



Human Genome & Genome Browser

Never Stand Still

Medicine

Dominik Beck

NHMRC Peter Doherty Fellow and Senior Lecturer

Lowy Cancer Research Centre, UNSW and Centre for Health Technology, UTS
SYDNEY NSW AUSTRALIA



What we will cover

Structure of the
human genome



Genomic information

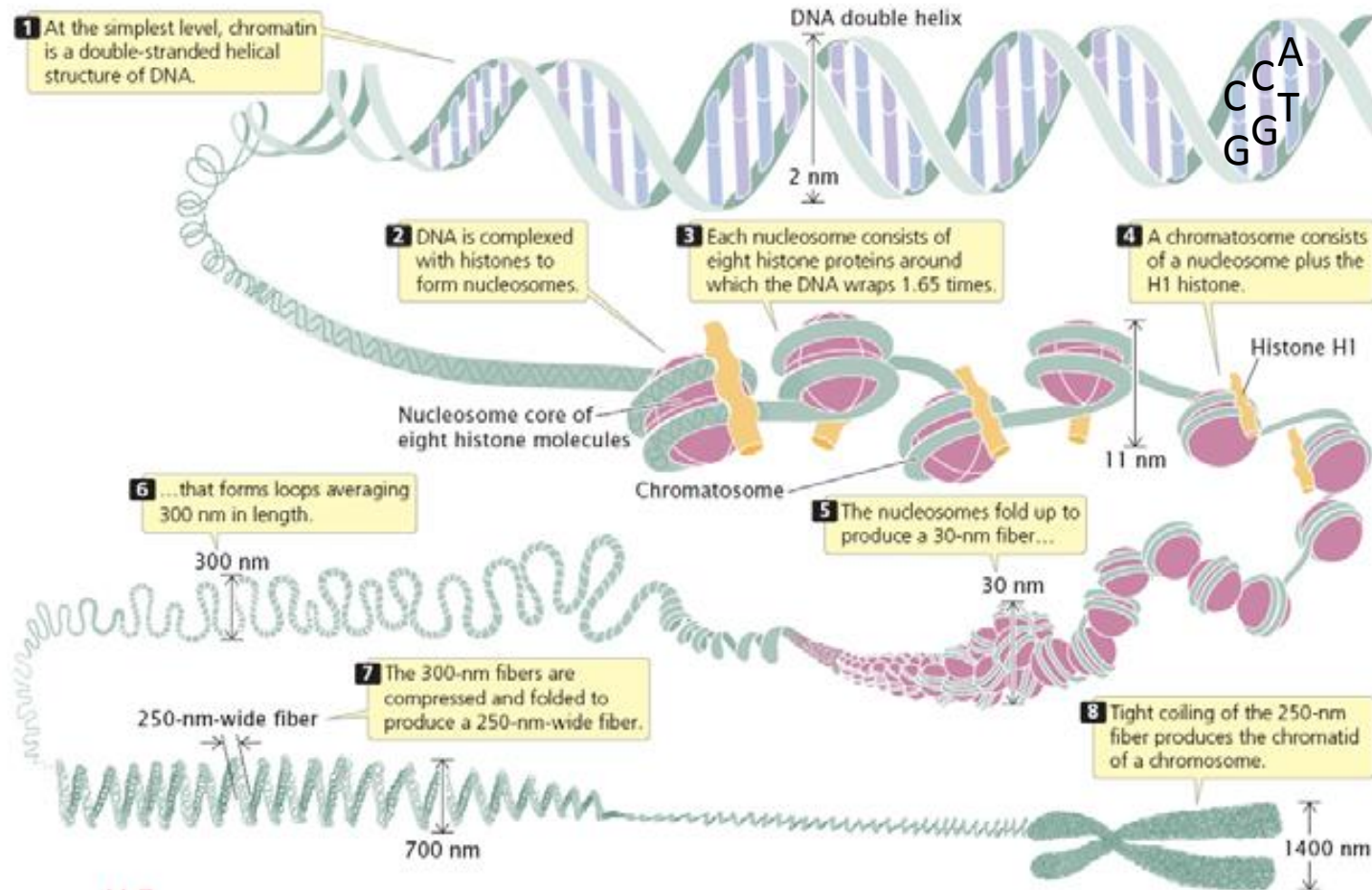


Data acquisition

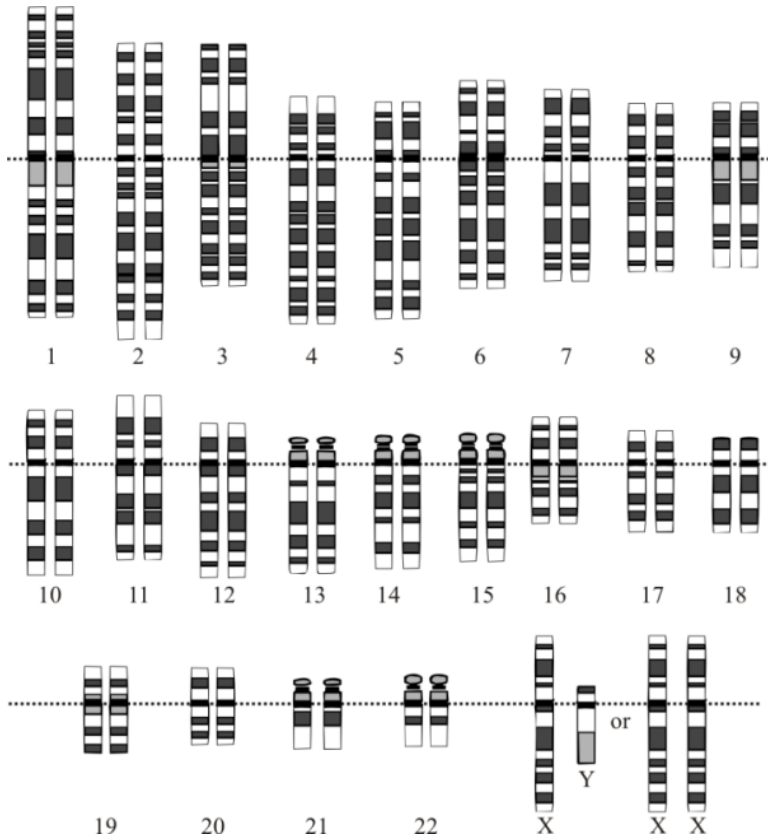


UCSC
Genome Browser

Structure of human genome

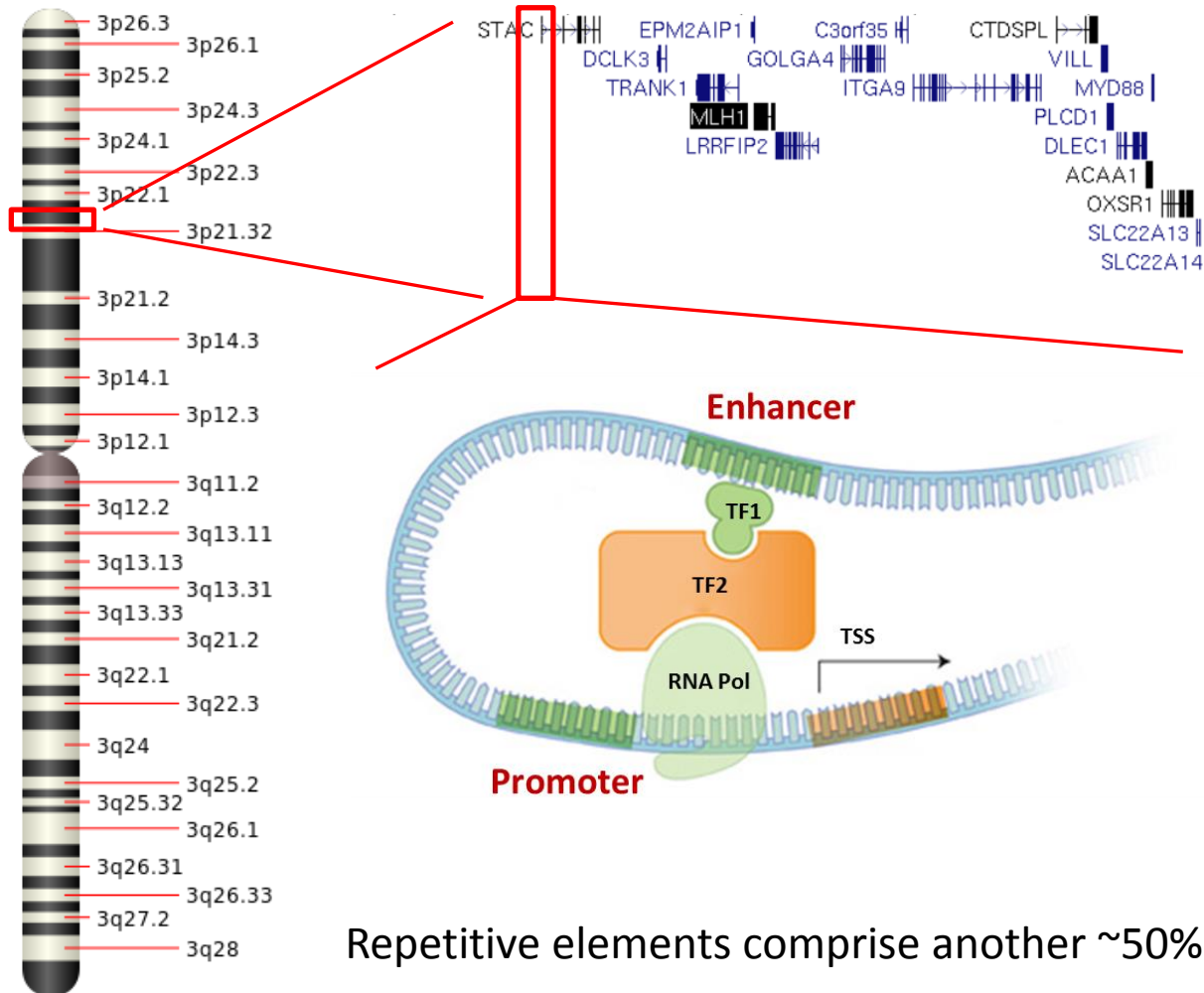


Structure of human genome



- Total of 23 pairs of chromosomes.
- Each chromosome is diploid.
- Each individual chromosome made up of double stranded DNA.
- ~3 billion bps (2m) compacted in a cell (15 μm)

Information in the genome



Genes:

~1.2% coding

~2% non-coding

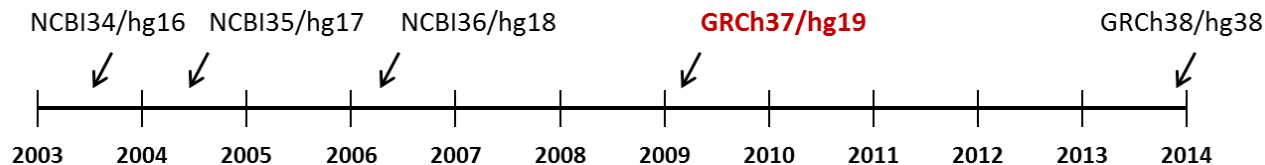
Regulatory regions:

~2%

Repetitive elements comprise another ~50% of the human genome

Reference human genome

- Human genomes vary significantly between individuals (~0.1%)
- Important things to note about the reference genome:
 - Is a composite sequence (i.e. does not correspond to anyone's genome)
 - Is haploid (i.e. only 1 sequence)
- Computationally, a reference genome is used.









Reference human genome

- Genomic data is most common represented in two ways:

1. Sequence data – fasta format (.fa or .fasta)

```
>chr1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ACAGTACTGGCGGATTATAGGGAAACACCCGGAGCATATGCTGTTTGGTC
TCAGtagactcctaataatgggattcctgggtttaaagtaaaaaataaa
tatgtttaatttgtgaactgattaccatcagaattgtactgttctgtatc
ccaccagcaatgtctaggaatgcctgtttctccaaaagtgtttactttt
....
```

2. Location data – bed format (.bed)

chr1	934343	935552	HES4	0	-
chr1	948846	949919	ISG15	0	+
...					
					
chromosome	start	end	name	score	strand

What we will cover

Structure of the human genome



Genomic information

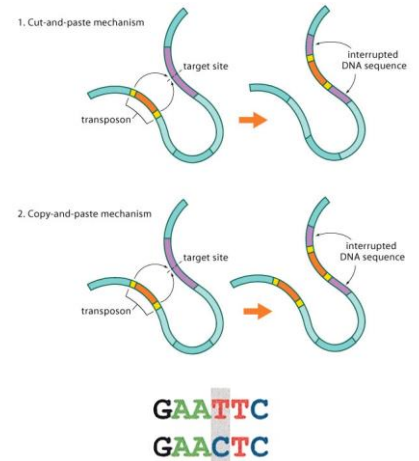
- DNA (Sequence variation)
- RNA (Genes & gene expression)
- Regulation\Epigenetics
 - DNA methylation
 - Histone modification
 - Transcription factor binding

DNA: Sequence variation



Variations in DNA sequence

- **Cytological level:**
 - Entire chromosome (e.g. chromosome numbers)
 - Partial chromosome (e.g. segmental duplications, rearrangements, and deletions)
- **Sub-chromosomal level:**
 - Transposable elements
 - Short Deletions/Insertions, Tandem repeats
- **Sequence level:**
 - Single Nucleotide Polymorphisms (SNPs)
 - Small Nucleotide Insertions and Deletions (Indels)

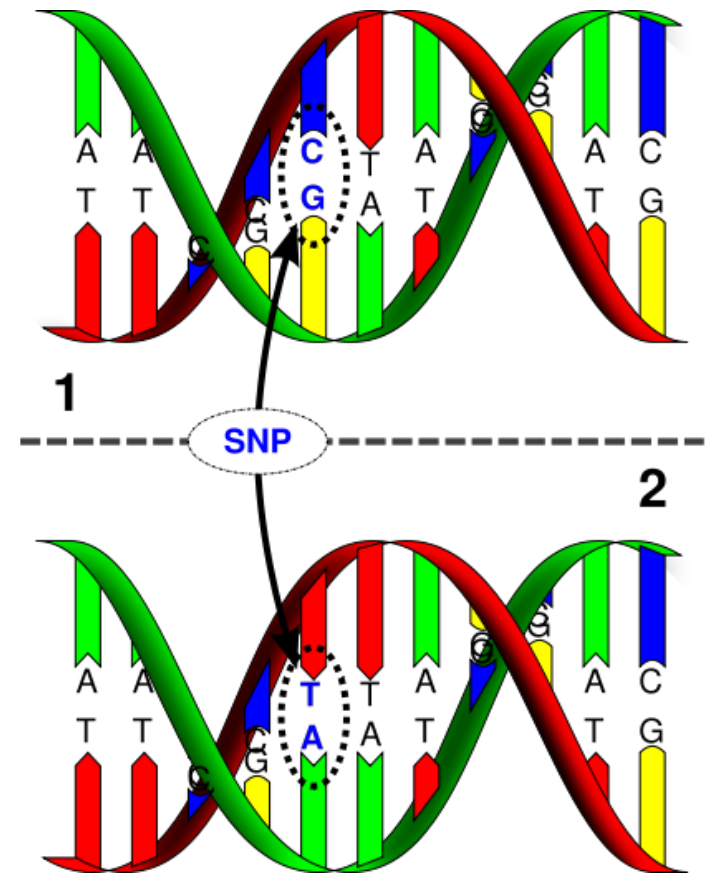


GAAATTC
GAACTC

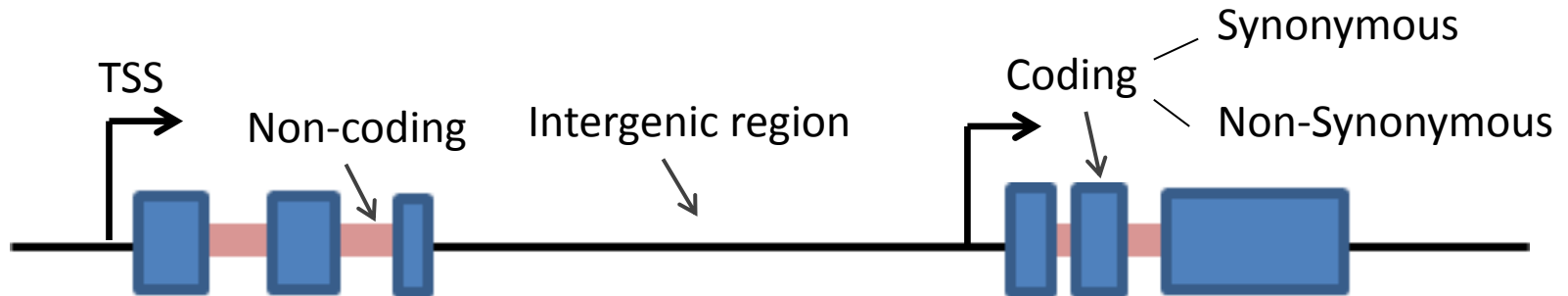
CATCGCGAATTCCCATCG
CATCG-----CATCG

Sequence variation

- Single nucleotide polymorphisms (SNPs)
 - DNA sequence variations that exist with members of a species.
 - They are inherited at birth and therefore present in all cells.
- Somatic mutations
 - Are somatic – i.e. only present in some cells
 - Mutations are often observed in cancer cells



Types of SNPs/Mutations



- Most SNPs and mutations fall in intergenic regions.
- Within genes, they can either fall in the non-coding or coding regions.
- Within coding regions, they can either not-change (synonymous) or change (**non-synonymous**) amino acids.

