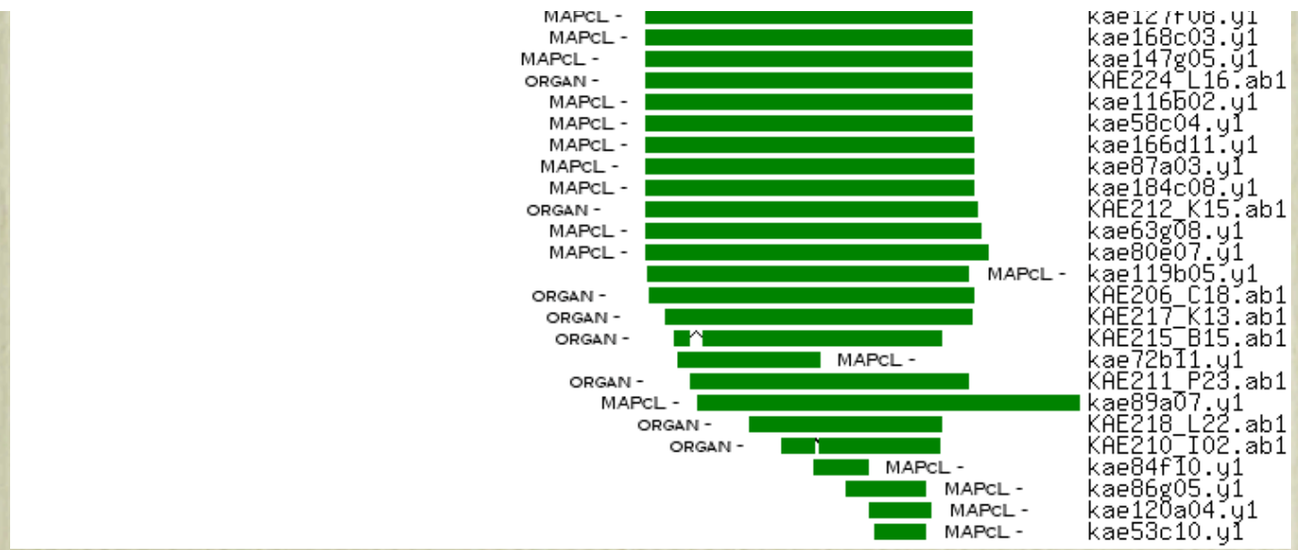
The MAPcL is Above Average

- Enriched for membrane protein genes
- Genes from critical organs removed
- Sequenced from the 5’ end

