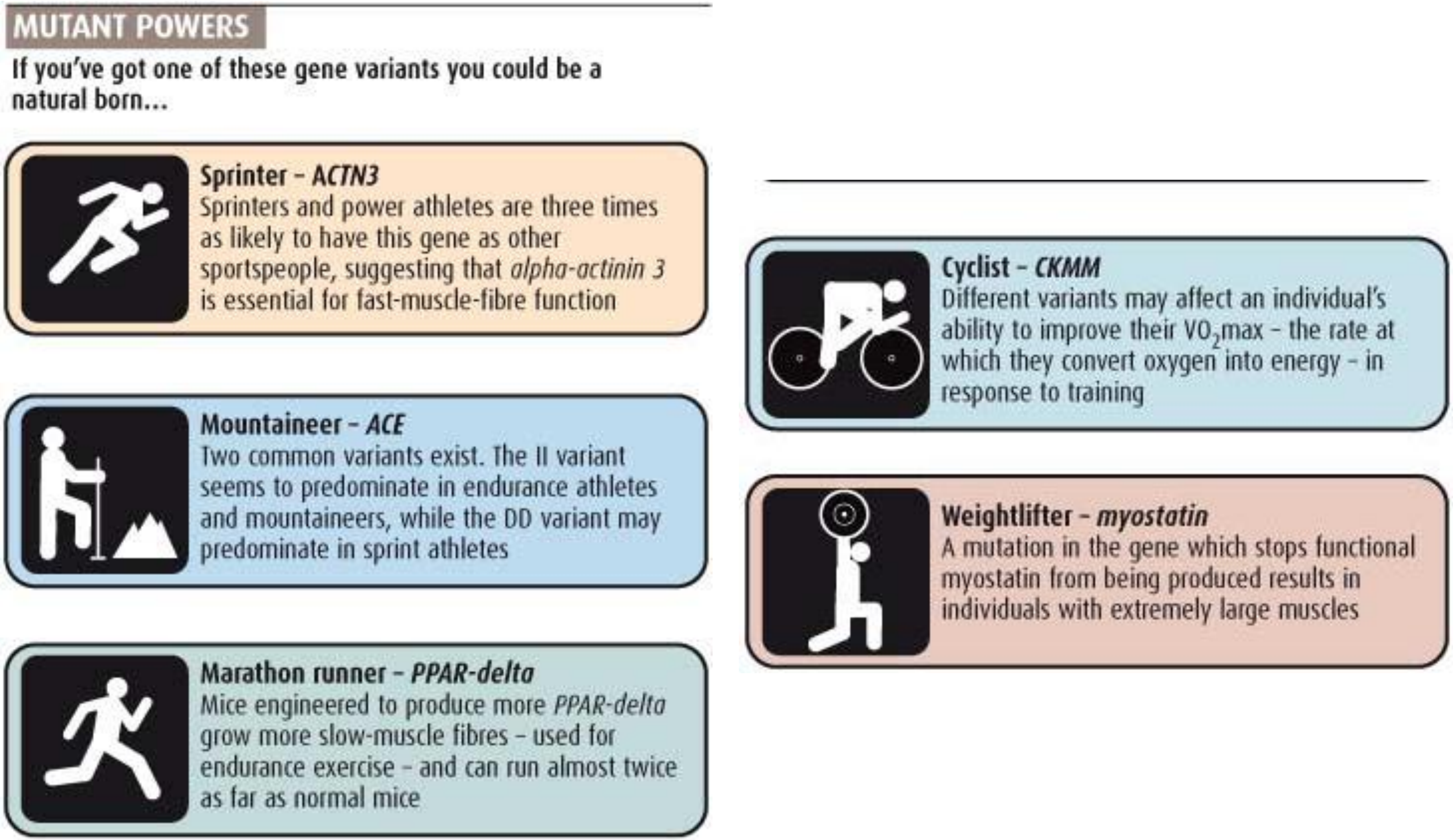
Epigenetics and gene regulation

Dr. Chris Evelo

BWE 15-05-2013

Genetic variations and sports



MUTANT POWERS

If you've got one of these gene variants you could be a natural born...

Sprinter – *ACTN3*
Sprinters and power athletes are three times as likely to have this gene as other sportspeople, suggesting that *alpha-actinin 3* is essential for fast-muscle-fibre function

Mountaineer – *ACE*
Two common variants exist. The II variant seems to predominate in endurance athletes and mountaineers, while the DD variant may predominate in sprint athletes

Marathon runner – *PPAR-delta*
Mice engineered to produce more *PPAR-delta* grow more slow-muscle fibres – used for endurance exercise – and can run almost twice as far as normal mice

Cyclist – *CKMM*
Different variants may affect an individual's ability to improve their VO_2max – the rate at which they convert oxygen into energy – in response to training

Weightlifter – *myostatin*
A mutation in the gene which stops functional myostatin from being produced results in individuals with extremely large muscles

Epigenetics and sports

Sports Medicine

February 2013, Volume 43, Issue 2, pp 93-110

Epigenetics in Sports

Tobias Ehlert, Perikles Simon, Dirk A. Moser

*We suggest that **epigenetic effects** may also play a considerable role in the determination of **athletic potential** and these effects will need to be studied using more sophisticated quantitative genetic models. In the future, epigenetic status and its potential influence on athletic performance will have to be considered, explored and validated using well controlled model systems before we can begin to extrapolate new findings to complex and heterogeneous human populations.*

Regulation of gene expression

1. Gene **transcription** regulation
   - Epigenetic regulation
     - DNA methylation
     - Histone modifications

2. mRNA **translation** regulation
   - microRNA

CONTENT

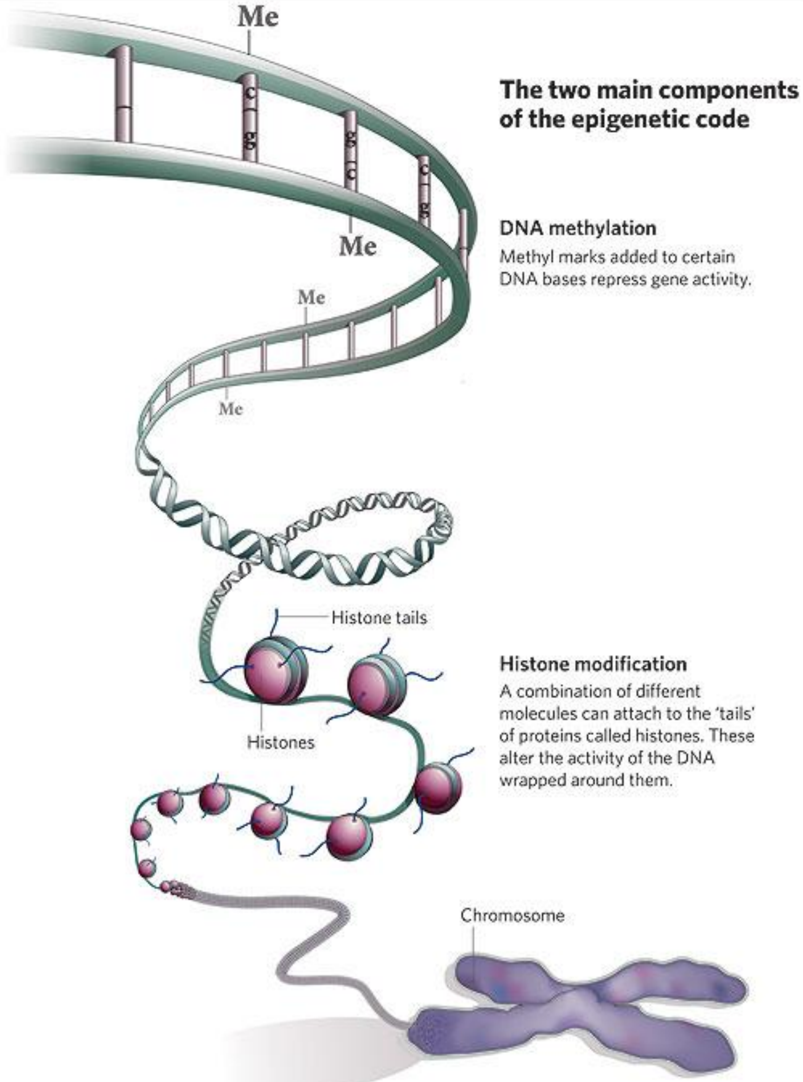
- What is Epigenetics?
  - Histone modifications
  - DNA methylation
- Biological relevance of epigenetics
- Epigenetics in UCSC
- Methods to measure DNA methylation
- Motif analysis
- microRNAs

Epigenetics/epigenomics

- **Epigenetics** refers to the study of changes in the regulation of gene activity and expression that are not dependent on gene DNA sequence.

- While epigenetics often refers to the study of single genes or sets of genes, **epigenomics** refers to more global analyses of epigenetic changes across the entire genome, so **genome-wide.**

Epigenetic regulation

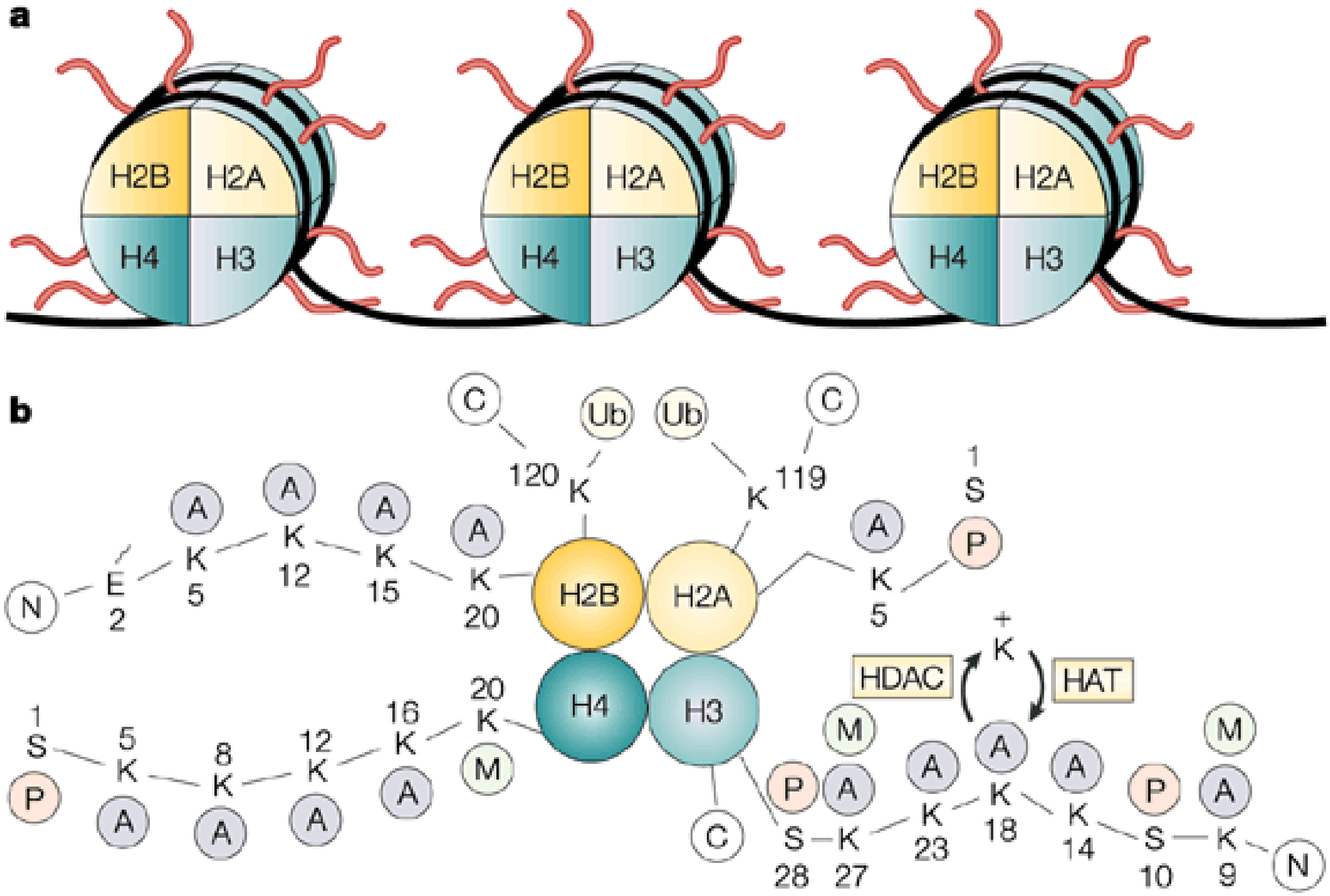

Me
c
g
g
c
c
g
Me
Me
Me
The two main components of the epigenetic code
DNA methylation
Methyl marks added to certain DNA bases repress gene activity.
Histone tails
Histones
Histone modification
A combination of different molecules can attach to the 'tails' of proteins called histones. These alter the activity of the DNA wrapped around them.
Chromosome

DNA methylation

Histone modifications

Qiu, J. *Epigenetics: Unfinished symphony*. **Nature** 2006, 441:143-145

Histones

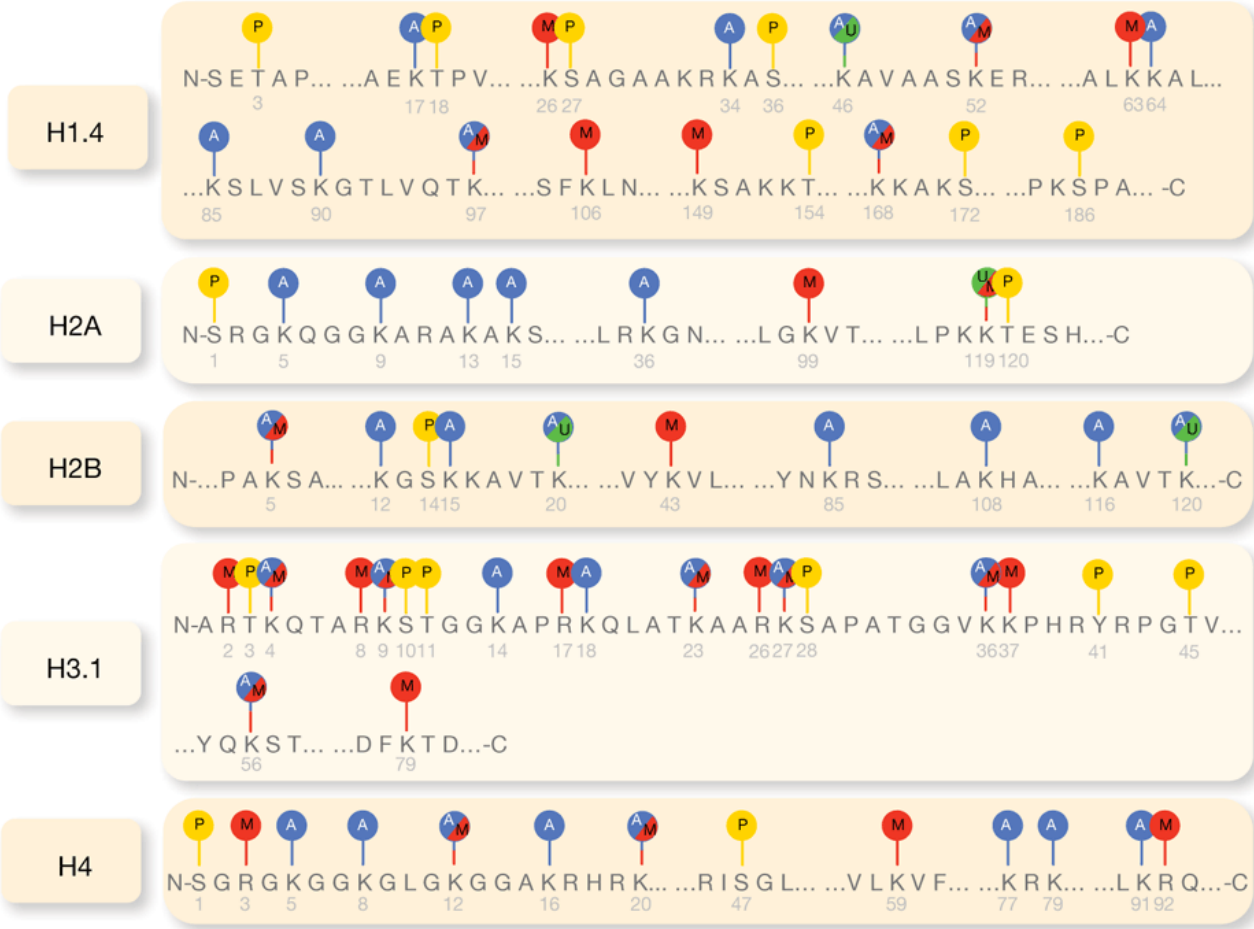


Nature Reviews | Cancer

Histone modifications I

- A combination of different molecules can attach to the tails of histones altering the activity of DNA wrapped around:
  - Methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, citrullination, and ADP-ribosylation

Histone modifications II


H1.4
N-S E T A P... ...A E K T P V... ...K S A G A A K R K A S... ...K A V A A S K E R... ...A L K K A L...
3 17 18 26 27 34 36 46 52 63 64
...K S L V S K G T L V Q T K... ...S F K L N... ...K S A K K T... ...K K A K S... ...P K S P A... -C
85 90 97 106 149 154 168 172 186
H2A
N-S R G K Q G G K A R A K A K S... ...L R K G N... ...L G K V T... ...L P K K T E S H...-C
1 5 9 13 15 36 99 119 120
H2B
N-...P A K S A... ...K G S K K A V T K... ...V Y K V L... ...Y N K R S... ...L A K H A... ...K A V T K...-C
5 12 14 15 20 43 85 108 116 120
H3.1
N-A R T K Q T A R K S T G G K A P R K Q L A T K A A R K S A P A T G G V K K P H R Y R P G T V...
2 3 4 8 9 10 11 14 17 18 23 26 27 28 36 37 41 45
...Y Q K S T... ...D F K T D...-C
56 79
H4
N-S G R G K G G K G L G K G G A K R H R K... ...R I S G L... ...V L K V F... ...K R K... ...L K R Q...-C
1 3 5 8 12 16 20 47 59 77 79 91 92
P
A
M
U

Epigenetic modifications and human disease, Portela and Esteller *Nature Biotechnology* 28, 1057–1068 (2010)

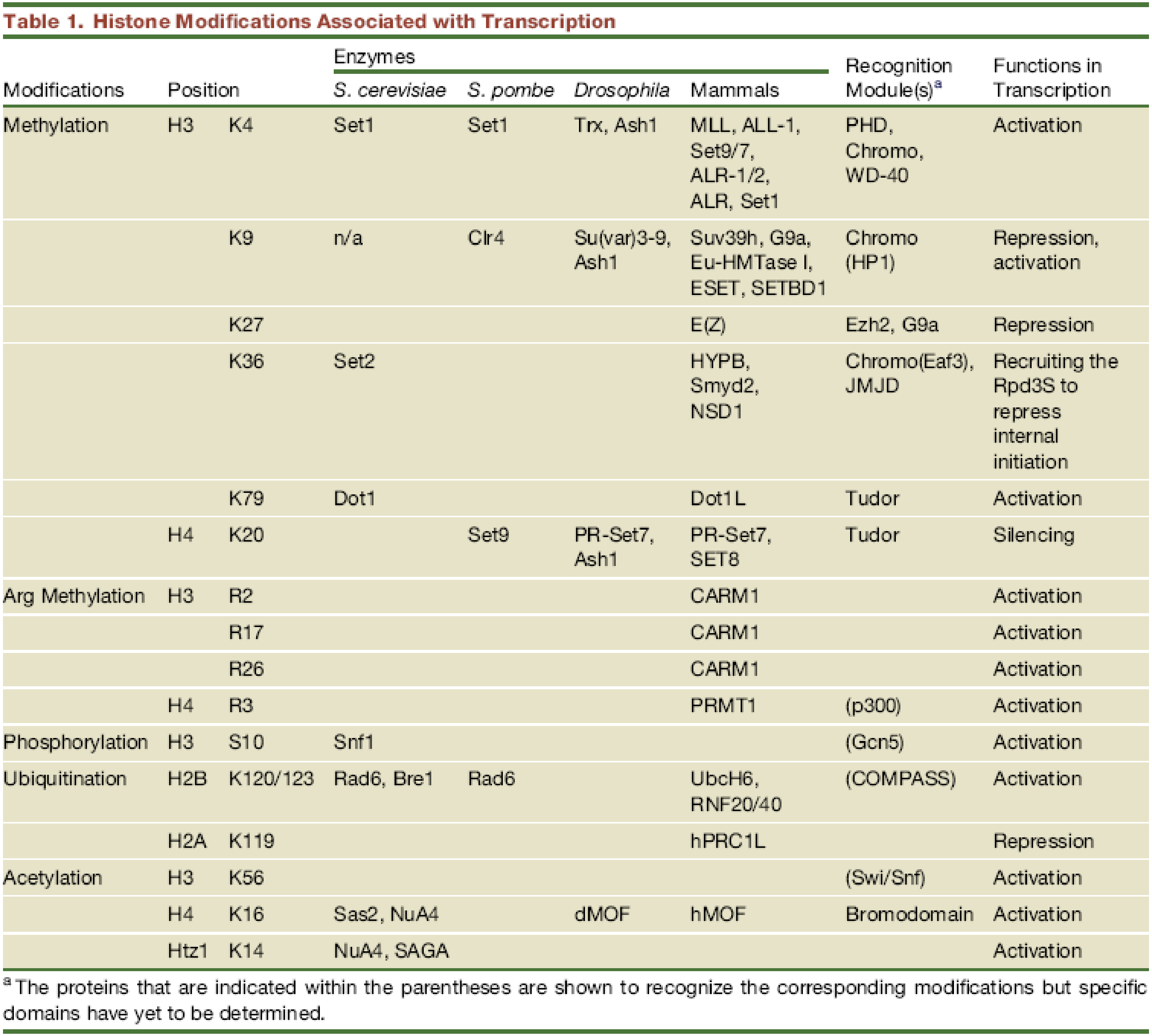


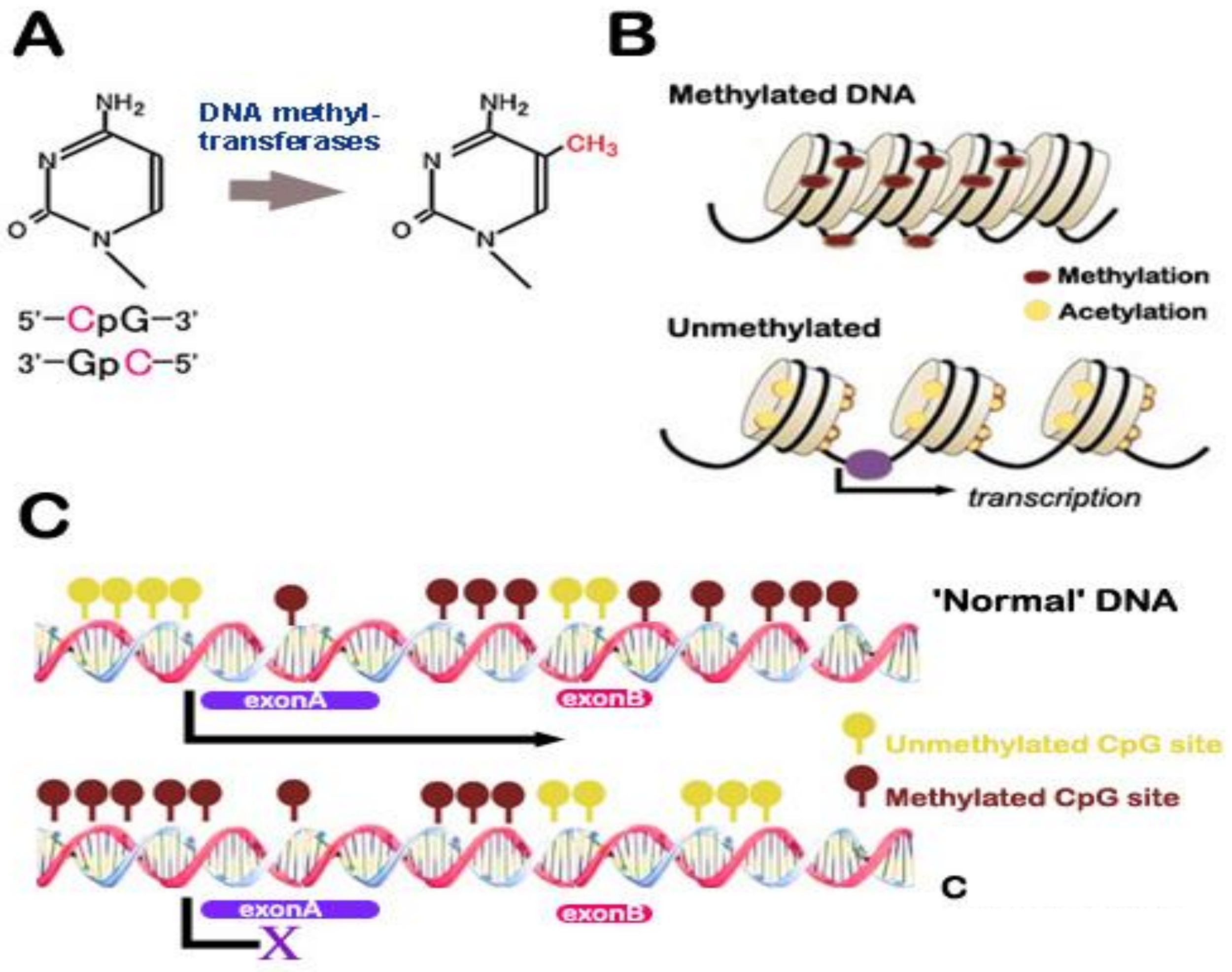
Table 1. Histone Modifications Associated with Transcription

			Enzymes					
Modifications	Position		*S. cerevisiae*	*S. pombe*	*Drosophila*	Mammals	Recognition Module(s)[a]	Functions in Transcription
Methylation	H3	K4	Set1	Set1	Trx, Ash1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromo, WD-40	Activation
		K9	n/a	Clr4	Su(var)3-9, Ash1	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Chromo (HP1)	Repression, activation
		K27				E(Z)	Ezh2, G9a	Repression
		K36	Set2			HYPB, Smyd2, NSD1	Chromo(Eaf3), JMJD	Recruiting the Rpd3S to repress internal initiation
		K79	Dot1			Dot1L	Tudor	Activation
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, SET8	Tudor	Silencing
Arg Methylation	H3	R2				CARM1		Activation
		R17				CARM1		Activation
		R26				CARM1		Activation
	H4	R3				PRMT1	(p300)	Activation
Phosphorylation	H3	S10	Snf1				(Gcn5)	Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	Rad6		UbcH6, RNF20/40	(COMPASS)	Activation
	H2A	K119				hPRC1L		Repression
Acetylation	H3	K56					(Swi/Snf)	Activation
	H4	K16	Sas2, NuA4		dMOF	hMOF	Bromodomain	Activation
	Htz1	K14	NuA4, SAGA					Activation

[a] The proteins that are indicated within the parentheses are shown to recognize the corresponding modifications but specific domains have yet to be determined.

Li e. al. (2007)
Cell 128, 707

DNA Methylation



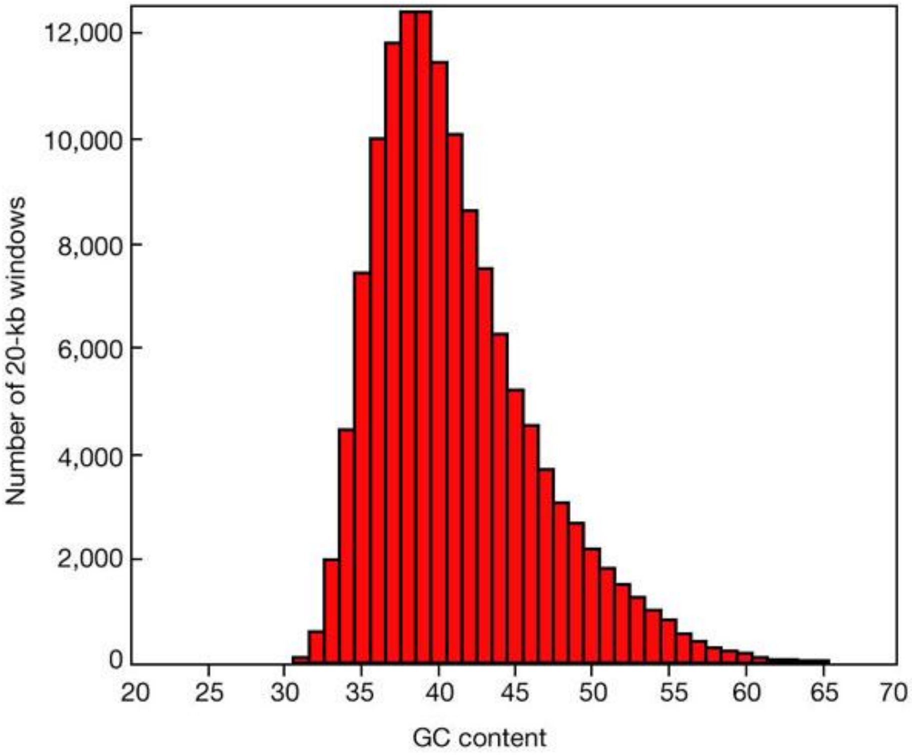
Hypomethylation
Hypermethylation

http://www.cellscience.com/reviews7/Taylor1.jpg

CpG islands

- CpG islands are clusters of '5-CG-3' di-nucleotides (CpGs)

- CpGs are underrepresented in the human genome, occurring at one fifth the expected frequency in genomic DNA



Source: IHGSC

CpG underrepresentation

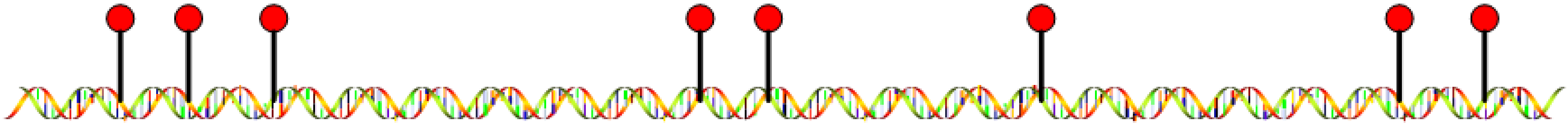
- Cause of underrepresentation:
  - CpG dinucleotides often are methylated on cytosine (m^{5}CpG)
  - m^{5}CpG can turn into to thymine through spontaneous deamination

- CpGs that are left in the genome, have thus been actively kept from mutating to thymine:
  - Implies functional relevance

CpG islands

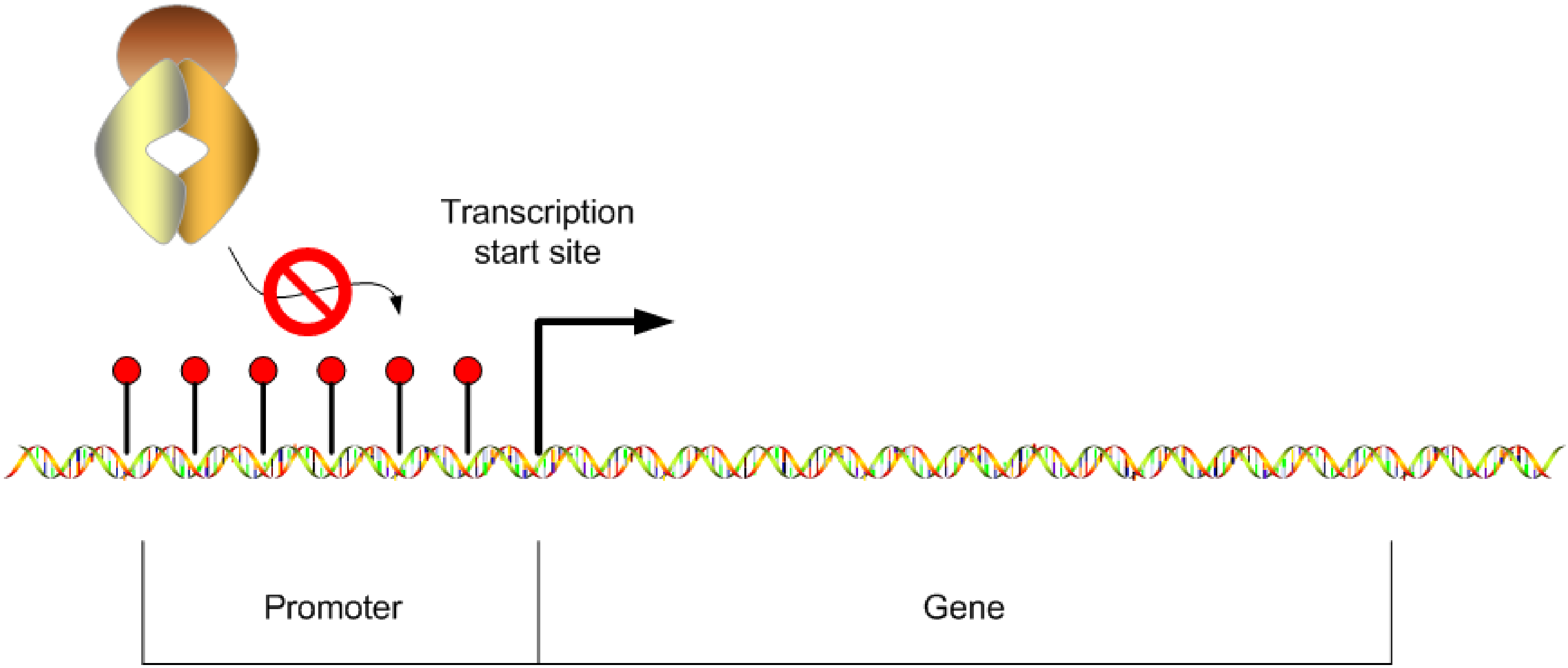
- Most CpGs are present in clusters called CpG islands (CGIs).

