

Deep Sequencing: Challenges and Opportunities for the Ensembl Genome Browser Project

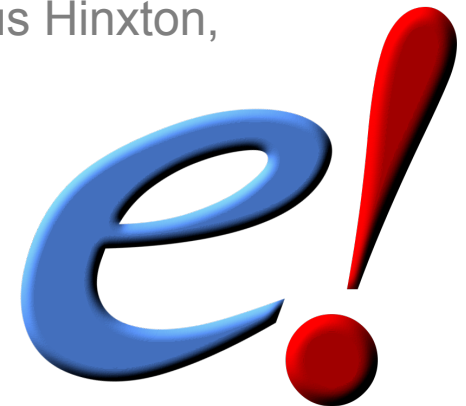
Introduction to ...
Making Sense of Genomes

Michael Schuster
EMBL-European Bioinformatics Institute
Wellcome Trust Genome Campus Hinxton,
Cambridge CB10 1SD
United Kingdom

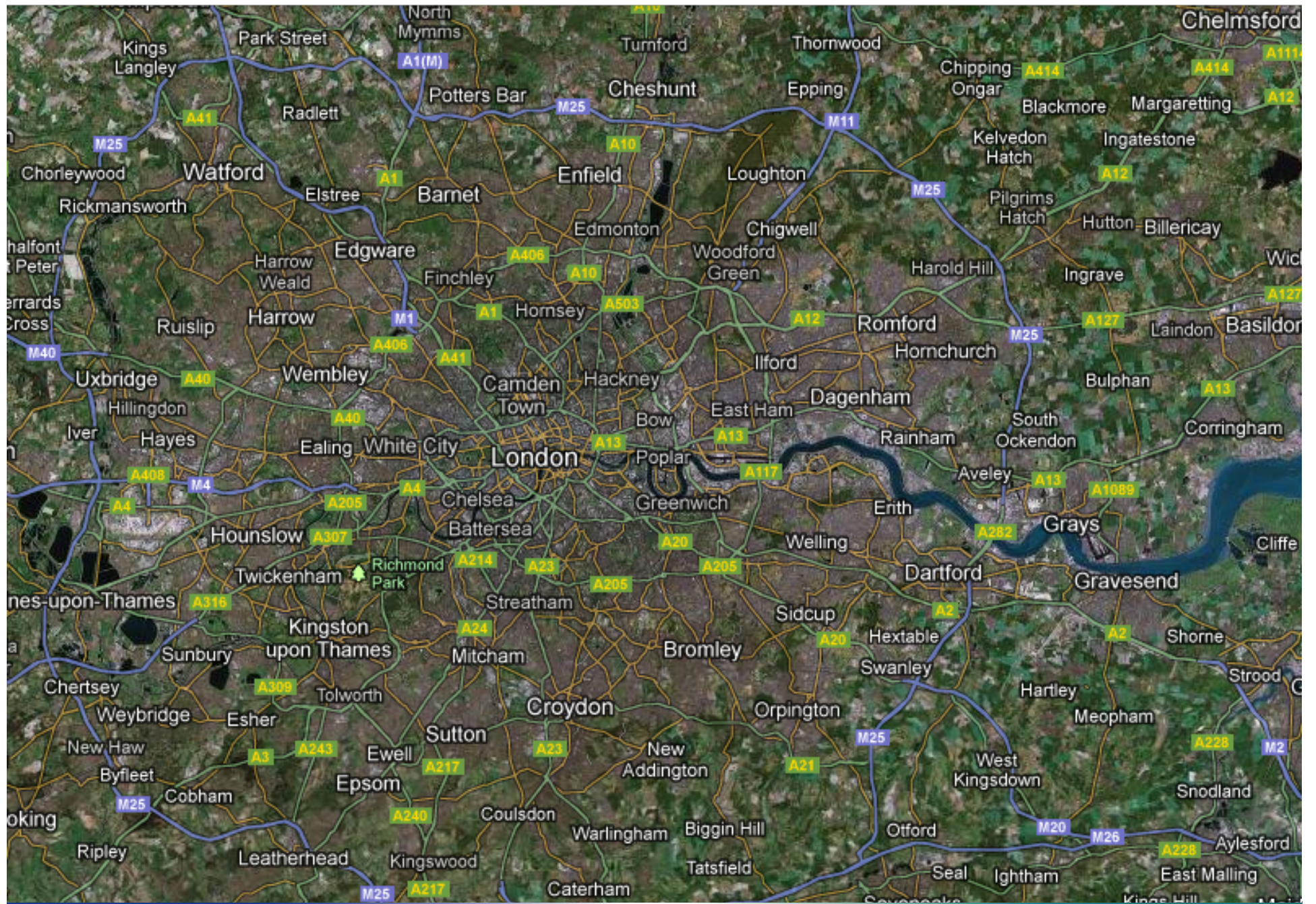
EMBL-EBI

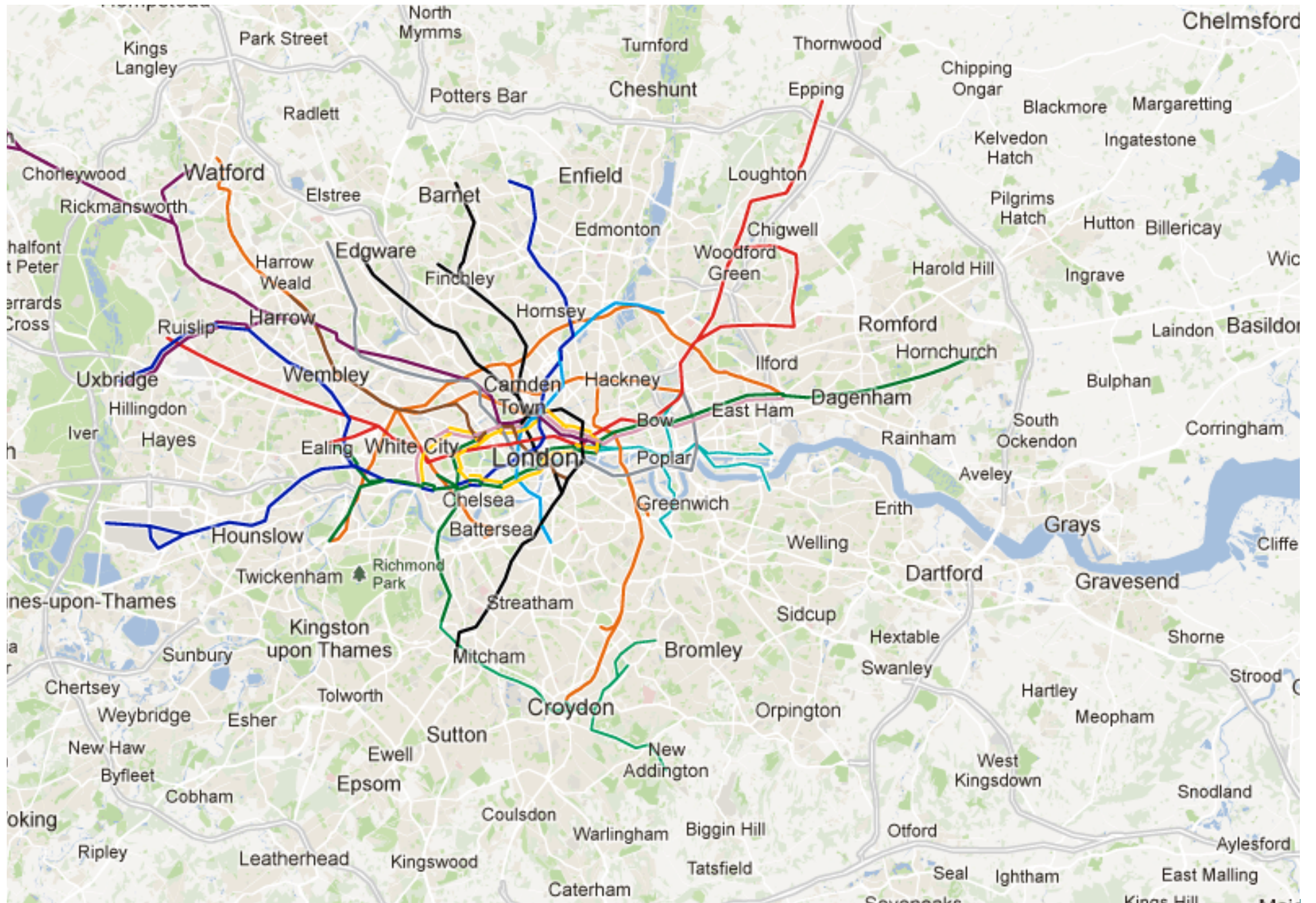


