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

UNSW AUSTRALIA
ADULT CANCER PROGRAM

LOWY CANCER RESEARCH CENTRE

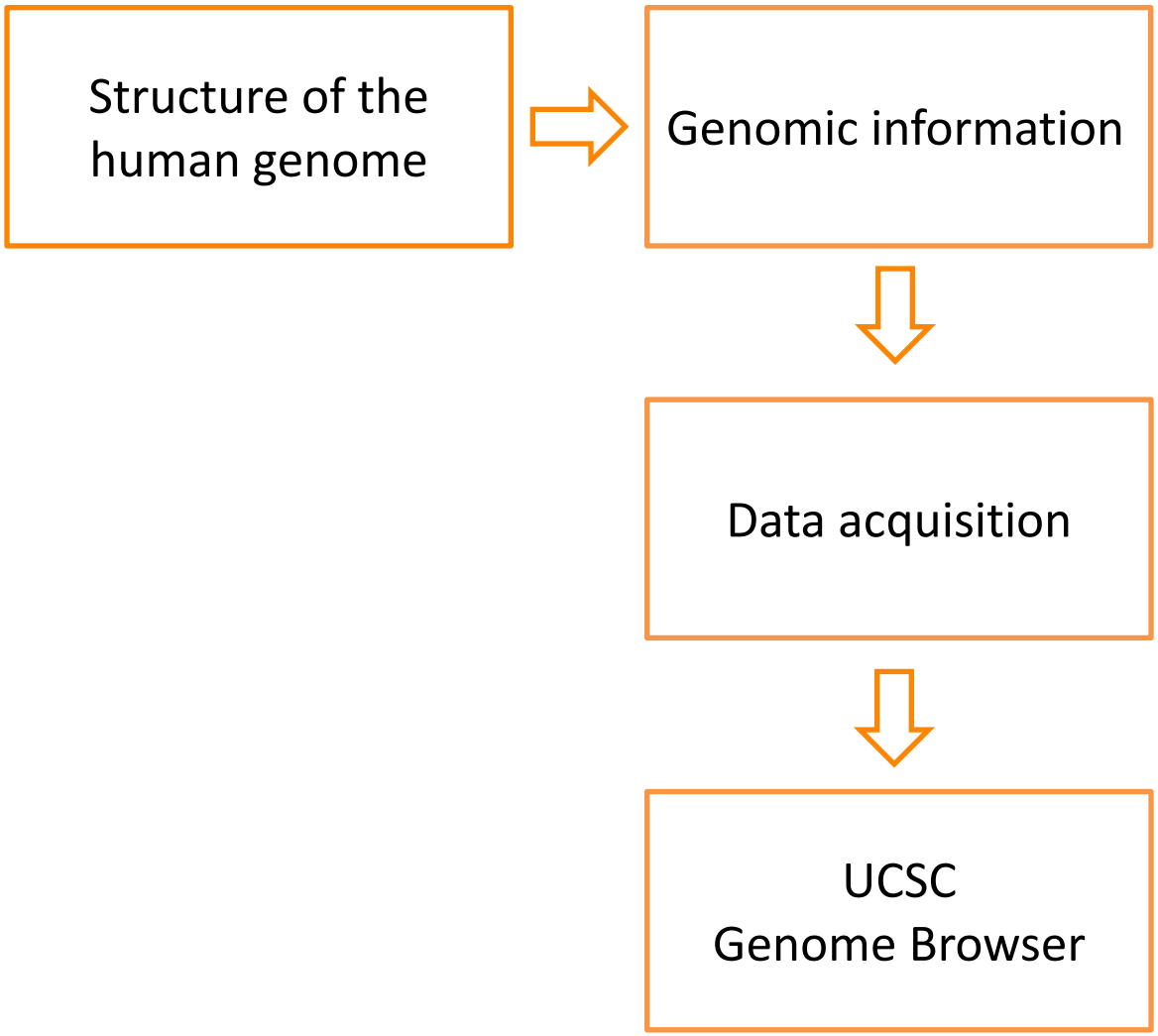
UTS UNIVERSITY OF TECHNOLOGY, SYDNEY

tcrn bridging the gap translational cancer research network

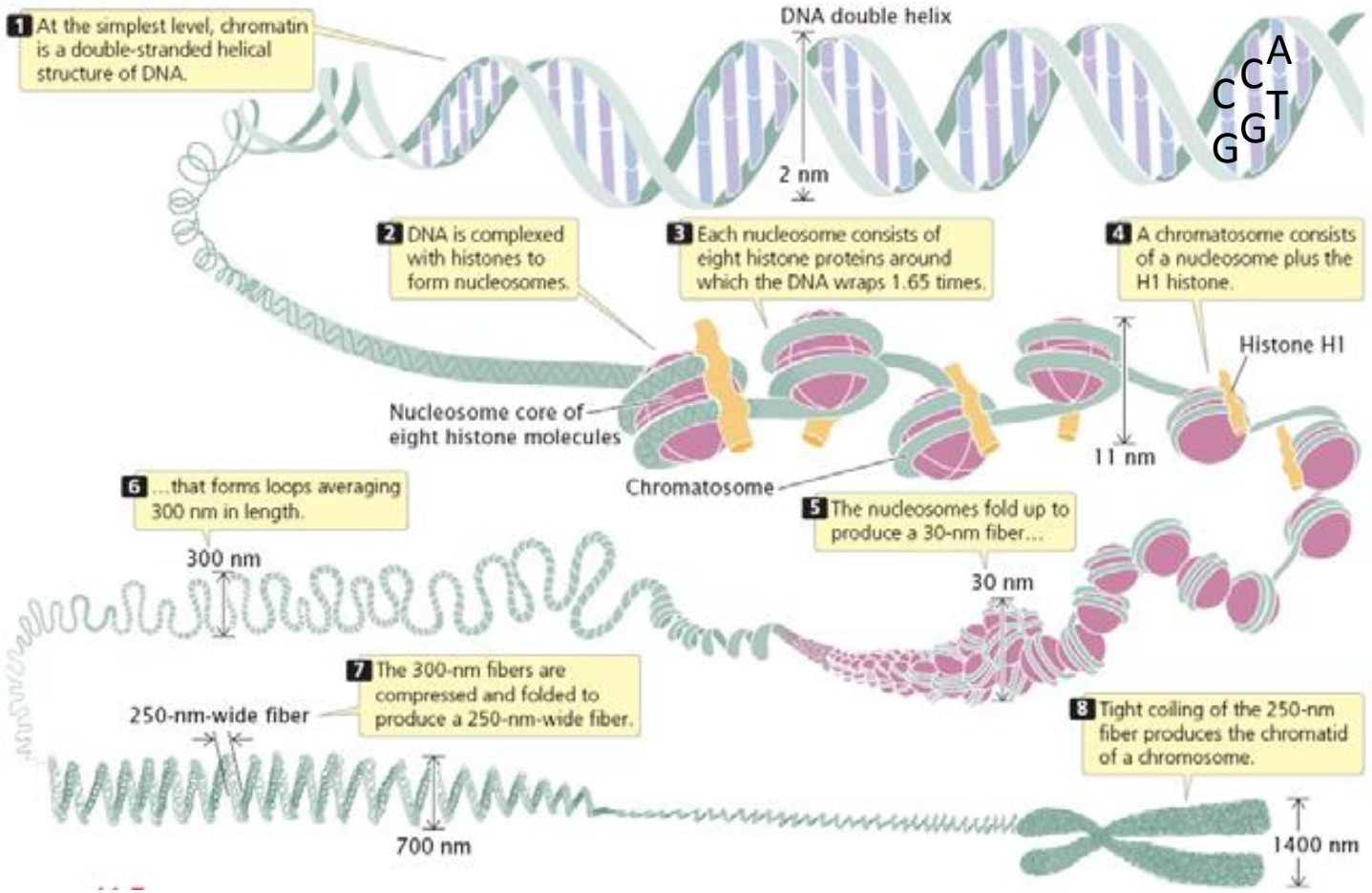
NSW GOVERNMENT

cancer institute NSW

What we will cover



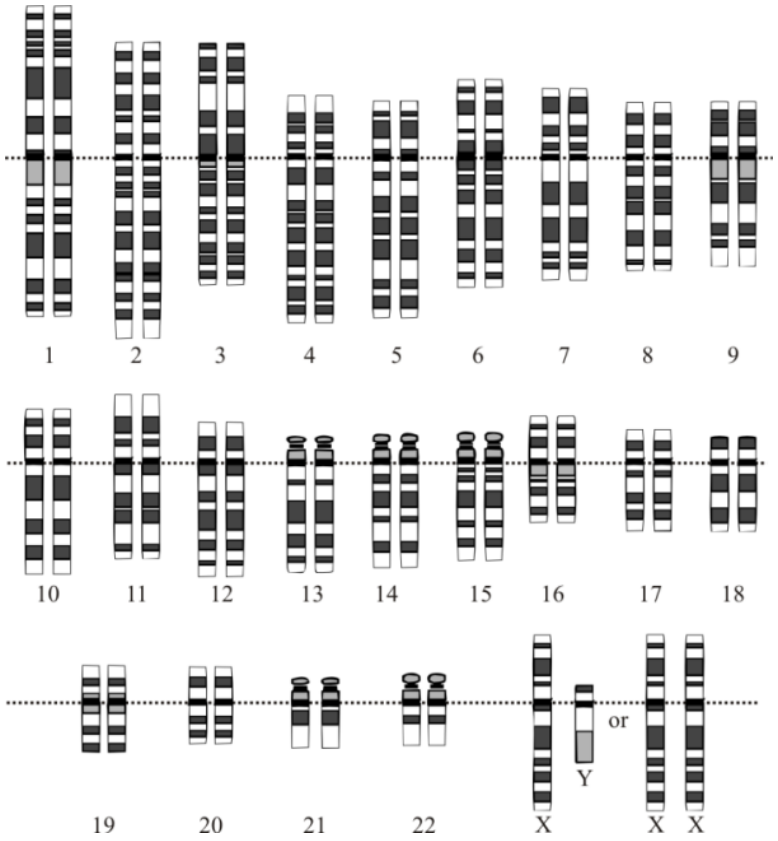
Structure of human genome



Annunziato A. 2008. DNA packaging: Nucleosomes and chromatin. *Nature Education* **1**(1).

UNSW THE UNIVERSITY OF NEW SOUTH WALES

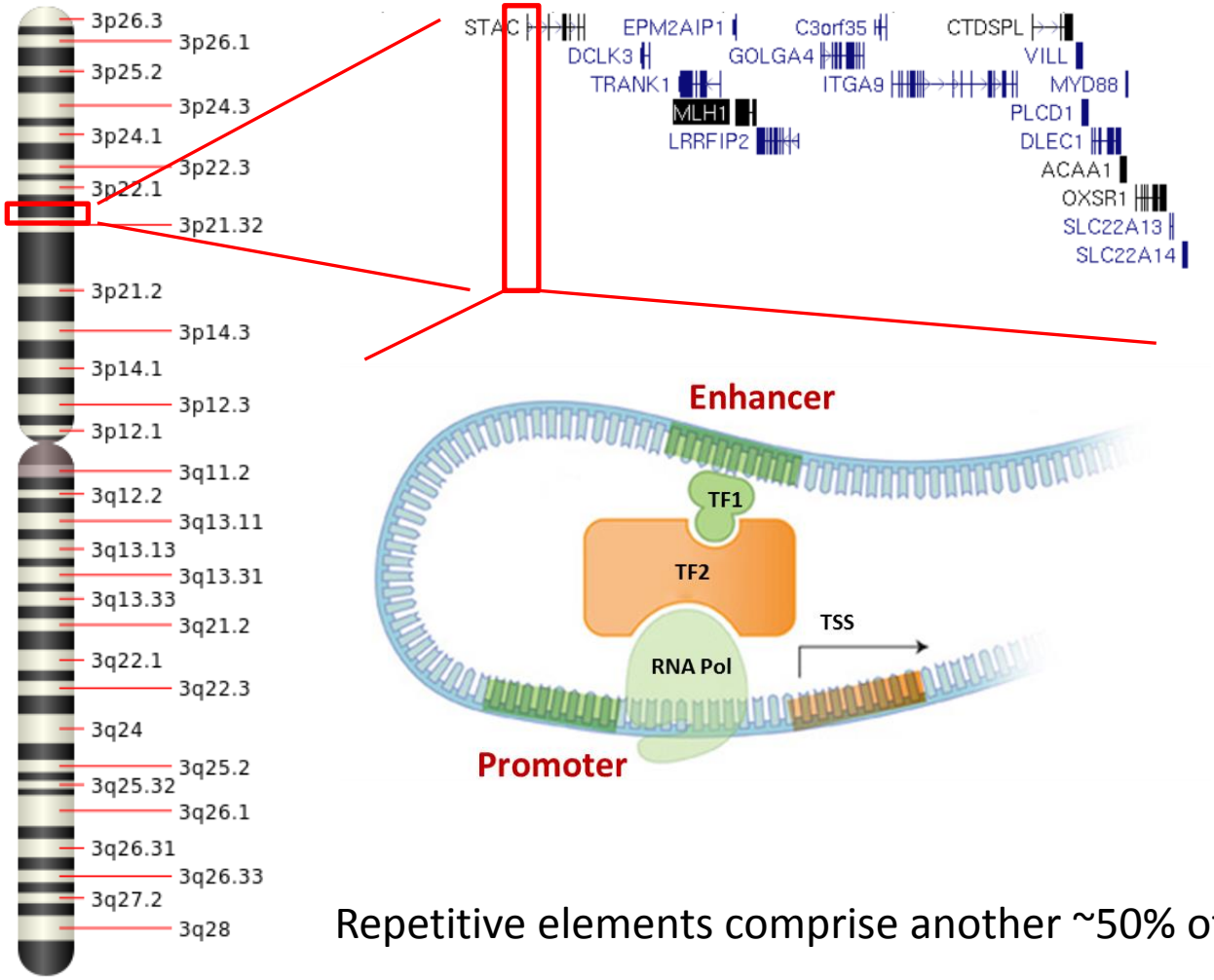
Structure of human genome



- 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)

Annunziato A. 2008. DNA packaging: Nucleosomes and chromatin. *Nature Education* **1**(1).

Information in the genome



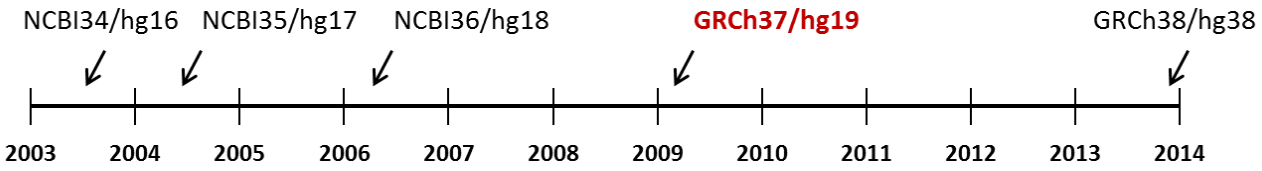
Genes:
~1.2% coding
~2% non-coding

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
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
ACAGTACTGGCGGATTATAGGGAAACACCCGGAGCATATGCTGTTTGGTC
TCAgtagactcctaaatatgggattcctgggtttaaaagtaaaaaataaa
tatgtttaatttgtgaactgattaccatcagaattgtactgttctgtatc
ccaccagcaatgtctaggaatgcctgtttctccacaaagtgtttacttttt
....
```

2. Location data – bed format (.bed)