- CGIs are located at various positions throughout genes, most notably in promoter regions, often in housekeeping genes



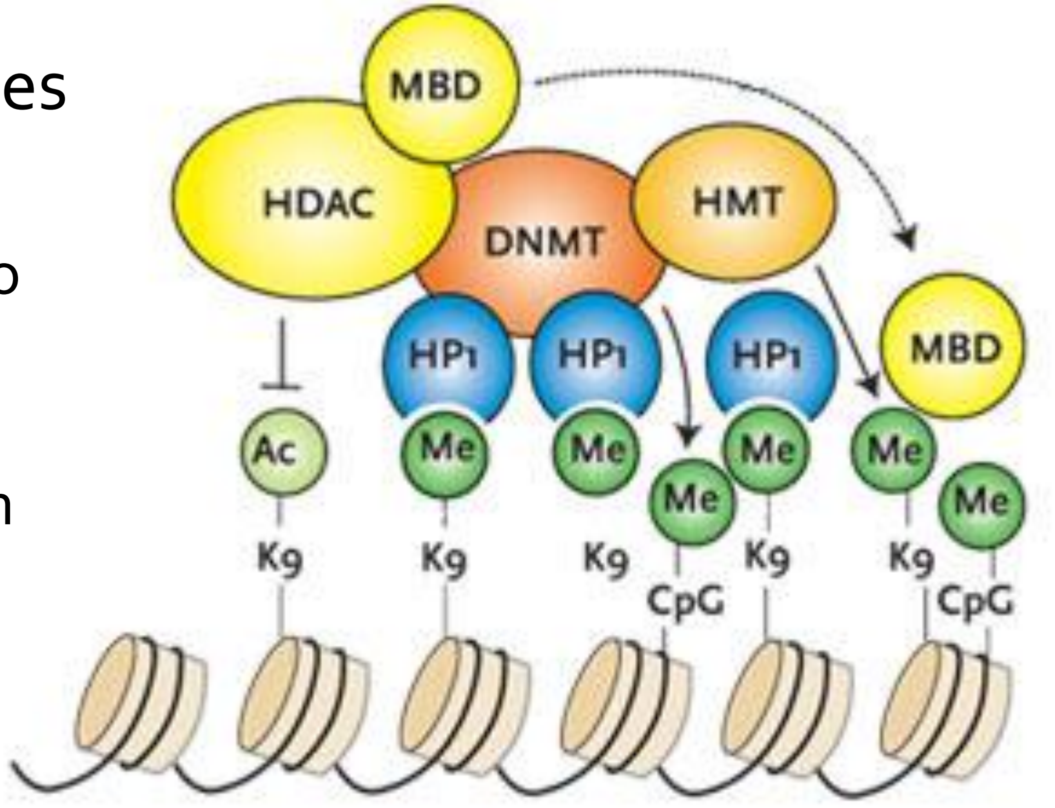
CpG island methylation (I)

- Methylation of promoter CGIs causes gene silencing:
  - Impedes TF binding directly: decrease in binding affinity



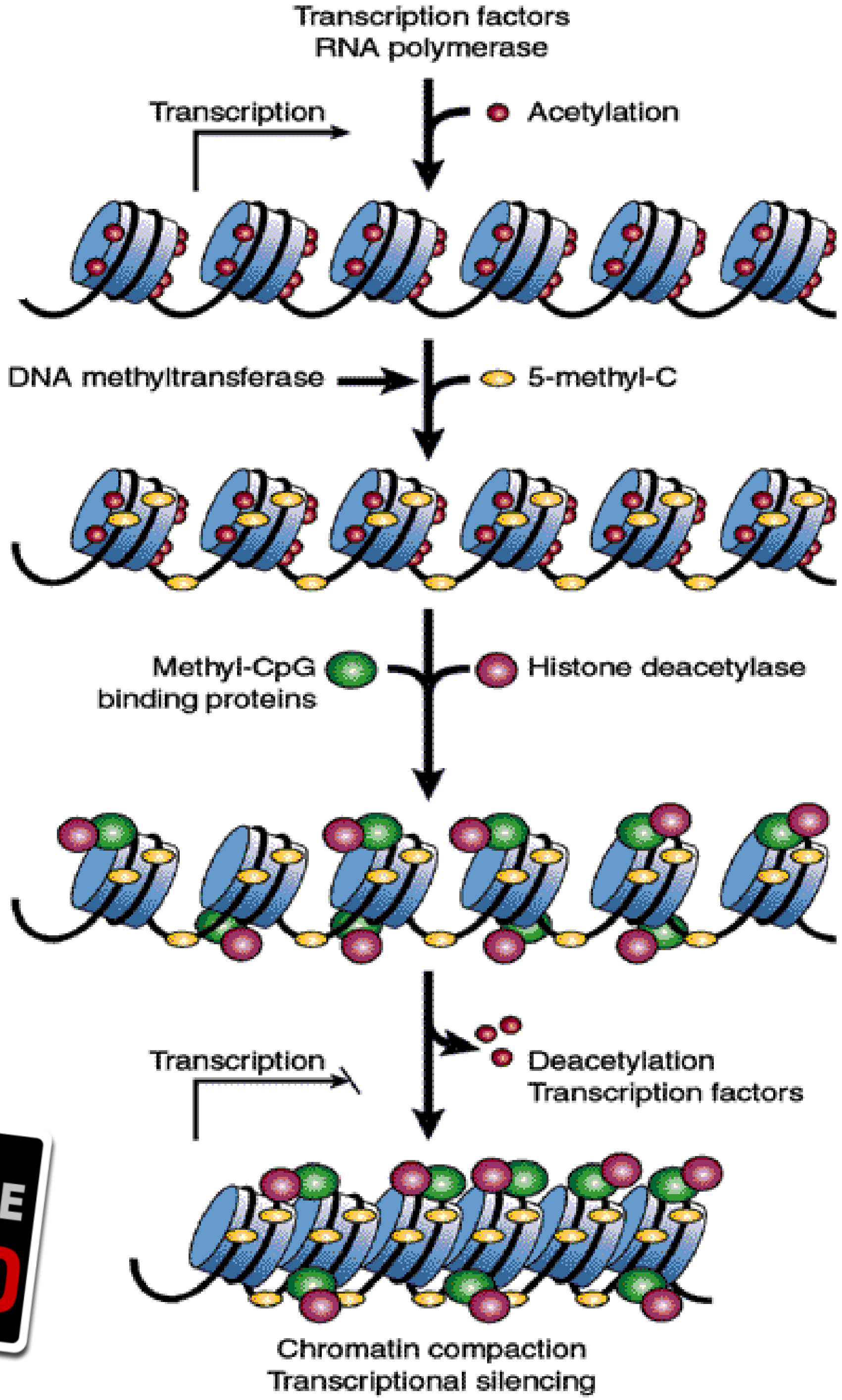
CpG island methylation (II)

- Methylation of promotor CGIs causes gene silencing:
  - MBD protein binds to methylated CGI, recruits histone modifiers resulting in closed chromatin structure

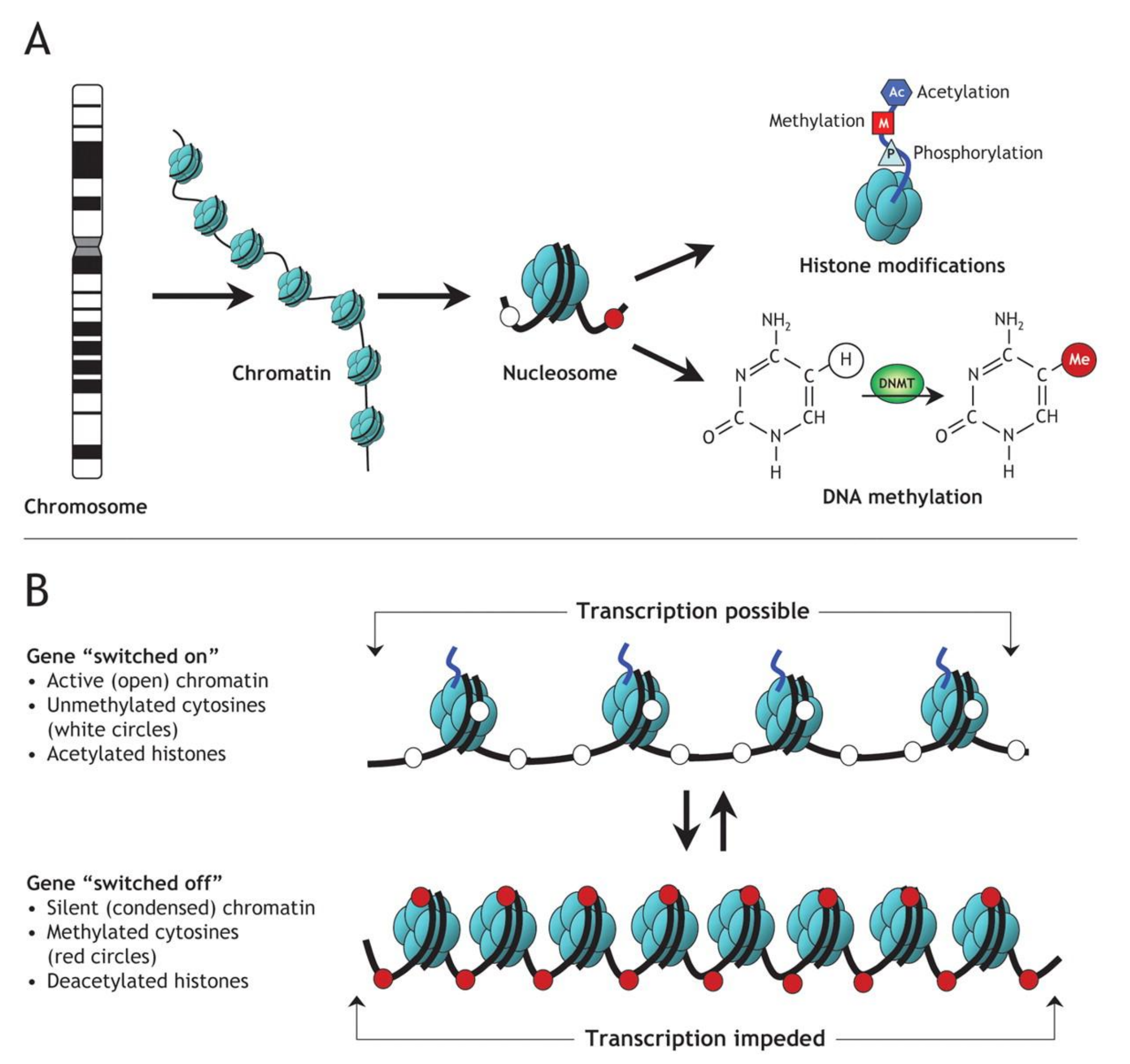


Source: *The Scientist* 2005, **19**(12)**:**18

Interplay between CpG methylation and histone modifications



Interplay between CpG methylation and histone modification



Gene "switched on"
- Active (open) chromatin
- Unmethylated cytosines (white circles)
- Acetylated histones

Gene "switched off"
- Silent (condensed) chromatin
- Methylated cytosines (red circles)
- Deacetylated histones

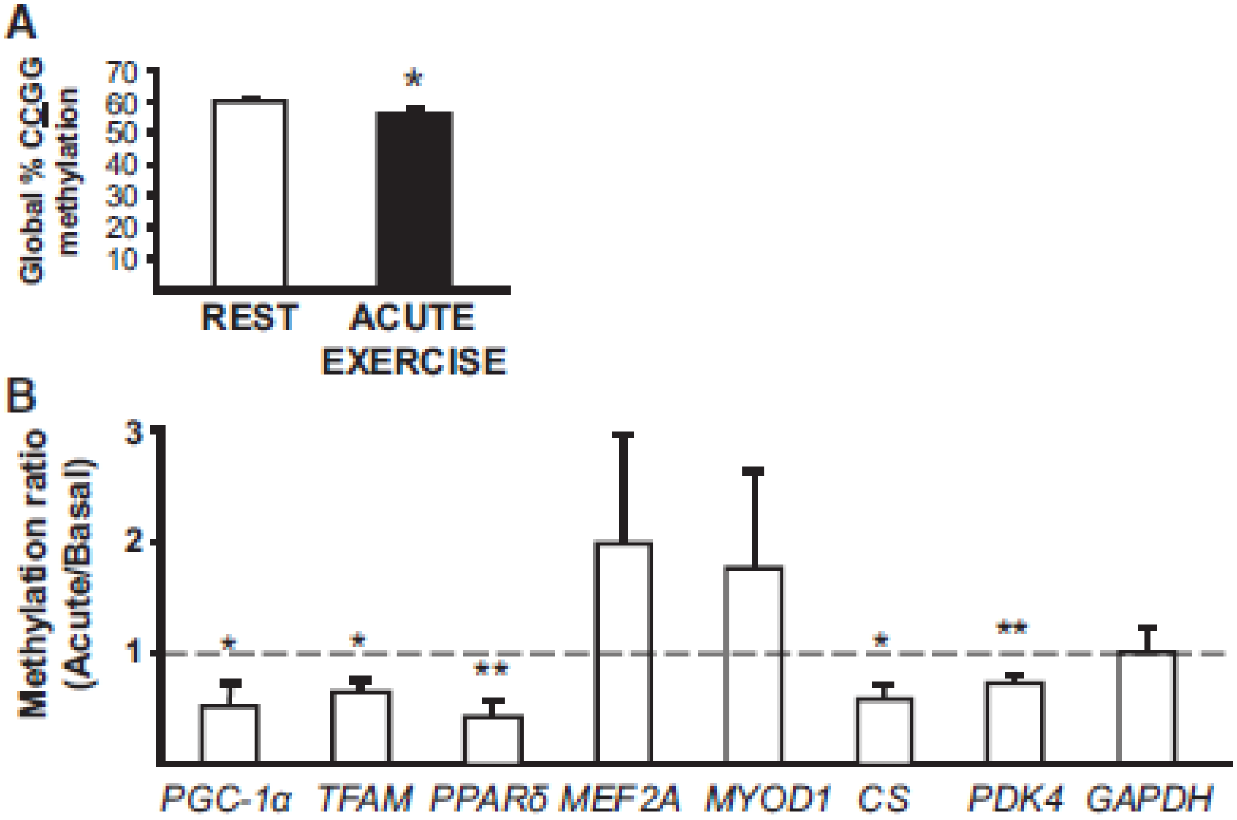
CpG island methylation (III)

- In general CpGs in a single CGI are either all methylated or all unmethylated:
  - Gradients across tissue for multiple copies

- When comparing phenotype *X* to phenotype *R*:
  - CGI **hypermethylation** (methylated in *X*, unmethylated in *R*)
  - CGI **hypomethylation** (vice versa)

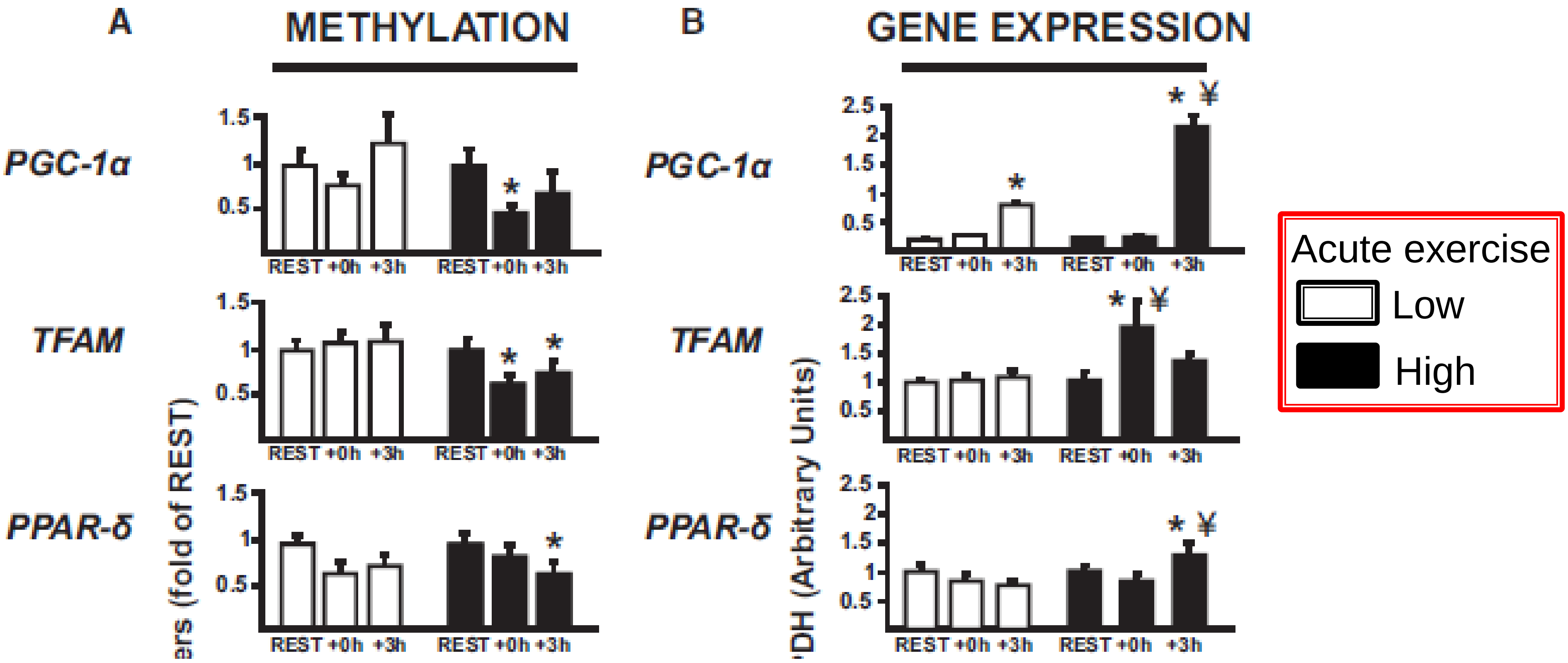
- Methylation blocks transcription, but de-methylation **does not** mediate transcription:
  - an appropriate (set of) transcription factor(s) is still required

Acute exercise remodels DNA methylation in skeletal muscle



Barres et al. (2012) Dynamic DNA methylation Remodeling after Exercise. *Cell Metabolism*

Exercise-induced promoter hypomethylation



Barres et al. (2012) Dynamic DNA methylation Remodeling after Exercise. *Cell Metabolism*

Natural Roles of DNA Methylation in Mammalian System

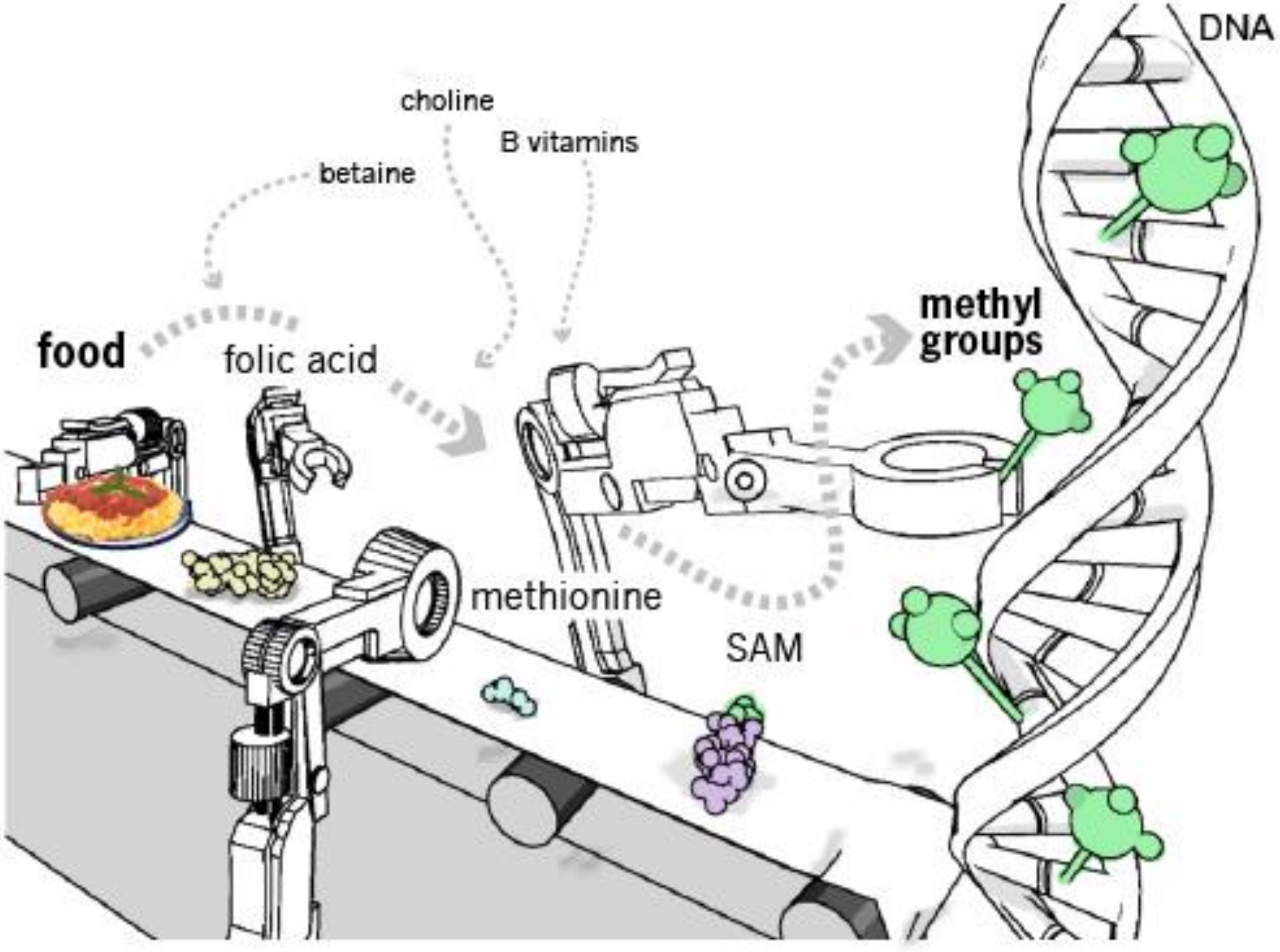
- Tissue specific expression controls
- Imprinting
- X chromosome inactivation
- Heterochromatin maintenance
- Developmental controls

DNA methylation and disease

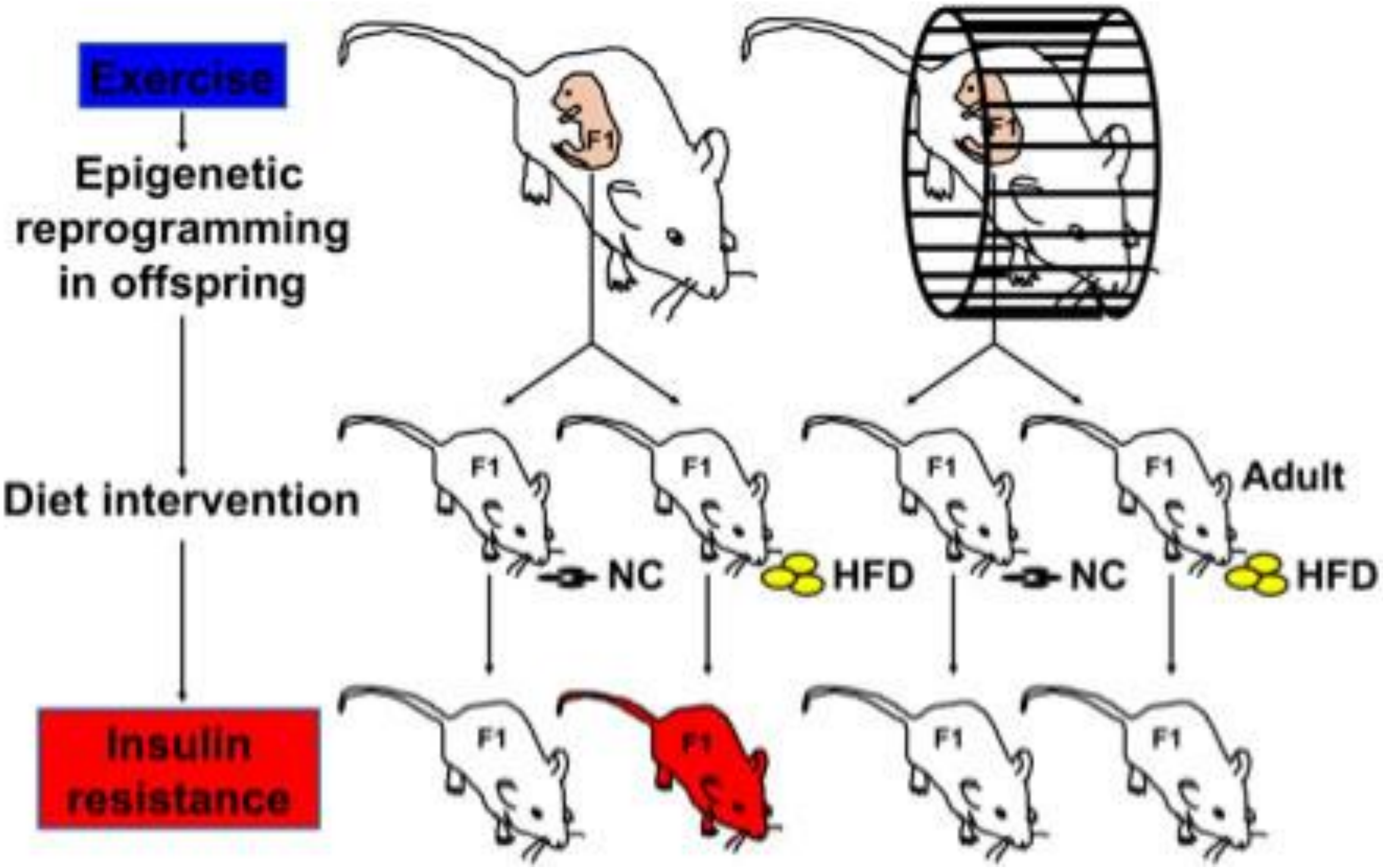
- Cancer:
  - **hypermethylation** of promotor CGIs is found in tumor suppressor genes
  - global **hypomethylation**: structural change

- CGI methylation profiles are used as biomarker profiles
  - Personalized medicine for cancer therapy (similar to SNPs)
  - Identify cancers of unknown origin based on CGI methylation profile

DNA methylation and diet



Epigenetics and maternal exercise



http://cvrc.virginia.edu/yan/

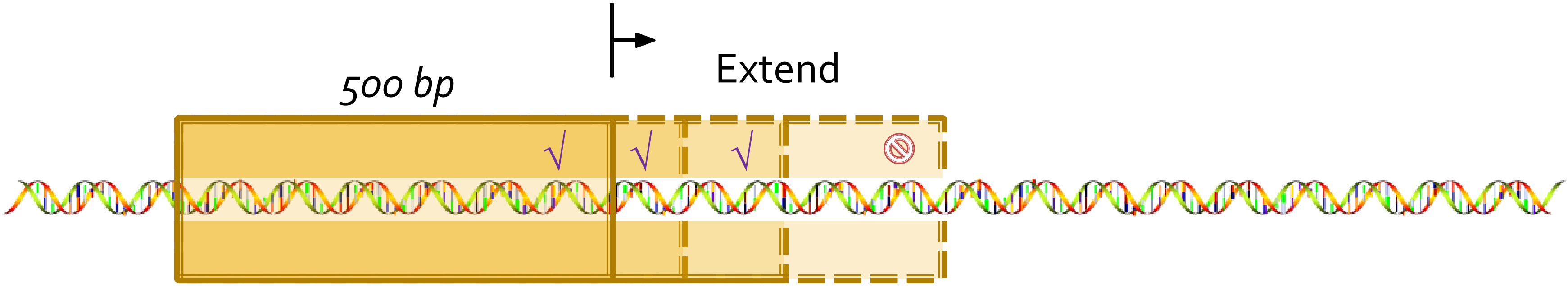
Finding CGIs and histone modifications in UCSC

Defining CpG islands

- Definition:
  - A CGI is a DNA sequence of at least 200 base pairs (bp) long with a GC content of at least 50% and a CpG observed/expected ratio of at least 0.6
  - observed/expected ratio = $\dfrac{[\text{Observed CpGs}] * [\text{Length of sequence}]}{[\text{No Of Cs} * \text{No of Gs}]}$
- A CpG island is genuine when it is proven to be functional:
  - susceptible to differential methylation
    - DNA methylation assay
  - with measurable effect on gene expression
    - Experimental validation of DNA methylation array results
    - Integration of DNA methylation microarray data with transcriptomics data

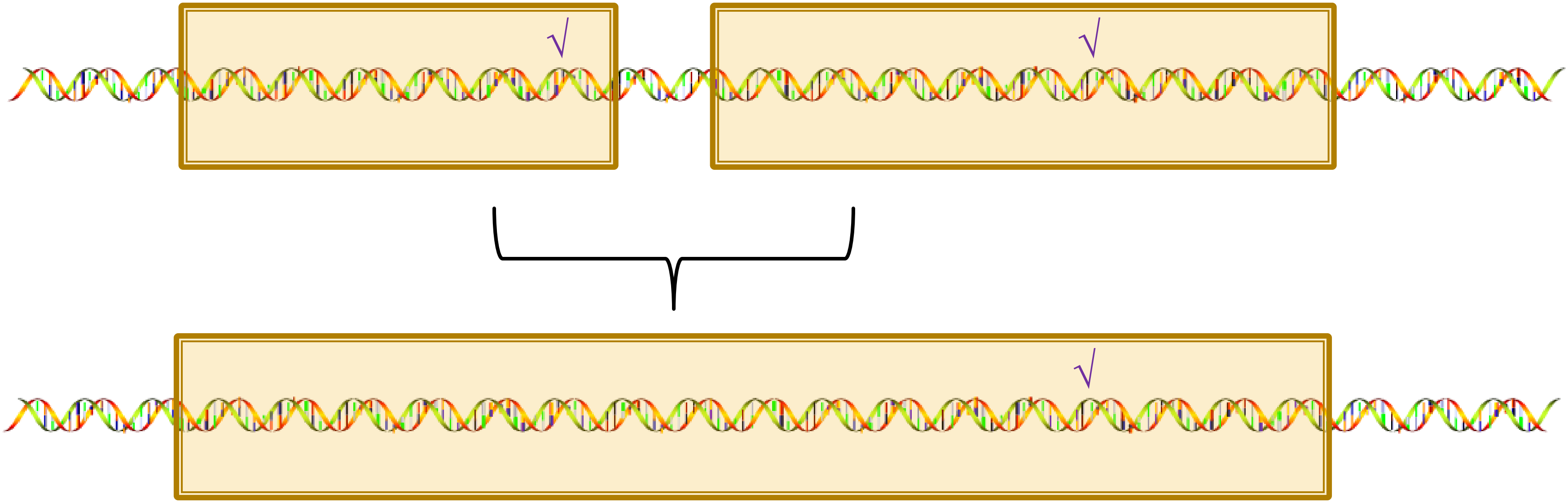
Finding CpG islands

- Algorithm outline:
  1. Move window with **minimum length (200 – 500 bp)** over the genome
  2. If sequence in window meets CGI criteria:
     1. **Extend** until it **no longer meets the criteria**
     2. Record the resulting sequence as **primary CpG island**

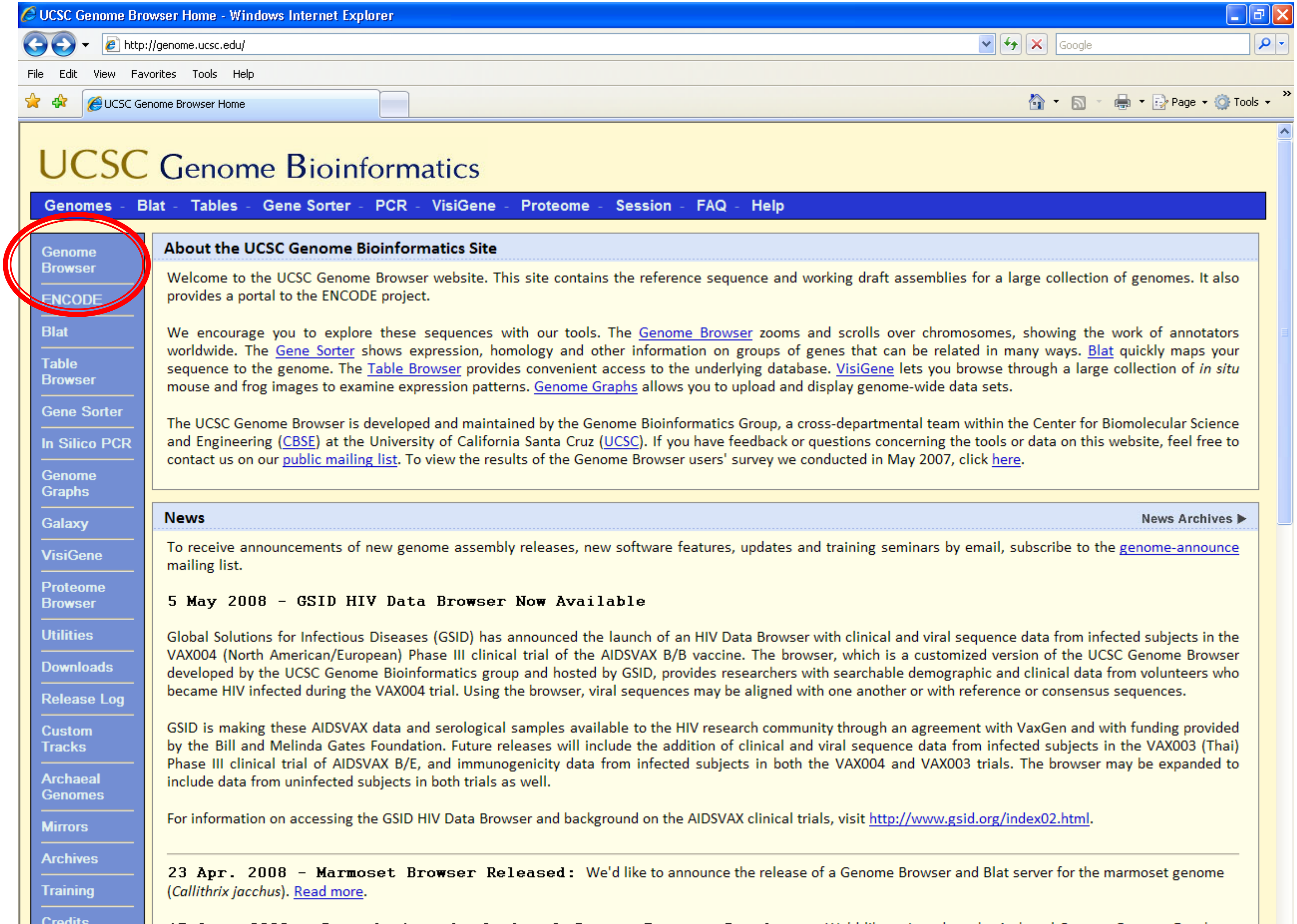


Finding CpG islands

- Algorithm outline:
  1. Move window to end of primary CpG island and repeat
  2. Final step: take close CGIs together



UCSC Genome Browser (http://genome.ucsc.edu/)



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 a portal to the ENCODE project.

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. To view the results of the Genome Browser users' survey we conducted in May 2007, click here.

News

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

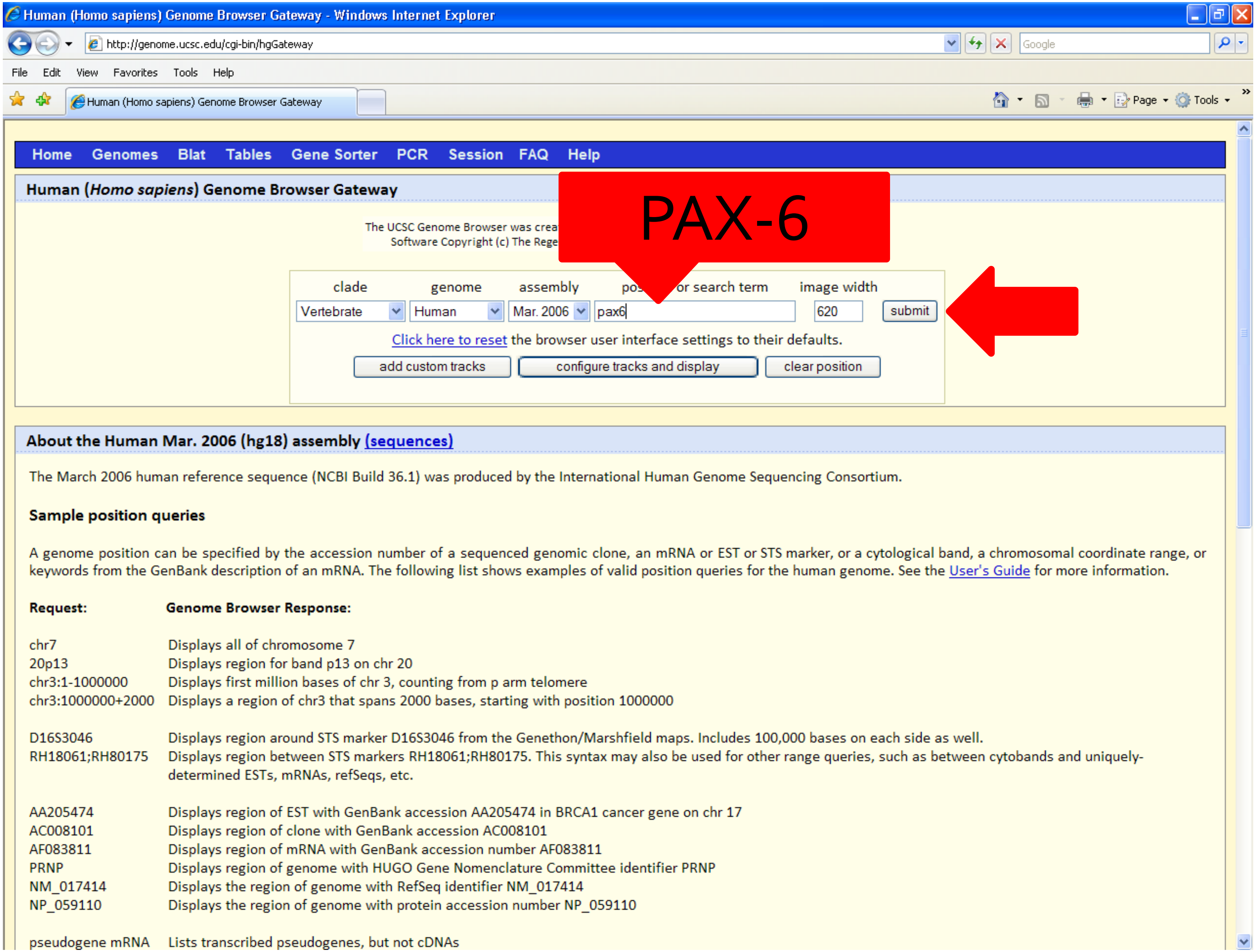
5 May 2008 - GSID HIV Data Browser Now Available

Global Solutions for Infectious Diseases (GSID) has announced the launch of an HIV Data Browser with clinical and viral sequence data from infected subjects in the VAX004 (North American/European) Phase III clinical trial of the AIDSVAX B/B vaccine. The browser, which is a customized version of the UCSC Genome Browser developed by the UCSC Genome Bioinformatics group and hosted by GSID, provides researchers with searchable demographic and clinical data from volunteers who became HIV infected during the VAX004 trial. Using the browser, viral sequences may be aligned with one another or with reference or consensus sequences.

GSID is making these AIDSVAX data and serological samples available to the HIV research community through an agreement with VaxGen and with funding provided by the Bill and Melinda Gates Foundation. Future releases will include the addition of clinical and viral sequence data from infected subjects in the VAX003 (Thai) Phase III clinical trial of AIDSVAX B/E, and immunogenicity data from infected subjects in both the VAX004 and VAX003 trials. The browser may be expanded to include data from uninfected subjects in both trials as well.

For information on accessing the GSID HIV Data Browser and background on the AIDSVAX clinical trials, visit http://www.gsid.org/index02.html.

23 Apr. 2008 - Marmoset Browser Released: We'd like to announce the release of a Genome Browser and Blat server for the marmoset genome (*Callithrix jacchus*). Read more.