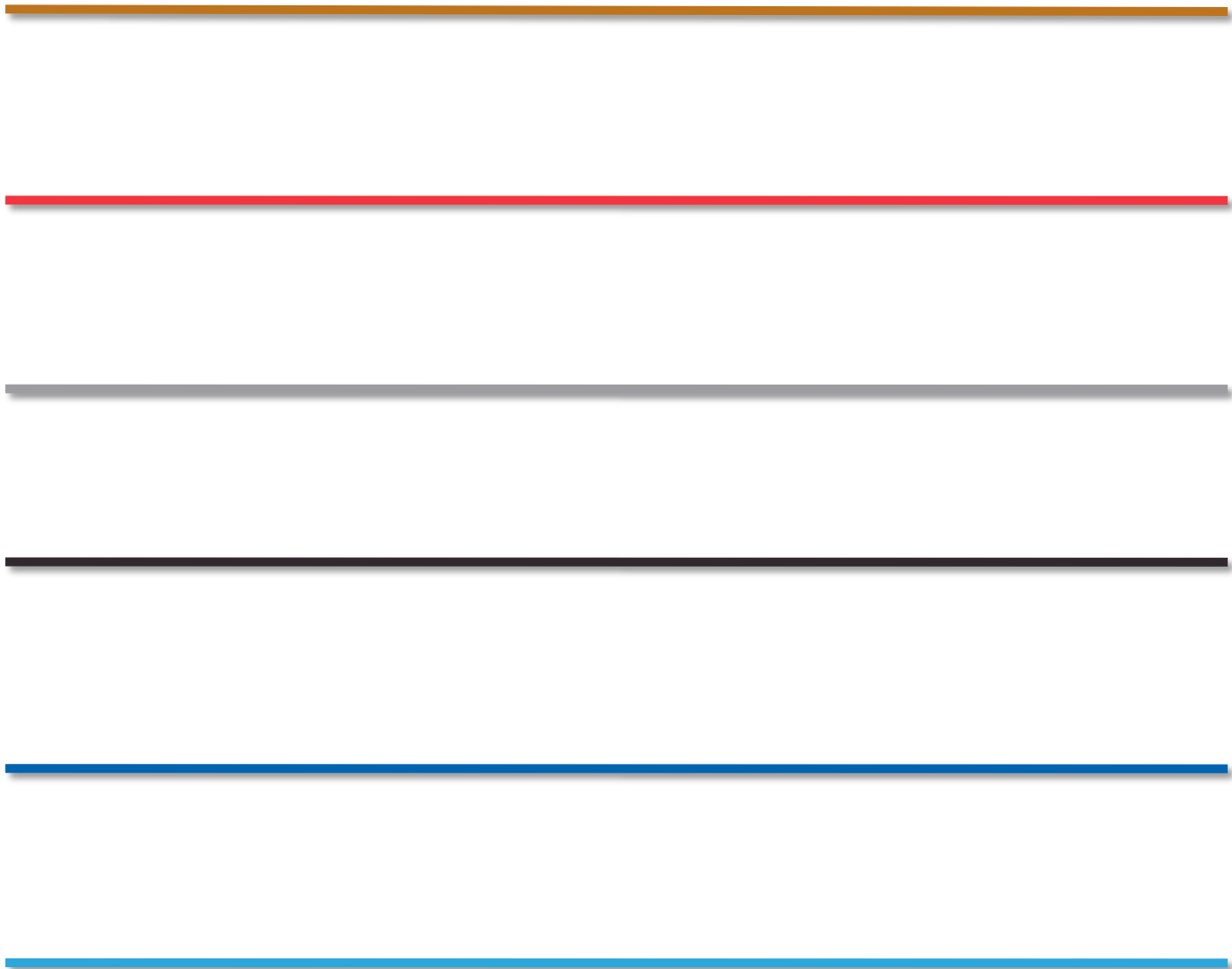
wellcome trust
sanger
institute

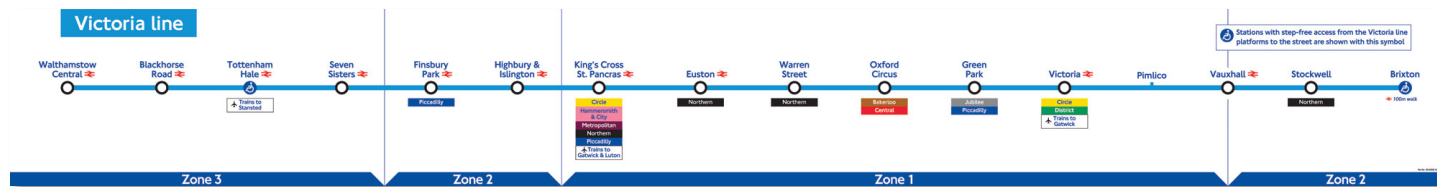
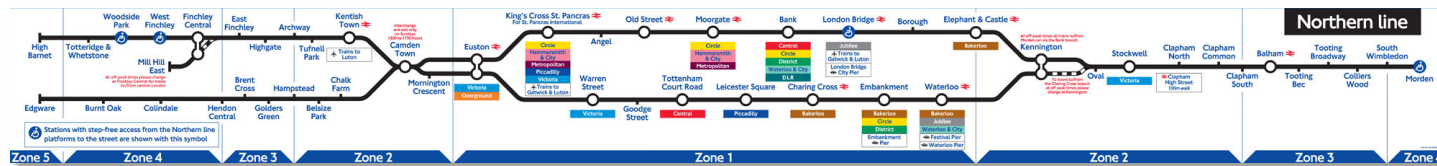
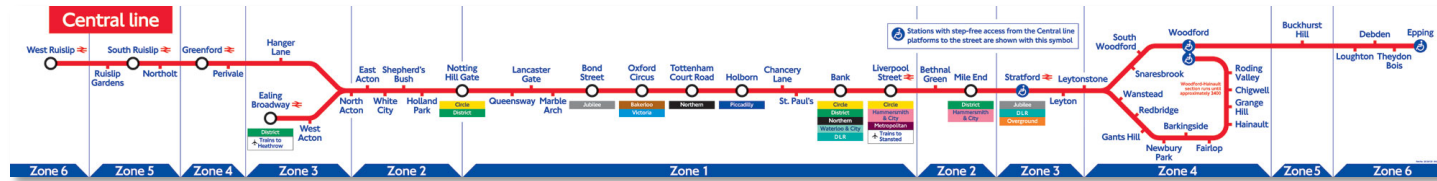