ATG	GAA	GCA	CGT
Met	Glu	Ala	Gly
↓			
ATG	GAC	GCA	CGT
Met	Asp	Ala	Gly

Effects of sequence variation

- Non-synonymous variants:
 - Missense (change protein structure)
 - Nonsense (truncates protein)
- Synonymous or non-coding variants:
 - Alter transcriptional/translational efficiency
 - Alter mRNA stability
 - Alter gene regulation (i.e. alter TF binding)
 - Alter RNA-regulation (i.e. affect miRNA binding)

Majority of sequence variation are neutral

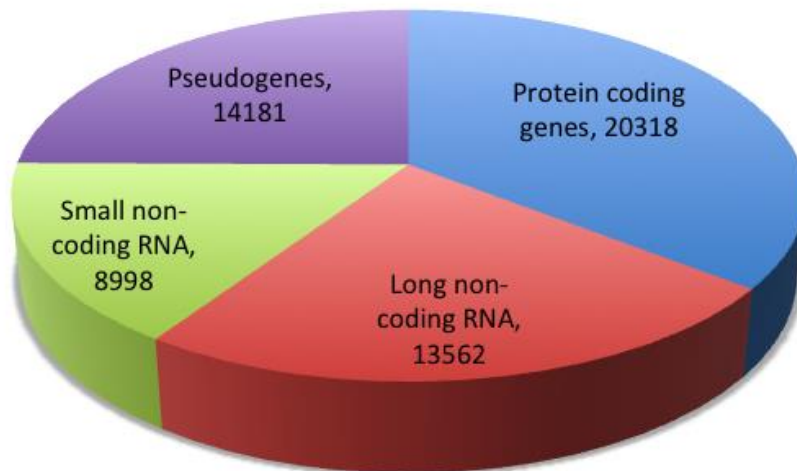


RNA: Genes and gene expression



Types of genes

- A gene is a functional unit of DNA that is transcribed into RNA.
- Total genes in the human genome – 57,445

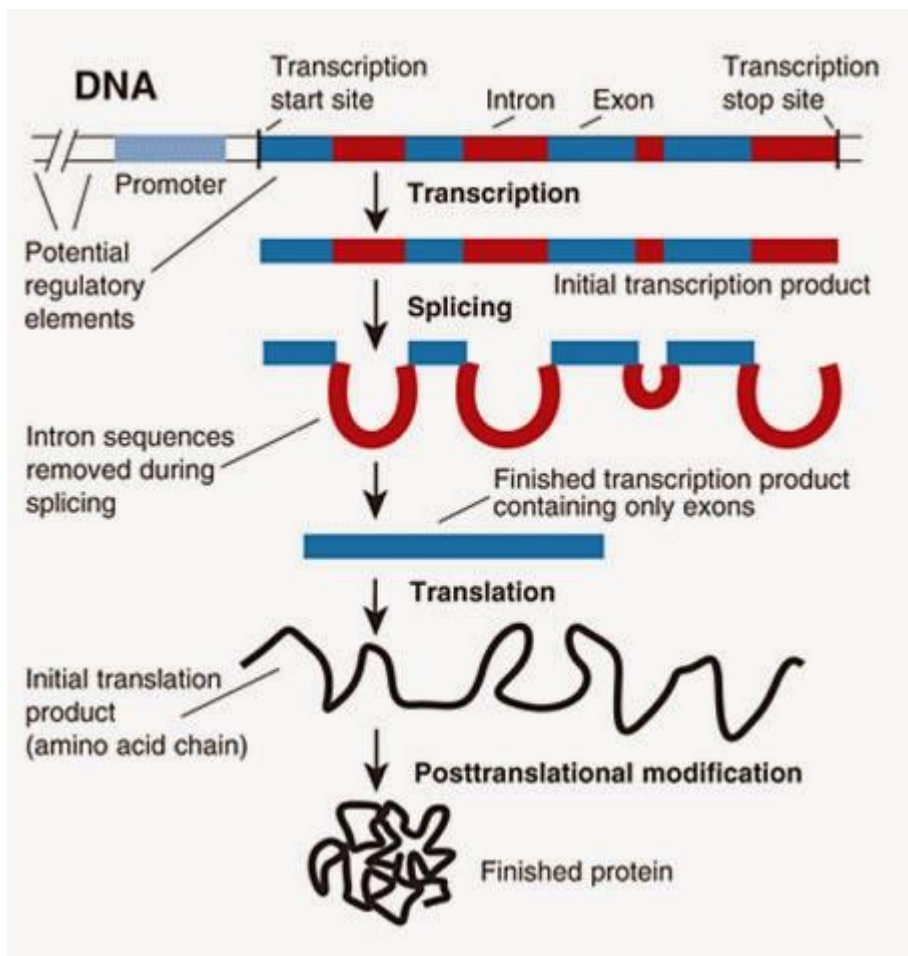


mRNA

miRNA

lncRNA

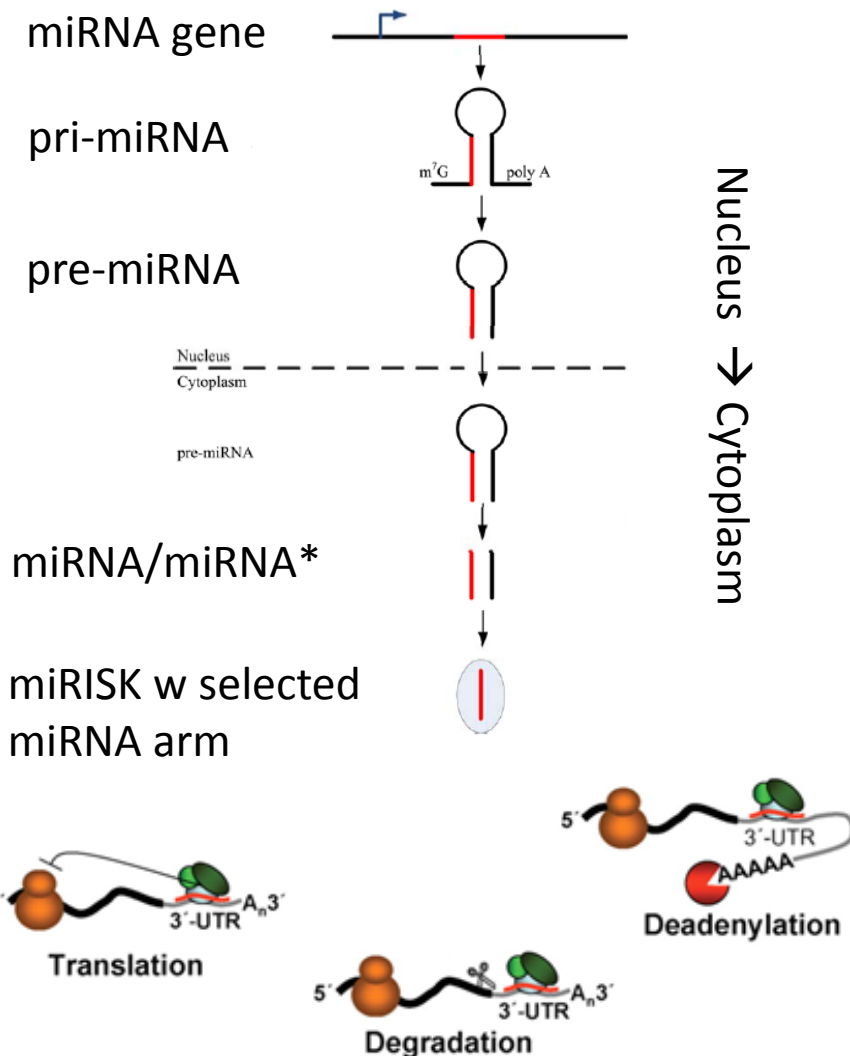
Protein coding genes



- Traditionally considered to be the most important functional unit of genomes.
- ~ 20,000 in the human genome.
- Due to splicing one gene can make many proteins.

Source: <http://www.news-medical.net>

MicroRNA (miRNA)



- Discovered in 1993.
- Plays a role in post-transcriptional regulation.
- Acts by either causing RNA degradation or inhibition of translation.
- Implicated in many aspects of health and disease including:
 - Development
 - Cancer
 - Heart disease

Long non-coding RNA (lncRNA)

- Recently described class of RNAs which often transcribed by PolIII promoters and often spliced.
- Unlike coding and miRNAs, lncRNA are less conserve.
- Non-coding transcripts > 200 nt in length.
- Many functions. Commonly recruitment of histone modifiers

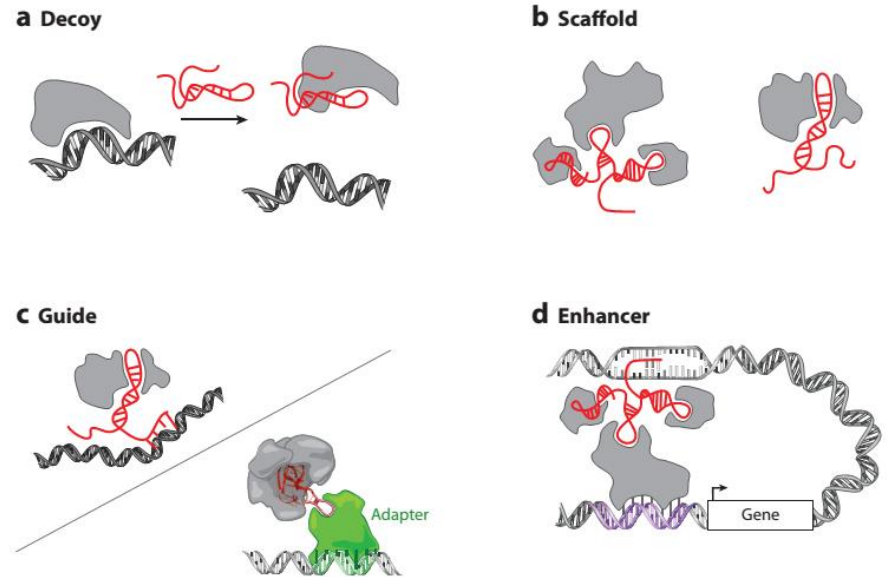
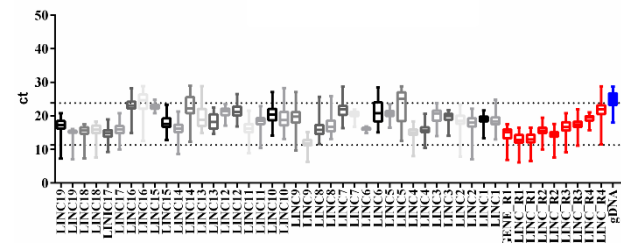
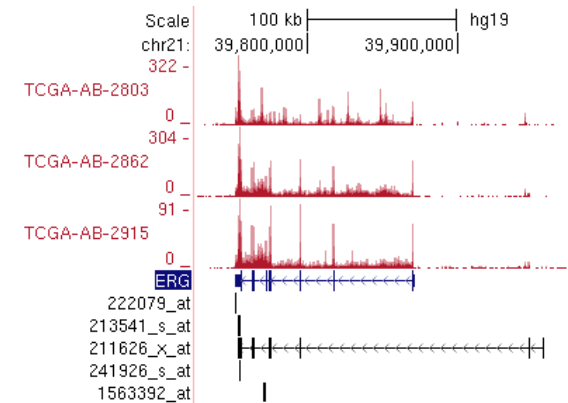
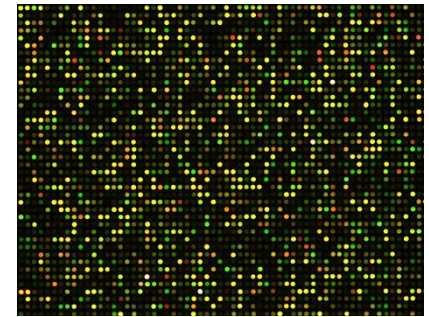


Figure 4

Models of long noncoding RNA (lncRNA) mechanisms of action. (a) The lncRNAs can act as decoys that titrate away DNA-binding proteins, such as transcription factors. (b) These lncRNAs may act as scaffolds to bring two or more proteins into a complex or spatial proximity and (c) may also act as guides to recruit proteins, such as chromatin modification enzymes, to DNA; this may occur through RNA-DNA interactions or through RNA interaction with a DNA-binding protein. (d) Such lncRNA guidance can also be exerted through chromosome looping in an enhancer-like model, where looping defines the *cis* nature and spread of the lncRNA effect.

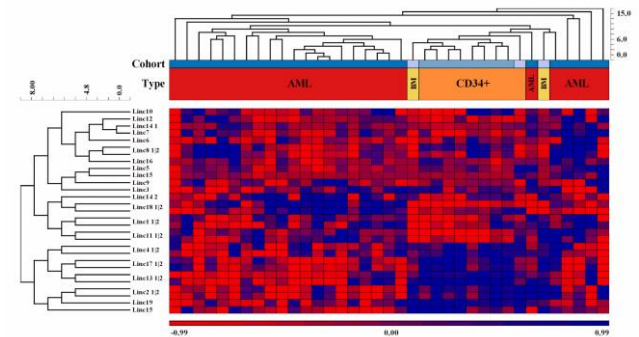
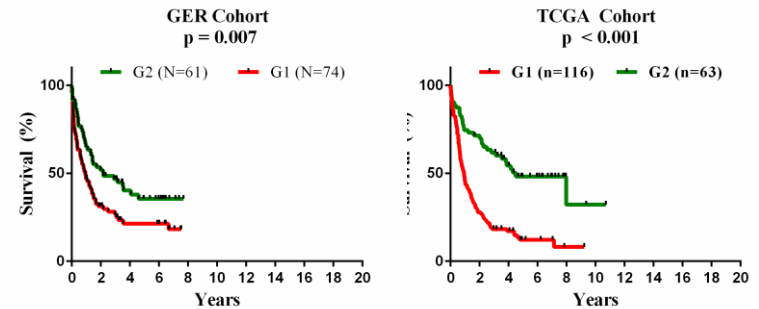
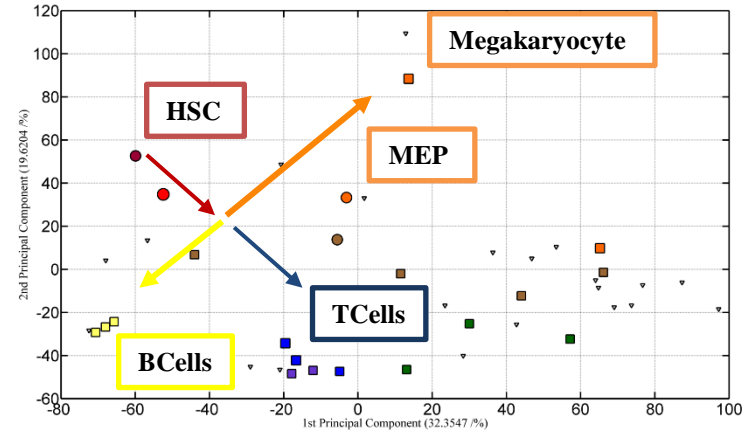
RNA expression

- Measuring the level of RNA in the sample.
- Generally microarray-, sequencing- or high-throughput PCR- based.
- Computation analysis and normalisation of expression data can be complicated.