Human (Homo sapiens) Genome Browser Gateway - Windows Internet Explorer
http://genome.ucsc.edu/cgi-bin/hgGateway
Google
File Edit View Favorites Tools Help
Human (Homo sapiens) Genome Browser Gateway
Page Tools
Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

PAX-6

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was crea
Software Copyright (c) The Rege

clade	genome	assembly	pos or search term	image width	
Vertebrate	Human	Mar. 2006	pax6	620	submit

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

add custom tracks
configure tracks and display
clear position

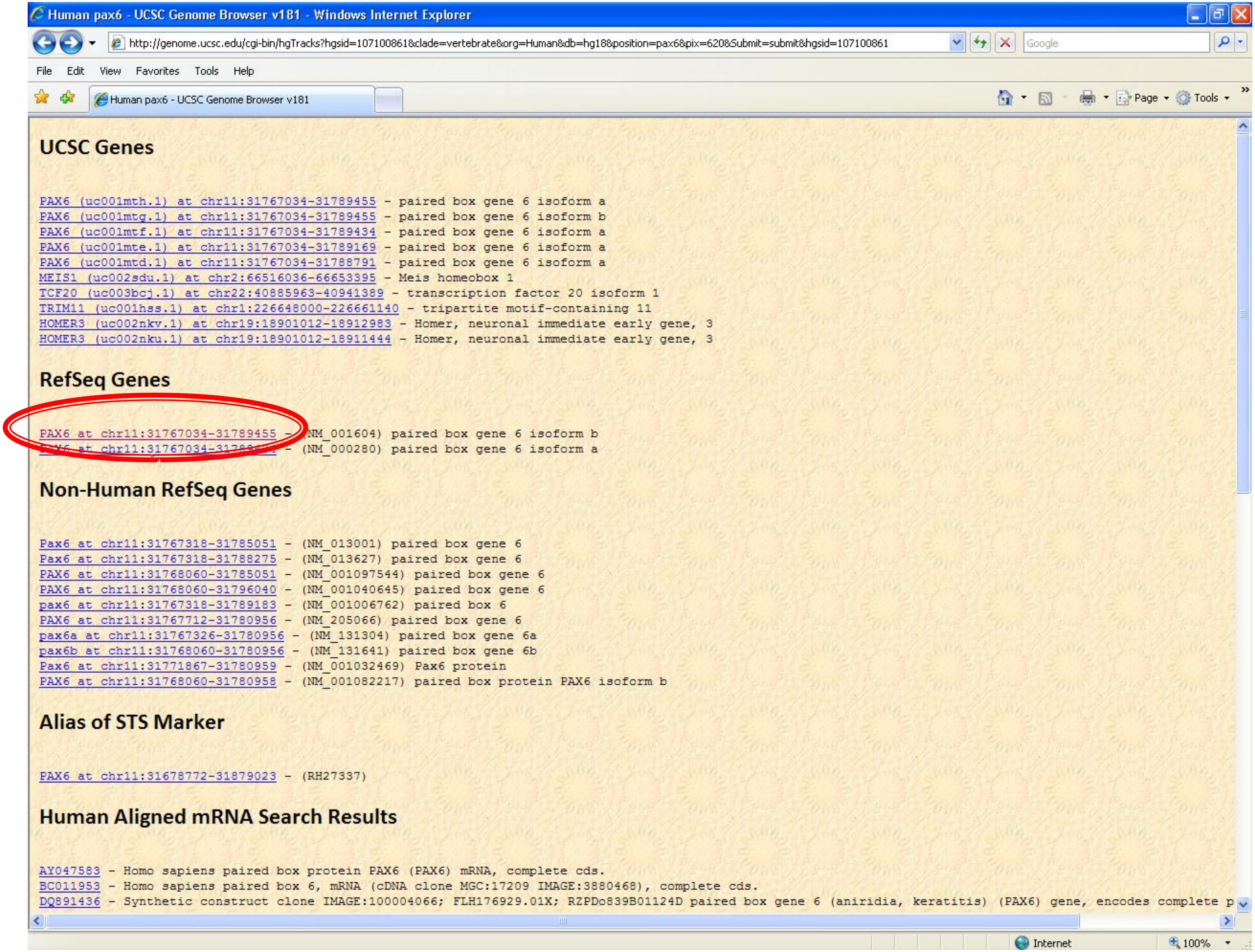
About the Human Mar. 2006 (hg18) assembly (sequences)

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Human Genome Sequencing Consortium.

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, or a cytological band, 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
20p13	Displays region for band p13 on chr 20
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
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175	Displays region between STS markers RH18061;RH80175. This syntax may also be used for other range queries, such as between cytobands and uniquely-determined ESTs, mRNAs, refSeqs, etc.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs



Human pax6 - UCSC Genome Browser v181 - Windows Internet Explorer
http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=107100861&clade=vertebrate&org=Human&db=hg18&position=pax6&pix=620&Submit=submit&hgsid=107100861
Google
File
Edit
View
Favorites
Tools
Help
Human pax6 - UCSC Genome Browser v181
Page
Tools
Internet
100%

UCSC Genes

```
PAX6 (uc001mth.1) at chr11:31767034-31789455 - paired box gene 6 isoform a
PAX6 (uc001mtg.1) at chr11:31767034-31789455 - paired box gene 6 isoform b
PAX6 (uc001mtf.1) at chr11:31767034-31789434 - paired box gene 6 isoform a
PAX6 (uc001mte.1) at chr11:31767034-31789169 - paired box gene 6 isoform a
PAX6 (uc001mtd.1) at chr11:31767034-31788791 - paired box gene 6 isoform a
MEIS1 (uc002sdu.1) at chr2:66516036-66653395 - Meis homeobox 1
TCF20 (uc003bcj.1) at chr22:40885963-40941389 - transcription factor 20 isoform 1
TRIM11 (uc001hss.1) at chr1:226648000-226661140 - tripartite motif-containing 11
HOMER3 (uc002nkv.1) at chr19:18901012-18912983 - Homer, neuronal immediate early gene, 3
HOMER3 (uc002nku.1) at chr19:18901012-18911444 - Homer, neuronal immediate early gene, 3
```

RefSeq Genes

```
PAX6 at chr11:31767034-31789455 - (NM_001604) paired box gene 6 isoform b
PAX6 at chr11:31767034-31789434 - (NM_000280) paired box gene 6 isoform a
```

Non-Human RefSeq Genes

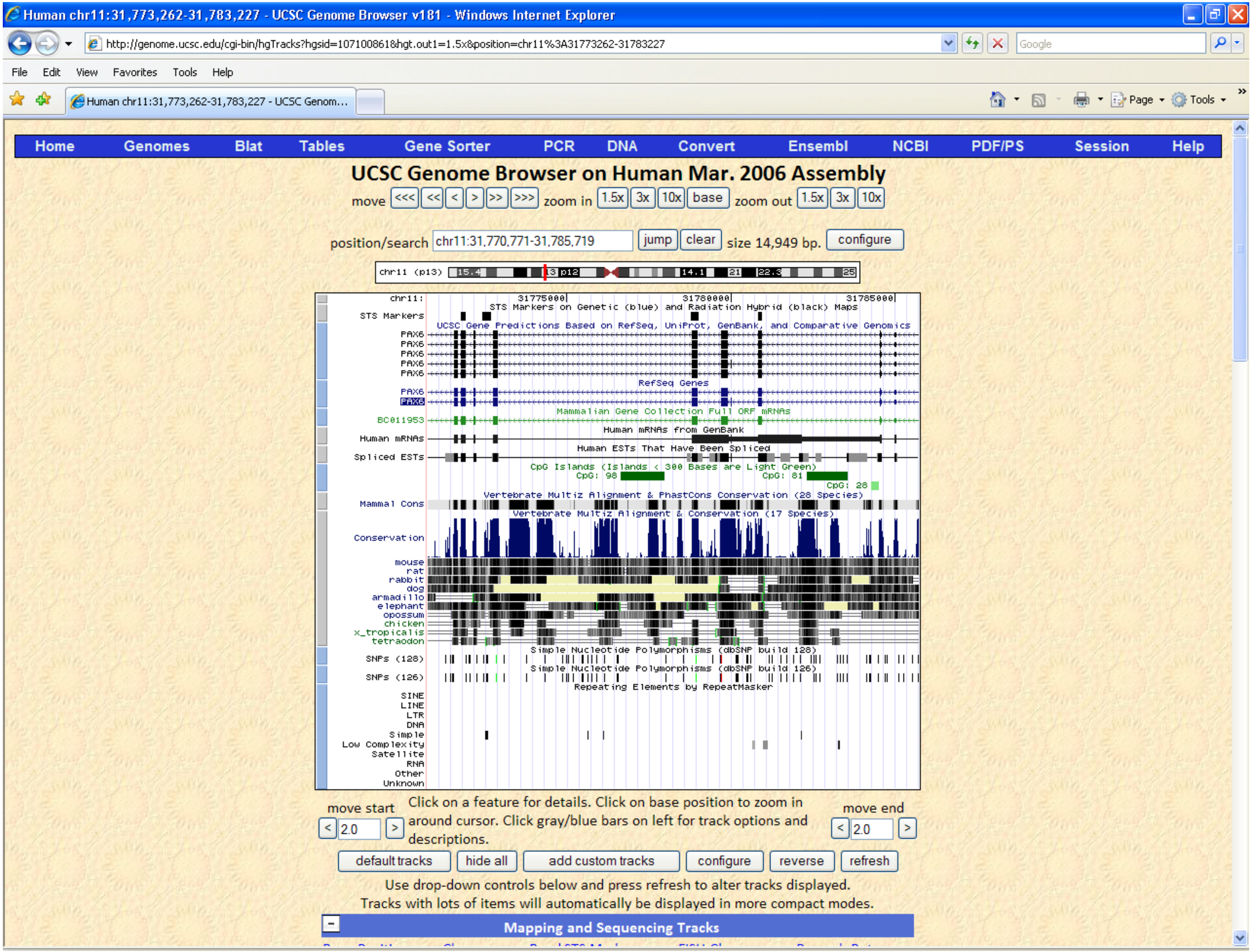
```
Pax6 at chr11:31767318-31785051 - (NM_013001) paired box gene 6
Pax6 at chr11:31767318-31788275 - (NM_013627) paired box gene 6
PAX6 at chr11:31768060-31785051 - (NM_001097544) paired box gene 6
PAX6 at chr11:31768060-31796040 - (NM_001040645) paired box gene 6
pax6 at chr11:31767318-31789183 - (NM_001006762) paired box 6
PAX6 at chr11:31767712-31780956 - (NM_205066) paired box gene 6
pax6a at chr11:31767326-31780956 - (NM_131304) paired box gene 6a
pax6b at chr11:31768060-31780956 - (NM_131641) paired box gene 6b
Pax6 at chr11:31771867-31780959 - (NM_001032469) Pax6 protein
PAX6 at chr11:31768060-31780958 - (NM_001082217) paired box protein PAX6 isoform b
```

Alias of STS Marker

```
PAX6 at chr11:31678772-31879023 - (RH27337)
```

Human Aligned mRNA Search Results

```
AY047583 - Homo sapiens paired box protein PAX6 (PAX6) mRNA, complete cds.
BC011953 - Homo sapiens paired box 6, mRNA (cDNA clone MGC:17209 IMAGE:3880468), complete cds.
DQ891436 - Synthetic construct clone IMAGE:100004066; FLH176929.01X; RZPDo839B01124D paired box gene 6 (aniridia, keratitis) (PAX6) gene, encodes complete p
```



Human chr11:31,773,262-31,783,227 - UCSC Genome Browser v181 - Windows Internet Explorer

http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=107100861&hgt.out1=1.5x&position=chr11%3A31773262-31783227

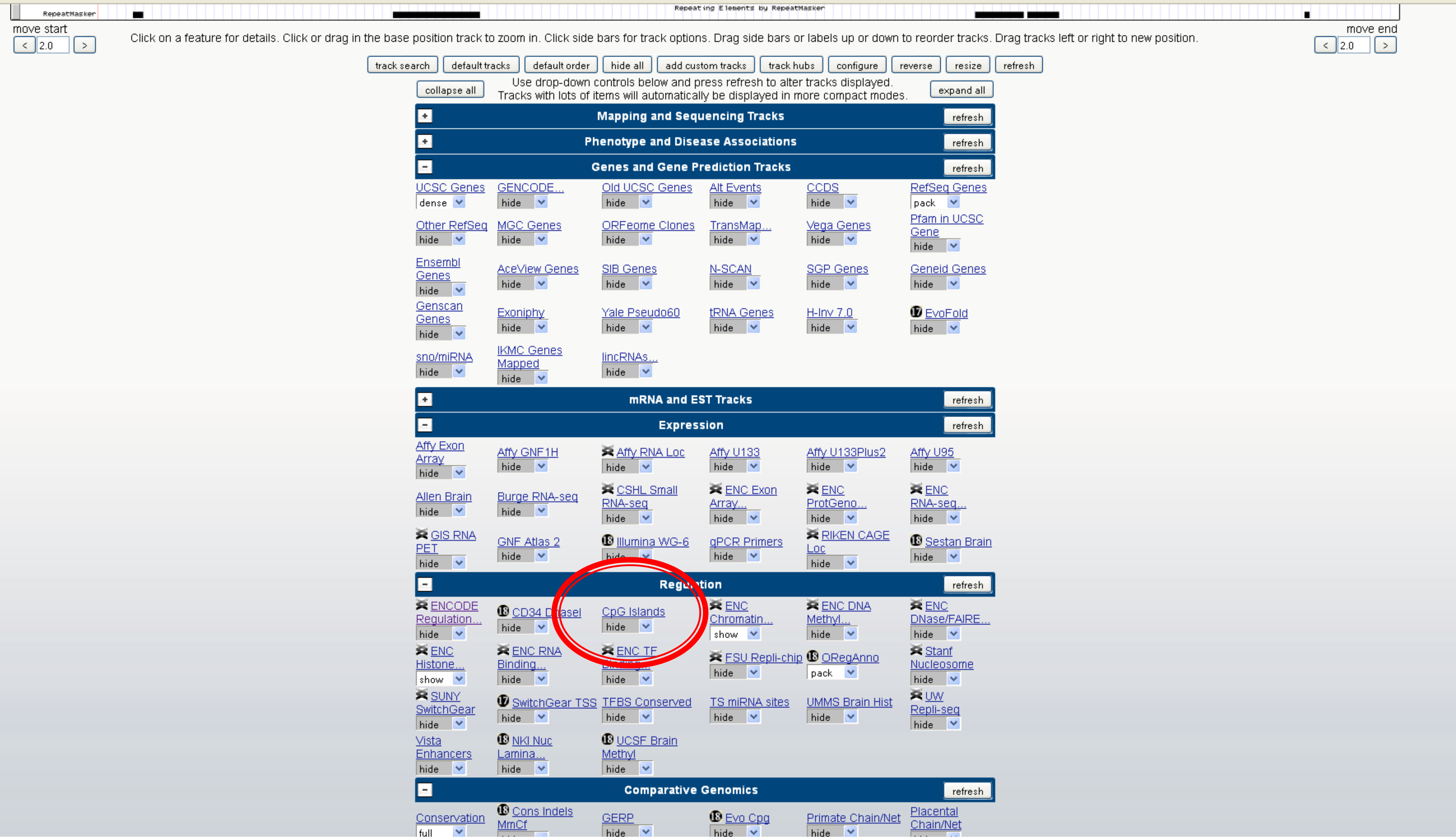
UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr11:31,770,771-31,785,719 size 14,949 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

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.

Mapping and Sequencing Tracks



move start < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

move end < 2.0 >

track search | default tracks | default order | hide all | add custom tracks | track hubs | configure | reverse | resize | refresh

collapse all | expand 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.

+ Mapping and Sequencing Tracks — refresh

+ Phenotype and Disease Associations — refresh

− Genes and Gene Prediction Tracks — refresh

UCSC Genes dense	GENCODE... hide	Old UCSC Genes hide	Alt Events hide	CCDS hide	RefSeq Genes pack
Other RefSeq hide	MGC Genes hide	ORFeome Clones hide	TransMap... hide	Vega Genes hide	Pfam in UCSC Gene hide
Ensembl Genes hide	AceView Genes hide	SIB Genes hide	N-SCAN hide	SGP Genes hide	Geneid Genes hide
Genscan Genes hide	Exoniphy hide	Yale Pseudo60 hide	tRNA Genes hide	H-Inv 7.0 hide	EvoFold hide
sno/miRNA hide	IKMC Genes Mapped hide	lincRNAs... hide			

+ mRNA and EST Tracks — refresh

− Expression — refresh

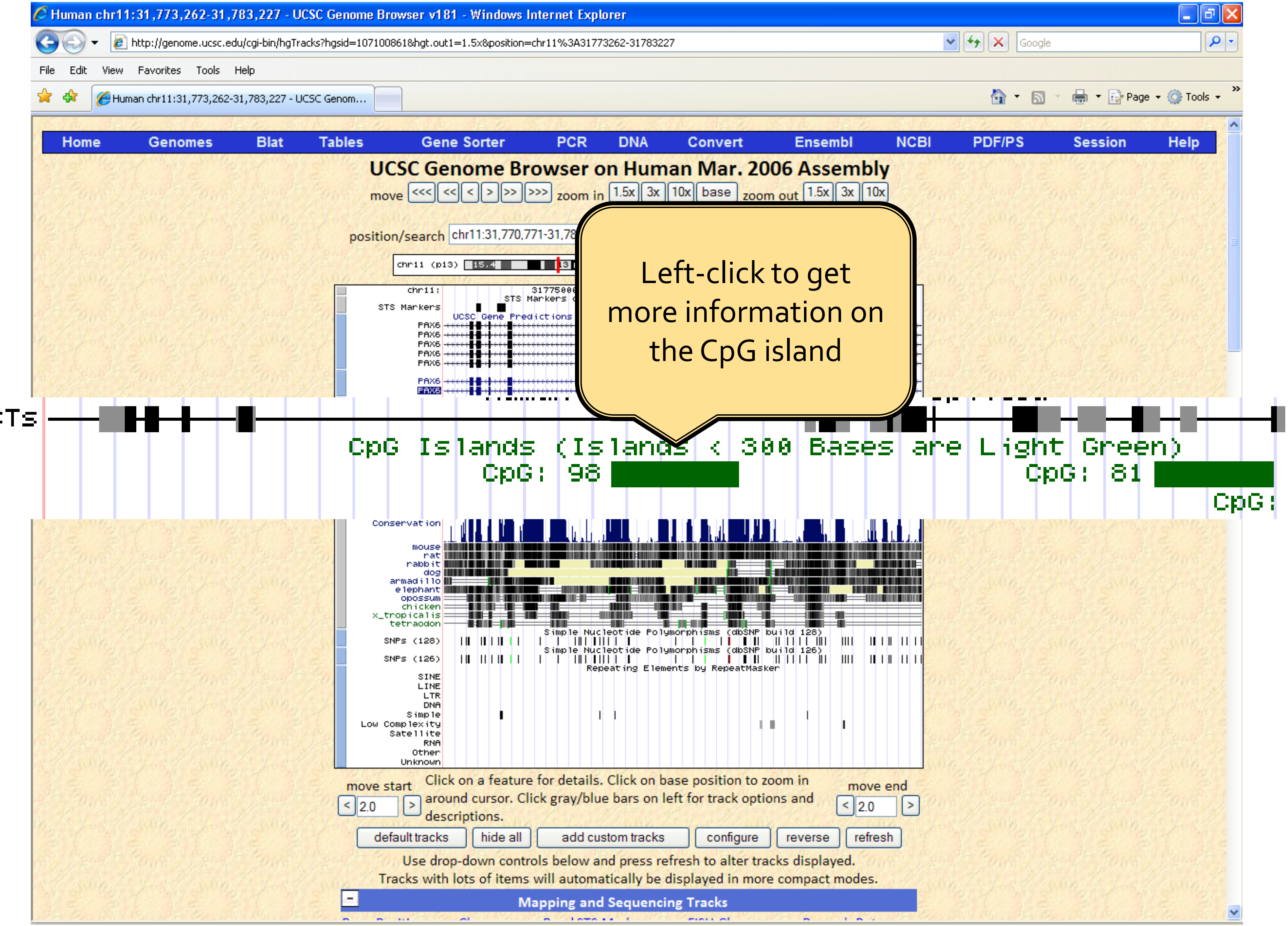
Affy Exon Array hide	Affy GNF1H hide	Affy RNA Loc hide	Affy U133 hide	Affy U133Plus2 hide	Affy U95 hide
Allen Brain hide	Burge RNA-seq hide	CSHL Small RNA-seq hide	ENC Exon Array... hide	ENC ProtGeno... hide	ENC RNA-seq... hide
GIS RNA PET hide	GNF Atlas 2 hide	Illumina WG-6 hide	qPCR Primers hide	RIKEN CAGE Loc hide	Sestan Brain hide

− Regulation — refresh

ENCODE Regulation... hide	CD34 DnaseI hide	CpG Islands hide	ENC Chromatin... show	ENC DNA Methyl... hide	ENC DNase/FAIRE... hide
ENC Histone... show	ENC RNA Binding... hide	ENC TF Binding... hide	FSU Repli-chip hide	ORegAnno pack	Stanf Nucleosome hide
SUNY SwitchGear hide	SwitchGear TSS hide	TFBS Conserved hide	TS miRNA sites hide	UMMS Brain Hist hide	UW Repli-seq hide
Vista Enhancers hide	NKI Nuc Lamina... hide	UCSF Brain Methyl hide			

− Comparative Genomics — refresh

Conservation full	Cons Indels MmCf	GERP hide	Evo Cpg hide	Primate Chain/Net hide	Placental Chain/Net

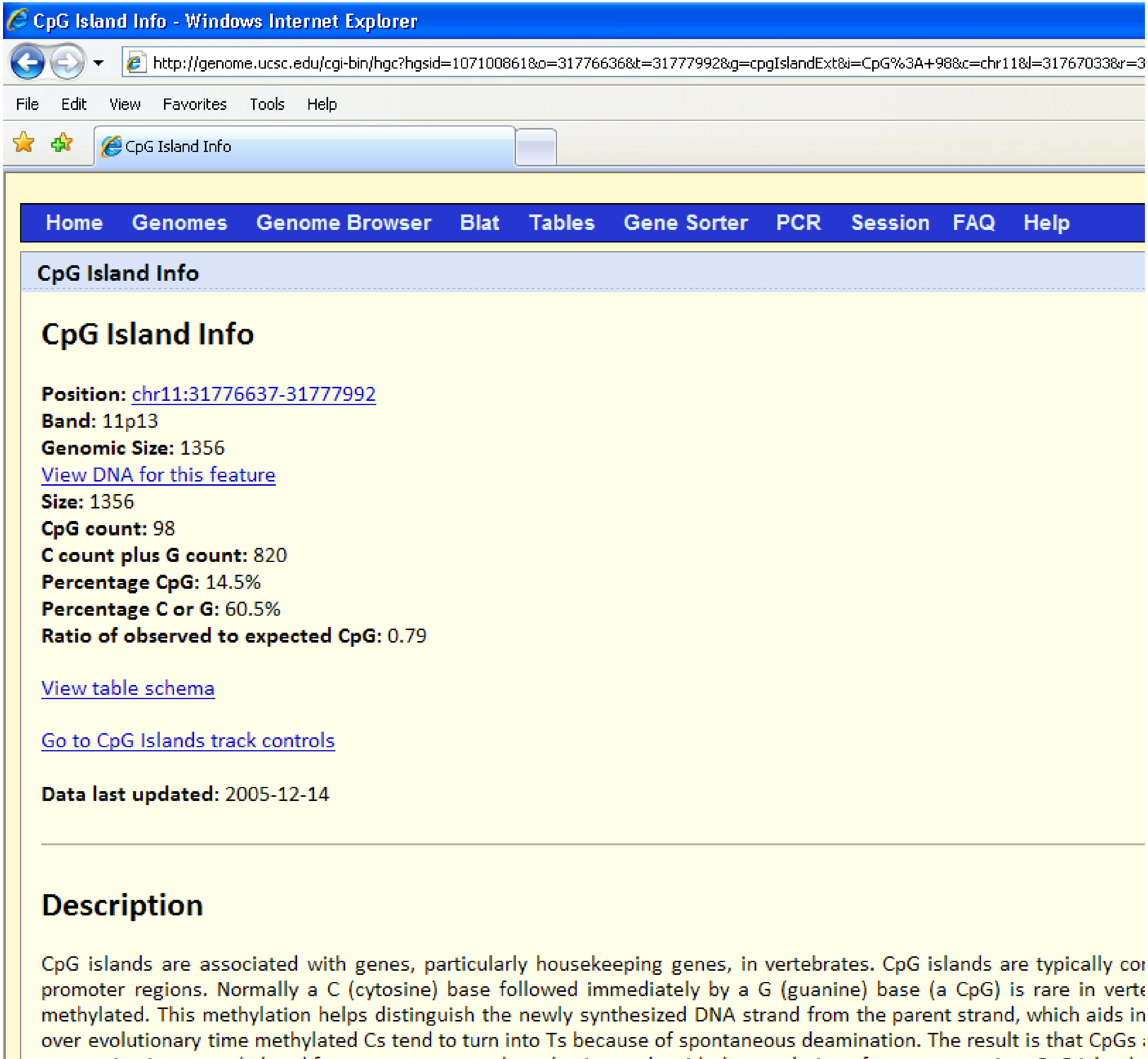


UCSC Genome Browser on Human Mar. 2006 Assembly

Left-click to get more information on the CpG island

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

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.



http://genome.ucsc.edu/cgi-bin/hgc?hgsid=107100861&o=31776636&t=31777992&g=cpgIslandExt&i=CpG%3A+98&c=chr11&l=31767033&r=3

CpG Island Info

Position: chr11:31776637-31777992
Band: 11p13
Genomic Size: 1356
View DNA for this feature
Size: 1356
CpG count: 98
C count plus G count: 820
Percentage CpG: 14.5%
Percentage C or G: 60.5%
Ratio of observed to expected CpG: 0.79

View table schema

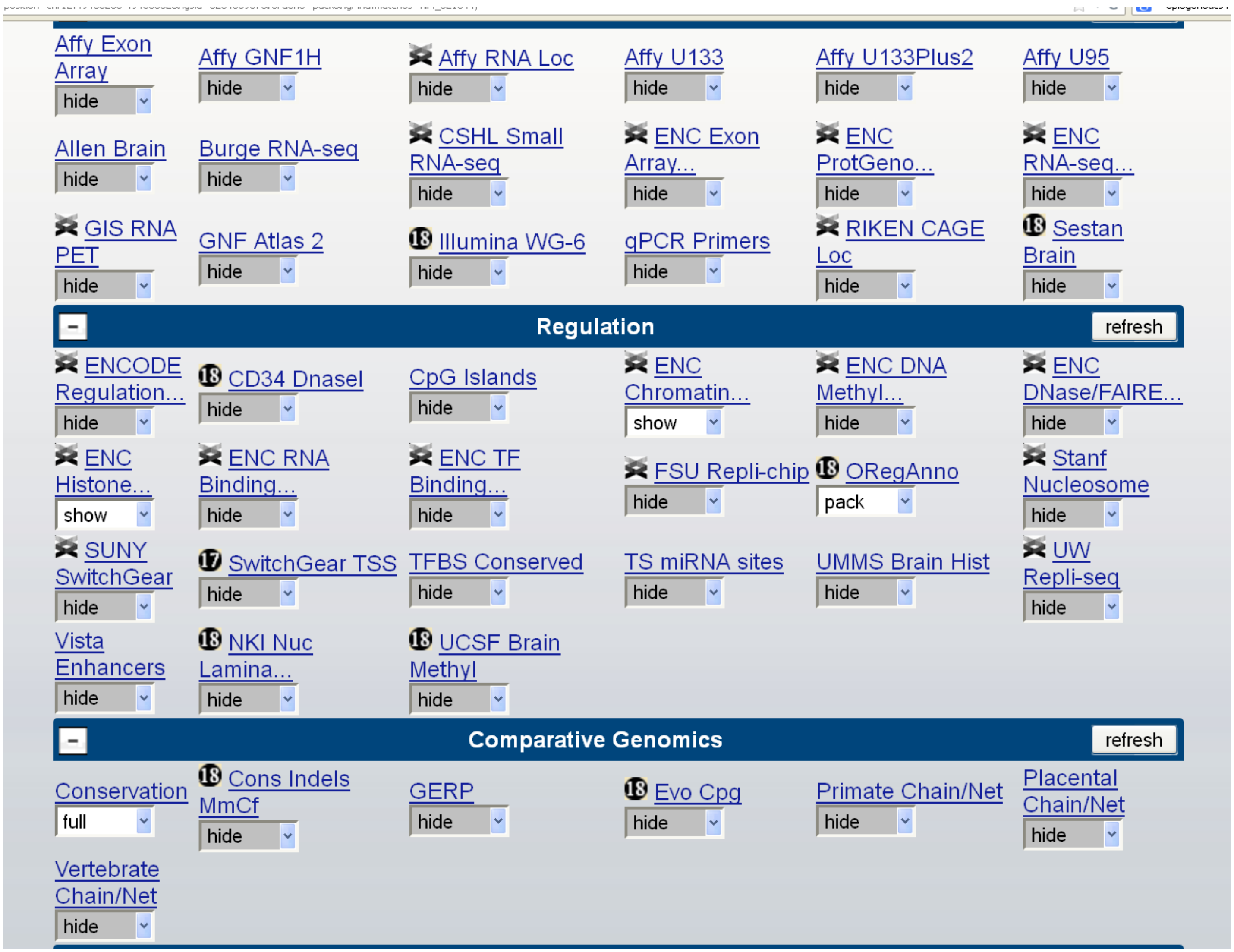
Go to CpG Islands track controls

Data last updated: 2005-12-14

Description

CpG islands are associated with genes, particularly housekeeping genes, in vertebrates. CpG islands are typically co
promoter regions. Normally a C (cytosine) base followed immediately by a G (guanine) base (a CpG) is rare in verte
methylated. This methylation helps distinguish the newly synthesized DNA strand from the parent strand, which aids in
over evolutionary time methylated Cs tend to turn into Ts because of spontaneous deamination. The result is that CpGs

ENCODE data in UCSC



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	CSHL Small RNA-seq hide	ENC Exon Array... hide	ENC ProtGeno... hide	ENC RNA-seq... hide
GIS RNA PET hide	GNF Atlas 2 hide	⑱ Illumina WG-6 hide	qPCR Primers hide	RIKEN CAGE Loc hide	⑱ Sestan Brain hide

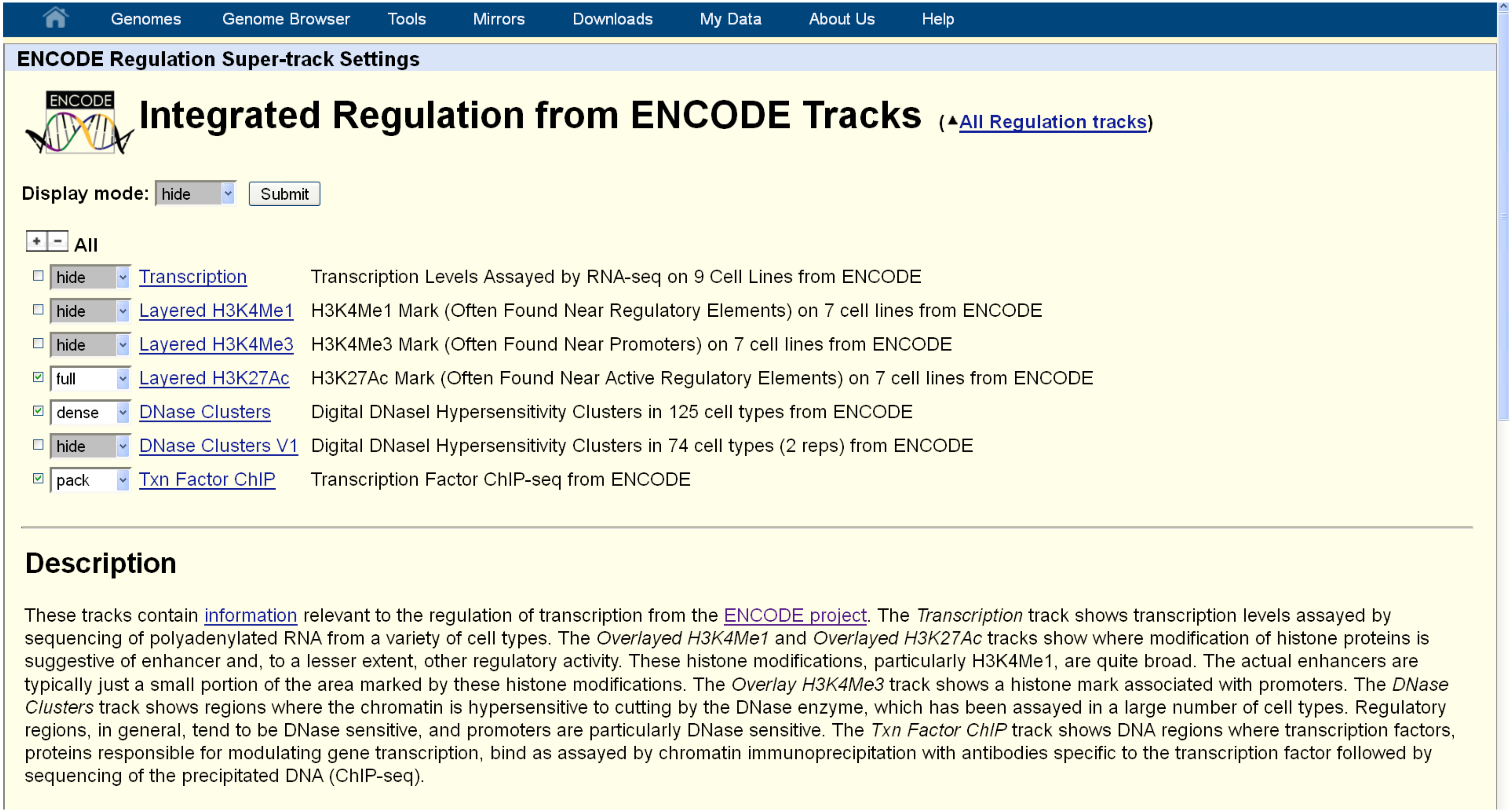
Regulation refresh

ENCODE Regulation... hide	⑱ CD34 DnaseI hide	CpG Islands hide	ENC Chromatin... show	ENC DNA Methyl... hide	ENC DNase/FAIRE... hide
ENC Histone... show	ENC RNA Binding... hide	ENC TF Binding... hide	FSU Repli-chip hide	⑱ ORegAnno pack	Stanf Nucleosome hide
SUNY SwitchGear hide	⑰ SwitchGear TSS hide	TFBS Conserved hide	TS miRNA sites hide	UMMS Brain Hist hide	UW Repli-seq hide
Vista Enhancers hide	⑱ NKI Nuc Lamina... hide	⑱ UCSF Brain Methyl hide			

Comparative Genomics refresh

Conservation full	⑱ Cons Indels MmCf hide	GERP hide	⑱ Evo Cpg hide	Primate Chain/Net hide	Placental Chain/Net hide
Vertebrate Chain/Net hide					

ENCODE regulation track



ENCODE Regulation Super-track Settings

Integrated Regulation from ENCODE Tracks (▲All Regulation tracks)

Display mode: hide Submit

All

☐	hide	Transcription	Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE
☐	hide	Layered H3K4Me1	H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
☐	hide	Layered H3K4Me3	H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
☑	full	Layered H3K27Ac	H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
☑	dense	DNase Clusters	Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
☐	hide	DNase Clusters V1	Digital DNaseI Hypersensitivity Clusters in 74 cell types (2 reps) from ENCODE
☑	pack	Txn Factor ChIP	Transcription Factor ChIP-seq from ENCODE

Description

These tracks contain information relevant to the regulation of transcription from the ENCODE project. The *Transcription* track shows transcription levels assayed by sequencing of polyadenylated RNA from a variety of cell types. The *Overlayed H3K4Me1* and *Overlayed H3K27Ac* tracks show where modification of histone proteins is suggestive of enhancer and, to a lesser extent, other regulatory activity. These histone modifications, particularly H3K4Me1, are quite broad. The actual enhancers are typically just a small portion of the area marked by these histone modifications. The *Overlay H3K4Me3* track shows a histone mark associated with promoters. The *DNase Clusters* track shows regions where the chromatin is hypersensitive to cutting by the DNase enzyme, which has been assayed in a large number of cell types. Regulatory regions, in general, tend to be DNase sensitive, and promoters are particularly DNase sensitive. The *Txn Factor ChIP* track shows DNA regions where transcription factors, proteins responsible for modulating gene transcription, bind as assayed by chromatin immunoprecipitation with antibodies specific to the transcription factor followed by sequencing of the precipitated DNA (ChIP-seq).