```
chr1   934343     935552     HES4       0          -
chr1   948846     949919     ISG15      0          +
...
```

chromosome start end name score strand

All about genomic formats here - http://genome.ucsc.edu/FAQ/FAQformat.html

What we will cover

Structure of the human genome

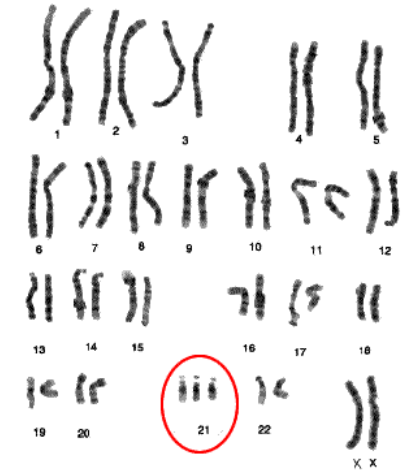
Genomic information

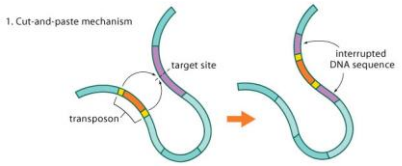
- DNA (Sequence variation)
- RNA (Genes & gene expression)
- Regulation\Epigenetics
  - DNA methylation
  - Histone modification
  - Transcription factor binding

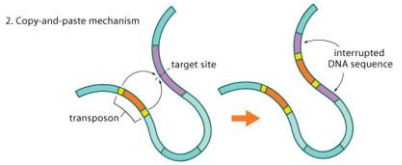
DNA: Sequence variation

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)

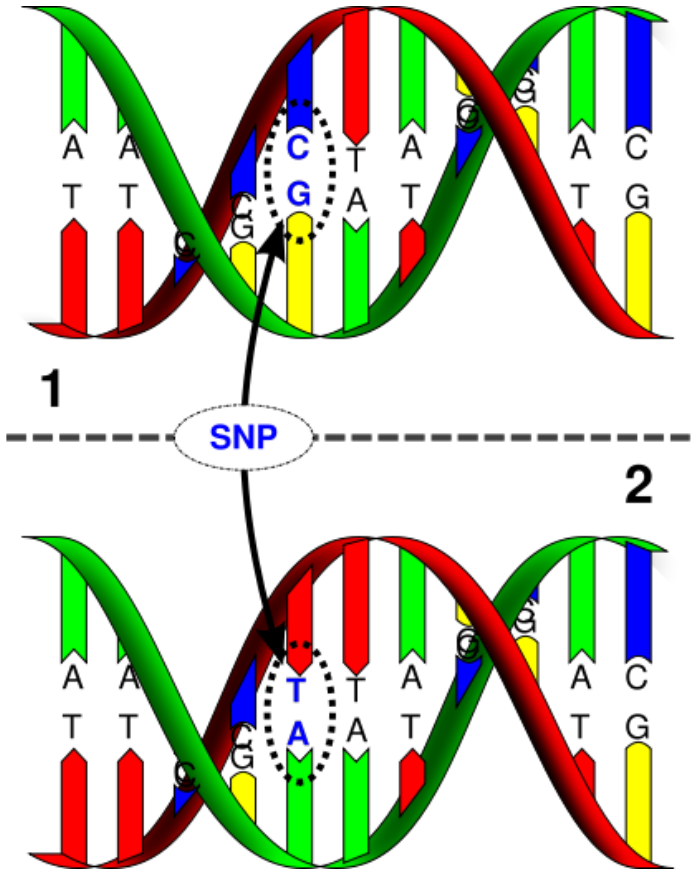




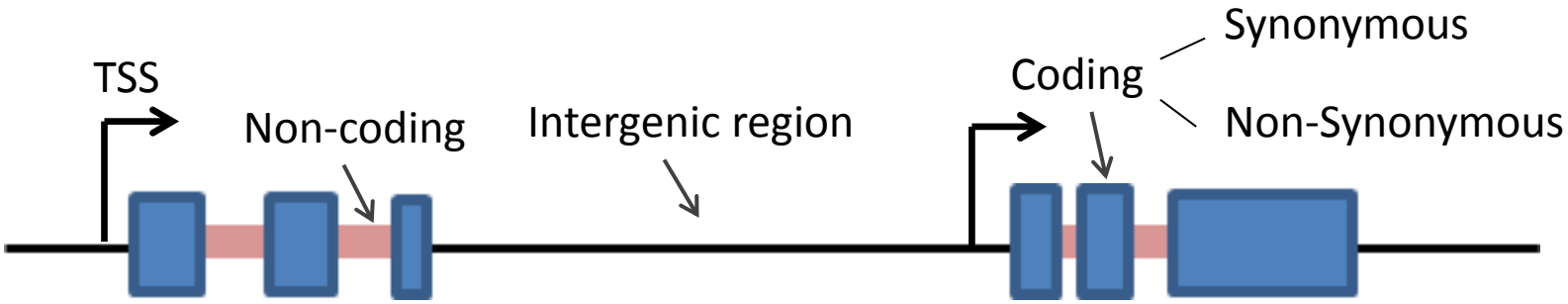


Sequence variation

- 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.

Effects of sequence variation

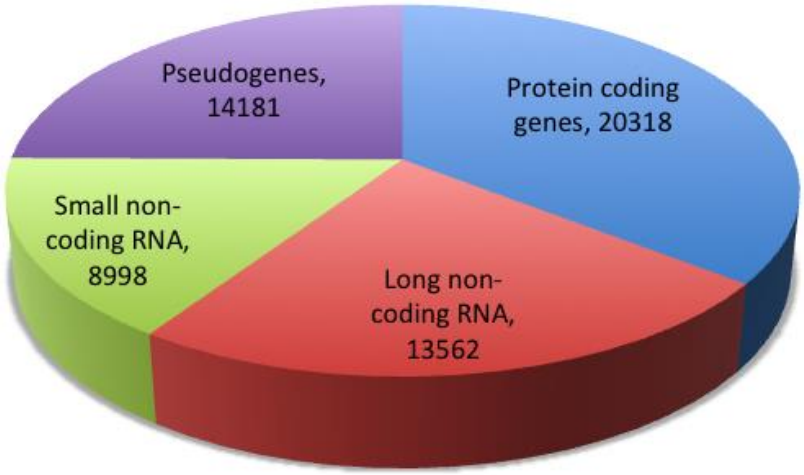
- Non-synonymous variants:
  - Missense (change protein structure)
  - Nonsense (truncates protein)
- Synonymous or non-coding variants:
  - Alter transcriptional/translational efficiency
  - Alter mRNA stability
  - Alter gene regulation (i.e. alter TF binding)
  - Alter RNA-regulation (i.e. affect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

Majority of sequence variation are neutral

RNA: Genes and gene expression

Types of genes

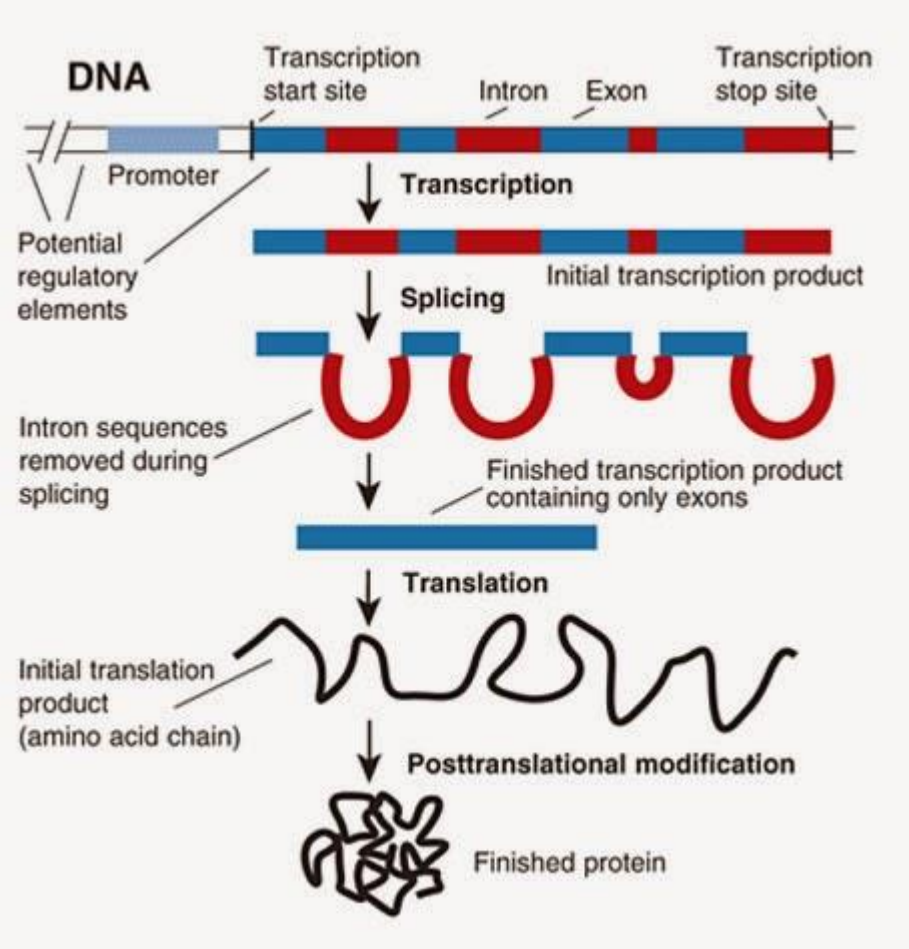
- A gene is a functional unit of DNA that is transcribed into RNA.
- Total genes in the human genome – 57,445



mRNA
miRNA
lncRNA

Source: GENCODE (version 18)

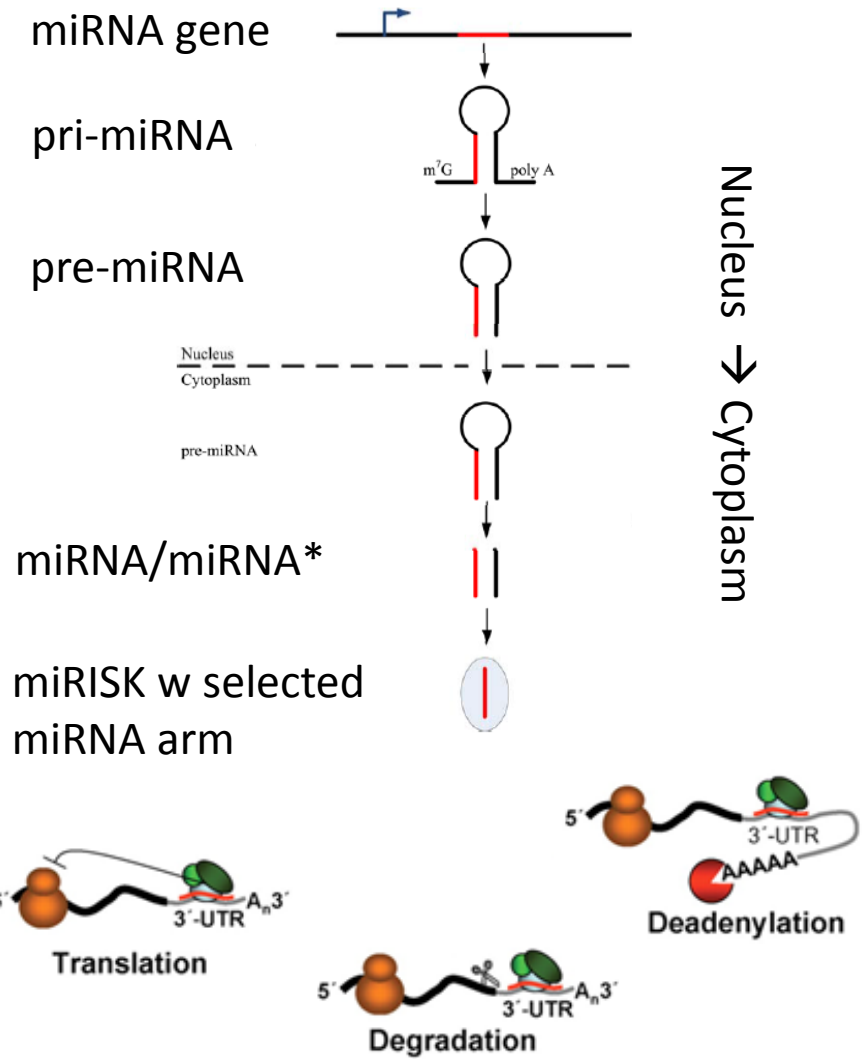
Protein coding genes



- 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 PolII 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

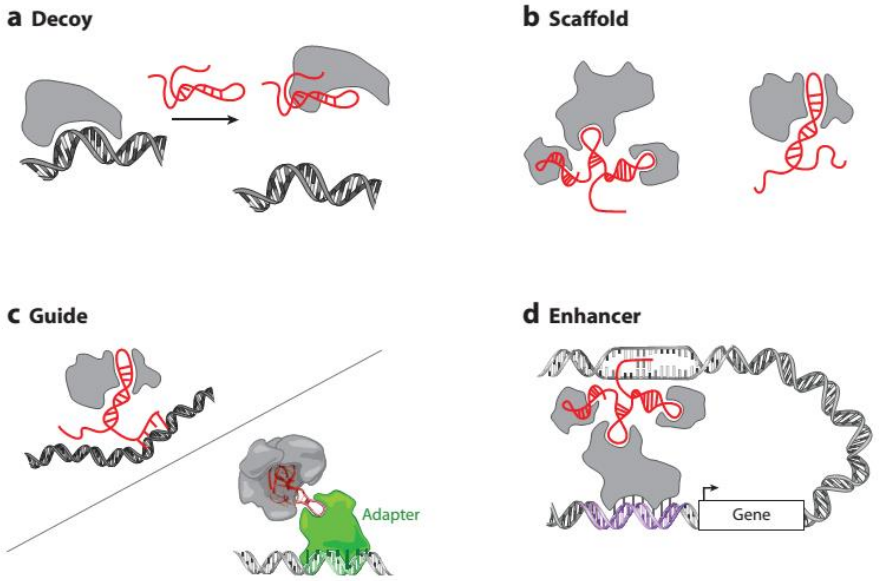
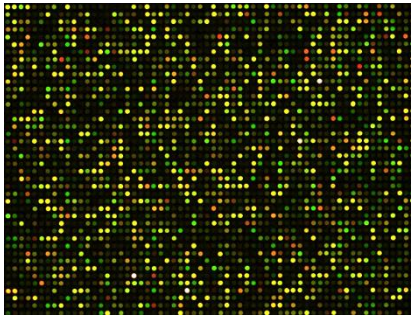


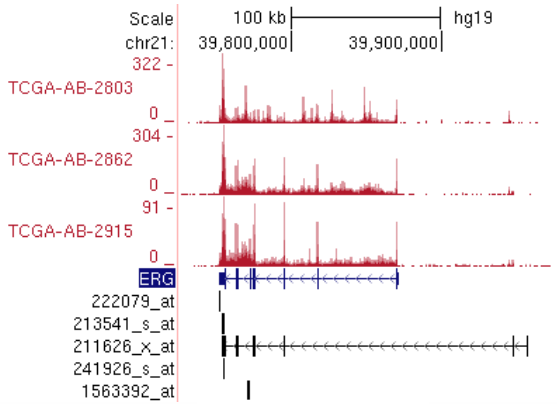
Figure 4

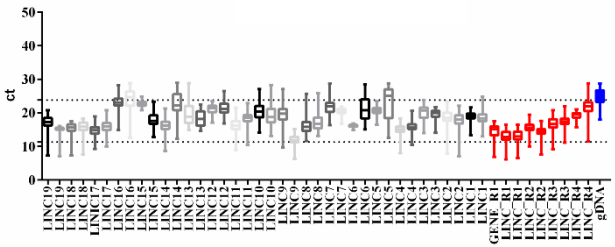
Models of long noncoding RNA (lncRNA) mechanisms of action. (*a*) The lncRNAs can act as decoys that titrate away DNA-binding proteins, such as transcription factors. (*b*) These lncRNAs may act as scaffolds to bring two or more proteins into a complex or spatial proximity and (*c*) may also act as guides 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