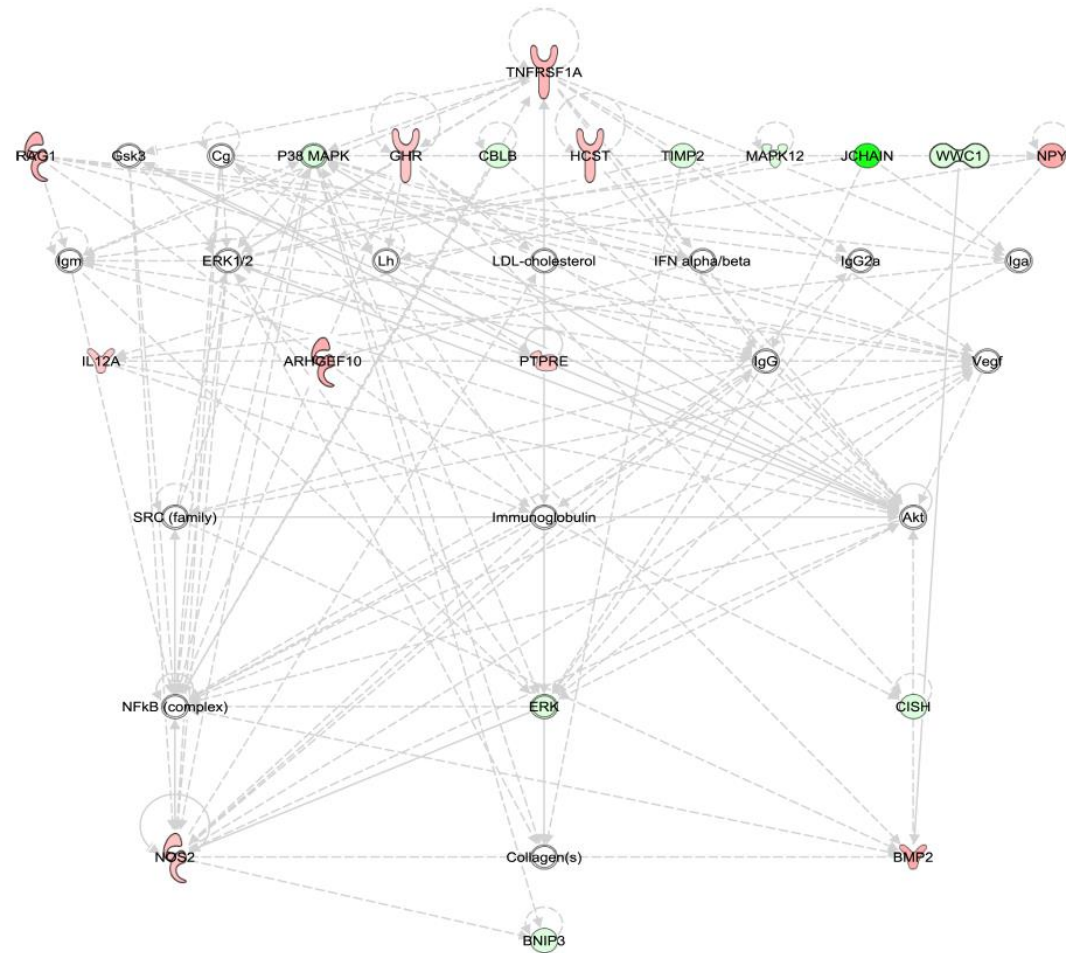
RNA expression applications

- Relatively cheap and fast readout of the functional state of a cell
- Association with clinical features
 - sequence variations
 - response to therapy
 - patient survival
 - ...
- Differential expression
 - between samples, or
 - between genes



RNA expression applications

- Differential expression of individual genes not necessarily informative.
- Genes are often grouped in gene-sets based on ontology or biological pathways.



Gene Regulation

Epigenetics



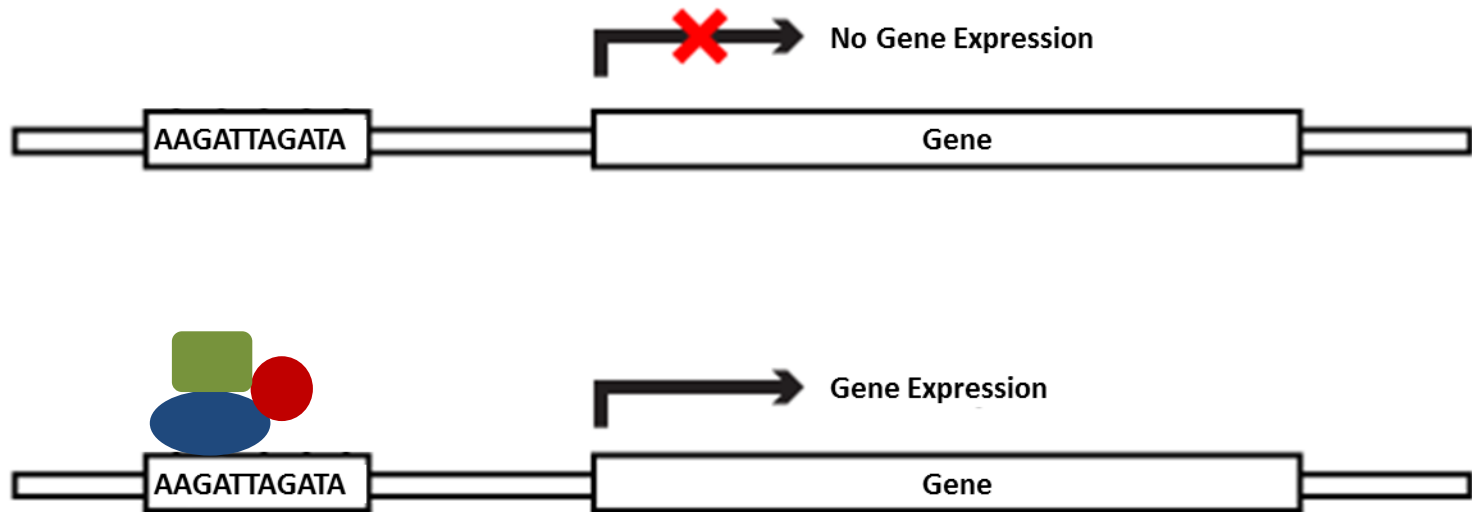
Epigenetics

- Mechanisms that alter cellular function **independent** to any changes in DNA sequence
- Mechanisms include:
 - Transcriptional regulation: Transcription Factors
 - Genome methylation
 - Histone modification / Nucleosome positioning
 - *Non-coding RNA*



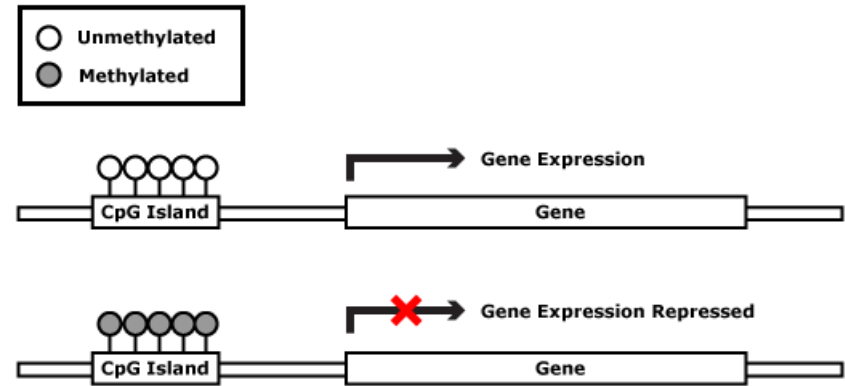
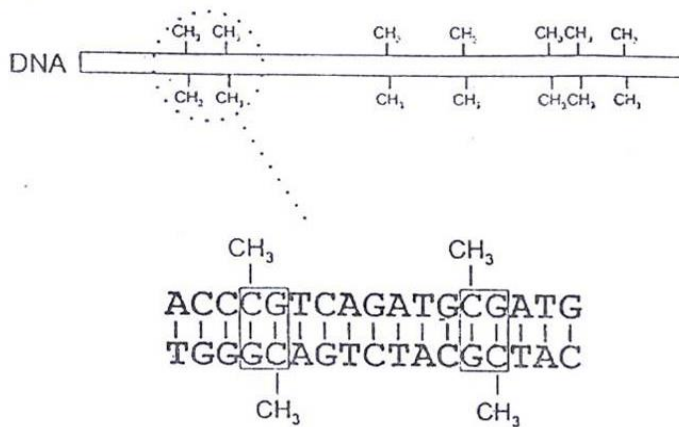
Transcriptional regulation

- Transcription factors are proteins that bind DNA to co-regulate gene expression.
- Typically binds at gene promoters or enhancers.

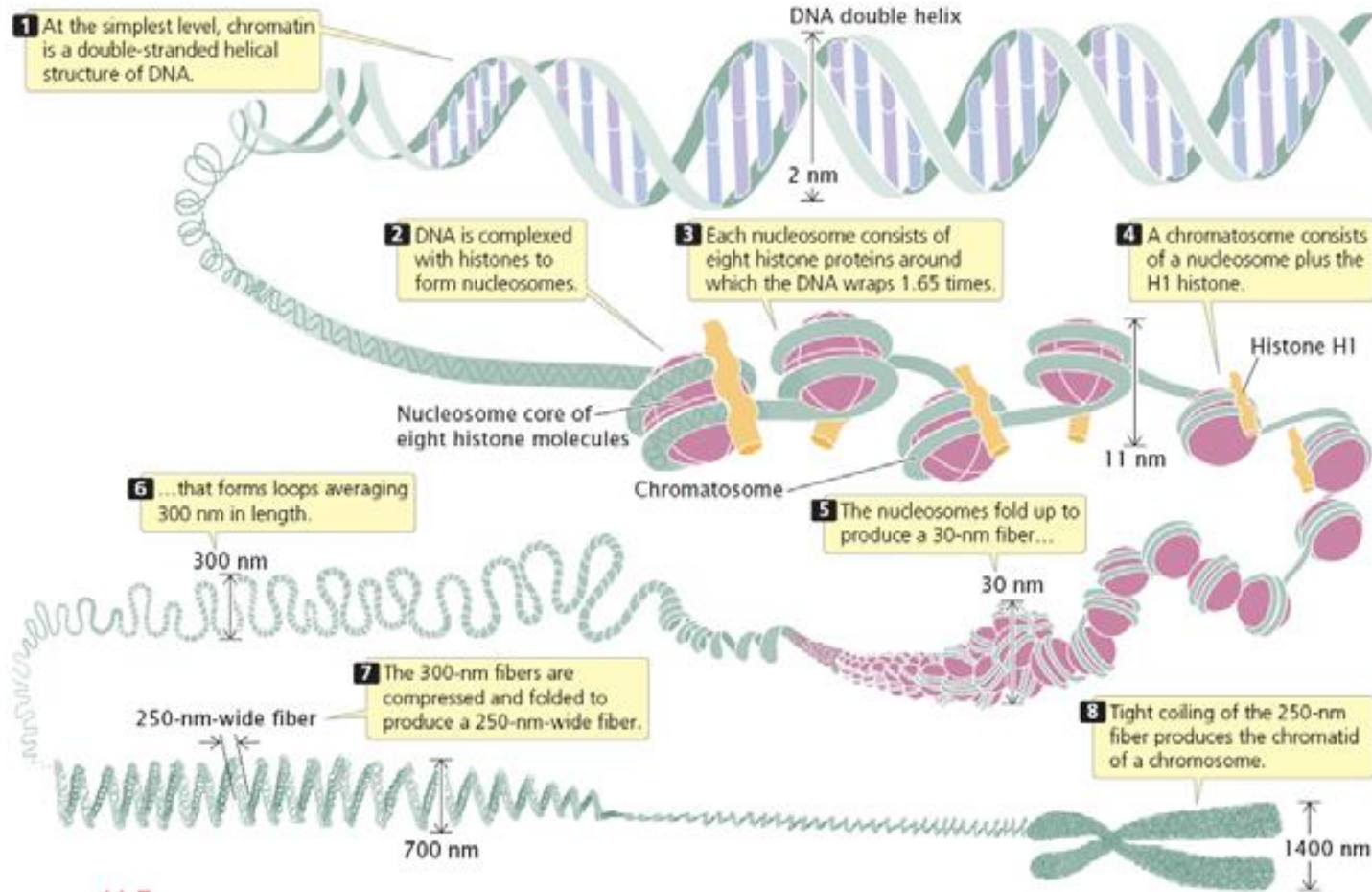


DNA methylation

- DNA is methylated on cytosine's in CpG dinucleotides



Nucleosomes & Histones



Histone acetylase
NRC: nucleosome
HAT: histone acetylase
IBP: insulator binding protein



nucleosome
condensation

What we will cover

Structure of the human genome



Genomic information

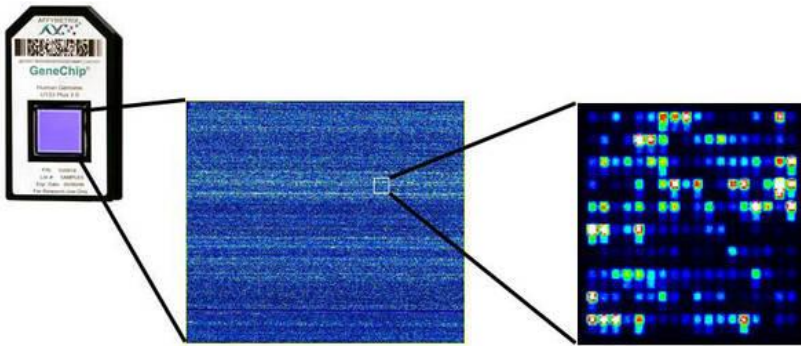
- DNA (Sequence variation)
- RNA (Genes & gene expression)
- Epigenetics
 - DNA methylation
 - Histone modification
 - Transcription factor binding



Data acquisition

- Microarrays
- Sequencing
- Chromatin IP

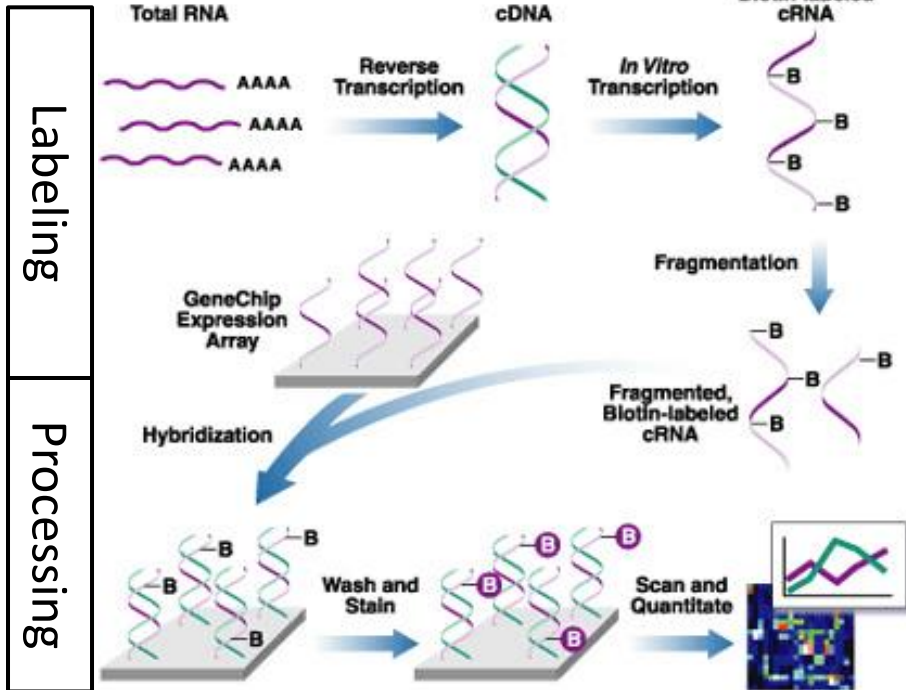
Array Technology



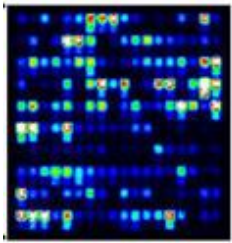
- Relies on fluorescence-based hybridisation of DNA against complementary probe on array.