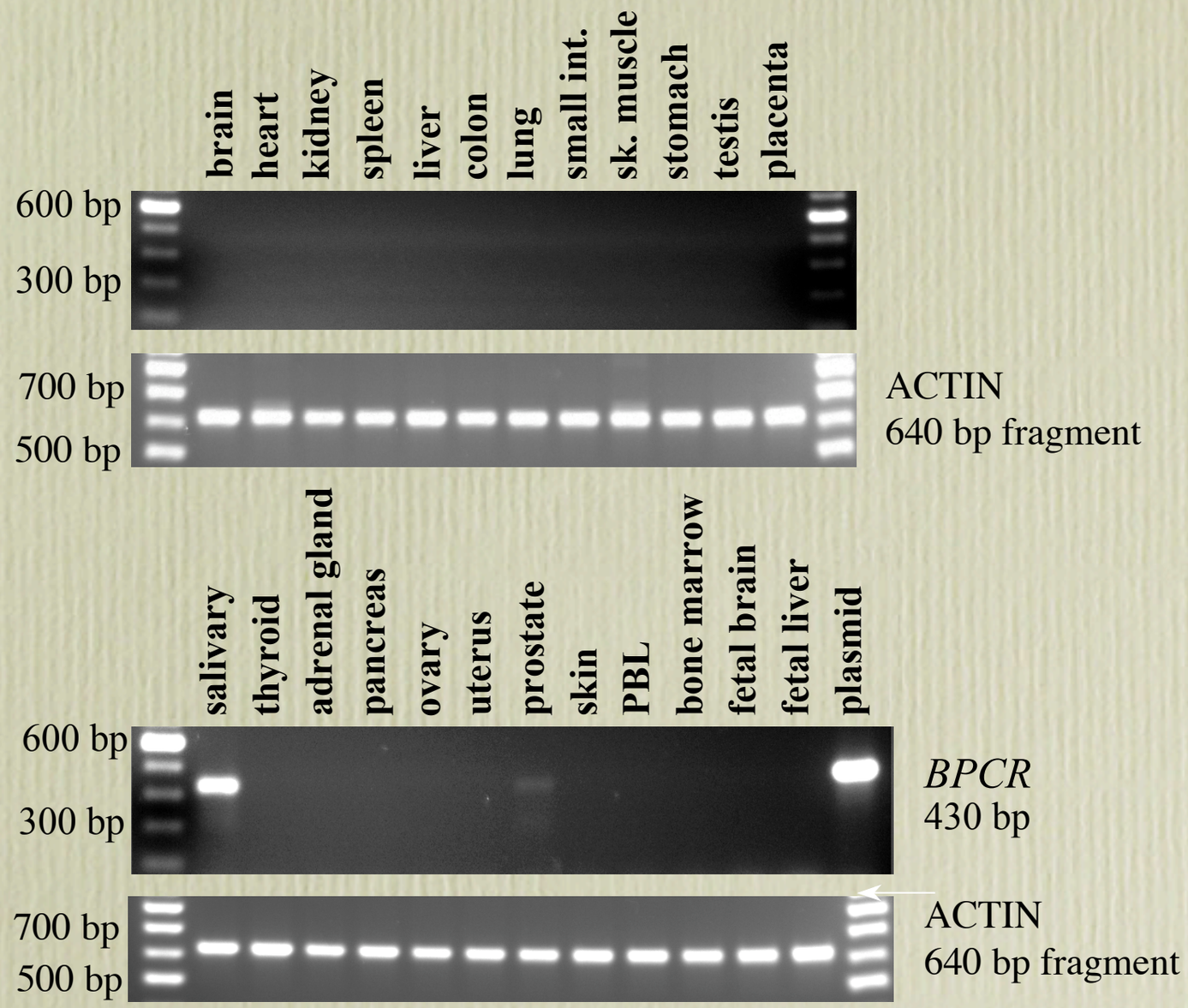
Breast Cancer Genes from MAPcL



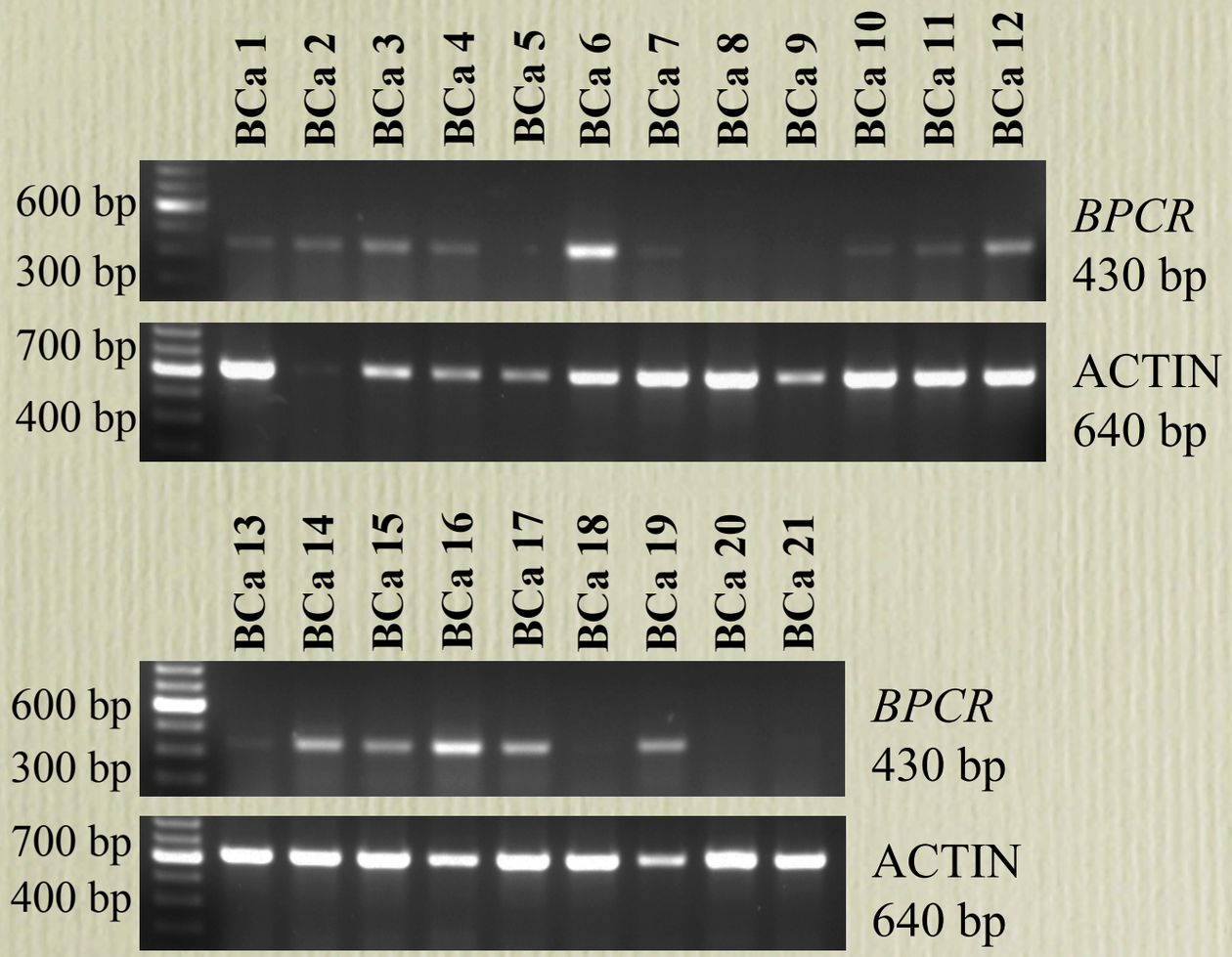
Breast Prostate and Salivary Gland Receptor - *BPSR*