Measuring regulatory events genome wide

Key approach: Enrichment analysis

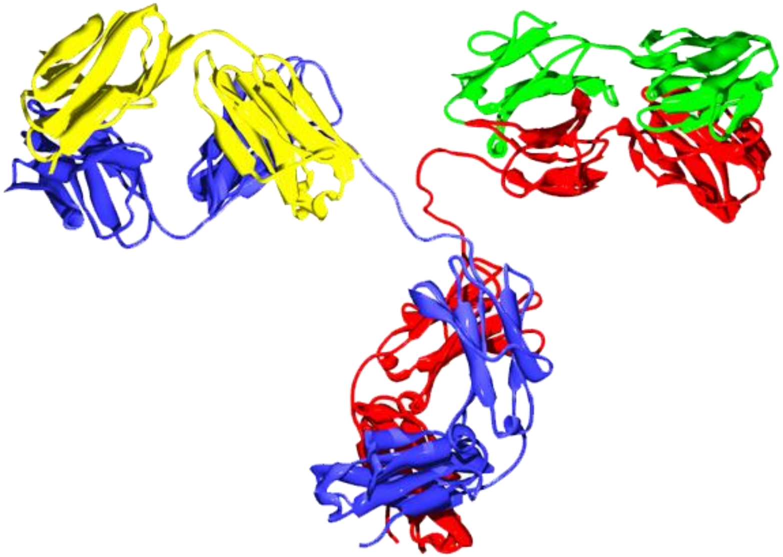
DNA sample that is biologically enriched for regulatory sequences

VS

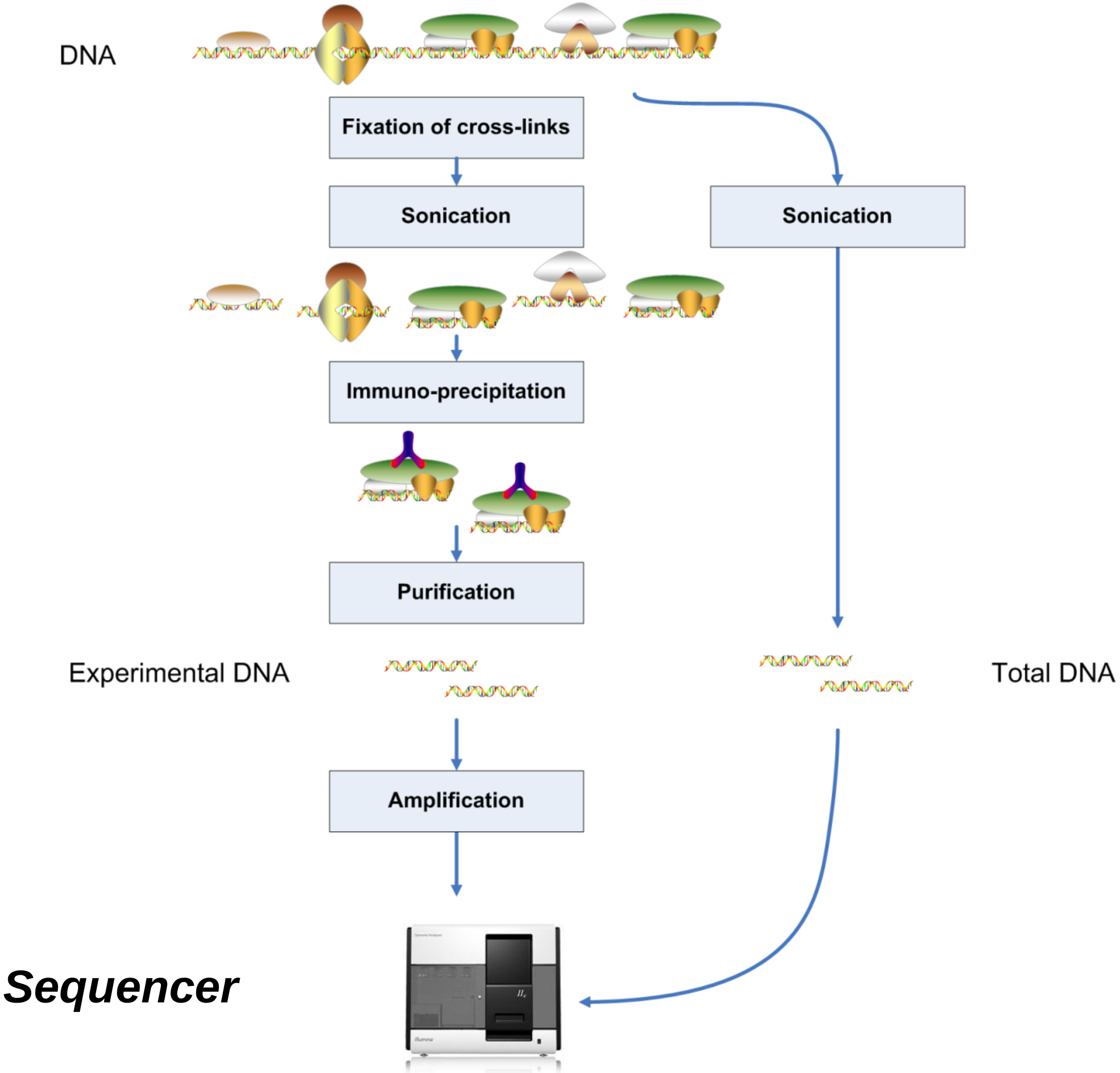
DNA reference sample containing all sequences found in the genome

Assays to determine enrichment

- General enrichment assay:
  - Chromatin immuno-precipitation (**ChIP**)
  - IP any DNA bound protein, as long as suitable anti-body is available



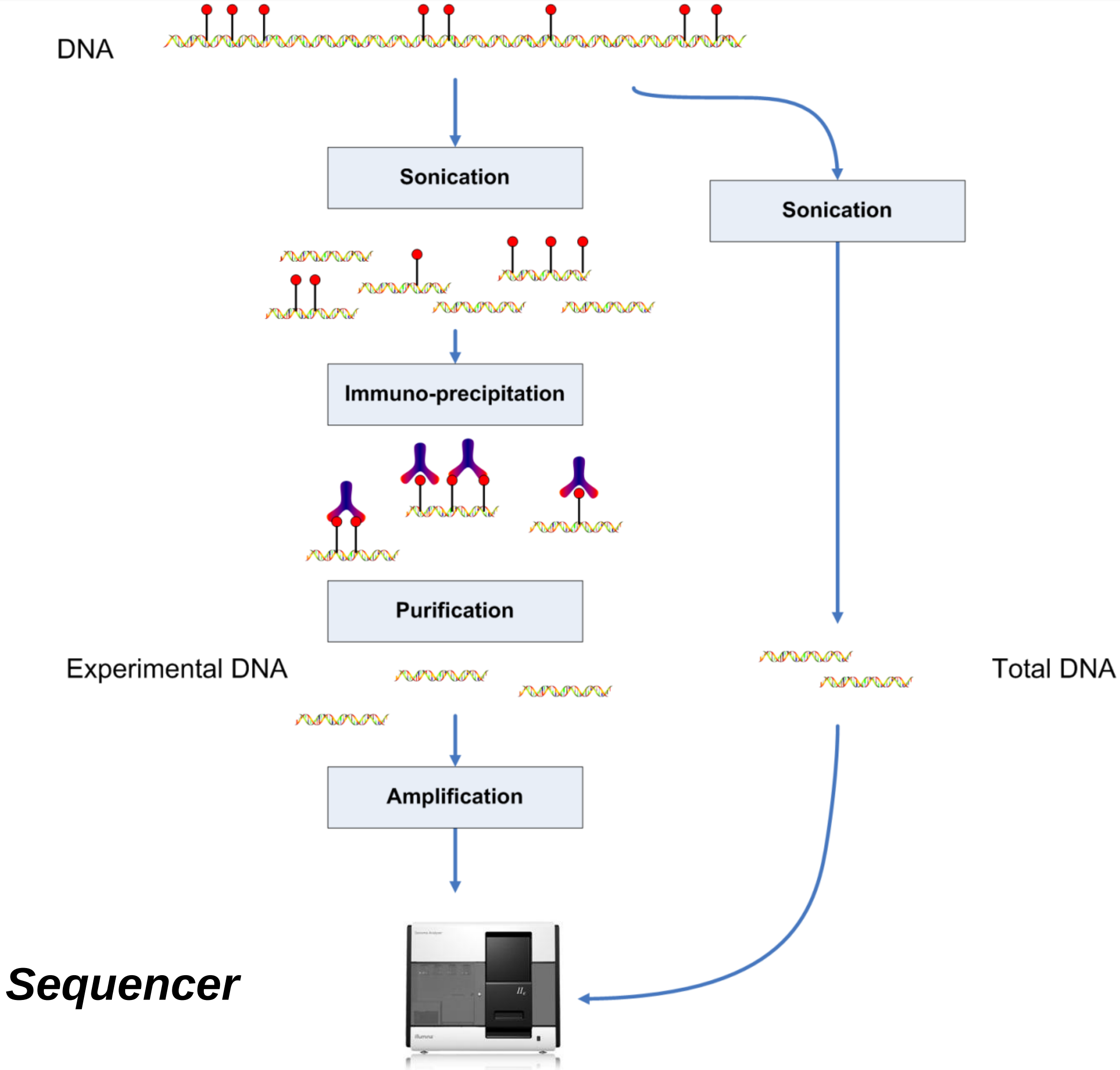
General enrichment assay



Assays to determine enrichment (2)

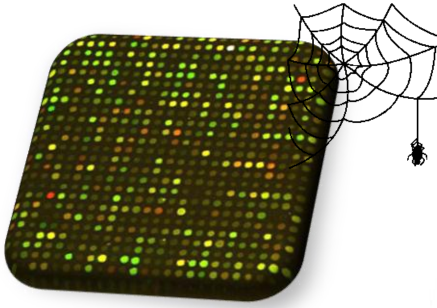
- DNA methylation:
  - Methyl-DNA immuno-precipitation (**MeDIP**)
    - IP methylated DNA directly
    - Biased towards CGI
  - Methylation sensitive restriction enzym based assay (e.g. **McrBc**)
    - Cut up methylated DNA, prevent it from being PCR amplified
    - Left with 'total DNA – methylated DNA'

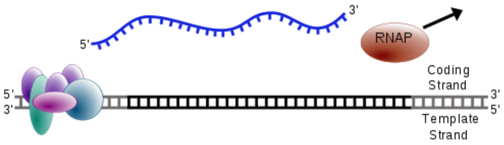
DNA methylation assay

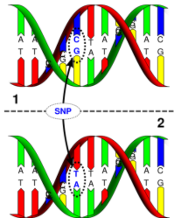


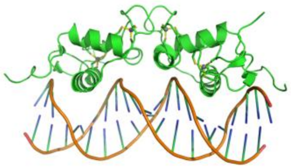
Technology

- Microarray technology
- Next generation sequencing
- Both have many applications
  - Gene expression
  - MicroRNA expression
  - Genetic variation
  - DNA methylation
  - DNA protein binding
  - …



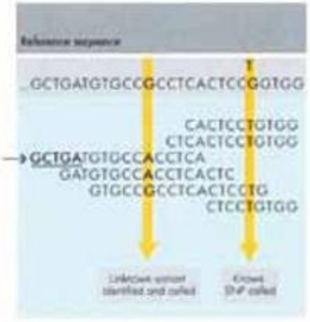






Next generation sequencing

- Sequence sonicated DNA sample:
  - Results: loads of short reads (30 ~ 50 bp)
- Map reads back to the genome (BLAST):
  - Usually keep unique hits only
- Annotate reads to genes



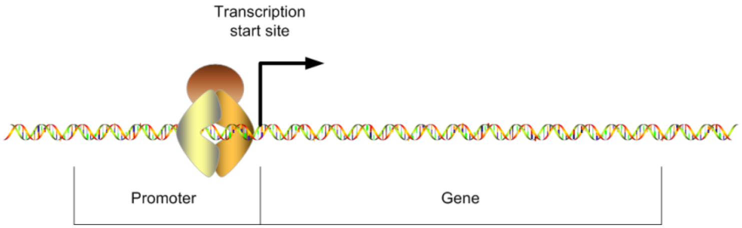
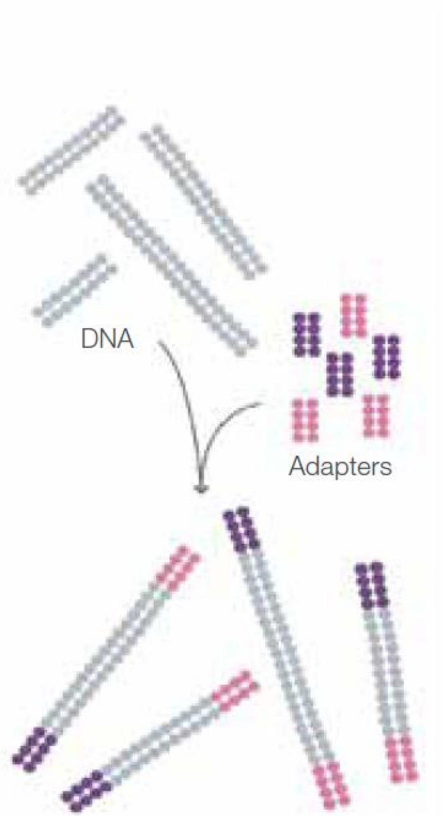
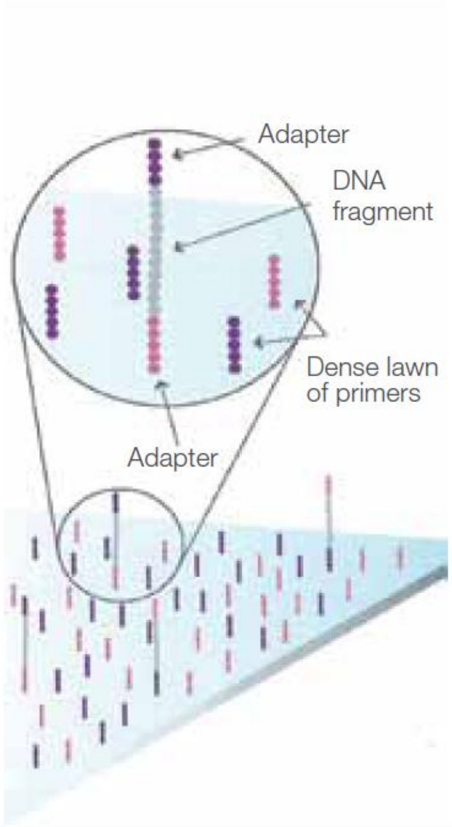


Figure 2: Prepare Genomic DNA Sample



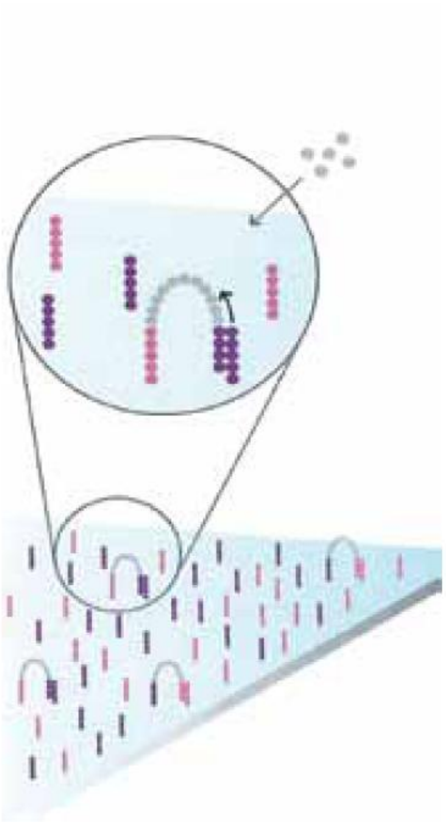
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Figure 3: Attach DNA to Surface



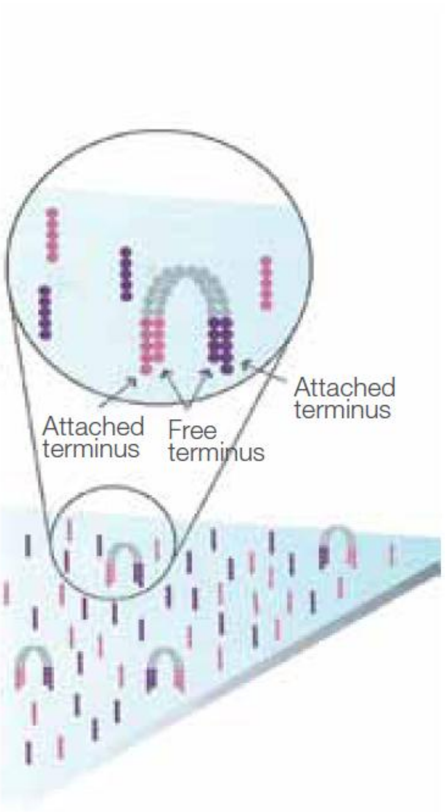
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

Figure 4: Bridge Amplification



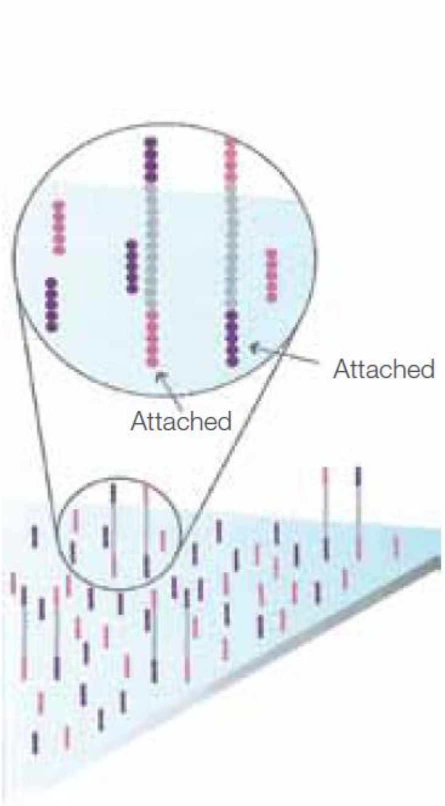
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

Figure 5: Fragments Become Double Stranded



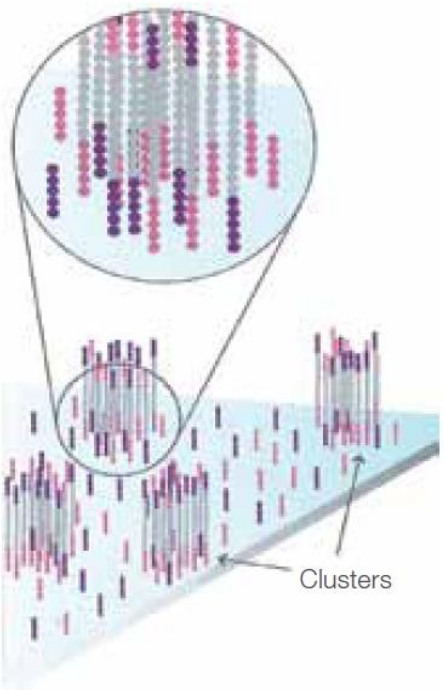
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

Figure 6: Denature the Double-Standed Molecules



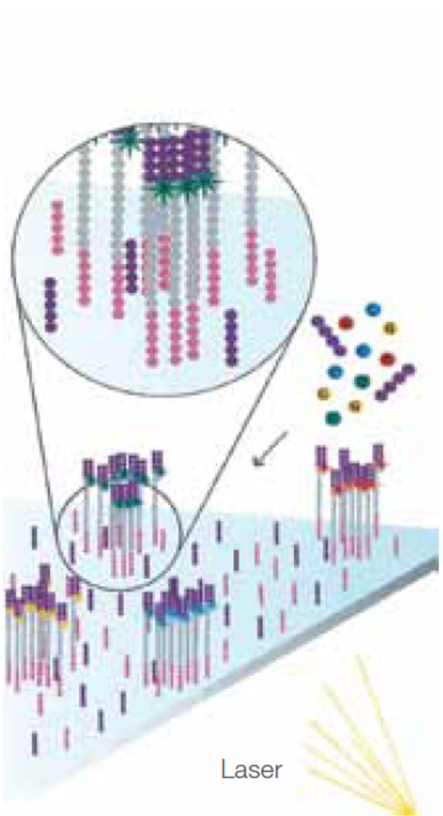
Denaturation leaves single-stranded templates anchored to the substrate.

Figure 7: Complete Amplification



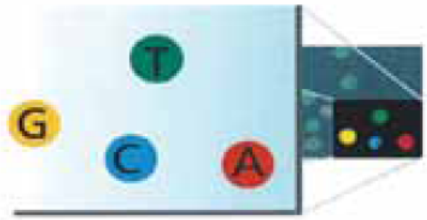
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Figure 8: Determine First Base



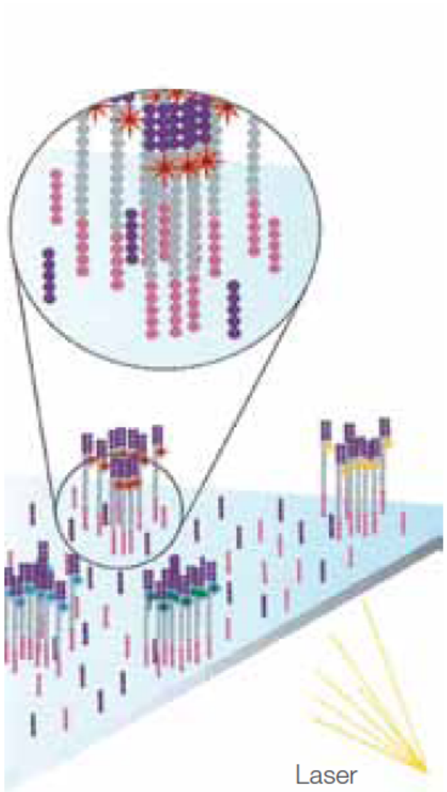
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

Figure 9: Image First Base



After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

Figure 10: Determine Second Base

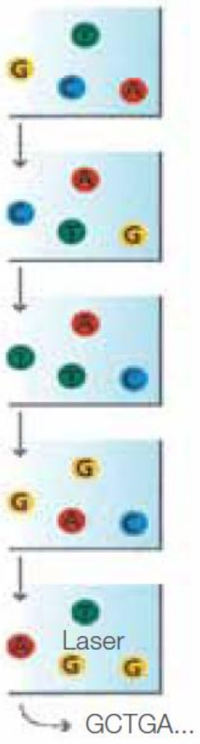


The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

Figure 11: Image Second Chemistry Cycle

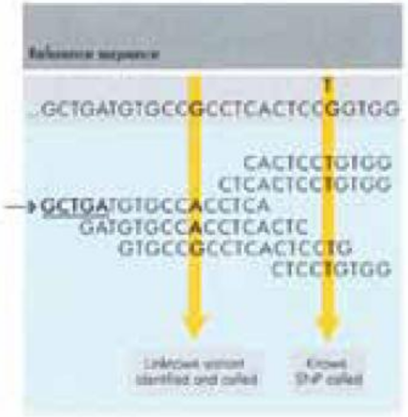
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

Figure 12: Sequencing Over Multiple Chemistry Cycles



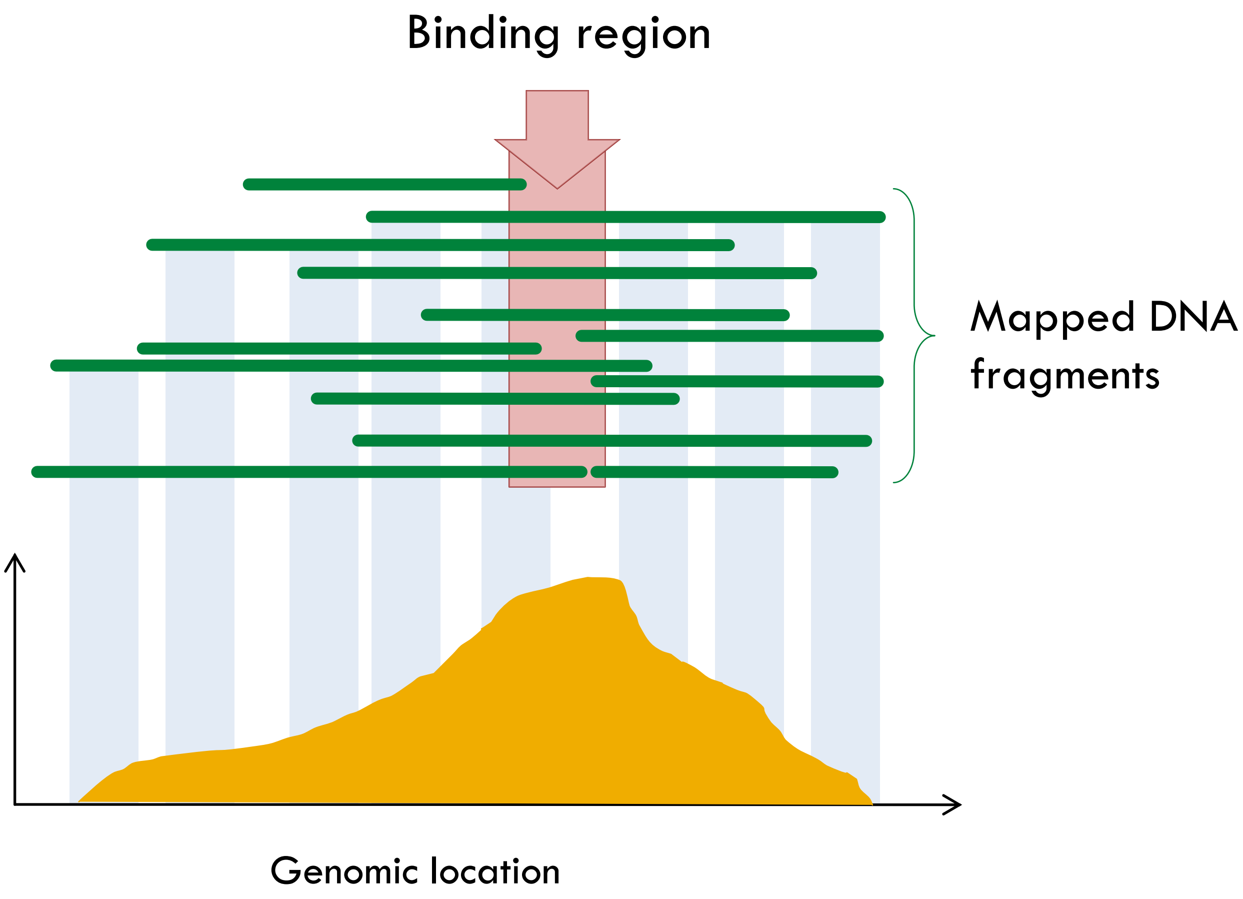
The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

Figure 13: Align Data



The data are aligned and compared to a reference, and sequencing differences are identified.

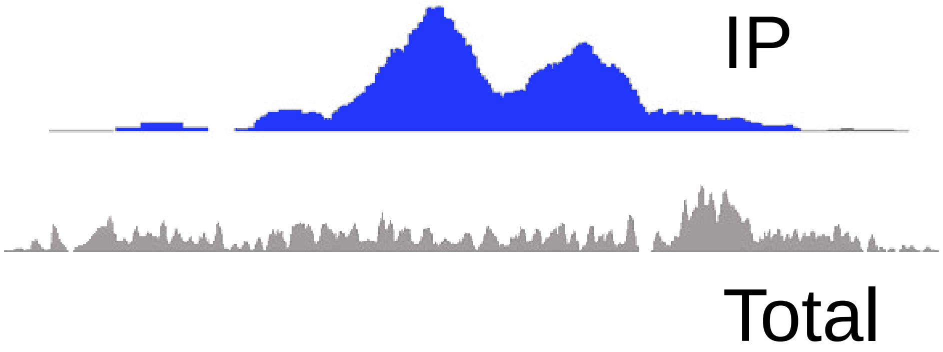
Summarize after mapping



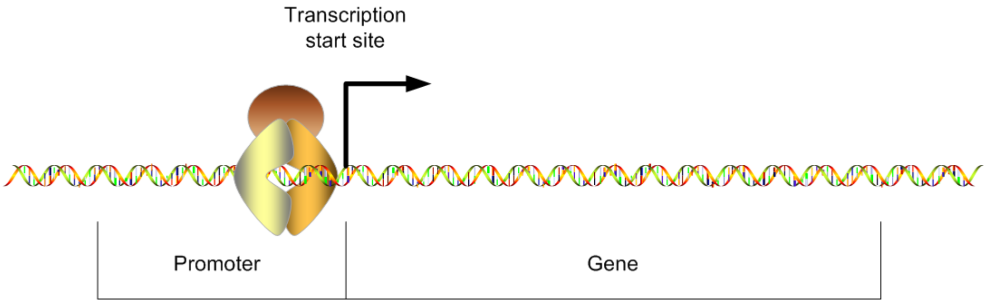
Signal = *Proportional to the number of target fragments*

Preprocessing ChIP-seq data

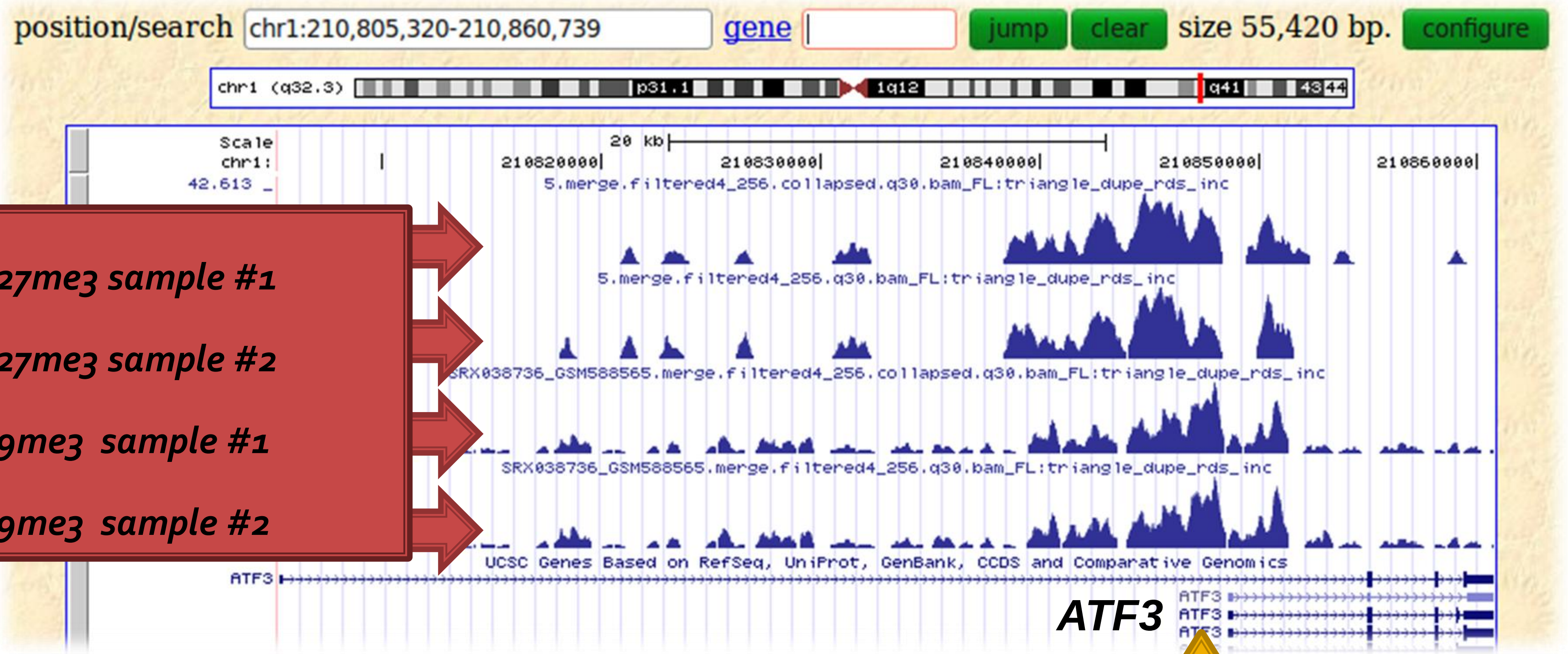
- Search for enriched regions in raw ChIP-seq data
  - IP compared to total DNA



- Annotate peaks to genes
  - Gene = whole genomic region +/- 2000 bp
  - Annotation retrieved from Ensembl (Biomart)



Result:



position/search chr1:210,805,320-210,860,739 gene jump clear size 55,420 bp. configure

5.merge.filtered4_256.collapsed.q30.bam_FL:triangle_dupe_rds_inc

5.merge.filtered4_256.q30.bam_FL:triangle_dupe_rds_inc

SRX038736_GSM588565.merge.filtered4_256.collapsed.q30.bam_FL:triangle_dupe_rds_inc

SRX038736_GSM588565.merge.filtered4_256.q30.bam_FL:triangle_dupe_rds_inc

UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

H3K27me3 sample #1

H3K27me3 sample #2

H3K9me3 sample #1

H3K9me3 sample #2

ATF3

- •located near TSS
- •all repressive marks
- •expectation: gene switched off

Biological interpretation

Essential steps

1. Integration with gene expression data
   - In most cases, you expect a strong correlation between gene expression and the investigated DNA binding protein, histone modification, DNA methylation levels, etc.

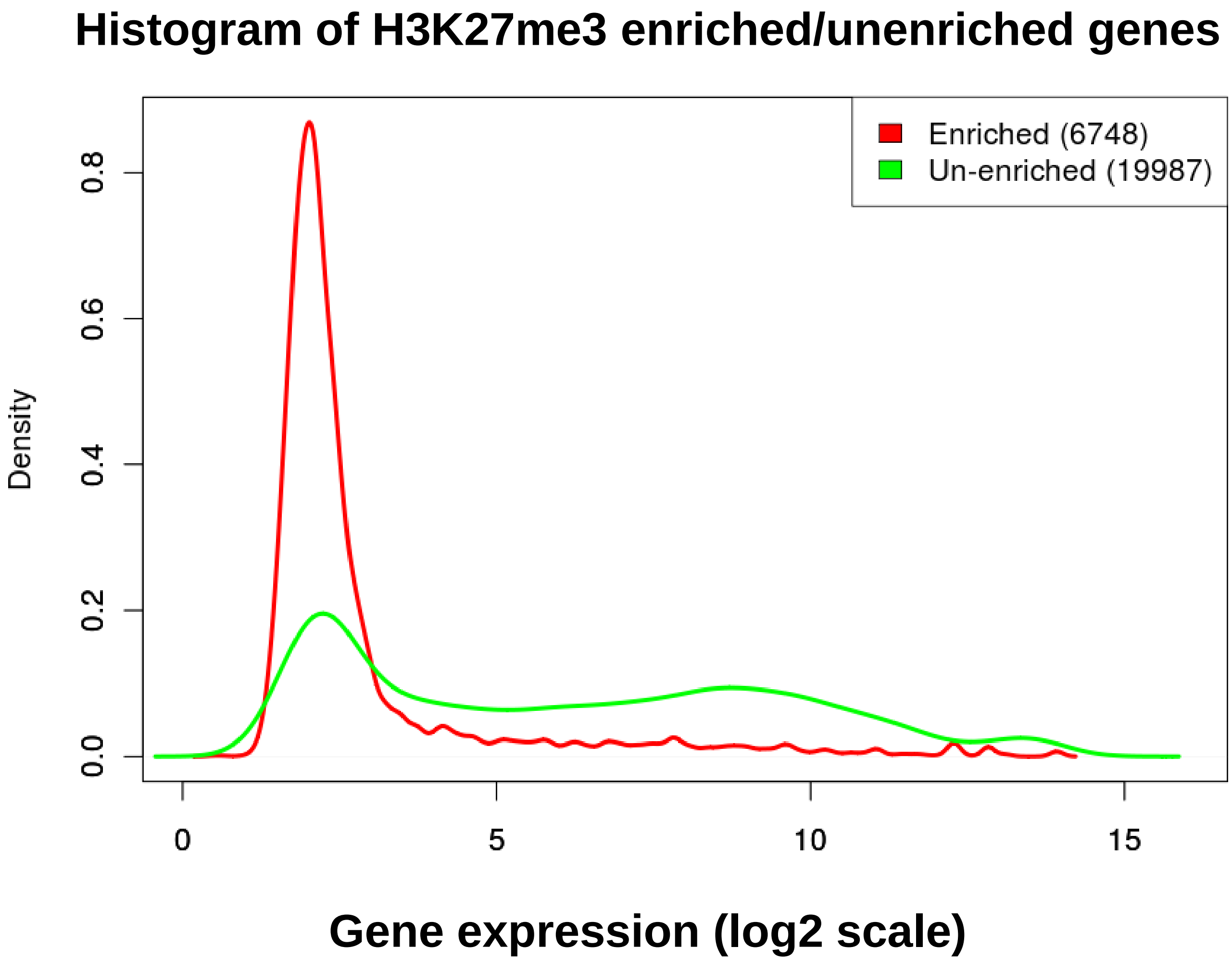
2. Sequence analysis of identified regulatory regions

Essential steps

1. Integration with gene expression data
   - In most cases, you expect a strong correlation between gene expression and the investigated DNA binding protein, histone modification, DNA methylation levels, etc.

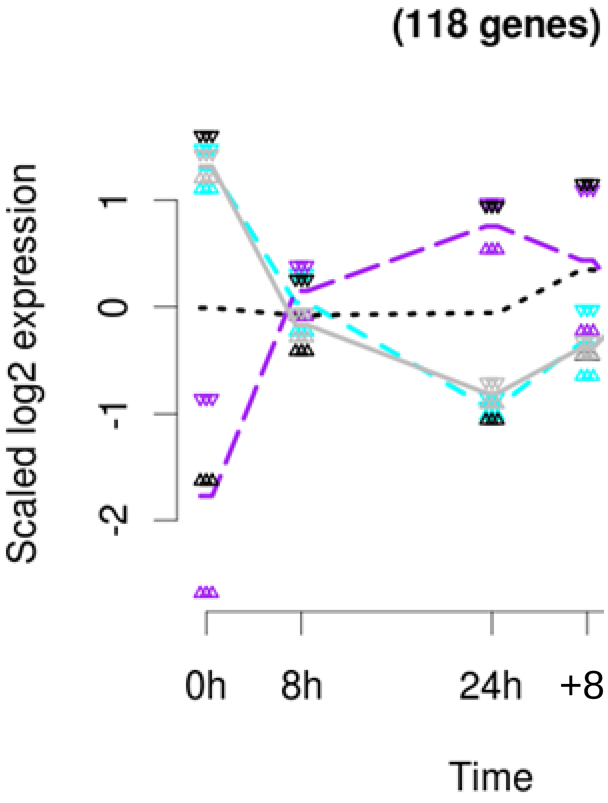
2. Sequence analysis of identified regulatory regions

Gene expression integration

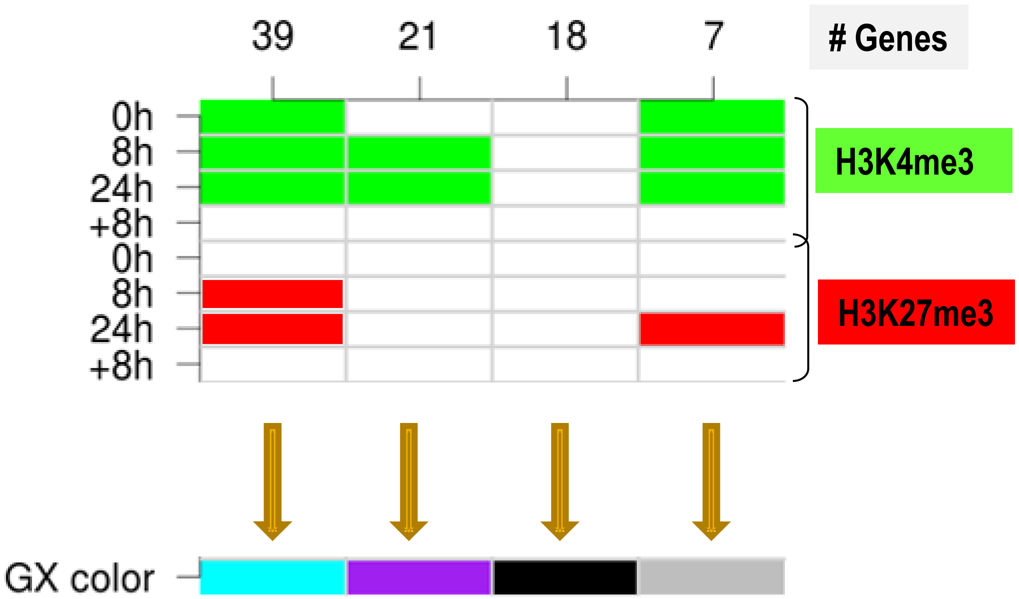


Gene expression integration (2)

Gene expression clusters



Histone mark occupancy in promoter



Essential steps

1. Integration with gene expression data
    - In most cases, you expect a strong correlation between gene expression and the investigated DNA binding protein, histone modification, DNA methylation levels, etc.