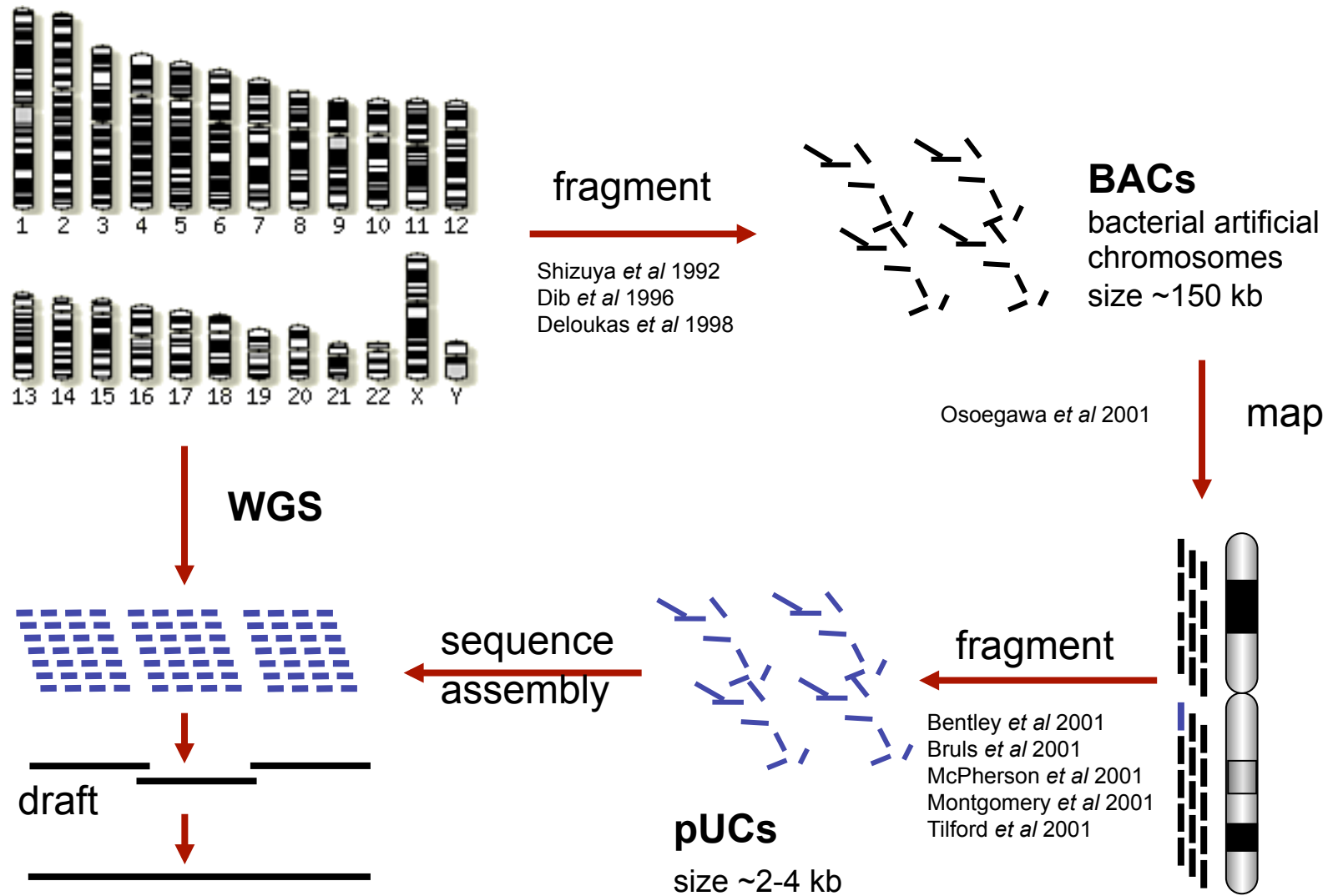




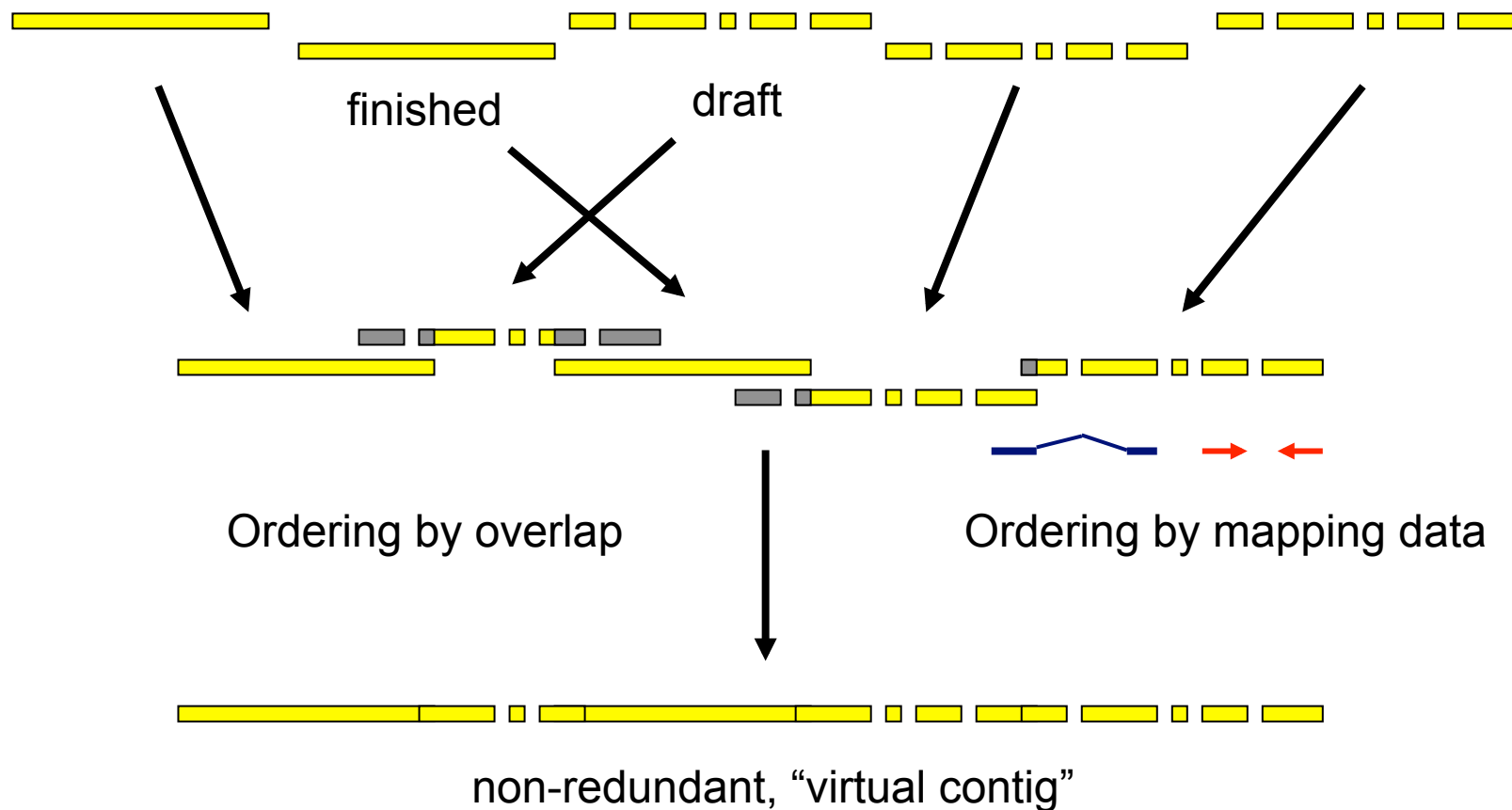




Mapping and Sequencing of the Human Genome



Genome Sequence Assembly



1560 base pairs out of ~3,200,000,000 ...