- Known molecule that can be converted to cDNA.
 - Expression array (probe for exonic DNA regions)
 - SNP array (probe for two alleles)
 - Methylation array (probe for bisulfide converted DNA)

- Limited by probes present on the array.



Array Technology



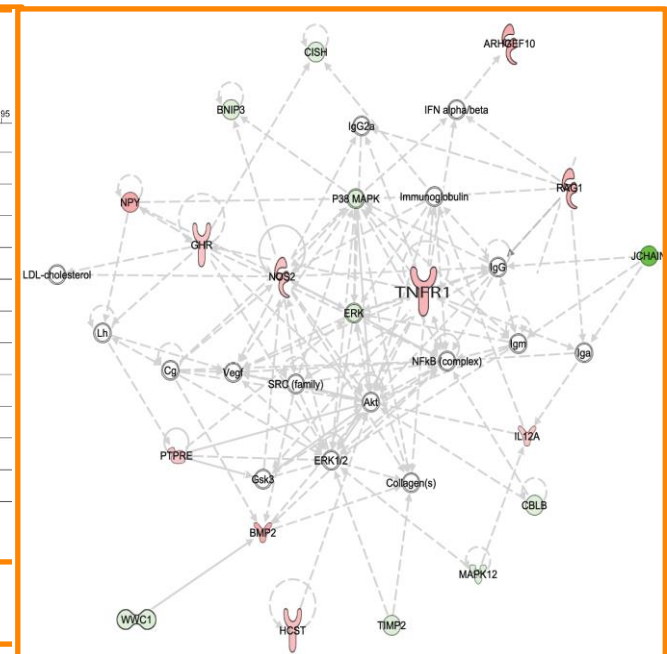
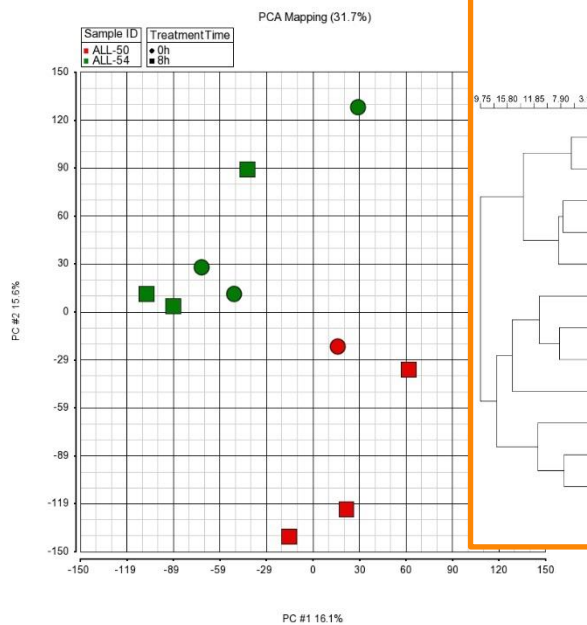
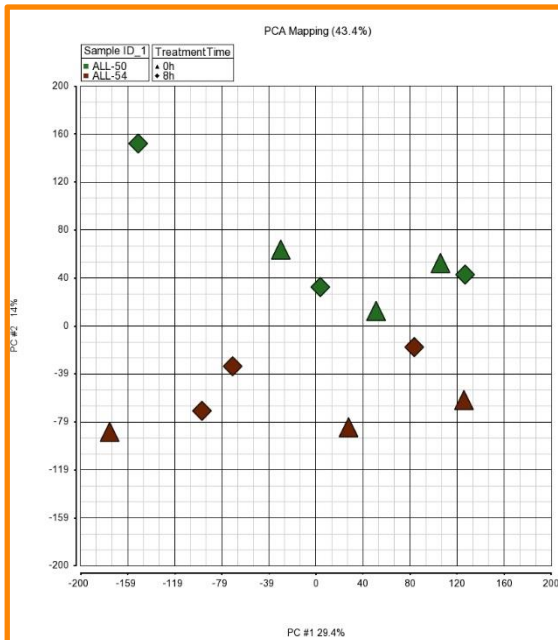
Images Processing

Quantification

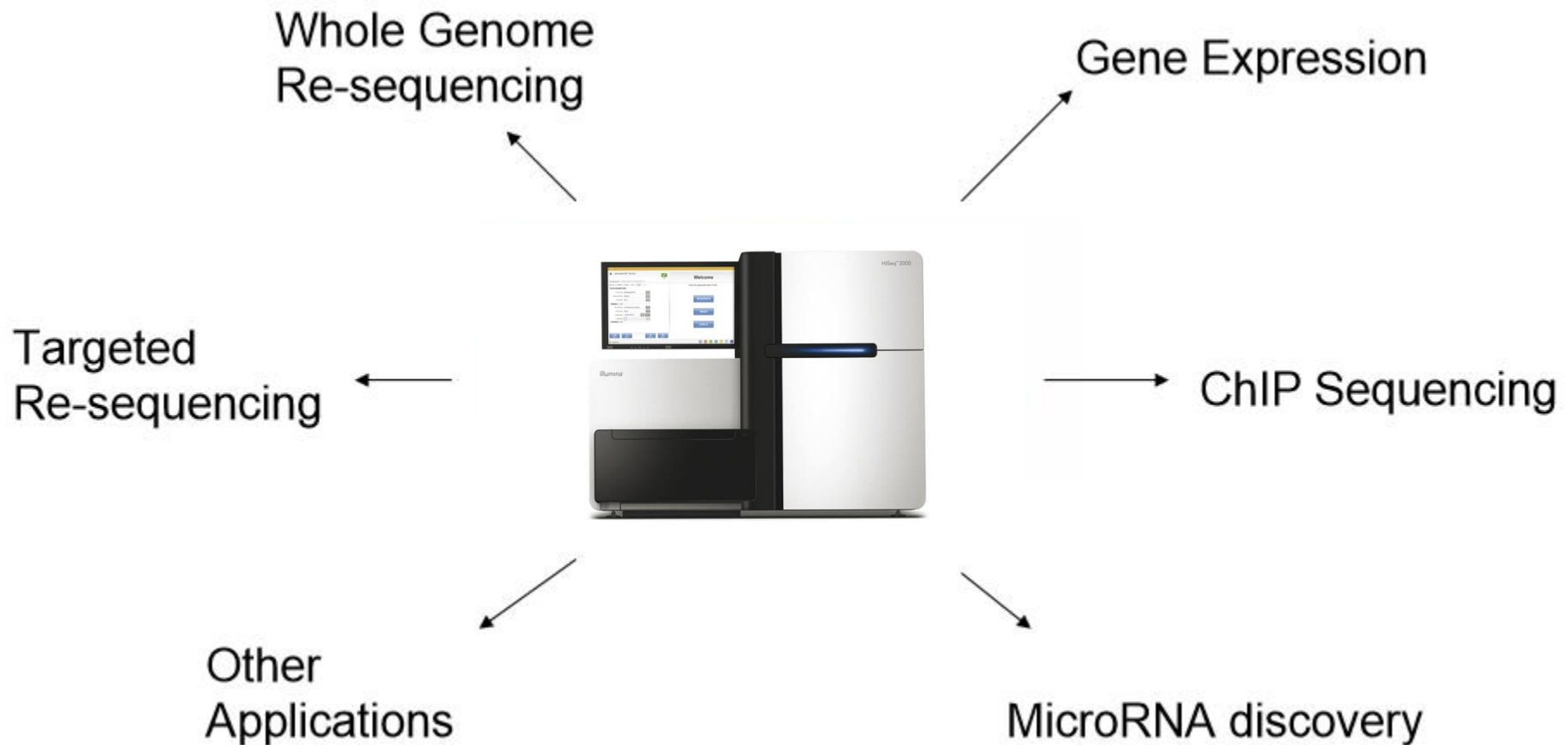


Pre-processing

Backgrd. Subs., Norm.



Next-generation sequencing



Next-generation sequencing (Illumina)



Next-generation sequencing (Illumina)

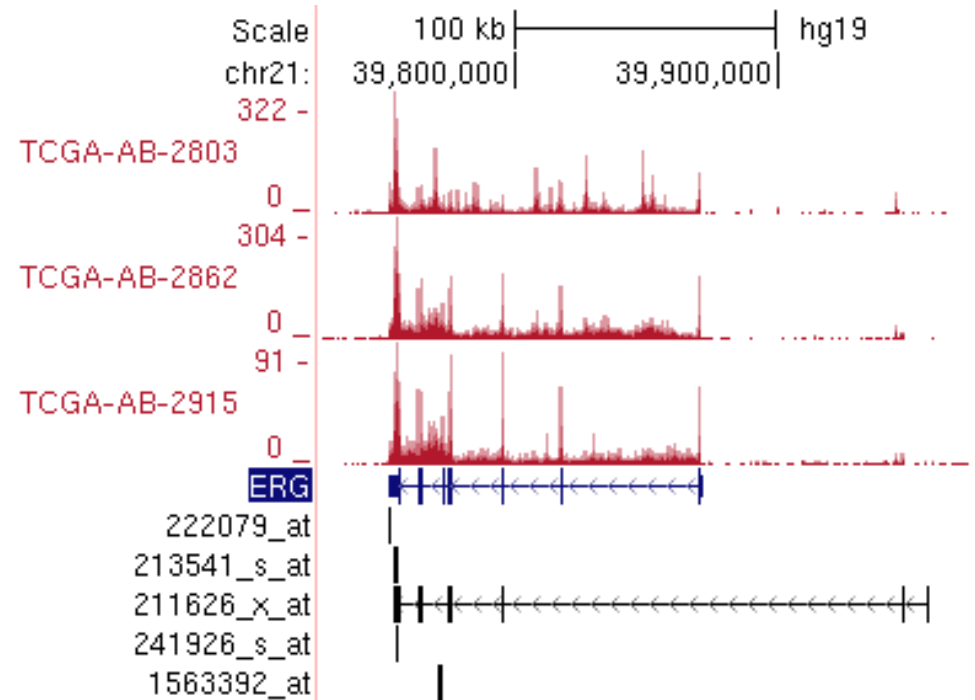
Alignment
human reference genome



Quantification
mRNA/miRNA/lncRNA



Quantification
mRNA/miRNA/lncRNA



Pros/cons of each technology

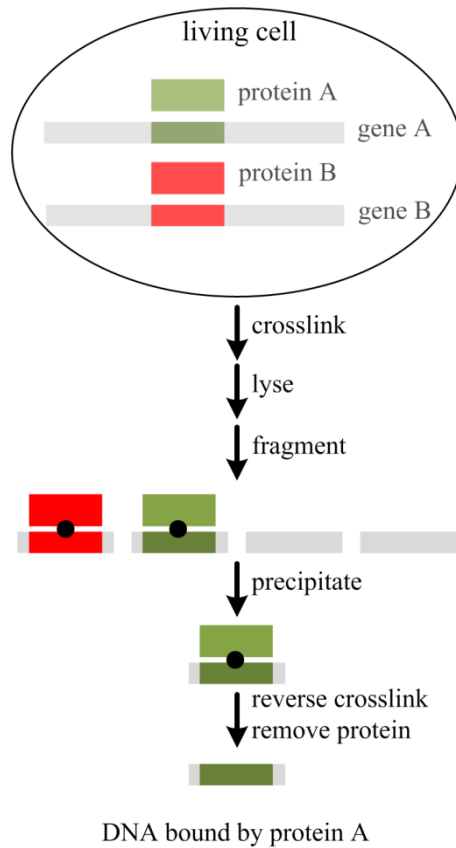
- NGS
 - Greater dynamic range (only limited by depth of sequencing)
 - Coverage of genome does not need to be limited.
 - Many more applications from sequencing data.
 - Data analysis and management can be challenging.
- Microarrays
 - Microarrays are still significantly cheaper.
 - Largest public datasets are likely to be microarray based.
 - Data analysis pipelines are well standardised.



Chromatin Immunoprecipitation Sequencing (ChIP-seq)

ChIP-seq of the seven transcription factors

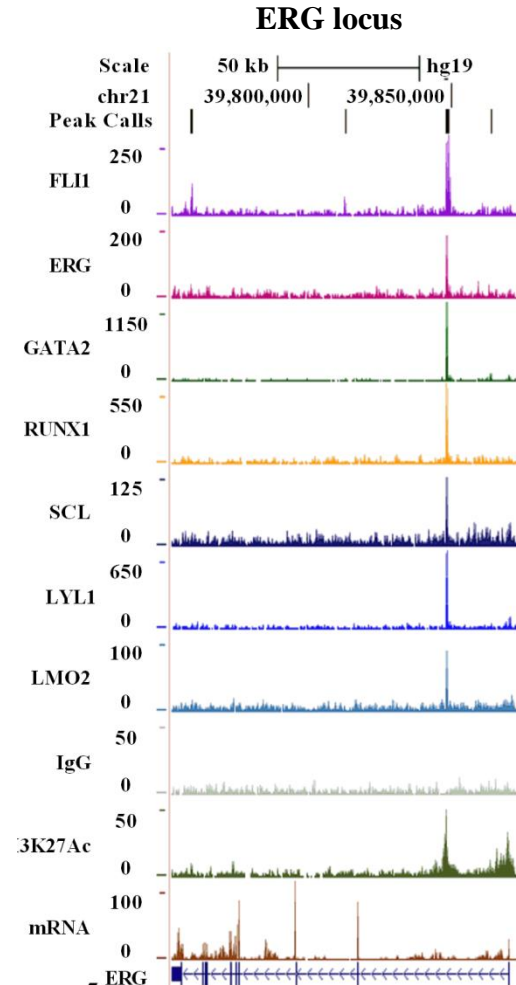
FLI1, ERG, GATA2, RUNX1, SCL, LYL1 and LMO



High-throughput sequencing



Bioinformatics



What we will cover

Structure of the
human genome



Genomic information



Data acquisition



UCSC
Genome Browser

- Background
- Genome Assemblies
- Annotation Tracks
- Associated Tools
- Practical Exercise



Genome Browser

<http://genome.ucsc.edu/>

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

Genome Browser
ENCODE
Neandertal
Blat
Table Browser
Gene Sorter
In Silico PCR
Genome Graphs
Galaxy
VisiGene
Utilities
Downloads
Release Log
Custom Tracks
Cancer Browser
Microbial Genomes
Mirrors
Training
Credits
Publications
Cite Us
Licenses
Jobs
Staff
Contact Us

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.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work 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 data 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 Genome Sciences and Policy. 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

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [Genome Browser News](#) mailing list.