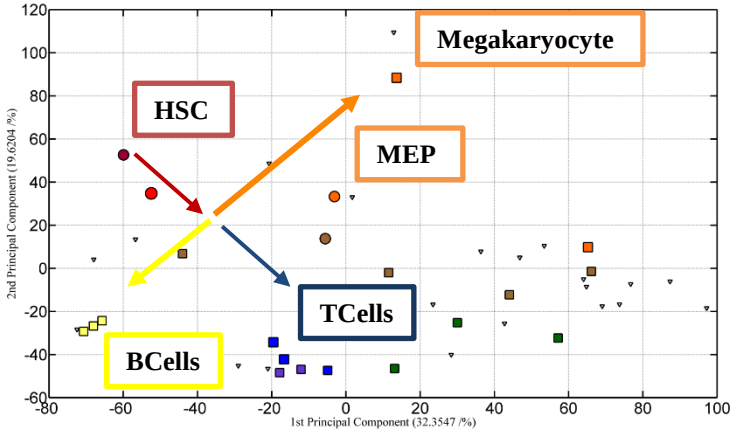
- Measuring the level of RNA in the sample.
- Generally microarray-, sequencing- or high-throughput PCR- based.
- Computation analysis and normalisation of expression data can be complicated.





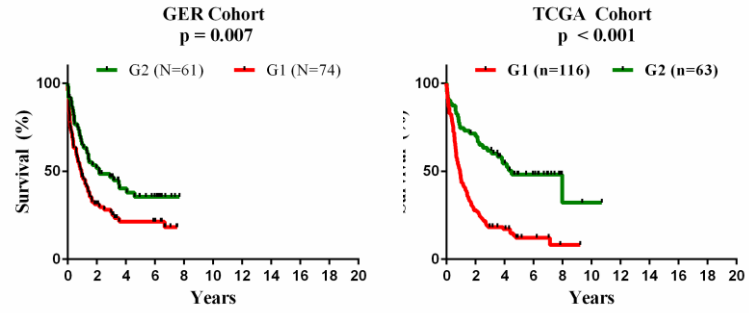
RNA expression applications

- Relatively cheap and fast readout of the functional state of a cell

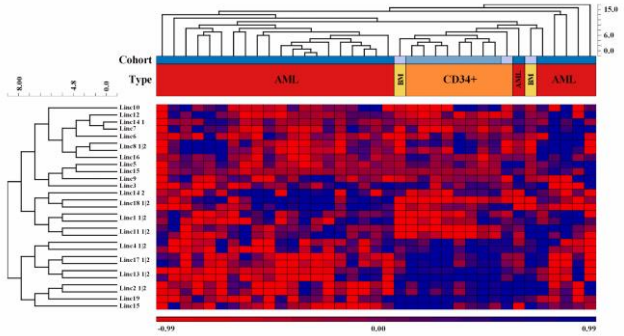


- Association with clinical features
  - sequence variations
  - response to therapy
  - patient survival

  …

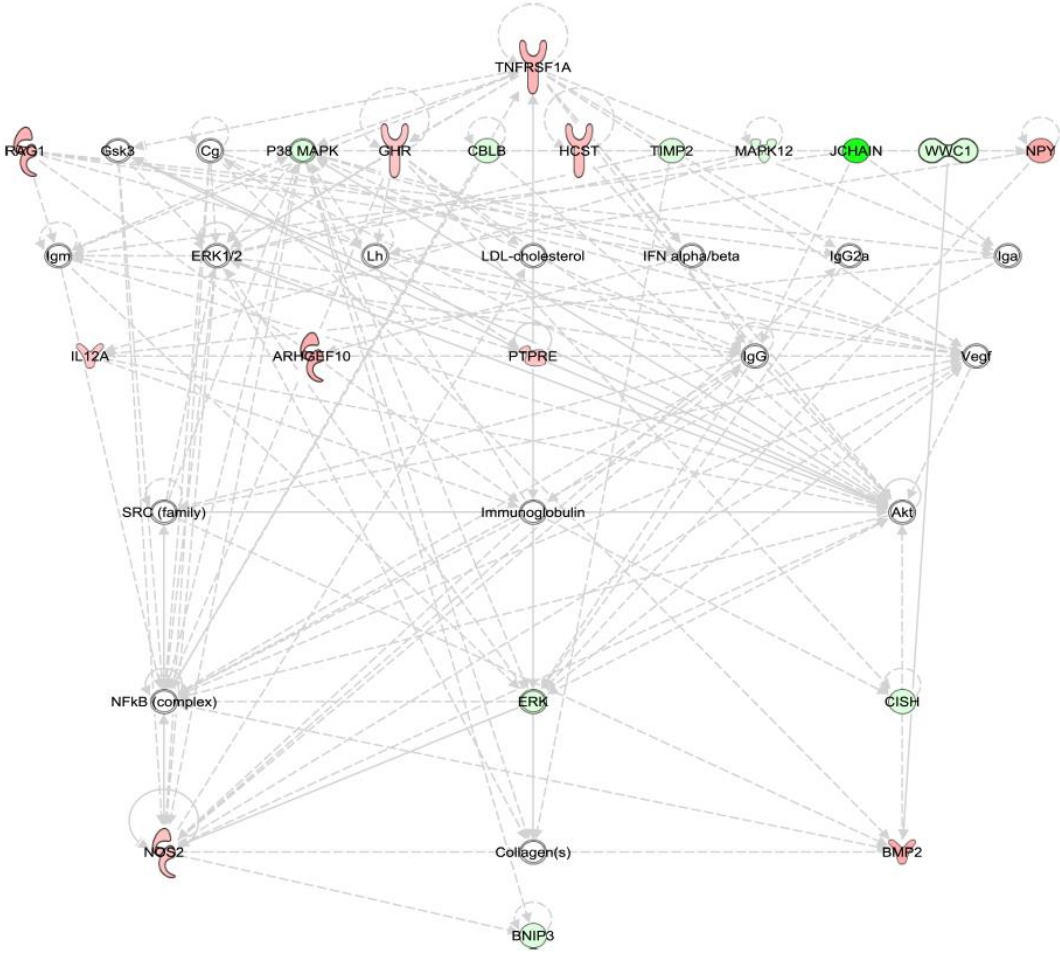


- Differential expression
  - between samples, or
  - between genes



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

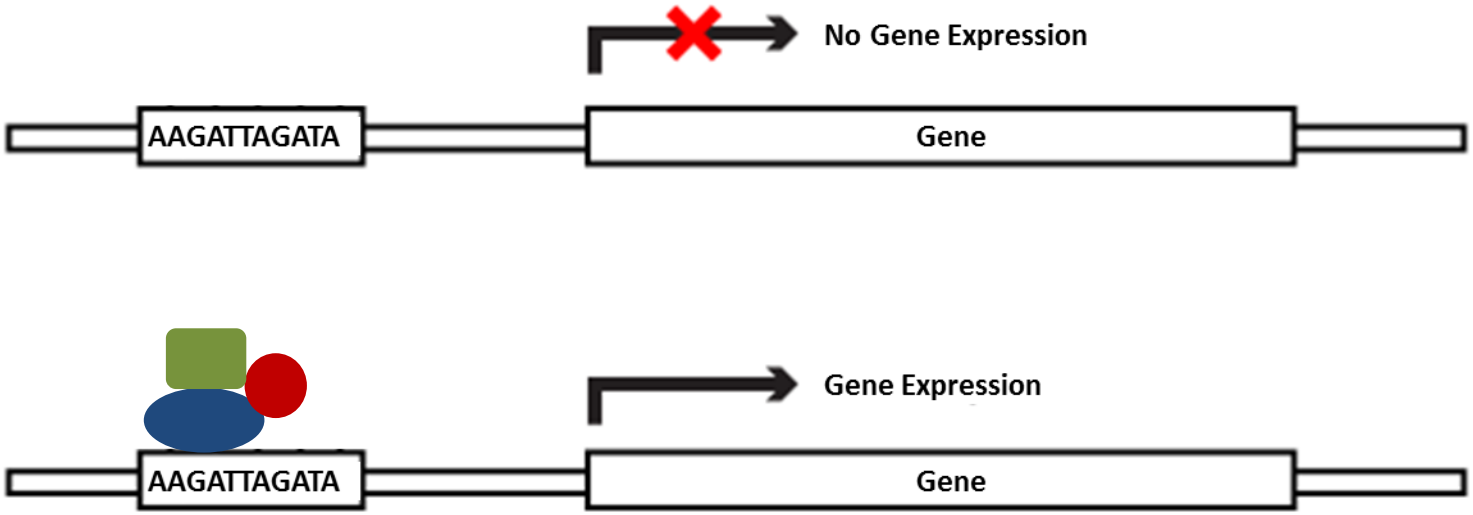
Epigenetics

- 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*

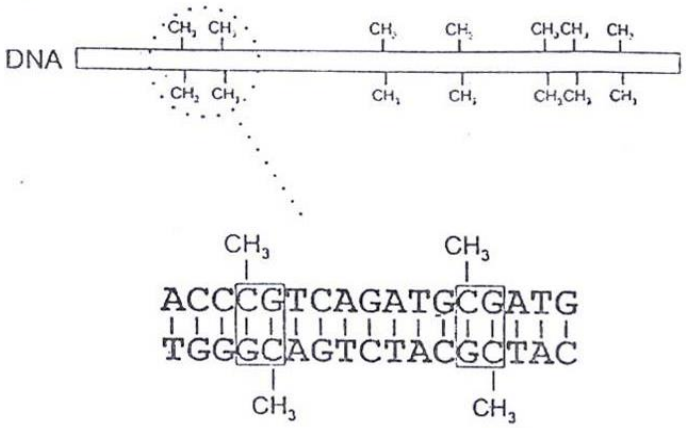
Transcriptional regulation

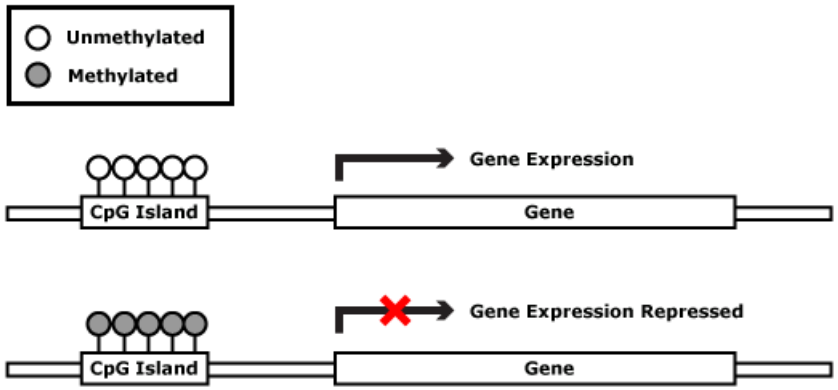
- Transcription factors are proteins that bind DNA to co-regulate gene expression.
- Typically binds at gene promoters or enhancers.



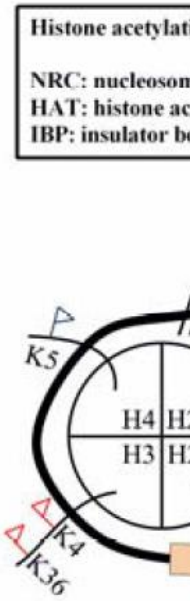
DNA methylation

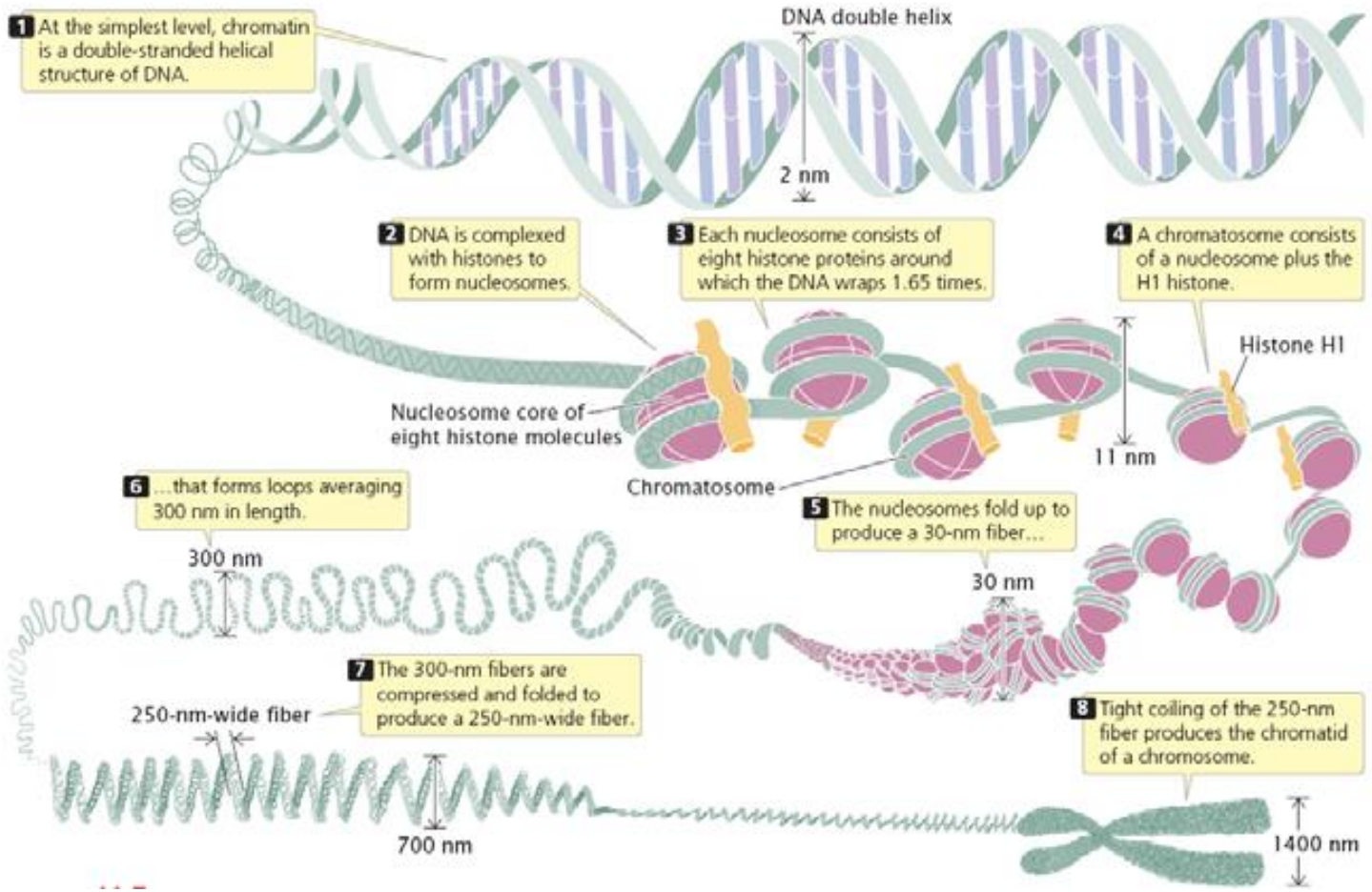
- DNA is methylated on cytosine's in CpG dinucleotides





Nucleosomes & Histones





n
on
ylation
ation

What we will cover

Structure of the human genome

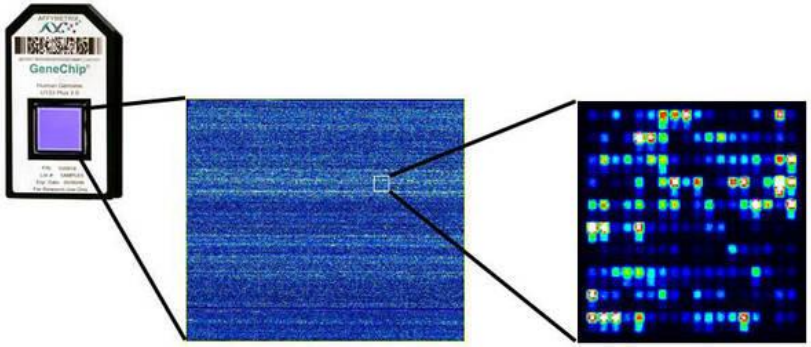
Genomic information

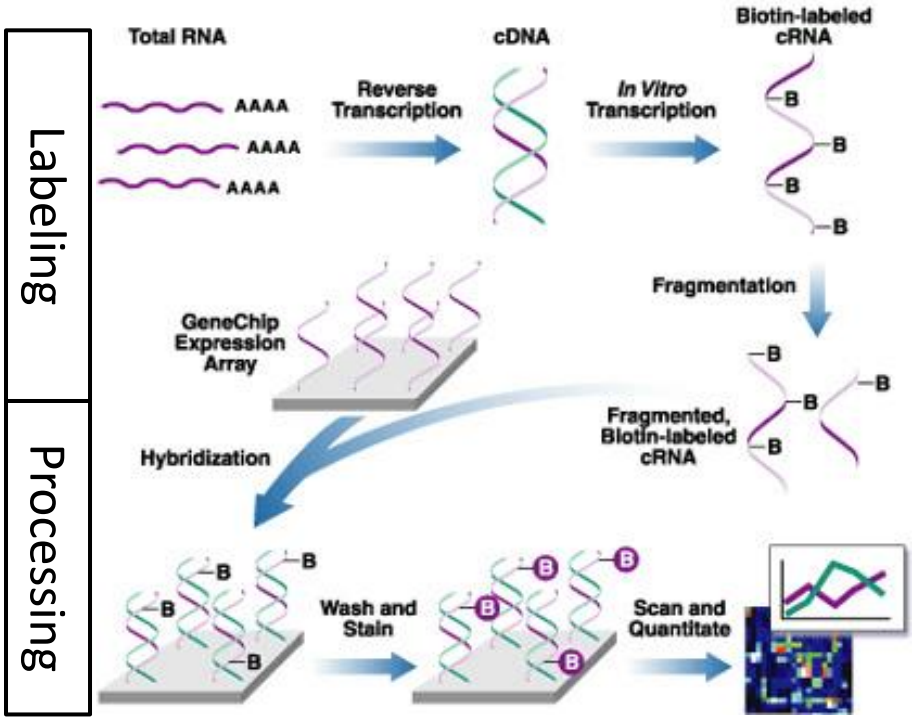
- DNA (Sequence variation)
- RNA (Genes & gene expression)
- Epigenetics
  - DNA methylation
  - Histone modification
  - Transcription factor binding

Data acquisition

- Microarrays
- Sequencing
- Chromatin IP

Array Technology

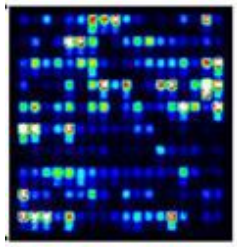




- Relies on fluorescence-based on hybridisation of DNA against complementary probe on array.
- Known molecule that can be converted to cDNA.
  - Expression array (probe for exonic DNA regions)
  - SNP array (probe for two alleles)
  - Methylation array (probe for bisulfide converted DNA)
- Limited by probes present on the array.

https://www.dkfz.de/gpcf/affymetrix_genechips.html

Array Technology