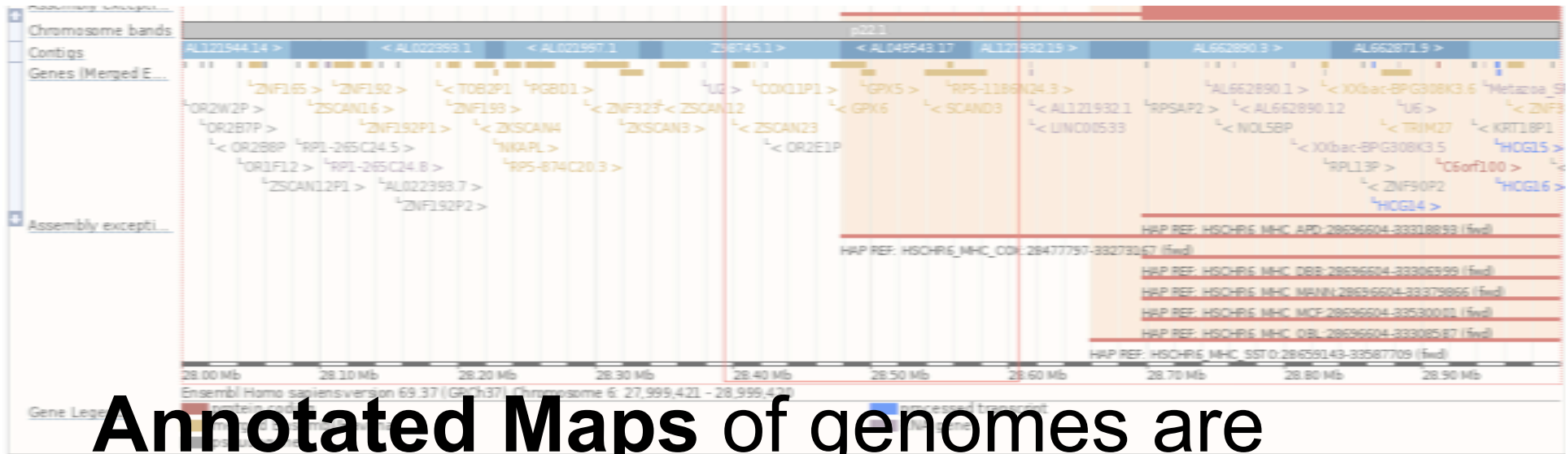
```
>chromosome:GRCh37:18:45357322:45458115:-1
AACCAAAATTTAGGTTAAACACTTATACCTGGCATCCCCTACCCGGGCCCTAAGTGCTCT
CCAAAGACCCACCTCTGGGCTTATTTACATCGCGAACCCGCCCTCGTTCAGCAACACGA
GGGCCCCGCCCGCTCATTTCATAGGGCGCGCCAGTCACGGGCGGGTCCTCGGGCTTC
TCCCGAACCCCTTTTCTCTCTCAGGCGCGTCCTCCGCGCTGCGCGGGGCCTCCCCGTGGG
ATGAGCACGTGTCCGCTGCCGGCCTCACGACGGACGGTGGACGAGGCGGCAGGCCCTAC
AGGCCCCACCCACGACGCTGGCGAGGGATCGGGCGGTCACCGGGAATCGTCTTAATGCGC
GGCAAGGCGCGGGCCTCTCCCTCTCCGCCCGTGAGCCCCGGTGGGAGCGCGCCGGCGCCC
AACTCAAGCGAAACCGCGGGCGTCCCGCCCCGCCCGGCCGCGCCCCGCCTCGCCTCACGC
TAGACTGGGGAGGCGGGACCAATCAGCGAGCGACGTCTCCCTTCCGATTCGAGGCCCCCG
ATGCGCGGCTCACACCCCGAGCTTCCCTCGTGCTGATTGGCTGCGGGCGCCGCCGGTCCG
GCCGGGAGGCGGGGCGGGCCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCC
CCGCGCCCCGCTCGCCCCCGGCCCTTCCCGCGGTGCTCGGCCTCGTTCCTTTCCTCCTC
CGCTCCCTCCGTCTTCCATACCCGCCCCGCGCGGGCTTTCGGCCGGCGTGCTCGCGCCCT
AACGGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGCCCCGGGGCTGCGCCG
GCGAATCGGCGGGGCCCGCGGCCAGGGTGGCAGGCGGGTCTACCCGCGCGGCCGCGGGC
GCGGAGAAGCAGCTCGCCAGCCAGCAGCCCGCCAGCCGCCGGGAGGTGGGTGCGTGGCGC
CGCGGCGGCCGGCGGGCCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGC
GCGGCCGTGCGGGCGGCCGGTAGGGCTGCGGGCGCGCCCTGAGGGGAGGAGGGGCAGCG
CGGGCGCGCGCTCCTCACCCCTCCTTCCCCGCGGGCGGCCAGGCTCCCTCCCCTC
CCCTCCCTCTCCTCCCCTCCCCTCCCCTCTTCCCCTACCCTCCCGCGCGCCCGGGCC
GCCGGCCGGGCCCGGGCCTGGGGGCGGGGCGGGAAGACGGCGGCCGGGAGTGTTTTAGT
TCCGCCTCCAATCGCCATTCCCCTCTTCCCCTCCAGCCCCCTCCATCCCATCGGAAGA
GGAAGGAACAAAAGGTCCCGGACCCCCGGATCTGACGGGGCGGGACCTGGCGCCACCTT
GCAGGTAAGCCTGGGCGCCC GCGGGCCTCCAGCTAGGGAAGTGTTTGCGTGCGTCCGCG
GCCGGGGCGATGGGCCGTGTACATGGCCGCTGCGGGTGGGGGCTGGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGGCCC GGGCGTGCGGGTTCGGGGCCGGAGAGCCGGGAAGGG
```

