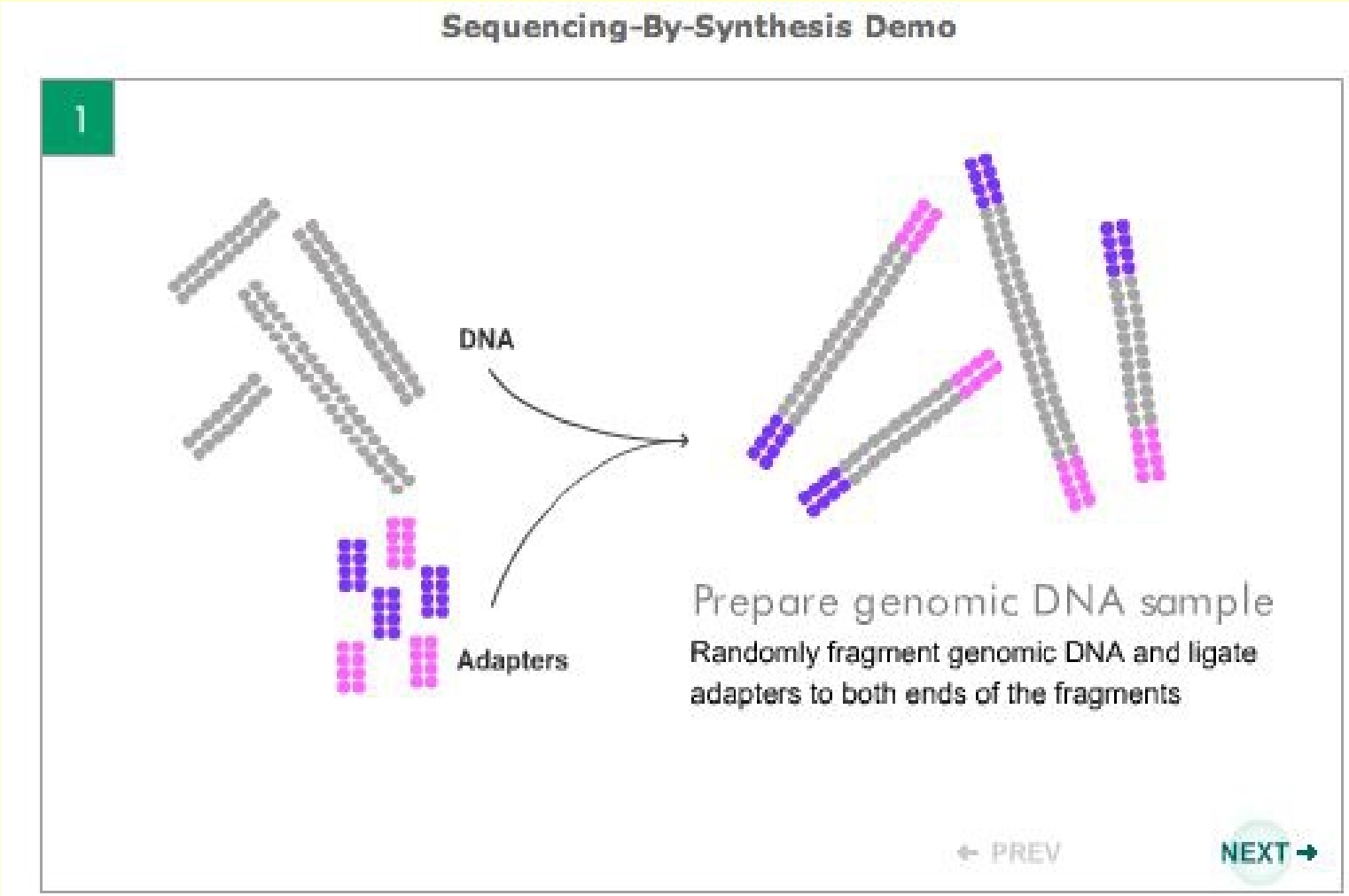
Next Generation Sequencing and Human Genome Databases

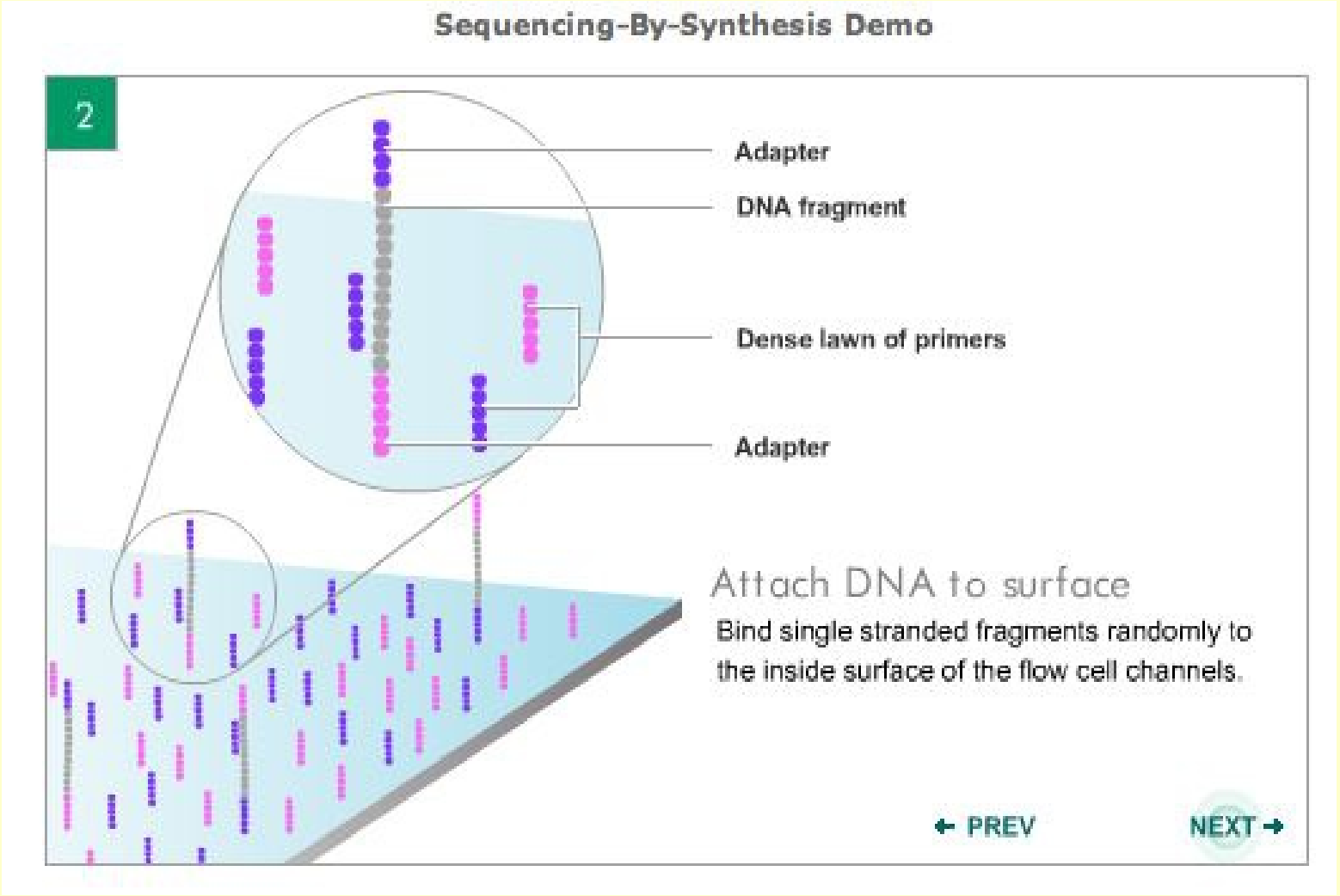
Genomics, Bioinformatics & Medicine
http://biochem158.stanford.edu/

Doug Brutlag
Professor Emeritus of Biochemistry & Medicine
Stanford University School of Medicine