2. Sequence analysis of identified regulatory regions

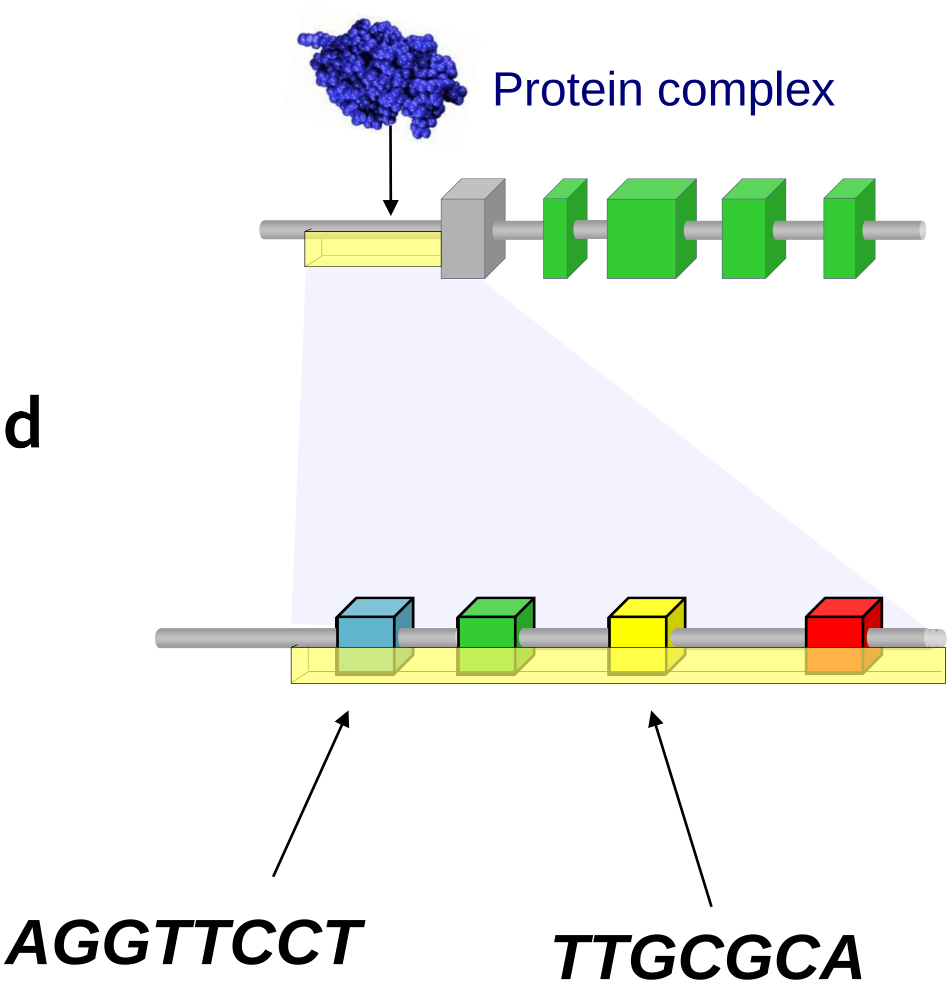
Motives for motif analysis

- **Validation** of **known** motifs
  - ChIP on protein X → scan for motif of protein X in enriched regions
  - DNA methylation array → scan for CpG islands in regions showing differential methylation
- Identifying **other** motifs
  - **Known**:
    - Scan for other transcription factor binding sites (which might be **functionally associated** with the ChIP'd protein)
  - **Novel**:
    - Identify novel motifs associated with the enriched regions

Transcription factor

- A transcription factor does **not** bind **randomly**

- They bind to **conserved** motifs of nucleotides called a **transcription factor binding site (TFBS)**



Transcription factor (2)

- Experimentally determined TFBSs are often referred to as **consensus sites**, which have a more statistical flavour (*caused by noise, variation, redundancy*):
  - By aligning multiple sequences (for instance ChIP-seq reads) a position weight matrix is constructed
  - The columns are the positions in the consensus site
  - The rows represent the relative frequency of each nucleotide for each position:

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	0	0.0000	0.8238	0.3333	0.3333	0	0.6667	0	0.3333	0	0	0.0000	1	0	0.0000	0.0875	0
C	0	0.3667	0.0000	0.3333	0.0000	1	0.0000	0	0.6667	0	0	0.6667	0	0	0.0762	0.5500	1
G	1	0.4500	0.1762	0.3333	0.6667	0	0.3333	0	0.0000	0	1	0.0000	0	1	0.0000	0.2749	0
T	0	0.1833	0.0000	0.0000	0.0000	0	0.0000	1	0.0000	1	0	0.3333	0	0	0.9238	0.0875	0

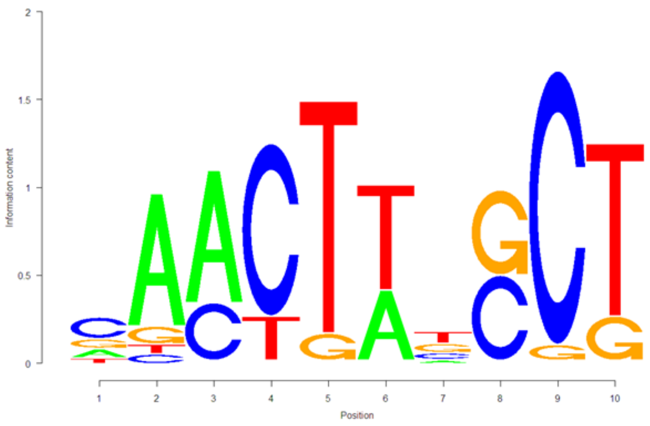
Transcription factor (3)

- Matrices are difficult to interpret. Hence, usually a **sequence logo** is created:
  - The **relative frequencies** are converted into **information entropies**. The information content at position *w* of a motif is given by:

$$ic(w)=\log_2(J)+\sum_{j=1}^{J} p_{jw}\log_2(p_{jw})$$

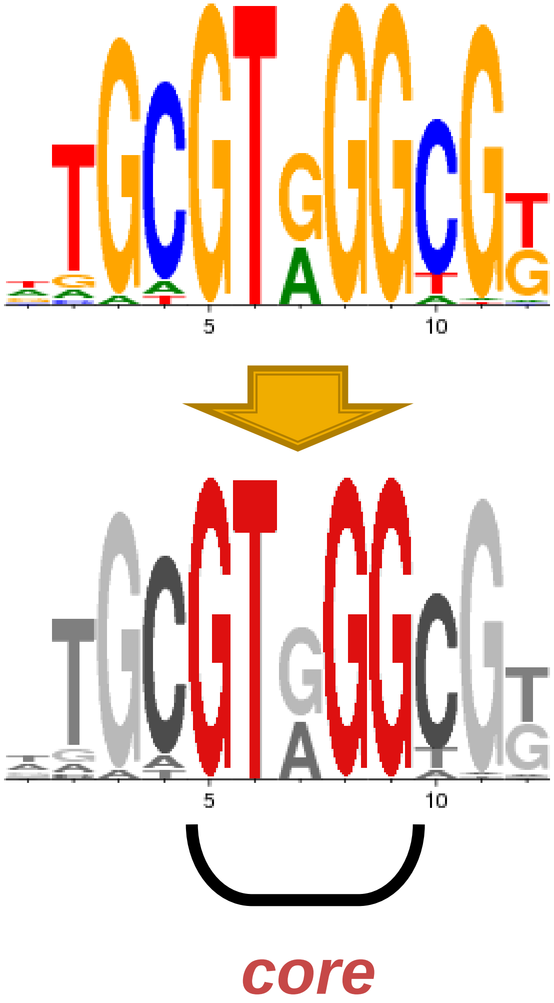
where *J* is the number of letters in the 'alphabet' (4 for DNA sequences)

  - In a sequence motif, the **height of a nucleotide letter** on a specific position corresponds to the relative **conservation** of that nucleotide on that position:



Scanning sequences for motifs

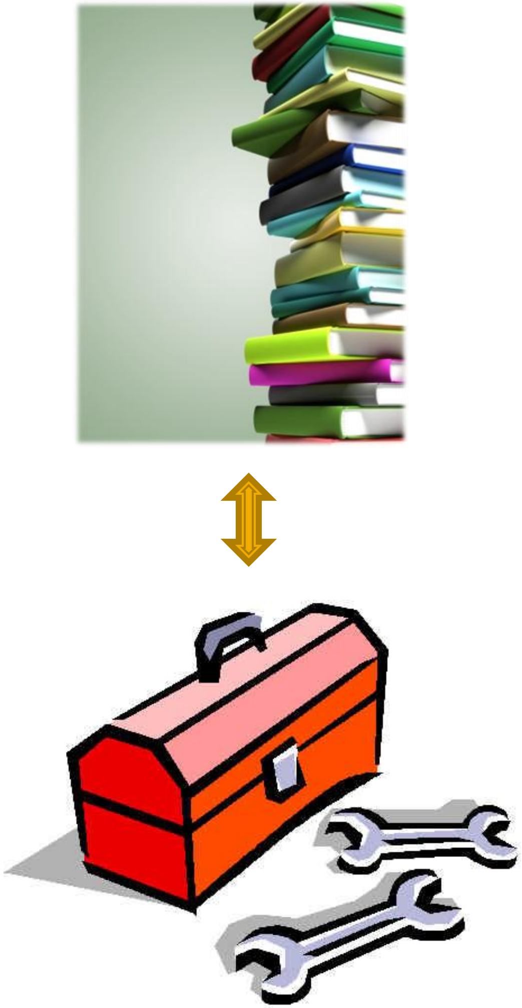
- Motifs are searched using algorithms
- In general to be called a 'hit':
  - 100% match with core
  - >70% match for whole motif



tools require databases require tools

- Databases:
  - TRANSFAC
  - JASPER_CORE

- Analysis tools:
  - CORE_TF
  - JASPER tools

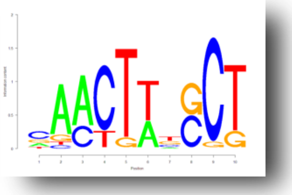


TRANSFAC

- TRANSFAC contains data on circa 10,000 transcription factors in species ranging from vertebrates viruses.

- It is the most comprehensive cross-species compilation of data regarding TFs:
  - Structural features of a factor
  - Expression pattern
  - Regulatory network
  - Functional properties (what does it do)
  - Interacting factors

bits
HPYHGESYANIDQHY
GHS
VK
N
C

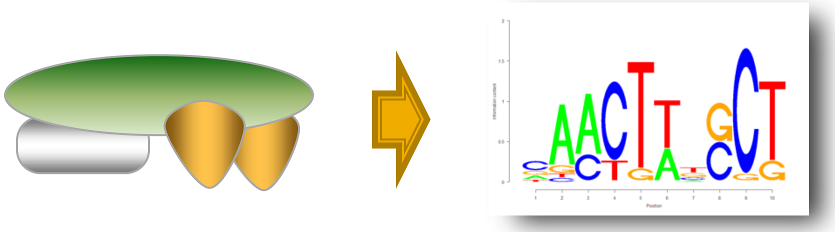


- Simple interface
  - Great database, not so great tools
  - Hard to curate the results you get

JASPAR

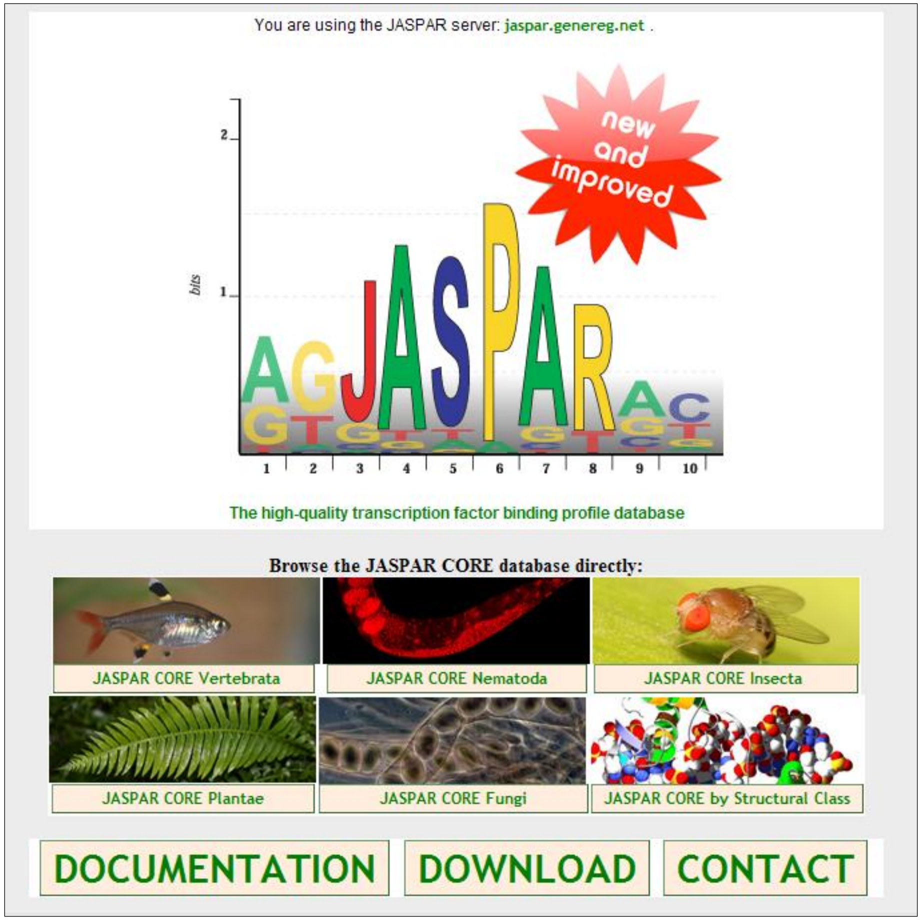
- http://jaspar.cgb.ki.se/
- The JASPAR database (JASPER_CORE) contains a curated, non-redundant set of profiles from published articles.
- One of the central goals with JASPAR_CORE is to give the single, "best" model for each transcription factor.

one factor, one model:



JASPAR (2)

- The prime difference to similar resources (TRANSFAC, etc) consist of the **open data access**, **non-redundancy** and **quality**



You are using the JASPAR server: jaspar.genereg.net .

The high-quality transcription factor binding profile database

Browse the JASPAR CORE database directly:

Pros and cons of JASPER

- **Pros**:
  - Open-access
  - Curated database
  - Motifs are fully annotated, including sequence logos
  - Various useful tools for transcription factor scanning

- **Cons**:
  - Curated database, but also relatively small
  - **Can only scan one sequence at a time!**

CORE_TF

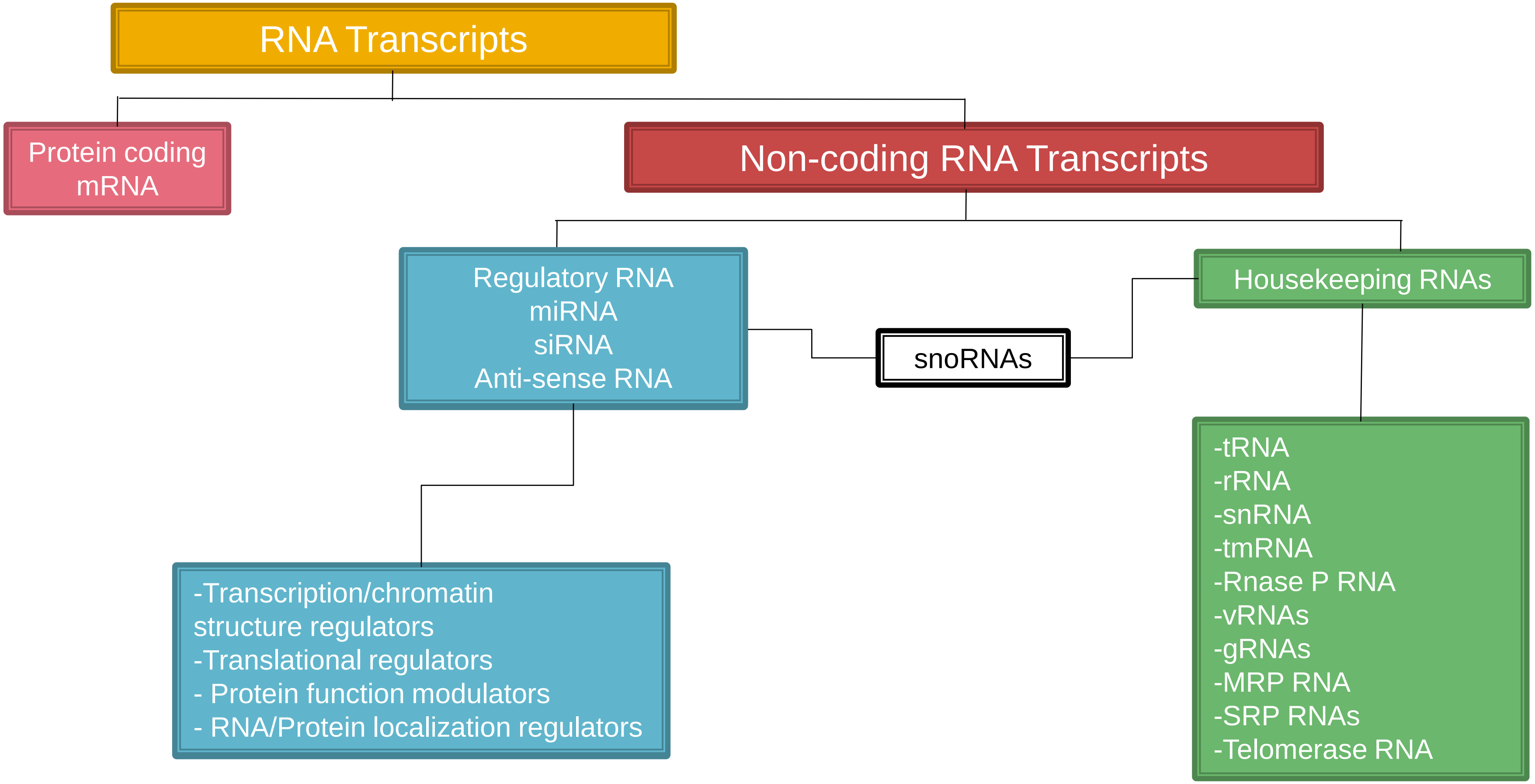
- http://grenada.lumc.nl/HumaneGenetica/CORE_TF/
- Uses the public **TRANSFAC** database
- Focused on overrepresentation analysis: what TFs are **overrepresented in** your **query compared** to a **random set**

Pros and cons of CORE_TF

- Pros:
  - Open-access
  - Can do TF overrepresentation analysis
  - Takes both sequences and IDs as input
- Cons:
  - No sequence logos of TF motifs
  - Additional information on the used motifs hidden from user -> ***find elsewhere***

miRNAs

Non-Coding RNA: Formerly known as *"JUNK"*



RNA Transcripts

Protein coding mRNA

Non-coding RNA Transcripts

Regulatory RNA
miRNA
siRNA
Anti-sense RNA

snoRNAs

Housekeeping RNAs

-Transcription/chromatin structure regulators
-Translational regulators
- Protein function modulators
- RNA/Protein localization regulators

-tRNA
-rRNA
-snRNA
-tmRNA
-Rnase P RNA
-vRNAs
-gRNAs
-MRP RNA
-SRP RNAs
-Telomerase RNA

microRNAs (miRNAs)

- Small non-coding RNAs, approximately 22 nt long.
- Regulate gene expression in a sequence-specific manner.
- The human genome may encode over 1000 miRNAs.
- May target about 60% of mammalian genes
- Abundant in many human cell types
- Well-conserved

myomiRs : muscle specific miRNAs

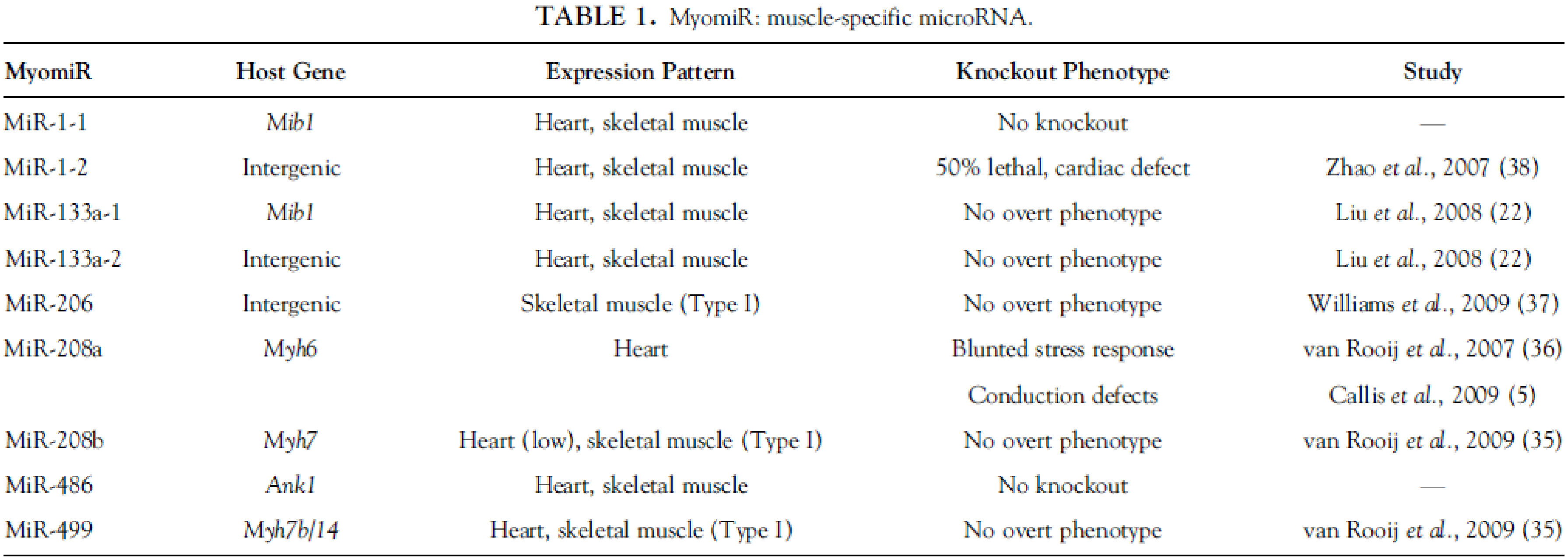


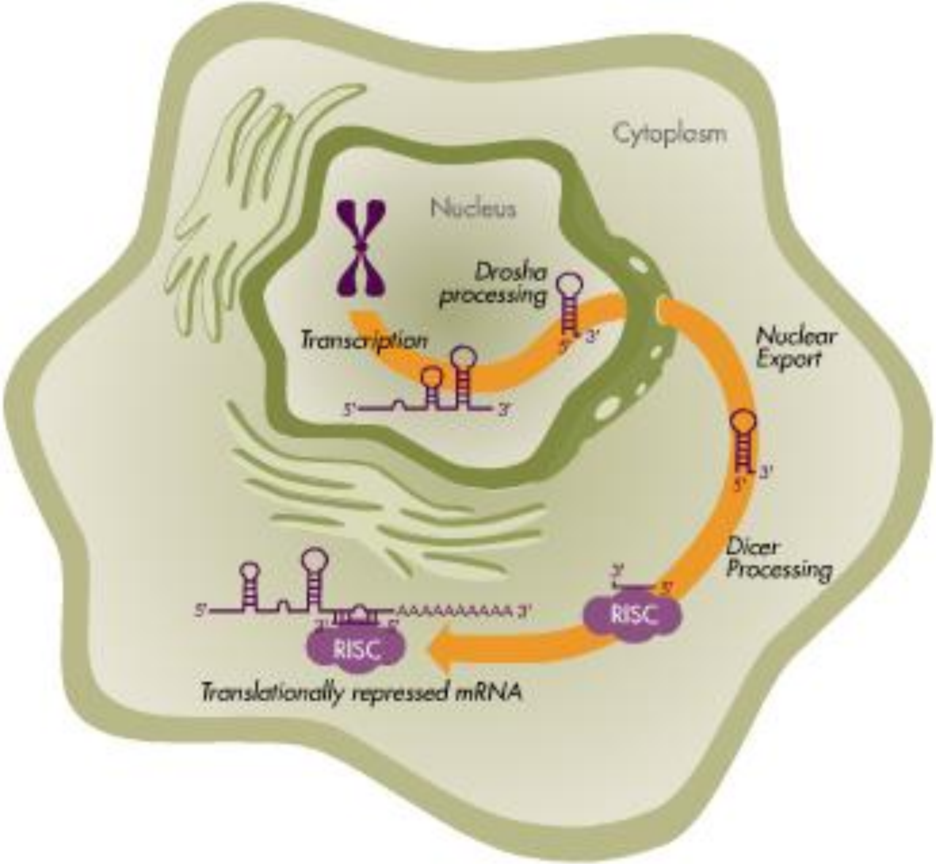
TABLE 1. MyomiR: muscle-specific microRNA.

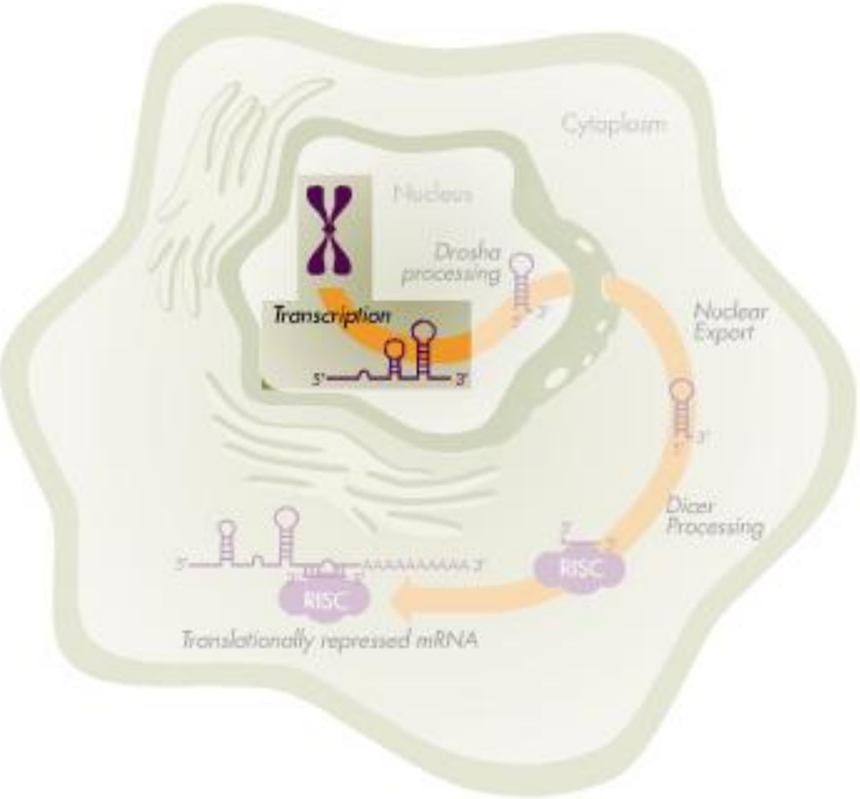
MyomiR	Host Gene	Expression Pattern	Knockout Phenotype	Study
MiR-1-1	*Mib1*	Heart, skeletal muscle	No knockout	—
MiR-1-2	Intergenic	Heart, skeletal muscle	50% lethal, cardiac defect	Zhao *et al.*, 2007 (38)
MiR-133a-1	*Mib1*	Heart, skeletal muscle	No overt phenotype	Liu *et al.*, 2008 (22)
MiR-133a-2	Intergenic	Heart, skeletal muscle	No overt phenotype	Liu *et al.*, 2008 (22)
MiR-206	Intergenic	Skeletal muscle (Type I)	No overt phenotype	Williams *et al.*, 2009 (37)
MiR-208a	*Myh6*	Heart	Blunted stress response	van Rooij *et al.*, 2007 (36)
			Conduction defects	Callis *et al.*, 2009 (5)
MiR-208b	*Myh7*	Heart (low), skeletal muscle (Type I)	No overt phenotype	van Rooij *et al.*, 2009 (35)
MiR-486	*Ank1*	Heart, skeletal muscle	No knockout	—
MiR-499	*Myh7b/14*	Heart, skeletal muscle (Type I)	No overt phenotype	van Rooij *et al.*, 2009 (35)

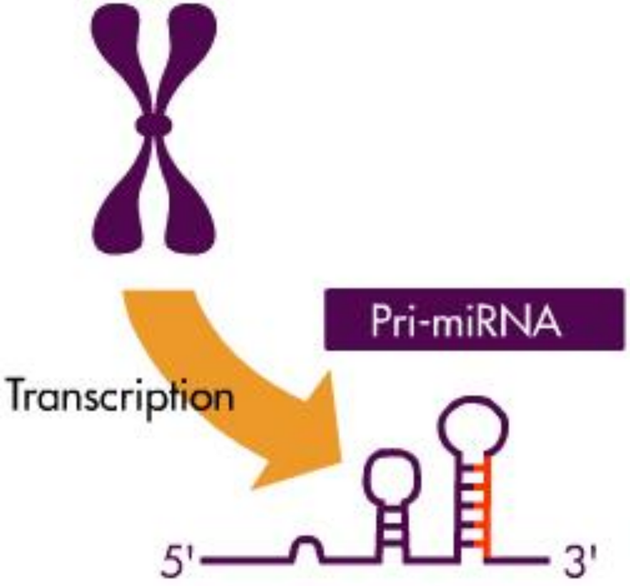
McCarthy (2011) The myomiR network in skeletal muscle placticity. *Exerc Sport Sci Rev.*

miRNA processing

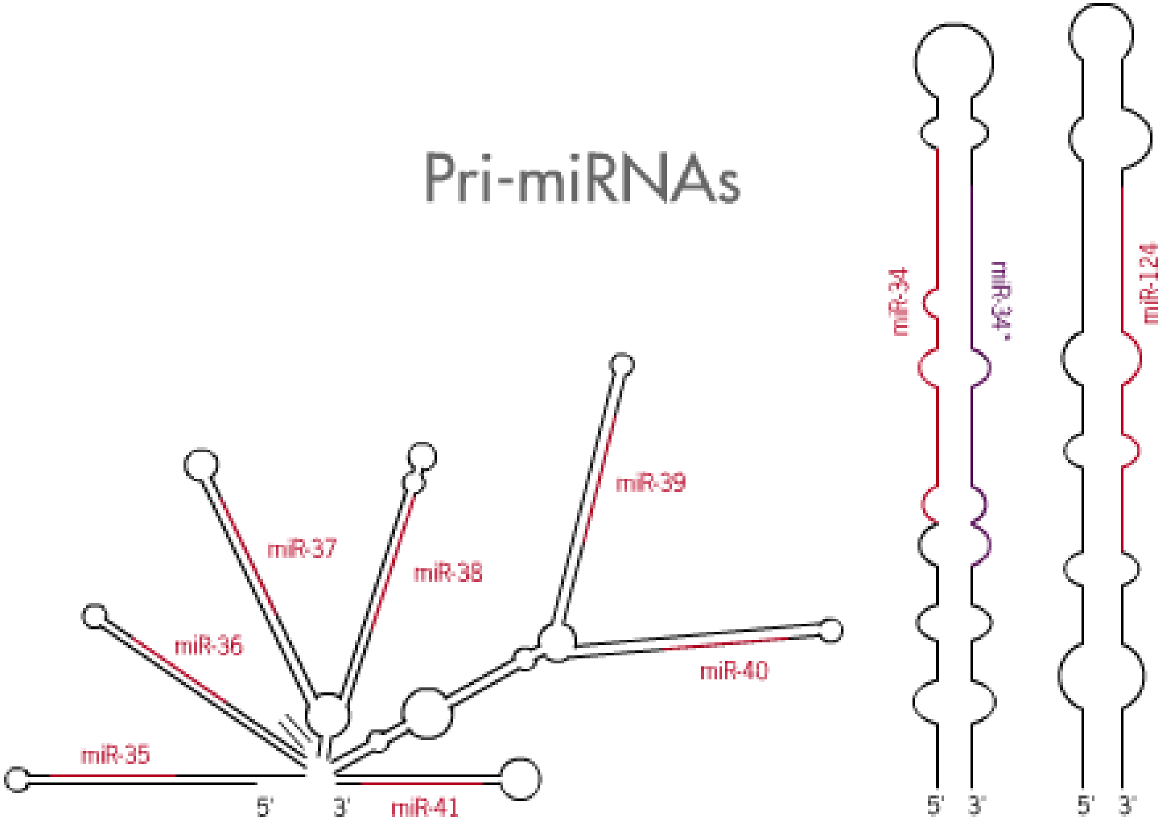
Single-stranded RNA which is 17-25 nucleotides long, regulating the expression of other genes.




Cytoplasm
Nucleus
Drosha processing
Transcription
5'
3'
Nuclear Export
5'
3'
Dicer Processing
RISC
5'
AAAAAAAAA 3'
RISC
Translationally repressed mRNA

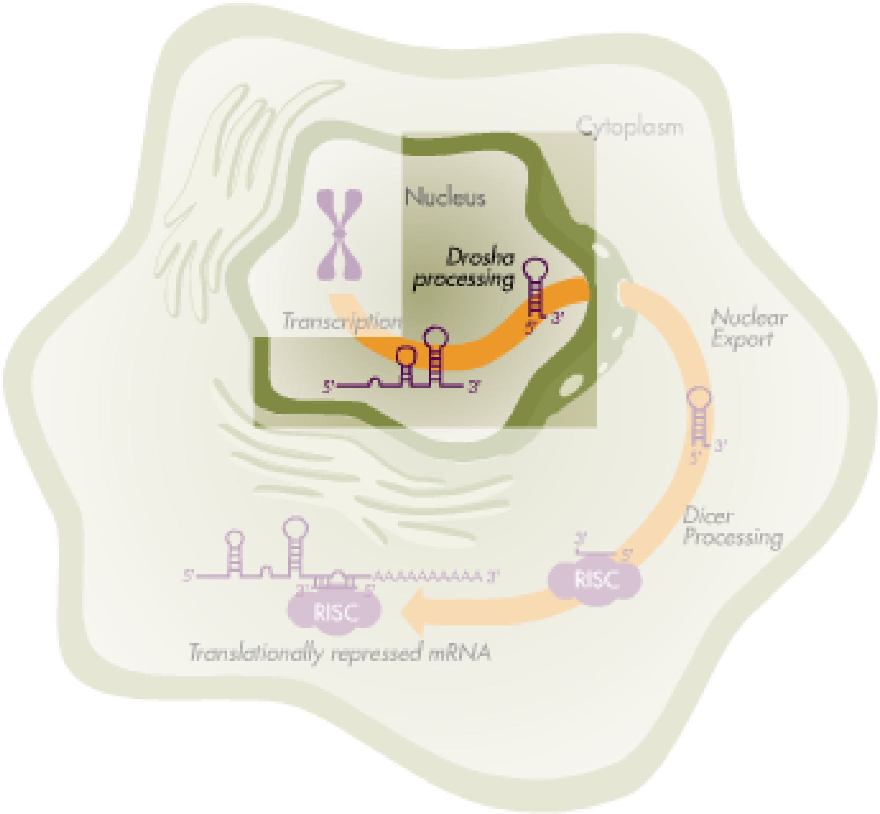

Pri-miRNA
Transcription
5'
3'

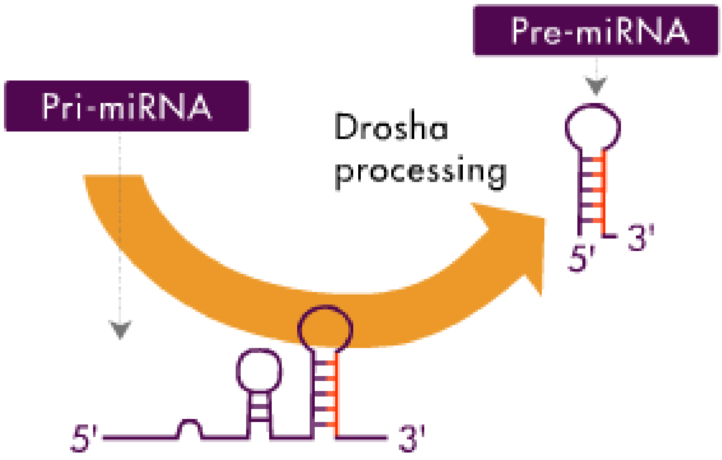
Approximately 60% of miRNAs are expressed independently, 15% are expressed in clusters, and 25% in introns.