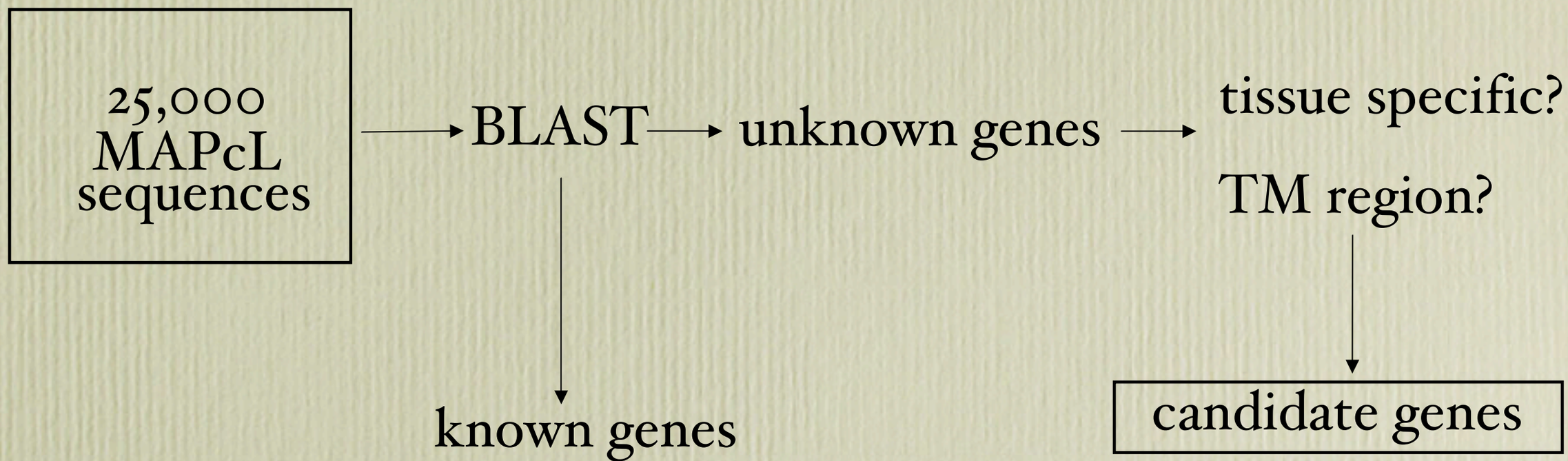
RT-PCR Analysis of *BPCR* in Normal Tissues