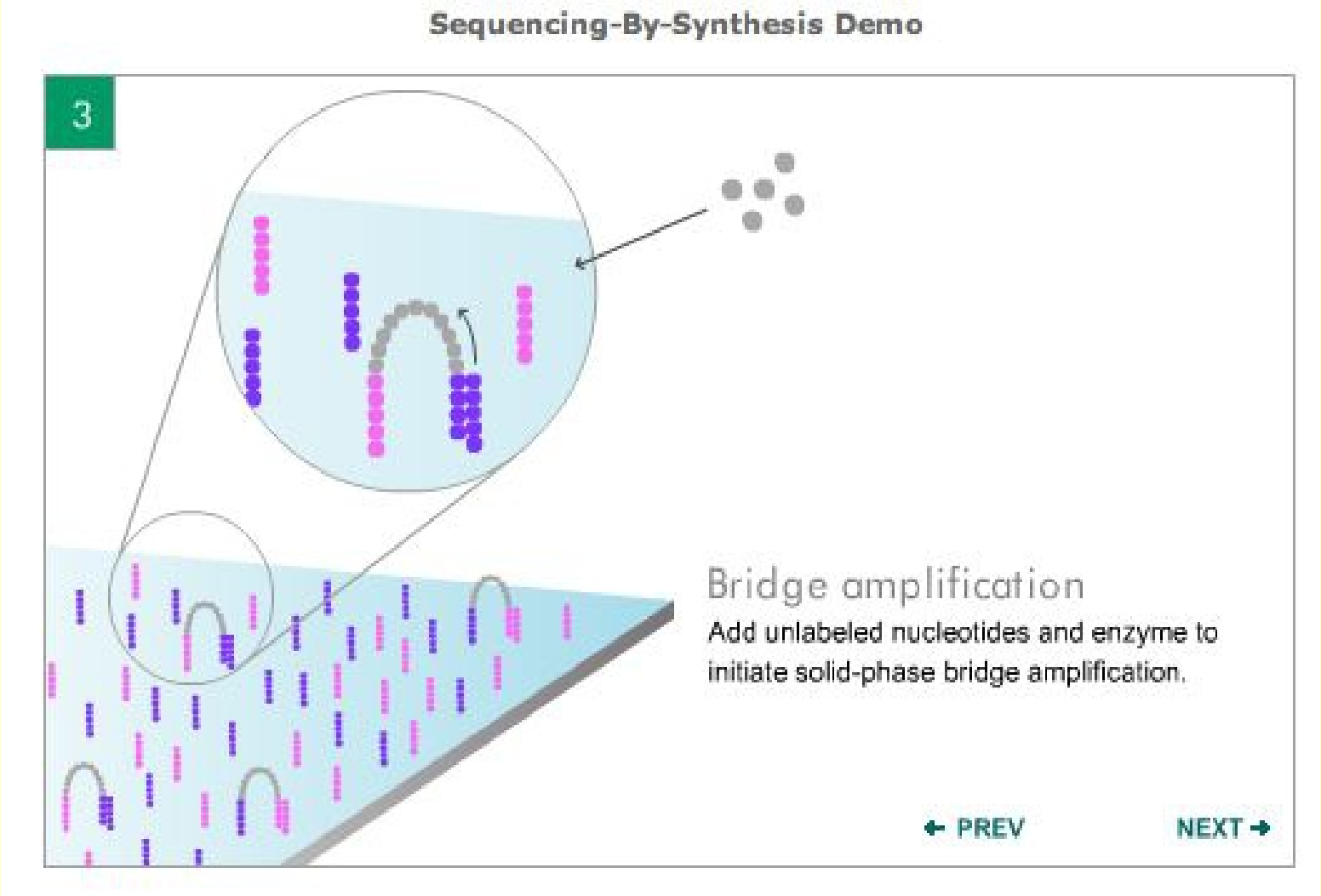
Illumina Solexa Sequencing Technology



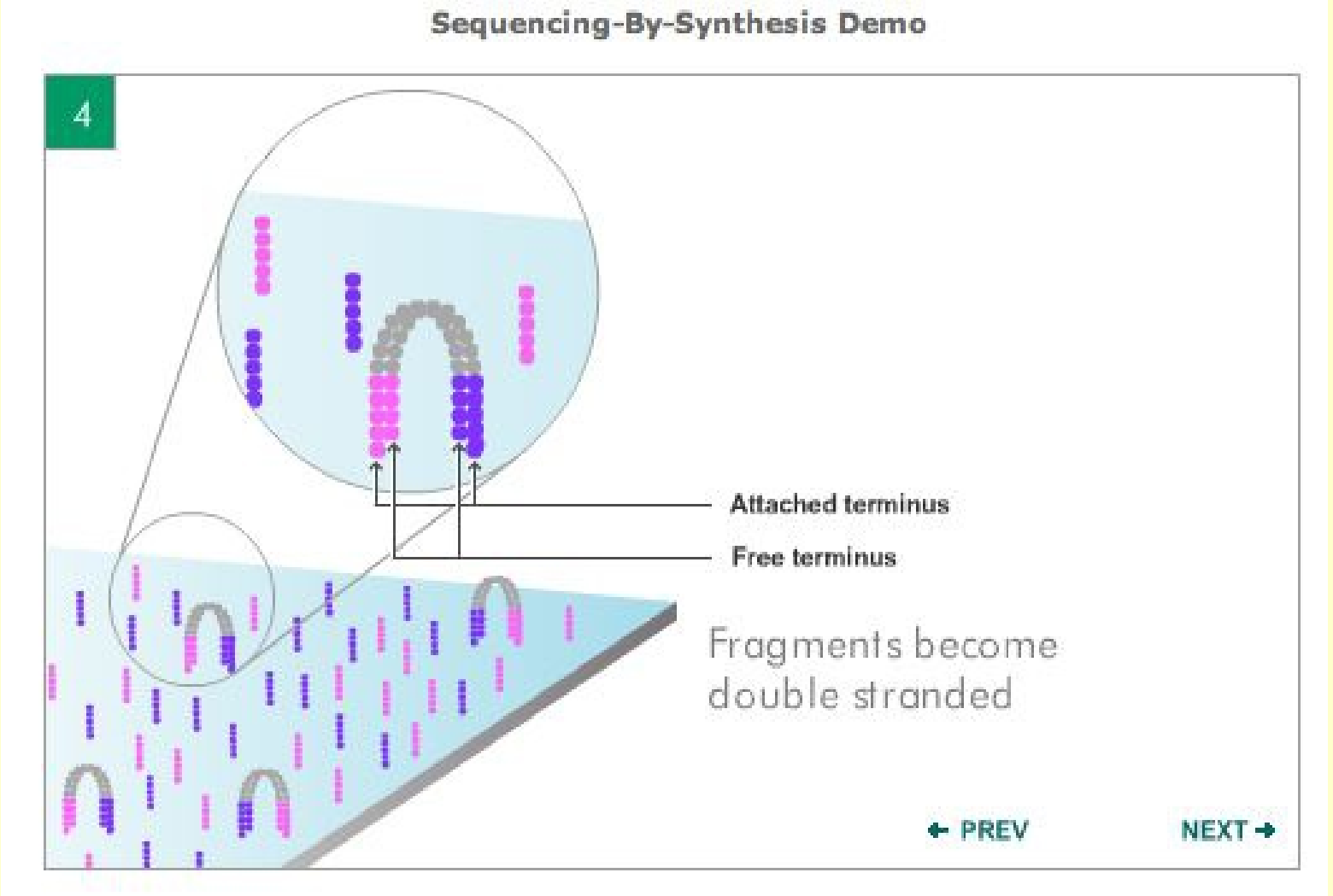
Illumina Solexa Sequencing Technology



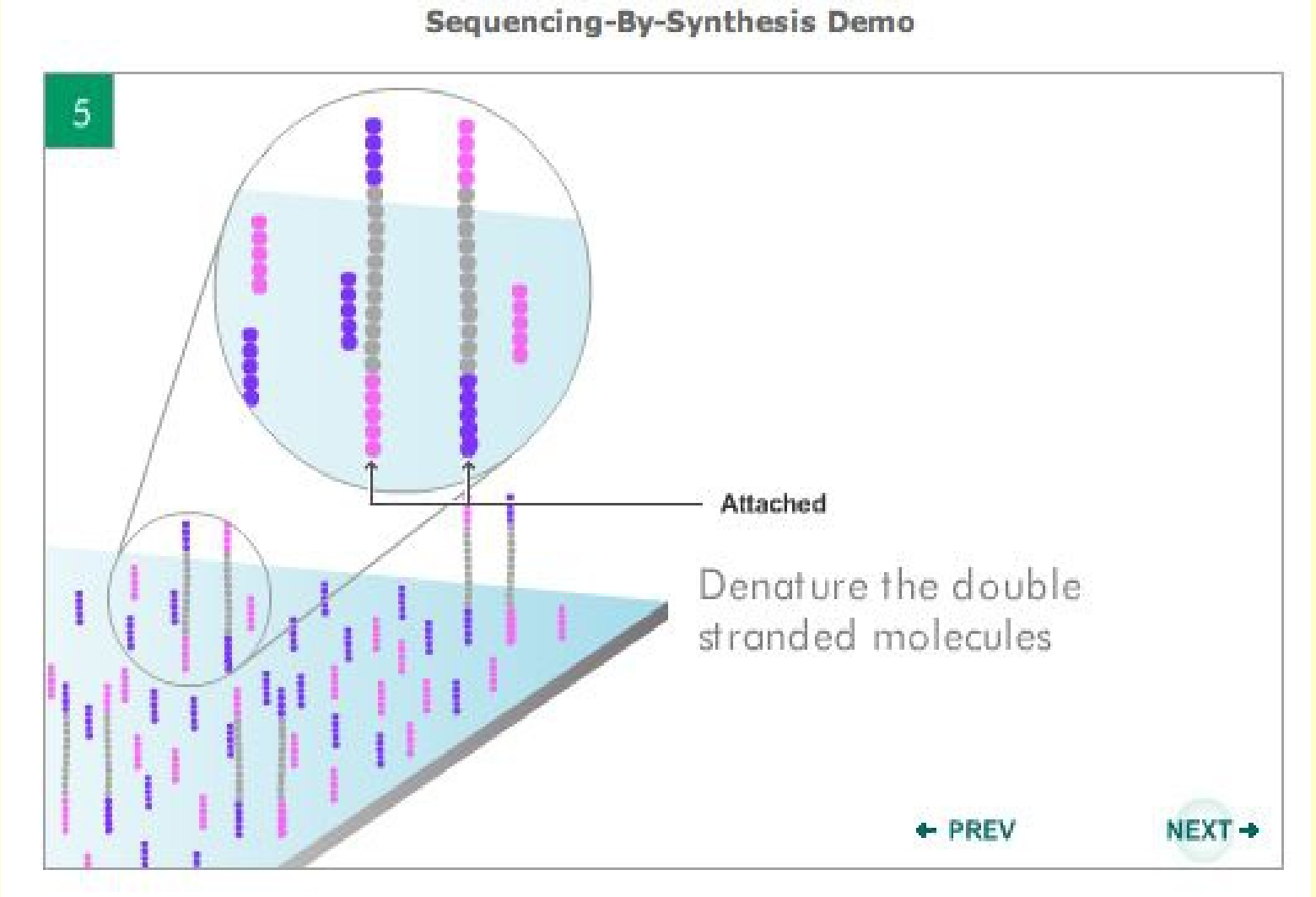
Illumina Solexa Sequencing Technology



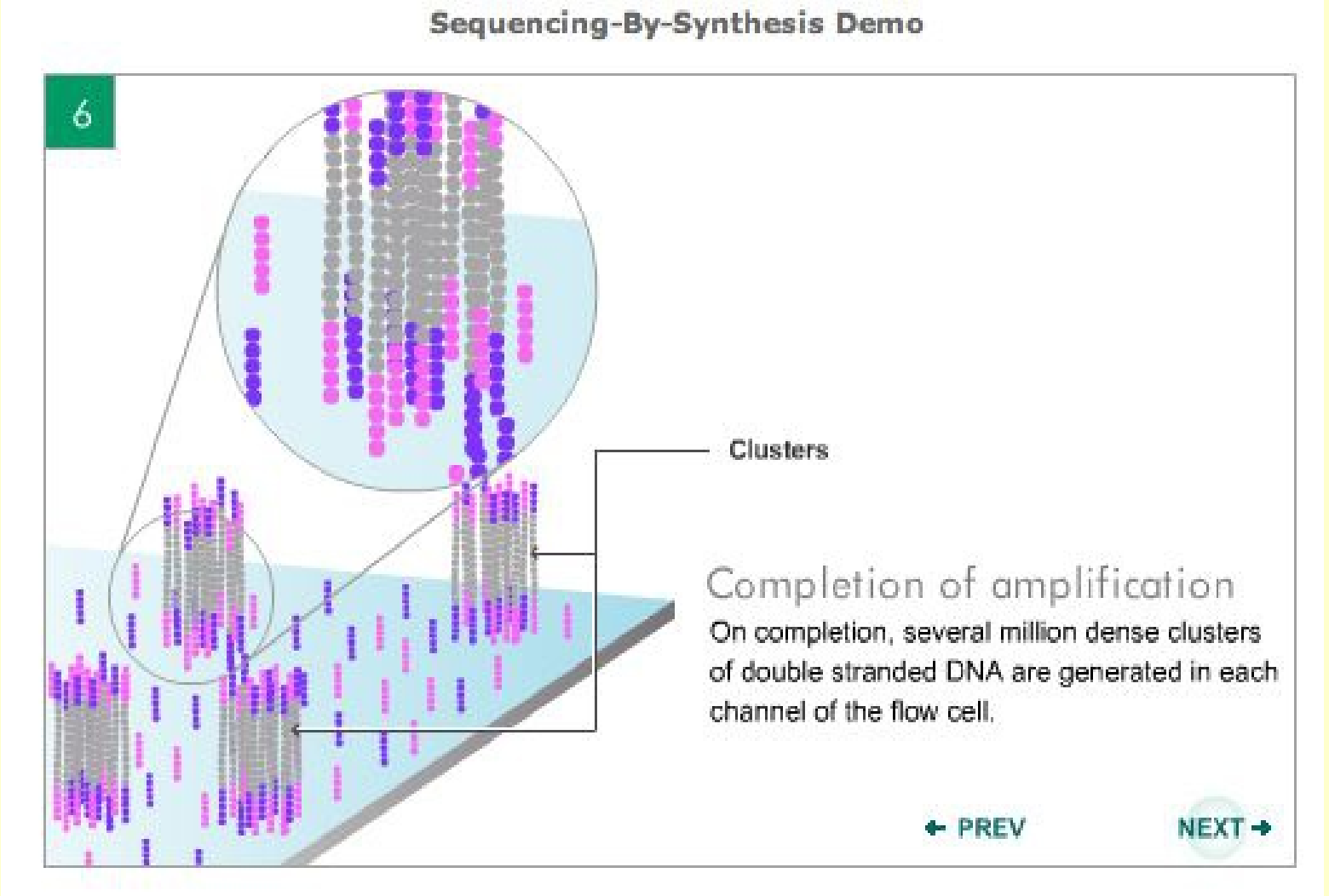
Illumina Solexa Sequencing Technology



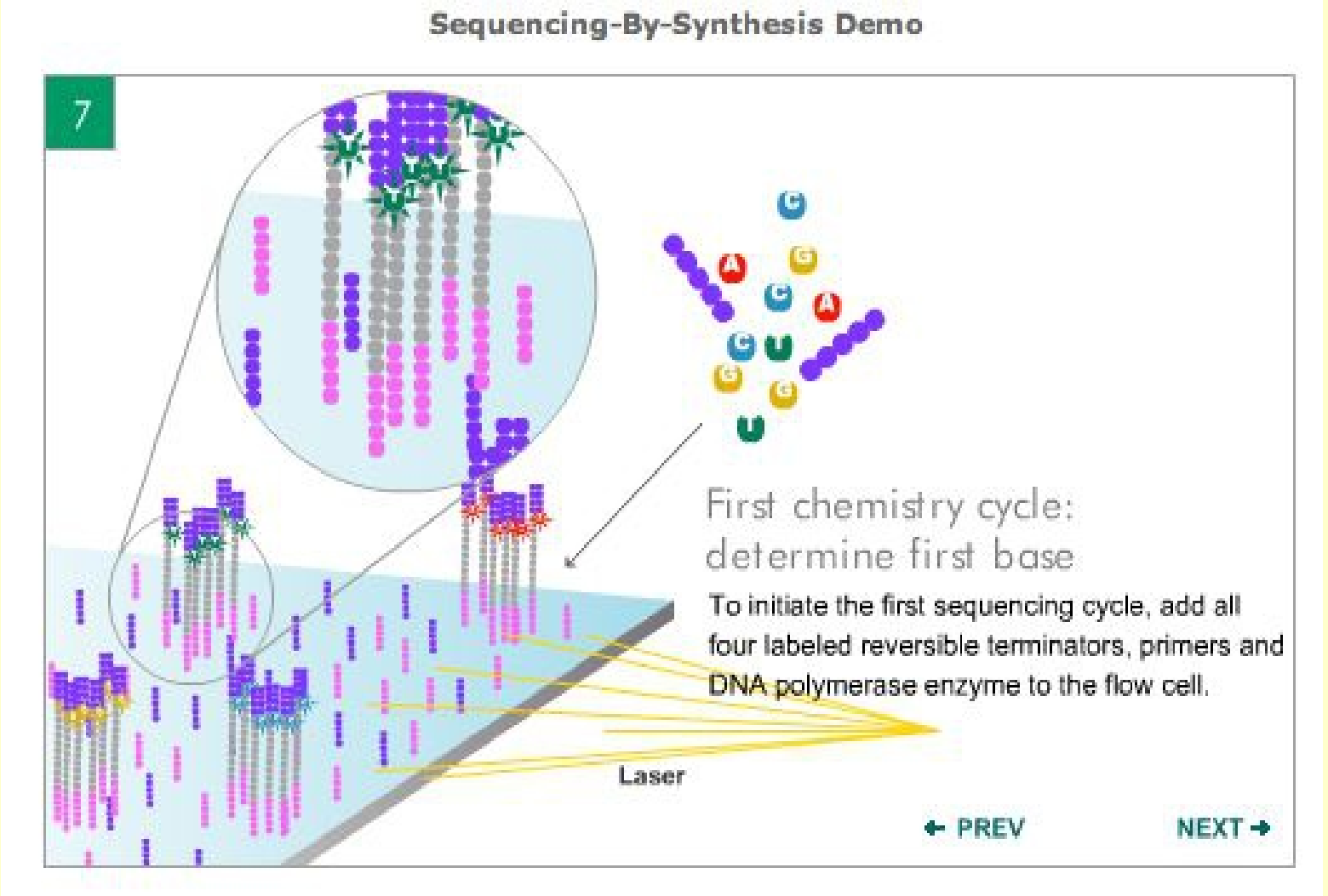
Illumina Solexa Sequencing Technology



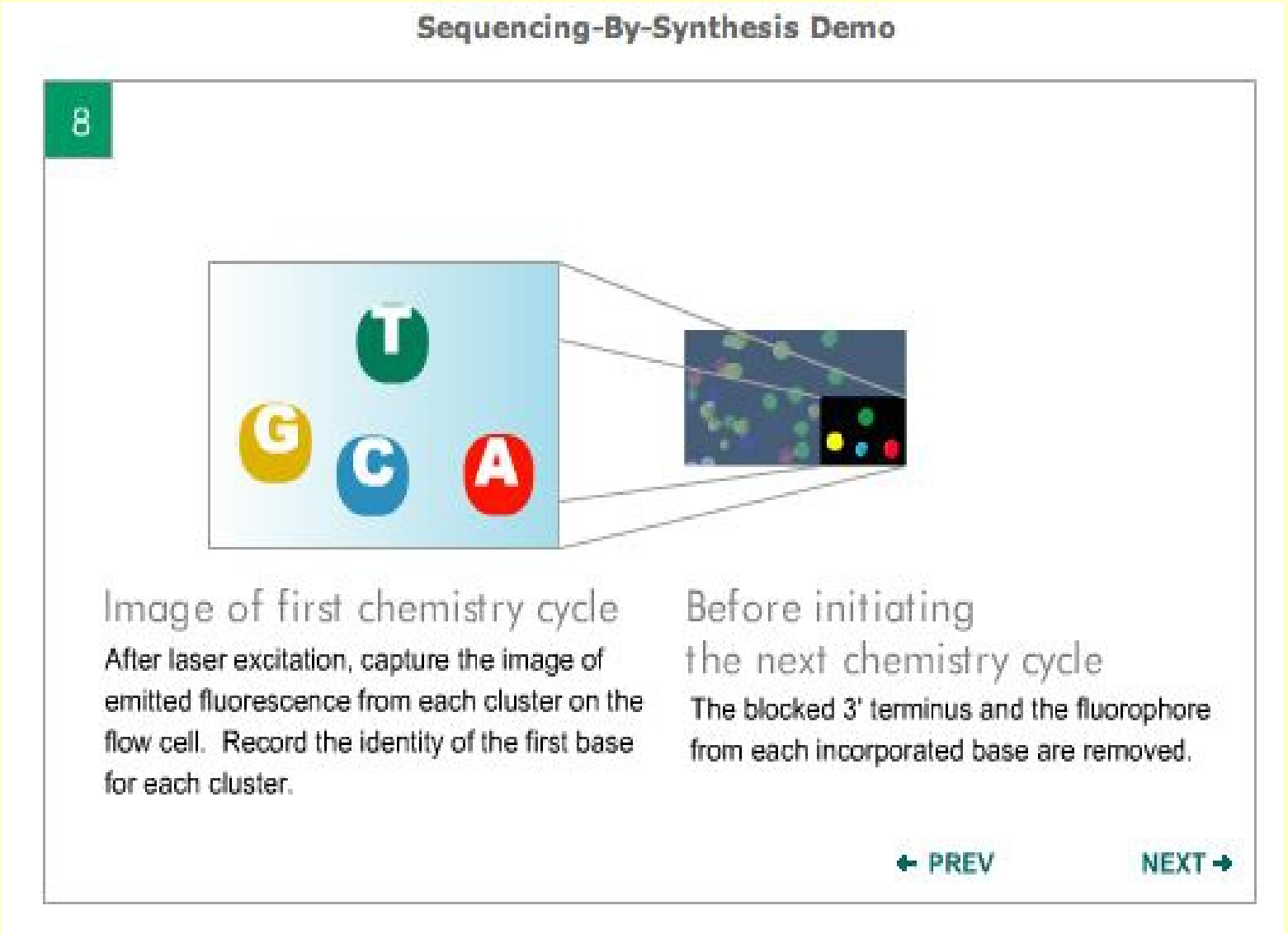
Illumina Solexa Sequencing Technology



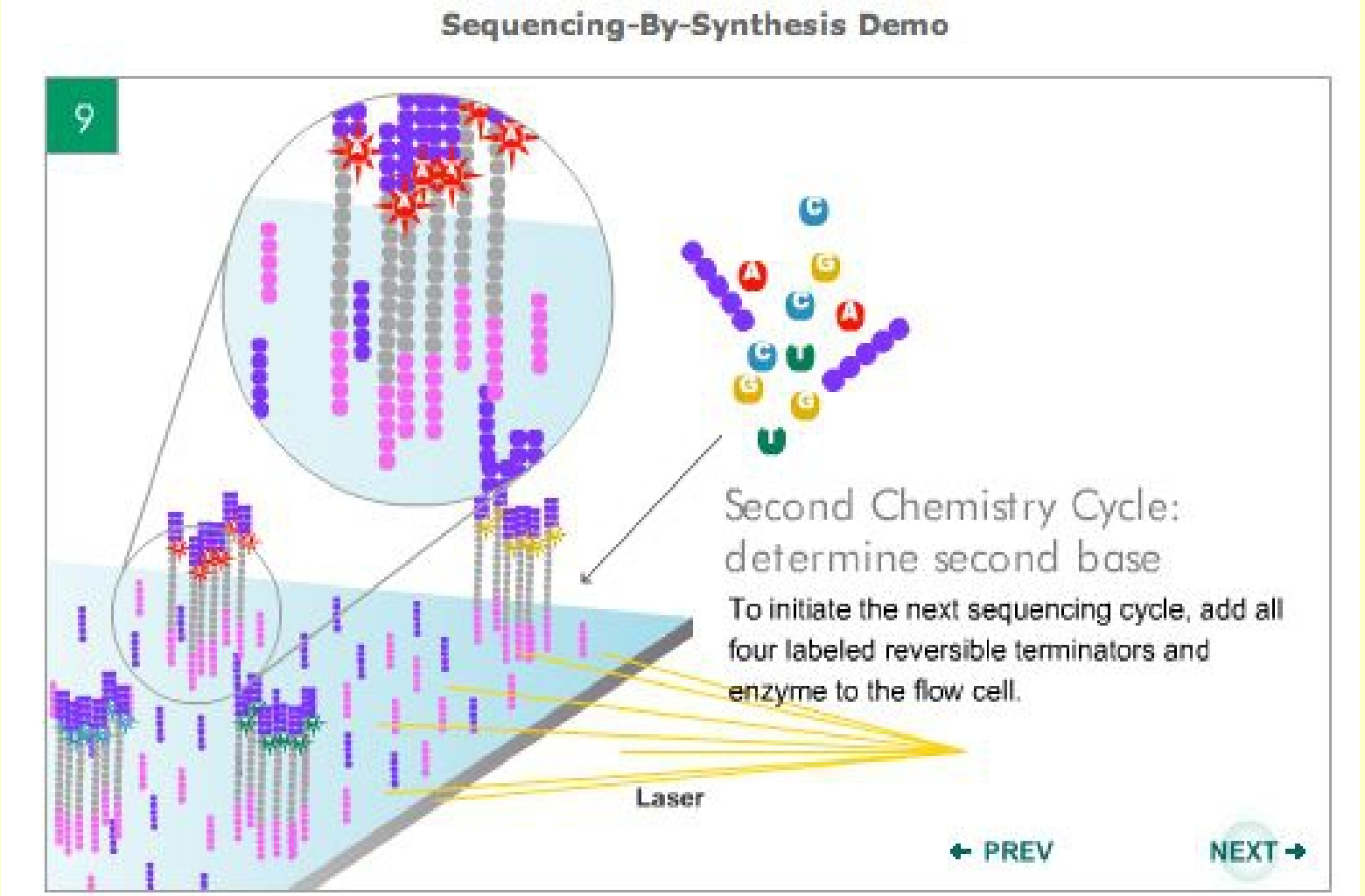
Illumina Solexa Sequencing Technology



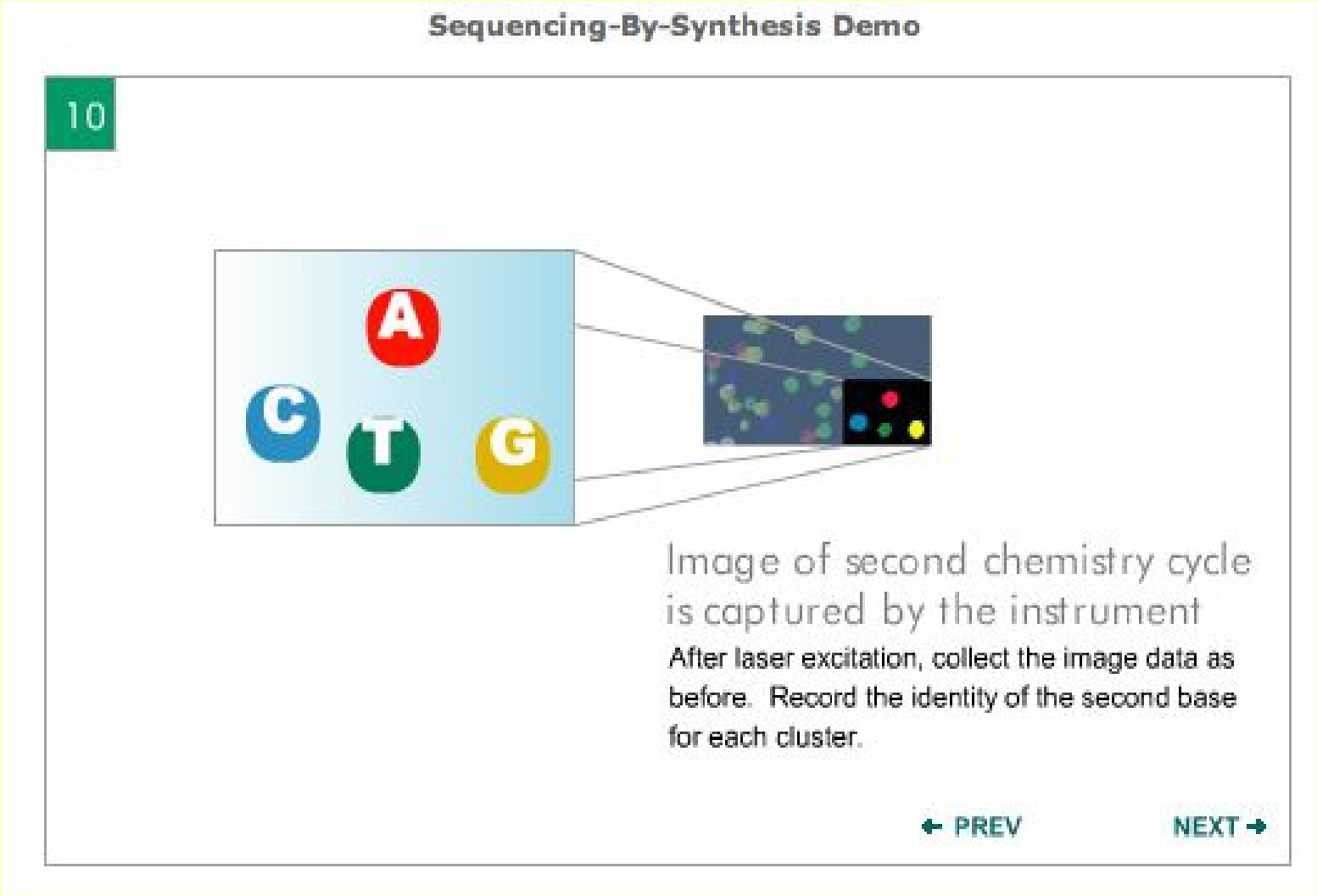
Illumina Solexa Sequencing Technology



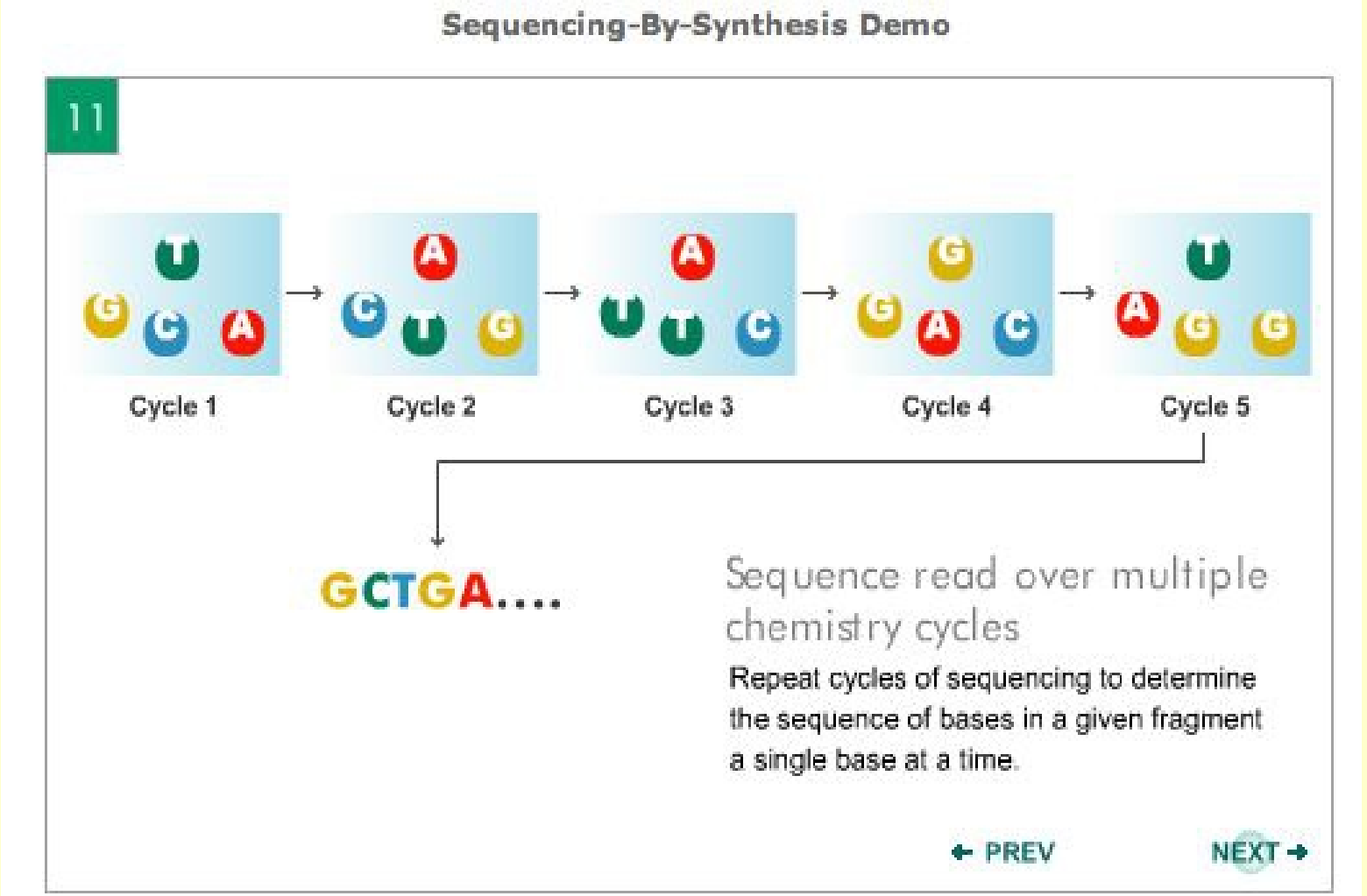
Illumina Solexa Sequencing Technology



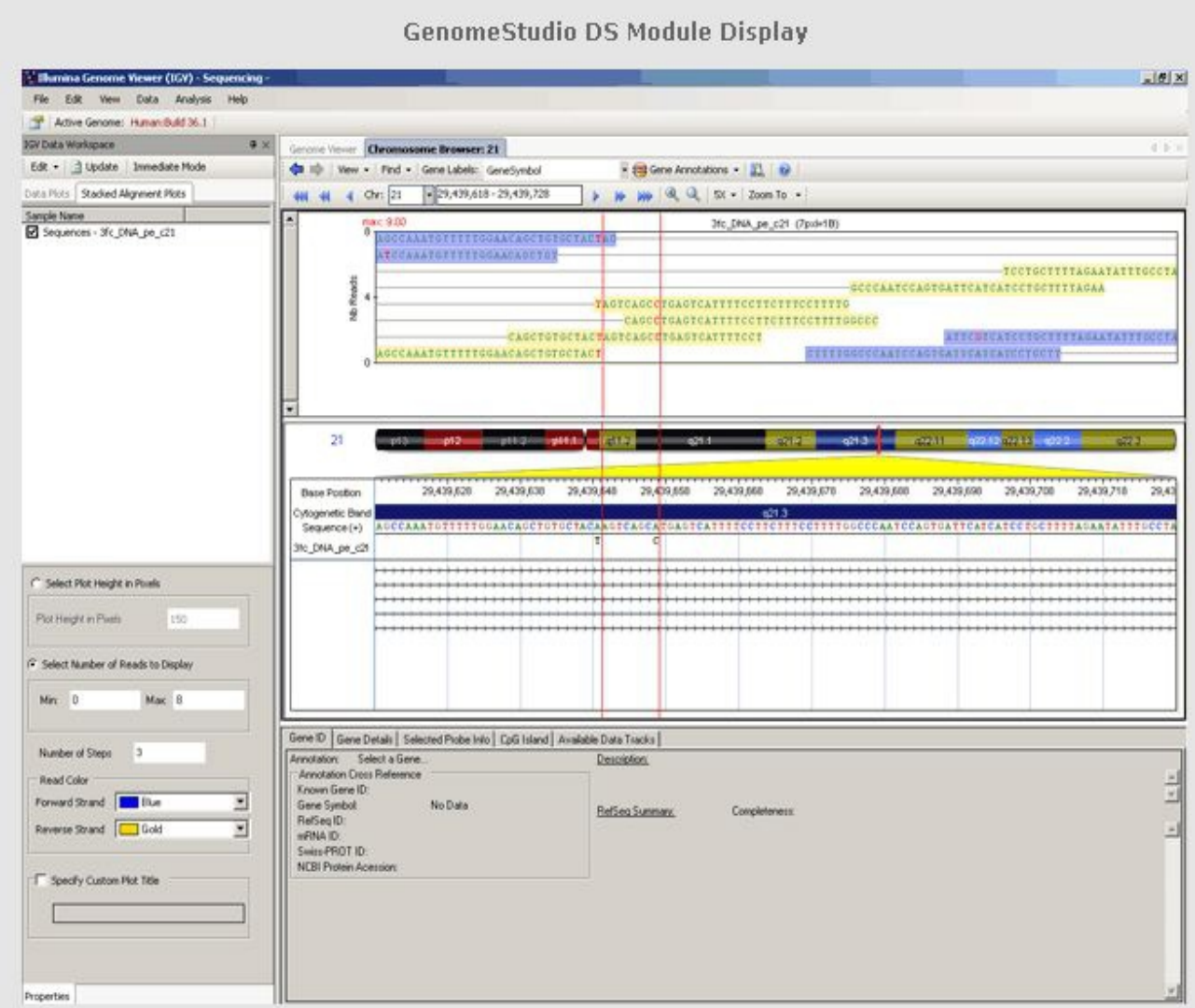
Illumina Solexa Sequencing Technology



Illumina Solexa Sequencing Technology



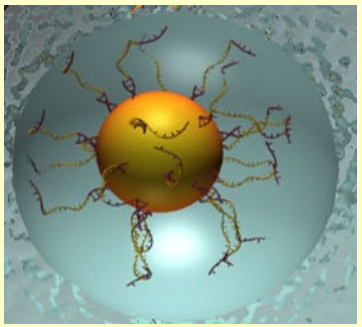
Illumina Solexa Sequencing Technology



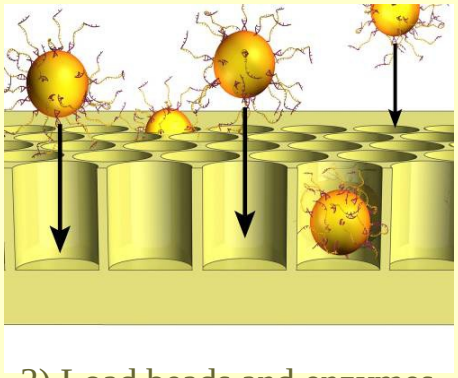
GenomeStudio displays SNPs identified by CASAVA based on alignment of reads against a reference sequence using the Illumina Chromosome Browser.

Life Sciences 454 Process Overview

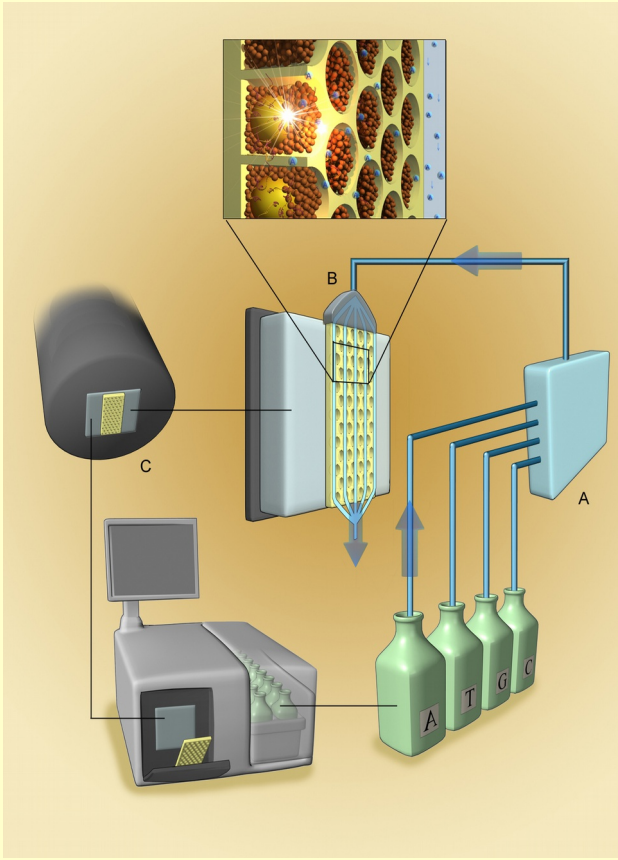
1) Prepare Adapter Ligated ssDNA Library



2) Clonal Amplification on 28 µ beads



3) Load beads and enzymes in PicoTiter Plate™

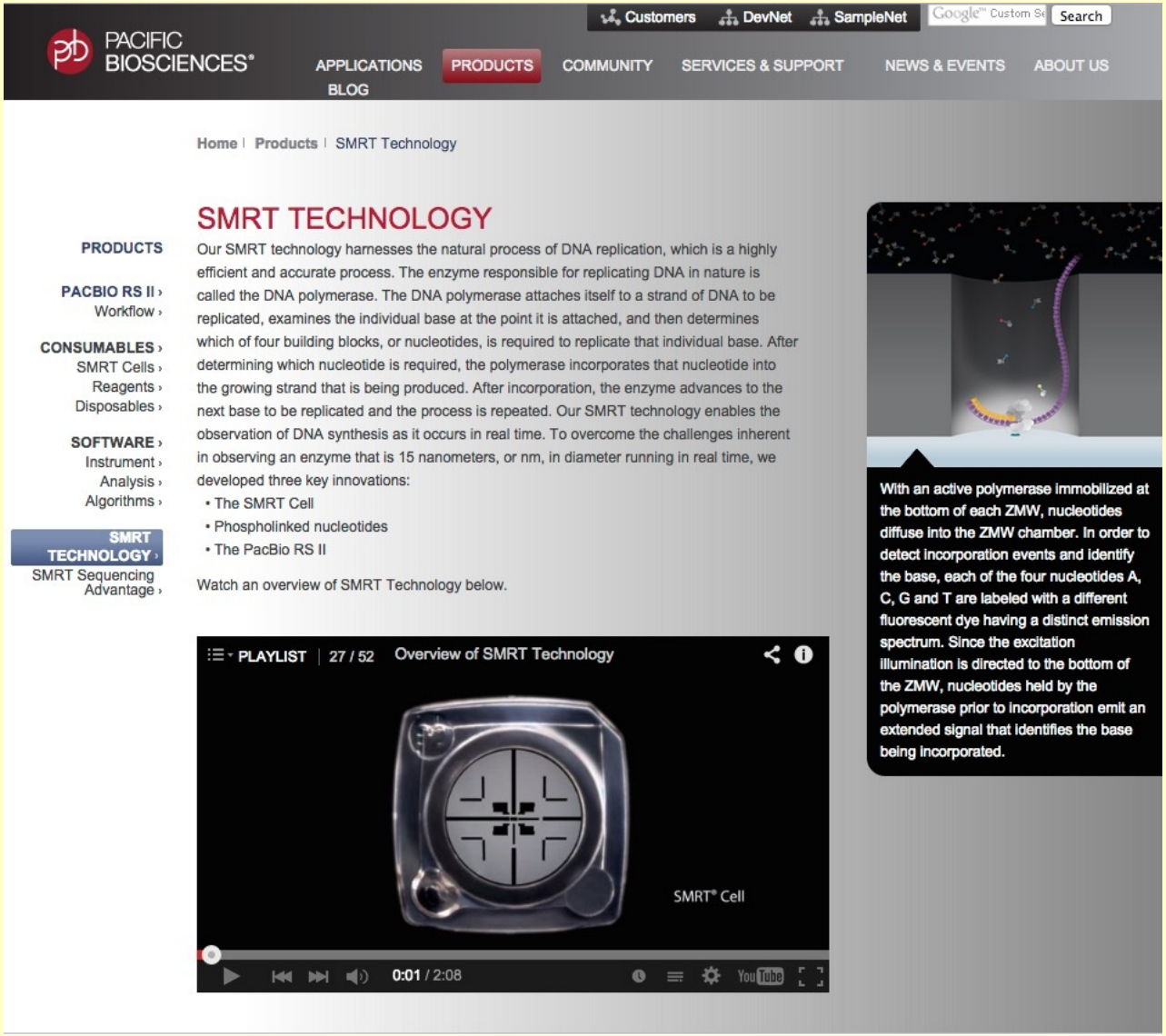


4) Perform Sequencing by synthesis on the 454 Instrument

Pacific Biosciences SMRT Sequencing

New PacBio Sequencing Technology Video

http://www.pacificbiosciences.com/products/smrt-technology/



PACIFIC BIOSCIENCES

Pacific Biosciences SMRT Sequencing

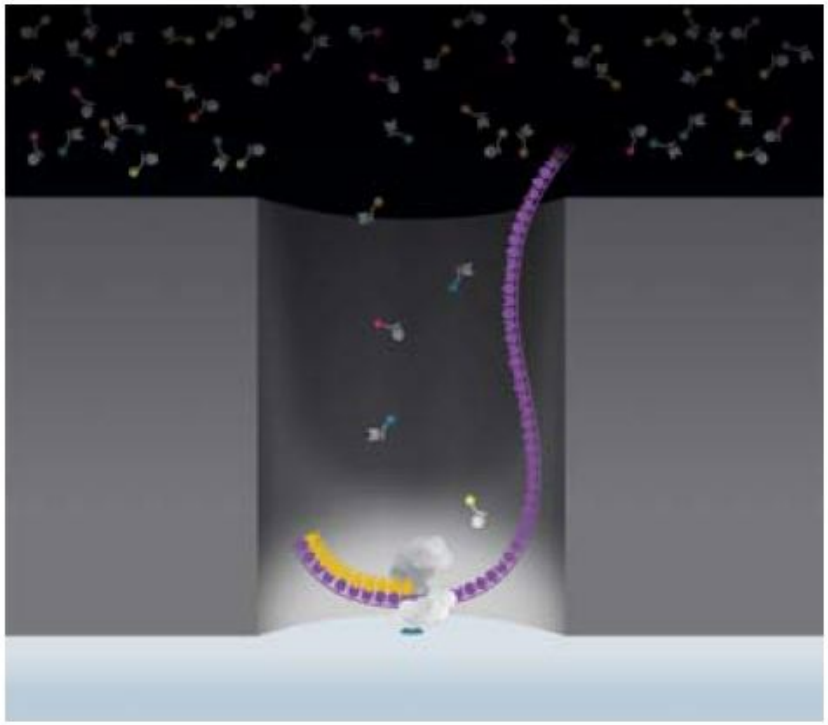


Figure 6. ZMW with DNA polymerase and phospholinked nucleotides

Phospholinked nucleotides are added into the ZMW at the high concentrations required for proper enzyme functioning.

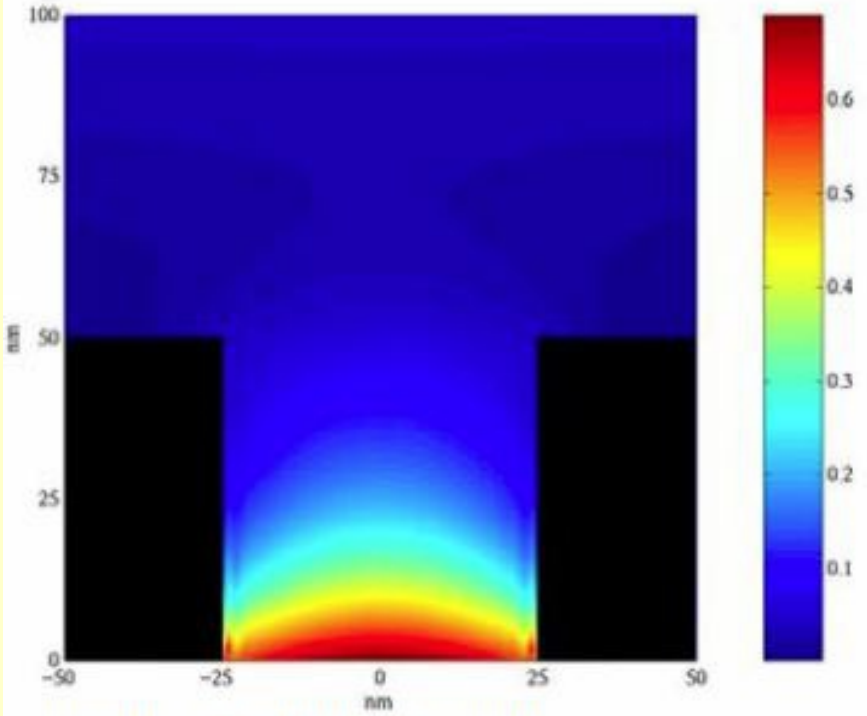


Figure 4. Detection volume

Attenuated light from the excitation beam penetrates only the lower 20-30 nm of each waveguide, creating a detection volume of 20 zeptoliters (10^{-21} liters).