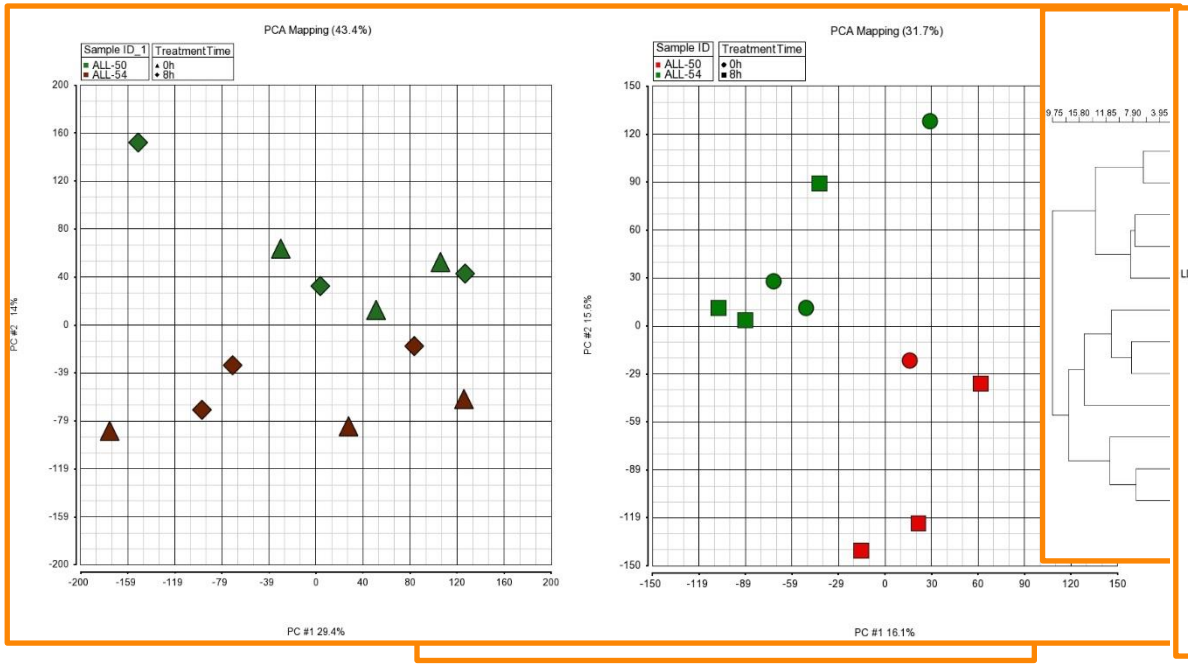


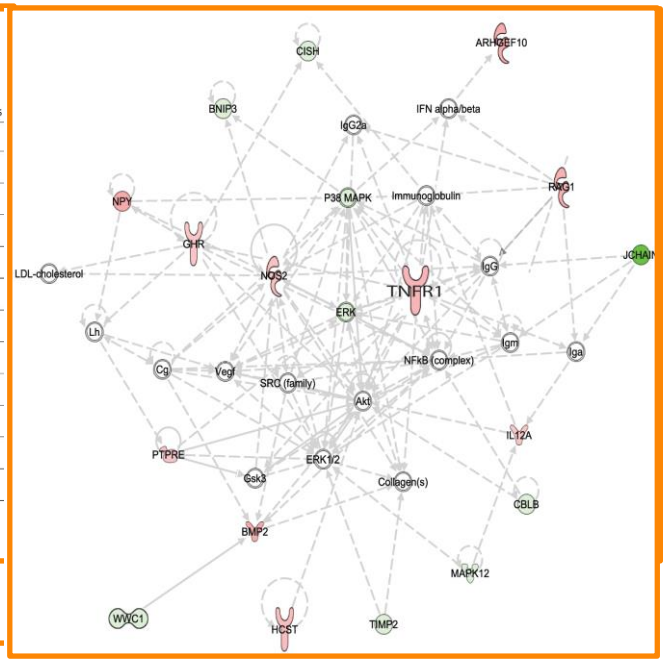
Images Processing

Quantification

Pre-processing

Backgrd. Subs., Norm.

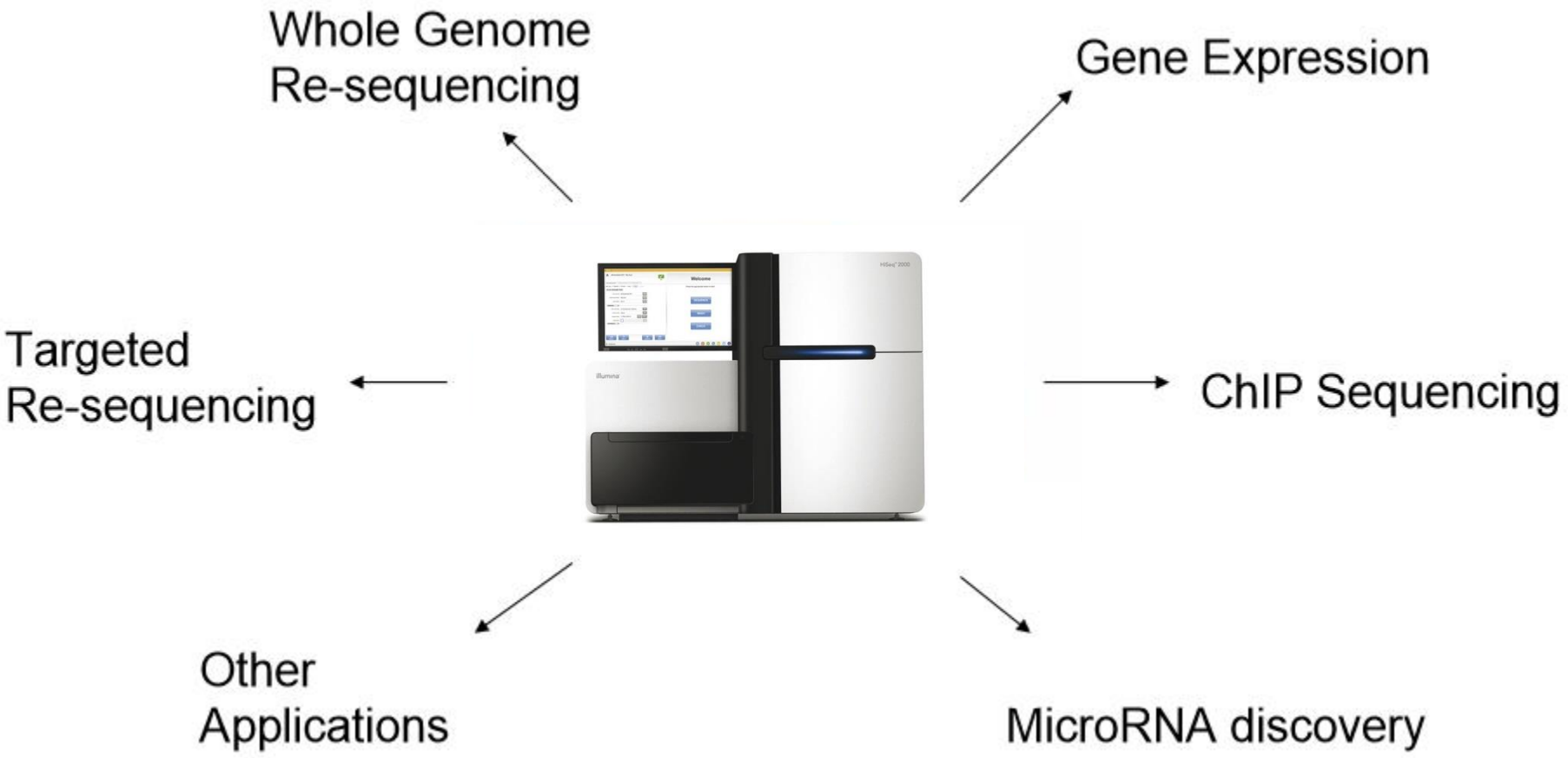




https://www.dkfz.de/gpcf/affymetrix_genechips.html

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Next-generation sequencing



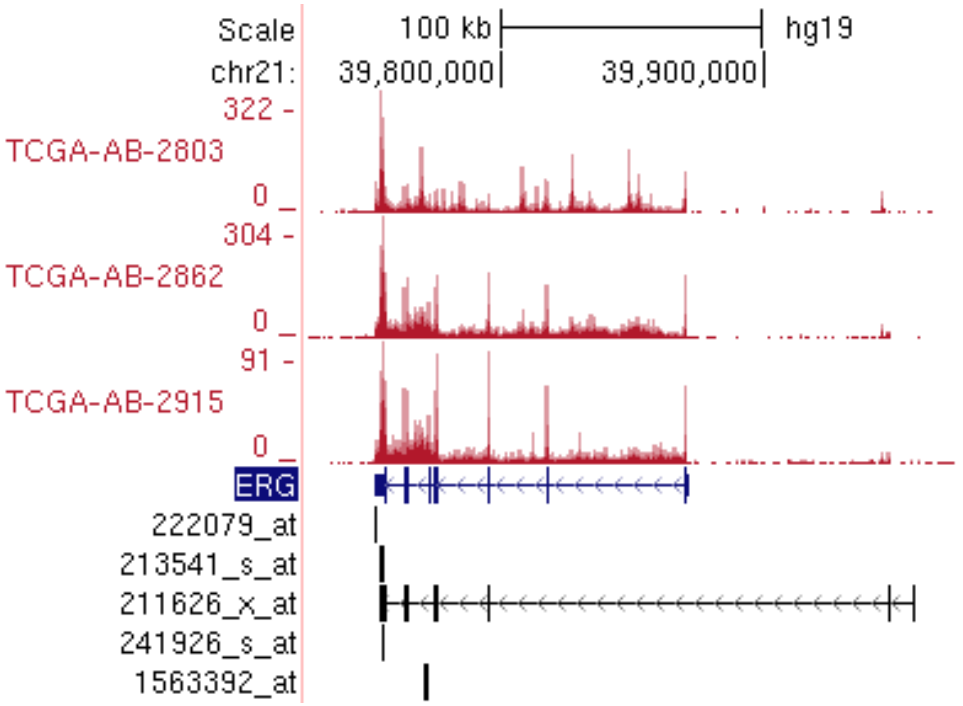
Next-generation sequencing (Illumina)

Next-generation sequencing (Illumina)

Alignment
human reference genome

Quantification
mRNA/miRNA/lncRNA

Quantification
mRNA/miRNA/lncRNA



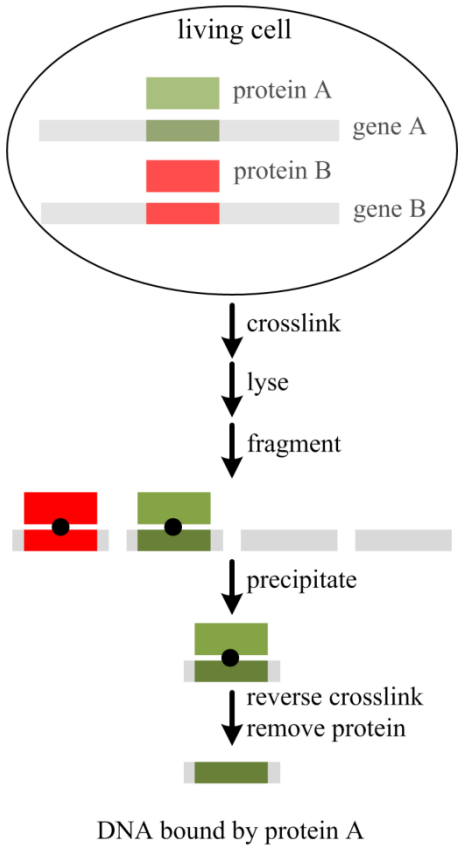
Pros/cons of each technology

- 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)

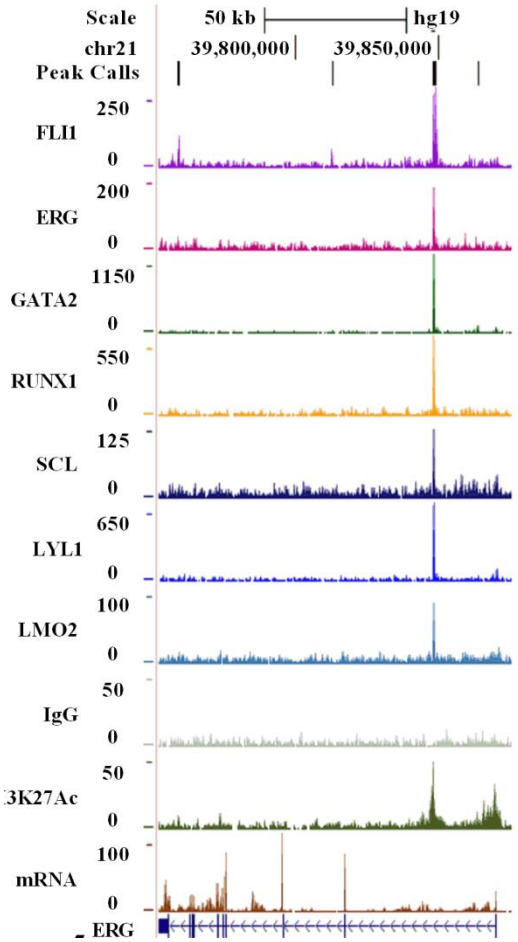
ChIP-seq of the seven transcription factors FLI1, ERG, GATA2, RUNX1, SCL, LYL1 and LMO



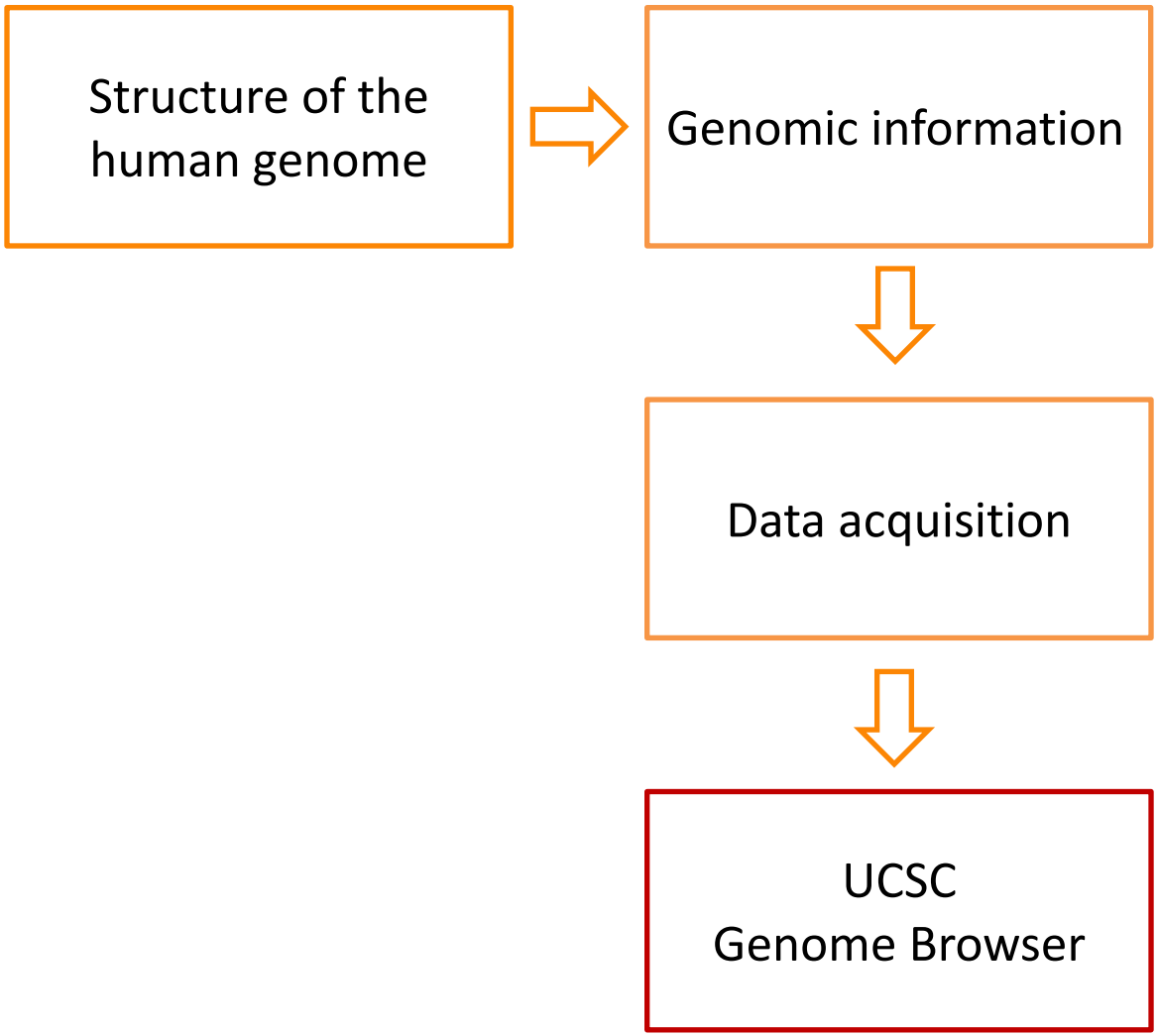
High- throughput sequencing

Bioinformatics

ERG locus



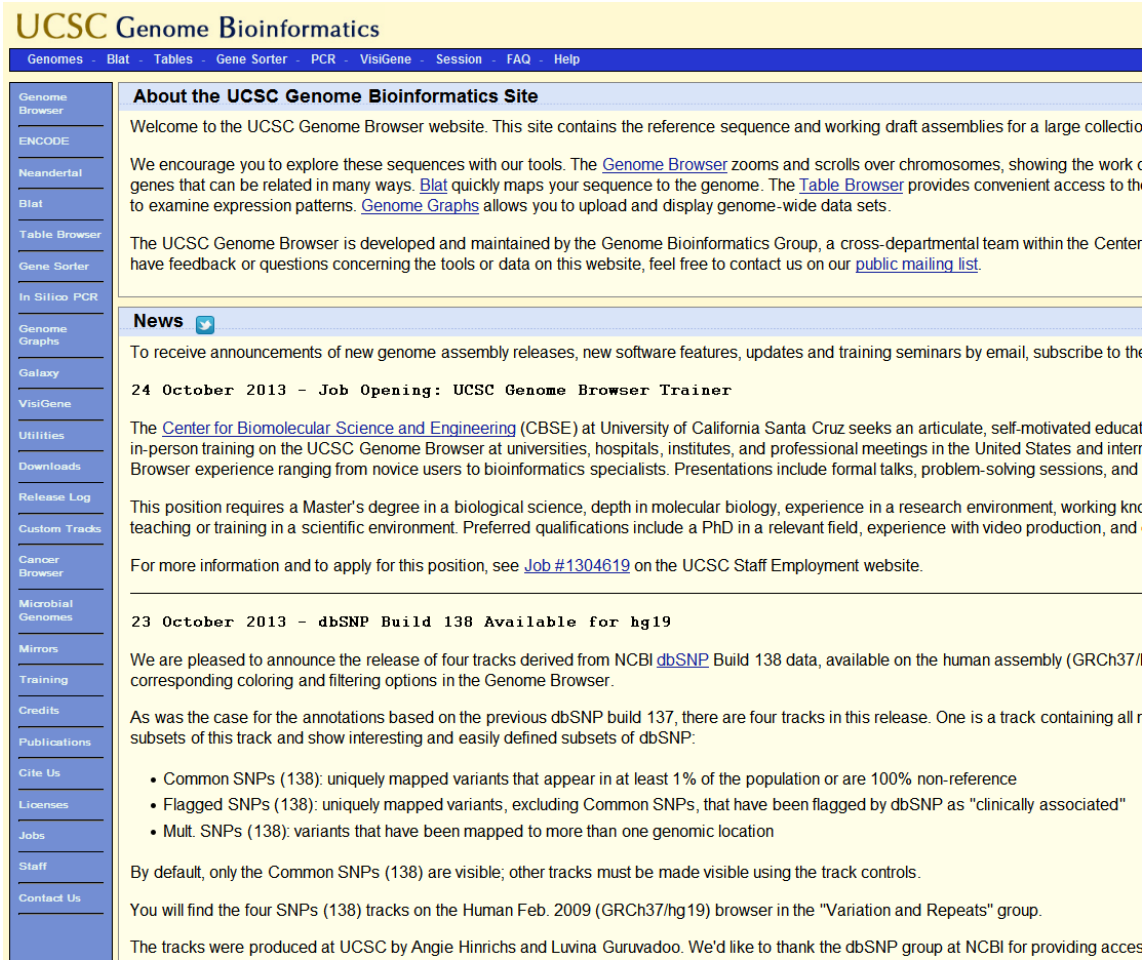
What we will cover



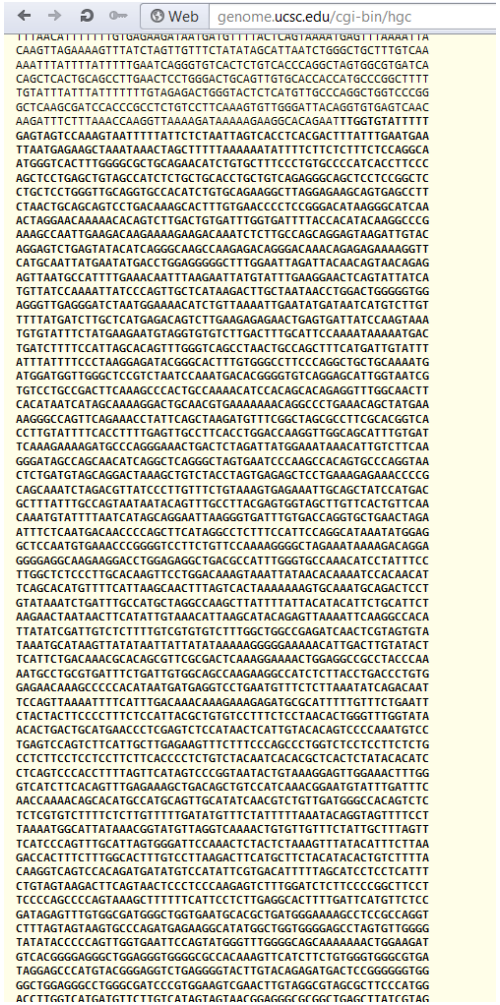
- Background
- Genome Assemblies
- Annotation Tracks
- Associated Tools
- Practical Exercise

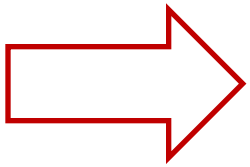
Genome Browser

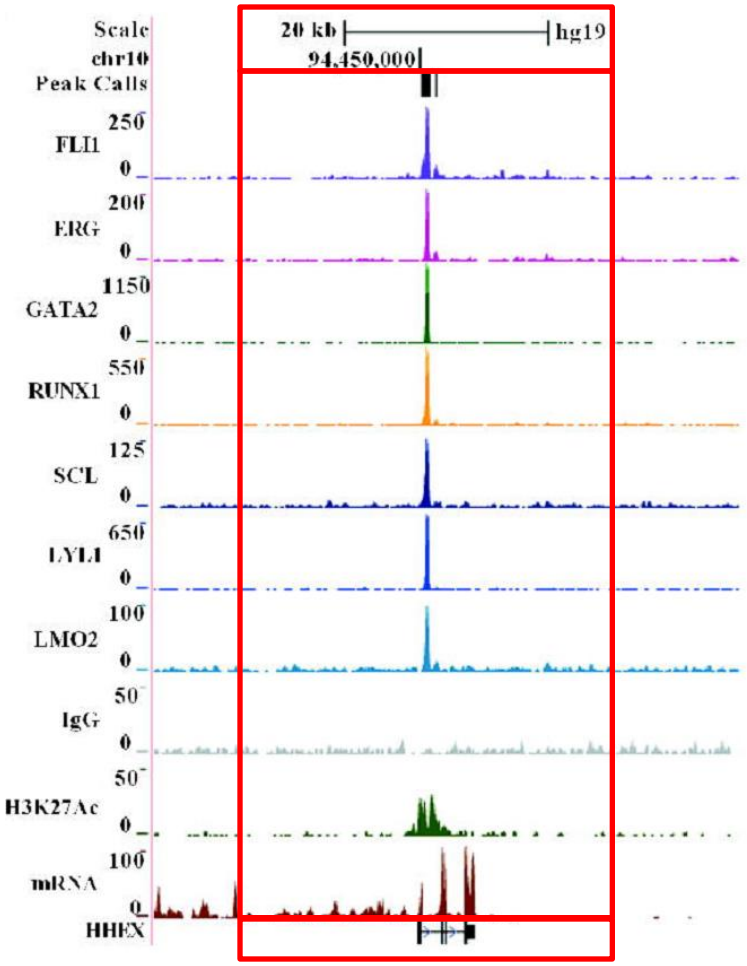
http://genome.ucsc.edu/



Background


Web
genome.ucsc.edu/cgi-bin/hgc




Scale
chr10
Peak Calls
20 kb
hg19
94,450,000
FLI1
250
0
ERG
200
0
GATA2
1150
0
RUNX1
550
0
SCL
125
0
LYL1
650
0
LMO2
100
0
IgG
50
0
H3K27Ac
50
0
mRNA
100
0
HHEX

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

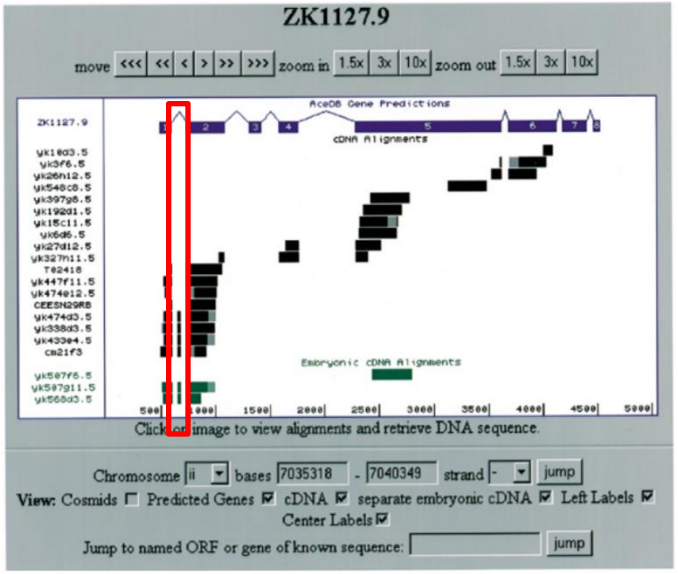
http://www.broadinstitute.org/igv/

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Background

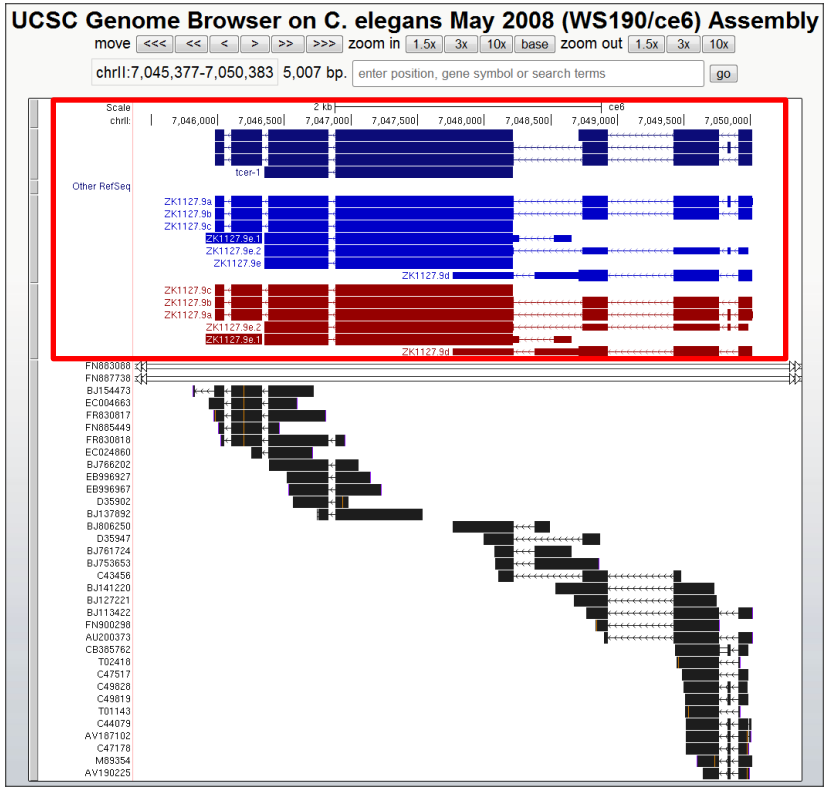
- Intronerator was developed by J. Kent to map the exon–intron structure of C. elegans RNAs mapped against genomic coordinates

Jim Kent



Background

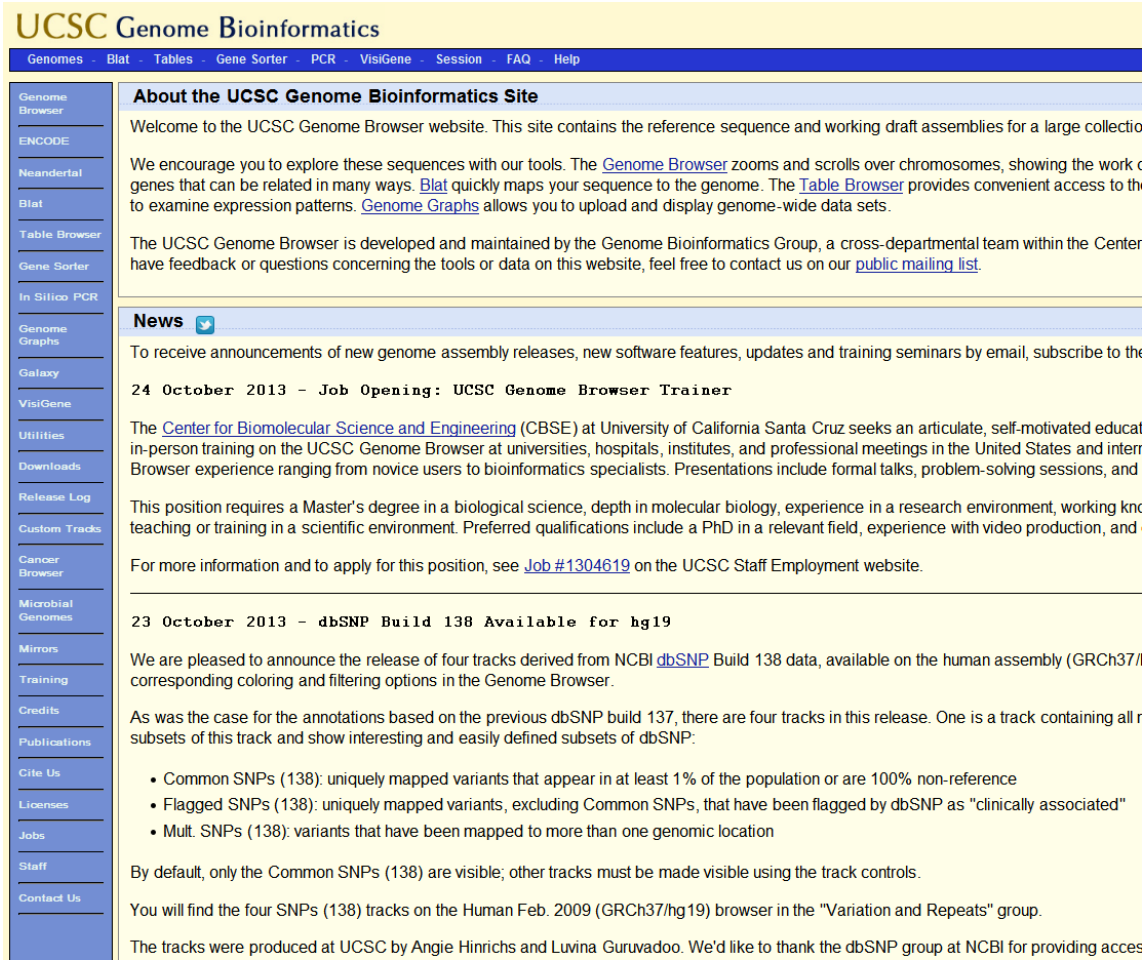
- Draft human genome sequence became available at the UCSC in 2000
- Intronerator was used as the graphics engine



UCSC Genome Browser

Genome Browser

http://genome.ucsc.edu/



Human (*Homo sapiens*) Genome Browser Gateway

group	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr10:123,227,429-123,343,066	FGFR2 submit