24 October 2013 - Job Opening: UCSC Genome Browser Trainer

The [Center for Biomolecular Science and Engineering](#) (CBSE) at University of California Santa Cruz seeks an articulate, self-motivated individual for in-person training on the UCSC Genome Browser at universities, hospitals, institutes, and professional meetings in the United States and international locations. The position requires a Master's degree in a biological science, depth in molecular biology, experience in a research environment, working knowledge of teaching or training in a scientific environment. Preferred qualifications include a PhD in a relevant field, experience with video production, and a minimum of 2 years of Genome Browser experience ranging from novice users to bioinformatics specialists. Presentations include formal talks, problem-solving sessions, and hands-on training.

This position requires a Master's degree in a biological science, depth in molecular biology, experience in a research environment, working knowledge of teaching or training in a scientific environment. Preferred qualifications include a PhD in a relevant field, experience with video production, and a minimum of 2 years of Genome Browser experience ranging from novice users to bioinformatics specialists. Presentations include formal talks, problem-solving sessions, and hands-on training.

For more information and to apply for this position, see [Job #1304619](#) on the UCSC Staff Employment website.

23 October 2013 - dbSNP Build 138 Available for hg19

We are pleased to announce the release of four tracks derived from NCBI [dbSNP](#) Build 138 data, available on the human assembly (GRCh37/hg19). The tracks include corresponding coloring and filtering options in the Genome Browser.

As was the case for the annotations based on the previous dbSNP build 137, there are four tracks in this release. One is a track containing all variants and three are subsets of this track and show interesting and easily defined subsets of dbSNP:

- Common SNPs (138): uniquely mapped variants that appear in at least 1% of the population or are 100% non-reference
- Flagged SNPs (138): uniquely mapped variants, excluding Common SNPs, that have been flagged by dbSNP as "clinically associated"
- Mult. SNPs (138): variants that have been mapped to more than one genomic location

By default, only the Common SNPs (138) are visible; other tracks must be made visible using the track controls.

You will find the four SNPs (138) tracks on the Human Feb. 2009 (GRCh37/hg19) browser in the "Variation and Repeats" group.

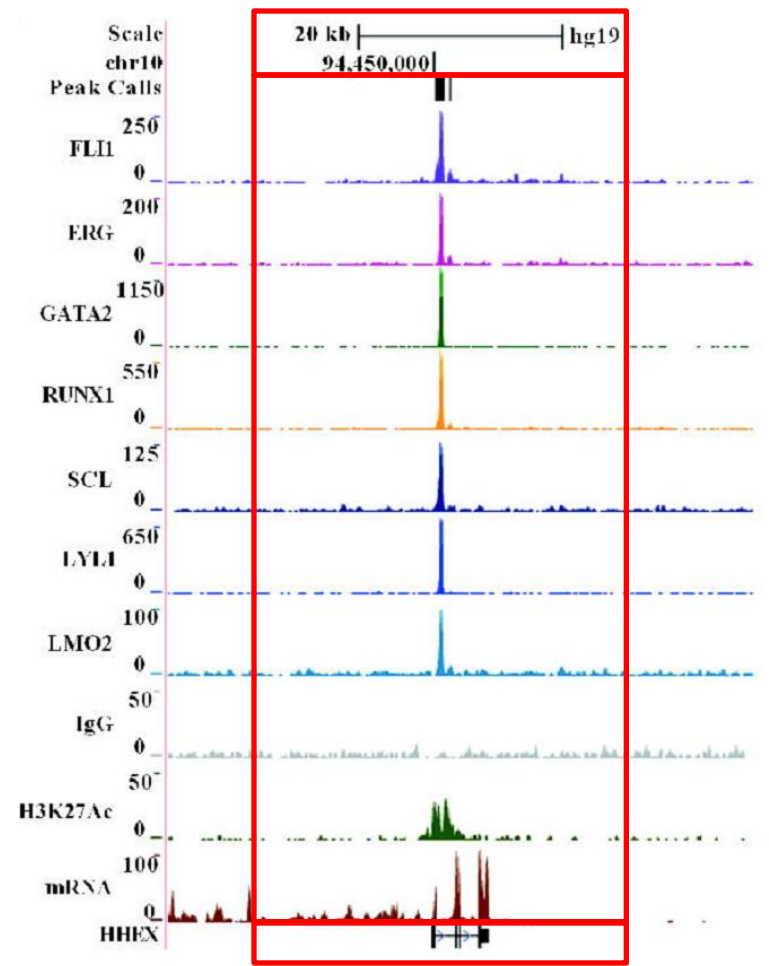
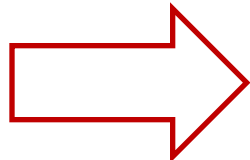
The tracks were produced at UCSC by Angie Hinrichs and Luvina Guruvadoo. We'd like to thank the dbSNP group at NCBI for providing access to the data.



Background

genome.ucsc.edu/cgi-bin/hg

```
TTTAAACATTTTTTTGTDRHRSHATATGTGTTTTTTCTCAGTAAATGRTTTTTAAATTTA  
CAAGTTAGAAAGGTTTATCTAGTTGTTTCTATATAGCATTAATCTGGGCTGCTTTGTCAA  
AAATTTATTTATTTTGAATCAGGGTGTCACTCTGTACCACAGGCTAGTGGCGTATCA  
CAGCTCACTGCAGCCTTGAACCTCTGGGACTGCAGTTGTGCCACCACCTGCCGGCTTT  
TGTATTTATTTATTTTTTGTAGAGACTGGGTACTCTCATGTGCCACAGGCTGGTCCCG  
GCTCAAGCGATCCACCAGCTCTGTCTTCAAAGTGTGGGATACAGGTTGAGTCAAC  
AAGTTCCTTTAAACCAAGGTTAAAGATAAAAGAAAGGACAGAGATTTGGTGTATTTTT  
GAGTGTCCAAAGTAAATTTTATCTCTAAATGTGCACCTACAGCTTATTTGAATGAA  
TTAATGAGAAGCTAAATAAACTAGCTTTTAAAAAATATTTCTCTCTCTCCAGGCA  
ATGGGTCACTTTGGGGGCTGCAGAACATCTGTGCTTTCCCTGTGCCCATCACTCTCC  
AGCTCTGAGCTGTAGCCATCTCTGCTGCACCTGTCTCAGAGGGCAGCTCTCCGGCTC  
CTGCTCTGGGTTGCAGGTGCCACATCTGTGCAGAAAGCTTAAAGGAAGCAGTGAAGCTT  
CTAACCTGCAGCAGCTTGCACAAAGCCTTTGTGAACCTCCGGACATAAGGGCATCAA  
ACTAGGAAACAAAACACAGTCTTGACTGTGATTTGGTAAATTTACACATAACAGGCGC  
AAAGCCAAITGAAGCAAGAAAAGAACAAATCTTGTCCAGCAGGAGTAAGATTTGTAC  
AGGAGTCTGAGTATACATCAGGGCAAGCCAGAGACAGGGCAACAGAGAAAAGGTT  
CATGCAATTTGAATATGACCTGGAGGGGCTTTGGAAATGAGTATCAACAGTAAACAG  
AGTTAATGCCATTTGAACAATTTAAGAATATGATTTGAAGAACTCAGTATATCA  
TGTTATCCAAAATTTCCAGTTGCTCATAGACTTGTCTAAATCCAGCTGGAGGGGGTGG  
AGGGTTGAGGAGCTAAATGGAAAACATCTGTTAAAATGAAATGAAATCATGCTCTGT  
TTTTATGATCTGTCTCATGAGACAGTCTGAAAGAGAGACTGAGTAAATTCAGATAA  
TGTGATTTTCTAATGAAGATGAGGTGTCTTGACTTTGCATTTCCAAAATAAAATGAC  
TGACTTTTTCCATTAGCACAGTTTGGTCAGCCTAACTGCCAGCTTCTCATGTTGATTT  
ATTTATTTTCCCTAAGGAGATACGGGCACTTTGTGGGCTTCCAGGCTGTGCAAAAATG  
ATGGATGGTGGGCTCGTCTAATCAAAATGACCGGGGTTCAGGAGCATTTGGTAATCG  
TGCTCTGCCGACTTCAAAGCCACTGCCAAAACATCCACAGCACAGGTTTGGCAACT  
CACATAATCATAGCAAAAGGACTGCAACTGAAAACAAAGCCCTGAAACAGCTATGAA  
AAGGGCCAGTTCAGAAACCTATTCAAGCTAAGATGTTTGGCTAGCCCTTCCGACGGTCA  
CCTTGATTTTCACTTTTGAAGTGTCTTCACTGGACCAAGGTTGGCAGGATTTGTGAT  
TCAAAGAAAAGATGCCAGGGAACTGACTAGATATGGAATAAATCATGTTCTCAA  
GGATAGCCAGCAACATCAGGCTCAGGGCTAGTGAATCCCAAGCACAGTCCACAGGTA  
CTCTGATGTAGCAGGACTAAAGCTGTCTACTAGTGAAGCTCTGAAAAGAAAACCCCG  
CAGCAAACTAGACGTTATCCCTTGTCTTGTAAAGTGAAGAAATGACGCTTCCATGAC  
GCTTTATTTGCCAGTAATACAGTTAGCTTACGAGTGTAGCTTGTGACCTGTCTCAA  
CAAATGATTTTAAATCATAGCAGGAATTAAGGGTGAATTTGTGACAGGTTGCACTAGA  
ATTTCTCAATGACAAACCCAGCTTCAAGGCTCTTCCATCCAGGCAATAATGGAG  
GCTCCAATGTAAACCCGGGCTCTTGTCTCAAAGGGGCTAGAATAAAAAGCAGGA  
GGGGAGGCAAGAAAGCCTGGAGAGGCTGACGCCATTTGGGTCGCAAACTCTATTTC  
TTGGCTCTCCCTTGCACAACTTCTGGACAAAATAAATAAACAAAATCCACAACT  
TCAGCACATGTTTCAATTAAGCACTTTAGTCACTAAAATAAGTGAAGTGCAGACTCCT  
GTATAAATCTGATTTGGCATGCTAGGCCAAGCTTATTTTATACATACATTTGCAATTC  
AAGAACATAAATCTCATATTGTAACATTAAGCATACAGGTTAAATTCAGGGCCACA  
TTATATCGATTTGTCTTTTGTGCTGTCTTGGCTGGCCGAGATCAACTGTAGTGTGA  
TAAATGCATAAGTTATATAATATATAAAAAGGGGAAAACATTTGACTGTATACT  
TCATTTGCACAAACGACAGGCTTGCACCTCAAAGGAAACTGGAGGCGCCATCCCAA  
AATGCTGCGTGATTTCTGATTGTGGCAGCCAAAGGCGCATCTCTTAACTGACCCTGTG  
GAGAACAAGGCCCCACATAATGAGGCTCTGAAATGTTTCTCTAAATACAGACAAT  
TCCAGTTAAATTTTCAATTTGACAAAACAAAAGAAAGATGCGCATTTTGTCTGAAAT  
CTACTACTTCCCTTCTCATTAGCTGTGTCTTCTCTTAACTGGGTTGGGATA  
ACACTGACTGATGAACCTCGAGTCTCATAACTCAITGTACACAGTCCCAAAATGTCC  
TGAGTCCAGCTTCTATTGCTTGAAGAGTCTTCTCCAGGCTGTGCTCTCTCTCTG  
CCTCTTCTCTCTCTCAAGCTCTGTCTCAATCACAGCTCACTCTATACACACT  
CTCAGTCCCACTTTTGTGTTATGATGCTCCGGTAACTGTAAAGGAGTTGAAAATTTGG  
GTCACTTCAAGTGTGAAAGCTGACAGCTGTCAATCAACGAAATGATTTGATTTCC  
AACAAAACAGCACATGCCATGAGTTGCATATCAAGCTGTGTTGATGGGCCACAGCTCT  
TCTGCTGCTTTTCTCTGTTTGTATGATGTTCTATTTTAAATACAGGAGTTTTCT  
TAAATGGCATATAAACGTTATGTTAGGTCAAAACGTGTTGTTCTATTTGCTTGTGTT  
TCACTCCAGTTGCAATAGTGGGATCCAAACTCACTCAAAAGTTATACATTTCTTAA  
GACCACTTTCTTGGCACTTTGTCTTAAAGCTTCAAGCTTCACTACATACACTGCTTTTA  
CAGGTCAGCTCAGGATGATGATGCTCAATTTGTGACATTTTGAATCTCTCTCATTT  
CTGTAGTAAGACTTCAGTAACCTCCCTCAAAGGACTTTGGATCTTCCCGGCTCTCT  
TCCACAGCCAGTAAGACTTTTTTCTTCTCTTGAAGGACTTTTGTTCATGTTCTCC  
GATAGATTTTGTGGCAGTGGCTGGTGAATGCACGCTGATGGGAAAAGCCTCCGCGAGT  
CTTTAGTAAAGTGCACAGATGAGAAGGCATAGGCTGGTGGGAGCCATGTTTGGGG  
TATATACCCCAAGTTGGTGAATCCAGTATGGGTTGGGCGCAAAAACATGGAAGAT  
GTACGGGGAAGGCTGGAGGGTGGGGGCGCACAAAGTTCATCTCTGGGTTGGGCGTA  
TAGGAGCCATGTACGGGAGCTGTAGGGGACTTGTACAGAGATGACTCCGGGGGGTGG  
GGCTGGAGGGCTGGGCGATCCGTTGGAAGTGCACATTTGAGGCGTAGCCCTCCATGG  
ACCTTGGCTATGATGTTCTTGTCTATAGTAAAGGAGGGGCGCGGTGAGCTTATGTAG
```



Background

Visualization of genomic data

- ❑ Graphical viewpoint on the very large amount of genomic sequence produced by the Human Genome Project.

Human Genome: 3,156,105,057 bp

- ❑ Focus turned from accumulating and assembling sequences to identifying and mapping functional landmarks

Genetic markers

Genes

SNPs

Points of regulation

- ❑ Visualization of Next-generation-sequencing data



Background

Client-side

Integrative Genomics Viewer*

- ❑ Application (Java) on the user's machine
- ❑ Often difficult to install
- ❑ Does not have the extensive third-party data of the other browsers
- ❑ Much faster than web-based browsers



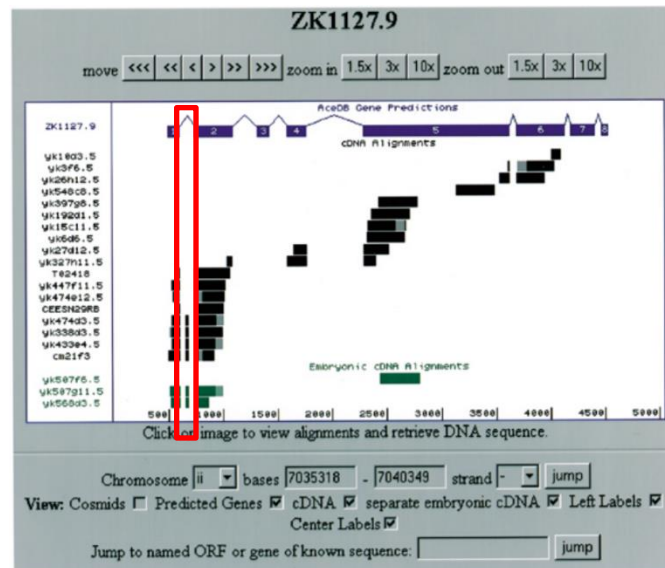
The screenshot shows the homepage of the Integrative Genomics Viewer (IGV) website. The browser address bar displays "www.broadinstitute.org/igv/". The page features a navigation menu on the left with links for Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below the menu is a search box and the Broad Institute logo. The main content area includes a large banner for "Integrative Genomics Viewer" with a background image of genomic tracks. Below the banner are sections for "What's New" (announcing an October 2013 workshop and a new release), "Citing IGV" (providing citation information), "Overview" (describing IGV as a high-performance visualization tool), "Downloads" (with a registration link), and "Funding" (listing funding sources like the National Cancer Institute and GenomeSpace). The footer contains logos for the National Cancer Institute, National Human Genome Research Institute, and GenomeSpace.