Phospholinked Fluorophores

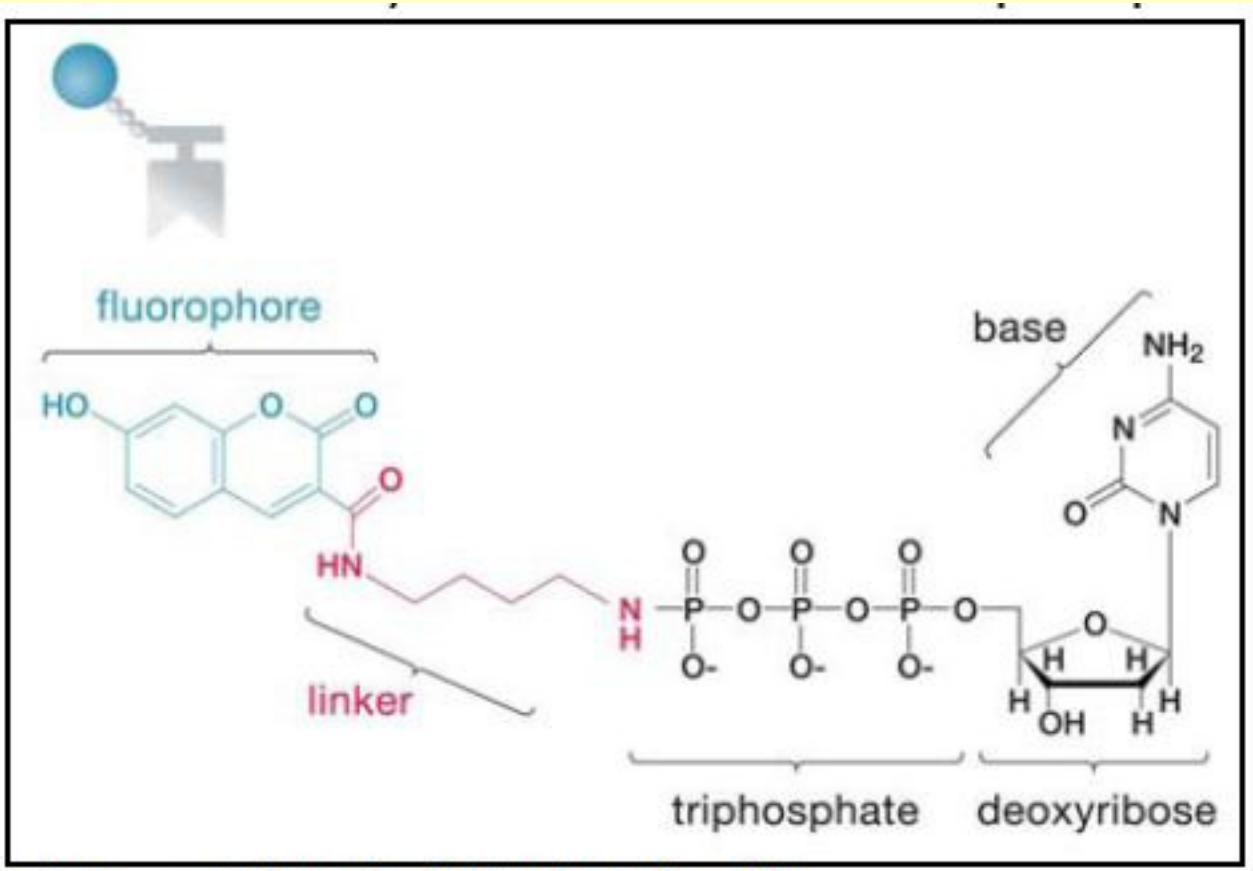
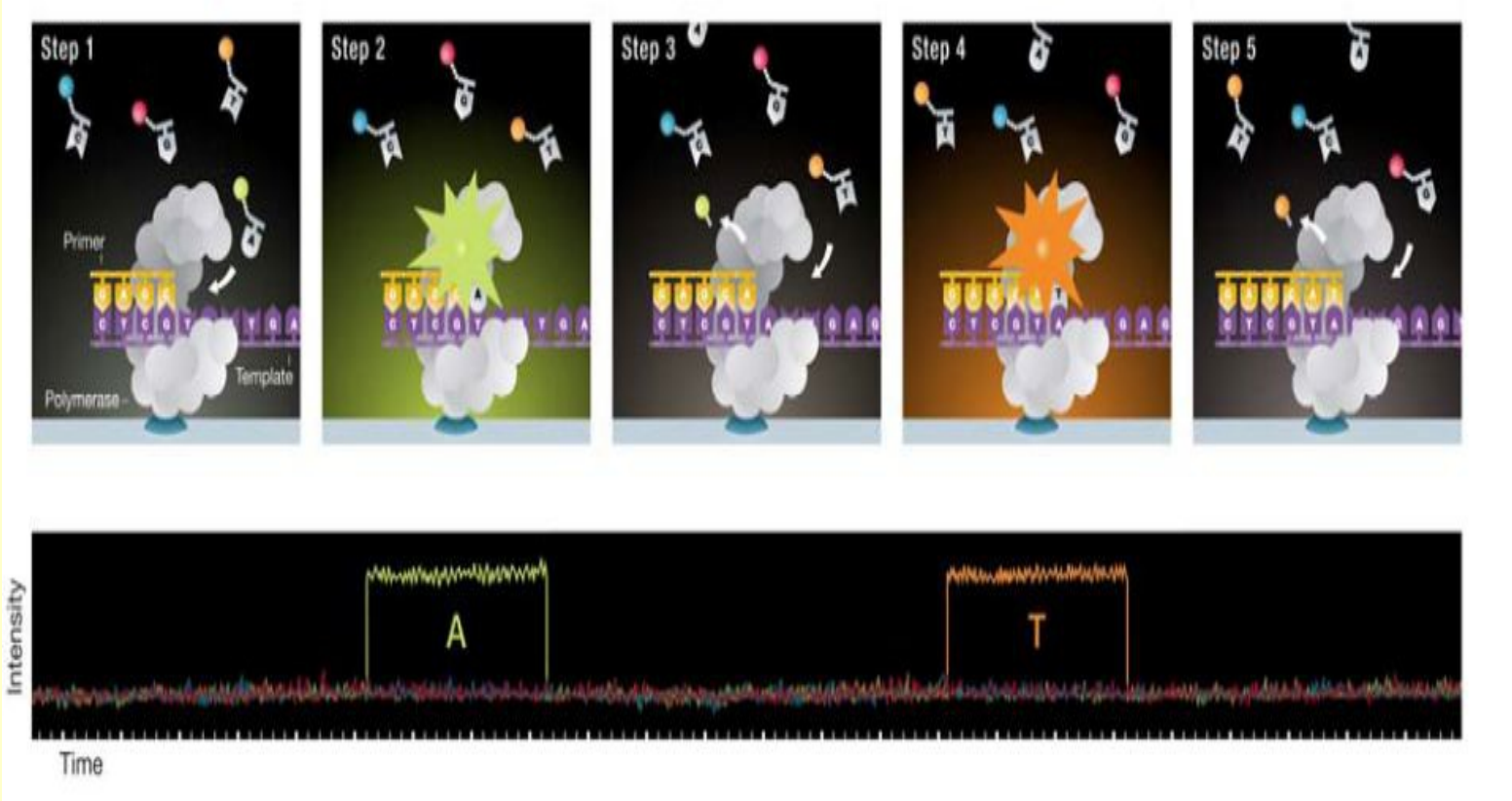


Figure 9. Phospholinked nucleotides

Phospholinked nucleotides have fluorophores attached to the triphosphate chain, which is naturally cleaved when the nucleotide is incorporated.

PACIFIC BIOSCIENCES

Processive Synthesis



Synthesis of Long Duplex DNA

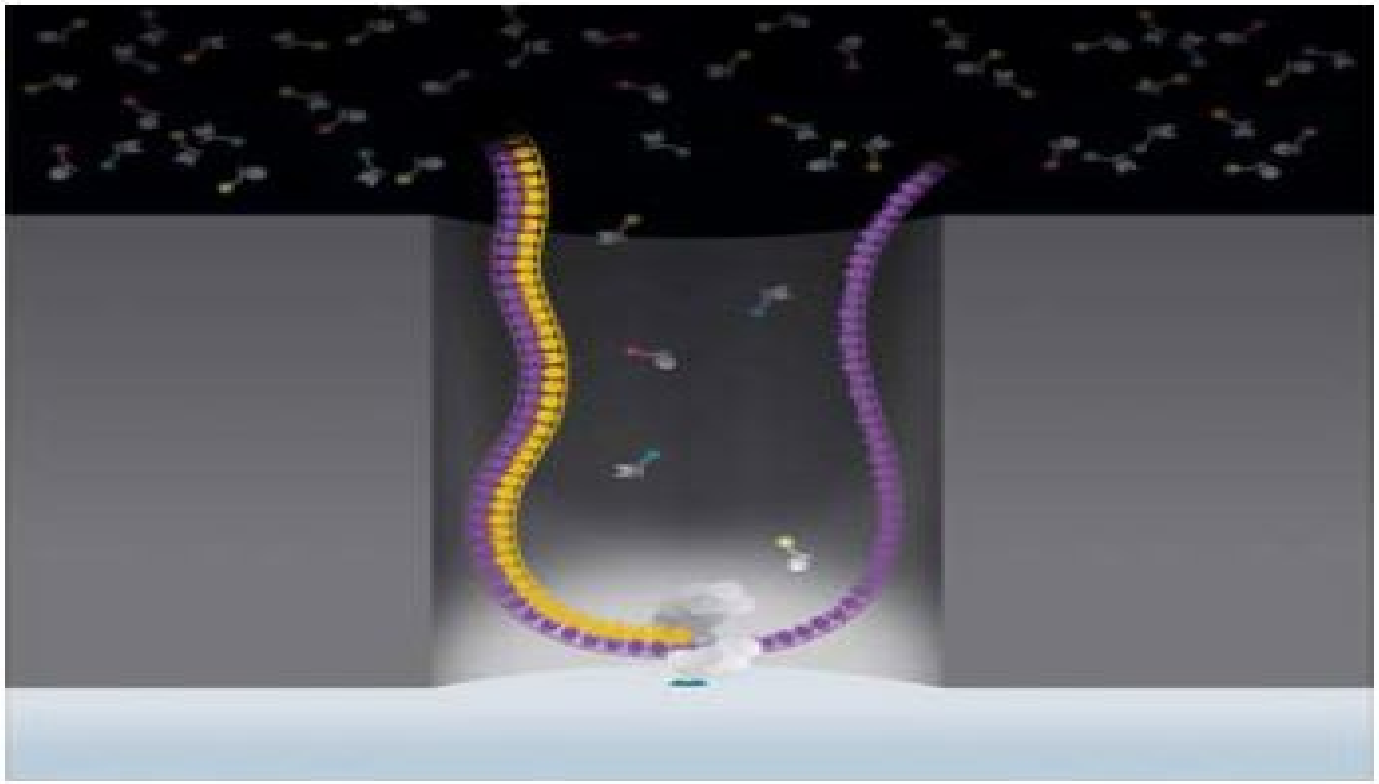
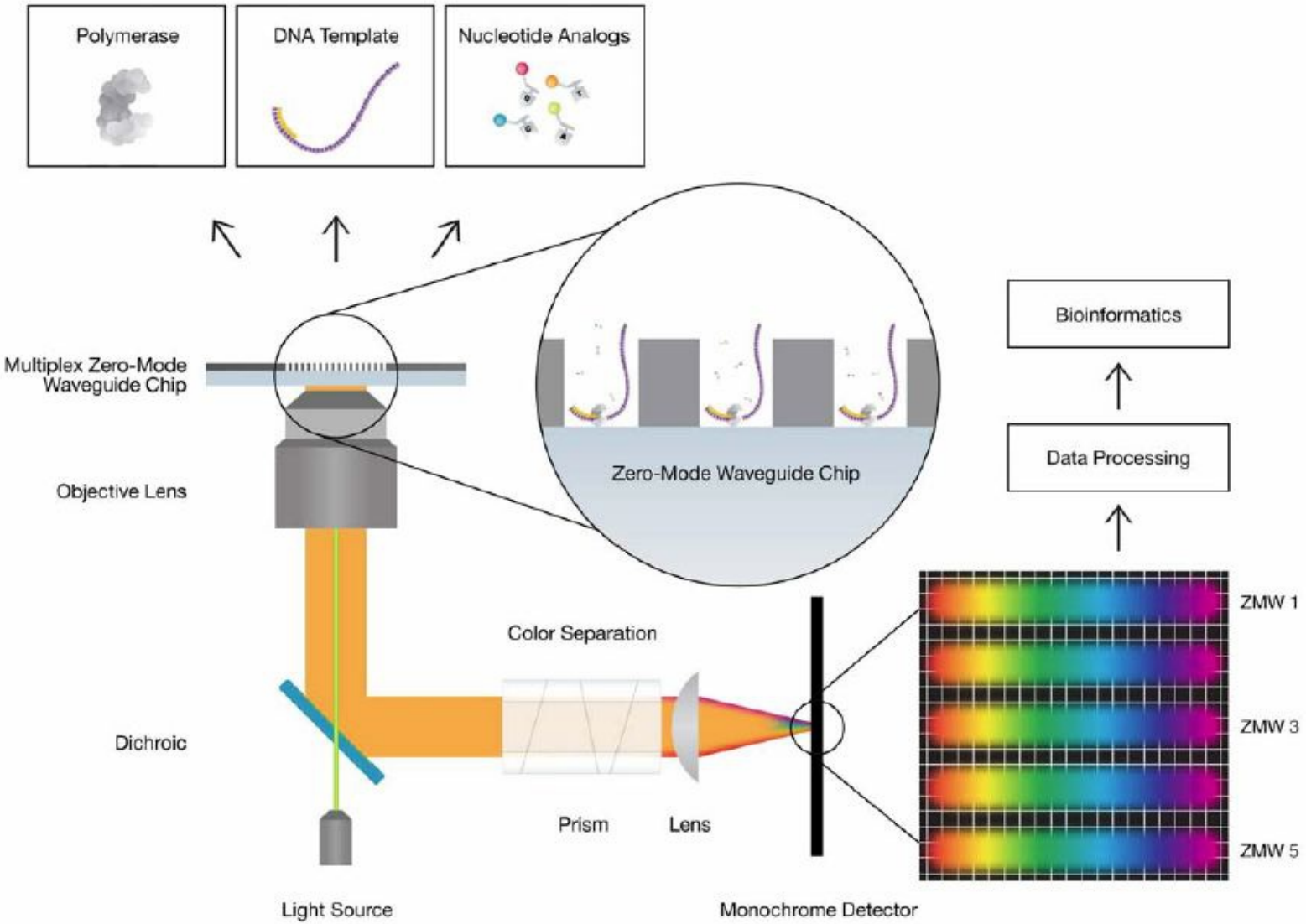


Figure 11. Synthesis of long DNA.

DNA polymerase processively incorporates nucleotides producing long, natural DNA.

PACIFIC BIOSCIENCES

Highly Parallel Optics System



PACIFIC BIOSCIENCES

Circular Templates Gives Redundant Sequencing and Accuracy

Circular Templates Gives Redundant Sequencing and Accuracy

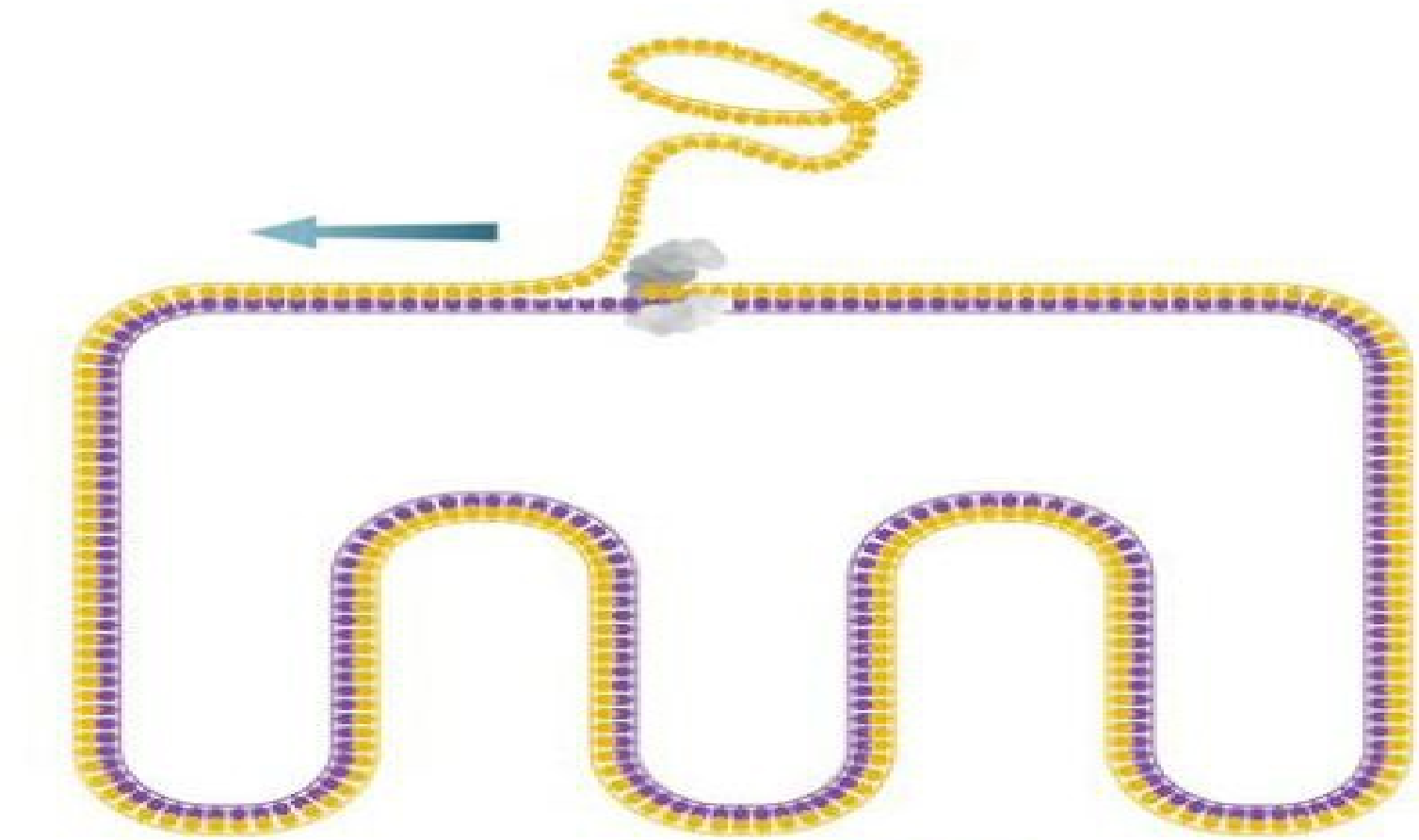
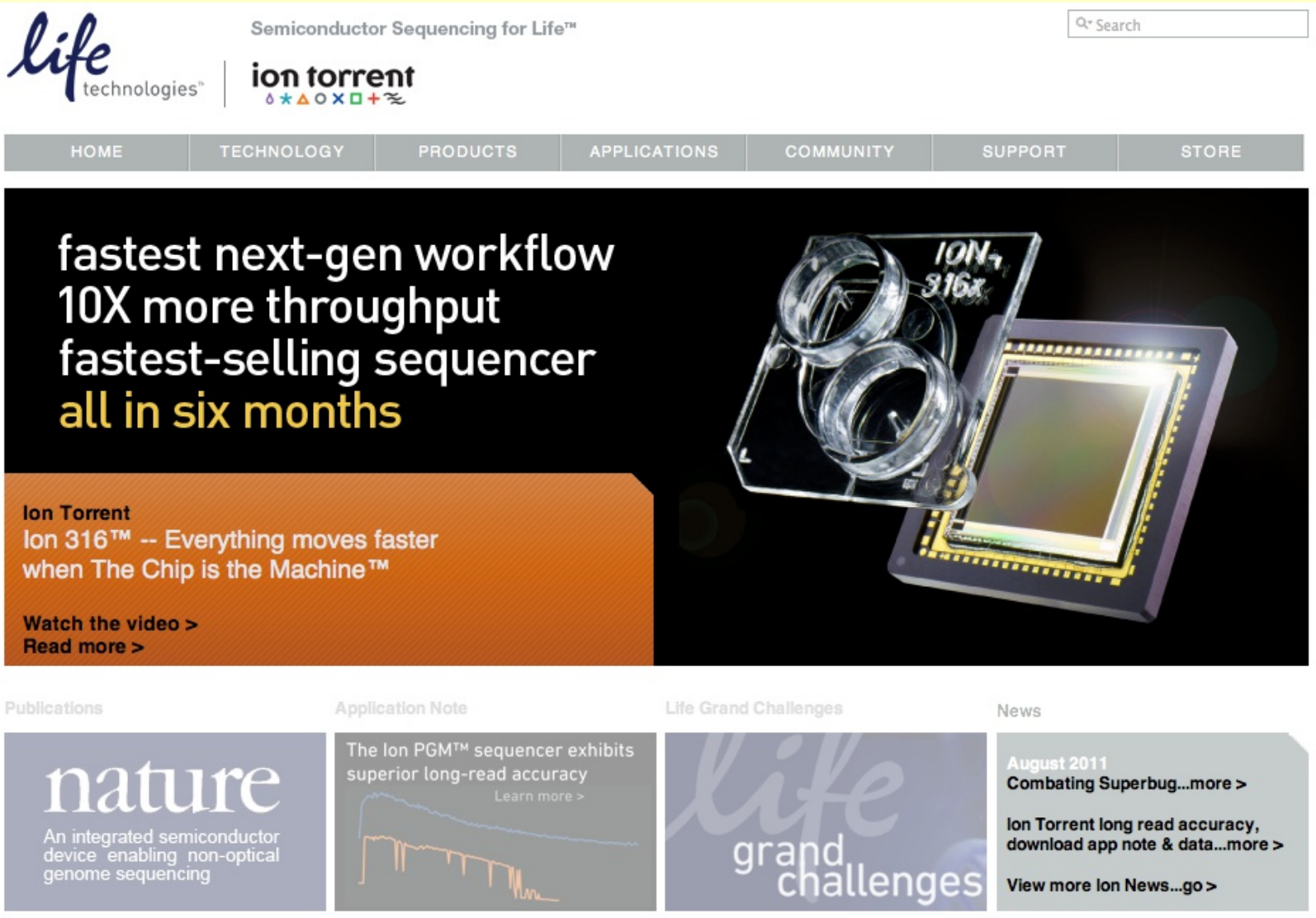


Figure 16. Molecular redundant sequencing

Ion Torrent Sequencing

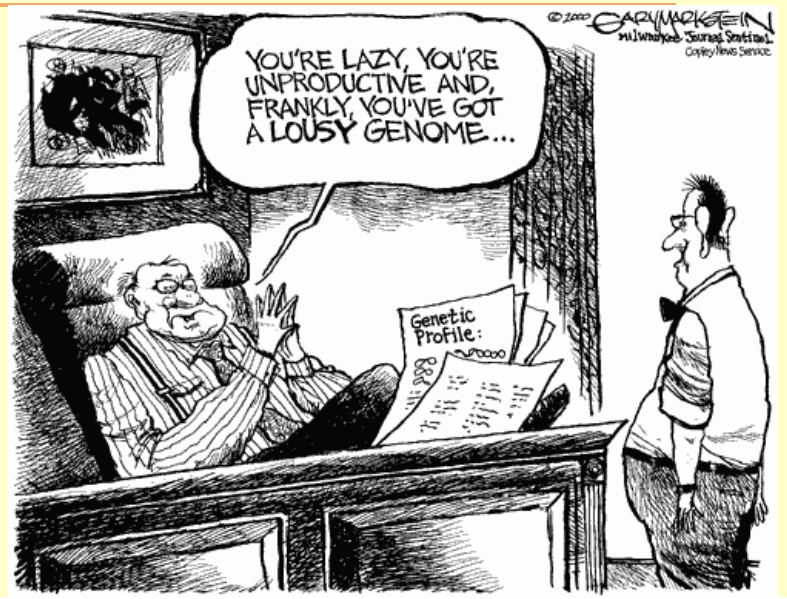


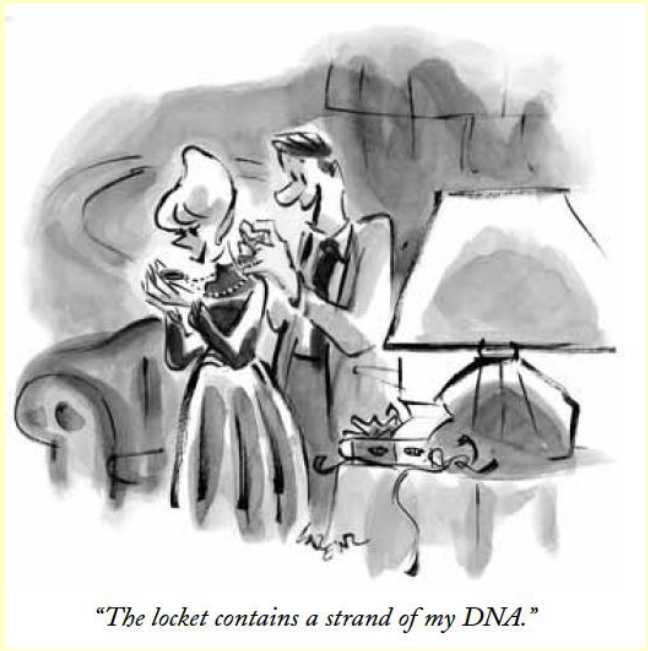
Ion Torrent Sequencing

Ion Torrent Sequencing

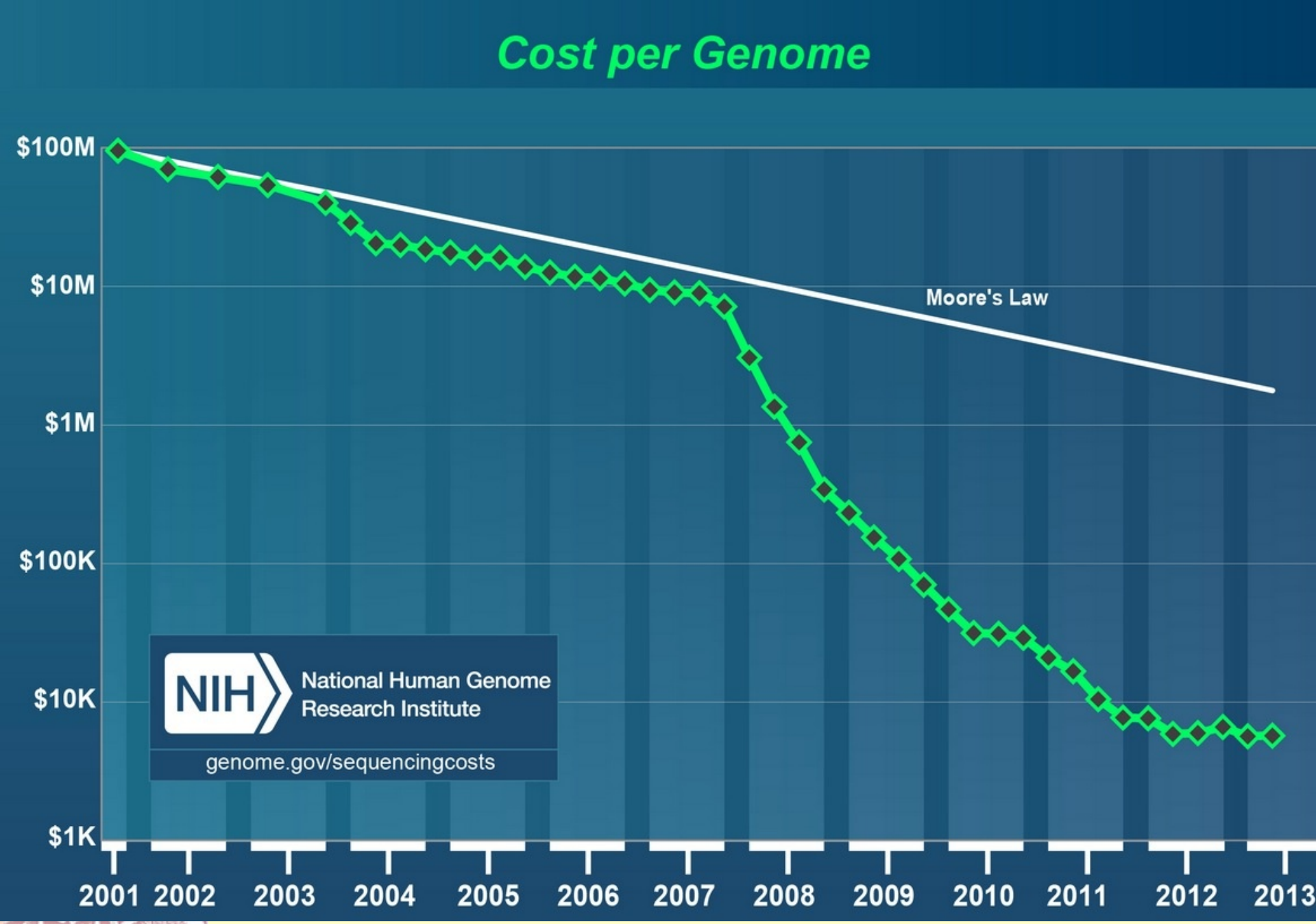
The Human Genome
How fast is the cost going down?

- 2006: $ 50 million
- 2008: $500,000
- 2009: $50,000
- 2010: $20,000
- 2011: $5,000
- 2012: $4,000
- 2013: $3,000
- 2014 $1,700



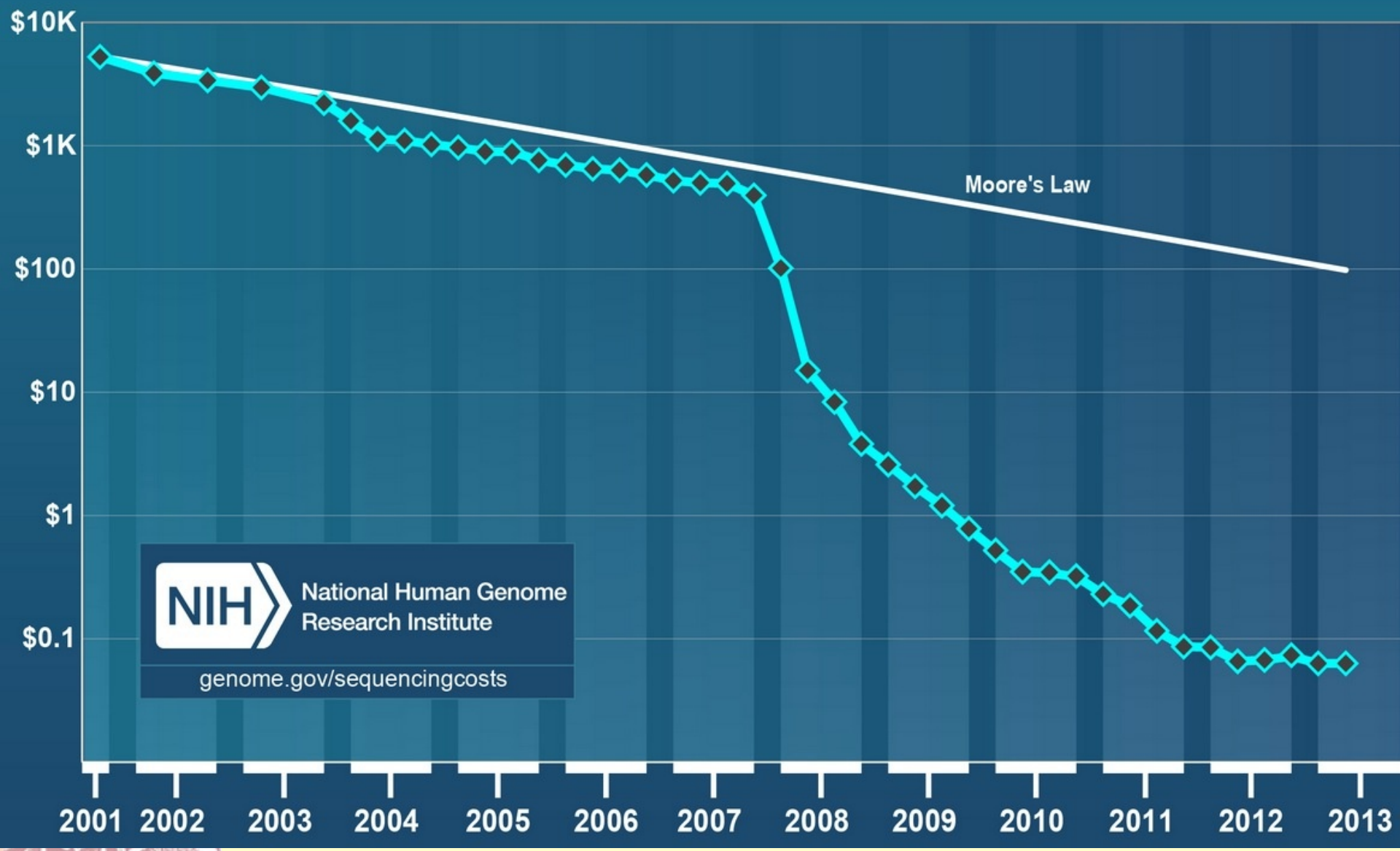

"The locket contains a strand of my DNA."