Annotation of Exons, Transcripts and Genes

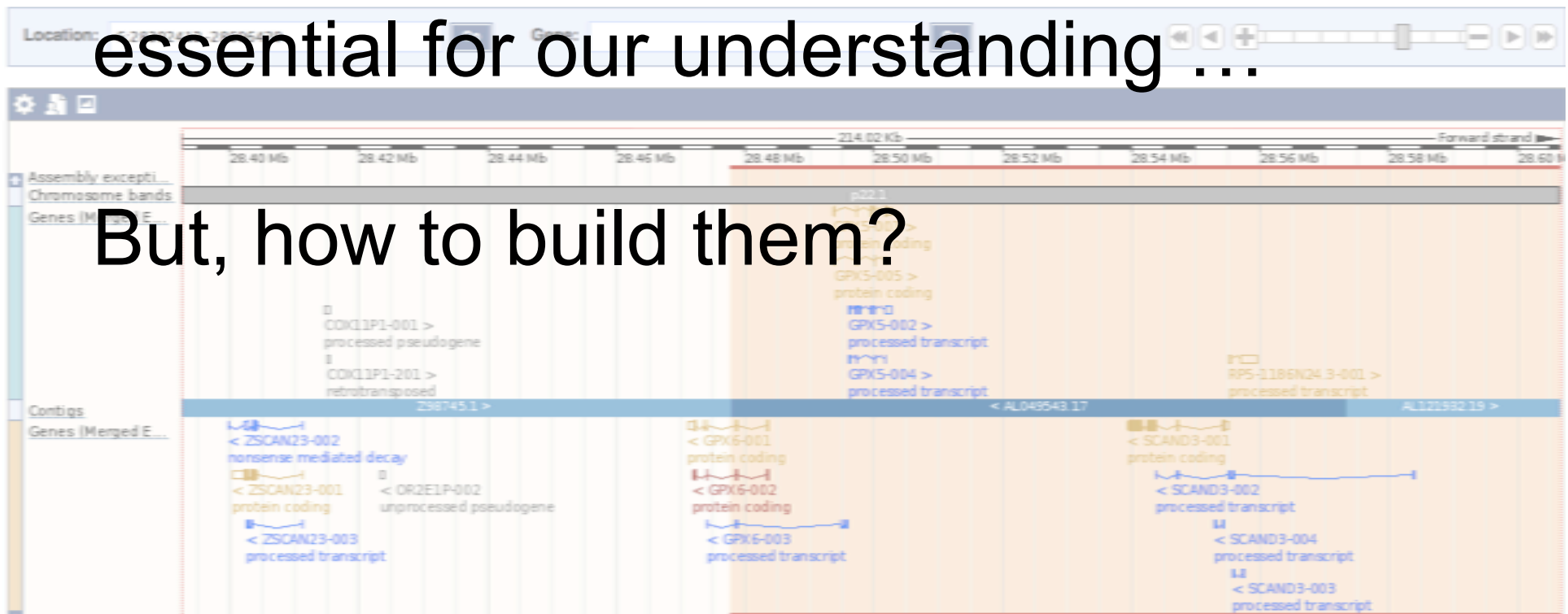
```
>chromosome:GRCh37:18:45357322:45458115:-1
AACCAAAATTTAGGTTAAACACTTATACCTGGCATCCCCTACCCGGGCCCTAAGTGCTCT
CCAAAGACCCACCTCTGGGCTTATTTACATCGCGAACCCGCCCTCGTTCAGCAACACGA
GGGCCCCGCCCGCTCATTTCATAGGGCGCGCCAGTCACGGGCGGGTCCTCGGGCTTC
TCCCGAACCCCTTTTCTCTCTCAGGCGCGTCCTCCGCGCTGCGCGGGGCCTCCCCGTGGG
ATGAGCACGTGTCCGCTGCCGGCCTCACGACGGACGGTGGACGAGGCGGCAGGCCCTAC
AGGCCCCACCCACGACGCTGGCGAGGGATCGGGCGGTACCCGGGAATCGTCTTAATGCGC
GGCAAGGCGCGGGCCTCTCCCTCTCCGCCCGTGAGCCCCGGTGGGAGCGCGCCGGCGCCC
AACTCAAGCGAAACCGCGGGCGTCCCGCCCCGCCCGGCCGCGCCCCGCCTCGCCTCACGC
TAGACTGGGGAGGCGGGACCAATCAGCGAGCGACGTCTCCCTTCCGATTTCGAGGCCCCCG
ATGCGCGGCTCACACCCCGAGCTTCCCTCGTGCTGATTGGCTGCGGGCGCCGCGGTCCG
GCCGGGAGGCGGGGCGGGCCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCC
CCGCGCCCCGCTCGCCCCCGGCCCTTCCCGCGGTGCTCGGCCTCGTTCCTTTCCTCCTC
CGCTCCCTCCGTCTTCCATACCCGCCCGCGCGGGCTTTCGGCCGGCGTGCTCGCGCCCT
AACGGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGCCCCGGGGCTGCGCCG
GCGAATCGGCGGGGCCCGCGGCCAGGGTGGCAGGCGGGTCTACCCGCGCGGCCGCGGGC
GCGGAGAAGCAGCTCGCCAGCCAGCAGCCCGCCAGCCGCGGGAGGTGGGTGCGTGGCGC
CGCGGCGGGCCGGCGGGCCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGC
GCGGCCGTGCGGGCGGCCGGTAGGGCTGCGGGCGCGCCCTGAGGGGAGGAGGGGCAGCG
CGGGCGCGCGCTCCTCACCCCTCCTTCCCCGCGGGCGGCGGCCAGGCTCCCTCCCCTC
CCCTCCCTCTCCTCCCCTCCCCTCCCCTCTTCCCCTACCCTCCCGCGCGCCCGGGCC
GCCGGCCGGGCCCGGGCCTGGGGGCGGGGCGGGAAGACGGCGGCCGGGAGTGTTCAGT
TCCGCCTCCAATCGCCATTCCCCTCTTCCCCTCCAGCCCCCTCCATCCATCGGAAGA
GGAAGGAACAAAAGGTCCCGGACCCCCGGATCTGACGGGGCGGGACCTGGCGCCACCTT
GCAGGTAAGCCTGGGCGCCCGCGGGCCTCCAGCTAGGGAAGTGTTCGCGTGCCTCCGCG
GCCGGGGCGATGGGCCGTGTACATGGCCGCTGCGGGTGGGGGCTGGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGGCCCGGGCGTGCGGGTTCGGGGCCGGAGAGCCGGGAAGGG
```