RT-PCR Analysis of *BPCR* Expression in Breast Cancers



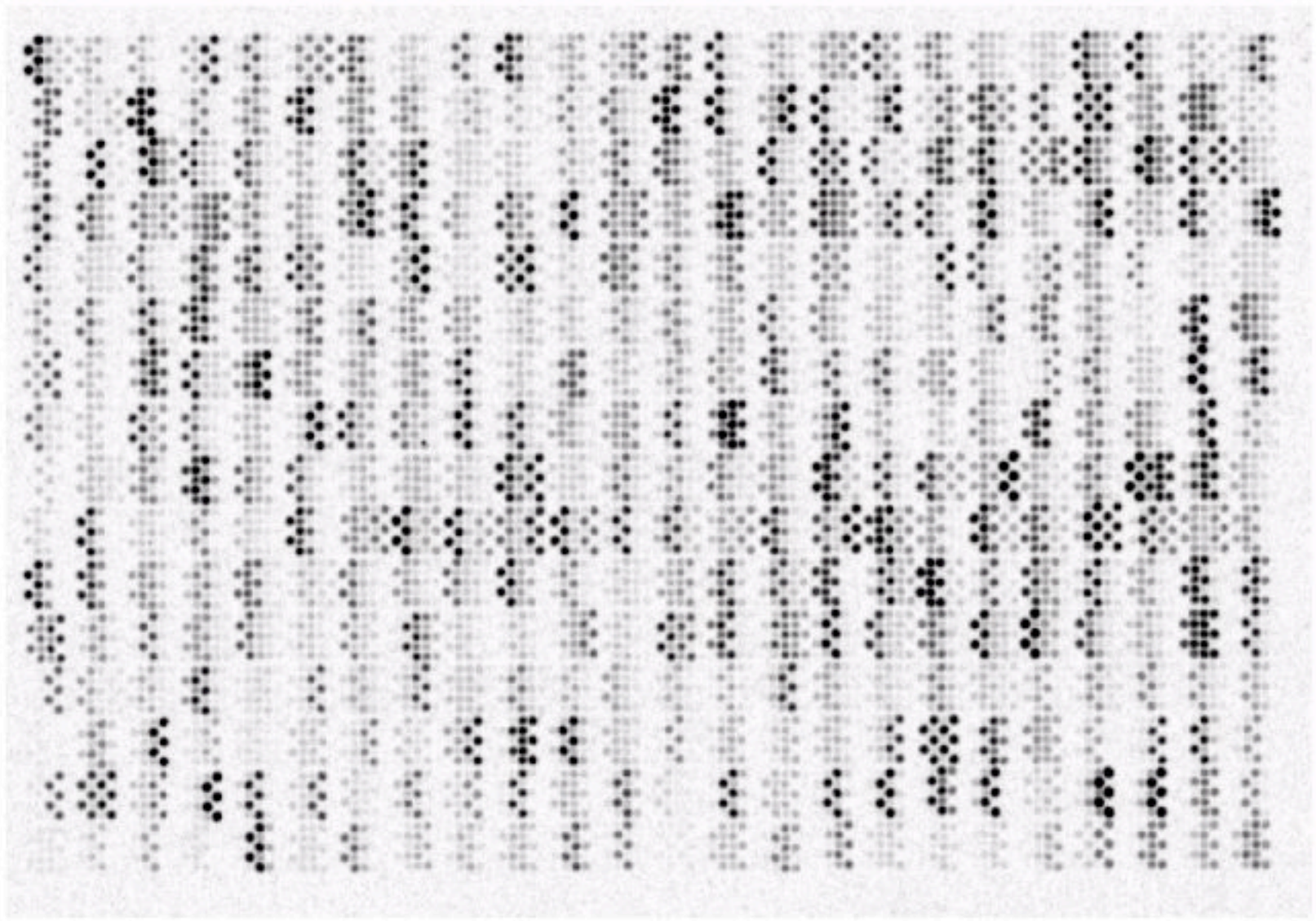
Analysis of MAPcL Sequences
How do we find new targets?



1400 unknown genes selected for array analysis

All at once...