Thanks to Seraf in Batzoglou



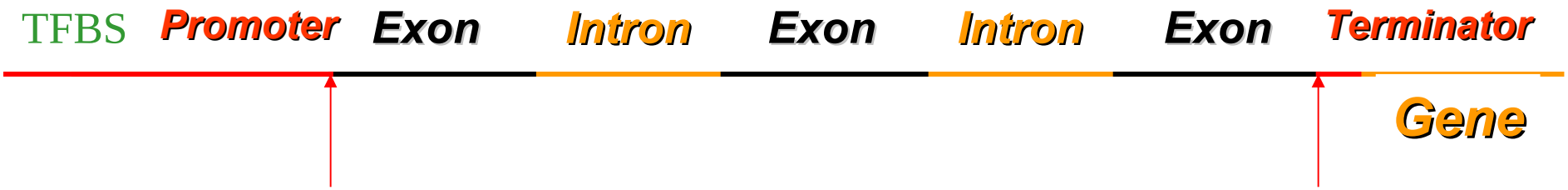
http://www.genome.gov/sequencingcosts/

Cost per Raw Megabase of DNA Sequence

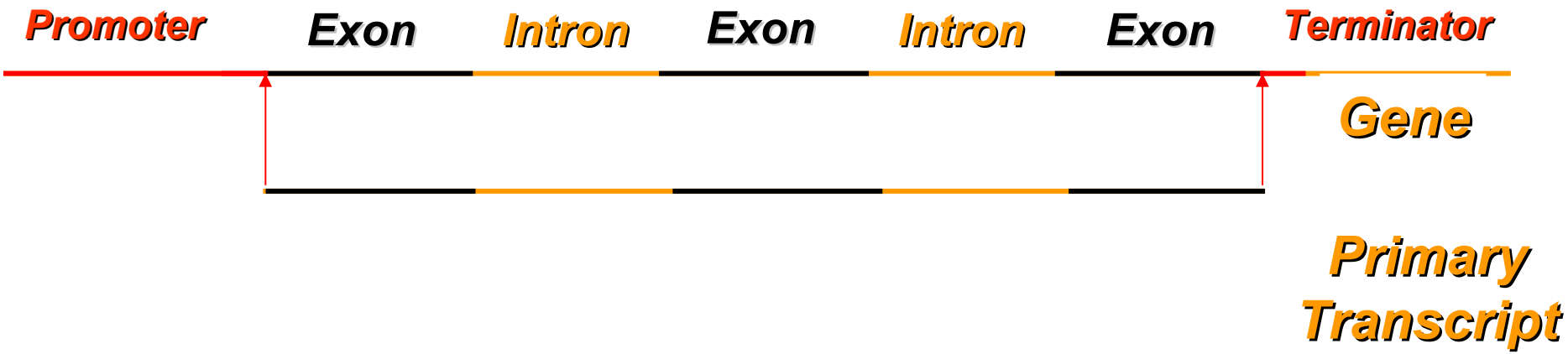


http://www.genome.gov/sequencingcosts/

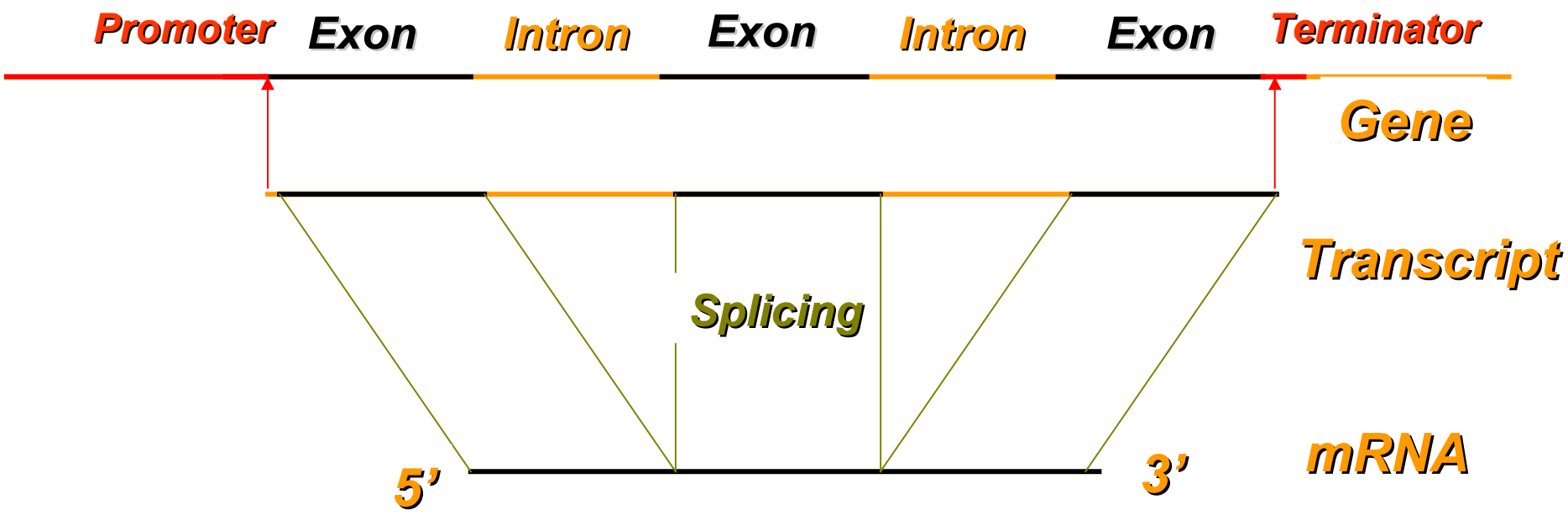
Components of a Typical Human Gene



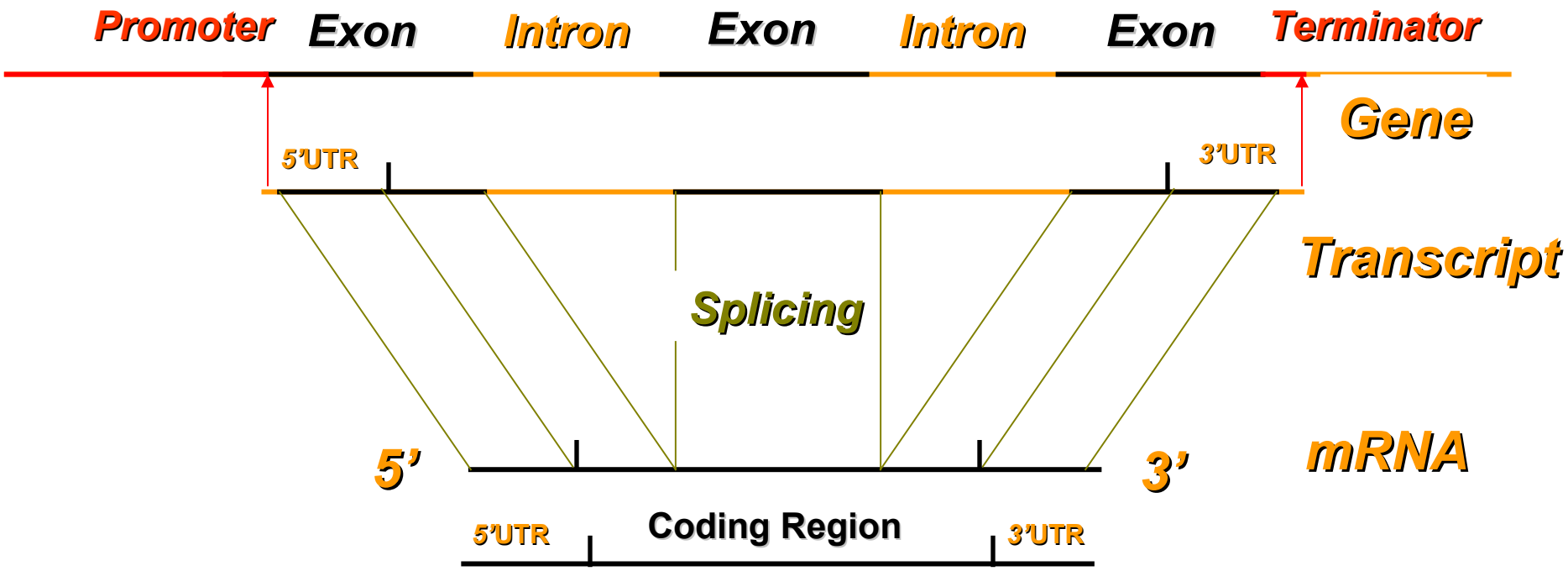
Active Genes are Transcribed into RNA



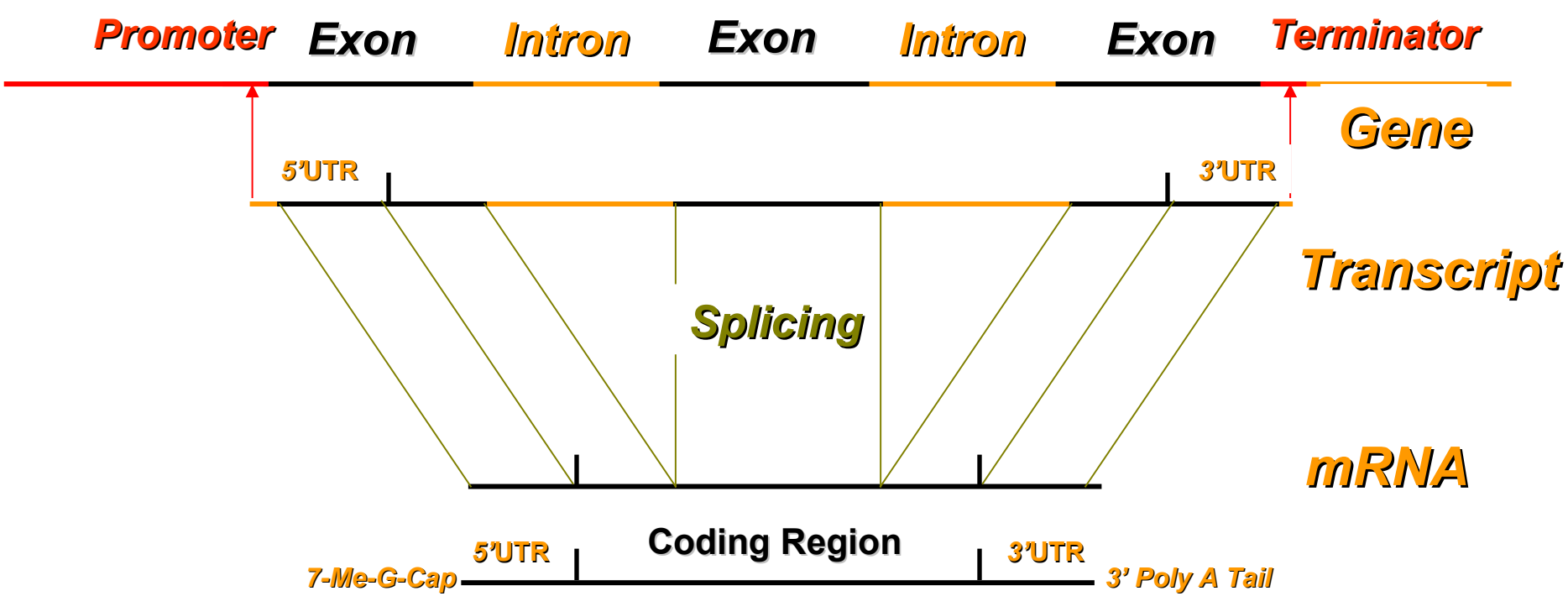
Splicing Transcript Yields Mature mRNA



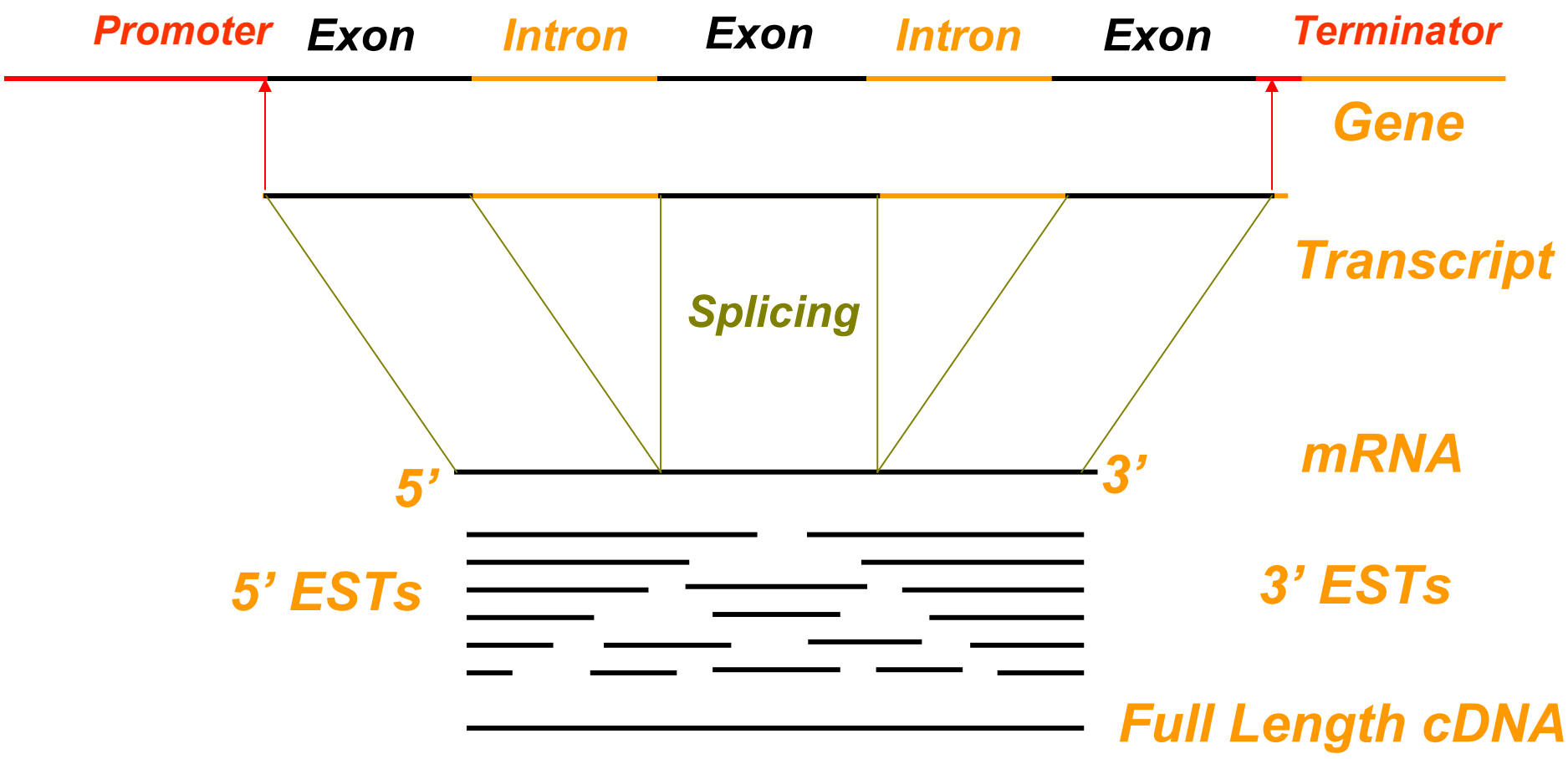
Mature mRNA contains Coding Region and 5′ and 3′ Untranslated Regions



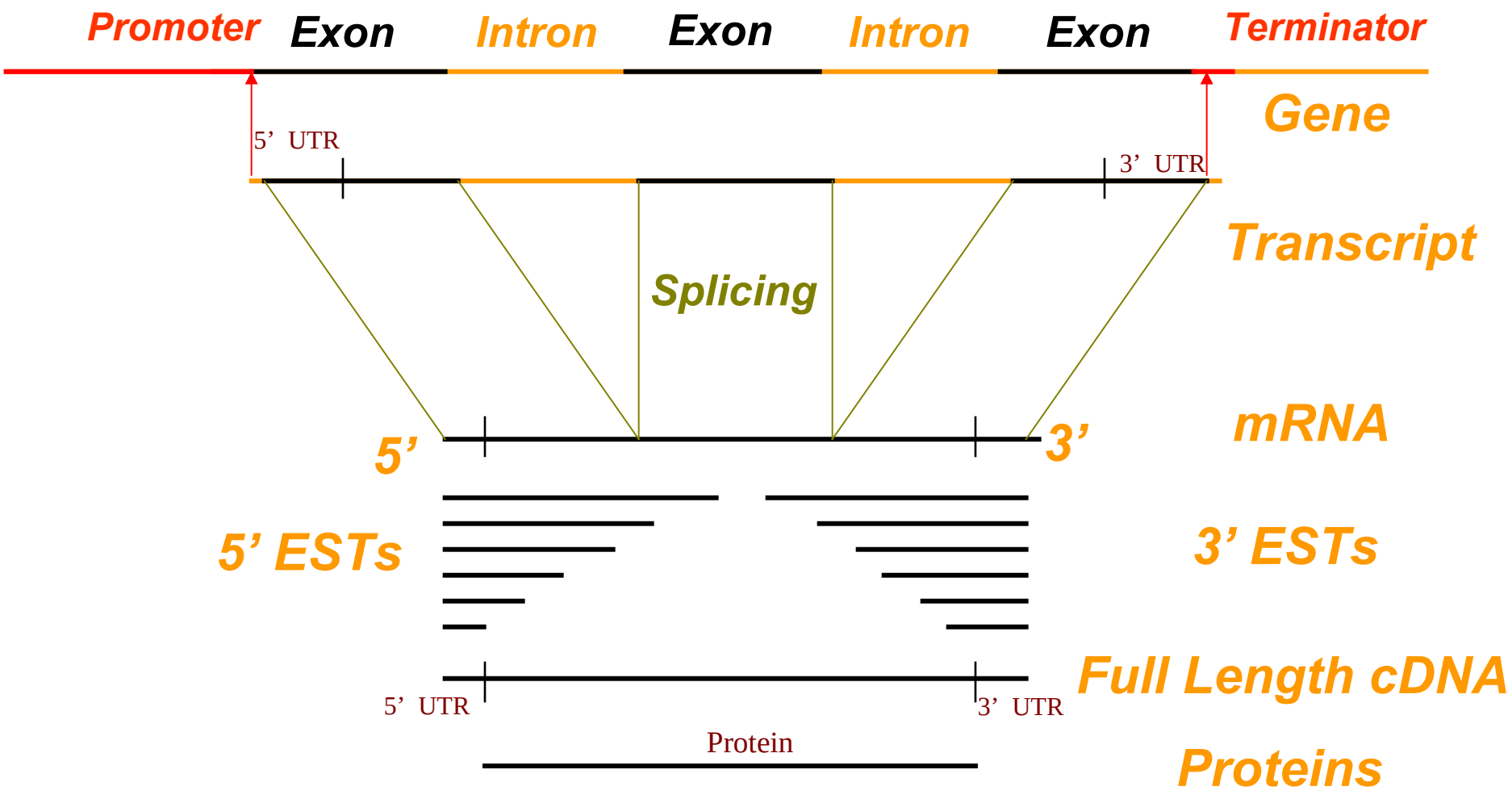
Mature mRNA contains 7-Methyl-Guanosine 5′ Cap and 3′ Poly A Tail



ESTs, Full Length cDNA
UniGene & RefSeq Databases



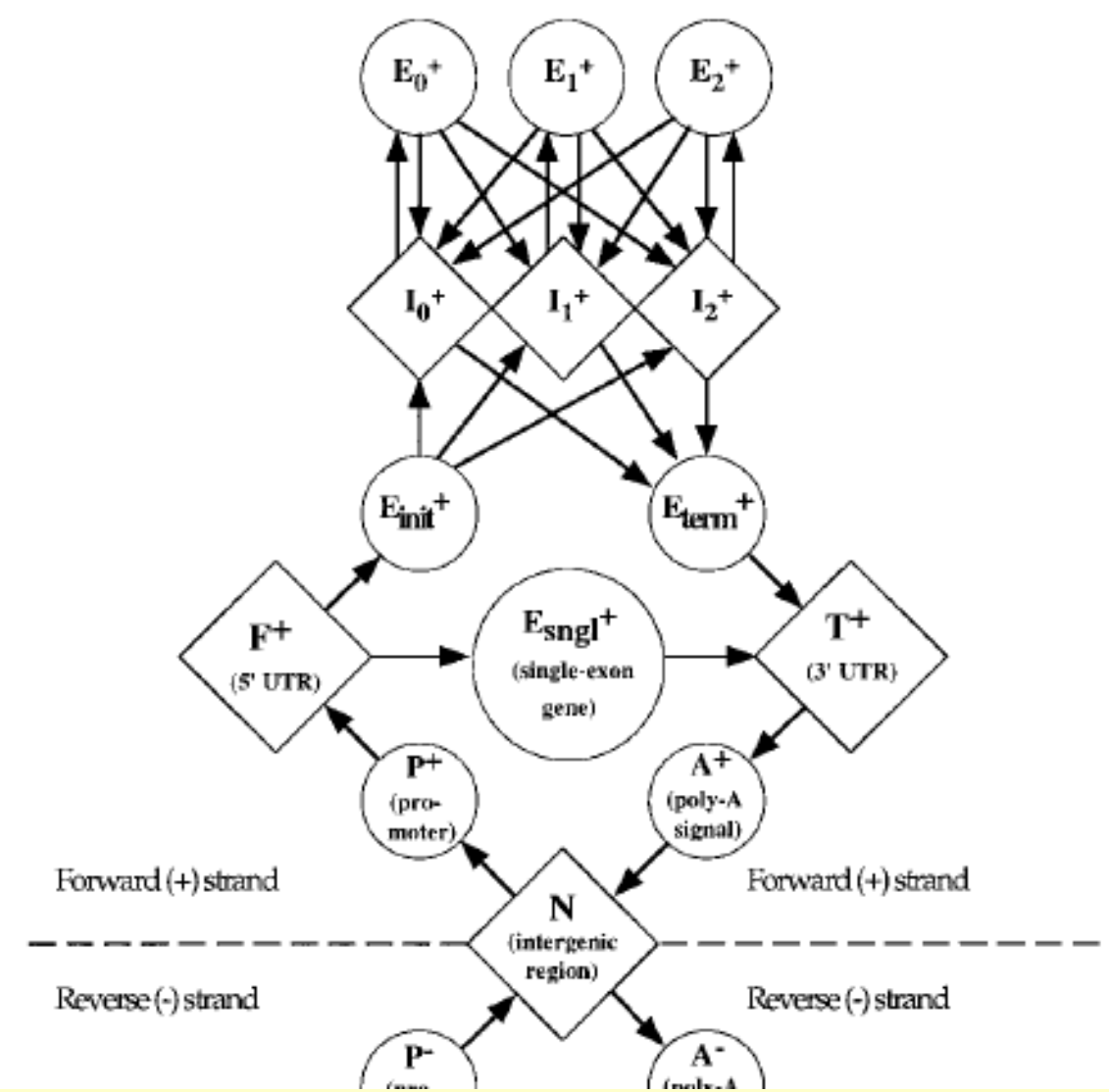
ESTs, Full Length cDNA
UniGene & RefSeq Databases