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

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

track search | manage custom tracks | track hubs | configure tracks and display

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.

Homo sapiens
(Graphic courtesy of CBSE)

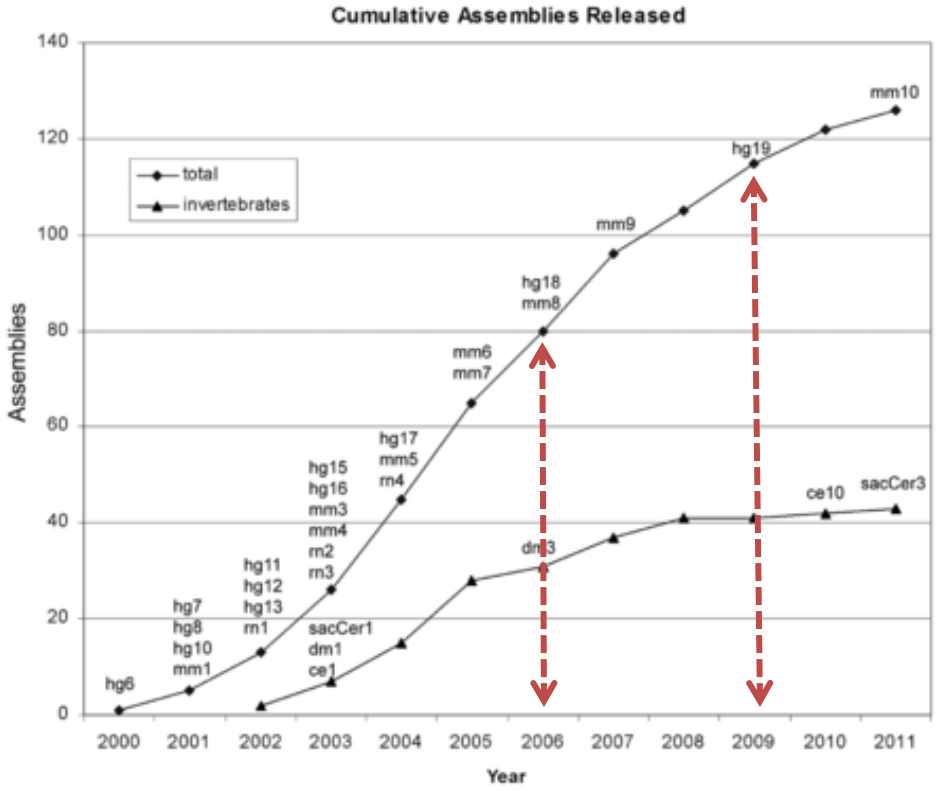
Sample position queries

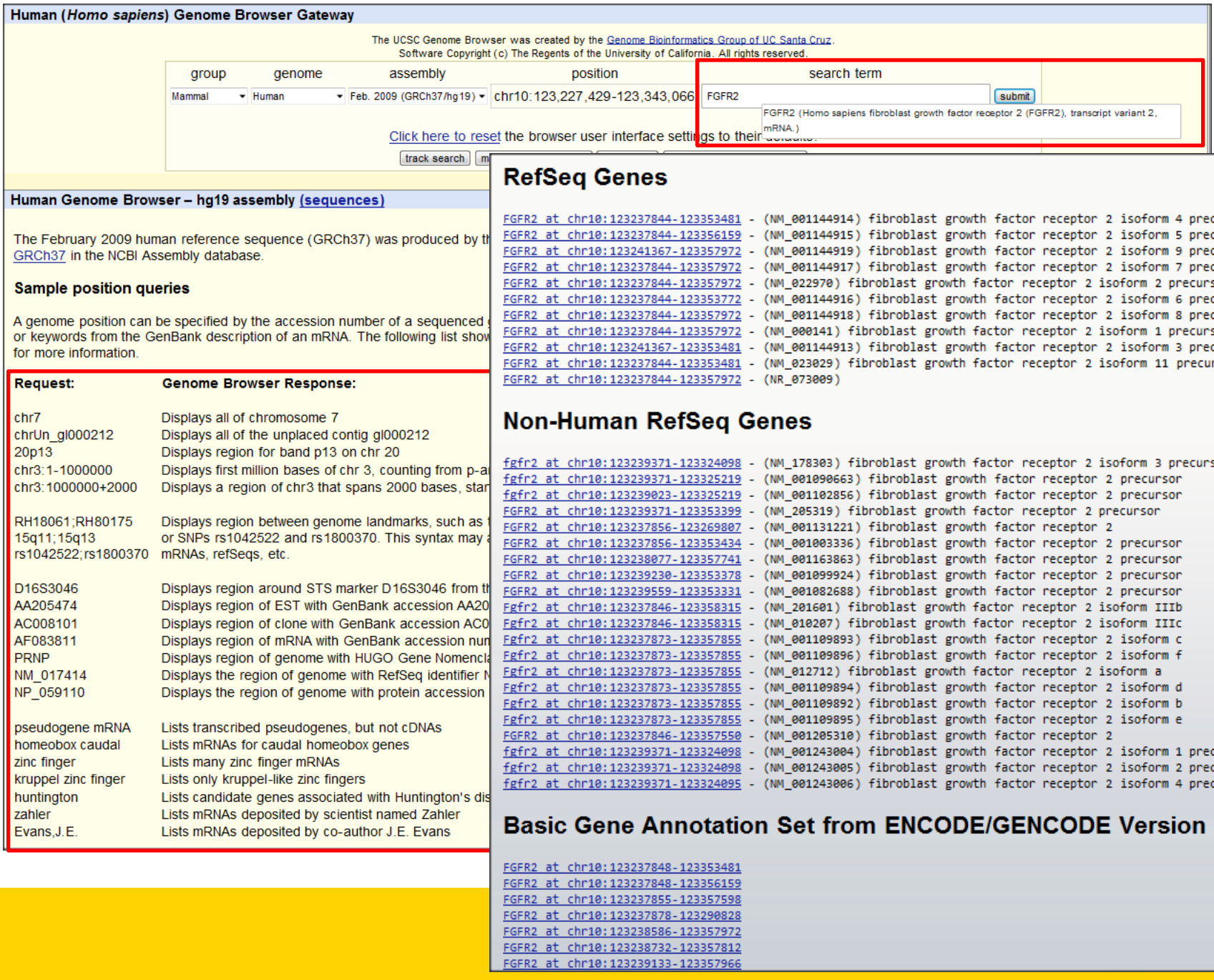
A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's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

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
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	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.
15q11;15q13	
rs1042522;rs1800370	
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

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: Mammal | genome: Human | assembly: Feb. 2009 (GRCh37/hg19) | position: chr10:123,227,429-123,343,066 | search term: FGFR2 submit

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

Click here to reset the browser user interface settings to their

track search

Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by th
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 show
for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-a
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, star
RH18061;RH80175	Displays region between genome landmarks, such as
15q11;15q13	or SNPs rs1042522 and rs1800370. This syntax may
rs1042522;rs1800370	mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from th
AA205474	Displays region of EST with GenBank accession AA20
AC008101	Displays region of clone with GenBank accession AC0
AF083811	Displays region of mRNA with GenBank accession nun
PRNP	Displays region of genome with HUGO Gene Nomencla
NM_017414	Displays the region of genome with RefSeq identifier N
NP_059110	Displays the region of genome with protein accession
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 dis
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans

RefSeq Genes

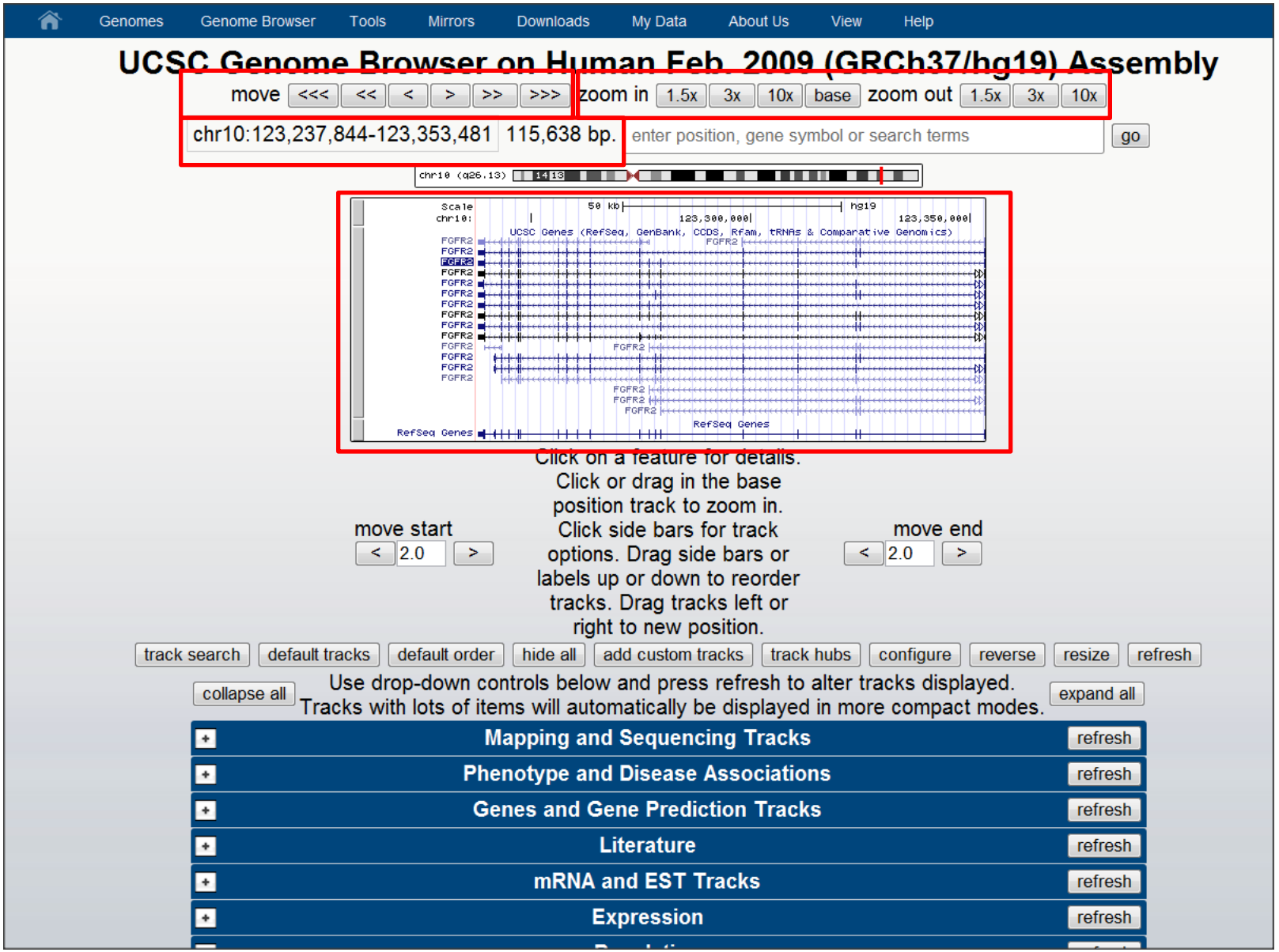
FGFR2 at chr10:123237844-123353481 - (NM_001144914) fibroblast growth factor receptor 2 isoform 4 precursor