Background

- Intronator was developed by J. Kent to map the exon–intron structure of *C. elegans* RNAs mapped against genomic coordinates

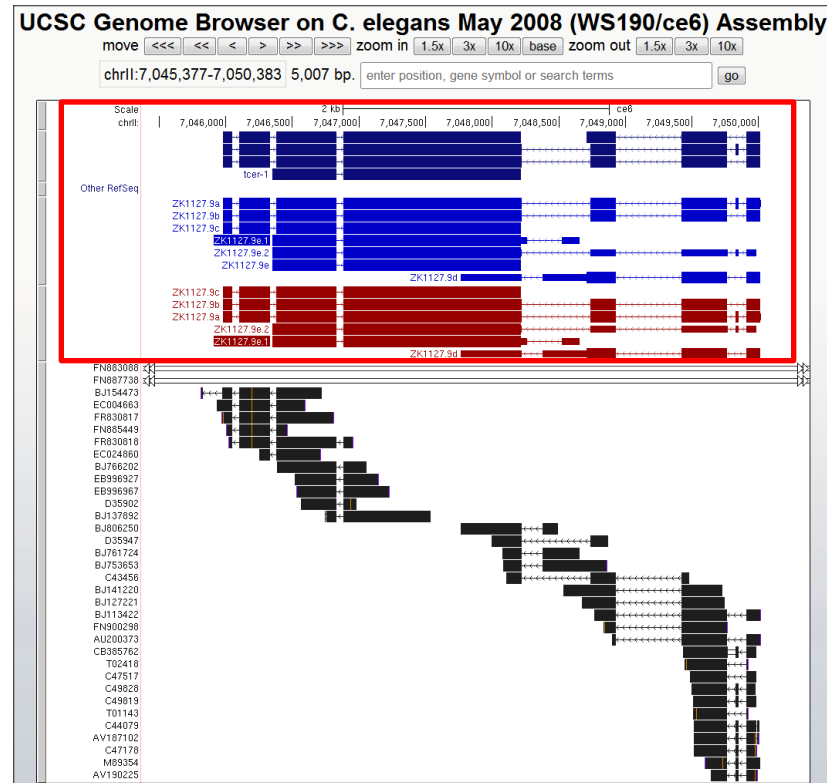


Jim Kent



Background

- ❑ Draft human genome sequence became available at the UCSC in 2000
- ❑ Intronerator was used as the graphics engine



UCSC Genome Browser



Genome Browser

<http://genome.ucsc.edu/>

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

[Genome Browser](#)
[ENCODE](#)
[Neandertal](#)
[Blat](#)
[Table Browser](#)
[Gene Sorter](#)
[In Silico PCR](#)
[Genome Graphs](#)
[Galaxy](#)
[VisiGene](#)
[Utilities](#)
[Downloads](#)
[Release Log](#)
[Custom Tracks](#)
[Cancer Browser](#)
[Microbial Genomes](#)
[Mirrors](#)
[Training](#)
[Credits](#)
[Publications](#)
[Cite Us](#)
[Licenses](#)
[Jobs](#)
[Staff](#)
[Contact Us](#)

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.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work 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 data 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 Genome Sciences and Policy. 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

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [Genome Browser News](#) mailing list.

24 October 2013 - Job Opening: UCSC Genome Browser Trainer

The [Center for Biomolecular Science and Engineering](#) (CBSE) at University of California Santa Cruz seeks an articulate, self-motivated educational professional to provide in-person training on the UCSC Genome Browser at universities, hospitals, institutes, and professional meetings in the United States and international locations. The position requires a Master's degree in a biological science, depth in molecular biology, experience in a research environment, working knowledge of video production, and a minimum of 5 years of genome browser experience ranging from novice users to bioinformatics specialists. Presentations include formal talks, problem-solving sessions, and hands-on training.

This position requires a Master's degree in a biological science, depth in molecular biology, experience in a research environment, working knowledge of video production, and a minimum of 5 years of genome browser experience ranging from novice users to bioinformatics specialists. Preferred qualifications include a PhD in a relevant field, experience with video production, and a minimum of 5 years of genome browser experience ranging from novice users to bioinformatics specialists.

For more information and to apply for this position, see [Job #1304619](#) on the UCSC Staff Employment website.

23 October 2013 - dbSNP Build 138 Available for hg19

We are pleased to announce the release of four tracks derived from NCBI [dbSNP](#) Build 138 data, available on the human assembly (GRCh37/hg19). The tracks include corresponding coloring and filtering options in the Genome Browser.

As was the case for the annotations based on the previous dbSNP build 137, there are four tracks in this release. One is a track containing all variants and the other three are subsets of this track and show interesting and easily defined subsets of dbSNP:

- Common SNPs (138): uniquely mapped variants that appear in at least 1% of the population or are 100% non-reference
- Flagged SNPs (138): uniquely mapped variants, excluding Common SNPs, that have been flagged by dbSNP as "clinically associated"
- Mult. SNPs (138): variants that have been mapped to more than one genomic location

By default, only the Common SNPs (138) are visible; other tracks must be made visible using the track controls.

You will find the four SNPs (138) tracks on the Human Feb. 2009 (GRCh37/hg19) browser in the "Variation and Repeats" group.

The tracks were produced at UCSC by Angie Hinrichs and Luvina Guruvadoo. We'd like to thank the dbSNP group at NCBI for providing access to the data.



Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term	
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr10:123,227,429-123,343,066	FGFR2	<input type="button" value="submit"/>

[Click here to reset](#) the browser user interface settings to their defaults.

FGFR2 (Homo sapiens fibroblast growth factor receptor 2 (FGFR2), transcript variant 2, mRNA.)

Human Genome Browser – hg19 assembly ([sequences](#))

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

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:

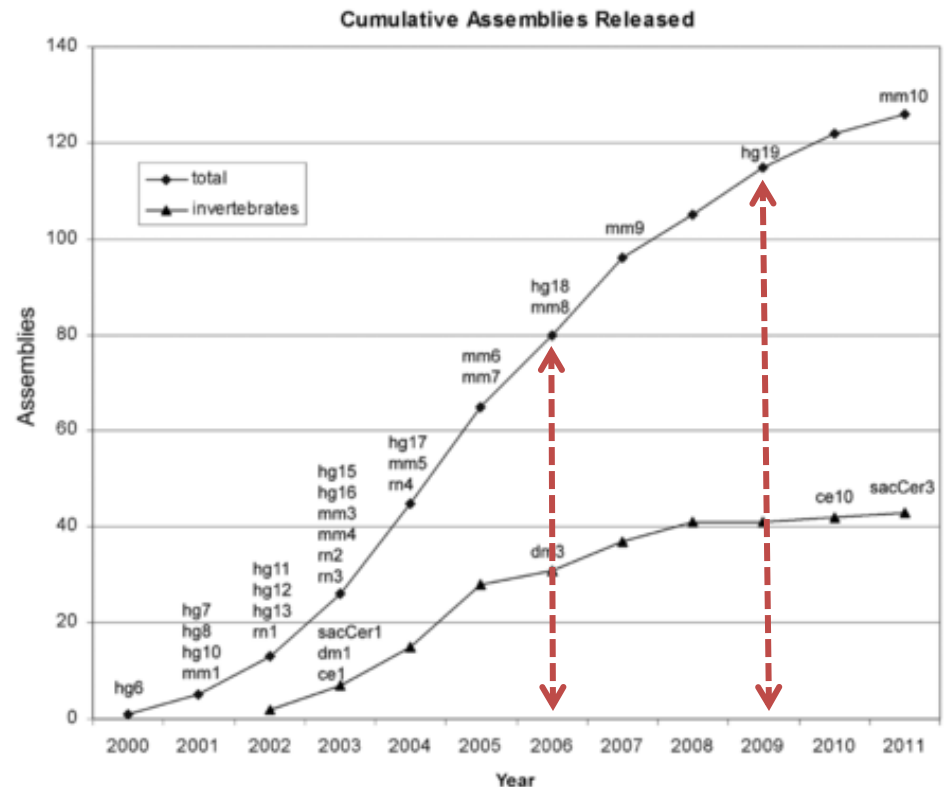
chr7	Displays all of chromosome 7
chrUn_gi000212	Displays all of the unplaced contig gi000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175 15q11;15q13 rs1042522;rs1800370	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11 to 15q13, or SNPs rs1042522 and rs1800370. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
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
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans



Homo sapiens
(Graphic courtesy of [CBSE](#))

Genome Assemblies

- ❑ Regular updates to genome assemblies to close gaps in genomic sequence, troubleshoot assembly problems and otherwise improve the genome assemblies
- ❑ Shifting coordinates for known sequences and a potential for confusion and error among researchers, particularly when reading literature based on older versions.
- ❑ Frequently used assemblies hg18/hg19
- ❑ New assemblies increase genomic coverage 6-fold and have been deposited in GenBank.
- ❑ 127 genome assemblies have been released on 58 organisms (April 2012)



Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr10:123,227,429-123,343,066	FGFR2 <input type="submit" value="submit"/>

[Click here to reset](#) the browser user interface settings to their defaults.

FGFR2 (Homo sapiens fibroblast growth factor receptor 2 (FGFR2), transcript variant 2, mRNA)

Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [GRCh37](#) in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced or keywords from the GenBank description of an mRNA. The following list shows for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gli000212	Displays all of the unplaced contig gli000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting at 1000000
RH18061;RH80175	Displays region between genome landmarks, such as bands or SNPs rs1042522 and rs1800370. This syntax may also be used for mRNAs, refSeqs, etc.
15q11;15q13	
rs1042522;rs1800370	
D16S3046	Displays region around STS marker D16S3046 from the genome
AA205474	Displays region of EST with GenBank accession AA205474
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 symbol 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
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans

RefSeq Genes

[FGFR2 at chr10:123237844-123353481](#) - (NM_001144914) fibroblast growth factor receptor 2 isoform 4 precursor
[FGFR2 at chr10:123237844-123356159](#) - (NM_001144915) fibroblast growth factor receptor 2 isoform 5 precursor
[FGFR2 at chr10:123241367-123357972](#) - (NM_001144919) fibroblast growth factor receptor 2 isoform 9 precursor
[FGFR2 at chr10:123237844-123357972](#) - (NM_001144917) fibroblast growth factor receptor 2 isoform 7 precursor
[FGFR2 at chr10:123237844-123357972](#) - (NM_022970) fibroblast growth factor receptor 2 isoform 2 precursor
[FGFR2 at chr10:123237844-123353772](#) - (NM_001144916) fibroblast growth factor receptor 2 isoform 6 precursor
[FGFR2 at chr10:123237844-123357972](#) - (NM_001144918) fibroblast growth factor receptor 2 isoform 8 precursor
[FGFR2 at chr10:123237844-123357972](#) - (NM_000141) fibroblast growth factor receptor 2 isoform 1 precursor
[FGFR2 at chr10:123241367-123353481](#) - (NM_001144913) fibroblast growth factor receptor 2 isoform 3 precursor
[FGFR2 at chr10:123237844-123353481](#) - (NM_023029) fibroblast growth factor receptor 2 isoform 11 precursor
[FGFR2 at chr10:123237844-123357972](#) - (NR_073009)

Non-Human RefSeq Genes

[fgfr2 at chr10:123239371-123324098](#) - (NM_178303) fibroblast growth factor receptor 2 isoform 3 precursor
[fgfr2 at chr10:123239371-123325219](#) - (NM_001090663) fibroblast growth factor receptor 2 precursor
[fgfr2 at chr10:123239023-123325219](#) - (NM_001102856) fibroblast growth factor receptor 2 precursor
[FGFR2 at chr10:123239371-123353399](#) - (NM_205319) fibroblast growth factor receptor 2 precursor
[FGFR2 at chr10:123237856-123269807](#) - (NM_001131221) fibroblast growth factor receptor 2
[FGFR2 at chr10:123237856-123353434](#) - (NM_001003336) fibroblast growth factor receptor 2 precursor
[FGFR2 at chr10:123238077-123357741](#) - (NM_001163863) fibroblast growth factor receptor 2 precursor
[FGFR2 at chr10:123239230-123353378](#) - (NM_001099924) fibroblast growth factor receptor 2 precursor
[FGFR2 at chr10:123239559-123353331](#) - (NM_001082688) fibroblast growth factor receptor 2 precursor
[Fgfr2 at chr10:123237846-123358315](#) - (NM_201601) fibroblast growth factor receptor 2 isoform IIIb
[Fgfr2 at chr10:123237846-123358315](#) - (NM_010207) fibroblast growth factor receptor 2 isoform IIIc
[Fgfr2 at chr10:123237873-123357855](#) - (NM_001109893) fibroblast growth factor receptor 2 isoform c
[Fgfr2 at chr10:123237873-123357855](#) - (NM_001109896) fibroblast growth factor receptor 2 isoform f
[Fgfr2 at chr10:123237873-123357855](#) - (NM_012712) fibroblast growth factor receptor 2 isoform a
[Fgfr2 at chr10:123237873-123357855](#) - (NM_001109894) fibroblast growth factor receptor 2 isoform d
[Fgfr2 at chr10:123237873-123357855](#) - (NM_001109892) fibroblast growth factor receptor 2 isoform b
[Fgfr2 at chr10:123237873-123357855](#) - (NM_001109895) fibroblast growth factor receptor 2 isoform e
[FGFR2 at chr10:123237846-123357550](#) - (NM_001205310) fibroblast growth factor receptor 2
[fgfr2 at chr10:123239371-123324098](#) - (NM_001243004) fibroblast growth factor receptor 2 isoform 1 precursor
[fgfr2 at chr10:123239371-123324098](#) - (NM_001243005) fibroblast growth factor receptor 2 isoform 2 precursor
[fgfr2 at chr10:123239371-123324098](#) - (NM_001243006) fibroblast growth factor receptor 2 isoform 4 precursor

Basic Gene Annotation Set from ENCODE/Gencode Version 17

[FGFR2 at chr10:123237848-123353481](#)
[FGFR2 at chr10:123237848-123356159](#)
[FGFR2 at chr10:123237855-123357598](#)
[FGFR2 at chr10:123237878-123290828](#)
[FGFR2 at chr10:123238586-123357972](#)
[FGFR2 at chr10:123238732-123357812](#)
[FGFR2 at chr10:123239133-123357966](#)



UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move



zoom in

1.5x

3x

10x

base

zoom out

1.5x

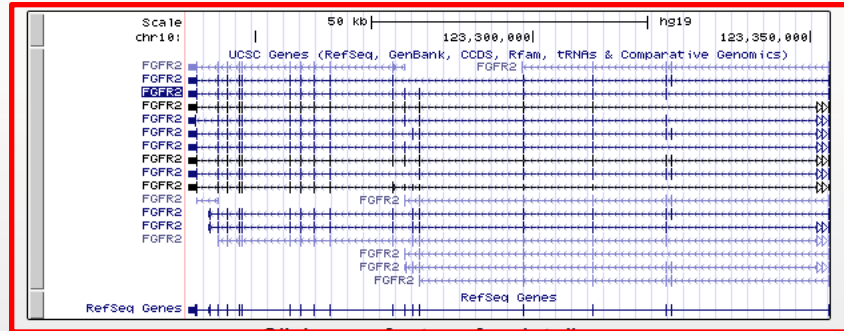
3x

10x

chr10:123,237,844-123,353,481 115,638 bp.

enter position, gene symbol or search terms

go



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 start



2.0



move end



2.0



track search

default tracks

default order

hide all

add custom tracks

track hubs

configure

reverse

resize

refresh

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
- Phenotype and Disease Associations refresh
- Genes and Gene Prediction Tracks refresh
- Literature refresh
- mRNA and EST Tracks refresh
- Expression refresh

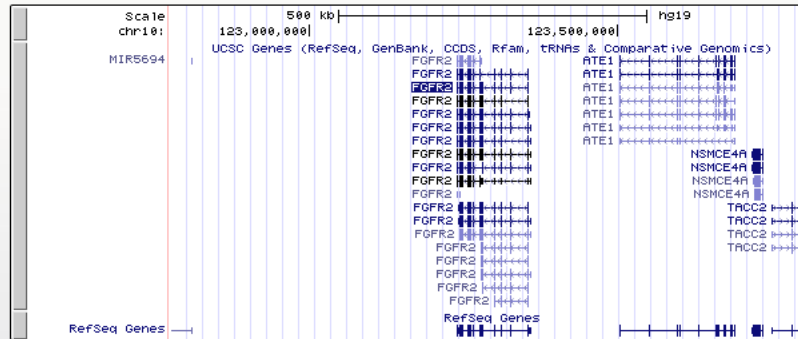




UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr10:122,775,292-123,816,033 1,040,742 bp.