1400 genes on one array



Sample 1

Summary

- We found what we were looking for
- High quality software plays a critical role
- C.S. degree directly relevant to life sciences research

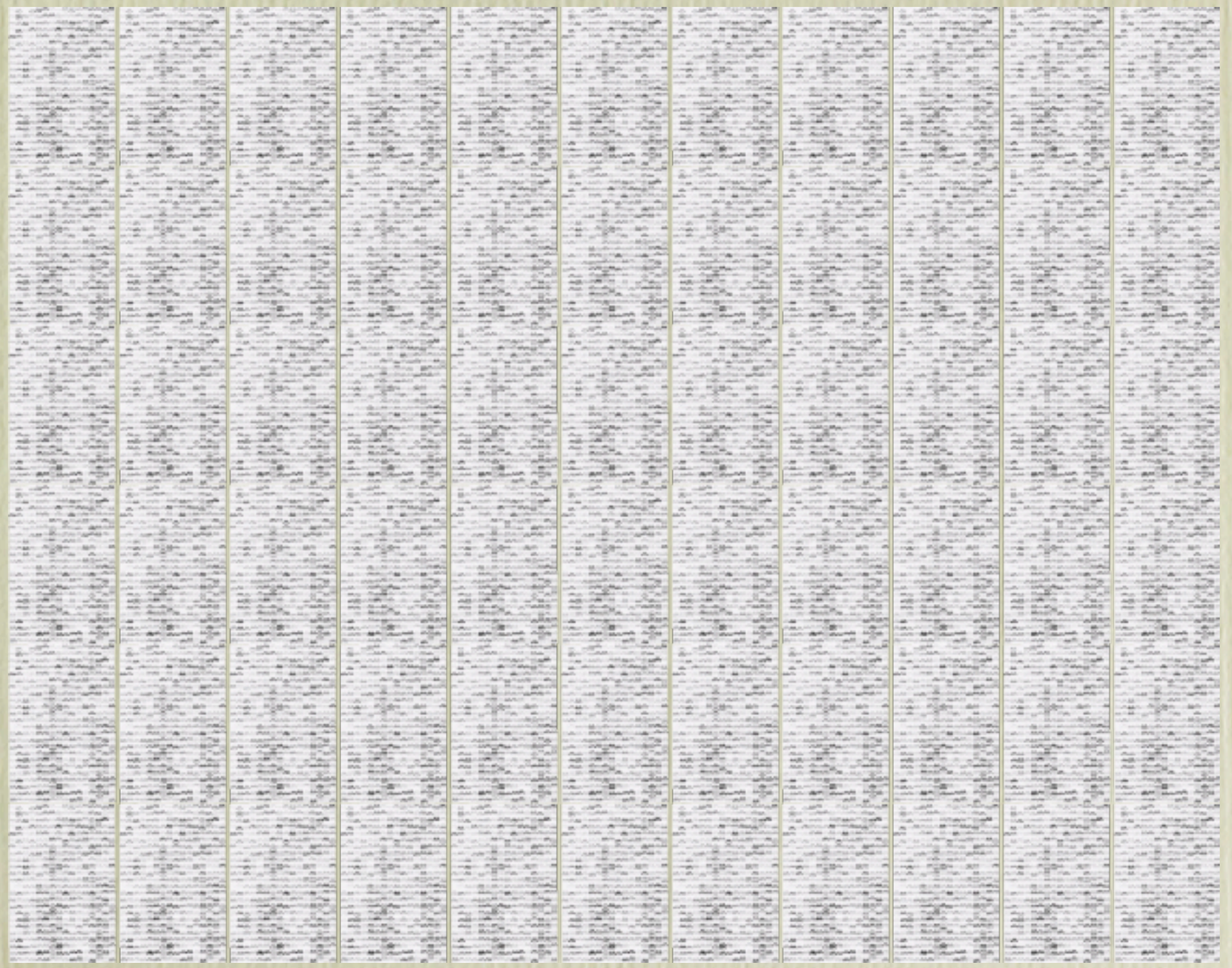
Kristi Egland
Ira Pastan
Bob Strausberg
Kevin Becker