FGFR2 at chr10:123237844-123356159 - (NM_001144915) fibroblast growth factor receptor 2 isoform 5 precursor
FGFR2 at chr10:123241367-123357972 - (NM_001144919) fibroblast growth factor receptor 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 chr10:123239371-123324095 - (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

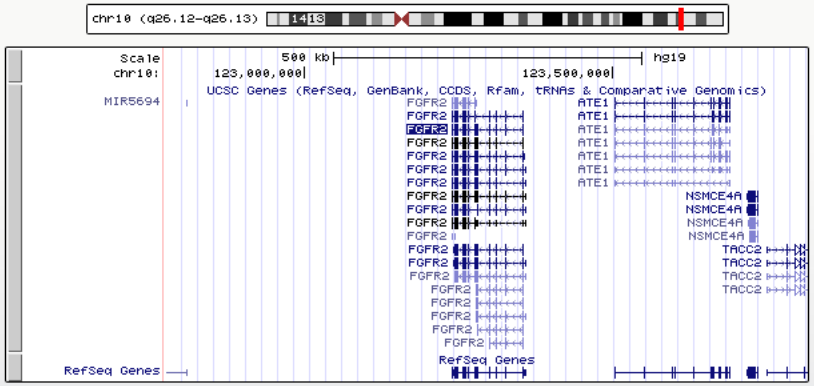


Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | About Us | View | Help

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. 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 >

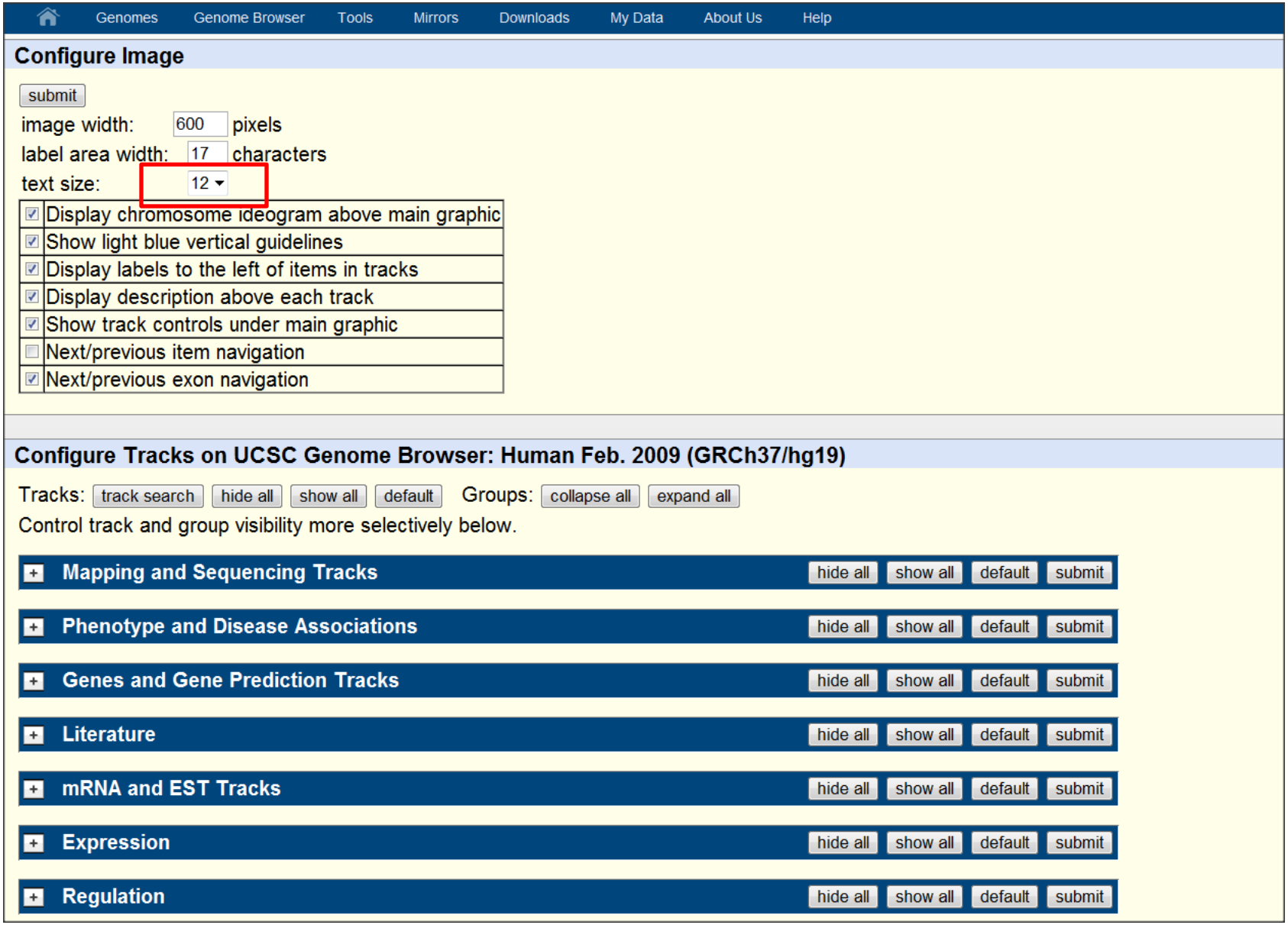
track search | default tracks | default order | hide all | add custom tracks | track hubs | configure | reverse | resize | refresh

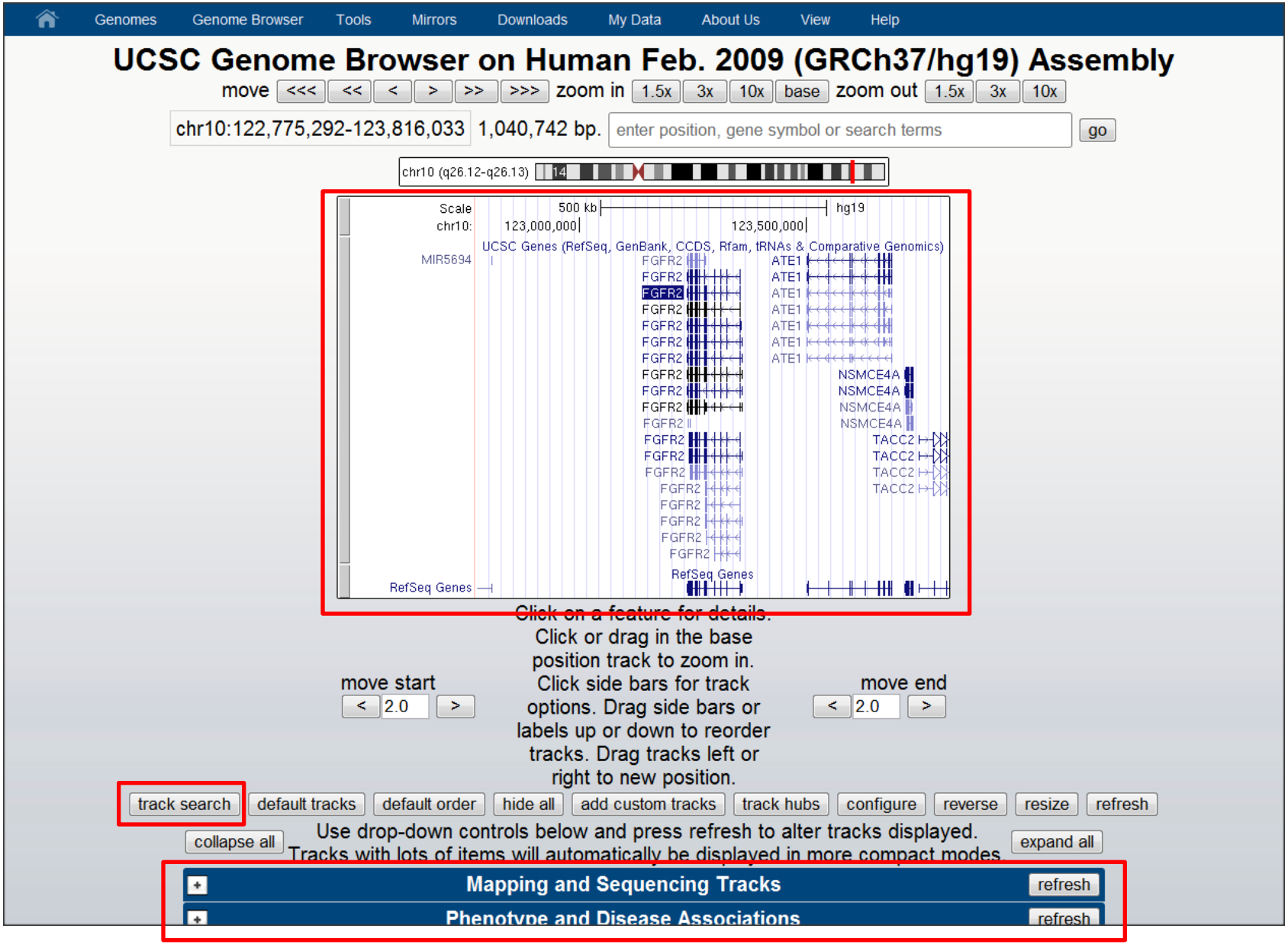
collapse all

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.

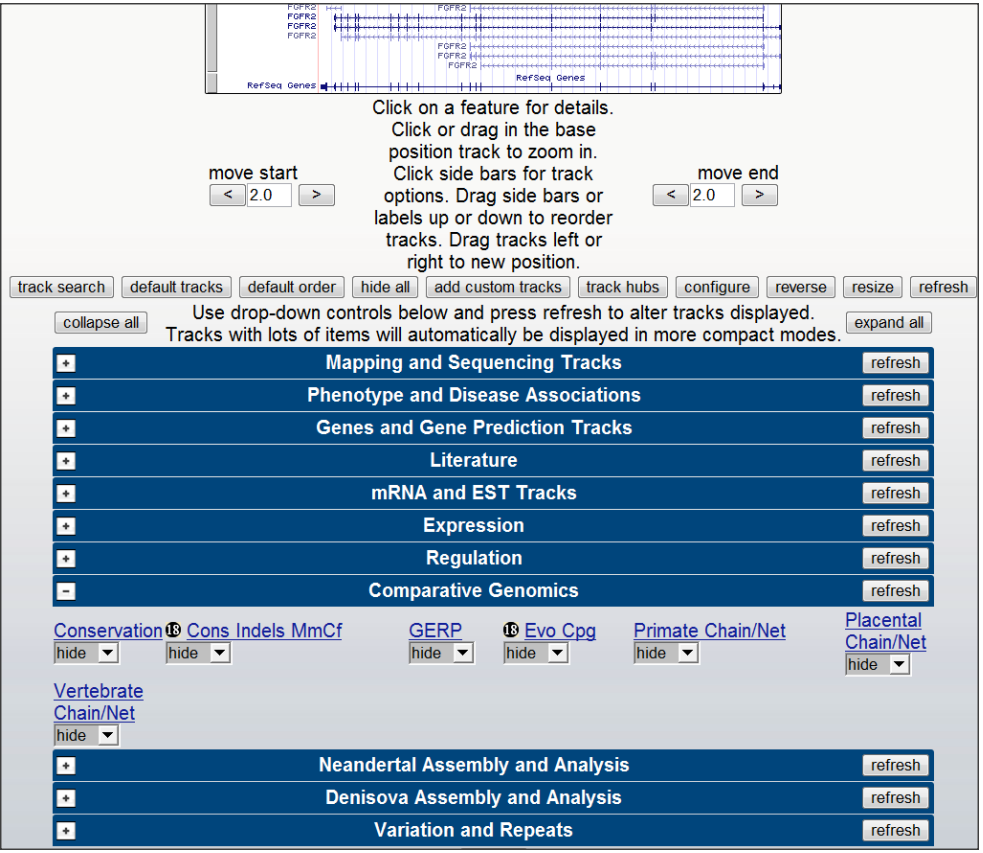
expand all

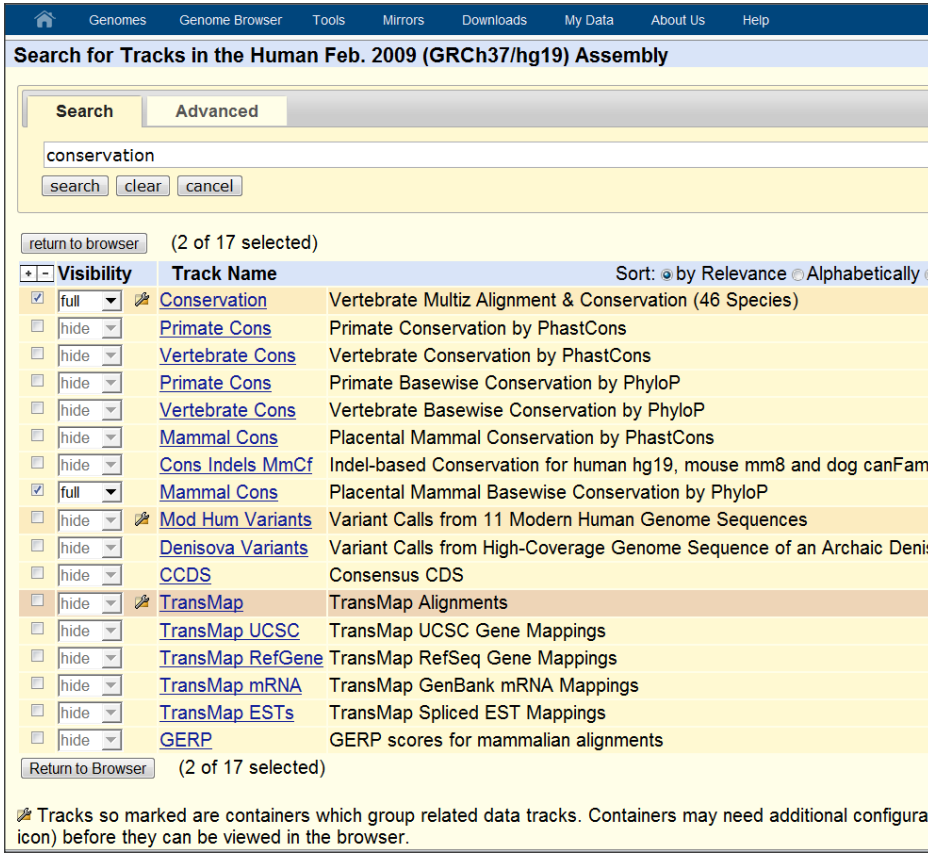
- Mapping and Sequencing Tracks — refresh
- Phenotype and Disease Associations — refresh
- Genes and Gene Prediction Tracks — refresh
- Literature — refresh
- mRNA and EST Tracks — refresh
- Expression — refresh




Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
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.
enter position, gene symbol or search terms
go
chr10 (q26.12-q26.13)
Scale
500 kb
hg19
chr10:
123,000,000
123,500,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
MIR5694
FGFR2
ATE1
NSMCE4A
TACC2
RefSeq Genes
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
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

Annotation tracks





UNSW
THE UNIVERSITY OF NEW SOUTH WALES

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

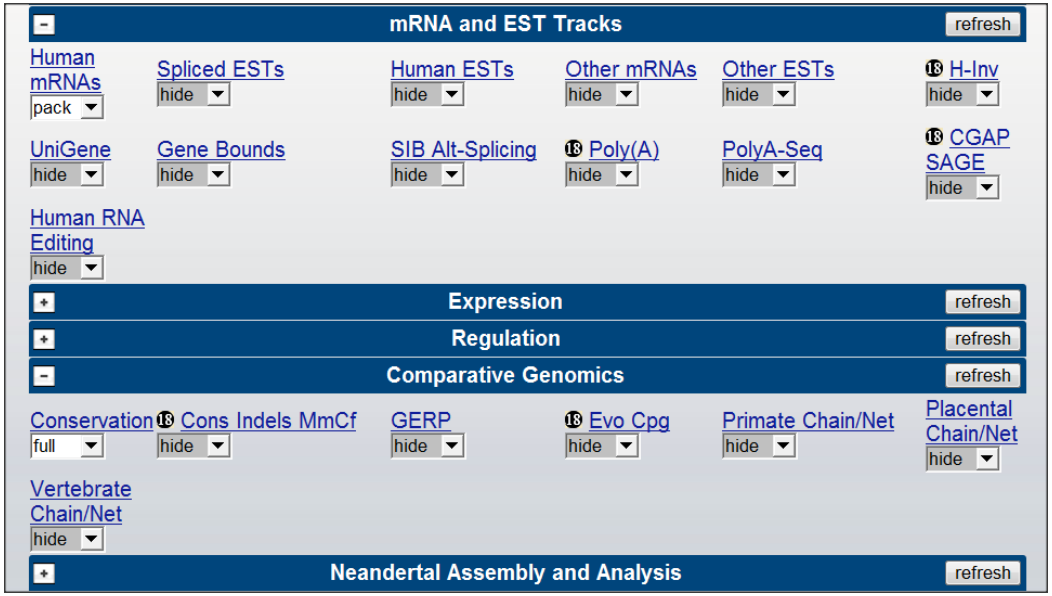
- Three different categories:
  - computed at UCSC
  - computed elsewhere and displayed at UCSC
  - computed and hosted entirely elsewhere

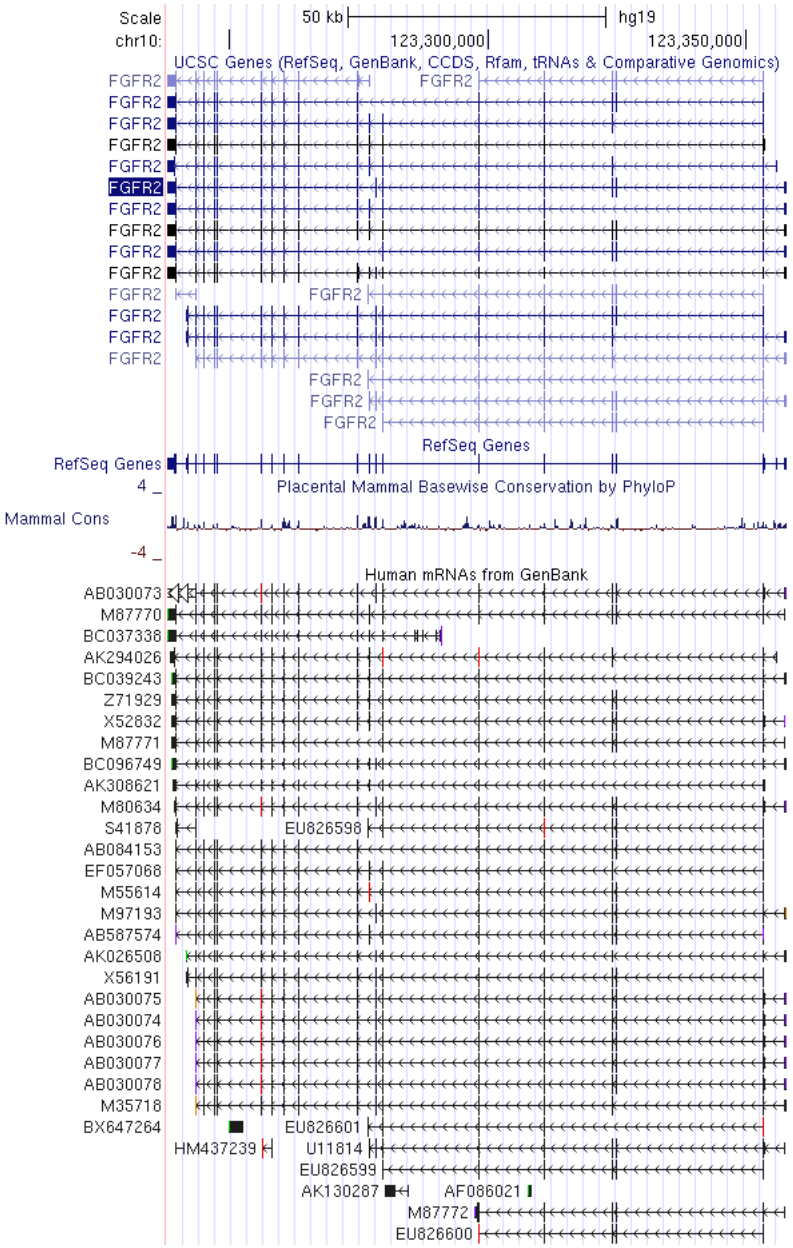
Annotation tracks computed at UCSC

- 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
H-Inv
hide
UniGene
hide
Gene Bounds
hide
SIB Alt-Splicing
hide
Poly(A)
hide
PolyA-Seq
hide
CGAP SAGE
hide
Human RNA Editing
hide
Expression
refresh
Regulation
refresh