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 start

< 2.0 >

move end

< 2.0 >

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

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
- Phenotype and Disease Associations
- Genes and Gene Prediction Tracks
- Literature
- mRNA and EST Tracks
- Expression





Configure Image

image width: pixelslabel area width: characterstext size:

<input checked="" type="checkbox"/>	Display chromosome ideogram above main graphic
<input checked="" type="checkbox"/>	Show light blue vertical guidelines
<input checked="" type="checkbox"/>	Display labels to the left of items in tracks
<input checked="" type="checkbox"/>	Display description above each track
<input checked="" type="checkbox"/>	Show track controls under main graphic
<input type="checkbox"/>	Next/previous item navigation
<input checked="" type="checkbox"/>	Next/previous exon navigation

Configure Tracks on UCSC Genome Browser: Human Feb. 2009 (GRCh37/hg19)

Tracks: Groups:

Control track and group visibility more selectively below.

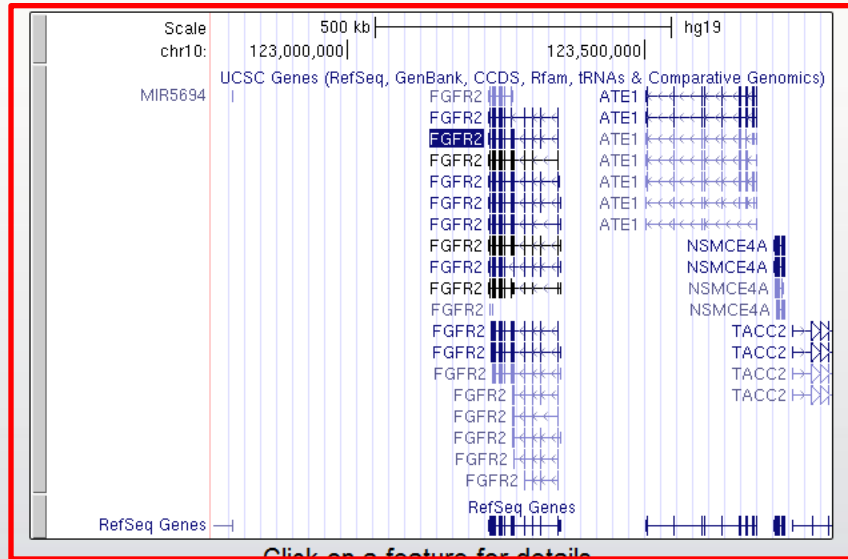
+ Mapping and Sequencing Tracks **+ Phenotype and Disease Associations** **+ Genes and Gene Prediction Tracks** **+ Literature** **+ mRNA and EST Tracks** **+ Expression** **+ Regulation**



UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr10:122,775,292-123,816,033 1,040,742 bp.



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 start < 2.0 >

move end < 2.0 >

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



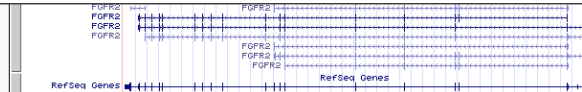
Mapping and Sequencing Tracks



Phenotype and Disease Associations

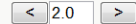


Annotation tracks

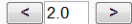


Click on a feature for details.
Click or drag in the base position track to zoom in.
Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

move start



move end



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

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
+	Phenotype and Disease Associations	refresh
+	Genes and Gene Prediction Tracks	refresh
+	Literature	refresh
+	mRNA and EST Tracks	refresh
+	Expression	refresh
+	Regulation	refresh
-	Comparative Genomics	refresh

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

Vertebrate Chain/Net

+	Neandertal Assembly and Analysis	refresh
+	Denisova Assembly and Analysis	refresh
+	Variation and Repeats	refresh

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Search for Tracks in the Human Feb. 2009 (GRCh37/hg19) Assembly

Search **Advanced**

conservation

search clear cancel

return to browser (2 of 17 selected)

Visibility	Track Name	Sort: by Relevance Alphabetically
<input checked="" type="checkbox"/> full	Conservation	Vertebrate Multiz Alignment & Conservation (46 Species)
<input type="checkbox"/> hide	Primate Cons	Primate Conservation by PhastCons
<input type="checkbox"/> hide	Vertebrate Cons	Vertebrate Conservation by PhastCons
<input type="checkbox"/> hide	Primate Cons	Primate Basewise Conservation by PhyloP
<input type="checkbox"/> hide	Vertebrate Cons	Vertebrate Basewise Conservation by PhyloP
<input type="checkbox"/> hide	Mammal Cons	Placental Mammal Conservation by PhastCons
<input type="checkbox"/> hide	Cons Indels MmCf	Indel-based Conservation for human hg19, mouse mm8 and dog canFam3
<input checked="" type="checkbox"/> full	Mammal Cons	Placental Mammal Basewise Conservation by PhyloP
<input type="checkbox"/> hide	Mod Hum Variants	Variant Calls from 11 Modern Human Genome Sequences
<input type="checkbox"/> hide	Denisova Variants	Variant Calls from High-Coverage Genome Sequence of an Archaic Denisovan
<input type="checkbox"/> hide	CCDS	Consensus CDS
<input type="checkbox"/> hide	TransMap	TransMap Alignments
<input type="checkbox"/> hide	TransMap UCSC	TransMap UCSC Gene Mappings
<input type="checkbox"/> hide	TransMap RefGene	TransMap RefSeq Gene Mappings
<input type="checkbox"/> hide	TransMap mRNA	TransMap GenBank mRNA Mappings
<input type="checkbox"/> hide	TransMap ESTs	TransMap Spliced EST Mappings
<input type="checkbox"/> hide	GERP	GERP scores for mammalian alignments

Return to Browser (2 of 17 selected)

Tracks so marked are containers which group related data tracks. Containers may need additional configuration (icon) before they can be viewed in the browser.



Annotation tracks

❑ The database may contain any data that can be mapped to genomic coordinates and therefore can be displayed in the Genome Browser

❑ Overview of tracks: <http://genome.ucsc.edu/cgi-bin/hgTracks>

❑ Three different categories:

- ❑ computed at UCSC
- ❑ computed elsewhere and displayed at UCSC
- ❑ computed and hosted entirely elsewhere



Annotation tracks computed at UCSC

- ❑ Comparative genomic annotations as well as Convert and liftOver capabilities
- ❑ mRNAs and ESTs in GenBank are aligned to the reference assembly in separate tracks (75 million GenBank RNAs and ESTs, ~3 billion bases of the human reference assembly → 2 CPU-years of computing time)
- ❑ The Conservation composite track displays the results of the multiz algorithm that aligns the results from up to 46 pairwise Blastz alignments to the reference assembly (e.g. hg19 human assembly consumed 10 CPU-years)



mRNA and EST Tracks
refresh

Human mRNAs pack	Spliced ESTs hide	Human ESTs hide	Other mRNAs hide	Other ESTs hide	<input checked="" type="checkbox"/> H-Inv hide
UniGene hide	Gene Bounds hide	SIB Alt-Splicing hide	<input checked="" type="checkbox"/> Poly(A) hide	PolyA-Seq hide	<input checked="" type="checkbox"/> CGAP SAGE hide

[Human RNA Editing](#)
hide

+ Expression
refresh

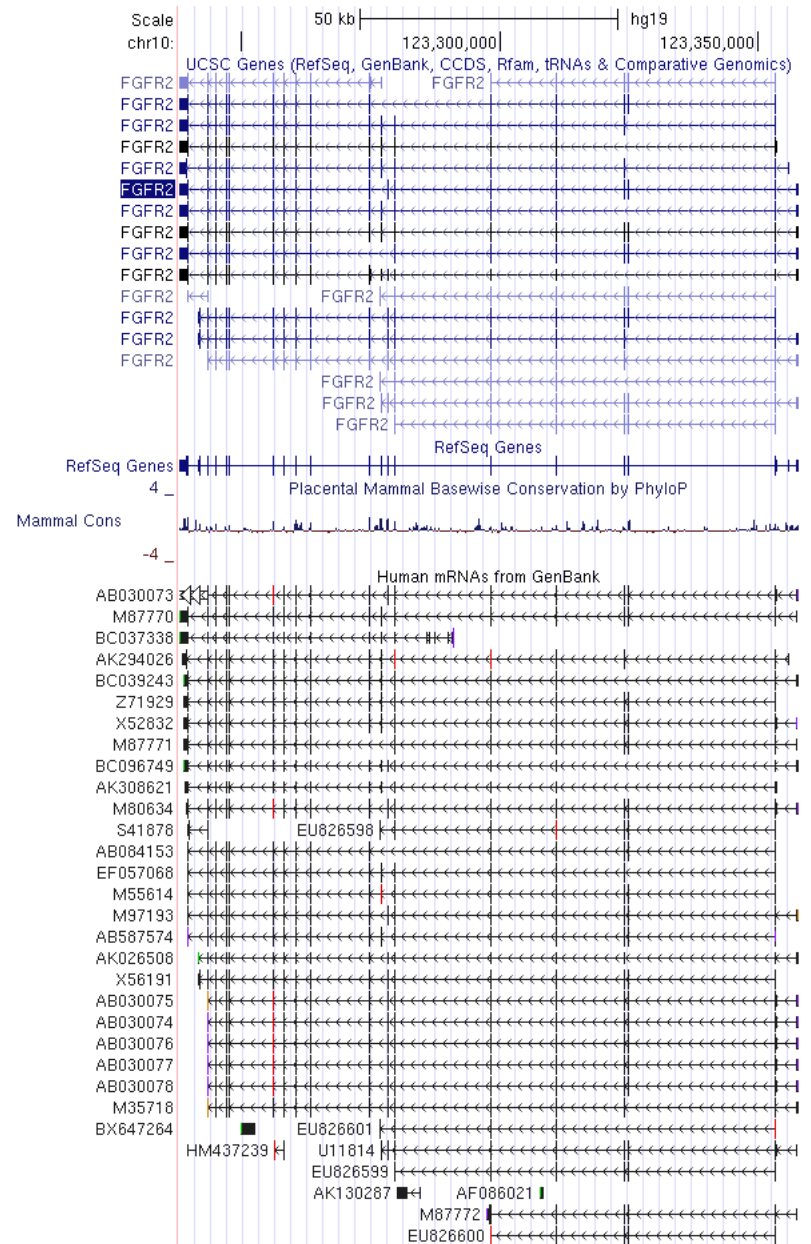
+ Regulation
refresh

- Comparative Genomics
refresh

Conservation full	<input checked="" type="checkbox"/> Cons Indels MmCf hide	GERP hide	<input checked="" type="checkbox"/> Evo Cpg hide	Primate Chain/Net hide	Placental Chain/Net hide
--------------------------------------	--	------------------------------	---	---	---

[Vertebrate Chain/Net](#)
hide

+ Neandertal Assembly and Analysis
refresh



Annotation tracks computed elsewhere and displayed at UCSC

Annotations that are not post-processed by the UCSC

- Probe sets for commercially available microarrays, copy-number variation from the Database of Genomic Variants or expression data from the GNF Expression Atlas
- Data Coordination Center for the ENCODE project allowing access to a large number of functional annotations in regards to gene regulation

Annotations that are post-processed by the UCSC

- dbSNP (Common SNPs, Flagged SNPs, Mult. SNPs)
- OMIM (OMIM Allelic Variant SNPs, OMIM Genes, OMIM Phenotypes)



Phenotype and Disease Associations refresh

[GAD View](#) [DECIPHER](#) **OMIM AV SNPs** [OMIM Genes](#) [OMIM Pheno Loci](#) [COSMIC](#)

[LOVD Variants](#) [HGMD Variants](#) [UniProt Variants](#) [ClinVar Variants](#) [GWAS Catalog](#) [ISCA](#)

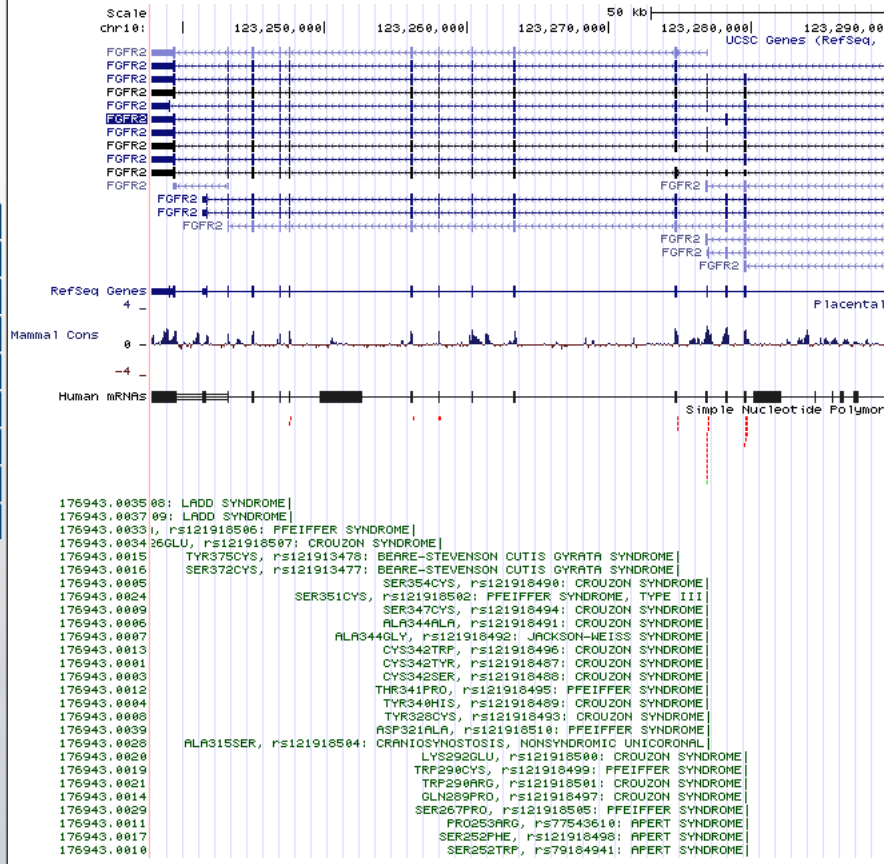
[Coriell CNVs](#) [RGD Human QTL](#) [RGD Rat QTL](#) [MGI Mouse QTL](#) [GeneReviews](#)