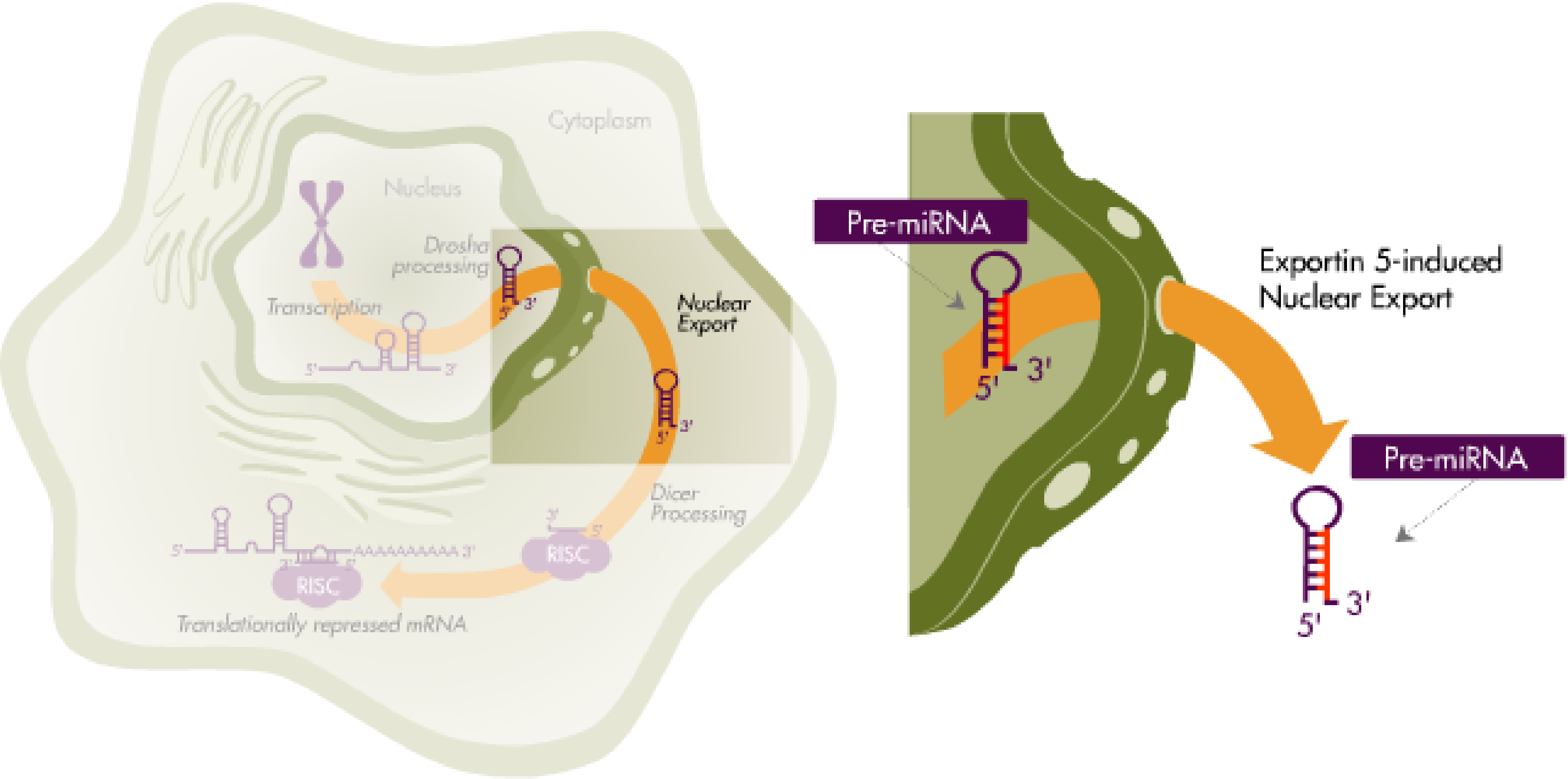
Pri-miRNAs

Drosha (a dsRNA-specific ribonuclease): Pri-miRNA → Pre-miRNA (70-100 nt)



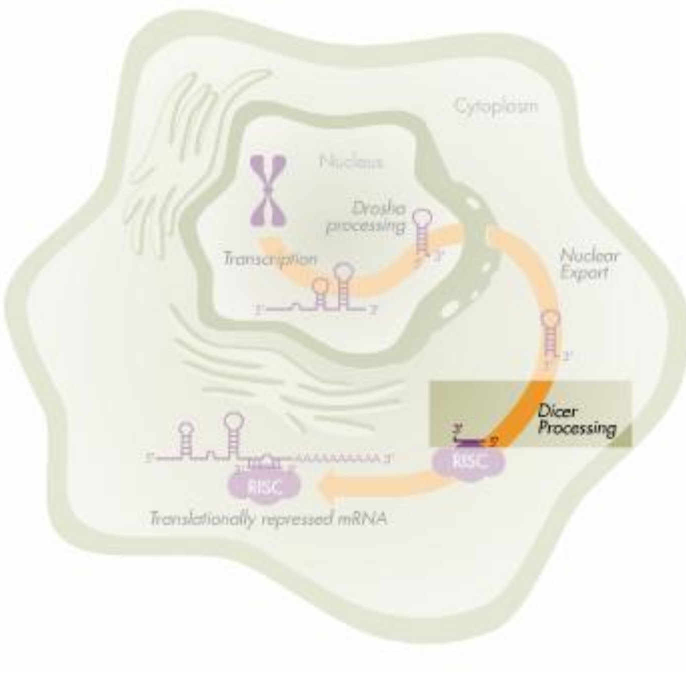


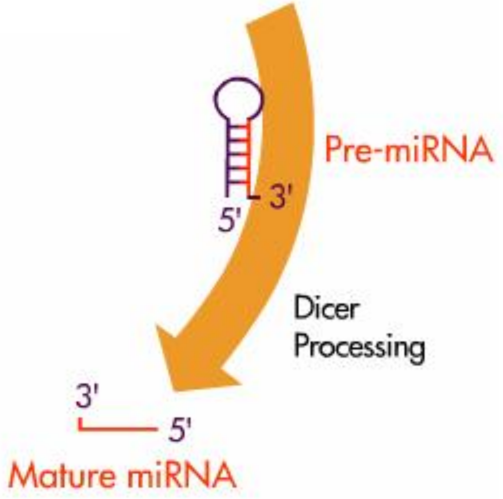
Exportin 5 - induced nuclear export:



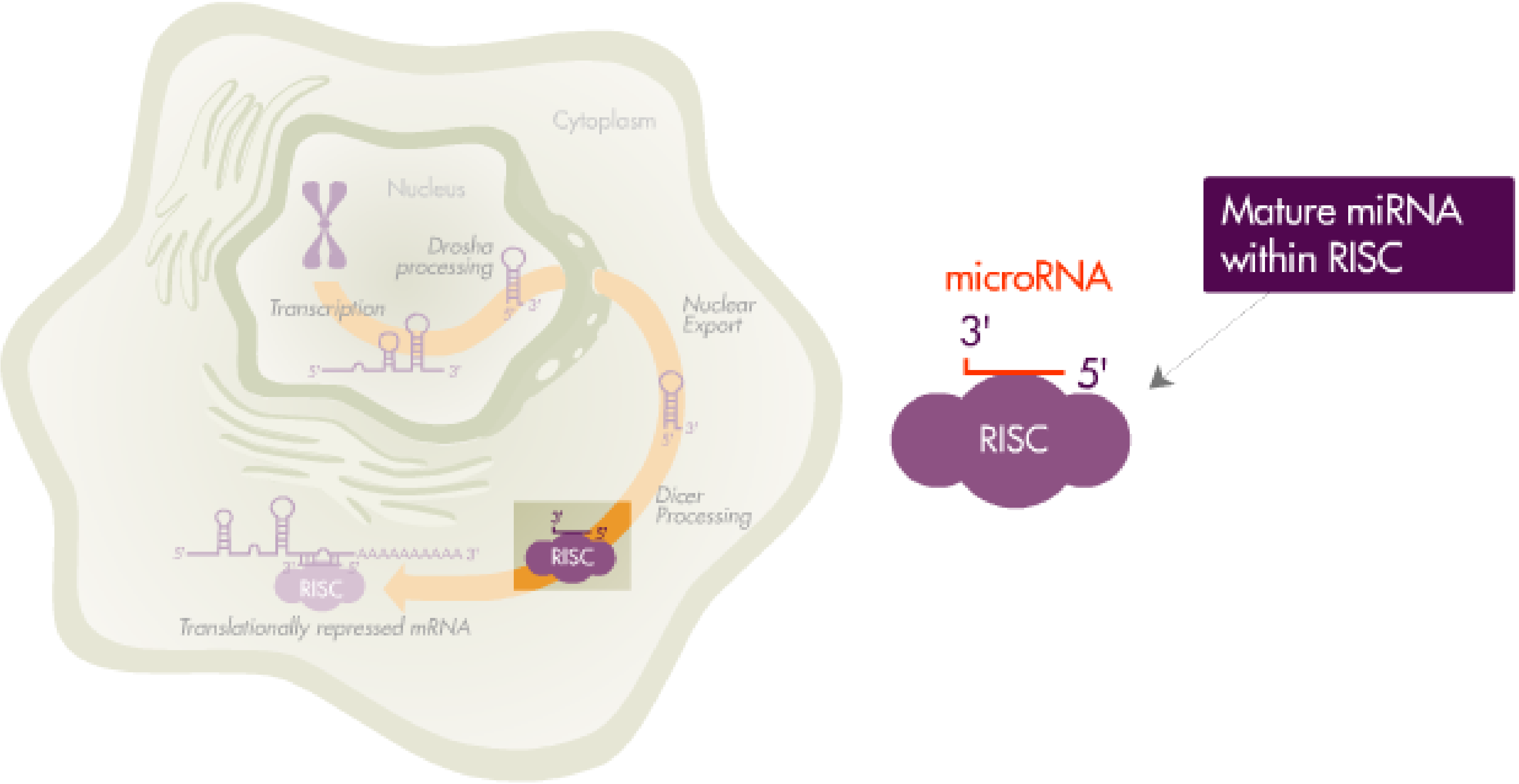
Dicer (a dsRNA-specific ribonuclease): Pre-miRNA → mature miRNA (17-25 nt)

```
    -uu c    a   cg    uc   a      c  ac
ccug   g caca acc  uaga  cga cuugug ug  c
||||   | |||| |||  ||||  ||| |||||| ||
ggau   c gugu ugg  aucu  guu gaacac ac  a
    ugu u    a   au    gu   c      -  gu
```

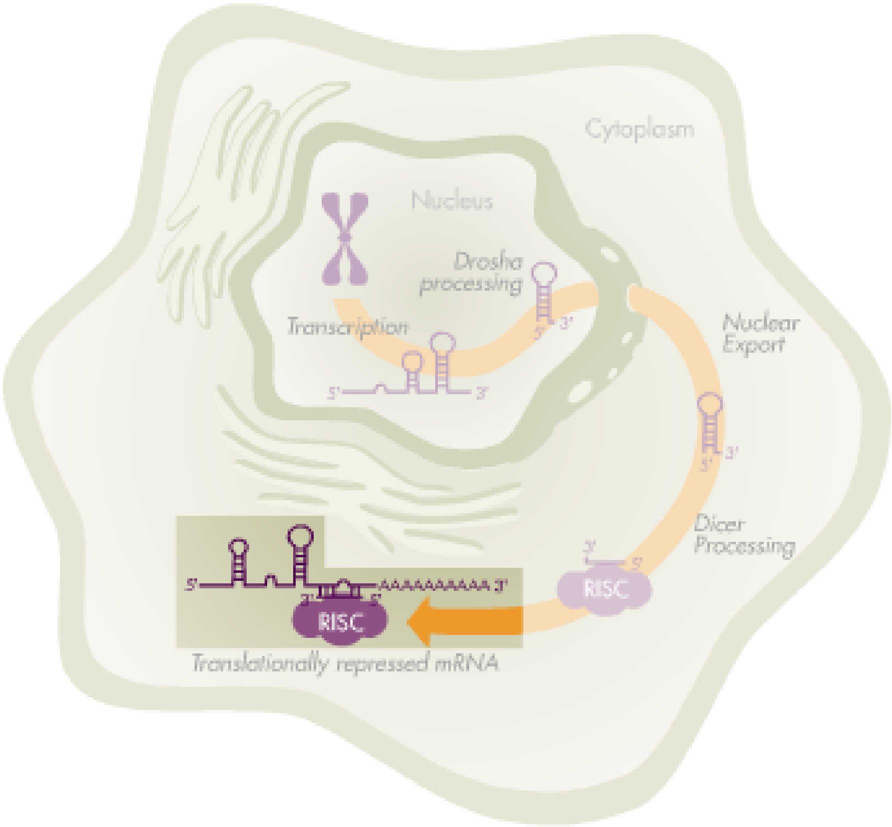


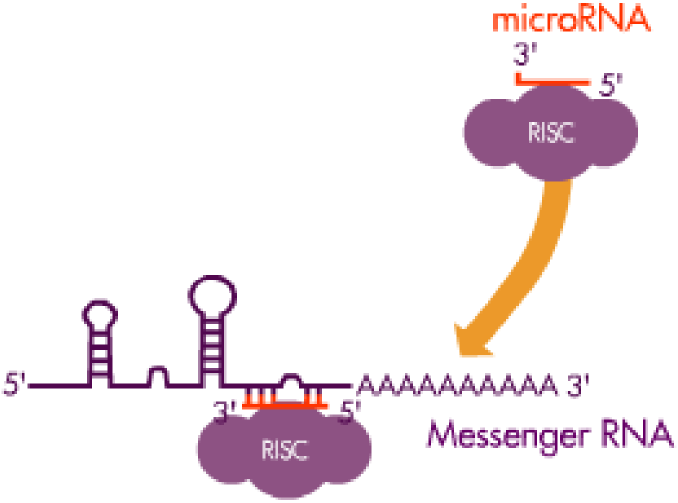


The miRNA is bound by a complex similar to RNA-Induced Silencing Complex (RISC) that participates in RNA interference (RNAi)



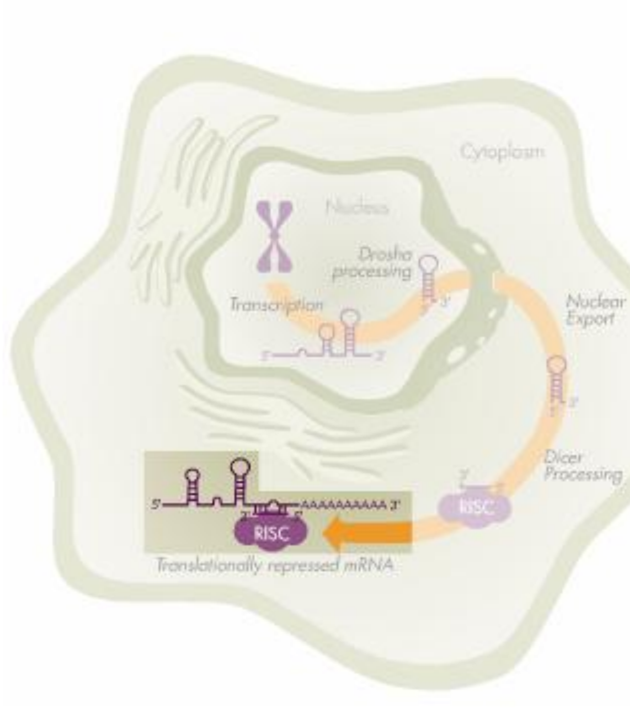
miRNA-RISC complex binds target mRNA:

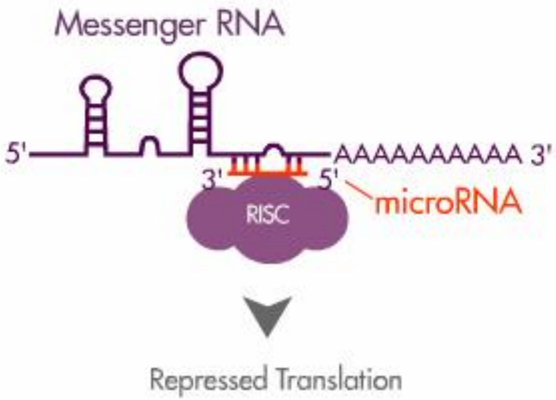




The annealing of the miRNA to the mRNA may

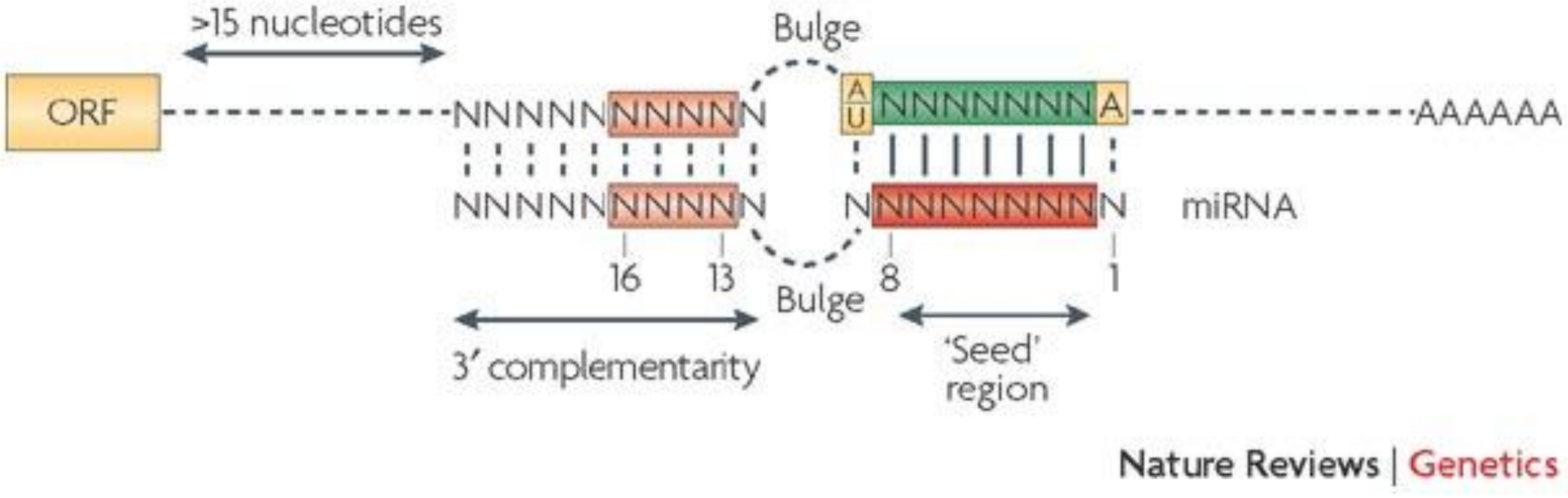
1. inhibit protein translation
2. facilitate cleavage of the mRNA.





```
             a   au    gu   c
5' ···ucugugu ugg  aucu  guu gaacacaacguaacug··· 3'
              |||  ||||  ||| ||||||
3'            acc  uaga  cga cuugug              5'
             a   cg    uc   a
```

miRNA-mRNA interactions

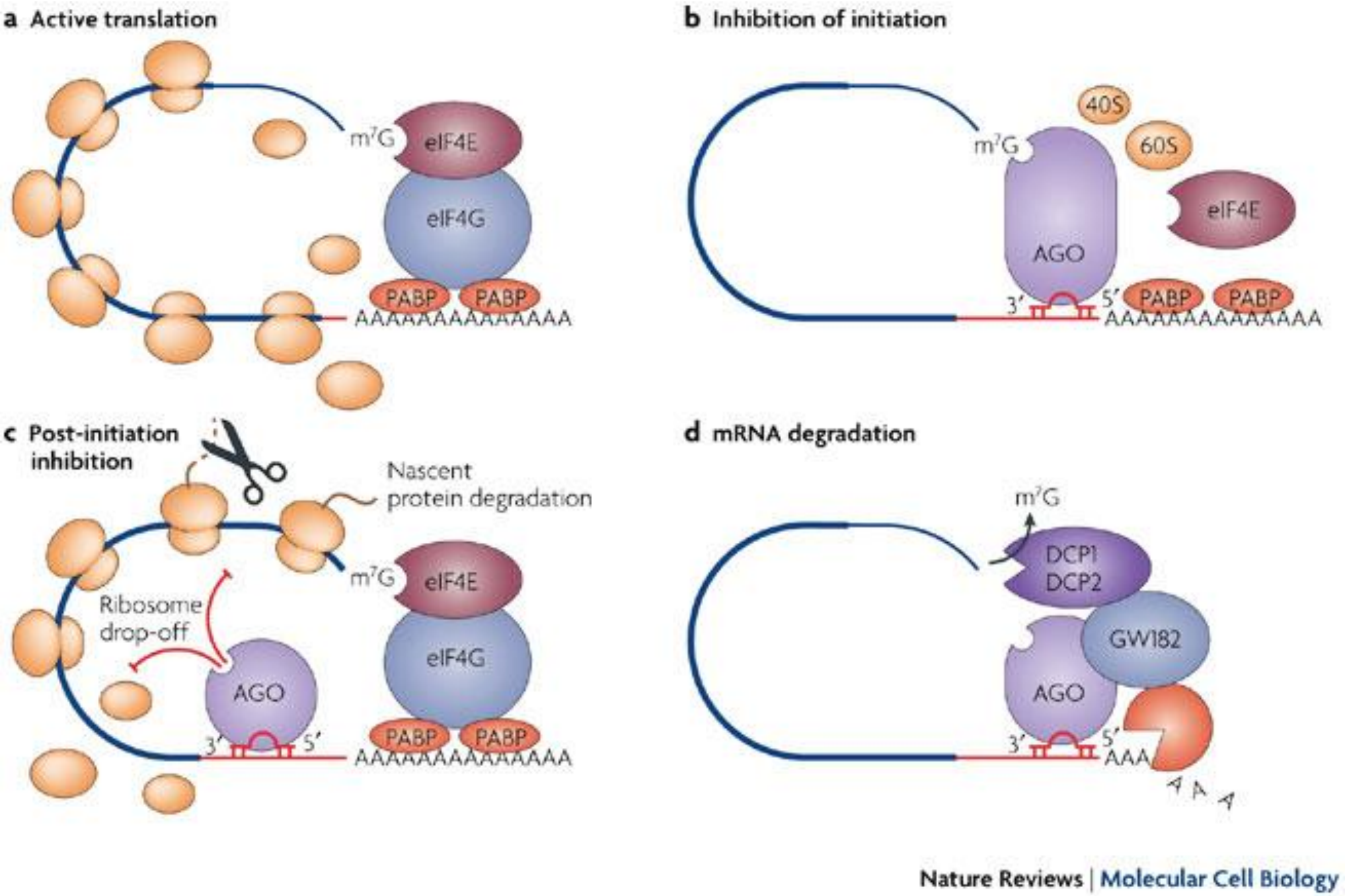


Witold Filipowicz, Suvendra N. Bhattacharyya* and Nahum Sonenberg‡*

miRNA function

- Involved in the post-transcriptional regulation of gene expression
- Important in development
- Metabolic regulation (miR-375 & insulin secretion)
- Multiple genomic loci (different expression **patterns**)

Differences in miRNA Mode of Action



Nature Reviews | Molecular Cell Biology

microRNA nomenclature (1)

General form for mature microRNA: hsa-miR-195

- Uncapitalized "mir-" refers to the pre-miRNA,
- A capitalized "miR-" refers to the mature form

- The prefix "mir“ or “miR” is followed by a dash and a number, the latter often indicating order of naming

- Species of origin is designated with a three-letter prefix, For example:
  hsa-miR-195 is a human (Homo sapiens) miRNA and
  mmu-miR-123 is a mouse (Mus musculus) miRNA.

microRNA nomenclature (2)

- **Distinct precursor** sequences and genomic loci but **identical mature** sequences:

hsa-miR-16-1 =	uagcagcacguaaauauuggcg
hsa-miR-16-2 =	uagcagcacguaaauauuggcg

- Lettered suffixes denote **closely related mature** sequences:

hsa-miR-15-a =	uagcagcacauaaugguuugug
hsa-miR-15-b =	uagcagcacau**c**augguuu**aca**

PREVIOUS

- Two sequences which originate from the **same predicted precursor**: **use relative abundancies:**
  miR-56 = the predominant product
  miR-56* = from the opposite arm of the precursor
- **predominant form unknown**:
  miR-142-5p = from the 5' arm
  miR-142-3p = from the 3' arm

NOW: only the 3p/5p annotation is used!

- **let-7 and lin-4** are exceptions to the numbering scheme, these names are retained for historical reasons.

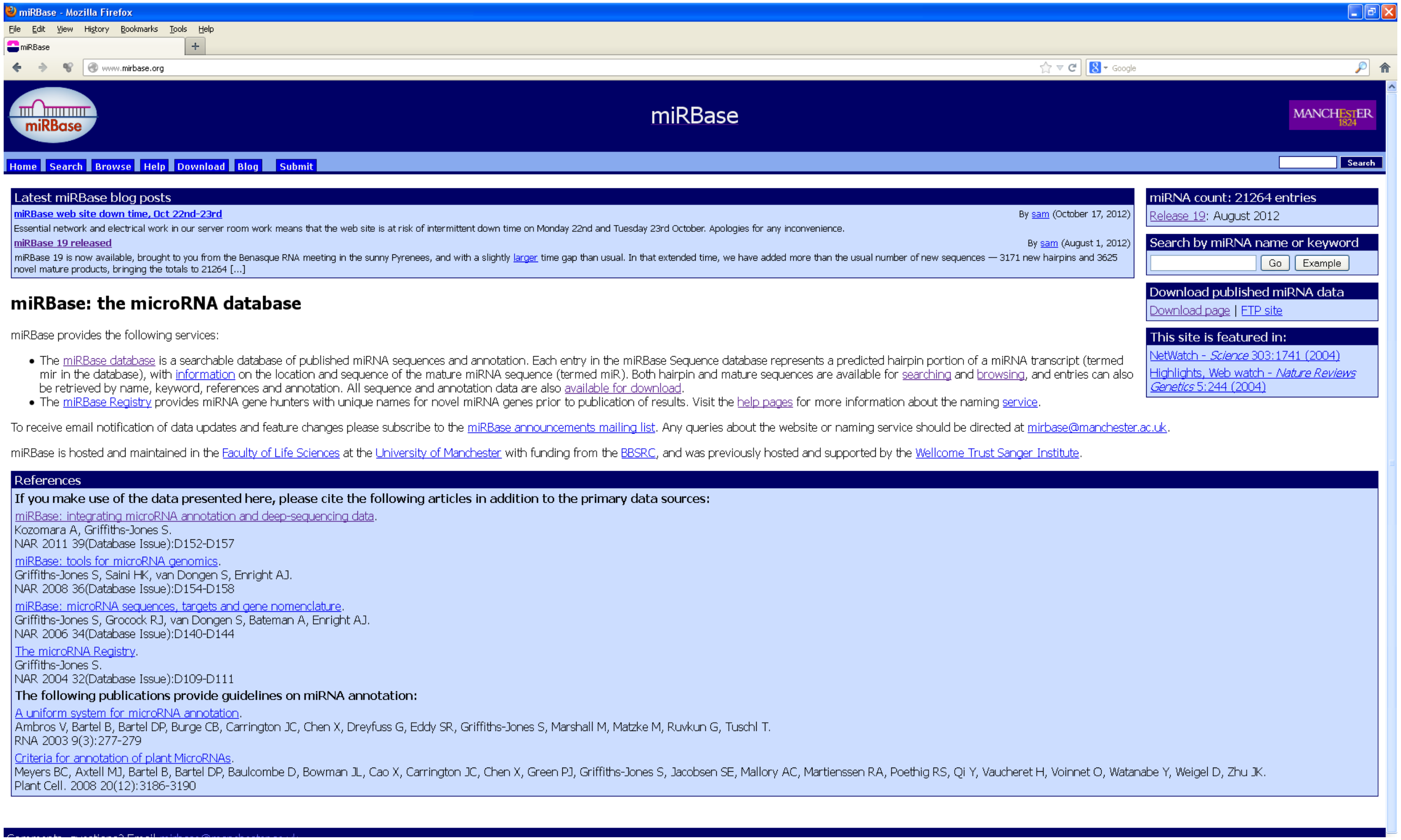
miRBase (www.mirbase.org)

The primary online repository for all microRNA sequences and annotation

miRBase version 19 contains:

- 21,264 hairpins
- 25,141 mature microRNAs
- 193 species

miRBase homepage


miRBase - Mozilla Firefox
File Edit View History Bookmarks Tools Help
miRBase
www.mirbase.org
Google
miRBase
MANCHESTER 1824
Home Search Browse Help Download Blog Submit
Search

miRBase

Latest miRBase blog posts

miRBase web site down time, Oct 22nd-23rd By sam (October 17, 2012)

Essential network and electrical work in our server room work means that the web site is at risk of intermittent down time on Monday 22nd and Tuesday 23rd October. Apologies for any inconvenience.

miRBase 19 released By sam (August 1, 2012)

miRBase 19 is now available, brought to you from the Benasque RNA meeting in the sunny Pyrenees, and with a slightly larger time gap than usual. In that extended time, we have added more than the usual number of new sequences — 3171 new hairpins and 3625 novel mature products, bringing the totals to 21264 [...]

miRBase: the microRNA database

miRBase provides the following services:

- The miRBase database is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also available for download.
- The miRBase Registry provides miRNA gene hunters with unique names for novel miRNA genes prior to publication of results. Visit the help pages for more information about the naming service.

To receive email notification of data updates and feature changes please subscribe to the miRBase announcements mailing list. Any queries about the website or naming service should be directed at mirbase@manchester.ac.uk.

miRBase is hosted and maintained in the Faculty of Life Sciences at the University of Manchester with funding from the BBSRC, and was previously hosted and supported by the Wellcome Trust Sanger Institute.

miRNA count: 21264 entries

Release 19: August 2012

Search by miRNA name or keyword

Go Example

Download published miRNA data

Download page | FTP site

This site is featured in:

NetWatch - *Science* 303:1741 (2004)

Highlights, Web watch - *Nature Reviews Genetics* 5:244 (2004)

References

If you make use of the data presented here, please cite the following articles in addition to the primary data sources:

miRBase: integrating microRNA annotation and deep-sequencing data.
Kozomara A, Griffiths-Jones S.
NAR 2011 39(Database Issue):D152-D157

miRBase: tools for microRNA genomics.
Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ.
NAR 2008 36(Database Issue):D154-D158

miRBase: microRNA sequences, targets and gene nomenclature.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ.
NAR 2006 34(Database Issue):D140-D144

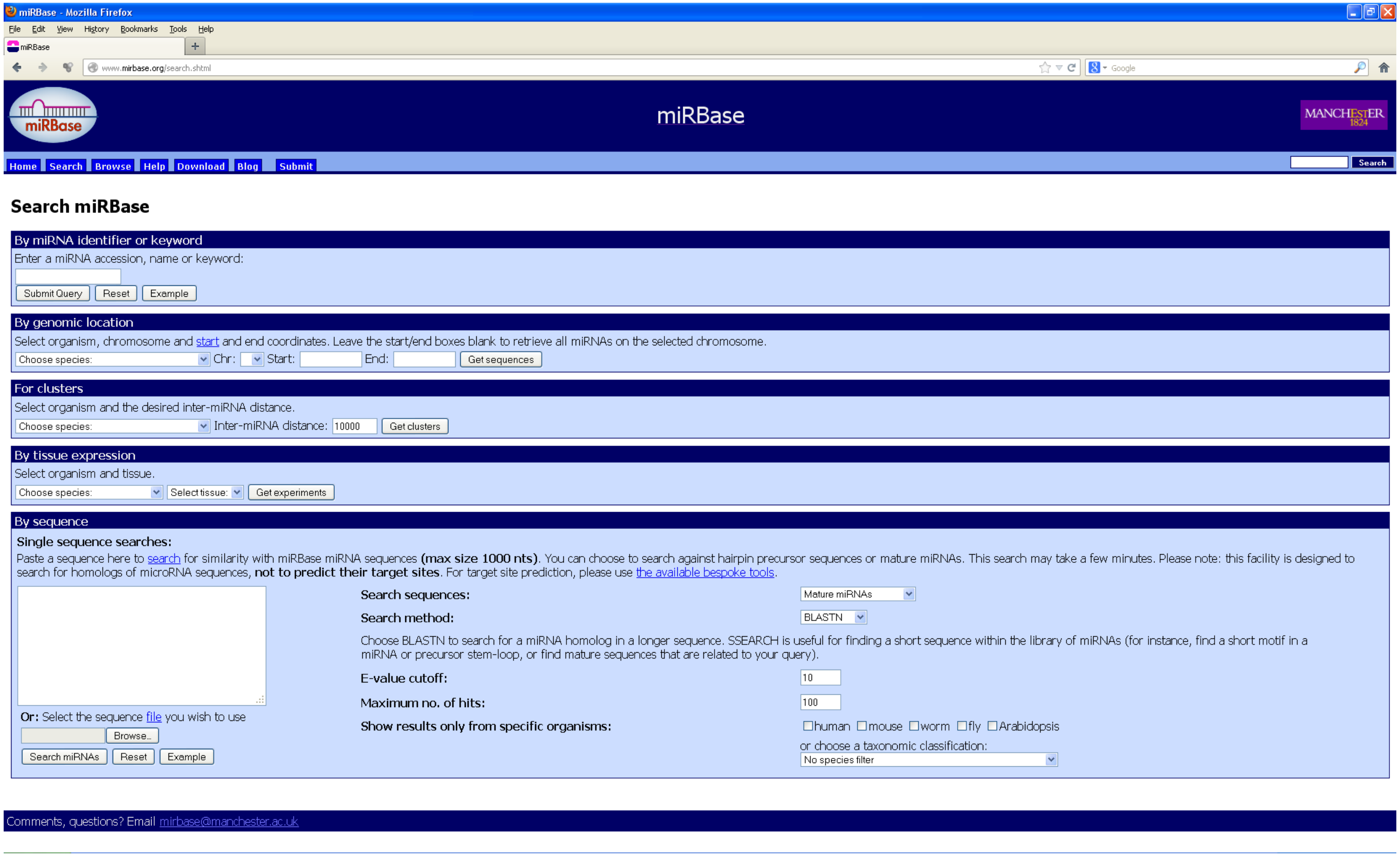
The microRNA Registry.
Griffiths-Jones S.
NAR 2004 32(Database Issue):D109-D111

The following publications provide guidelines on miRNA annotation:

A uniform system for microRNA annotation.
Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T.
RNA 2003 9(3):277-279

Criteria for annotation of plant MicroRNAs.
Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK.
Plant Cell. 2008 20(12):3186-3190

miRBase Search


miRBase - Mozilla Firefox
File Edit View History Bookmarks Tools Help
miRBase
www.mirbase.org/search.shtml
Google
miRBase
miRBase
MANCHESTER 1824
Home Search Browse Help Download Blog Submit
Search

Search miRBase

By miRNA identifier or keyword

Enter a miRNA accession, name or keyword:

[Submit Query] [Reset] [Example]

By genomic location

Select organism, chromosome and start and end coordinates. Leave the start/end boxes blank to retrieve all miRNAs on the selected chromosome.

Choose species: Chr: Start: End: [Get sequences]

For clusters

Select organism and the desired inter-miRNA distance.

Choose species: Inter-miRNA distance: 10000 [Get clusters]

By tissue expression

Select organism and tissue.

Choose species: Select tissue: [Get experiments]

By sequence

Single sequence searches:

Paste a sequence here to search for similarity with miRBase miRNA sequences **(max size 1000 nts)**. You can choose to search against hairpin precursor sequences or mature miRNAs. This search may take a few minutes. Please note: this facility is designed to search for homologs of microRNA sequences, **not to predict their target sites**. For target site prediction, please use the available bespoke tools.

Search sequences: Mature miRNAs

Search method: BLASTN

Choose BLASTN to search for a miRNA homolog in a longer sequence. SSEARCH is useful for finding a short sequence within the library of miRNAs (for instance, find a short motif in a miRNA or precursor stem-loop, or find mature sequences that are related to your query).

E-value cutoff: 10

Maximum no. of hits: 100

Show results only from specific organisms: □human □mouse □worm □fly □Arabidopsis

or choose a taxonomic classification: No species filter

Or: Select the sequence file you wish to use

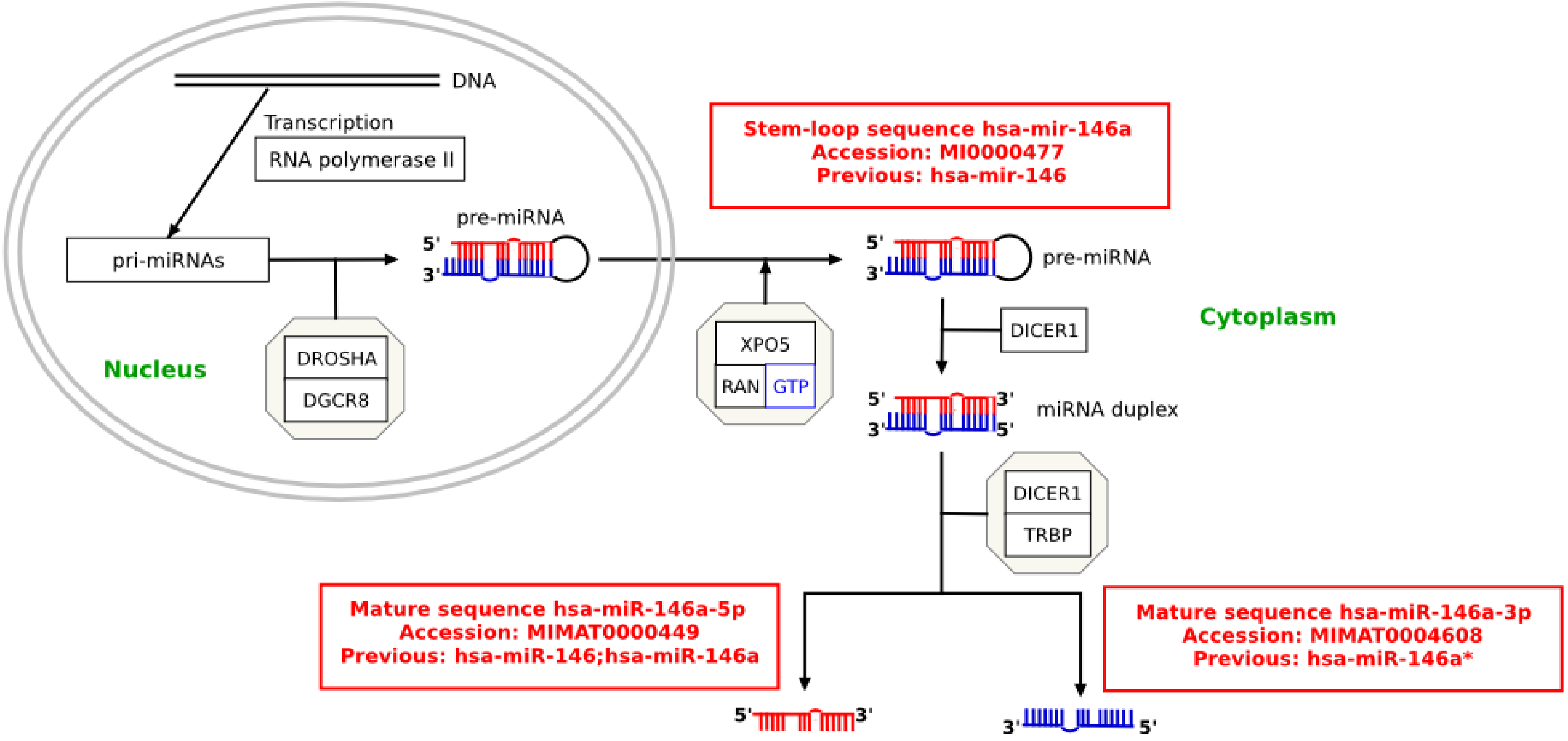
[Browse...]