Comparative Genomics
refresh
Conservation
full
Cons Indels MmCf
hide
GERP
hide
Evo Cpg
hide
Primate Chain/Net
hide
Placental Chain/Net
hide
Vertebrate Chain/Net
hide
Neandertal Assembly and Analysis
refresh


Scale
50 kb
hg19
chr10:
123,300,000
123,350,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
FGFR2
RefSeq Genes
RefSeq Genes
4
Placental Mammal Basewise Conservation by PhyloP
Mammal Cons
-4
Human mRNAs from GenBank
AB030073
M87770
BC037338
AK294026
BC039243
Z71929
X52832
M87771
BC096749
AK308621
M80634
S41878
EU826598
AB084153
EF057068
M55614
M97193
AB587574
AK026508
X56191
AB030075
AB030074
AB030076
AB030077
AB030078
M35718
BX647264
EU826601
HM437239
U11814
EU826599
AK130287
AF086021
M87772
EU826600

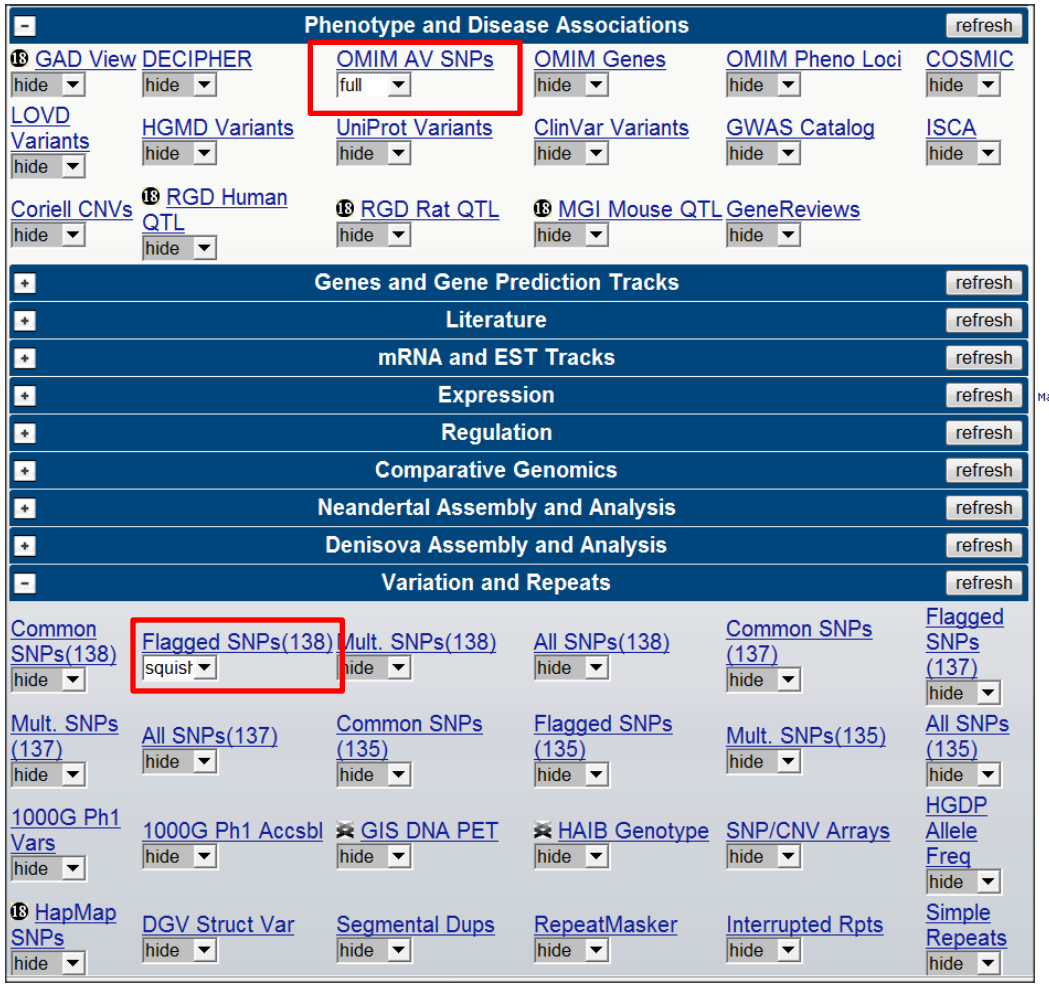
Annotation tracks computed elsewhere and displayed at UCSC

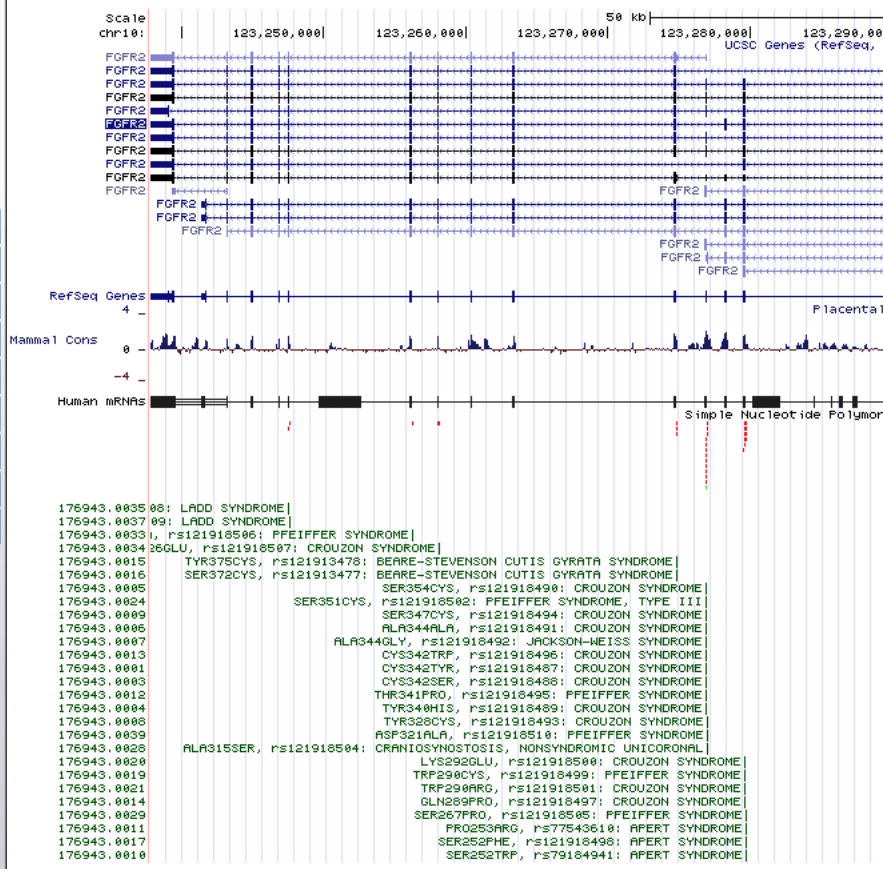
Annotations that are not post-processed by the UCSC

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

Annotations that are post-processed by the UCSC

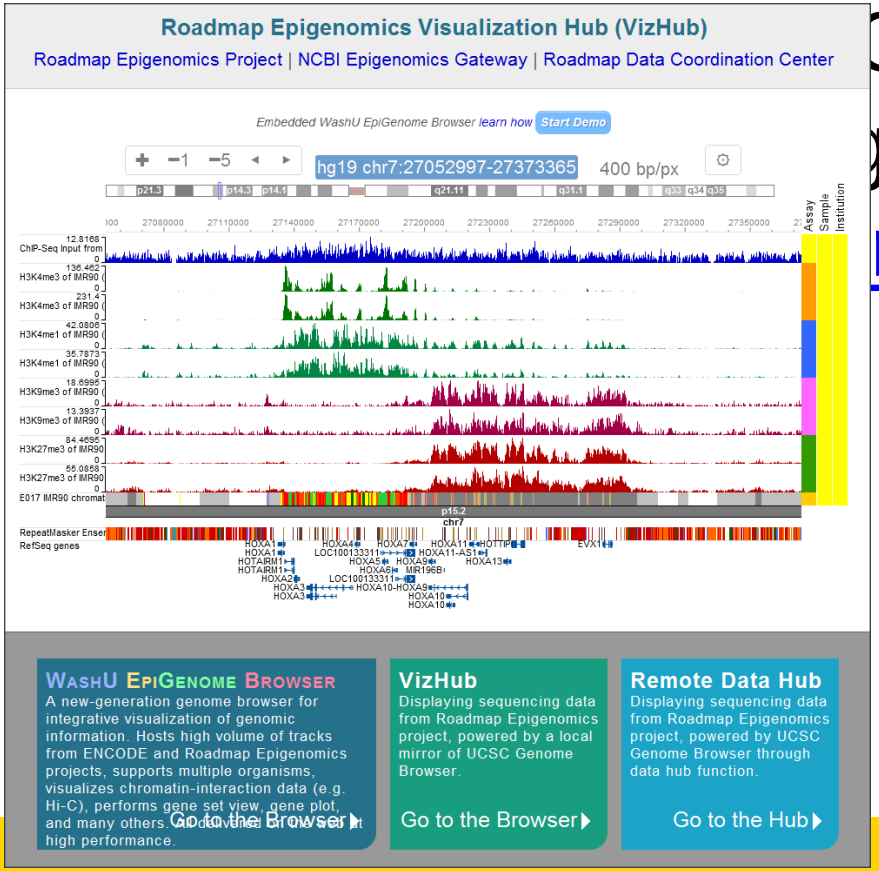
- dbSNP (Common SNPs, Flagged SNPs, Mult. SNPs)
- OMIM (OMIM Allelic Variant SNPs, OMIM Genes, OMIM Phenotypes)

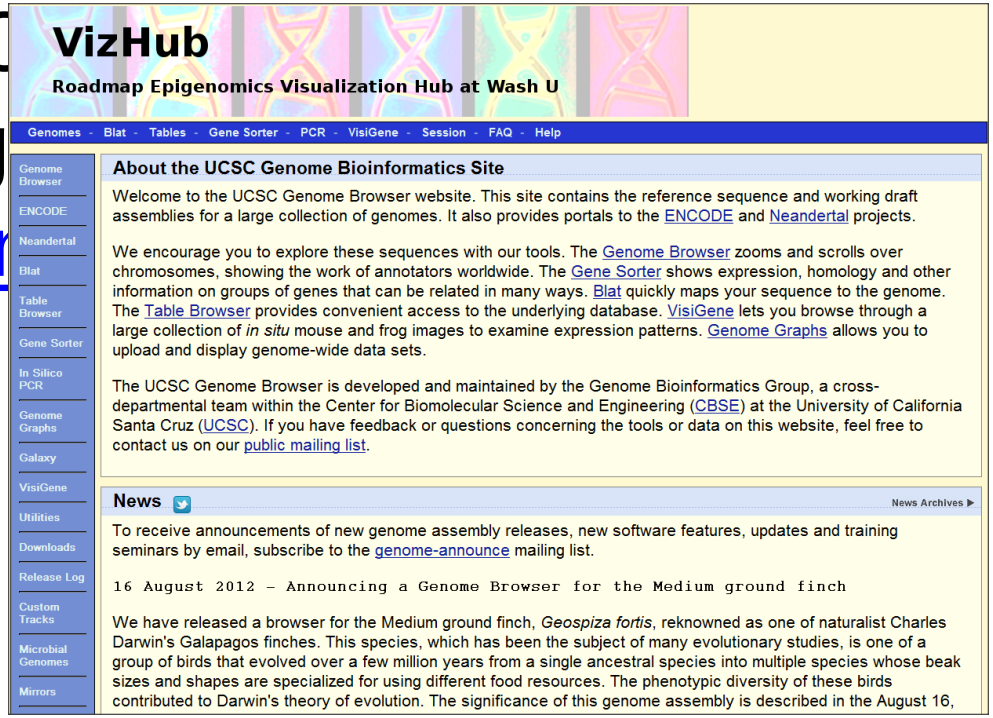




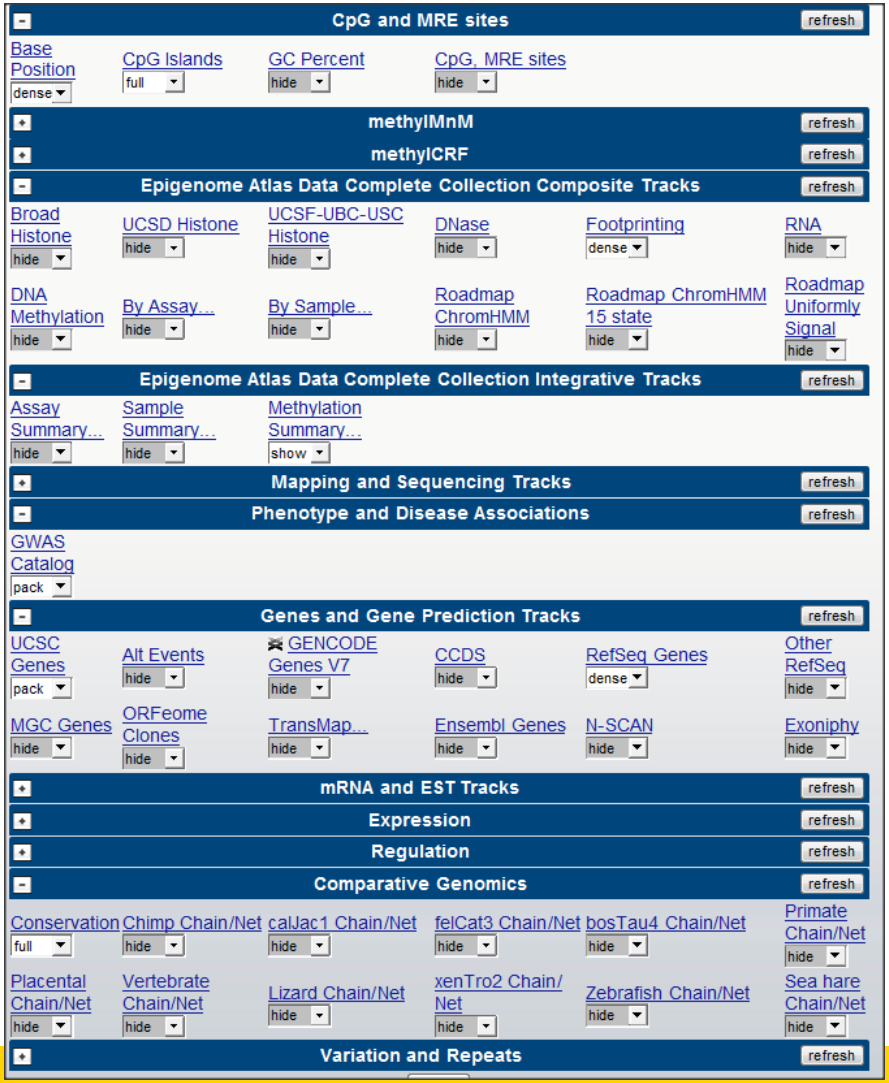
Annotation tracks computed and hosted elsewhere

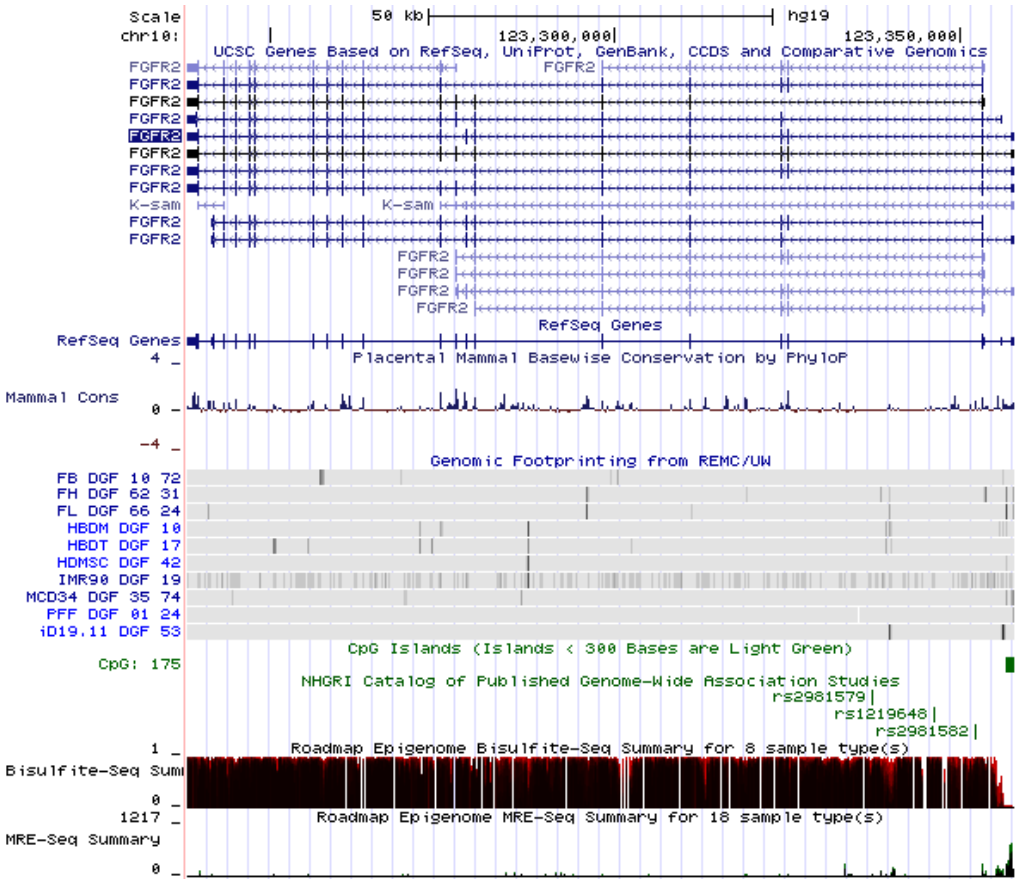
- Data tracks are hosted remotely (no data





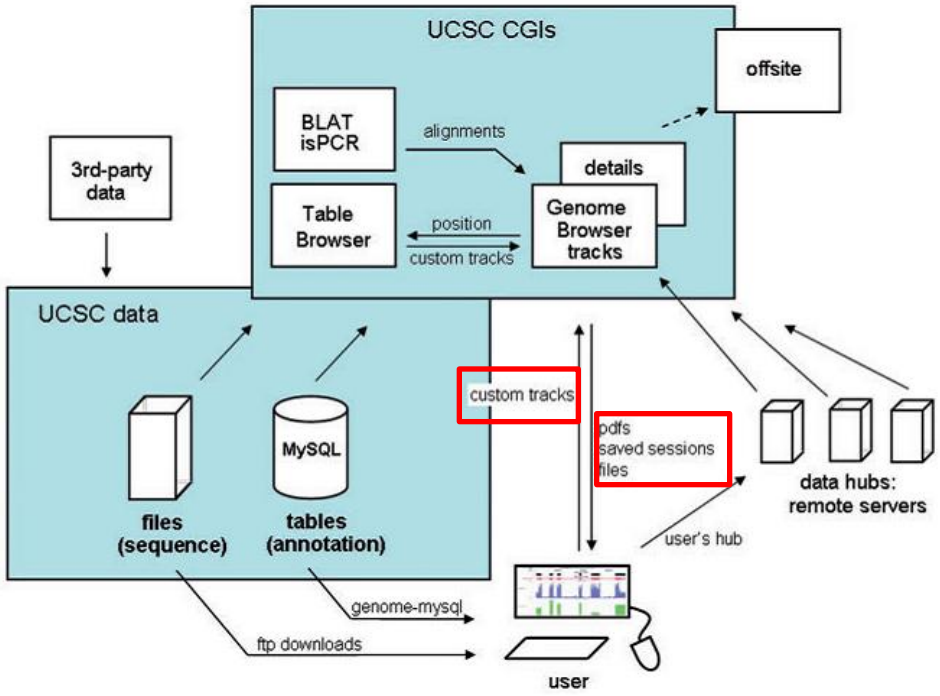
Tracks from the Epigenome project



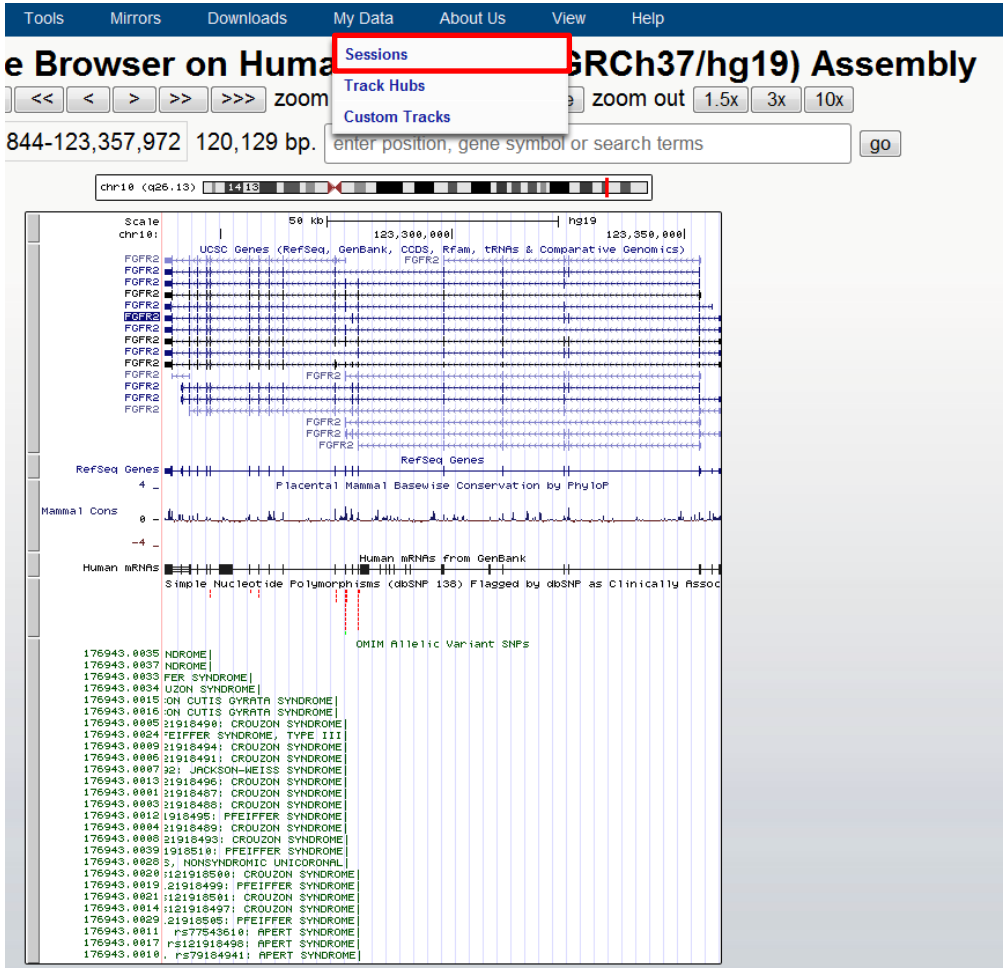


Associated Tools

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



Sessions



Save Settings

Save current settings as named session:

name: hg19 allow this session to be loaded by others submi

Save current settings to a local file:

file: UCSC_Session.txt file type returned: plain text submi

(leave file blank to get output in browser window)

Restore Settings

Use settings from another user's saved session:

user: session name: submit

Use settings from a local file: "C:\Users\z3265235 Choose... submit

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

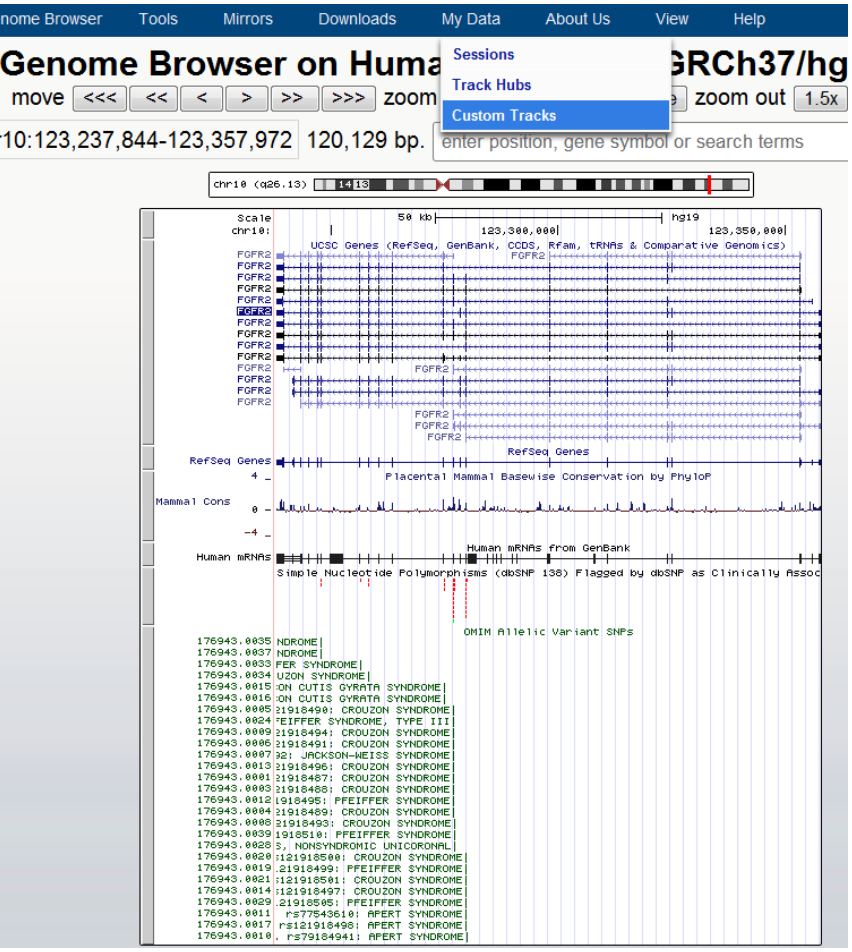
Sharing Sessions

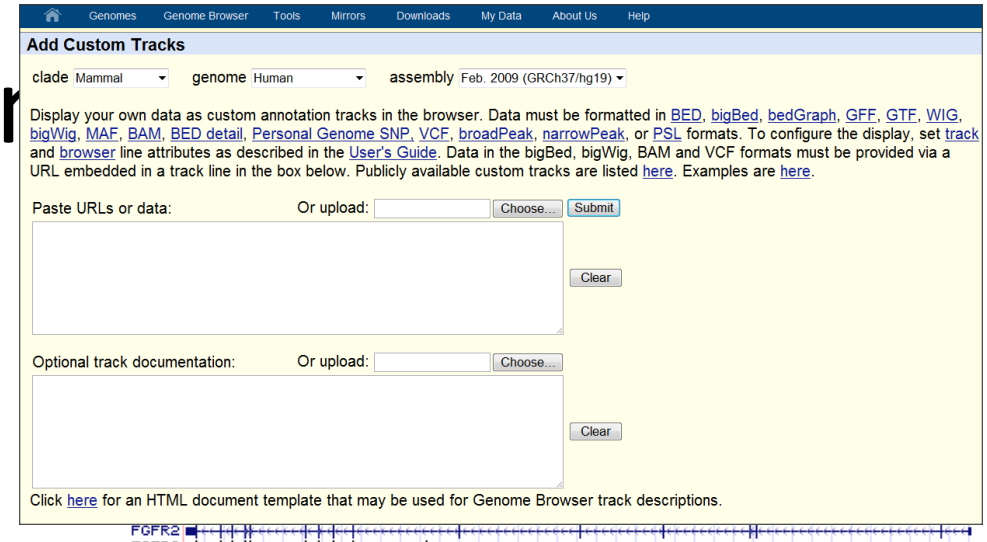