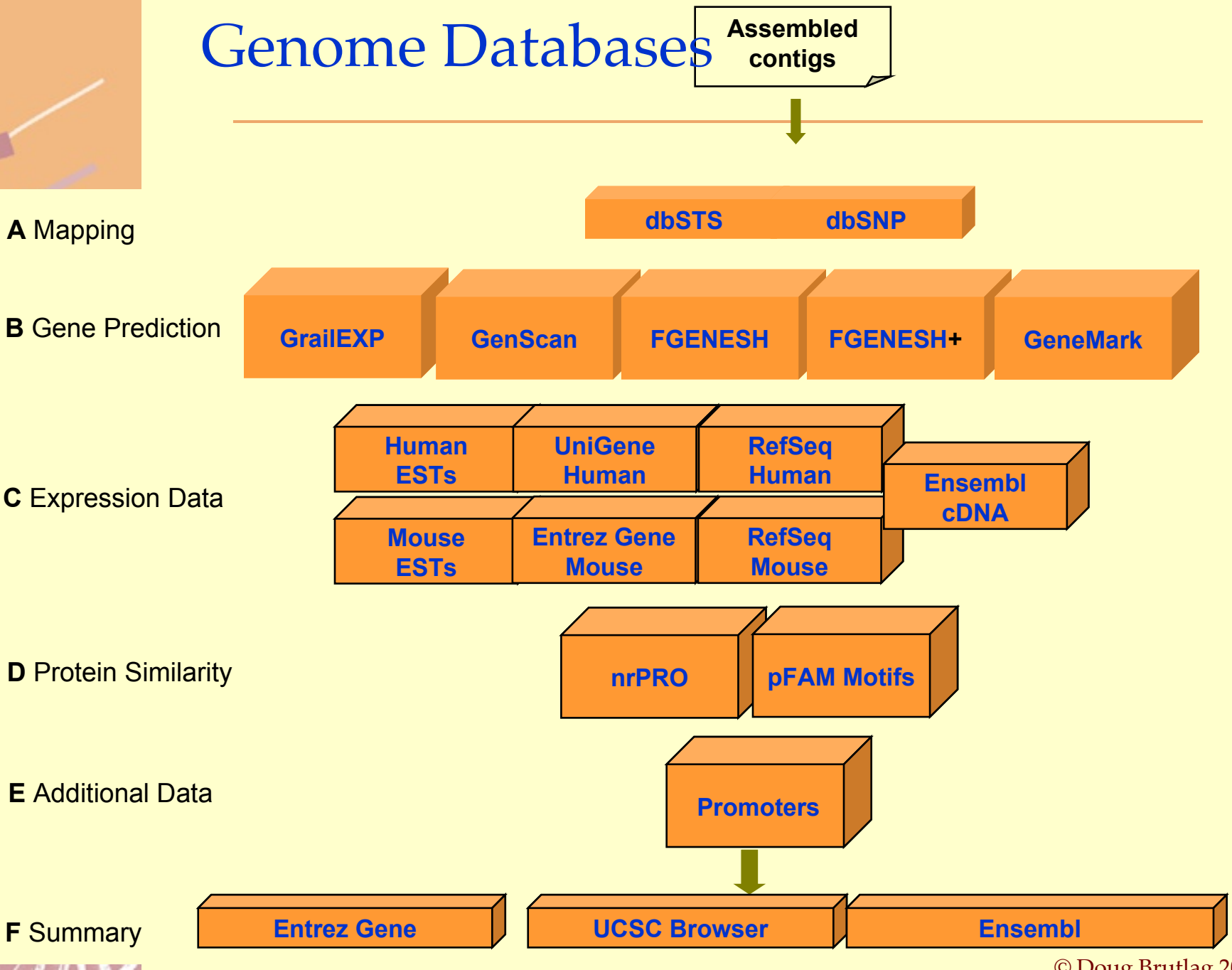
GENSCAN Gene Model

http://genes.mit.edu/GENSCAN.html

Hidden Markov models of gene structure

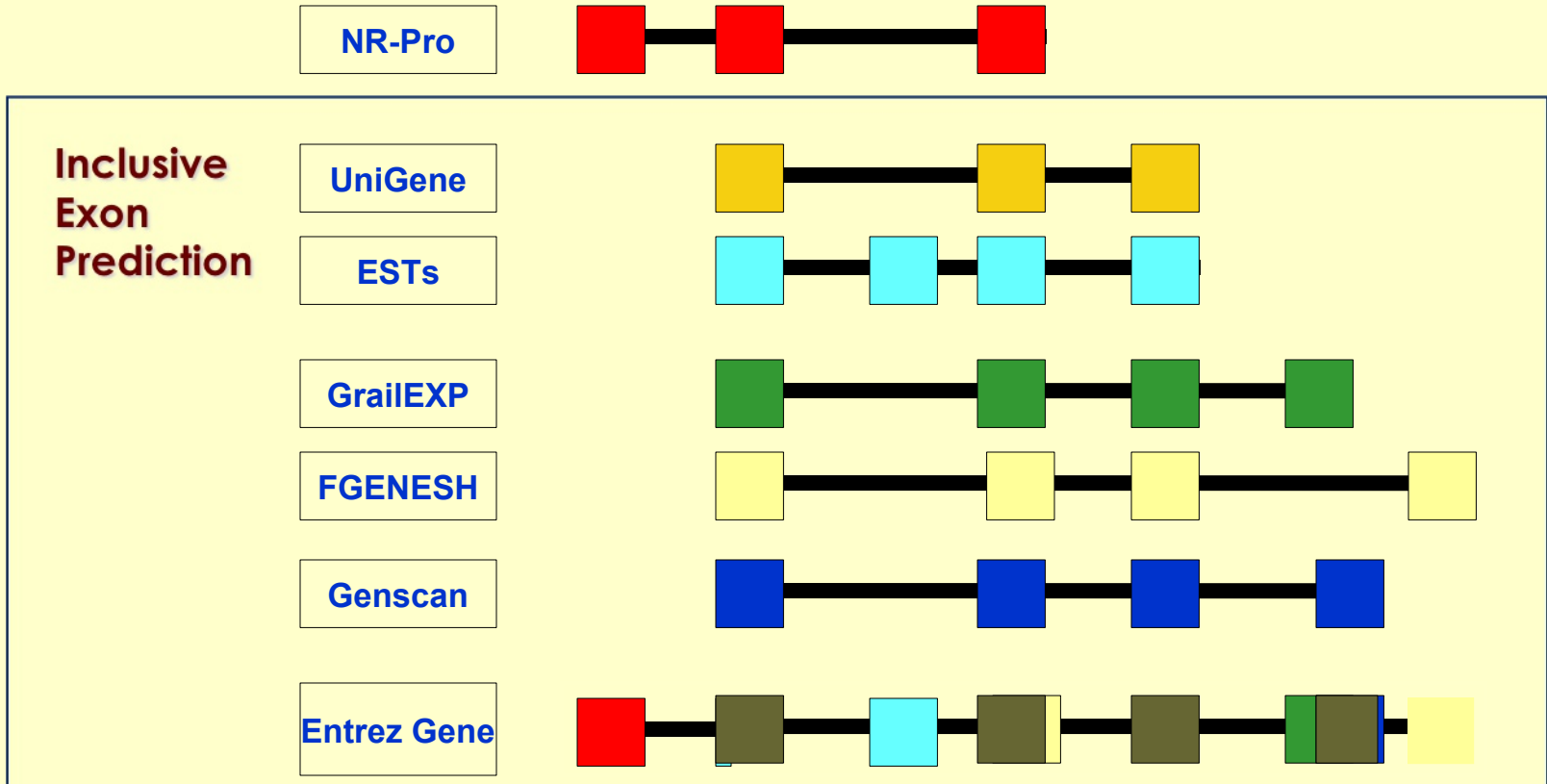


Genome Databases

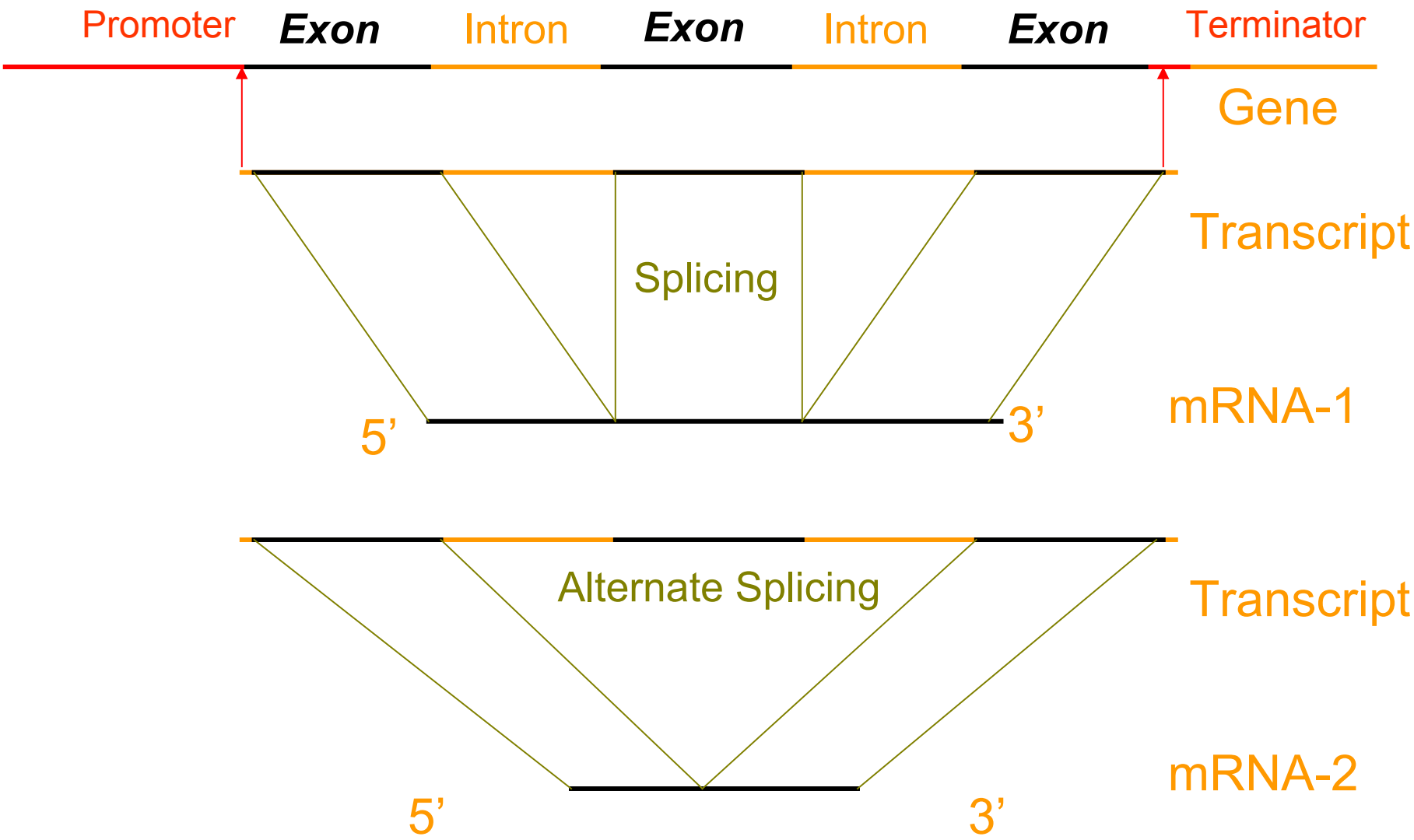


Entrez Gene Loci

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene



Alternative Splicing Generates Distinct Proteins in Different Tissues



NCBI Genomes

http://www.ncbi.nlm.nih.gov/genome

NCBI Resources How To brutlag My NCBI Sign Out

Genome Genome dog Search

Limits Advanced Help

Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

Using Genome

- Help
- Browse by Organism
- Download / FTP
- Submit a genome

Custom resources

- Human Genome
- Microbes
- Organelles
- Plants
- Viruses

Other Resources

- Assembly
- BioProject
- BioSample
- Map Viewer
- Protein Clusters

Genome Tools

- BLAST the Human Genome
- Genomic groups BLAST
- NCBI remap
- Genome Decoration Page

Genome Annotation and Analysis

- Eukaryotic Genome Annotation
- Prokaryotic Genome Annotation
- PASC (Pairwise Sequence Comparison)
- TaxPlot (3-way Genome Comparison)

External Resources

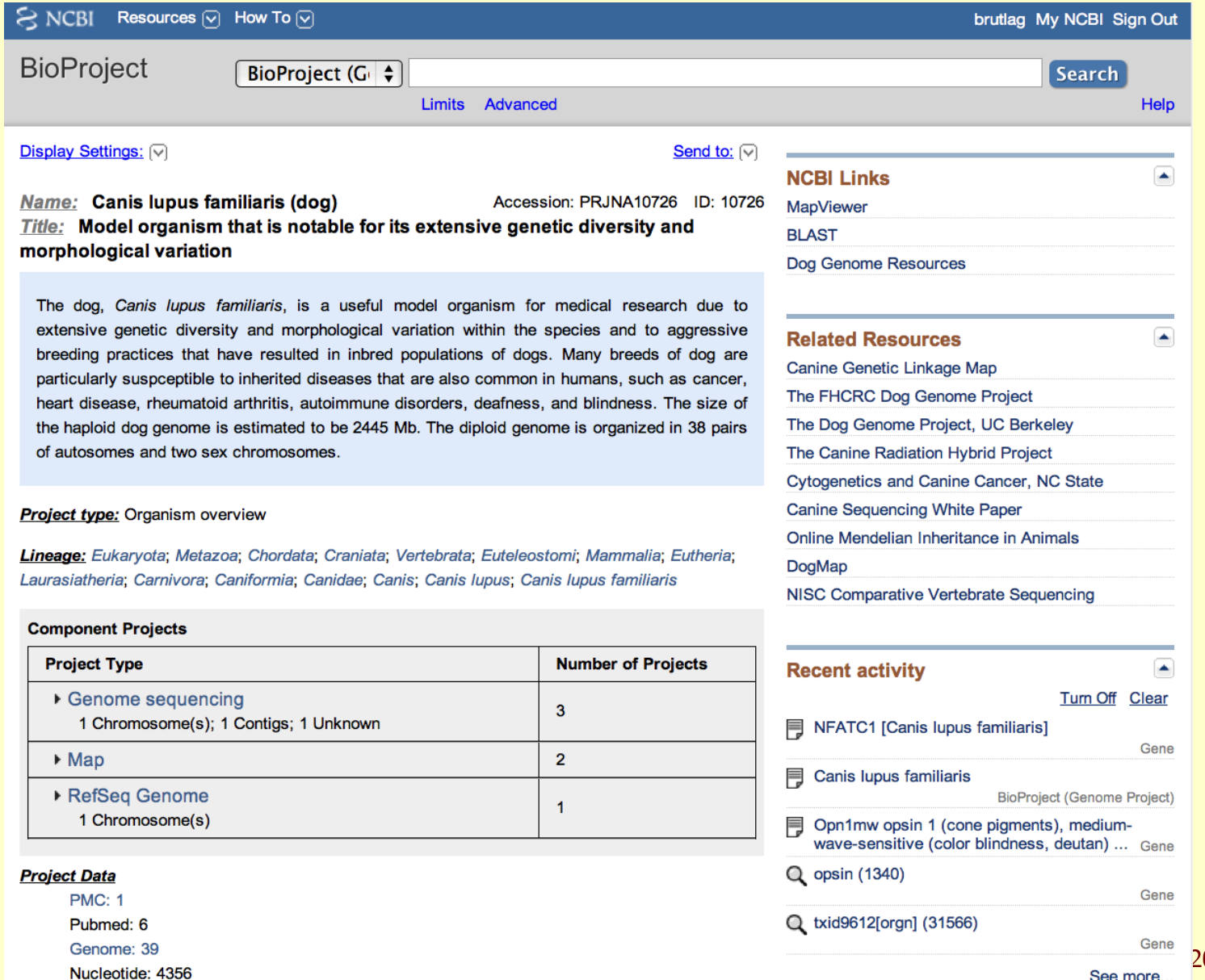
- GOLD - Genomes Online Database
- Ensembl Genome Browser
- Bacteria Genomes at Sanger
- Large-Scale Genome Sequencing (NHGRI)

Eukaryote Genome s

http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi

Canis lupus familiaris Genome

http://www.ncbi.nlm.nih.gov/sites/entrez?db=bioproject&cmd=Retrieve&dopt=O



NCBI Resources How To brutlag My NCBI Sign Out

BioProject BioProject (G Search

Limits Advanced Help

Display Settings: Send to:

Name: **Canis lupus familiaris (dog)** Accession: PRJNA10726 ID: 10726

Title: **Model organism that is notable for its extensive genetic diversity and morphological variation**

The dog, *Canis lupus familiaris*, is a useful model organism for medical research due to extensive genetic diversity and morphological variation within the species and to aggressive breeding practices that have resulted in inbred populations of dogs. Many breeds of dog are particularly suspceptible to inherited diseases that are also common in humans, such as cancer, heart disease, rheumatoid arthritis, autoimmune disorders, deafness, and blindness. The size of the haploid dog genome is estimated to be 2445 Mb. The diploid genome is organized in 38 pairs of autosomes and two sex chromosomes.

Project type: Organism overview

Lineage: *Eukaryota*; *Metazoa*; *Chordata*; *Craniata*; *Vertebrata*; *Euteleostomi*; *Mammalia*; *Eutheria*; *Laurasiatheria*; *Carnivora*; *Caniformia*; *Canidae*; *Canis*; *Canis lupus*; *Canis lupus familiaris*

Component Projects

Project Type	Number of Projects
Genome sequencing 1 Chromosome(s); 1 Contigs; 1 Unknown	3
Map	2
RefSeq Genome 1 Chromosome(s)	1

Project Data

PMC: 1
Pubmed: 6
Genome: 39
Nucleotide: 4356

NCBI Links

MapViewer
BLAST
Dog Genome Resources

Related Resources

Canine Genetic Linkage Map
The FHCRC Dog Genome Project
The Dog Genome Project, UC Berkeley
The Canine Radiation Hybrid Project
Cytogenetics and Canine Cancer, NC State
Canine Sequencing White Paper
Online Mendelian Inheritance in Animals
DogMap
NISC Comparative Vertebrate Sequencing

Recent activity

Turn Off Clear

NFATC1 [Canis lupus familiaris] Gene
Canis lupus familiaris BioProject (Genome Project)
Opn1mw opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan) ... Gene
opsin (1340) Gene
txid9612[orgn] (31566) Gene

See more

NCBI Entrez Gene

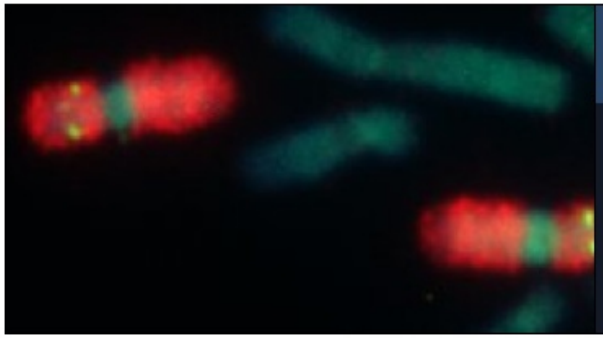
http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene

NCBI Resources How To brutlag My NCBI Sign Out

Gene

Gene human opsin Search

Limits Advanced Help



Welcome to Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene

- Gene Quick Start
- FAQ
- Download/FTP
- RefSeq Mailing List
- Gene News

Gene Tools

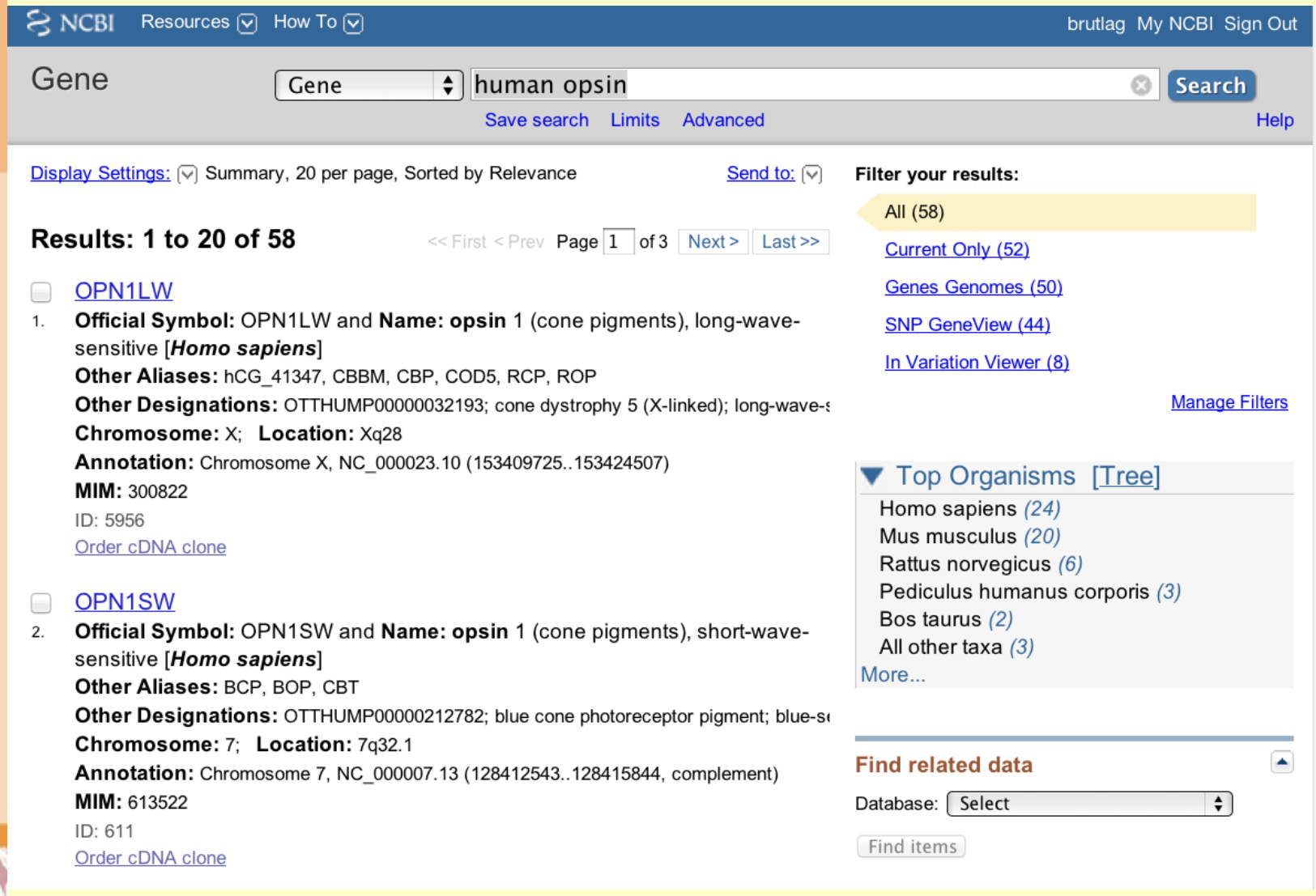
- Submit GeneRIFs
- Submit Correction
- Statistics
- BLAST
- Genome Workbench
- Splign

Other Resources

- HomoloGene
- OMIM
- RefSeq
- RefSeqGene
- UniGene
- Protein Clusters

NCBI Entrez Gene: Human Opsin

http://www.ncbi.nlm.nih.gov/gene?term=human%20opsin



Entrez Gene: Human Opsin OPN1MW

http://www.ncbi.nlm.nih.gov/gene/2652

NCBI Resources How To brutlag My NCBI

Gene Gene Search

Limits Advanced Help

Display Settings: Full Report Send to:

OPN1MW opsin 1 (cone pigments), medium-wave-sensitive [*Homo sapiens*]

Gene ID: 2652, updated on 29-Sep-2011

Summary

Official Symbol	OPN1MW provided by HGNC
Official Full Name	opsin 1 (cone pigments), medium-wave-sensitive provided by HGNC
Primary source	HGNC:4206
See related	Ensembl:ENSG00000147380; HPRD:02365; MIM:300821
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	CBD; GCP; GOP; CBBM; COD5; OPN1MW1; OPN1MW2; MGC176615; MGC177321; MGC198468; MGC198469
Summary	This gene encodes for a light absorbing visual pigment of the opsin gene family. The encoded protein is called green cone photopigment or medium-wavelength sensitive opsin. Opsins are G-protein coupled receptors with seven transmembrane domains, an N-terminal extracellular domain, and a C-terminal cytoplasmic domain. The long-wavelength opsin gene and multiple copies of the medium-wavelength opsin gene are tandemly arrayed on the X chromosome and frequent unequal recombination and gene conversion may occur between these sequences. X chromosomes may have fusions of the medium- and long-wavelength opsin genes or may have more than one copy of these genes. Defects in this gene are the cause of deutanopic colorblindness. [provided by RefSeq, Mar 2009]

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- General gene info
- General protein info
- Reference sequences
- Related sequences
- Additional links

Links

- Order cDNA clone
- BioAssay, by Gene target
- BioAssays, Gene target, Active
- BioProjects
- BioSystems
- Books
- CCDS
- Conserved Domains
- dbVar
- Full text in PMC
- Genome

Entrez Gene: Human Opsin OPN1MW

http://www.ncbi.nlm.nih.gov/gene/2652

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence NC_000023 chromosome X reference GRCh37.p5 Primary Assembly

Go to nucleotide Graphics FASTA GenBank

153,445,944 : 153,464,491 (18,548 bases shown, positive strand) Open Full View

Flip Strands | Configure

PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
RefSeq Proteins
RefSeq RNAs
RefSeqGene
SNP
SNP: GeneView
SNP: Genotype
SNP: VarView
Taxonomy
UniGene
UniSTS

Links to other resources

AceView
Ensembl
Evidence Viewer
GeneTests for MIM: 300821
GeneTests for MIM: 303800
HGNC

NCBI

Human genome overview page (Build 37.3)
Human genome overview page (Build 36.3)
Map Viewer Home

Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Region Shown:
153,378K
153,533K
Go

out
zoom
in

You are here:

Ideogram

default
master

NCBI Map Viewer

PubMed | Entrez | BLAST | OMIM | Taxonomy | Structure

Search | Find | Find in This View | Advanced Search

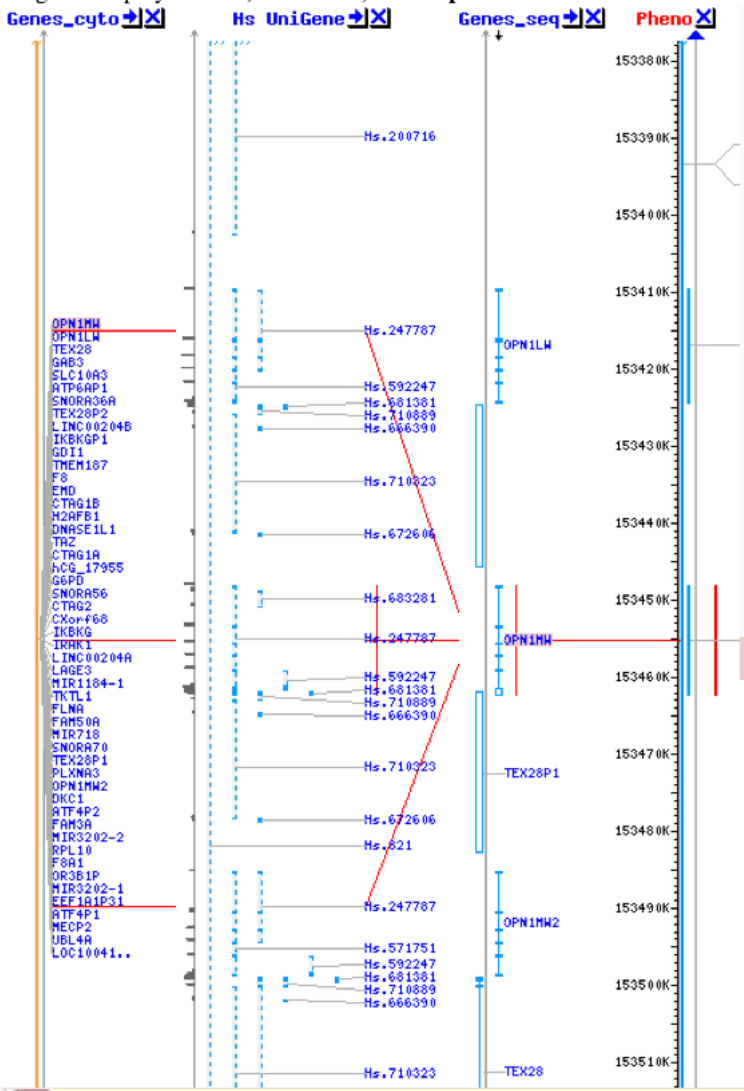
Homo sapiens (human) Build 37.3 (Current) — BLAST human sequences

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 [X] Y MT

Query: 2652[gene_id] [clear]

Master Map: Phenotype (includes QTLs) — Summary of Maps — Maps & Options

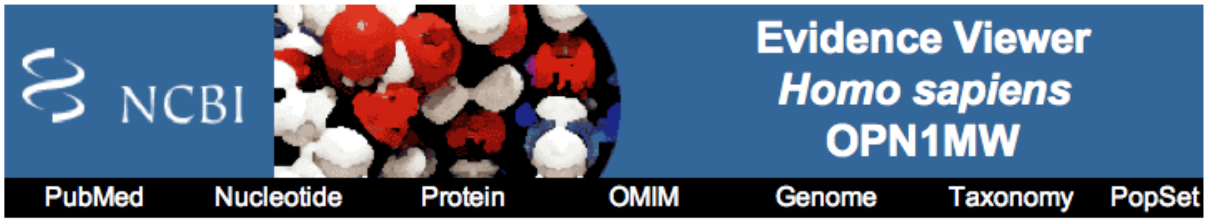
Region Displayed: **153,378K-153,533K bp** — Download/View Sequence/Evidence



Symbol	Go to	Description
BMND4	Info	Bone mineral density QTL 4
ODPF	Info	Osseous dysplasia, digital, with facial pigmentary defects and multiple f
OPN1LW	Info	
OPN1MW	Info	Colorblindness, deutan, 303800; Blue cone monochromacy, 303700

Evidence Viewer for OPN1MW

http://www.ncbi.nlm.nih.gov/sutils/evv.cgi?taxid=9606&contig=NT_167198.1&gene=OPN1MW



Key for display of mRNAs aligning in this region:

MapView
Evidence Viewer Help

- Genomic sequence (C)
- model exons, single (M)
- mRNA exons, single (G, R)
- model exons, overlapping (M)
- mRNA exons, overlapping (G, R)

C = contig; M = model mRNA; R = RefSeq mRNA; G = GenBank mRNA

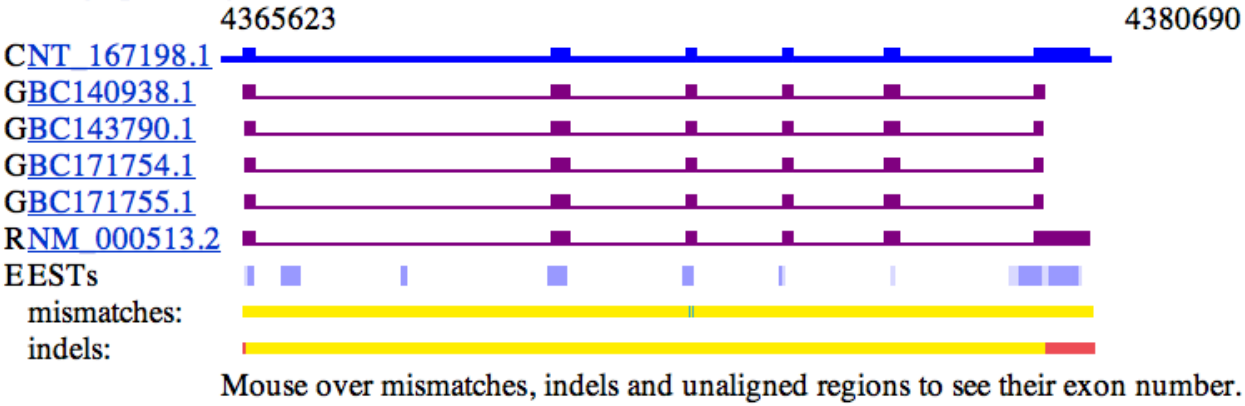
R = new since last genome build; R = updated since last genome build

EST density key (E):

- 1 EST
- 2-5 ESTs
- 6-20 ESTs
- 21-99 ESTs
- >100 ESTs

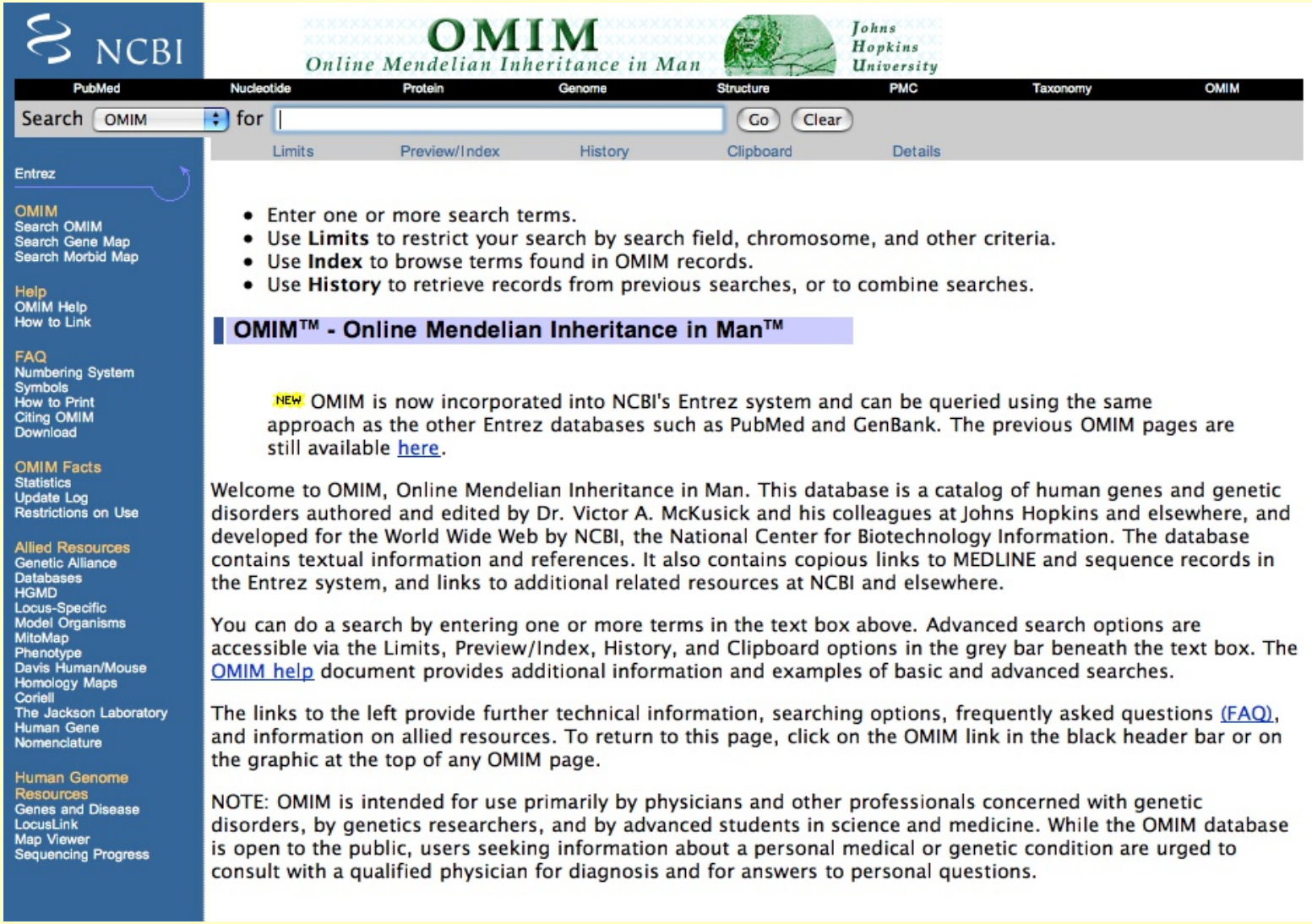
6 exons and 1 gene found in this genomic region spanning 15068 bp.