[Search miRNAs] [Reset] [Example]

Comments, questions? Email mirbase@manchester.ac.uk

Identifiers in miRBase

- In addition to a name or ID, each miRBase Sequence entry has a unique **accession number**.
  stem-loop sequence: MI0000069
  mature sequence: MIMAT0000068

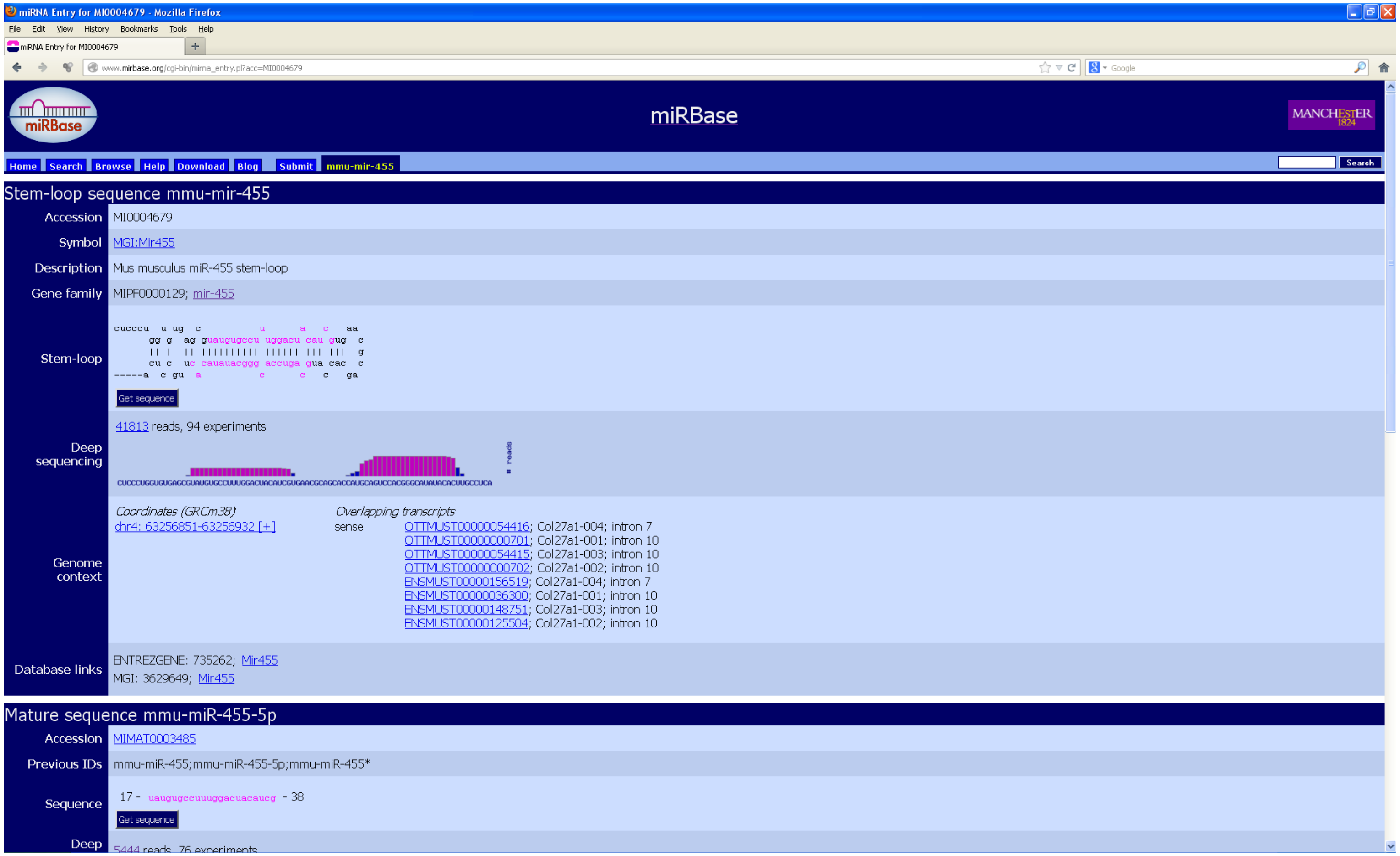


Stem-loop sequence hsa-mir-146a
Accession: MI0000477
Previous: hsa-mir-146

Mature sequence hsa-miR-146a-5p
Accession: MIMAT0000449
Previous: hsa-miR-146;hsa-miR-146a

Mature sequence hsa-miR-146a-3p
Accession: MIMAT0004608
Previous: hsa-miR-146a*

miRBase: mmu-mir-455 entry



miRNA Entry for MI0004679 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
miRNA Entry for MI0004679
www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0004679
Google
miRBase
miRBase
MANCHESTER 1824
Home Search Browse Help Download Blog Submit mmu-mir-455
Search

Stem-loop sequence mmu-mir-455

Accession	MI0004679
Symbol	MGI:Mir455
Description	Mus musculus miR-455 stem-loop
Gene family	MIPF0000129; mir-455

Stem-loop

```
cucccu  u ug   c          u      a   c   aa
      gg g  ag guaugugccu uggacu cau gug  c
      || |  || |||||||||| |||||| ||| |||  g
      cu c  uc cauauacggg accuga gua cac  c
-----a  c gu  a          c      c   c   ga
```

Get sequence

Deep sequencing

41813 reads, 94 experiments

CUCCCUGGUGUGAGCGUAUGUGCCUUUGGACUACAUCGUGAACGCAGCACCAUGCAGUCCACGGGCAUAUACACUUGCCUCA
# reads

Genome context

Coordinates (GRCm38)
chr4: 63256851-63256932 [+]

Overlapping transcripts

sense	OTTMUST00000054416; Col27a1-004; intron 7
	OTTMUST00000000701; Col27a1-001; intron 10
	OTTMUST00000054415; Col27a1-003; intron 10
	OTTMUST00000000702; Col27a1-002; intron 10
	ENSMUST00000156519; Col27a1-004; intron 7
	ENSMUST00000036300; Col27a1-001; intron 10
	ENSMUST00000148751; Col27a1-003; intron 10
	ENSMUST00000125504; Col27a1-002; intron 10

Database links

ENTREZGENE: 735262; Mir455
MGI: 3629649; Mir455

Mature sequence mmu-miR-455-5p

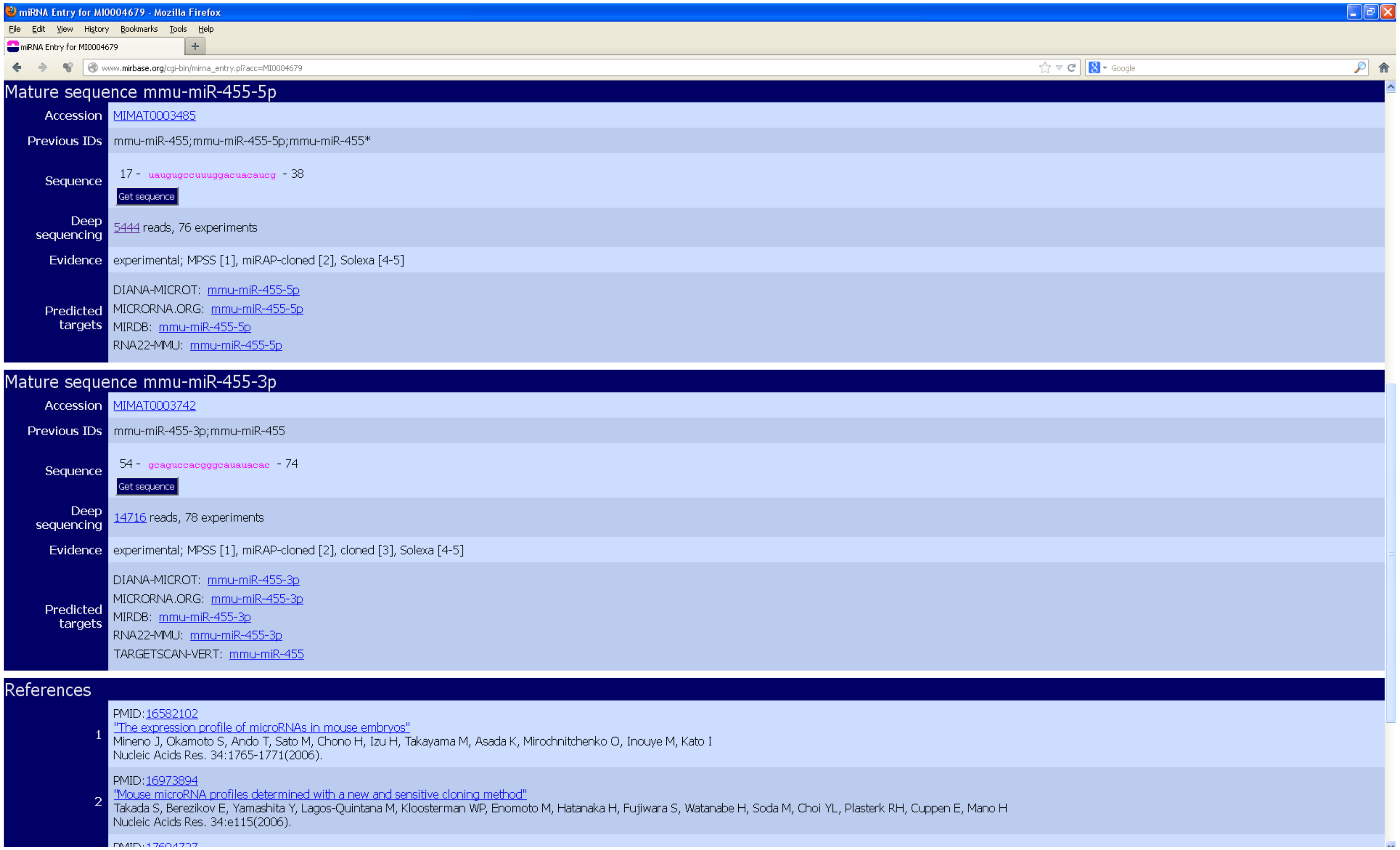
Accession	MIMAT0003485
Previous IDs	mmu-miR-455;mmu-miR-455-5p;mmu-miR-455*
Sequence	17 - uaugugccuuuggacuacaucg - 38

Get sequence

Deep

5444 reads, 76 experiments

miRBase: mmu-mir-455 entry


miRNA Entry for MI0004679 - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
miRNA Entry for MI0004679
www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0004679
Google

Mature sequence mmu-miR-455-5p

Accession	MIMAT0003485
Previous IDs	mmu-miR-455;mmu-miR-455-5p;mmu-miR-455*
Sequence	17 - uaugugccuuuggacuacaucg - 38 Get sequence
Deep sequencing	5444 reads, 76 experiments
Evidence	experimental; MPSS [1], miRAP-cloned [2], Solexa [4-5]
Predicted targets	DIANA-MICROT: mmu-miR-455-5p MICRORNA.ORG: mmu-miR-455-5p MIRDB: mmu-miR-455-5p RNA22-MMU: mmu-miR-455-5p

Mature sequence mmu-miR-455-3p

Accession	MIMAT0003742
Previous IDs	mmu-miR-455-3p;mmu-miR-455
Sequence	54 - gcaguccacgggcauauacac - 74 Get sequence
Deep sequencing	14716 reads, 78 experiments
Evidence	experimental; MPSS [1], miRAP-cloned [2], cloned [3], Solexa [4-5]
Predicted targets	DIANA-MICROT: mmu-miR-455-3p MICRORNA.ORG: mmu-miR-455-3p MIRDB: mmu-miR-455-3p RNA22-MMU: mmu-miR-455-3p TARGETSCAN-VERT: mmu-miR-455

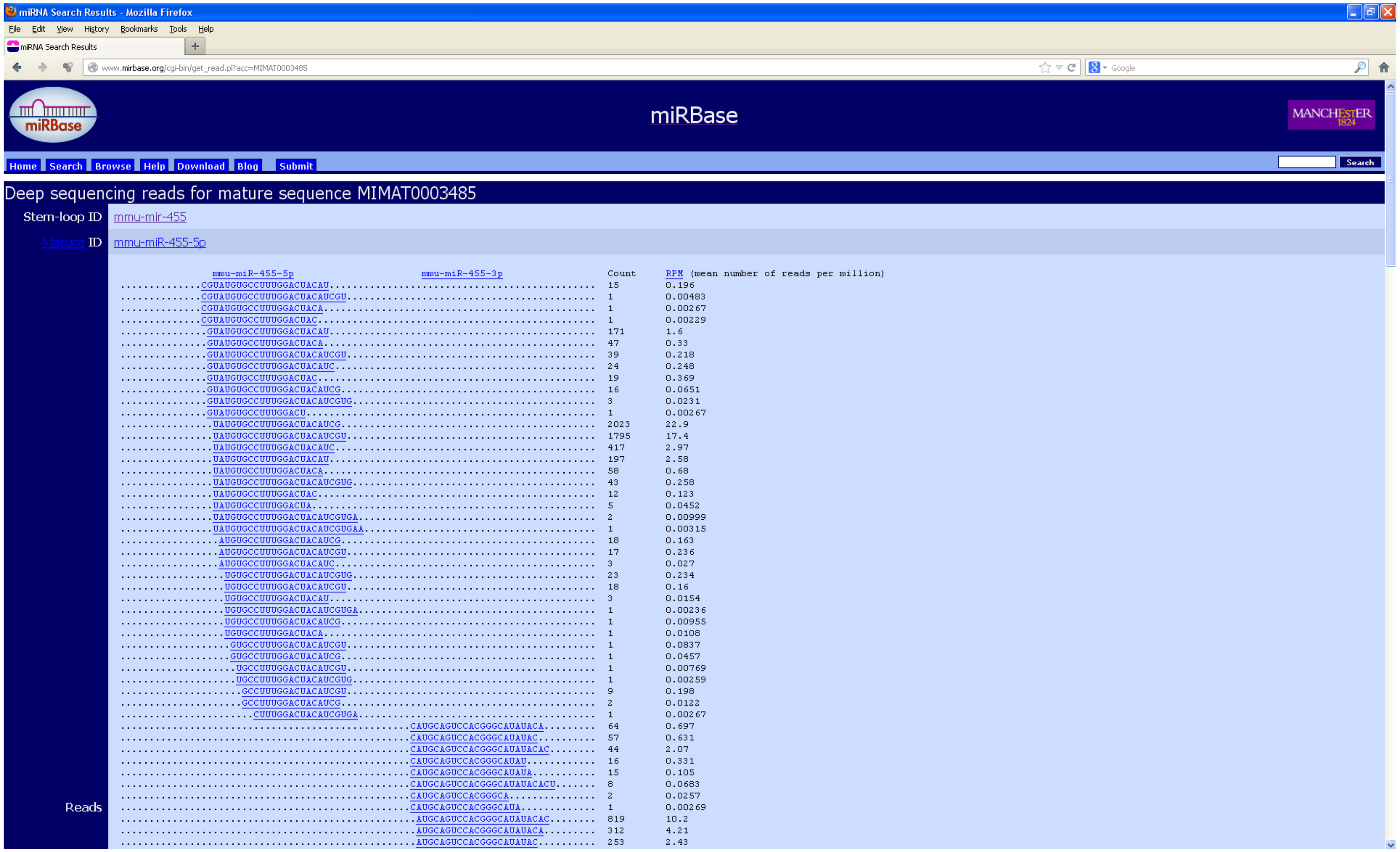
References

1 PMID:16582102
"The expression profile of microRNAs in mouse embryos"
Mineno J, Okamoto S, Ando T, Sato M, Chono H, Izu H, Takayama M, Asada K, Mirochnitchenko O, Inouye M, Kato I
Nucleic Acids Res. 34:1765-1771(2006).

2 PMID:16973894
"Mouse microRNA profiles determined with a new and sensitive cloning method"
Takada S, Berezikov E, Yamashita Y, Lagos-Quintana M, Kloosterman WP, Enomoto M, Hatanaka H, Fujiwara S, Watanabe H, Soda M, Choi YL, Plasterk RH, Cuppen E, Mano H
Nucleic Acids Res. 34:e115(2006).

PMID:17604727

mmu-miR-455-5p deep sequencing


miRNA Search Results - Mozilla Firefox
File Edit View History Bookmarks Tools Help
miRNA Search Results
www.mirbase.org/cgi-bin/get_read.pl?acc=MIMAT0003485
Google
miRBase
miRBase
MANCHESTER 1824
Home
Search
Browse
Help
Download
Blog
Submit
Search

Deep sequencing reads for mature sequence MIMAT0003485

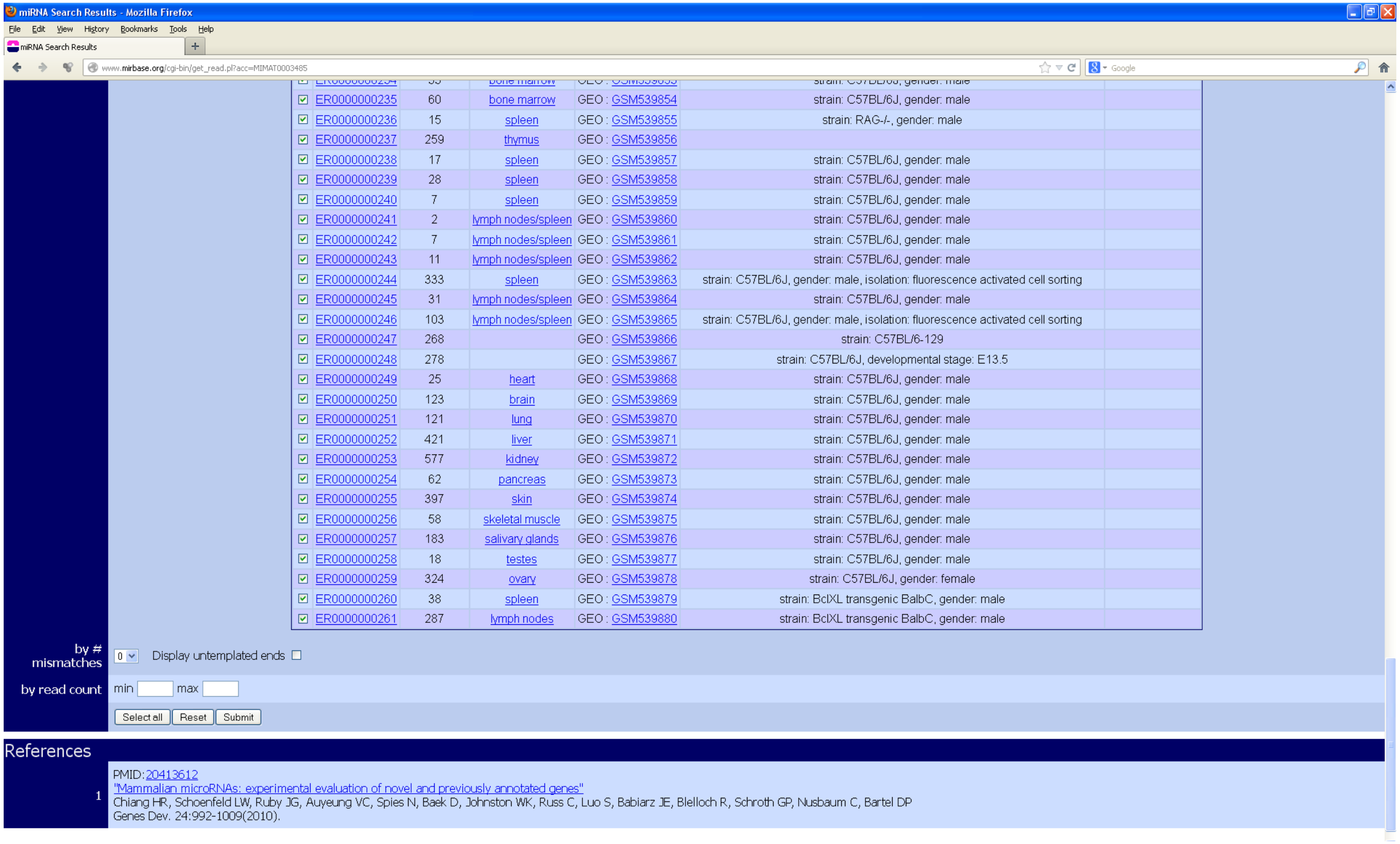
Stem-loop ID mmu-mir-455

Mature ID mmu-miR-455-5p

Reads

```
               mmu-miR-455-5p                     mmu-miR-455-3p                  Count     RPM (mean number of reads per million)
..............CGUAUGUGCCUUUGGACUACAU..............................................  15        0.196
..............CGUAUGUGCCUUUGGACUACAUCGU...........................................  1         0.00483
..............CGUAUGUGCCUUUGGACUACA...............................................  1         0.00267
..............CGUAUGUGCCUUUGGACUAC................................................  1         0.00229
...............GUAUGUGCCUUUGGACUACAU..............................................  171       1.6
...............GUAUGUGCCUUUGGACUACA...............................................  47        0.33
...............GUAUGUGCCUUUGGACUACAUCGU...........................................  39        0.218
...............GUAUGUGCCUUUGGACUACAUC.............................................  24        0.248
...............GUAUGUGCCUUUGGACUAC................................................  19        0.369
...............GUAUGUGCCUUUGGACUACAUCG............................................  16        0.0651
...............GUAUGUGCCUUUGGACUACAUCGUG..........................................  3         0.0231
...............GUAUGUGCCUUUGGACU..................................................  1         0.00267
................UAUGUGCCUUUGGACUACAUCG............................................  2023      22.9
................UAUGUGCCUUUGGACUACAUCGU...........................................  1795      17.4
................UAUGUGCCUUUGGACUACAUC.............................................  417       2.97
................UAUGUGCCUUUGGACUACAU..............................................  197       2.58
................UAUGUGCCUUUGGACUACA...............................................  58        0.68
................UAUGUGCCUUUGGACUACAUCGUG..........................................  43        0.258
................UAUGUGCCUUUGGACUAC................................................  12        0.123
................UAUGUGCCUUUGGACUA.................................................  5         0.0452
................UAUGUGCCUUUGGACUACAUCGUGA.........................................  2         0.00999
................UAUGUGCCUUUGGACUACAUCGUGAA........................................  1         0.00315
.................AUGUGCCUUUGGACUACAUCG............................................  18        0.163
.................AUGUGCCUUUGGACUACAUCGU...........................................  17        0.236
.................AUGUGCCUUUGGACUACAUC.............................................  3         0.027
..................UGUGCCUUUGGACUACAUCGUG..........................................  23        0.234
..................UGUGCCUUUGGACUACAUCGU...........................................  18        0.16
..................UGUGCCUUUGGACUACAU..............................................  3         0.0154
..................UGUGCCUUUGGACUACAUCGUGA.........................................  1         0.00236
..................UGUGCCUUUGGACUACAUCG............................................  1         0.00955
..................UGUGCCUUUGGACUACA...............................................  1         0.0108
...................GUGCCUUUGGACUACAUCGU...........................................  1         0.0837
...................GUGCCUUUGGACUACAUCG............................................  1         0.0457
....................UGCCUUUGGACUACAUCGU...........................................  1         0.00769
....................UGCCUUUGGACUACAUCGUG..........................................  1         0.00259
.....................GCCUUUGGACUACAUCGU...........................................  9         0.198
.....................GCCUUUGGACUACAUCG............................................  2         0.0122
.......................CUUUGGACUACAUCGUGA.........................................  1         0.00267
..................................................CAUGCAGUCCACGGGCAUAUACA.........  64        0.697
..................................................CAUGCAGUCCACGGGCAUAUAC..........  57        0.631
..................................................CAUGCAGUCCACGGGCAUAUACAC........  44        2.07
..................................................CAUGCAGUCCACGGGCAUAU............  16        0.331
..................................................CAUGCAGUCCACGGGCAUAUA...........  15        0.105
..................................................CAUGCAGUCCACGGGCAUAUACACU.......  8         0.0683
..................................................CAUGCAGUCCACGGGCA...............  2         0.0257
..................................................CAUGCAGUCCACGGGCAUA.............  1         0.00269
...................................................AUGCAGUCCACGGGCAUAUACAC........  819       10.2
...................................................AUGCAGUCCACGGGCAUAUACA.........  312       4.21
...................................................AUGCAGUCCACGGGCAUAUAC..........  253       2.43
```

mmu-miR-455-5p deep sequencing


miRNA Search Results - Mozilla Firefox
File Edit View History Bookmarks Tools Help
miRNA Search Results
www.mirbase.org/cgi-bin/get_read.pl?acc=MIMAT0003485
Google

☑	ER0000000234	55	bone marrow	GEO : GSM539853	strain: C57BL/6J, gender: male
☑	ER0000000235	60	bone marrow	GEO : GSM539854	strain: C57BL/6J, gender: male
☑	ER0000000236	15	spleen	GEO : GSM539855	strain: RAG-/-, gender: male
☑	ER0000000237	259	thymus	GEO : GSM539856	
☑	ER0000000238	17	spleen	GEO : GSM539857	strain: C57BL/6J, gender: male
☑	ER0000000239	28	spleen	GEO : GSM539858	strain: C57BL/6J, gender: male
☑	ER0000000240	7	spleen	GEO : GSM539859	strain: C57BL/6J, gender: male
☑	ER0000000241	2	lymph nodes/spleen	GEO : GSM539860	strain: C57BL/6J, gender: male
☑	ER0000000242	7	lymph nodes/spleen	GEO : GSM539861	strain: C57BL/6J, gender: male
☑	ER0000000243	11	lymph nodes/spleen	GEO : GSM539862	strain: C57BL/6J, gender: male
☑	ER0000000244	333	spleen	GEO : GSM539863	strain: C57BL/6J, gender: male, isolation: fluorescence activated cell sorting
☑	ER0000000245	31	lymph nodes/spleen	GEO : GSM539864	strain: C57BL/6J, gender: male
☑	ER0000000246	103	lymph nodes/spleen	GEO : GSM539865	strain: C57BL/6J, gender: male, isolation: fluorescence activated cell sorting
☑	ER0000000247	268		GEO : GSM539866	strain: C57BL/6-129
☑	ER0000000248	278		GEO : GSM539867	strain: C57BL/6J, developmental stage: E13.5
☑	ER0000000249	25	heart	GEO : GSM539868	strain: C57BL/6J, gender: male
☑	ER0000000250	123	brain	GEO : GSM539869	strain: C57BL/6J, gender: male
☑	ER0000000251	121	lung	GEO : GSM539870	strain: C57BL/6J, gender: male
☑	ER0000000252	421	liver	GEO : GSM539871	strain: C57BL/6J, gender: male
☑	ER0000000253	577	kidney	GEO : GSM539872	strain: C57BL/6J, gender: male
☑	ER0000000254	62	pancreas	GEO : GSM539873	strain: C57BL/6J, gender: male
☑	ER0000000255	397	skin	GEO : GSM539874	strain: C57BL/6J, gender: male
☑	ER0000000256	58	skeletal muscle	GEO : GSM539875	strain: C57BL/6J, gender: male
☑	ER0000000257	183	salivary glands	GEO : GSM539876	strain: C57BL/6J, gender: male
☑	ER0000000258	18	testes	GEO : GSM539877	strain: C57BL/6J, gender: male
☑	ER0000000259	324	ovary	GEO : GSM539878	strain: C57BL/6J, gender: female
☑	ER0000000260	38	spleen	GEO : GSM539879	strain: BclXL transgenic BalbC, gender: male
☑	ER0000000261	287	lymph nodes	GEO : GSM539880	strain: BclXL transgenic BalbC, gender: male

by # mismatches 0 Display untemplated ends □

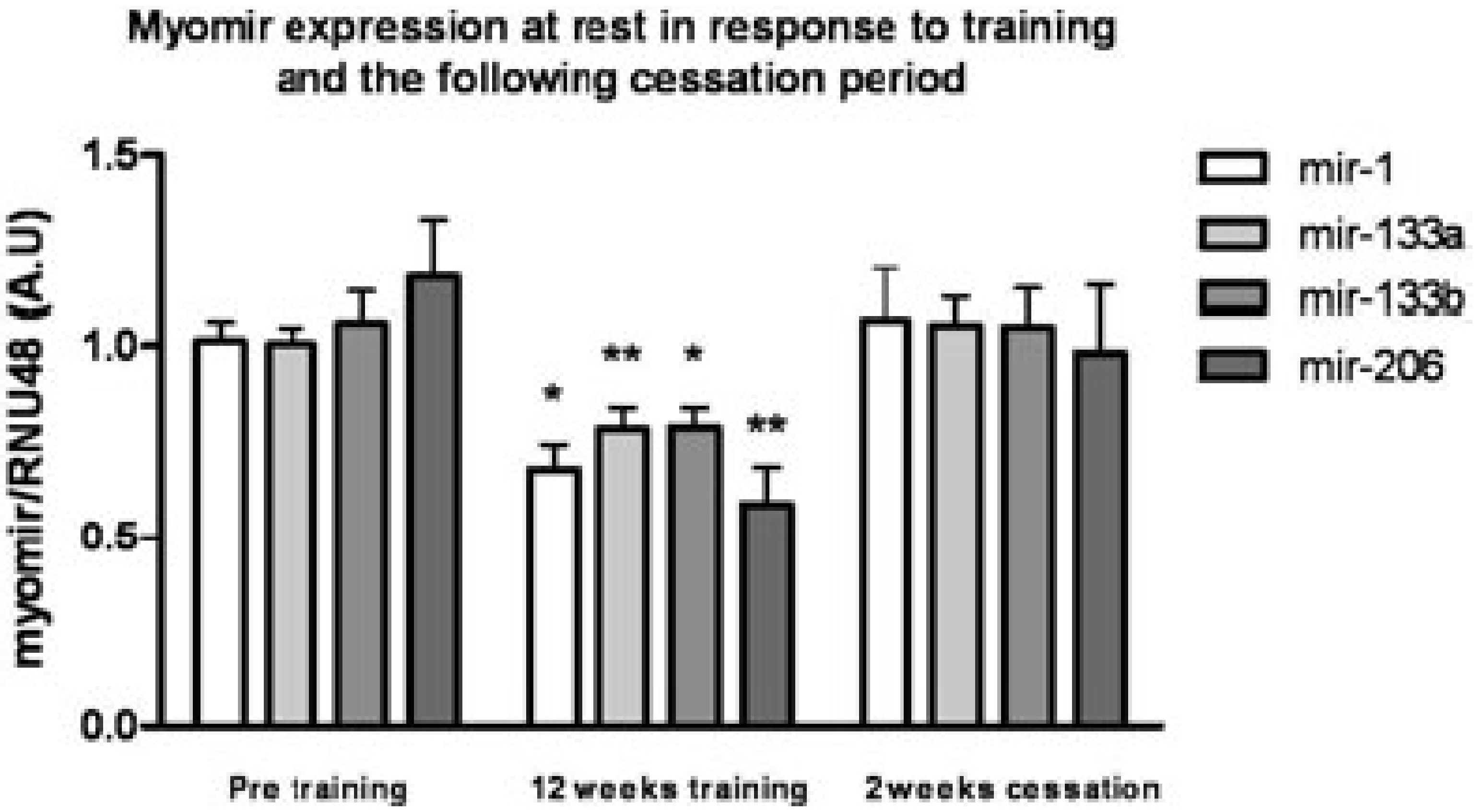
by read count min max

Select all Reset Submit

References

1 PMID:20413612
"Mammalian microRNAs: experimental evaluation of novel and previously annotated genes"
Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP
Genes Dev. 24:992-1009(2010).

miRNAs and endurance training

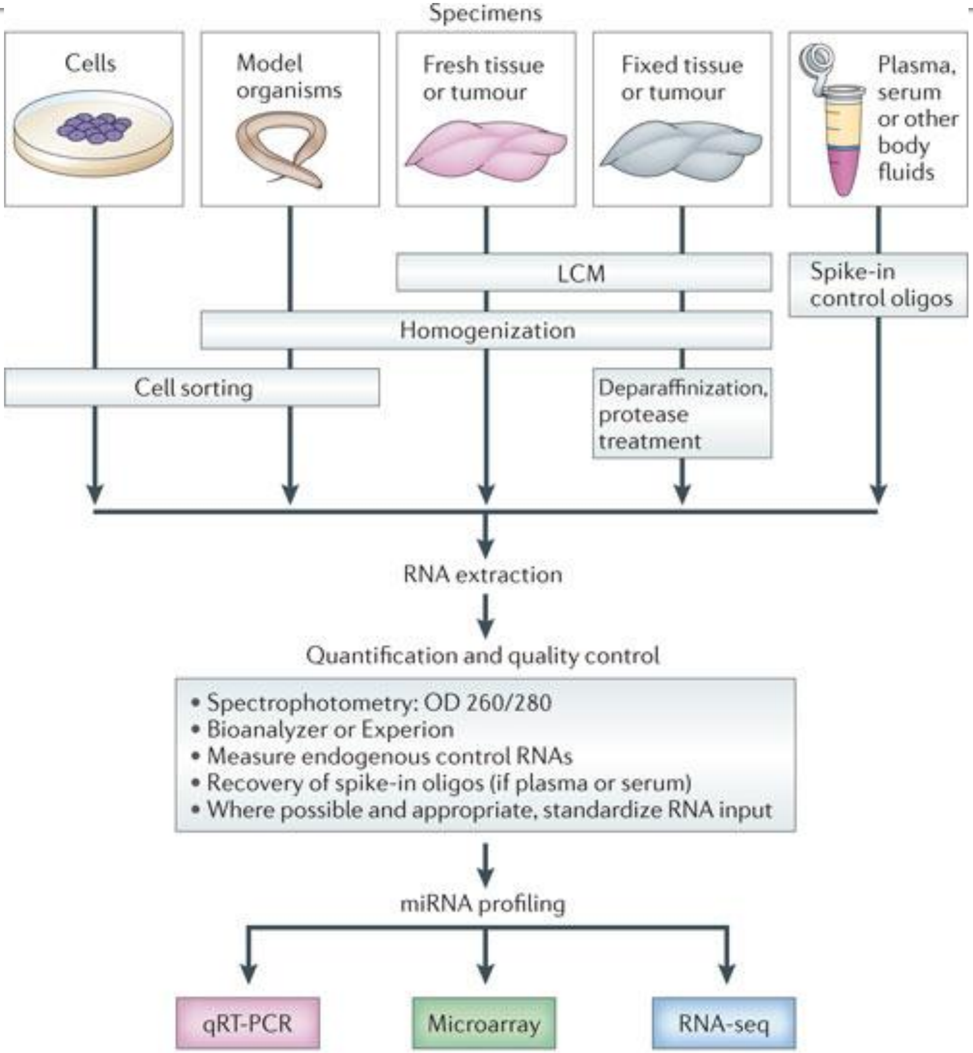


Nielsen et a. (2010) Muscle specific microRNAs and exercise. *J Physiol*

miRNA and disease

- Cancer:
  - Several miRNAs have been found to be overexpressed in specific types of cancer.
  - Patterns of miRNA activity can be used to distinguish several types of cancers: **biomarker profiles**
  - Useful to identify cancers of unknown origin
- Heart disease:
  - Specific miRNAs change in diseased human hearts: **biomarker profiles**

microRNA profiling (1)



Specimens

Cells

Model organisms

Fresh tissue or tumour

Fixed tissue or tumour

Plasma, serum or other body fluids

LCM

Spike-in control oligos

Homogenization

Cell sorting

Deparaffinization, protease treatment

RNA extraction

Quantification and quality control

- Spectrophotometry: OD 260/280
- Bioanalyzer or Experion
- Measure endogenous control RNAs
- Recovery of spike-in oligos (if plasma or serum)
- Where possible and appropriate, standardize RNA input

miRNA profiling

qRT-PCR

Microarray

RNA-seq

Nature Reviews | Genetics

microRNA profiling (2)

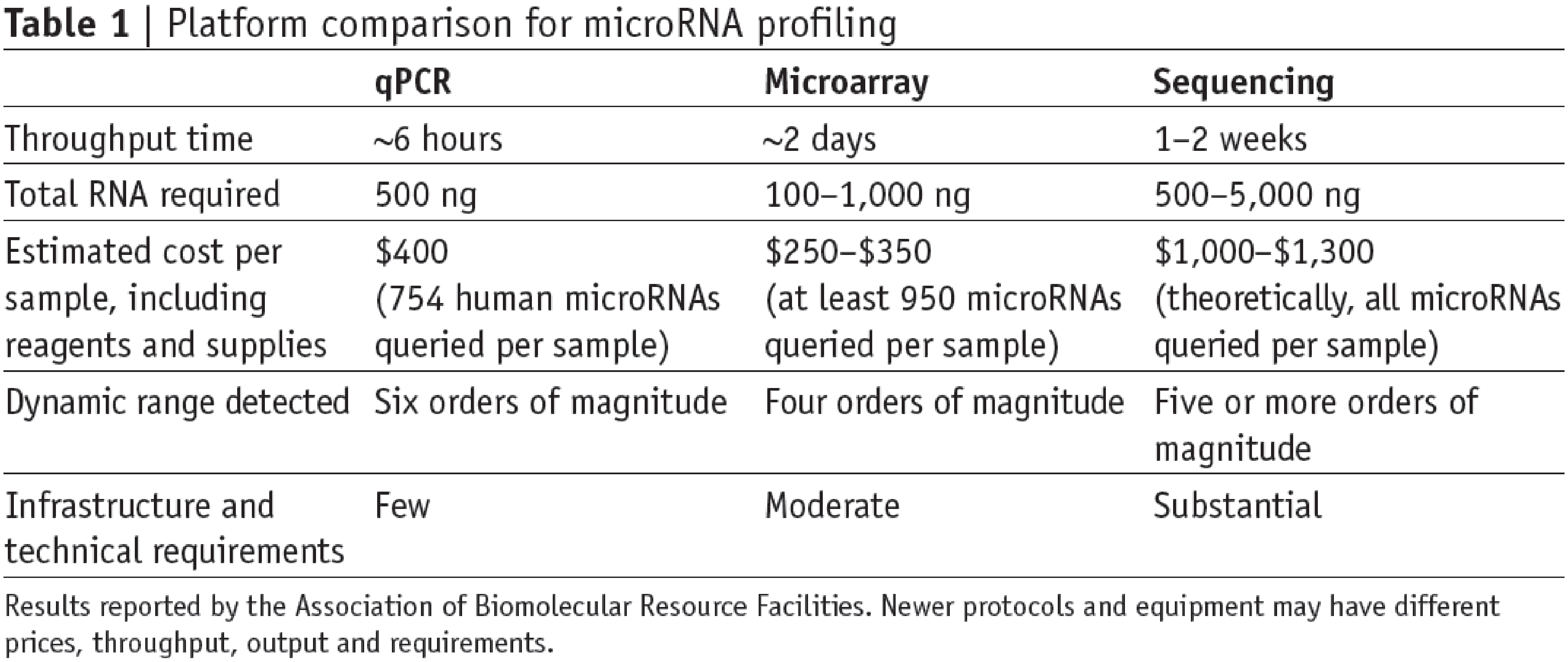


Table 1 | Platform comparison for microRNA profiling

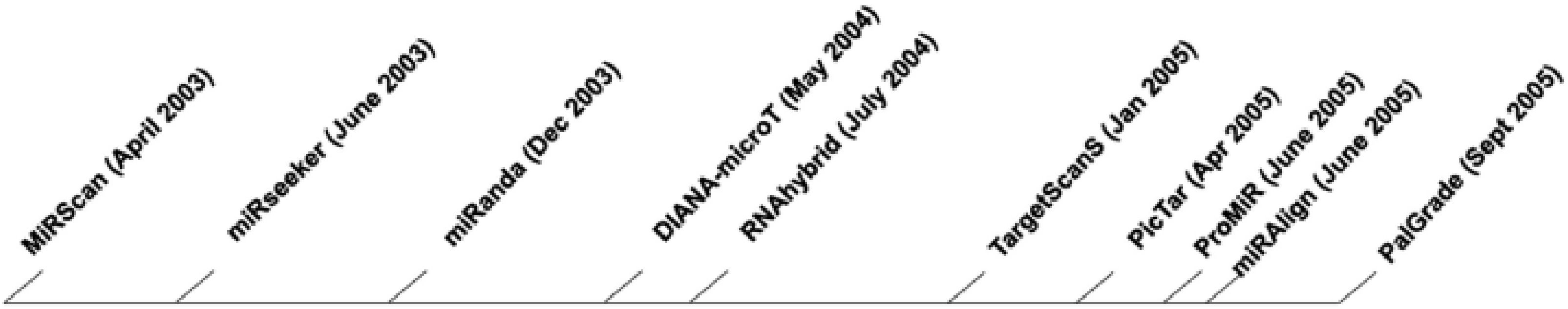
	qPCR	Microarray	Sequencing
Throughput time	~6 hours	~2 days	1–2 weeks
Total RNA required	500 ng	100–1,000 ng	500–5,000 ng
Estimated cost per sample, including reagents and supplies	$400 (754 human microRNAs queried per sample)	$250–$350 (at least 950 microRNAs queried per sample)	$1,000–$1,300 (theoretically, all microRNAs queried per sample)
Dynamic range detected	Six orders of magnitude	Four orders of magnitude	Five or more orders of magnitude
Infrastructure and technical requirements	Few	Moderate	Substantial

Results reported by the Association of Biomolecular Resource Facilities. Newer protocols and equipment may have different prices, throughput, output and requirements.