There are several ways to share saved sessions with others.

- Each previously saved named session appears with Browser and Email links. that session loaded. The resulting Genome Browser page can be bookmarke Email link invokes your email tool with a message containing the Genome Bro
- If you have saved your settings to a local file, you can send email to others w 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** where **U** is the URL mySession.txt . In this type of link, you can replace "hgSession" with "hgTrack

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Custom





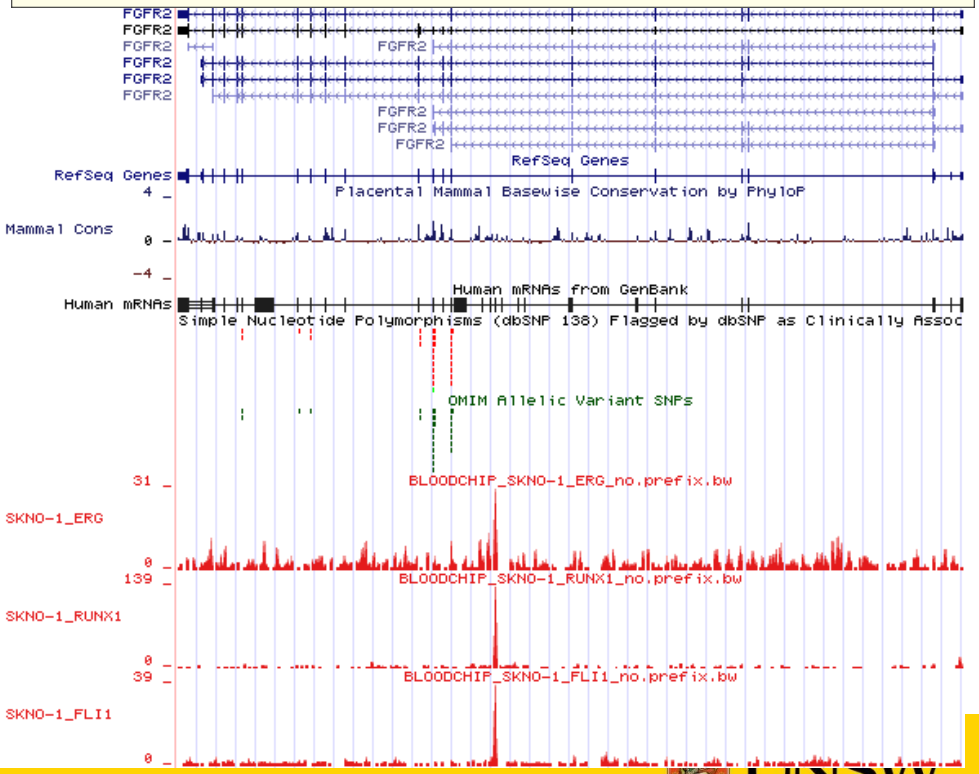
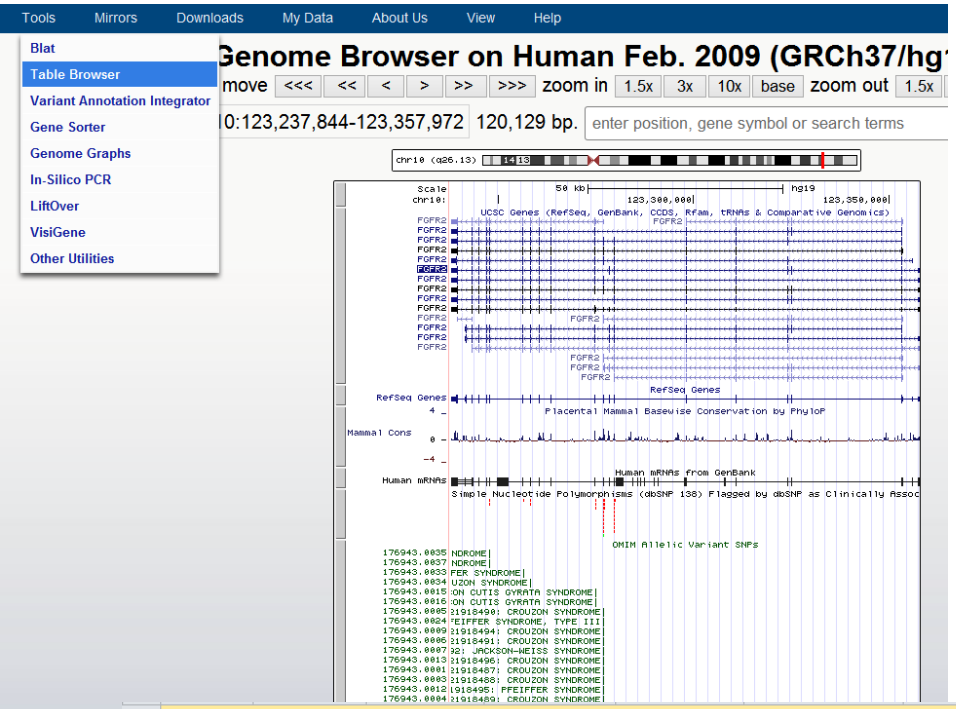
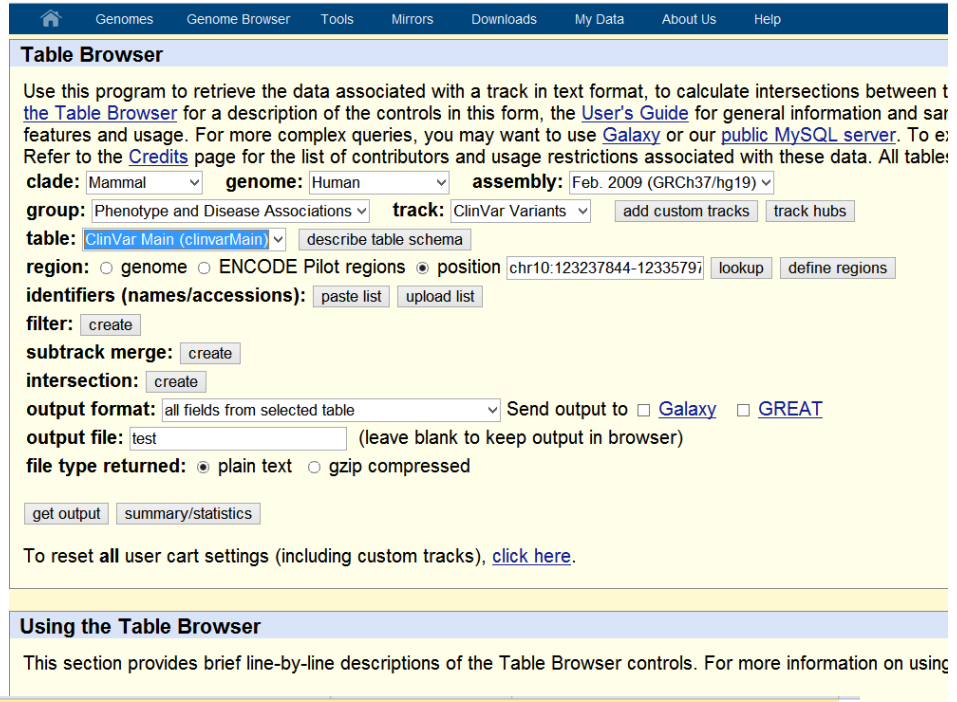


Table Browser





	#chrom	chromStart	chromEnd	name	type	clinSign	phenotype	origin	otherIds
2	chr10	123247548	123247549	FGFR2:c.1942G	single nucleoti	pathogenic		germline	OMIM Allelic Variant:176943.0035
3	chr10	123247608	123247609	FGFR2:c.1882G	single nucleoti	pathogenic		germline	OMIM Allelic Variant:176943.0037
4	chr10	123256214	123256215	FGFR2:c.1694A	single nucleoti	pathogenic	GeneReviews:NBK1455, MedGen:C1863356,	germline	OMIM Allelic Variant:176943.0033
5	chr10	123258104	123258105	FGFR2:c.1576A	single nucleoti	pathogenic	OMIM:123500, Orphanet:207, SNOMED CT:28861008, MedGen:C1865070, OMIM:609579, Orphanet:168624	germline	OMIM Allelic Variant:176943.0034
6	chr10	123274745	123274746	FGFR2:c.1172T	single nucleoti	pathogenic	MedGen:C3281247, OMIM:614592,	germline	OMIM Allelic Variant:176943.0043
7	chr10	123274776	123274777	FGFR2:c.1141T	single nucleoti	pathogenic	MedGen:C3281247, OMIM:614592,	germline	OMIM Allelic Variant:176943.0044
8	chr10	123274793	123274794	FGFR2:c.1124A	single nucleoti	other, pathogenic	, GeneReviews:NBK1455, MedGen:C1852406, OMIM:123790, Orphanet:1555	somatic, uncertain	OMIM Allelic Variant:176943.0015
9	chr10	123274802	123274803	FGFR2:c.1115C	single nucleoti	pathogenic	GeneReviews:NBK1455, MedGen:C1852406, OMIM:123790, Orphanet:1555	germline	OMIM Allelic Variant:176943.0016
10	chr10	123276855	123276856	FGFR2:c.1061C	single nucleoti	pathogenic	OMIM:123500, Orphanet:207, SNOMED CT:28861008	germline	OMIM Allelic Variant:176943.0005