View graphic only

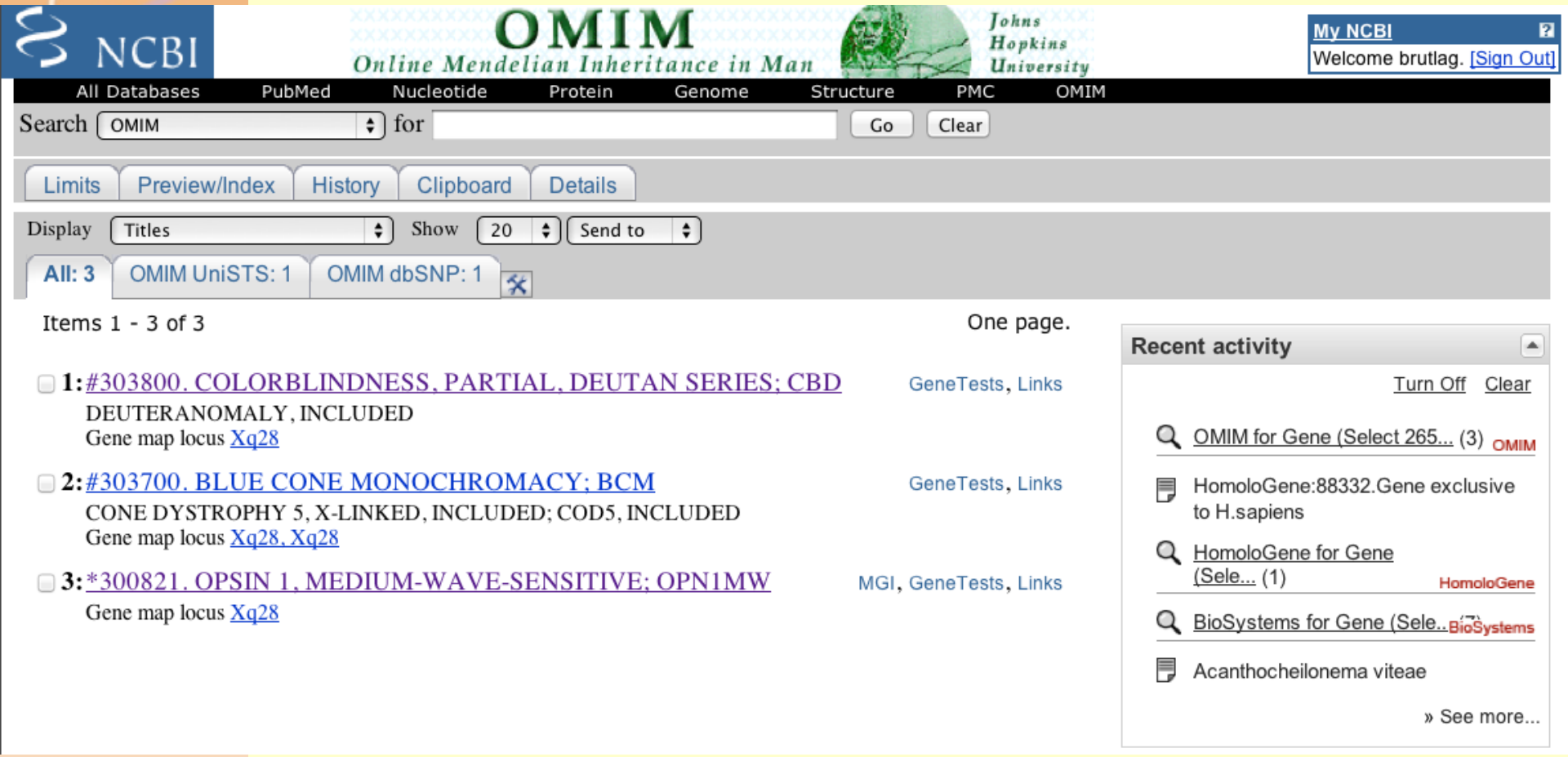


OMIM Home Page

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM



Colorblindness in OMIM



Colorblindness in OMIM

http://omim.org/entry/303800

#303800

COLORBLINDNESS, PARTIAL, DEUTAN SERIES; CBD

Alternative titles; symbols

DEUTAN COLORBLINDNESS; DCB
DEUTERANOPIA
GREEN COLORBLINDNESS

Other entities represented in this entry:

DEUTERANOMALY, INCLUDED

Phenotype Gene Relationships

Location	Phenotype	Phenotype MIM number	Gene/Locus	Gene/Locus MIM number
Xq28	Colorblindness, deutan	303800	OPN1MW	300821

Clinical Synopsis

TEXT

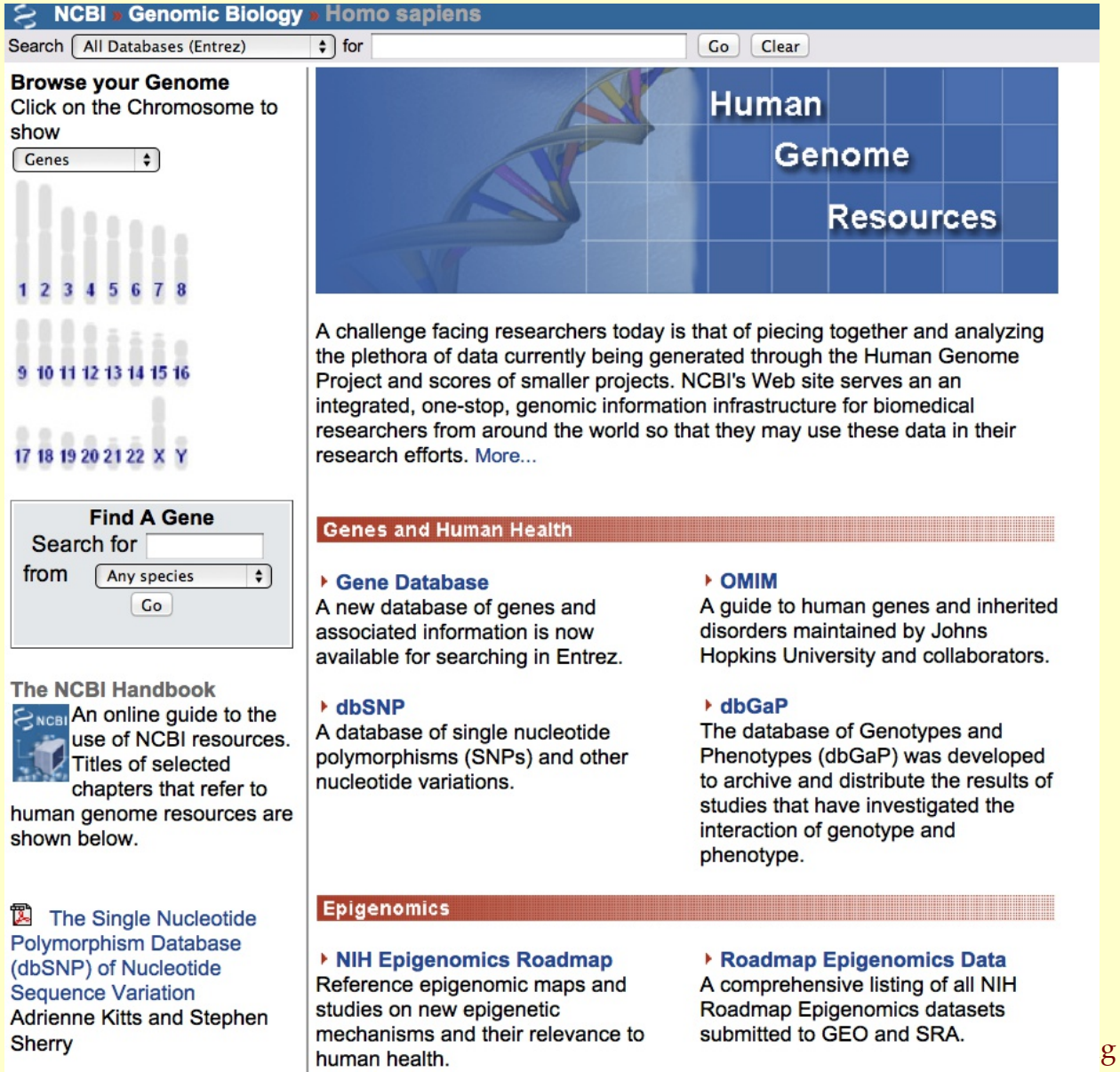
A number sign (#) is used with this entry because deutan colorblindness is caused by mutation in the OPN1MW gene (300821), which encodes green cone pigment.

Description

Normal color vision in humans is trichromatic, being based on 3 classes of cone that are maximally sensitive to light at approximately 420 nm (blue cones; 613522), 530 nm (green cones; 300821), and 560 nm (red cones; 300822). Comparison by neural circuits of light absorption by the 3 classes of cone photoreceptors allows perception of red, yellow, green, and blue colors individually or in various combinations. Dichromatic color vision is severely defective color vision based on the use of only 2 types of photoreceptors, blue plus green (protanopia; see 303900) or blue plus red (deuteranopia). Anomalous trichromacy is trichromatic color vision based on a blue, green, and an anomalous red-like photoreceptor (protanomaly), or a blue, red, and an anomalous green-like photoreceptor (deuteranomaly). The color vision defect is generally mild but may in

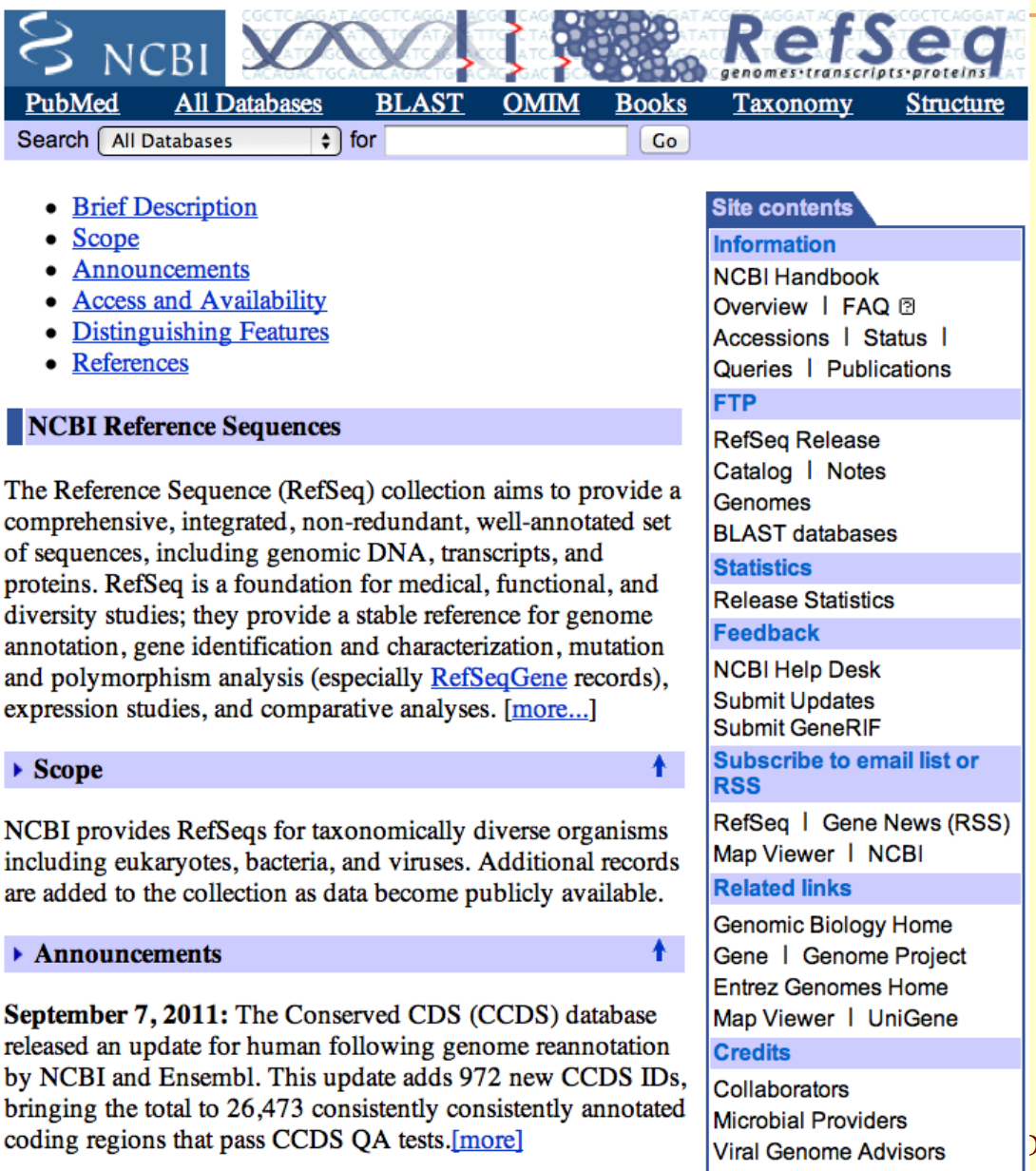
Human Genome Resources

http://www.ncbi.nlm.nih.gov/genome/guide/human/



RefSeq

http://www.ncbi.nlm.nih.gov/RefSeq/

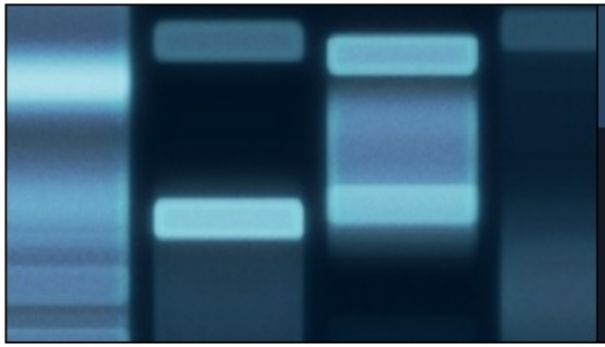


RefSeq Gene

http://www.ncbi.nlm.nih.gov/refseq/rsg/

NCBI Resources How To brutlag My NCBI Sign Out

RefSeqGene Gene Search



RefSeqGene

RefSeqGene defines genomic sequences to be used as reference standards for well-characterized genes and is part of the Locus Reference Genomic (LRG) Project.

Using RefSeqGene

- About
- Browse Genes with RefSeqGene Sequences
- See RefSeqGene Sequences in the Nucleotide Database
- Download Data
- New RefSeqGene Sequences (last 30 days)

Tools

- Clinical Remap
- Genome Workbench
- Map Viewer
- RefSeqGene BLAST

Related Sites

- Entrez Gene
- RefSeq
- Locus Reference Genomic (LRG)
- Variation Databases (dbSNP, dbVar)
- Online Mendelian Inheritance In Man
- Contact Us

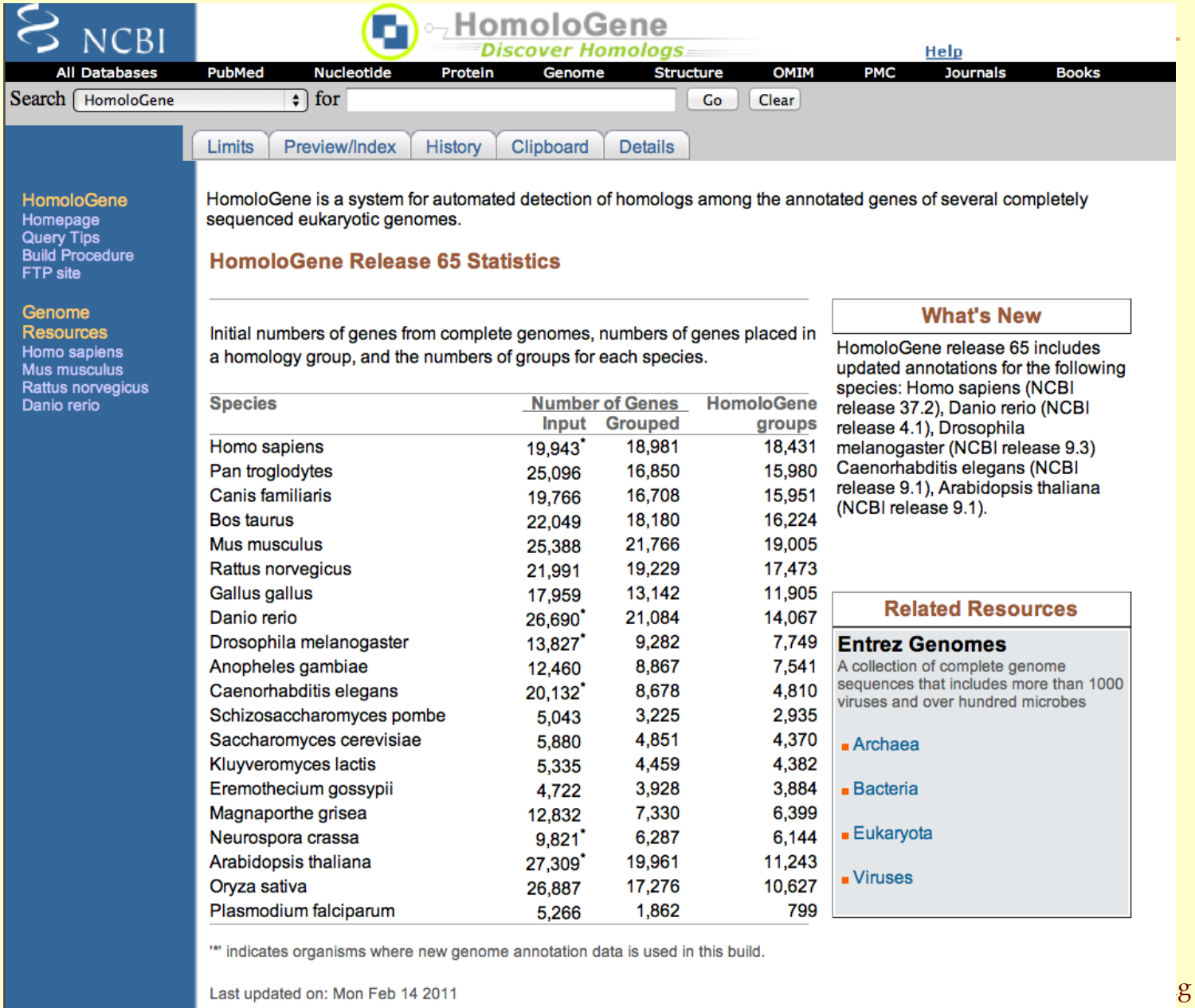
UniGene: An Organized View of the Transcriptome.

Each UniGene entry is a set of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location.

Species	UniGene Entries
Chordata	
Mammalia	
Bos taurus (cow)	43,448
Canis lupus familiaris (dog)	27,853
Equus caballus (horse)	8,348
Homo sapiens (human)	123,891
Macaca fascicularis (crab-eating macaque)	11,980
Macaca mulatta (rhesus monkey)	15,359
Monodelphis domestica (gray short-tailed opossum)	359
Mus musculus (mouse)	79,119
Ornithorhynchus anatinus (platypus)	1,831
Oryctolagus cuniculus (rabbit)	6,576
Ovis aries (sheep)	14,659
Papio anubis (olive baboon)	12,110
Pongo abelii (Sumatran orangutan)	6,996
Rattus norvegicus (Norway rat)	65,511
Sus scrofa (pig)	51,670
Trichosurus vulpecula (silver-gray brushtail possum)	11,771
Actinopterygii	
Danio rerio (zebrafish)	56,944
Fundulus heteroclitus (killifish)	4,618
Gadus morhua (Atlantic cod)	14,542
Gasterosteus aculeatus (three spined stickleback)	18,938
Oncorhynchus mykiss (rainbow trout)	25,025
Oryzias latipes (Japanese medaka)	22,552
Pimephales promelas (fathead minnow)	21,765
Salmo salar (Atlantic salmon)	36,314
Takifugu rubripes (pufferfish)	3,809
Amphibia	
Xenopus laevis (African clawed frog)	35,077

NCBI Homologene Database

http://www.ncbi.nlm.nih.gov/homologene



NCBI

HomoloGene
Discover Homologs

Help

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search HomoloGene for Go Clear

Limits | Preview/Index | History | Clipboard | Details

HomoloGene
Homepage
Query Tips
Build Procedure
FTP site

Genome Resources
Homo sapiens
Mus musculus
Rattus norvegicus
Danio rerio

HomoloGene is a system for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes.

HomoloGene Release 65 Statistics

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species.

Species	Number of Genes		HomoloGene groups
	Input	Grouped	
Homo sapiens	19,943*	18,981	18,431
Pan troglodytes	25,096	16,850	15,980
Canis familiaris	19,766	16,708	15,951
Bos taurus	22,049	18,180	16,224
Mus musculus	25,388	21,766	19,005
Rattus norvegicus	21,991	19,229	17,473
Gallus gallus	17,959	13,142	11,905
Danio rerio	26,690*	21,084	14,067
Drosophila melanogaster	13,827*	9,282	7,749
Anopheles gambiae	12,460	8,867	7,541
Caenorhabditis elegans	20,132*	8,678	4,810
Schizosaccharomyces pombe	5,043	3,225	2,935
Saccharomyces cerevisiae	5,880	4,851	4,370
Kluyveromyces lactis	5,335	4,459	4,382
Eremothecium gossypii	4,722	3,928	3,884
Magnaporthe grisea	12,832	7,330	6,399
Neurospora crassa	9,821*	6,287	6,144
Arabidopsis thaliana	27,309*	19,961	11,243
Oryza sativa	26,887	17,276	10,627
Plasmodium falciparum	5,266	1,862	799

'*' indicates organisms where new genome annotation data is used in this build.

Last updated on: Mon Feb 14 2011

What's New

HomoloGene release 65 includes updated annotations for the following species: Homo sapiens (NCBI release 37.2), Danio rerio (NCBI release 4.1), Drosophila melanogaster (NCBI release 9.3) Caenorhabditis elegans (NCBI release 9.1), Arabidopsis thaliana (NCBI release 9.1).

Related Resources

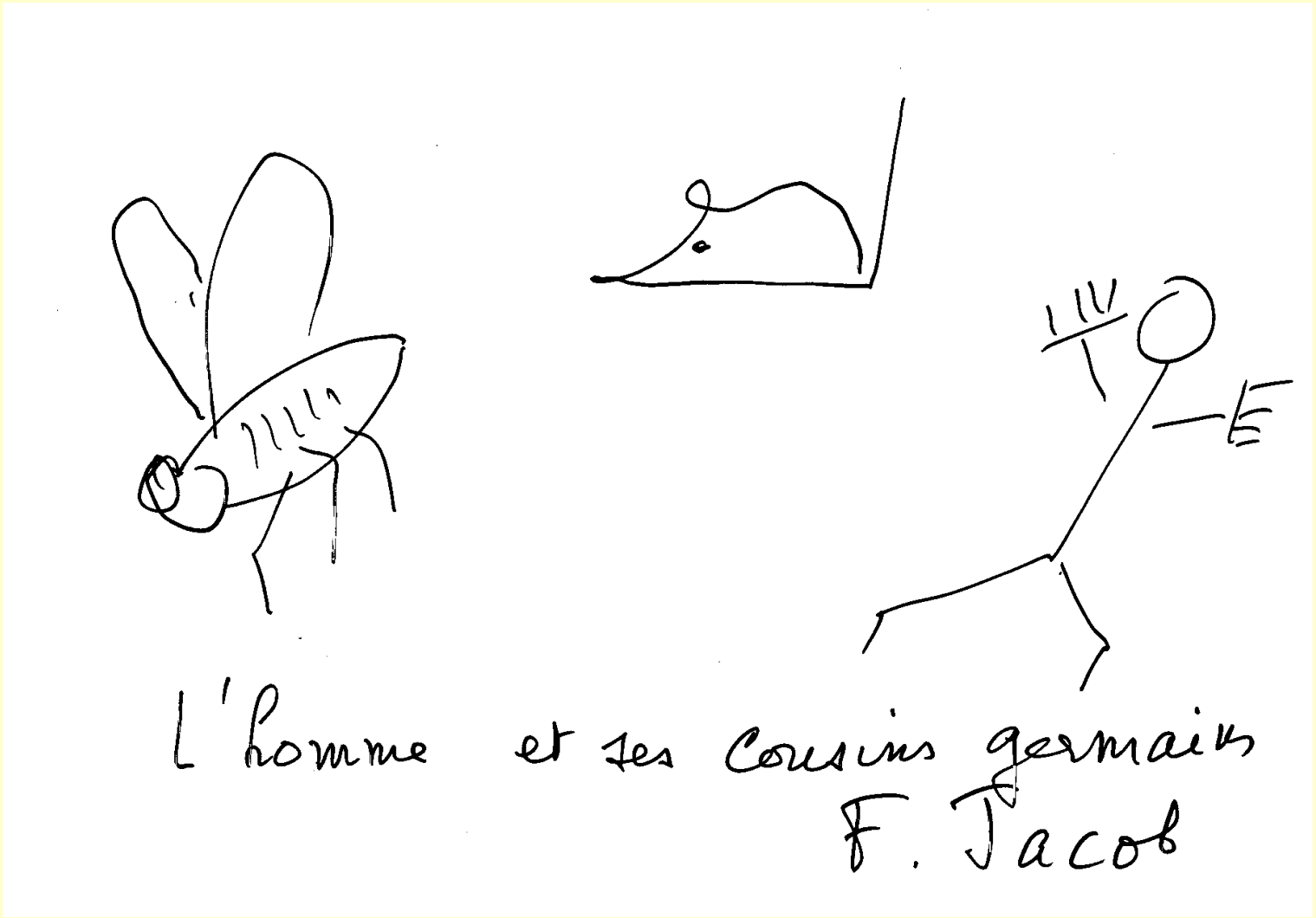
Entrez Genomes

A collection of complete genome sequences that includes more than 1000 viruses and over hundred microbes