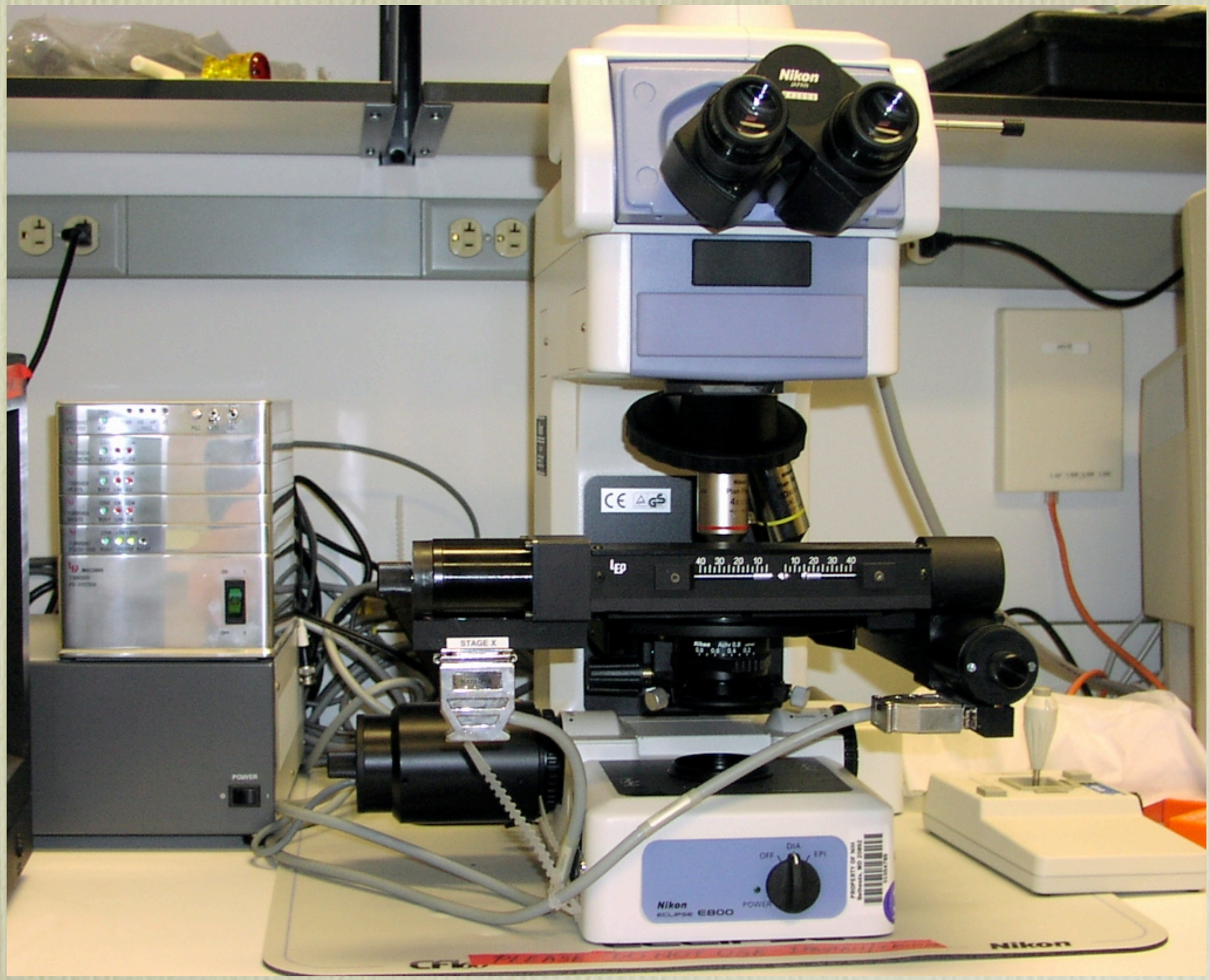
NATIONAL® CANCER INSTITUTE

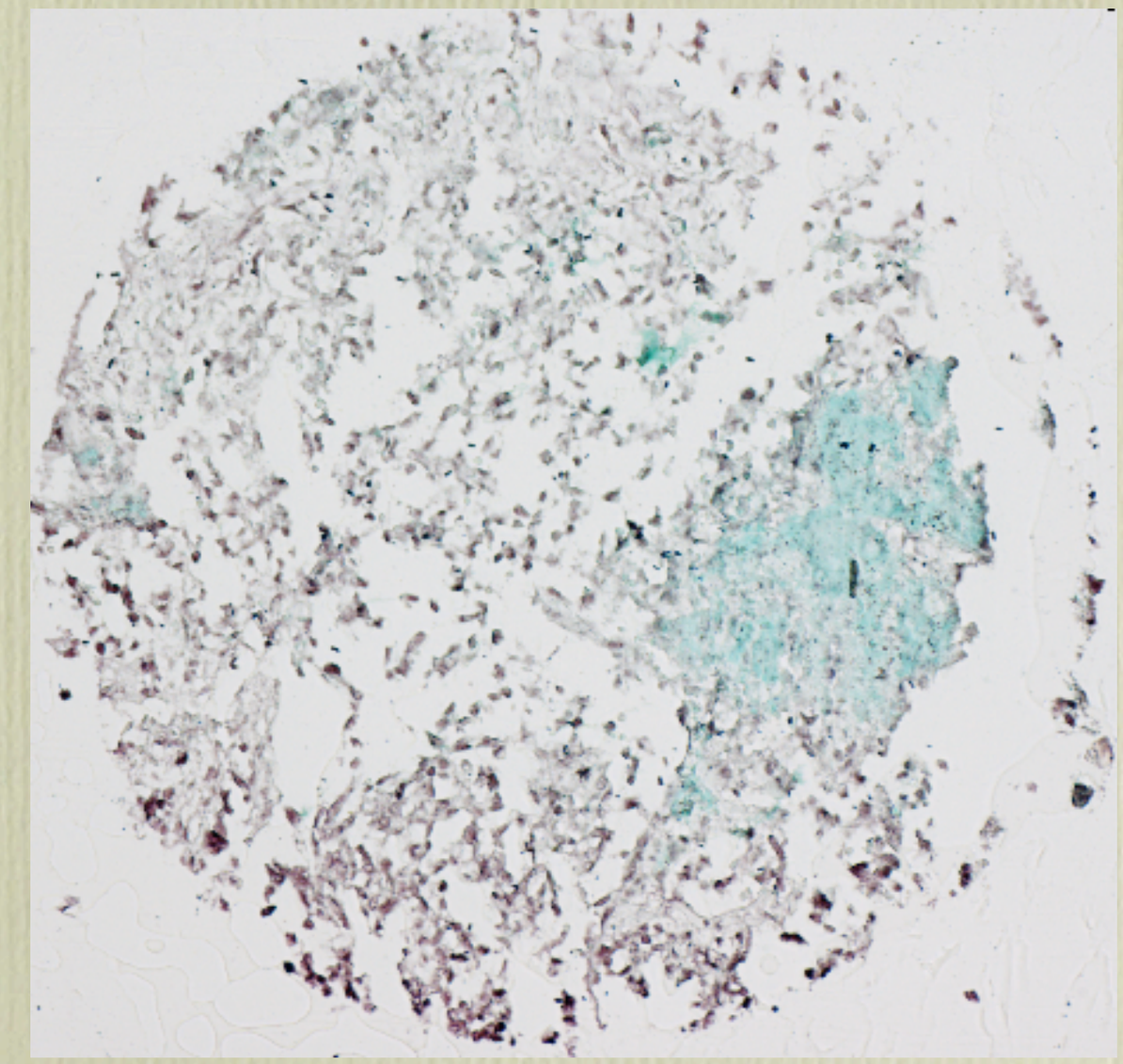
NATIONAL INSTITUTES OF HEALTH

Too Much

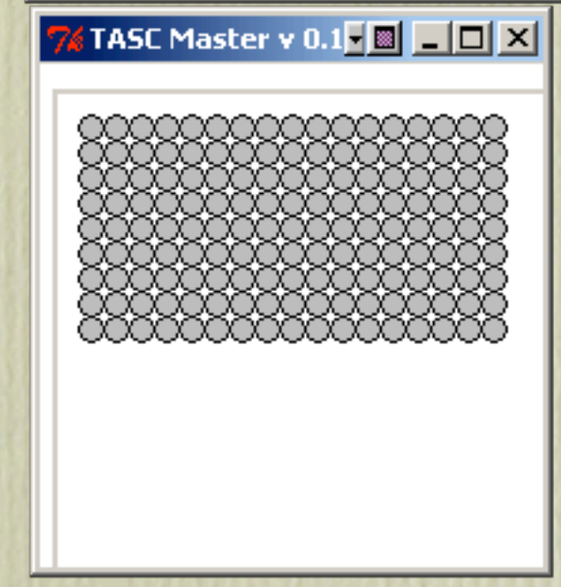
134 arrays x 10,000 spots = 1.3M data points




Nikon
STAGE X
POWER
OFF
DIA
EPI
Nikon
ECLIPSE E800
POWER
Nikon




TASC Master Control Panel
Slide Template
Setup
Pics Directory
Scan
Quit


TASC Master v 0.1

- RefSeq array - expression to pathway
- dbEST zScore compare to RefSeq array
- Expression browser - all info
- Array experiment interface for public databases
- Lit search tool hooked to othe rdatabases
- Promoter sequence extract, analyze, compare