Baker, Monya (2010). "MicroRNA profiling: separating signal from noise". *Nature Methods* **7** (9): 687–69

miRNA target prediction

Target prediction history



- **Solutions make use of:**
  1. alignment algorithms
  2. conservation rates
  3. thermodynamics

Alignment scores

Nucleotides 2–7 of the miRNA ('seed region') need to be perfectly complementary:

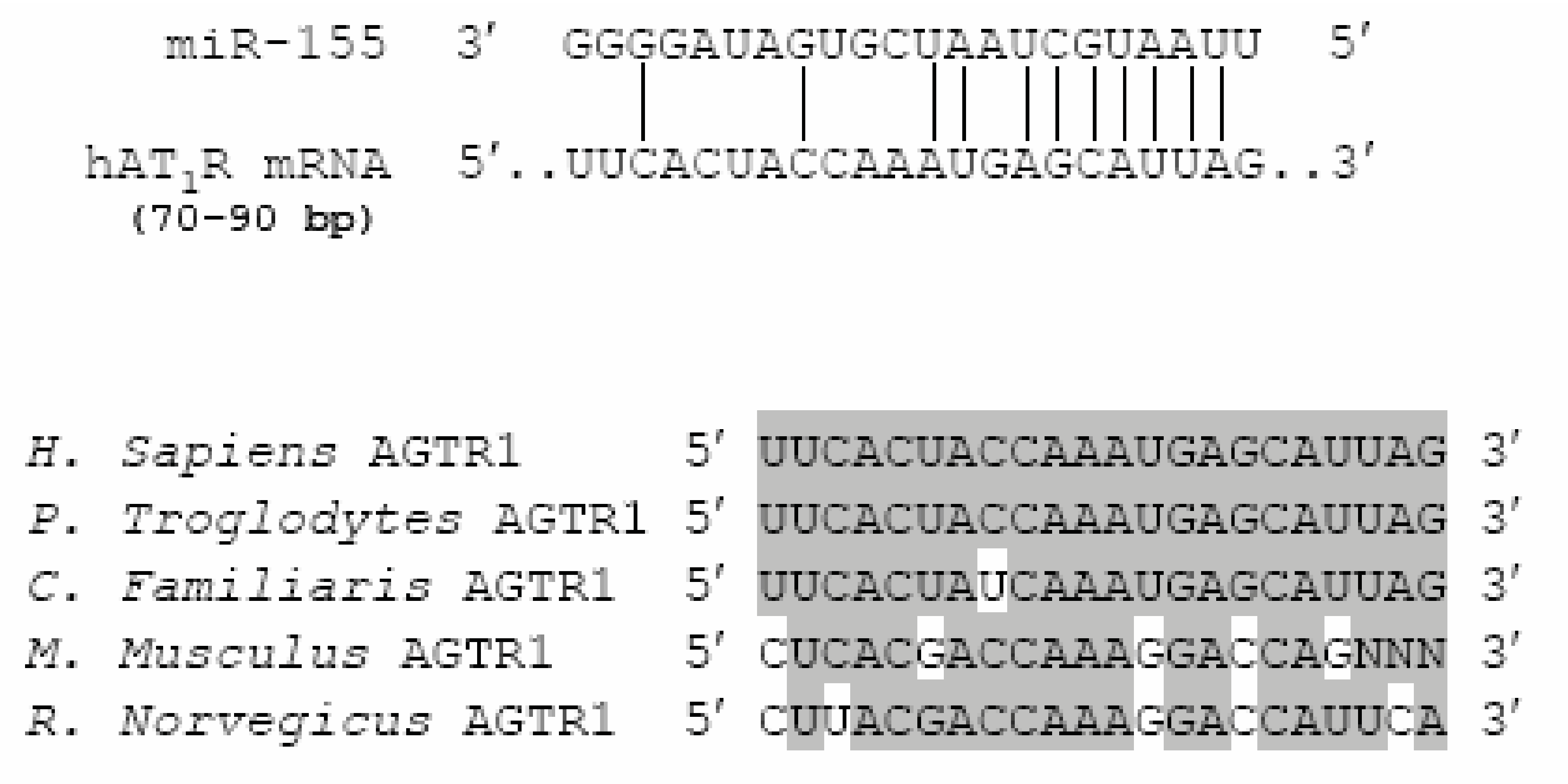
```
5' ..GGAACAA-AUUUC-UCUGGGACAGCCCUAAAUGCAGAC.. 3'
             | ||| ||  || |||||||
         3' CUUAAGUAGU-CCGGUCGGGAA 5'
```

Good

```
5' ..AUCGAUU-AUAAC-UCUGGCACAAUCCAGCCCAGACG.. 3'
             |   | ||  |  ||  ||
         3' CUUAAGUAGU-CCGGUCGGGAA 5'
```

Not so good

Conservation

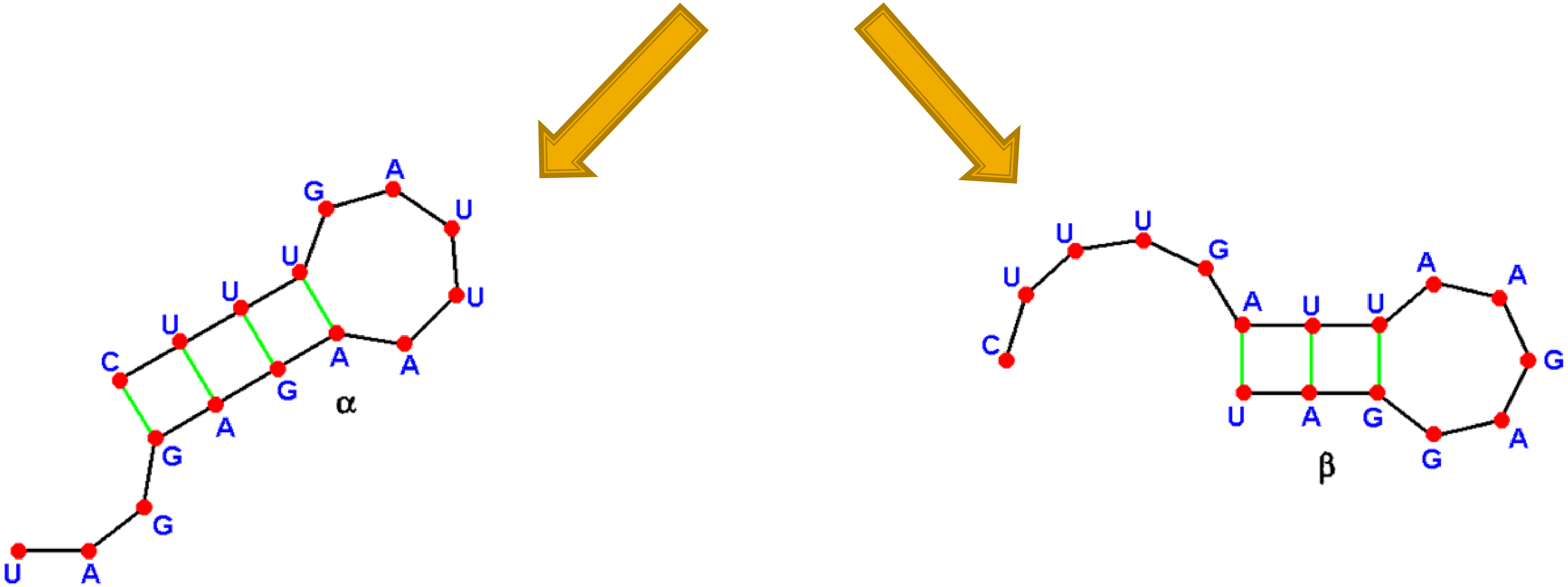


```
    miR-155   3'  GGGGAUAGUGCUAAUCGUAAUU  5'
                      |    |    || | | | | | ||
 hAT1R mRNA   5'..UUCACUACCAAAUGAGCAUUAG..3'
  (70-90 bp)
```

```
H. Sapiens AGTR1         5' UUCACUACCAAAUGAGCAUUAG 3'
P. Troglodytes AGTR1     5' UUCACUACCAAAUGAGCAUUAG 3'
C. Familiaris AGTR1      5' UUCACUAUCAAAUGAGCAUUAG 3'
M. Musculus AGTR1        5' CUCACGACCAAAGGACCAGNNN 3'
R. Norvegicus AGTR1      5' CUUACGACCAAAGGACCAUUCA 3'
```

Thermodynamics: free energy

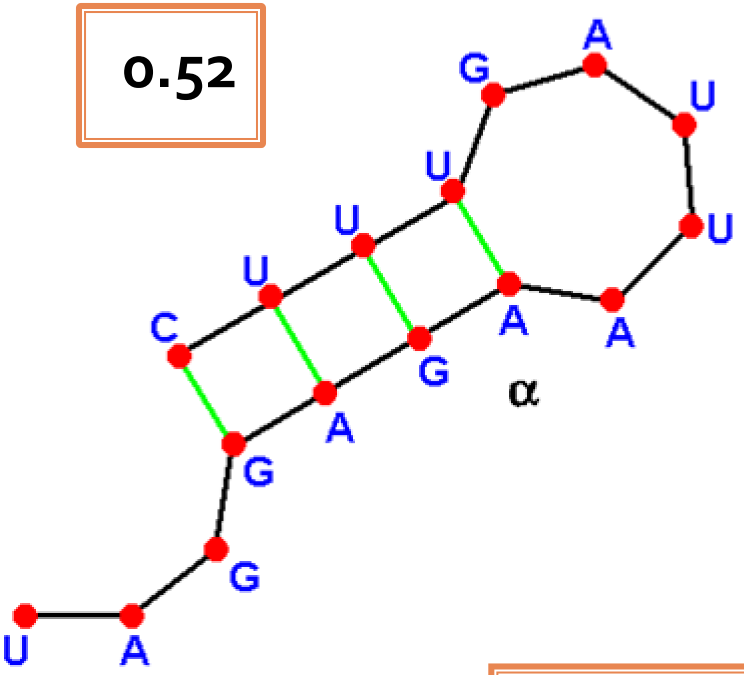
UAGGAGAAUUAGUUUC



Stronger bond = Higher free energy

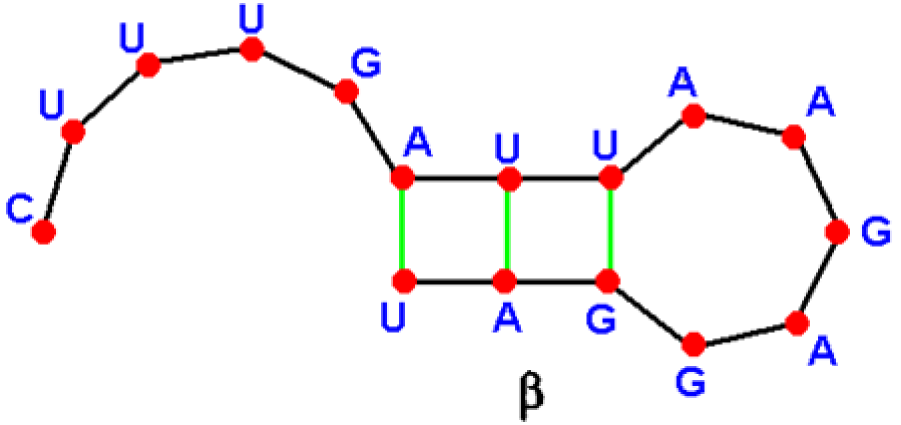
Different configurations

0.52



α

0.03



β

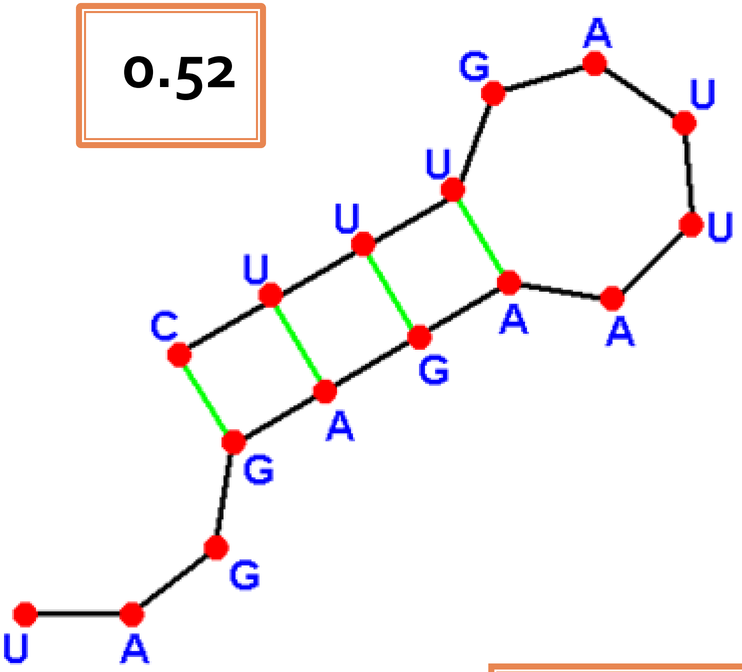
Different configurations

=

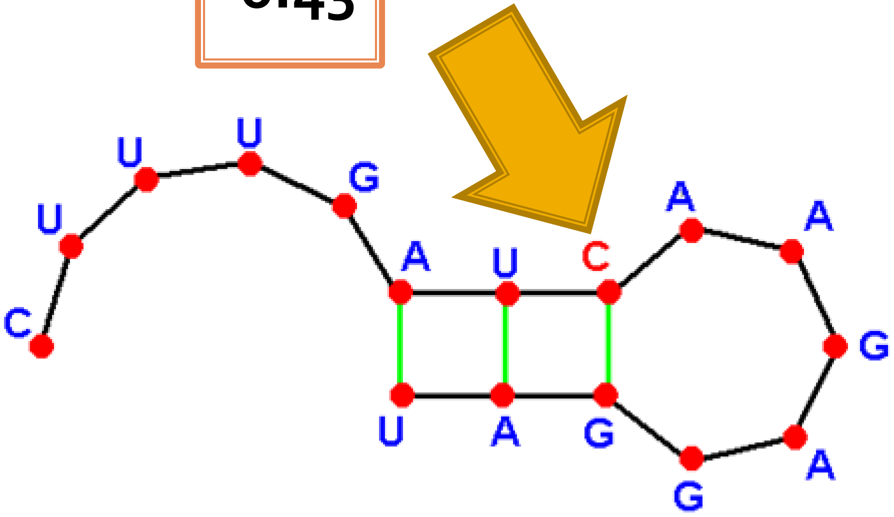
Different free energy

Different configurations

0.52



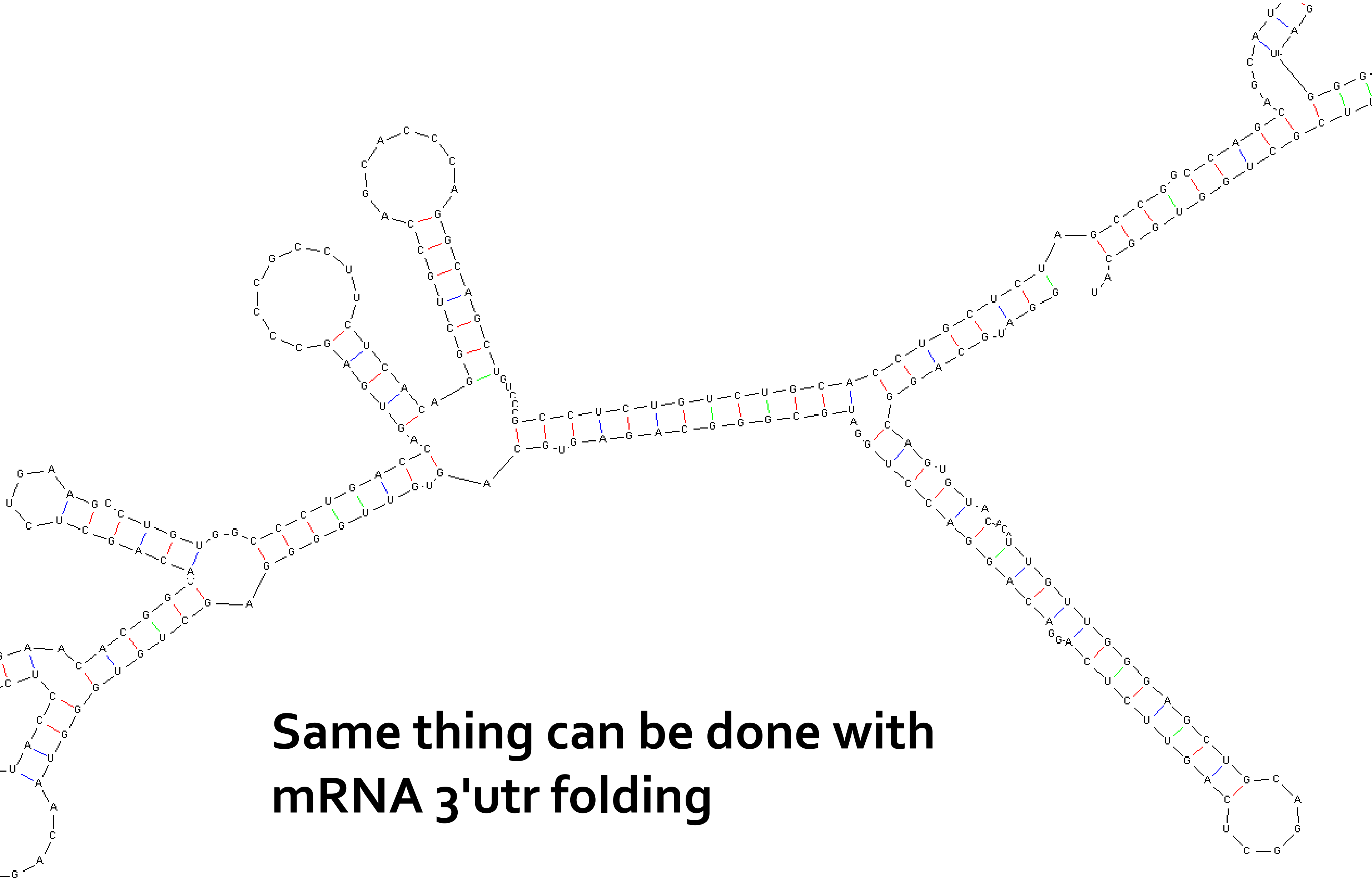
0.43



Different configurations

=

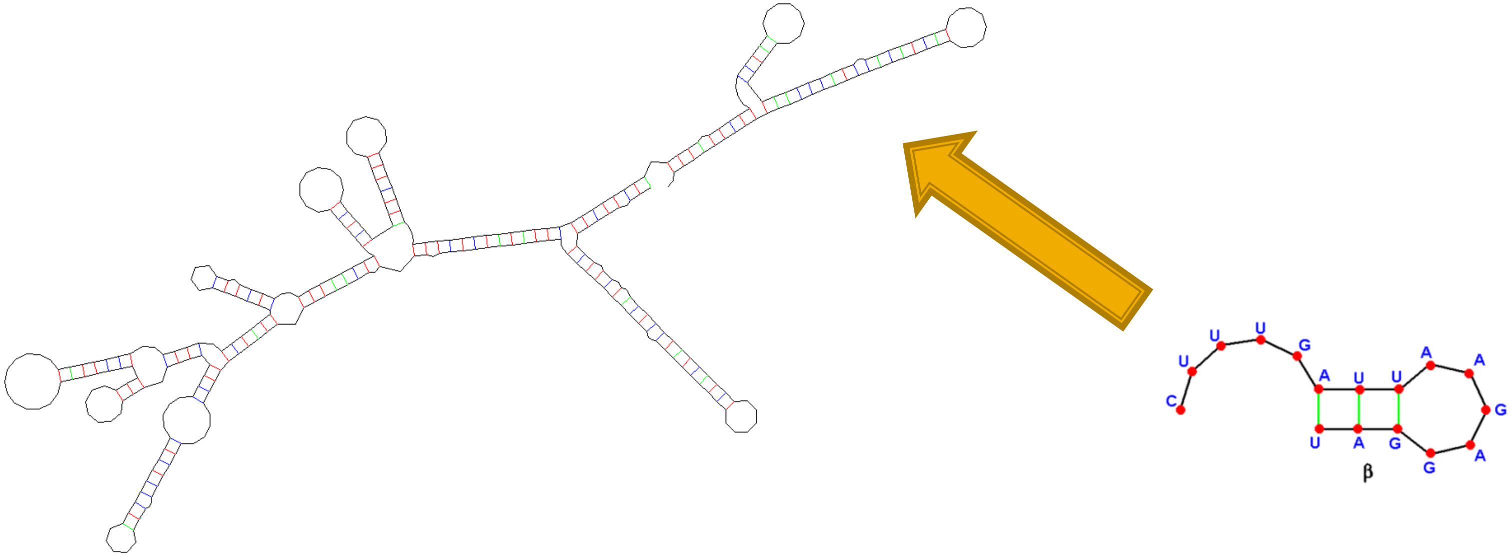
Different free energy



Same thing can be done with mRNA 3'utr folding

Potential target sites

- Thermodynamical requirements:
  - low energy of self-binding for both the miRNA and mRNA
  - high energy of resulting 3'utr - miRNA binding



Database overview

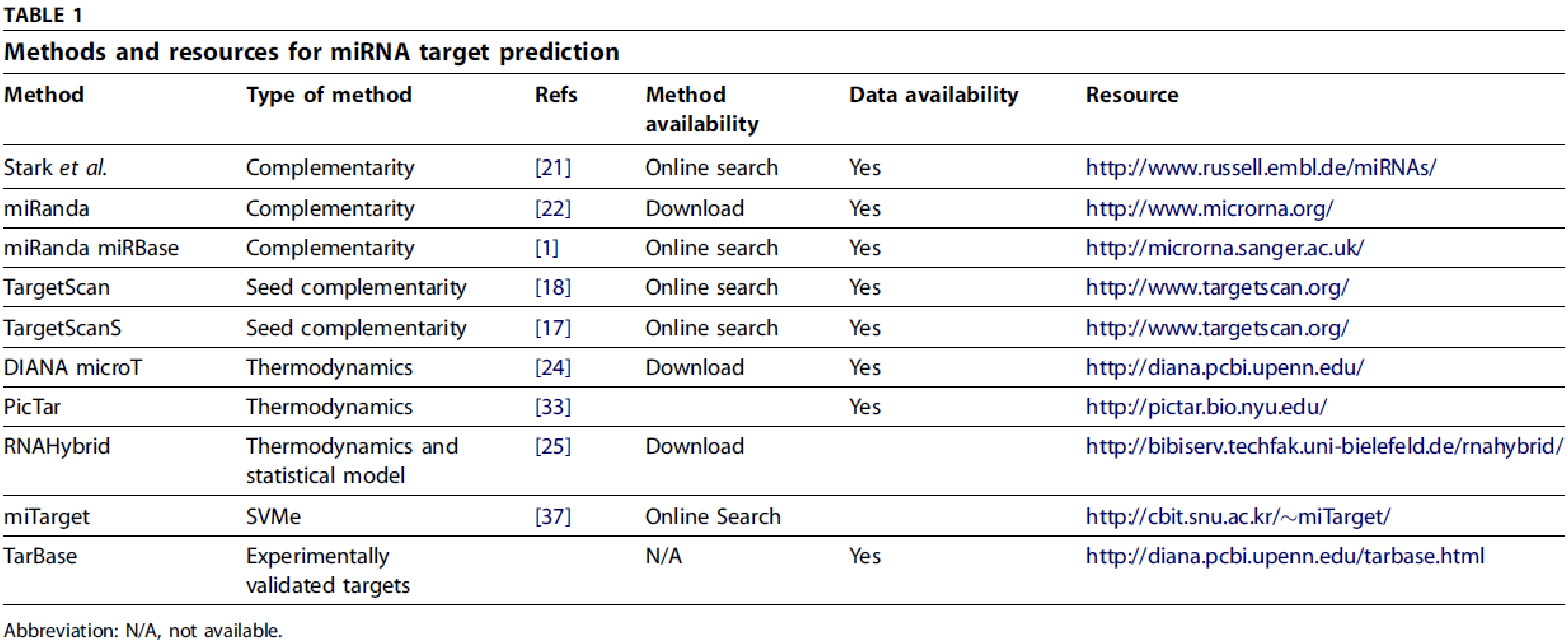


TABLE 1

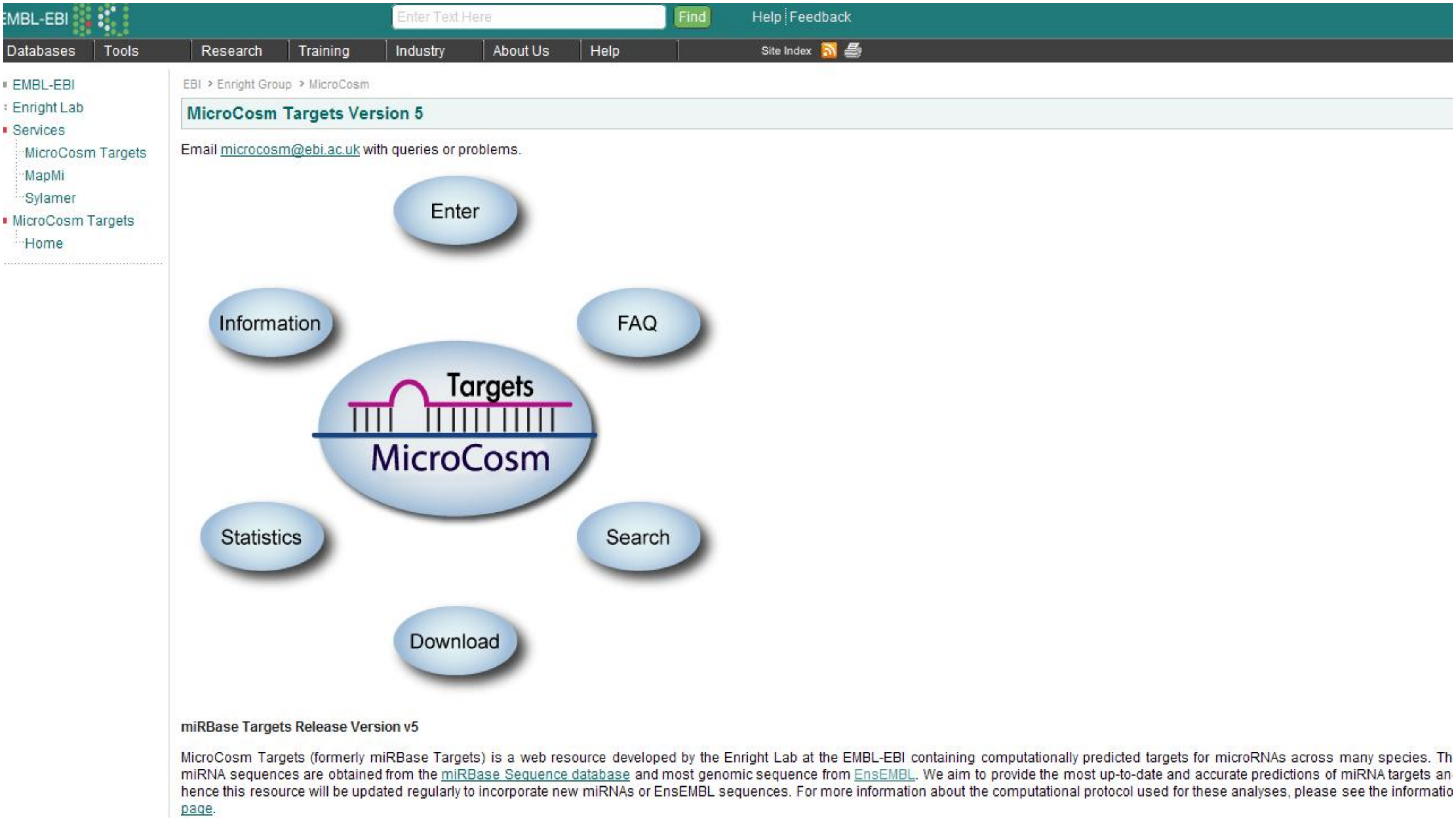
Methods and resources for miRNA target prediction

Method	Type of method	Refs	Method availability	Data availability	Resource
Stark *et al.*	Complementarity	[21]	Online search	Yes	http://www.russell.embl.de/miRNAs/
miRanda	Complementarity	[22]	Download	Yes	http://www.microrna.org/
miRanda miRBase	Complementarity	[1]	Online search	Yes	http://microrna.sanger.ac.uk/
TargetScan	Seed complementarity	[18]	Online search	Yes	http://www.targetscan.org/
TargetScanS	Seed complementarity	[17]	Online search	Yes	http://www.targetscan.org/
DIANA microT	Thermodynamics	[24]	Download	Yes	http://diana.pcbi.upenn.edu/
PicTar	Thermodynamics	[33]		Yes	http://pictar.bio.nyu.edu/
RNAHybrid	Thermodynamics and statistical model	[25]	Download		http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
miTarget	SVMe	[37]	Online Search		http://cbit.snu.ac.kr/~miTarget/
TarBase	Experimentally validated targets		N/A	Yes	http://diana.pcbi.upenn.edu/tarbase.html

Abbreviation: N/A, not available.

Mazière, P, et al. (2007) Prediction of microRNA targets. Drug discovery today, Vol. 12 (11-12): 452-8.

Scanning is done by going to the MicroCosm Targets website (linked on front page of miRBase)



MicroCosm Targets Version 5

Email microcosm@ebi.ac.uk with queries or problems.

miRBase Targets Release Version v5

MicroCosm Targets (formerly miRBase Targets) is a web resource developed by the Enright Lab at the EMBL-EBI containing computationally predicted targets for microRNAs across many species. Th miRNA sequences are obtained from the miRBase Sequence database and most genomic sequence from EnsEMBL. We aim to provide the most up-to-date and accurate predictions of miRNA targets an hence this resource will be updated regularly to incorporate new miRNAs or EnsEMBL sequences. For more information about the computational protocol used for these analyses, please see the informatio page.

All miRNA hits for *Rattus norvegicus* and let-7a

500 hits found.

Page 1 of 10
1 2 3 4 5 6 7 8 9 10 next >>

Gene Name	Transcript	Gene	Description	GO Terms	Total Score	Total Energy	Best P value	Total Sites	No. Cons Species	No. miRNAs
▽	▽	▽	▽		▽	▽	▽	▽	▽	▽
NP_001013247.1	ENSRNOT00000014386	ENSRNOG00000010673	Era (G-protein)-like 1 (E. coli) (predicted) [Source:RefSeq_peptide;Acc:NP_001013247]	□ 🟦 □	**138**	**-214**	**9.15469e-10**	**8**	**5**	**6 [+]**
Q71KM5_RAT	ENSRNOT00000028130	ENSRNOG00000020733	CRAMP (Fragment). [Source:Uniprot/SPTREMBL;Acc:Q71KM5]	🟪 □ □	**37**	**-23**	**6.3227e-09**	**2**	**4**	**25 [+]**
	ENSRNOT00000032786	ENSRNOG00000007654	leucine-rich repeats and immunoglobulin-like domains 3 [Source:RefSeq_peptide;Acc:NP_700356]leucine-rich repeats and immunoglobulin-like domains 3 [Source:RefSeq_peptide;Acc:NP_700356] BY ORTHOLOGY TO:ENST00000320743	□ □ □	**64**	**-78**	**1.08549e-08**	**4**	**10**	**5 [+]**
ACADS_RAT	ENSRNOT00000001556	ENSRNOG00000001177	Acyl-CoA dehydrogenase, short-chain specific, mitochondrial precursor (EC 1.3.99.2) (SCAD) (Butyryl-CoA dehydrogenase). [Source:Uniprot/SWISSPROT;Acc:P15651]	🟩 🟦 🟩	**178**	**-221**	**1.97468e-08**	**11**	**8**	**16 [+]**
XP_213226.1	ENSRNOT00000005899	ENSRNOG00000004461	PREDICTED: similar to 2810417J12Rik protein [Source:RefSeq_peptide_predicted;Acc:XP_213226]	□ □ □	**37**	**-25**	**2.08358e-08**	**2**	**4**	**15 [+]**
XP_216873.1	ENSRNOT00000006903	ENSRNOG00000005102	PREDICTED: similar to RIKEN cDNA 2900091E11 [Source:RefSeq_peptide_predicted;Acc:XP_216873]	□ □ □	**82**	**-60**	**3.00314e-08**	**5**	**7**	**36 [+]**
NP_001004211.1	ENSRNOT00000006642	ENSRNOG00000004670	DEAD (Asp-Glu-Ala-Asp) box polypeptide 56 [Source:RefSeq_peptide;Acc:NP_001004211]	□ 🟩 □	**95**	**-103**	**3.15554e-08**	**6**	**4**	**30 [+]**
NP_001020047.1	ENSRNOT00000005376	ENSRNOG00000003964	RIKEN cDNA 1110014D18 gene (1110014D18Rik), mRNA [Source:RefSeq_dna;Acc:NM_026746]RIKEN cDNA 1110014D18 gene (1110014D18Rik), mRNA [Source:RefSeq_dna;Acc:NM_026746] BY ORTHOLOGY TO:ENSMUST00000079703	🟩 🟩 🟦	**69**	**-70**	**3.89717e-08**	**4**	**4**	**13 [+]**
CBPB2_RAT	ENSRNOT00000014909	ENSRNOG00000010935	Carboxypeptidase B2 precursor (EC 3.4.17.20) (Carboxypeptidase U) (Thrombin-activatable fibrinolysis inhibitor) (TAFI) (Carboxypeptidase R) (CPR). [Source:Uniprot/SWISSPROT;Acc:Q9EQV9]	🟩 🟩 □	**20**	**-7**	**6.12503e-08**	**1**	**2**	**10 [+]**
XP_343784.1	ENSRNOT00000004733	ENSRNOG00000003554	PREDICTED: similar to Pig-a precursor [Source:RefSeq_peptide_predicted;Acc:XP_343784]	□ □ □	**90**	**-62**	**1.13116e-07**	**5**	**7**	**21 [+]**
NP_001008889.1	ENSRNOT00000030476	ENSRNOG00000025704	HIV-induced protein-7-like protease [Source:RefSeq_peptide;Acc:NP_001008889]	□ 🟩 □	**80**	**-94**	**1.1802e-07**	**5**	**9**	**5 [+]**
XP_220798.3	ENSRNOT00000036814	ENSRNOG00000027711	PREDICTED: similar to ubiquitin specific protease 32 [Source:RefSeq_peptide_predicted;Acc:XP_220798]	🟩 🟩 □	**34**	**-33**	**1.27342e-07**	**2**	**8**	**3 [+]**

Practical session, May 16th