- Archaea
- Bacteria
- Eukaryota
- Viruses

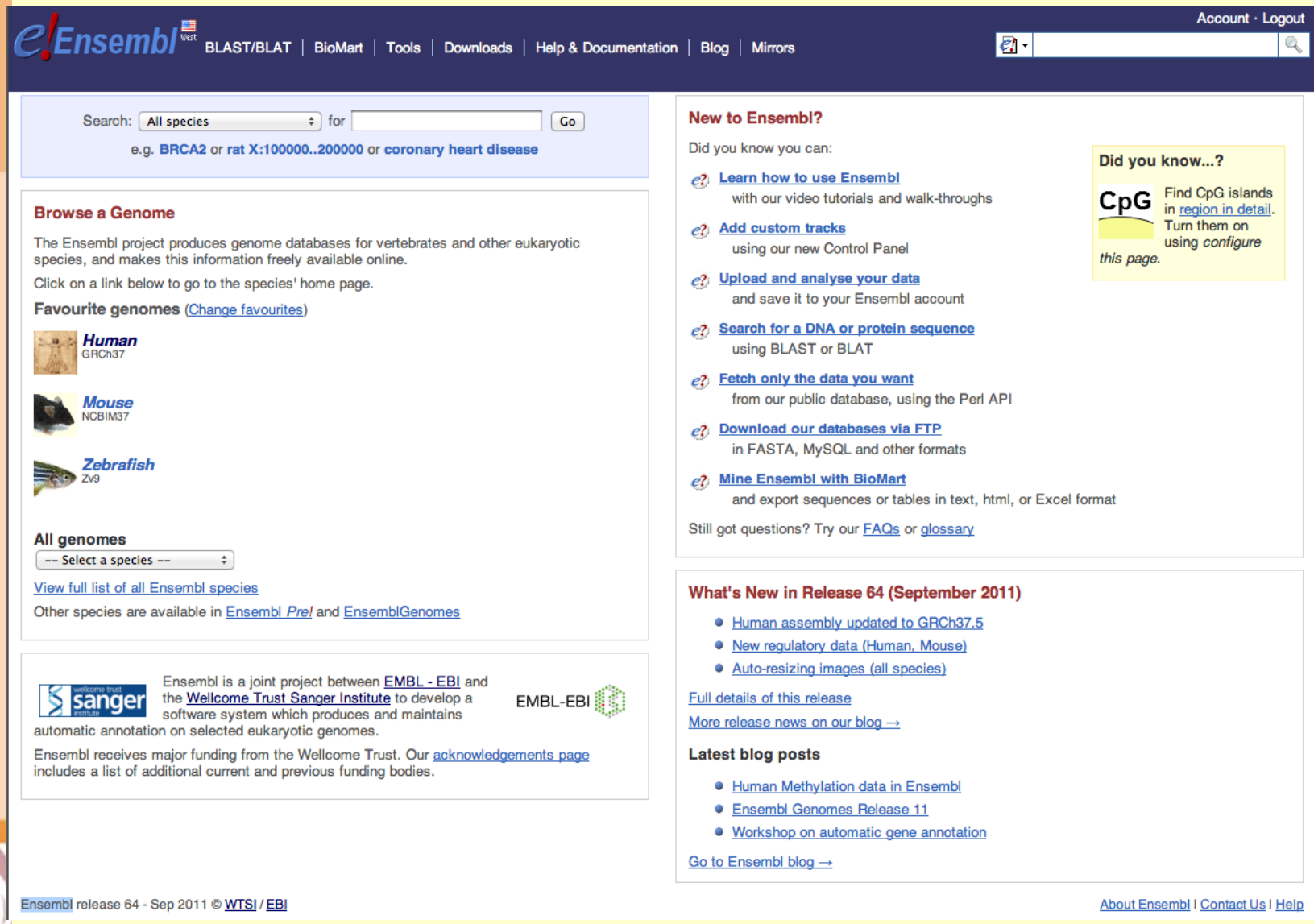
g 2014

Comparative Genomics



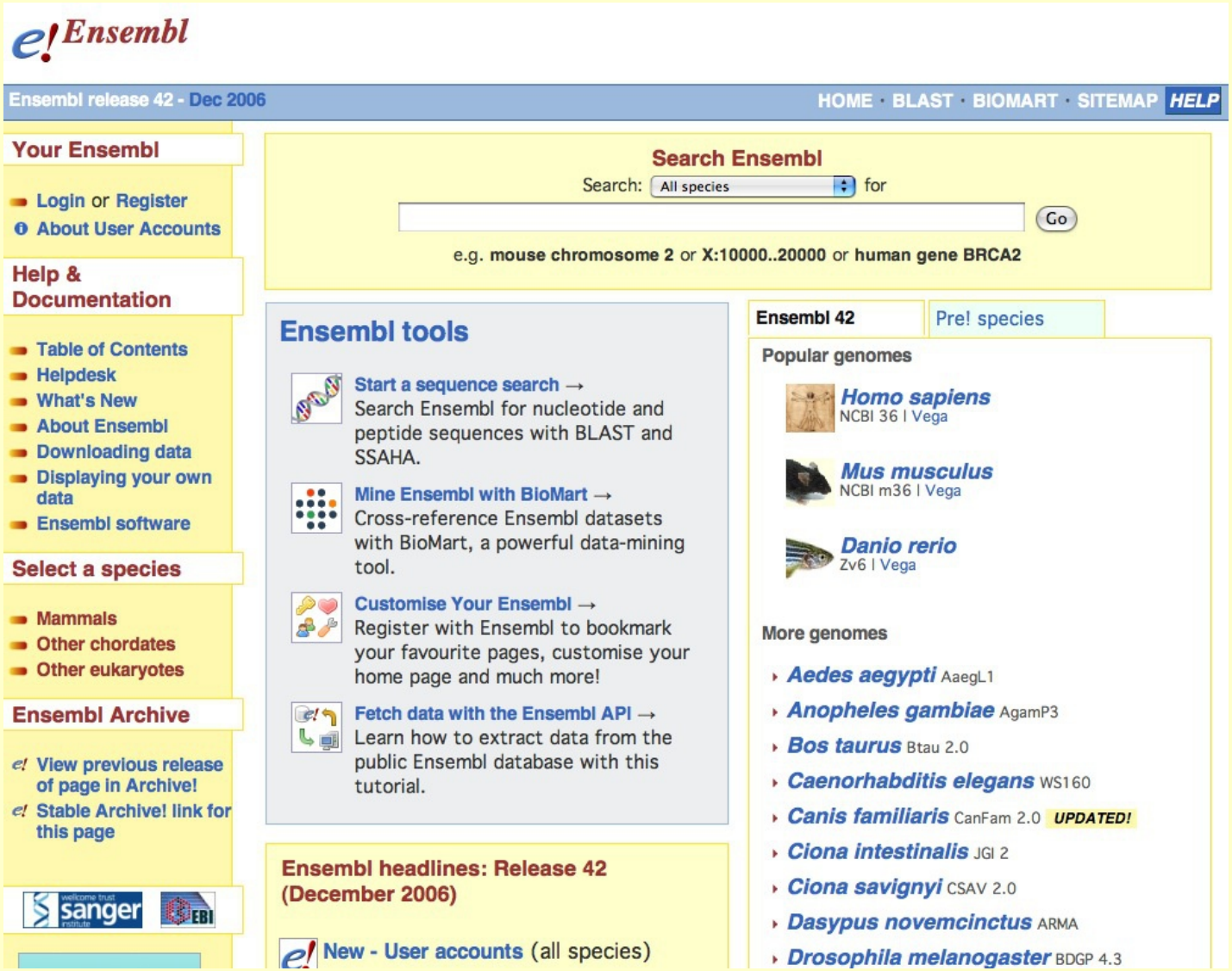
Ensembl Home Page

http://www.ensembl.org/



EBI Genomes Home Page

http://www.ensembl.org/



Ensembl Human Genome

http://www.ensembl.org/Homo_sapiens/

Account · Logout

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37)

About this species
- Description
- Genome Statistics
  - Assembly and Genebuild
  - Top 40 InterPro hits
  - Top 500 InterPro hits
- What's New
- Sample entry points
  - Karyotype
  - Location (6:133017695-1331
  - Gene (BRCA2)
  - Transcript (FOXP2-203)
  - Variation (rs1333049)
  - Regulation (ENSR00001348

Configure this page

Manage your data

Export data

Bookmark this page

Search Ensembl Human

Search for: opsin Go

e.g. BRCA2 or 6:133017695-133161157 or osteoarthritis

Description

Human (*Homo sapiens*)

Assembly

This site provides a data set based on the February 2009 *Homo sapiens* high coverage assembly GRCh37 (GCA_000001405.6) from the Genome Reference Consortium. This assembly is used by UCSC to create their hg19 database. The data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model of exonerate.

This release of the assembly has the following properties:

- 27478 contigs.
- contig length total 3.2 Gb.
- chromosome length total 3.1 Gb.

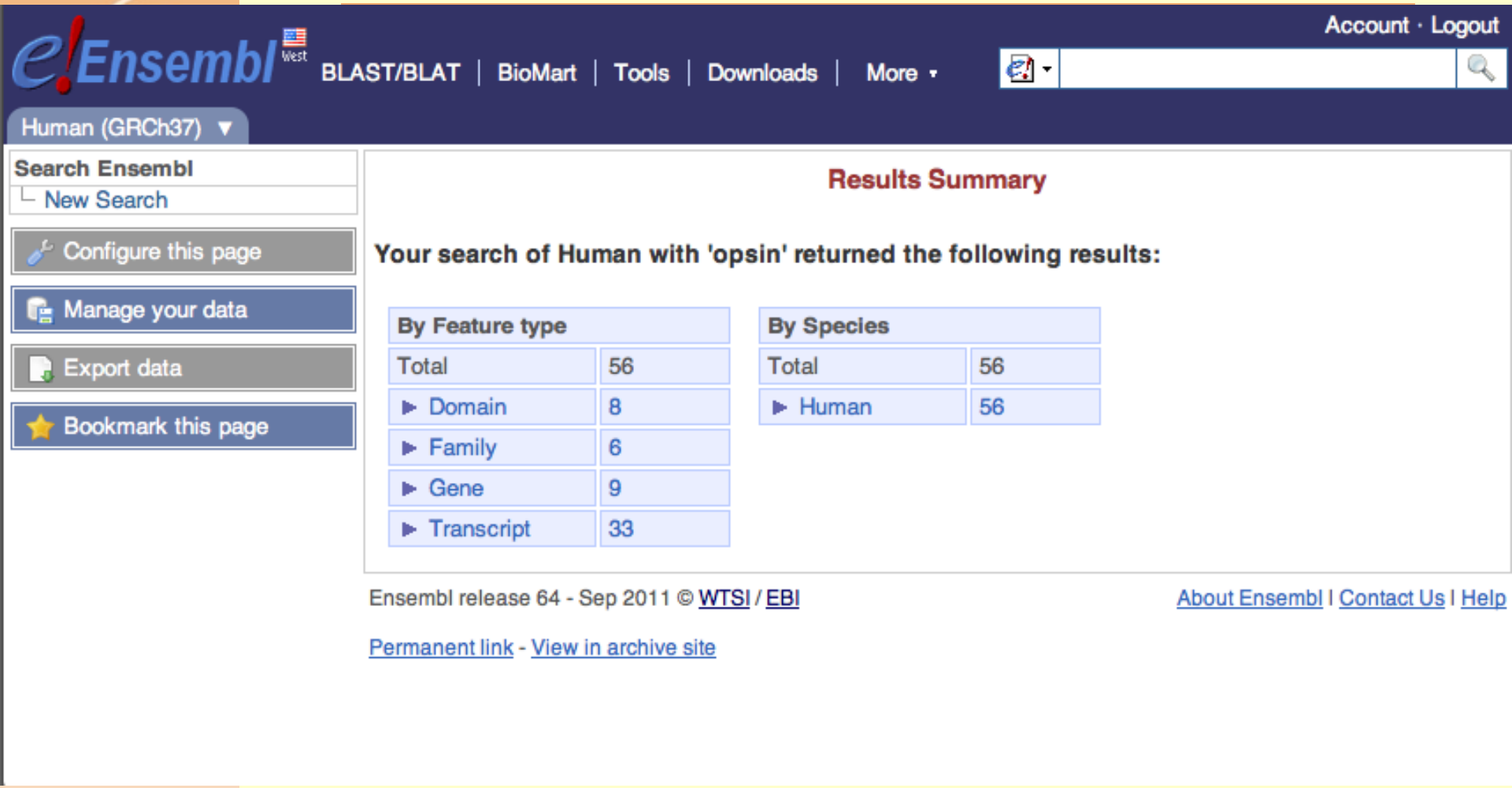
It also includes nine haplotypic regions, mainly in the MHC region of chromosome 6.

As the GRC maintains and improves the assembly, patches are being introduced. Patch release five (GRCh37.p5) was included in Ensembl release 64. Currently, assembly patches are of two types:

- Novel patch: new sequences that add alternative sequence at a loci and will remain as haplotypes in the next major assembly release by GRC
- Fix patch: sequences that correct the reference sequence and will replace the given region of the reference assembly at the next major assembly release by GRC

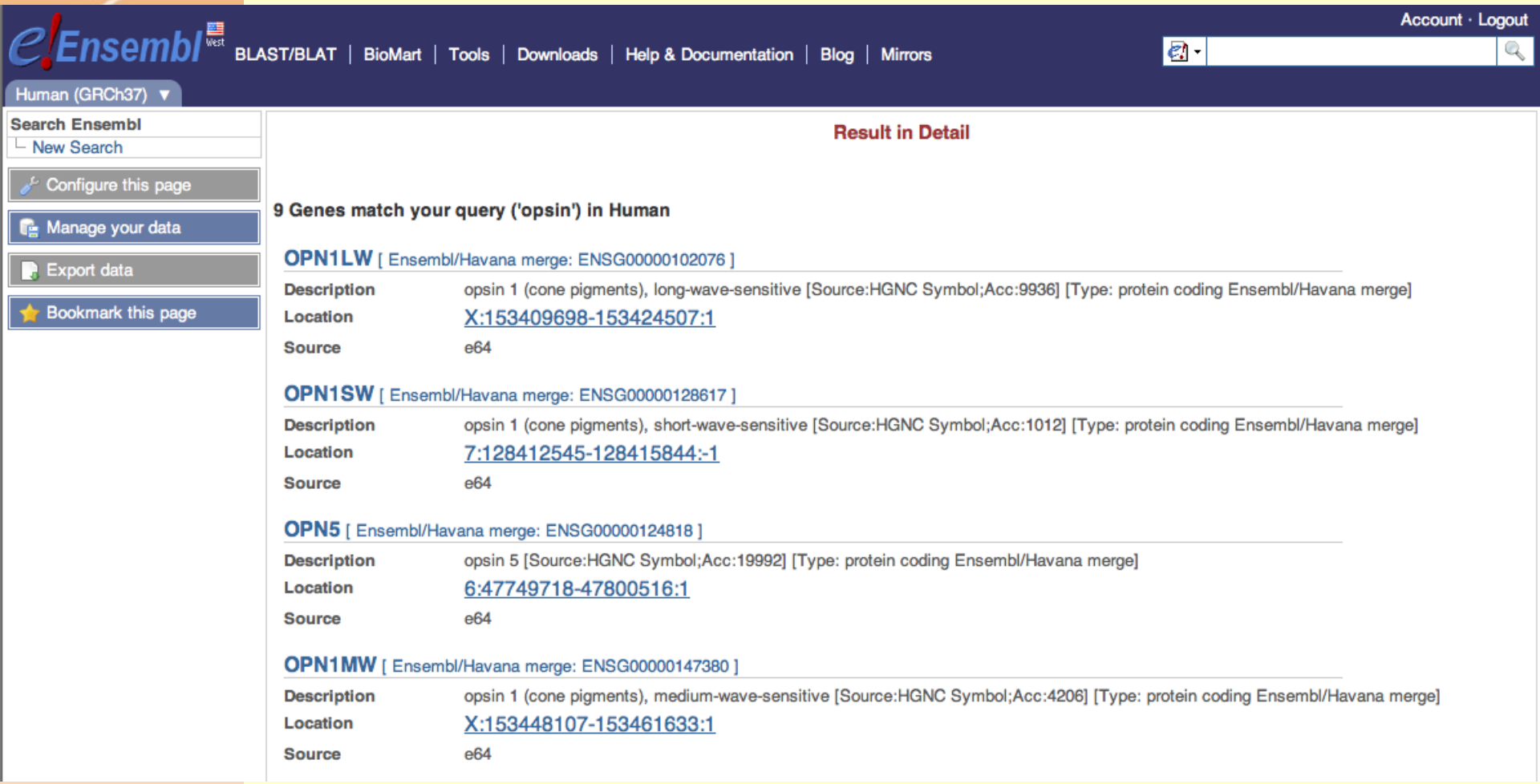
Ensembl Human Opsin Search

http://uswest.ensembl.org/Homo_sapiens/Search/Results?species=Homo_sapiens;idx=;q



Ensembl Human Opsin Genes

http://uswest.ensembl.org/Homo_sapiens/Search/Results?species=Homo_sapiens;idx=;q



Configure this page
Manage your data
Export data
Bookmark this page

Gene: OPN1MW (ENSG00000147380)

Description opsin 1 (cone pigments), medium-wave-sensitive [Source:HGNC Symbol;Acc:4206]

Location Chromosome X: 153,448,107-153,461,633 forward strand.

Transcripts There are 3 transcripts in this gene

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
OPN1MW-001	ENST00000369935	1257	ENSP00000358951	364	Protein coding	CCDS14743
OPN1MW-002	ENST00000430054	590	ENSP00000394838	164	Protein coding	-
OPN1MW-003	ENST00000468495	692	No protein product	-	Processed transcript	-

Transcript and Gene level displays

In Ensembl we provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary help

Name OPN1MW (HGNC Symbol)

Synonyms CBBM, CBD, COD5, GCP, OPN1MW1 [To view all Ensembl genes linked to the name click here.]

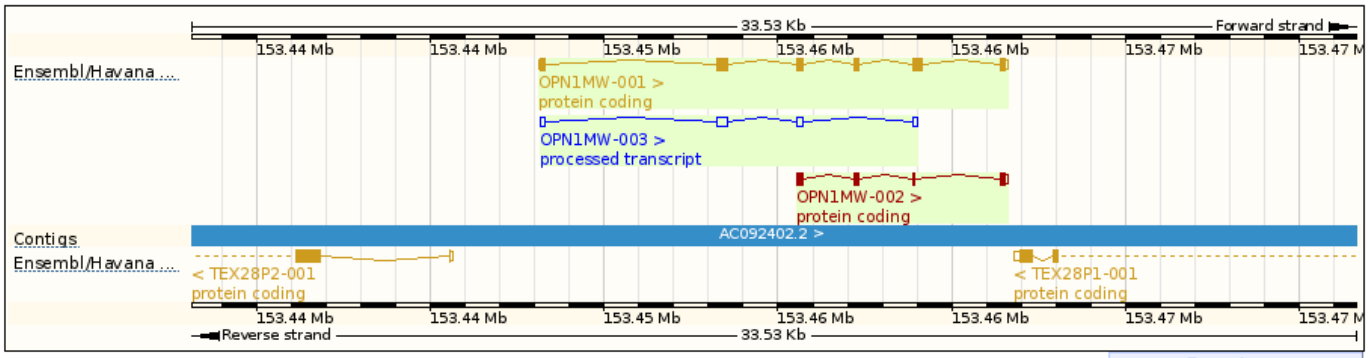
CCDS This gene is a member of the Human CCDS set: CCDS14743

Gene type Known protein coding

Prediction Method Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.

Alternative genes **This gene corresponds to the following database identifiers:**

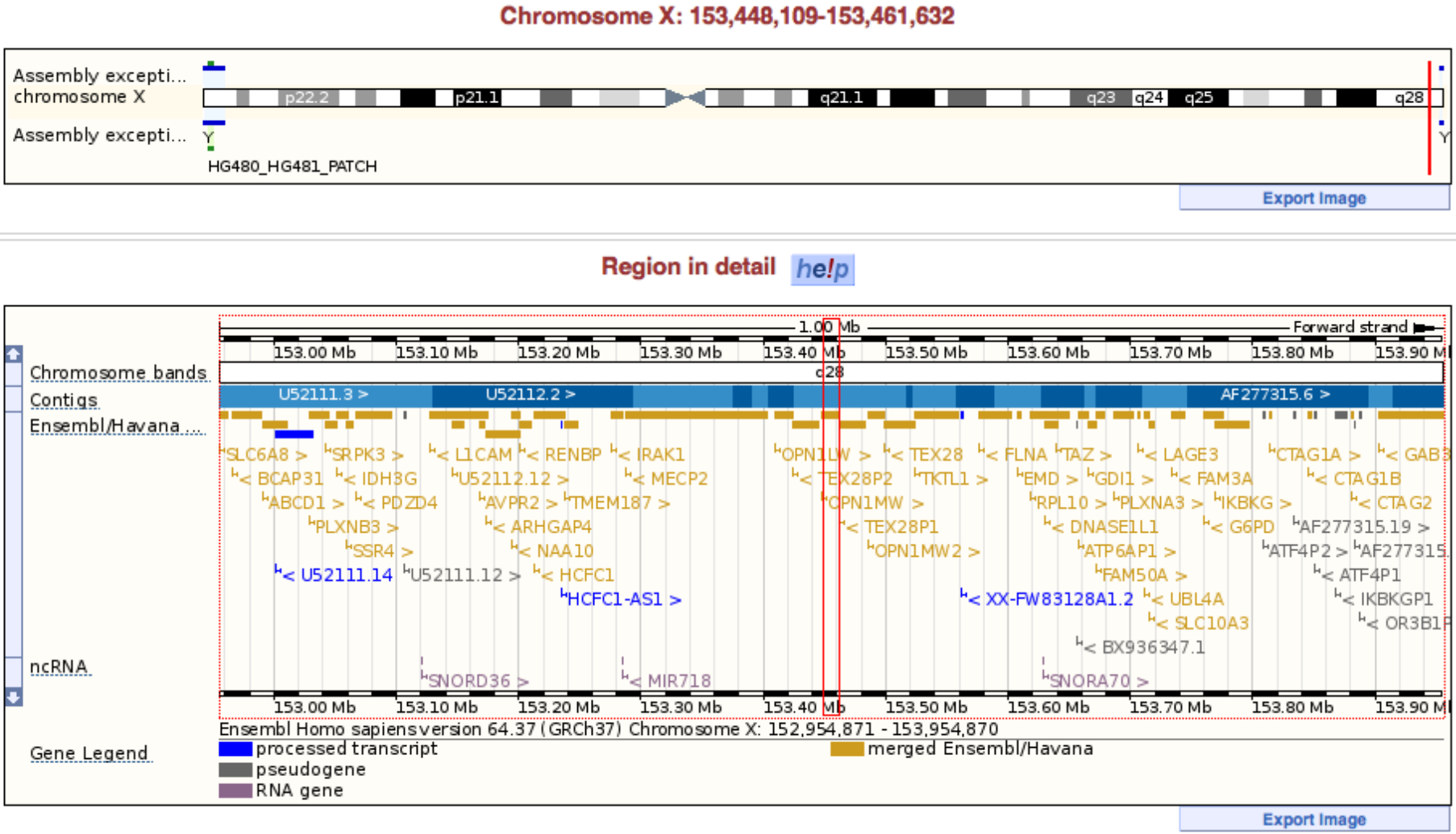
Havana gene: OTTHUMG00000022652 (version 6) [view all locations]



Export Image

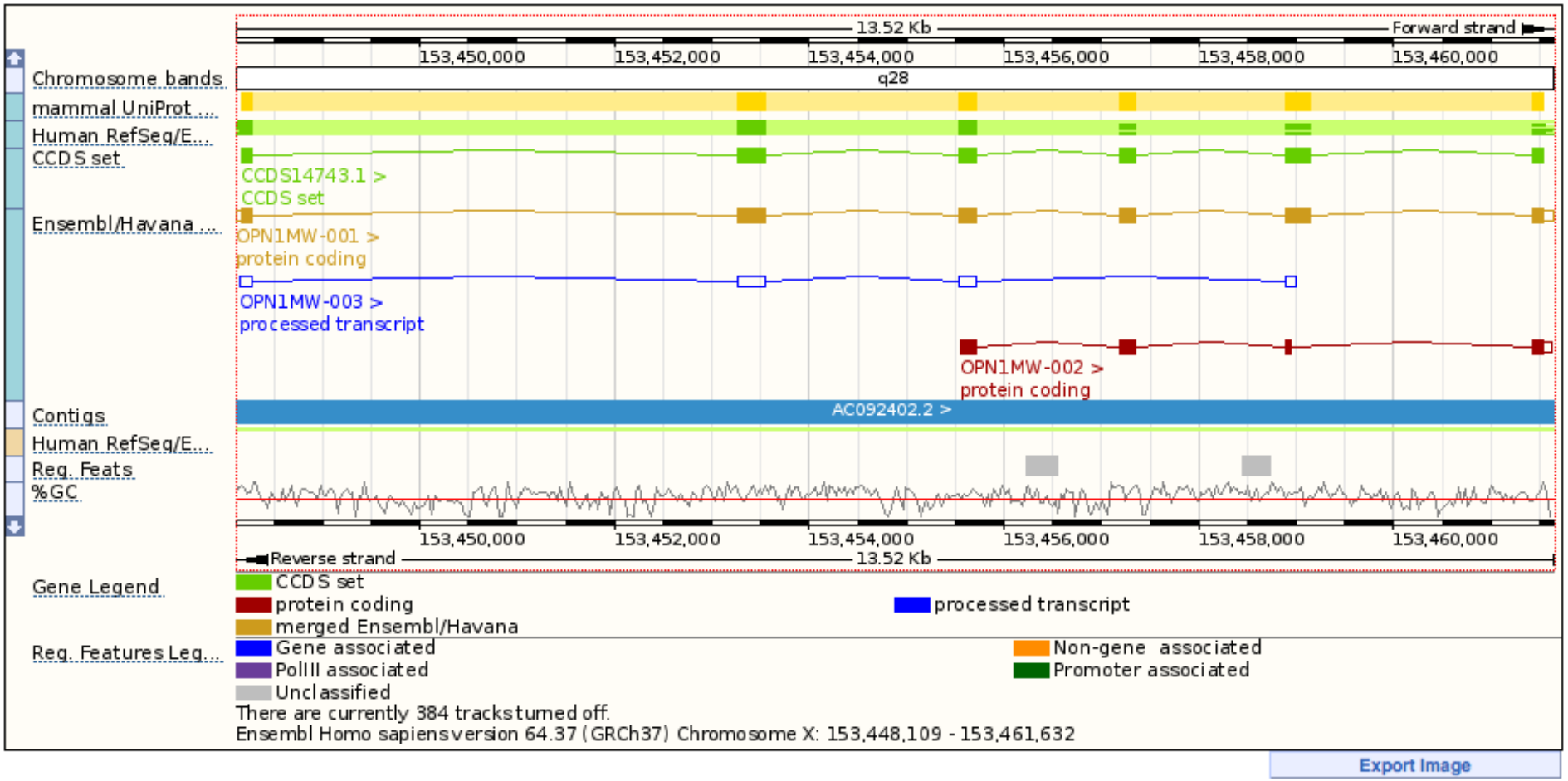
Ensembl Opsin OPN1MW Gene Location

http://uswest.ensembl.org/Homo_sapiens/Location/View?h=Havana%20gene;r=X



Ensembl OPN1MW Transcripts

http://uswest.ensembl.org/Homo_sapiens/Location/View?h=Havana%20gene;r=X



Ensembl OPN1MW Opsin Protein

http://uswest.ensembl.org/Homo_sapiens/Transcript/ProteinSummary?db=core;g=ENSG0000

Transcript: OPN1MW-001 (ENST00000369935)

Description opsin 1 (cone pigments), medium-wave-sensitive [Source:HGNC Symbol;Acc:4206]

Location Chromosome X: 153,448,107-153,461,633 forward strand.

Gene This transcript is a product of gene ENSG00000147380 - There are 3 transcripts in this gene

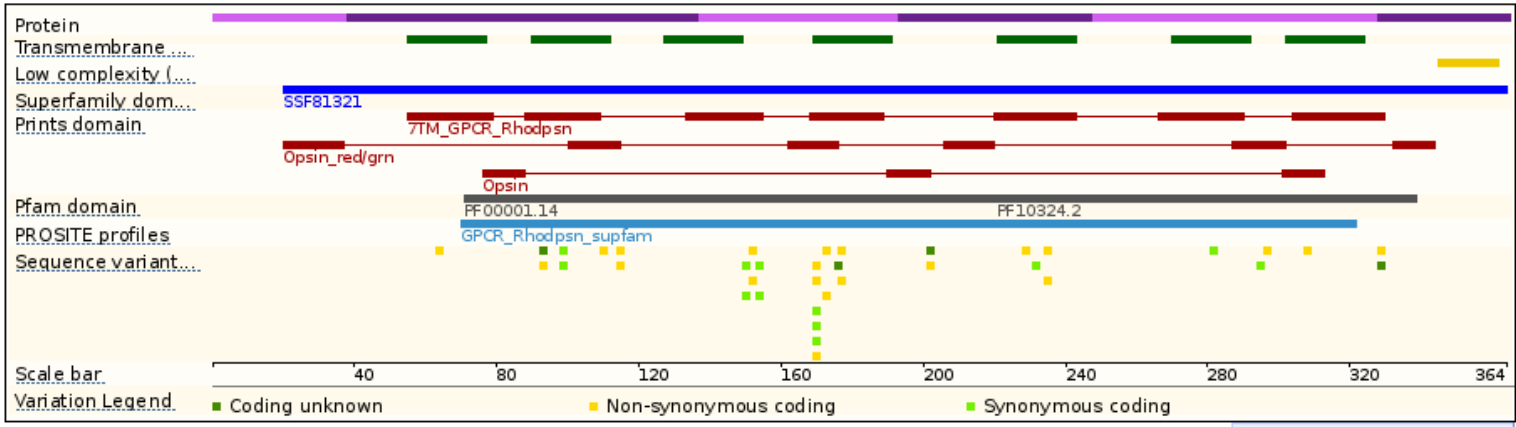
Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
OPN1MW-001	ENST00000369935	1257	ENSP00000358951	364	Protein coding	CCDS14743
OPN1MW-002	ENST00000430054	590	ENSP00000394838	164	Protein coding	-
OPN1MW-003	ENST00000468495	692	No protein product	-	Processed transcript	-

Transcript and Gene level displays

Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated with. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Protein summary help



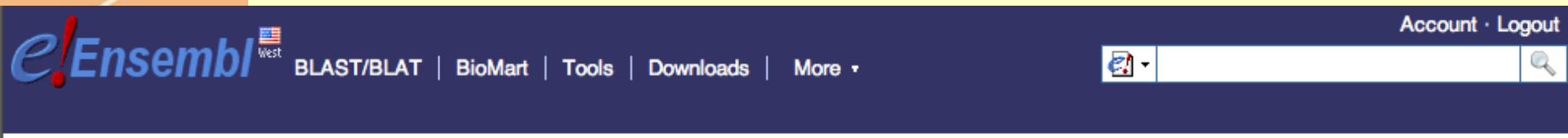
Export Image

Statistics

Ave. residue weight: 111.495 g/mol
Charge: 11.5
Isoelectric point: 8.7067
Molecular weight: 40,584.28 g/mol
Number of residues: 364 aa

Ensembl Tutorials

http://uswest.ensembl.org/info/website/tutorials/index.html



Help & Documentation > Tutorials

Ensembl Tutorials and Worked Examples

Online tutorials

- Introduction to Browsing Chordate Genomes - a 3-hour comprehensive interactive tutorial aimed at beginners

Video tutorials

The tutorials listed below are Flash animations of some of our training presentations. We are gradually adding to the list, so please check back regularly.

Note that we are now hosting all our tutorials on YouTube (and 优酷网 for users in China) for ease of maintenance. A selection of tutorials is also available on the EBI E-Video website.