Genes and Gene Prediction Tracks refresh

[Literature](#) refresh
[mRNA and EST Tracks](#) refresh
[Expression](#) refresh
[Regulation](#) refresh
[Comparative Genomics](#) refresh
[Neandertal Assembly and Analysis](#) refresh
[Denisova Assembly and Analysis](#) refresh

Variation and Repeats refresh

Common SNPs(138) <input type="button" value="hide"/>	Flagged SNPs(138) <input type="button" value="squish"/>	Mult. SNPs(138) <input type="button" value="hide"/>	All SNPs(138) <input type="button" value="hide"/>	Common SNPs(137) <input type="button" value="hide"/>	Flagged SNPs(137) <input type="button" value="hide"/>
Mult. SNPs(137) <input type="button" value="hide"/>	All SNPs(137) <input type="button" value="hide"/>	Common SNPs(135) <input type="button" value="hide"/>	Flagged SNPs(135) <input type="button" value="hide"/>	Mult. SNPs(135) <input type="button" value="hide"/>	All SNPs(135) <input type="button" value="hide"/>
1000G Ph1 Vars <input type="button" value="hide"/>	1000G Ph1 Accsbl <input type="button" value="hide"/>	GIS DNA PET <input type="button" value="hide"/>	HAIB Genotype <input type="button" value="hide"/>	SNP/CNV Arrays <input type="button" value="hide"/>	HGDP Allele Freq <input type="button" value="hide"/>
HapMap SNPs <input type="button" value="hide"/>	DGV Struct Var <input type="button" value="hide"/>	Segmental Dups <input type="button" value="hide"/>	RepeatMasker <input type="button" value="hide"/>	Interrupted Rpts <input type="button" value="hide"/>	Simple Repeats <input type="button" value="hide"/>



Annotation tracks computed and hosted elsewhere

❑ Data tracks are hosted remotely (no data)

Roadmap Epigenomics Visualization Hub (VizHub)

Roadmap Epigenomics Project | NCBI Epigenomics Gateway | Roadmap Data Coordination Center

Embedded WashU EpiGenome Browser [learn how](#) [Start Demo](#)

hg19 chr7:27052997-27373365 400 bp/px

Assay Sample Institution

RepeatMasker Ensembl RefSeq genes

WASHU EPIGENOME BROWSER
A new-generation genome browser for integrative visualization of genomic information. Hosts high volume of tracks from ENCODE and Roadmap Epigenomics projects, supports multiple organisms, visualizes chromatin-interaction data (e.g. Hi-C), performs gene set view, gene plot, and many others. [Go to the Browser](#) ▶

VizHub
Displaying sequencing data from Roadmap Epigenomics project, powered by a local mirror of UCSC Genome Browser. [Go to the Browser](#) ▶

Remote Data Hub
Displaying sequencing data from Roadmap Epigenomics project, powered by UCSC Genome Browser through data hub function. [Go to the Hub](#) ▶

VizHub

Roadmap Epigenomics Visualization Hub at Wash U

Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Session - FAQ - Help

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 [Neanderthal](#) 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

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

16 August 2012 – Announcing a Genome Browser for the Medium ground finch

We have released a browser for the Medium ground finch, *Geospiza fortis*, renowned as one of naturalist Charles Darwin's Galapagos finches. This species, which has been the subject of many evolutionary studies, is one of a group of birds that evolved over a few million years from a single ancestral species into multiple species whose beak sizes and shapes are specialized for using different food resources. The phenotypic diversity of these birds contributed to Darwin's theory of evolution. The significance of this genome assembly is described in the August 16,



Tracks from the Epigenome project

CpG and MRE sites refresh

Base Position CpG Islands GC Percent CpG_MRE sites

methylMnM refresh

methylCRF refresh

Epigenome Atlas Data Complete Collection Composite Tracks refresh

Broad Histone UCSD Histone Histone DNase Footprinting RNA

DNA Methylation By Assay... By Sample... Roadmap ChromHMM Roadmap ChromHMM 15 state Roadmap Uniformly Signal

Epigenome Atlas Data Complete Collection Integrative Tracks refresh

Assay Summary... Sample Summary... Methylation Summary...

Mapping and Sequencing Tracks refresh

Phenotype and Disease Associations refresh

GWAS Catalog

Genes and Gene Prediction Tracks refresh

UCSC Genes Alt Events GENCODE Genes V7 CCDS RefSeq Genes Other RefSeq

MGC Genes ORFeome Clones TransMap... Ensembl Genes N-SCAN Exoniphy

mRNA and EST Tracks refresh

Expression refresh

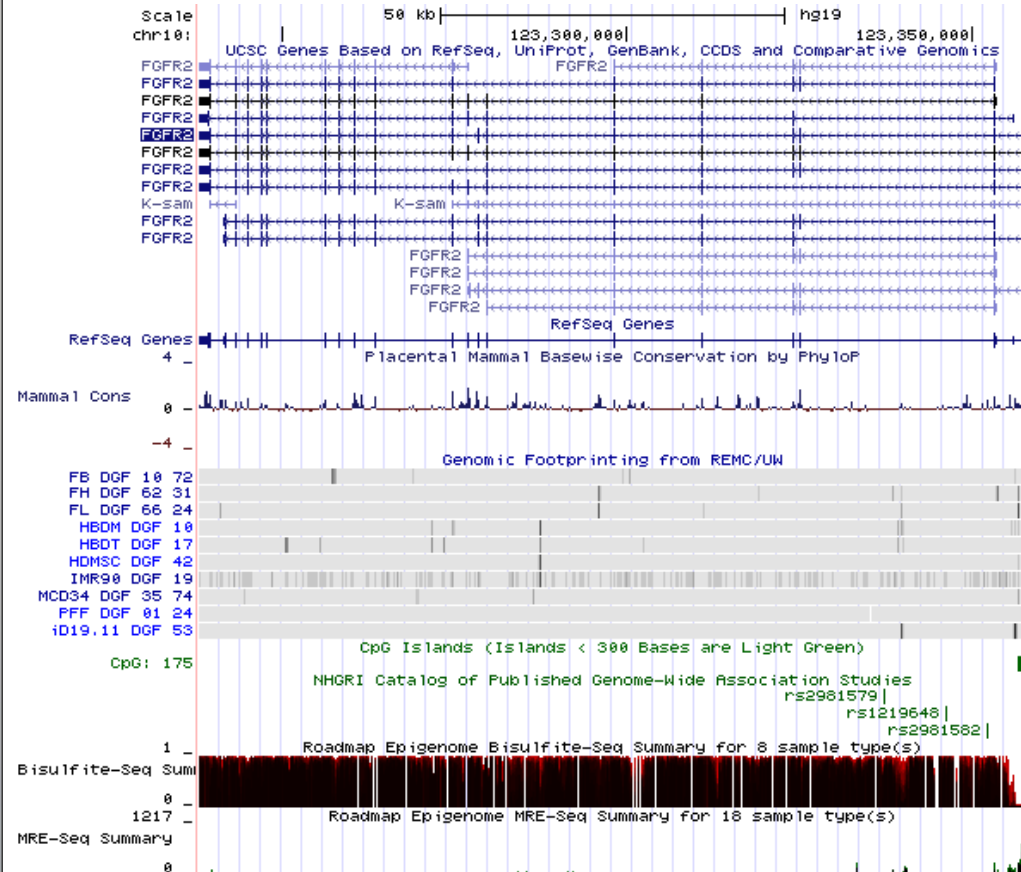
Regulation refresh

Comparative Genomics refresh

Conservation Chimp Chain/Net caJac1 Chain/Net felCat3 Chain/Net bosTau4 Chain/Net Primate Chain/Net

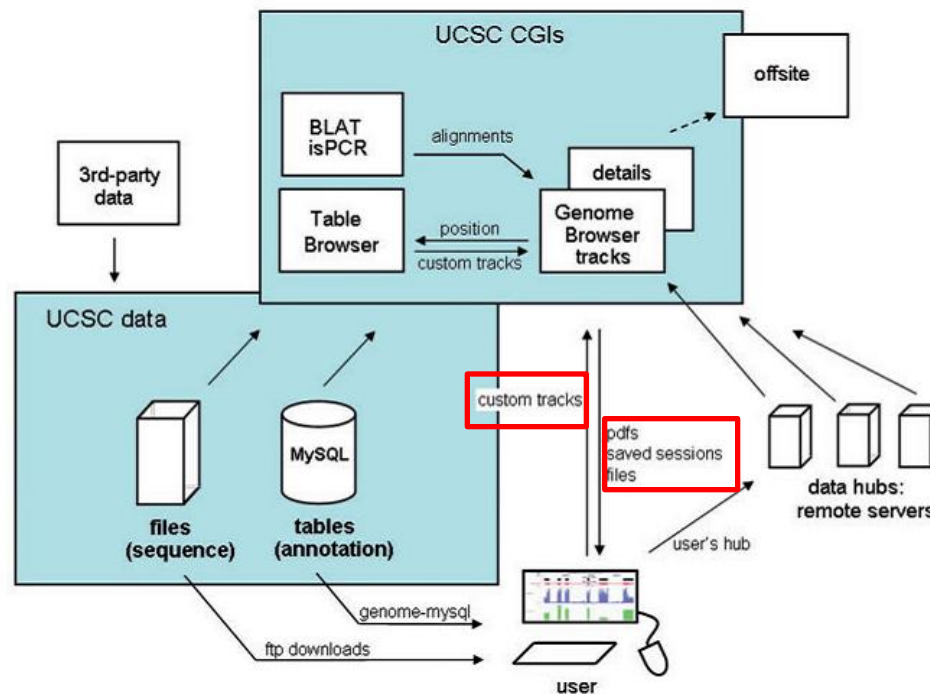
Placental Chain/Net Vertebrate Chain/Net Lizard Chain/Net xenTro2 Chain/Net Zebrafish Chain/Net Sea hare Chain/Net

Variation and Repeats refresh



Associated Tools

- ❑ Tools other than the main graphic image account for 42% of traffic on the UCSC server



Sessions

Tools Mirrors Downloads My Data About Us View Help

Genome Browser on Human Reference Genome (GRCh37/hg19) Assembly

844-123,357,972 120,129 bp. enter position, gene symbol or search terms go

Scale 50 kb chr10 (hg19) 123,300,000 hg19 123,350,000

UCSC Genes (RefSeq, GenBank, CDS, Refam, tRNAs & Comparative Genomics)

RefSeq Genes

Mammal Cons

Human mRNAs from GenBank

Simple Nucleotide Polymorphisms (dbSNP 138) Flagged by dbSNP as Clinically Assoc

OMIM Allelic Variant SNPs

176943.0035 NDROME|
 176943.0037 NDROME|
 176943.0033 FER SYNDROME|
 176943.0034 UZON SYNDROME|
 176943.0015 CON CUTIS GYRATA SYNDROME|
 176943.0016 CON CUTIS GYRATA SYNDROME|
 176943.0005 1918490: CROUZON SYNDROME|
 176943.0024 PFEIFFER SYNDROME, TYPE III|
 176943.0009 1918494: CROUZON SYNDROME|
 176943.0005 1918491: CROUZON SYNDROME|
 176943.0007 32: JACKSON-WEISS SYNDROME|
 176943.0013 1918496: CROUZON SYNDROME|
 176943.0011 1918487: CROUZON SYNDROME|
 176943.0003 1918488: CROUZON SYNDROME|
 176943.0012 1918495: PFEIFFER SYNDROME|
 176943.0004 1918489: CROUZON SYNDROME|
 176943.0002 1918492: CROUZON SYNDROME|
 176943.0039 1918510: PFEIFFER SYNDROME|
 176943.0028 5, NONSYNDROMIC UNICORONIL|
 176943.0020 1121918590: CROUZON SYNDROME|
 176943.0019 1918499: PFEIFFER SYNDROME|
 176943.0021 1121918591: CROUZON SYNDROME|
 176943.0014 1121918497: CROUZON SYNDROME|
 176943.0029 21918505: PFEIFFER SYNDROME|
 176943.0011 rs77543610: APERT SYNDROME|
 176943.0017 rs121918498: APERT SYNDROME|
 176943.0010 rs79184941: APERT SYNDROME|

Save Settings

Save current settings as named session:

name: allow this session to be loaded by others

Save current settings to a local file:

file: file type returned:

(leave file blank to get output in browser window)

Restore Settings

Use settings from another user's saved session:

user: session name:

Use settings from a local file:

Use settings from a URL (http://..., ftp://...):

Sharing Sessions

There are several ways to share saved sessions with others.

- Each previously saved named session appears with Browser and Email links. That session loaded. The resulting Genome Browser page can be bookmarked. Email link invokes your email tool with a message containing the Genome Browser URL.
- If you have saved your settings to a local file, you can send email to others with genome.ucsc.edu/cgi-bin/hgSession.
- If a saved settings file is available from a web server, you can send email to [hgSession?hgS_doLoadUrl=submit&hgS_loadUrlName=U](mailto:genome.ucsc.edu/cgi-bin/hgSession?hgS_doLoadUrl=submit&hgS_loadUrlName=U) where **U** is the URL of the file, e.g. <http://myServer.com/mySession.txt>. In this type of link, you can replace "hgSession" with "hgTrack".



Custom

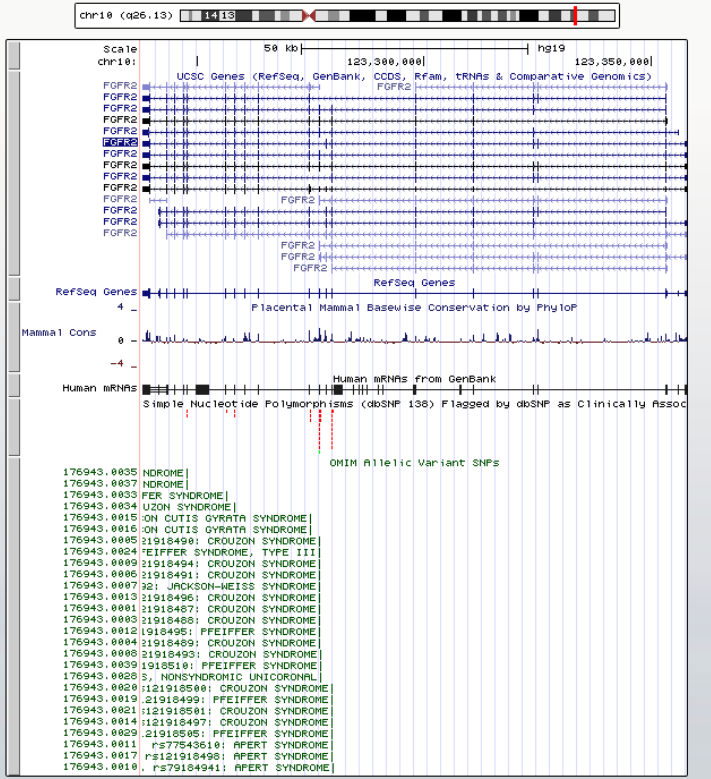
Genome Browser Tools Mirrors Downloads My Data About Us View Help

Genome Browser on Human GRCh37/hg19

move <<< << < > >> >>> zoom zoom out 1.5x

10:123,237,844-123,357,972 120,129 bp. enter position, gene symbol or search terms

- Sessions
- Track Hubs
- Custom Tracks



Add Custom Tracks

clade Mammal genome Human assembly Feb. 2009 (GRCh37/hg19)

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [bigBed](#), [bedGraph](#), [GFF](#), [GTF](#), [WIG](#), [bigWig](#), [MAF](#), [BAM](#), [BED detail](#), [Personal Genome SNP](#), [VCF](#), [broadPeak](#), [narrowPeak](#), or [PSL](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Data in the bigBed, bigWig, BAM and VCF formats must be provided via a URL embedded in a track line in the box below. Publicly available custom tracks are listed [here](#). Examples are [here](#).

Paste URLs or data: Or upload:

Optional track documentation: Or upload:

Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.