Title	Running time (minutes)
The Ensembl Genome Browser	10:00
EnsemblGenomes - Extending Ensembl across the taxonomic space	4:54
Comparative Genomics	9:58
SNPs and other Variations - 1 of 2	7:06
SNPs and other Variations - 2 of 2	6:30
Introduction to BioMart	4.27
BioMart: Variation IDs to HGNC Symbols	2:58
Clip: Exons and Introns	1:12
Clip: Export Sequence	1:08
Clip: Genome Variation	0:37
Clip: Transcriptomics (ArrayExpress)	1:12
Clip: View Conserved Sequence	1:26
Clip: View External Data (DAS)	1:53

UCSC Genome Home Page

http://genome.ucsc.edu/

UCSC Genome Bioinformatics

Genome Browser
ENCODE
Neandertal
Blat
Table Browser
Gene Sorter
In Silico PCR
Genome Graphs
Galaxy
VisiGene
Proteome Browser
Utilities
Downloads
Release Log
Custom Tracks

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the ENCODE and Neandertal projects.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list.

News

News Archives ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

10 October 2011 - Updated Yeast Browser Released

We are happy to announce an updated Yeast Genome Browser for the initial release of *Saccharomyces cerevisiae*, sacCer3. The April 2011 *Saccharomyces cerevisiae* genome assembly (Saccharomyces cerevisiae S288c assembly from Saccharomyces Genome Database (GCA_000146055.2)) was produced by the *Saccharomyces* Genome Database (SGD™) project.

Chromosomes available in this assembly: chrI, chrII, chrIII, chrIV ... etc ... chrXVI, and chrM. See also: SGD™ genome snapshot/overview.

Downloads of the yeast data and annotations may be obtained from the UCSC Genome Browser FTP server or Downloads page. The *S. cerevisiae* annotation tracks were generated by UCSC and collaborators worldwide.

We'd like to thank the Saccharomyces Genome Database (SGD™). The *S. cerevisiae* Genome Browser and annotation tracks were produced by Hiram Clawson, Greg Roe, and Steve Heitner. See the Credits page for a detailed list of the organizations and individuals who contributed to this release.

UCSC Genome Browser

http://genome.ucsc.edu/cgi-bin/hgGateway

Home | Genomes | Blat | Tables | Gene Sorter | PCR | Session | FAQ | Help

Human (*Homo sapiens*) Genome Browser Gateway

clade	genome	assembly	position or search term	gene	
Mammal	Human	Feb. 2009 (GRCh37/hg19)	opsin		submit

Click here to reset the browser user interface settings to their defaults.

track search | add custom tracks | track hubs | configure tracks and display | clear position

About the Human Feb. 2009 (GRCh37/hg19) assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the Genome Reference Consortium.

Homo sapiens
(Graphic courtesy of CBSE)

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.

UCSC

UCSC Genes

OPN4 (uc010qmk.1) at chr10:88414314-88426605 - opsin 4 isoform 2
OPN4 (uc009xsx.1) at chr10:88420233-88423870 - opsin 4 isoform 1
OPN1MW (uc004fkd.2) at chrX:153485203-153499469 - opsin 1 (cone pigments), medium-wave-sensitive
OPN1MW (uc004fkb.2) at chrX:153448085-153462351 - opsin 1 (cone pigments), medium-wave-sensitive
OPN1LW (uc004fjz.3) at chrX:153409725-153424505 - opsin 1 (cone pigments), long-wave-sensitive
OPN1SW (uc003vnt.3) at chr7:128412545-128415844 - opsin 1 (cone pigments), short-wave-sensitive
OPN5 (uc003ozd.2) at chr6:47754862-47794114 - opsin 5 isoform 2
OPN5 (uc003ozc.2) at chr6:47749798-47794114 - opsin 5 isoform 1
OPN4 (uc001kdq.2) at chr10:88414314-88426214 - opsin 4 isoform 1
OPN4 (uc001kdp.2) at chr10:88414314-88426214 - opsin 4 isoform 2
OPN3 (uc001hza.2) at chr1:241756453-241803701 - opsin 3
RGR (uc001kde.1) at chr10:86004809-86018944 - retinal G-protein coupled receptor isoform 3
RGR (uc001kdd.1) at chr10:86004809-86018944 - retinal G-protein coupled receptor isoform 1
RGR (uc001kdc.1) at chr10:86004809-86018944 - retinal G-protein coupled receptor isoform 2
VSX1 (uc010qde.1) at chr20:25052132-25062767 - visual system homeobox 1 isoform b
VSX1 (uc002wue.2) at chr20:25052132-25062767 - visual system homeobox 1 isoform b
VSX1 (uc002wuq.1) at chr20:25059195-25062767 - visual system homeobox 1 isoform b
VSX1 (uc010qdf.1) at chr20:25052132-25062767 - visual system homeobox 1 isoform b
VSX1 (uc010qdd.1) at chr20:25052132-25062767 - visual system homeobox 1 isoform b
VSX1 (uc002wuf.2) at chr20:25056100-25062767 - visual system homeobox 1 isoform a
CRX (uc002phq.3) at chr19:48325099-48346584 - cone-rod homeobox protein
RHO (uc003emt.2) at chr3:129247482-129254186 - rhodopsin
RRH (uc003hzv.2) at chr4:110749150-110765859 - peropsin
ARR3 (uc004dyb.2) at chrX:69488185-69501690 - arrestin 3, retinal (X-arrestin)
GRK7 (uc011bnd.1) at chr3:141497043-141535890 - G-protein-coupled receptor kinase 7 precursor

RefSeq Genes

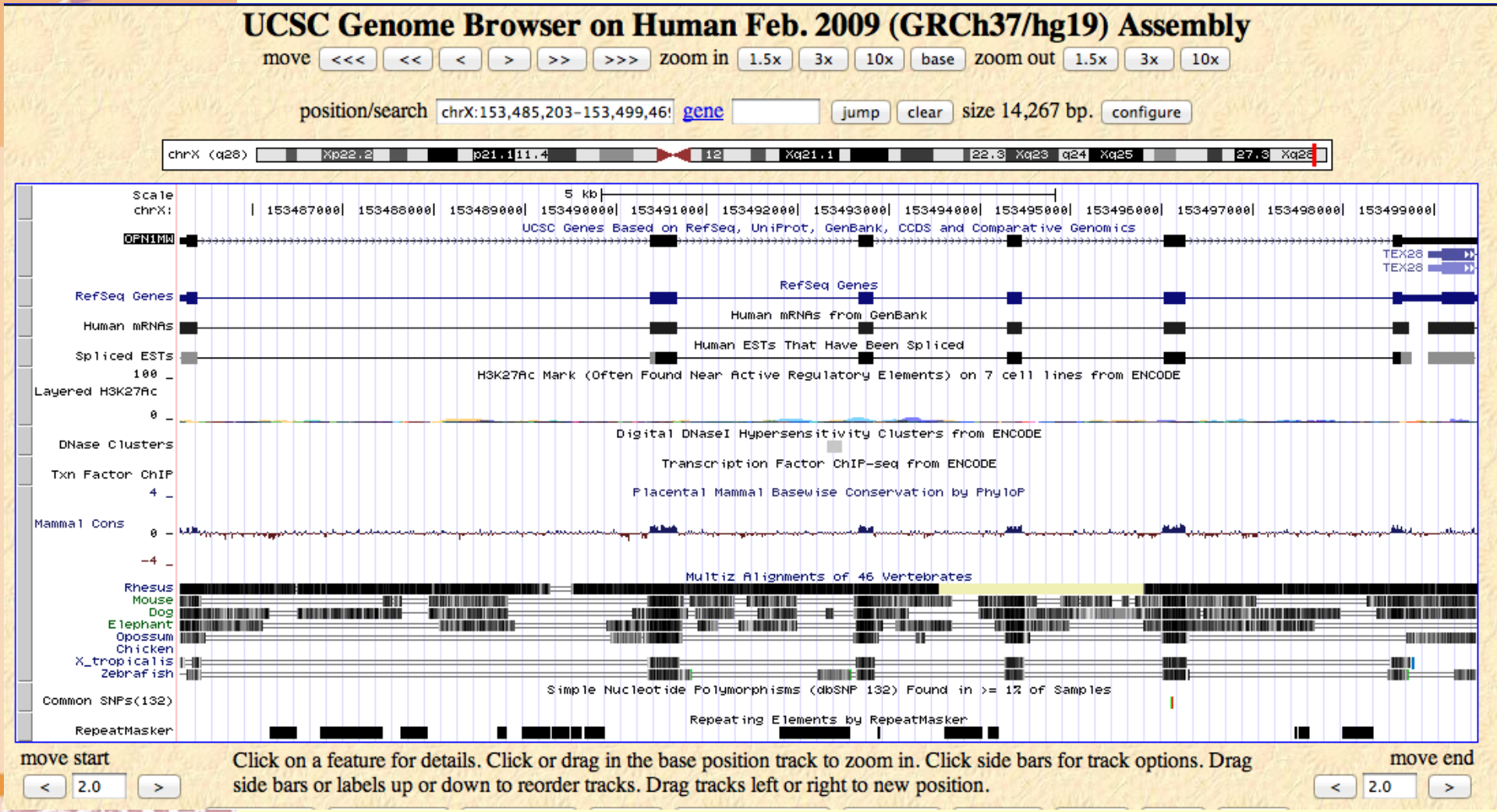
OPN1MW at chrX:153485203-153499470 - (NM_000513) medium-wave-sensitive opsin 1
OPN1MW at chrX:153448085-153462352 - (NM_000513) medium-wave-sensitive opsin 1
OPN5 at chr6:47749775-47794116 - (NM_181744) opsin-5
OPN1SW at chr7:128412543-128415844 - (NM_001708) short-wave-sensitive opsin 1
OPN1LW at chrX:153409725-153424507 - (NM_020061) long-wave-sensitive opsin 1
OPN1MW2 at chrX:153485203-153498755 - (NM_001048181) medium-wave-sensitive opsin 1
OPN1MW2 at chrX:153448085-153461637 - (NM_001048181) medium-wave-sensitive opsin 1
OPN3 at chr1:241756452-241803701 - (NM_014322) opsin-3

Non-Human RefSeq Genes

OPN5 at chr6:47749803-47779473 - (NM_001206080) opsin-5
OPN1SW at chr7:128413703-128415835 - (NM_001076704) blue-sensitive opsin
LOC751972 at chr3:129247577-129252528 - (NM_001076696) RH2 opsin

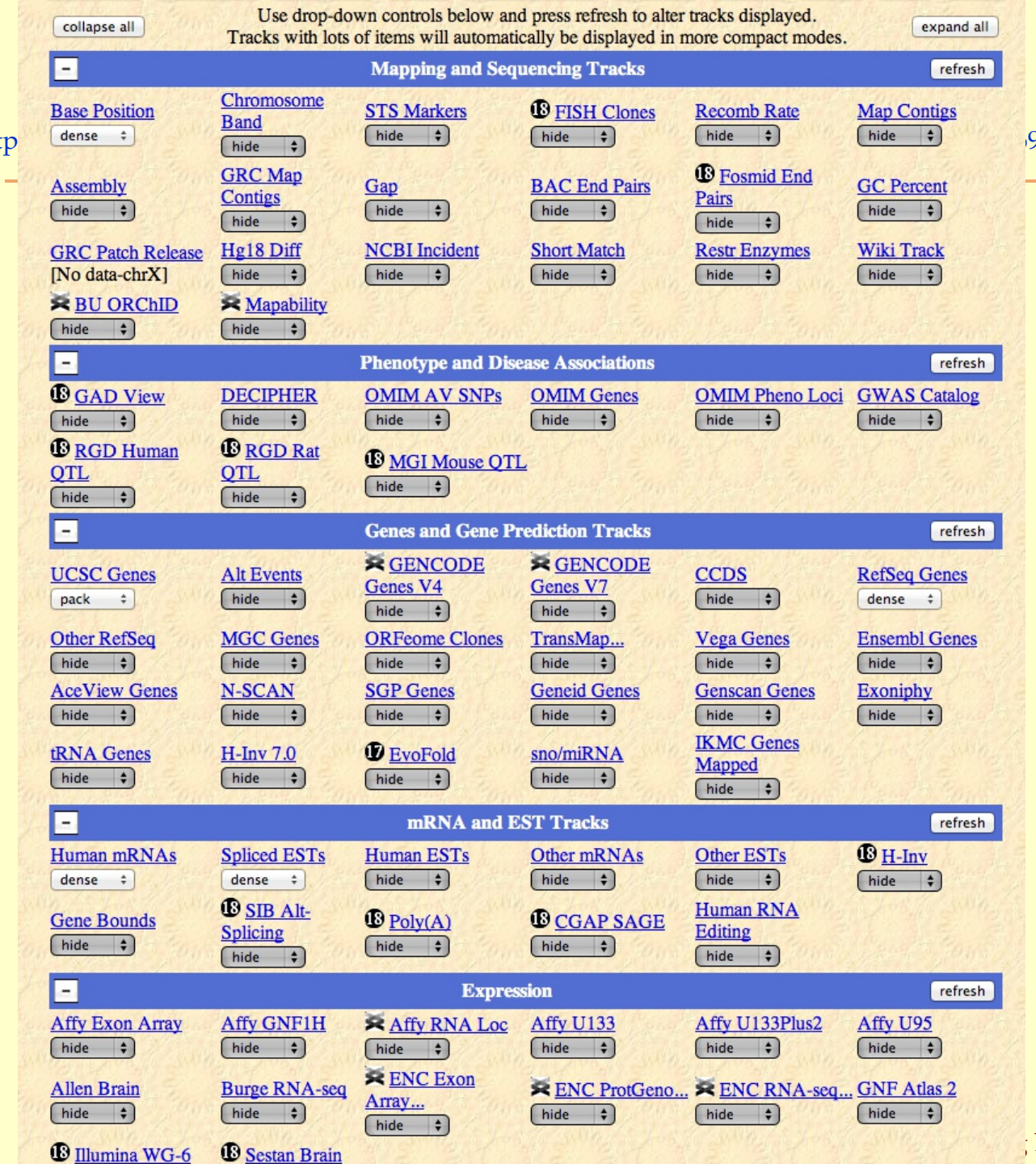
UCSC Genome Browser

http://genome.ucsc.edu/cgi-bin/hgTracks?position=chrX:153485203-153499469&hgsid=216983641&known



UCSC

http ...983641&known...


collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

expand all

Mapping and Sequencing Tracks refresh

Base Position: dense	Chromosome Band: hide	STS Markers: hide	18 FISH Clones: hide	Recomb Rate: hide	Map Contigs: hide
Assembly: hide	GRC Map Contigs: hide	Gap: hide	BAC End Pairs: hide	18 Fosmid End Pairs: hide	GC Percent: hide
GRC Patch Release [No data-chrX]	Hg18 Diff: hide	NCBI Incident: hide	Short Match: hide	Restr Enzymes: hide	Wiki Track: hide
BU ORChID: hide	Mapability: hide				

Phenotype and Disease Associations refresh

18 GAD View: hide	DECIPHER: hide	OMIM AV SNPs: hide	OMIM Genes: hide	OMIM Pheno Loci: hide	GWAS Catalog: hide
18 RGD Human QTL: hide	18 RGD Rat QTL: hide	18 MGI Mouse QTL: hide			

Genes and Gene Prediction Tracks refresh

UCSC Genes: pack	Alt Events: hide	GENCODE Genes V4: hide	GENCODE Genes V7: hide	CCDS: hide	RefSeq Genes: dense
Other RefSeq: hide	MGC Genes: hide	ORFeome Clones: hide	TransMap...: hide	Vega Genes: hide	Ensembl Genes: hide
AceView Genes: hide	N-SCAN: hide	SGP Genes: hide	Geneid Genes: hide	Genscan Genes: hide	Exoniphy: hide
tRNA Genes: hide	H-Inv 7.0: hide	17 EvoFold: hide	sno/miRNA: hide	IKMC Genes Mapped: hide	

mRNA and EST Tracks refresh

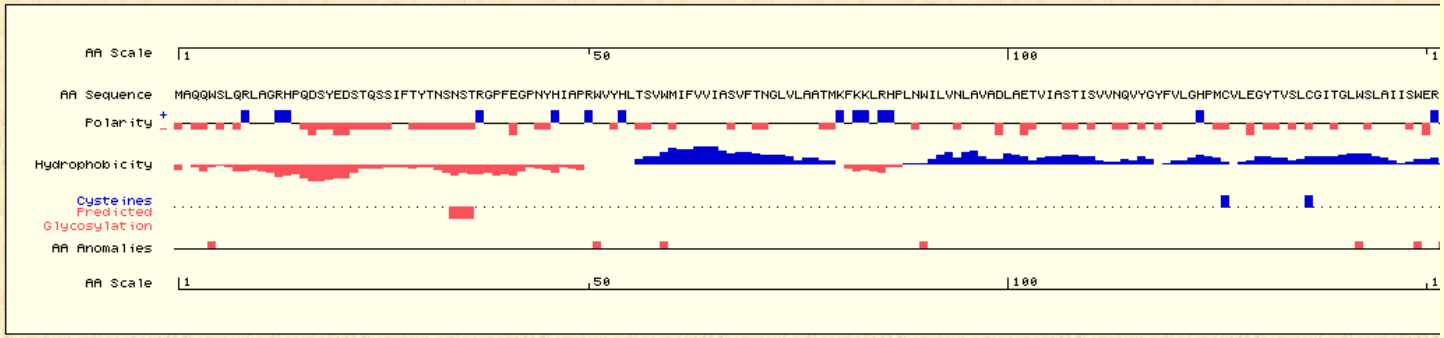
Human mRNAs: dense	Spliced ESTs: dense	Human ESTs: hide	Other mRNAs: hide	Other ESTs: hide	18 H-Inv: hide
Gene Bounds: hide	18 SIB Alt-Splicing: hide	18 Poly(A): hide	18 CGAP SAGE: hide	Human RNA Editing: hide	

Expression refresh

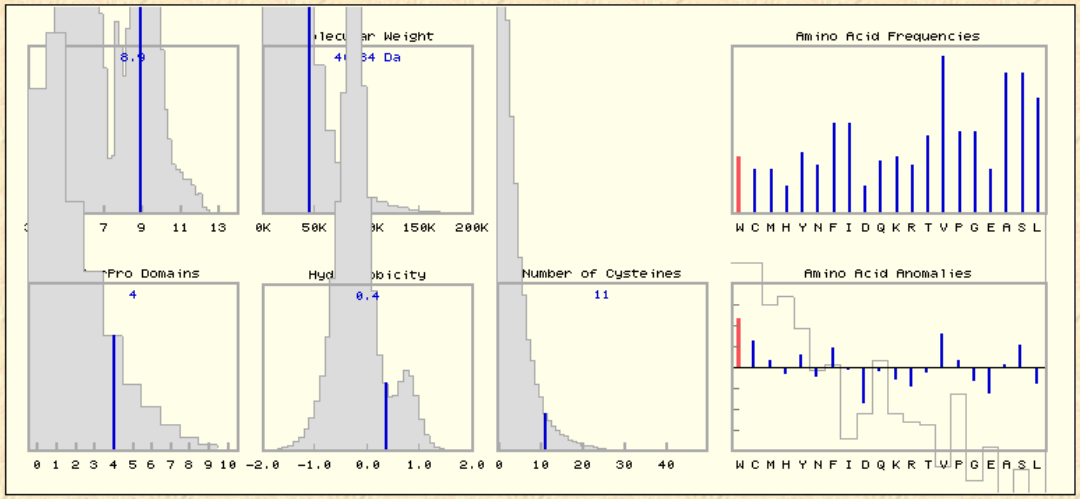
Affy Exon Array: hide	Affy GNF1H: hide	Affy RNA Loc: hide	Affy U133: hide	Affy U133Plus2: hide	Affy U95: hide
Allen Brain: hide	Burge RNA-seq: hide	ENC Exon Array...: hide	ENC ProtGeno...: hide	ENC RNA-seq...: hide	GNF Atlas 2: hide
18 Illumina WG-6	18 Sestan Brain				

Protein P04001 (aka OPSG_HUMAN) RecName: Full=Green-sensitive opsin; AltName: Full=Green cone photoreceptor pigment;
Organism: Homo sapiens (human)

Move <<< << < > >> >>> Current scale: FULL Rescale to 1/6 1/2 FULL



Explanation of Protein Tracks



Explanation of Protein Property Histograms

UCSC Links:

- BLAT - Homo sapiens (human)

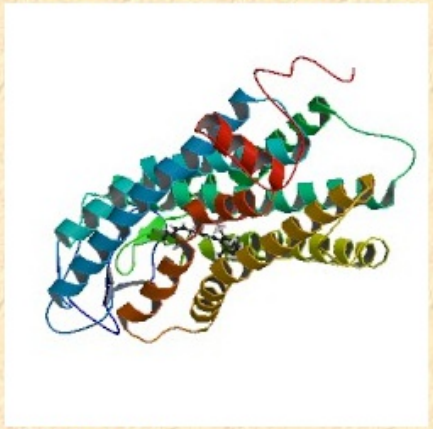
InterPro Domains: Graphical view of domain structure

- IPR000276 - 7TM_GPCR_Rhodpsn
- IPR017452 - GPCR_Rhodpsn_supfam
- IPR001760 - Opsin
- IPR000378 - Opsin_red/grn

Pfam Domains:

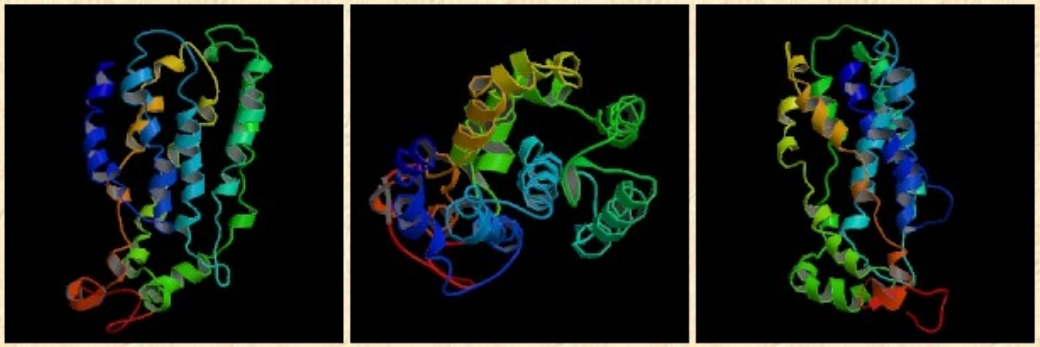
- PF00001 - 7 transmembrane receptor (rhodopsin family)

Protein Data Bank (PDB) 3-D Structure



1KPW - Model

ModBase Predicted Comparative 3D Structure on P04001



Front Top Side

The pictures above may be empty if there is no ModBase structure for the protein. The ModBase structure frequently covers just a fragment of the protein. You may be asked to log onto ModBase the first time you click on the pictures. It is simplest after logging in to just click on the picture again to get to the specific info on that model.

Total amino acids: 364

FASTA record:

```
>P04001|OPSG_HUMAN|RecName: Full=Green-sensitive opsin; AltName: Full=Green cone photoreceptor pigment;
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNSNSTRGPFEGPNYHIAPR
WVYHLTSVWMIFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADL
AETVIASTISVVNQVYGYFVLGHPMCVLEGYTVSLCGITGLWSLAIISWE
RWMVVCKPFGNVRFDAKLAIVGIAFSWIWAAVWTAPPIFGWSRYWPHGLK
TSCGPDVFSGSSYPGVQSYMIVLMVTCCITPLSIIVLCYLQVWLAIRAVA
KQQKESESTQKAEKEVTRMVVVMVLAFCFCWGPYAFFACFAAANPGYPFH
PLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDGSELSSA
SKTEVSSVSSVSPA
```

UCSC

UCSC Help File

http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html

UCSC Genome Bioinformatics

Home - Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Proteome - Session - FAQ

Genome Browser User Guide

Table of Contents:

Search the Genome Browser help pages: Submit

Search the Genome mailing list archives: Submit

See also the Open Helix tutorial and training materials.

Questions and feedback are welcome.

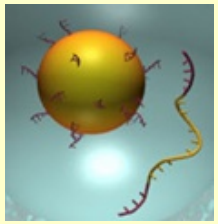
What does the Genome Browser do?

454 LIFE SCIENCES

Emulsion Based Clonal Amplification

Single test tube generation of millions of clonally amplified sequencing templates
No cloning and colony picking

Adapter carrying library DNA



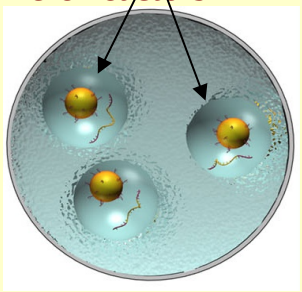
Anneal DNA template to capture beads

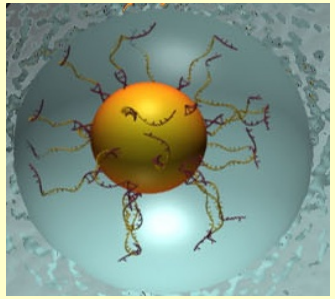
+ PCR Reagents

+ Emulsion Oil

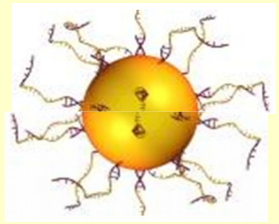
"Water-in-oil" emulsion

Micro-reactors





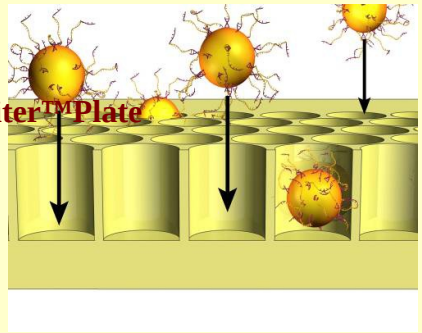
Perform emulsion PCR



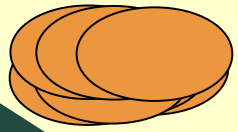
Break micro-reactors
Isolate DNA containing beads

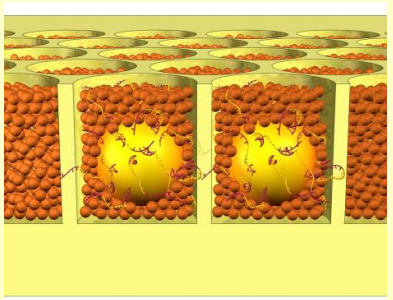
Depositing DNA Beads into the PicoTiter™Plate

Load beads into PicoTiter™Plate

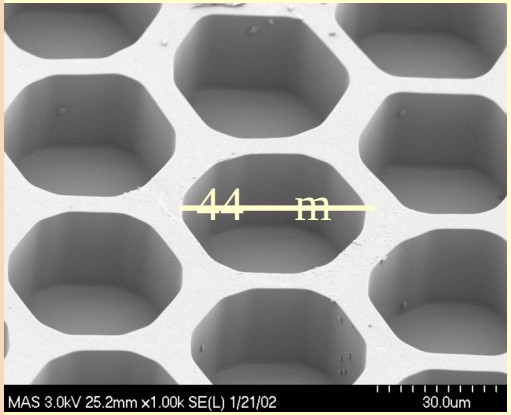


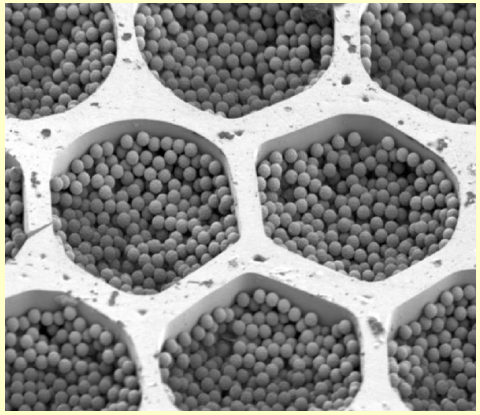
Load Enzyme Beads





Centrifuge Step

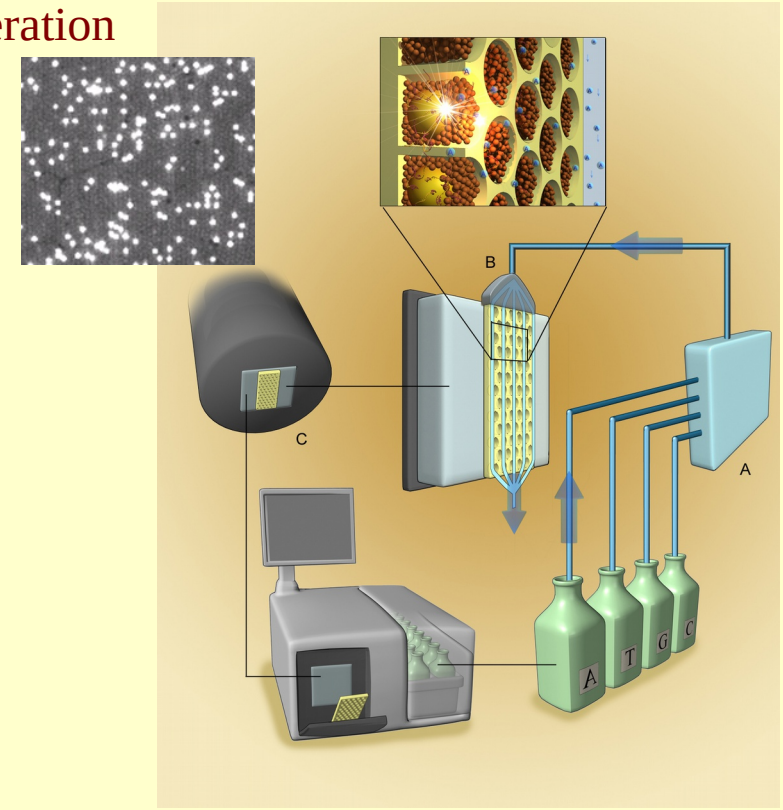




- 70x75mm array of fused optical f bers with etched wells
- 1.6 million wells monitored optically through f ber

Sequencing-By-Synthesis

- Simultaneous sequencing of the entire genome in hundreds of thousands of picoliter-size wells
- Pyrophosphate signal generation



Flowgrams and BaseCalling