Annotation of Genetic Variation

```
>chromosome:GRCh37:18:45357322:45458115:-1
AACCAAAATTTAGGTTAAACACTTATACCTGGCATCCYCTACCCGGGCCCTAAGTGCTCT
CCAAAACACCCACCTCTGGGCTTATTTACATCGGAACCCGCCCTCGTTCAGCAACACRA
GGGCCCCGCCCKCTCATTTCATAGGGCGCGCCAGTCACGGGGCCRGGTCCTCGGGCTTC
TCCCGAACCCCTTTTCYCTCTCAGGCSCGTCTCCGCGCTGCGCGGGGCCTCCCCGTGGG
ATGAGCACGTGTCCGCTGCCGGCCTCACGACGGACGGTGGACGAGGCGGCAGGCCCCYAC
AGGCCCCACCCACGACGCTGGCGAGGGATCGGGCGGTCACCRGGGAATCGTCTTAATGCGC
GGCAAGGCGCGGGCCTCTCCCTCTCCGCCCGTGAGCCCCGGTGRGAGCGCGCCGGCGCCC
AACTCAAGYGAAACCGCGGGCGTCCCGCCCCGCCCGGCCGCGCCCCGCTCGCCTCACGC
TAGACTGGGGAGGCGGGACCAATCAGCGAGCGACGTCTCCCTTCCGATTTCGAGGCCCCCG
ATGCGCGGCTCACACCCCGAGCTTCCCTCGTGCTGATTGGCTGCGGGCGCCGCGGTCCG
GCCGGGAGGCGGGGCGGGCCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCC
CCGCGCCCCGCTCGCCCCCGGCCCTTCCCGCGGTGCTCGGCCTCGTTCCTTTCCTCCTC
CGCTCCCTCCGTCTTCCATACCCGCCCCGCGCGGCTTTCGGCCGGCGTGCTCGCGCCCT
AACGGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGCCCCGGGGCTGCGCCG
GCGAATCGGCGGGGCCCGCGGCCAGGGTGGCAGGCGGGTCTACCCGCGCGGCCGCGGGC
GCGGAGAAGCAGCTCGCCAGCCAGCAGCCCGCCAGCCGCGGGAGGTGGGTGCGTGGCGC
CGCGGCGGGCCGGCGGGCCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGC
GCGGCCGTGCGGGCGGCCGGTAGGGCTGCGGGCGCGCCCTGAGGGGAGGAGGGGCAGCG
CGGGCGCGCGCTCCTCACCCCTCCTTCCCCGCGGGCGGCGGCCAGGCTCCCTCCCCTC
CCCTCCCTCCTCCCCTCCCCTCCCCTCTTCCCCTACCCTCCCGCGCGCCCGGGCC
GCCGGCCGGGCCCGGGCCTGGGGGCGGGGCGGGAAGACGGCGGCCGGGAGTGTTCAGT
TCCGCTCCAATCGCCATTCYCTCTTCCCCTCCAGCCCCCTCCATCCRTCAGGAAGA
GGAAGGAACAAAAGGTCCCGGACCCCCGGATCTGACGGGGCGGGACCTGGYGCCACCTT
GCAGGTAAGCCTGGGCGCCMCGGGCCKCCAGCTAGGGAAGTGTTCYGTGCGTCCGCG
GCCGGGGCGATGGGCCGTGTACATGGCCGCTGCGGGTGGGGGCTGGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGGCCCGGGCGTGCGGGTTTCGGGGCCGGAGAGCCGGGAAGGG
```



Annotated Maps of genomes are essential for our understanding ...

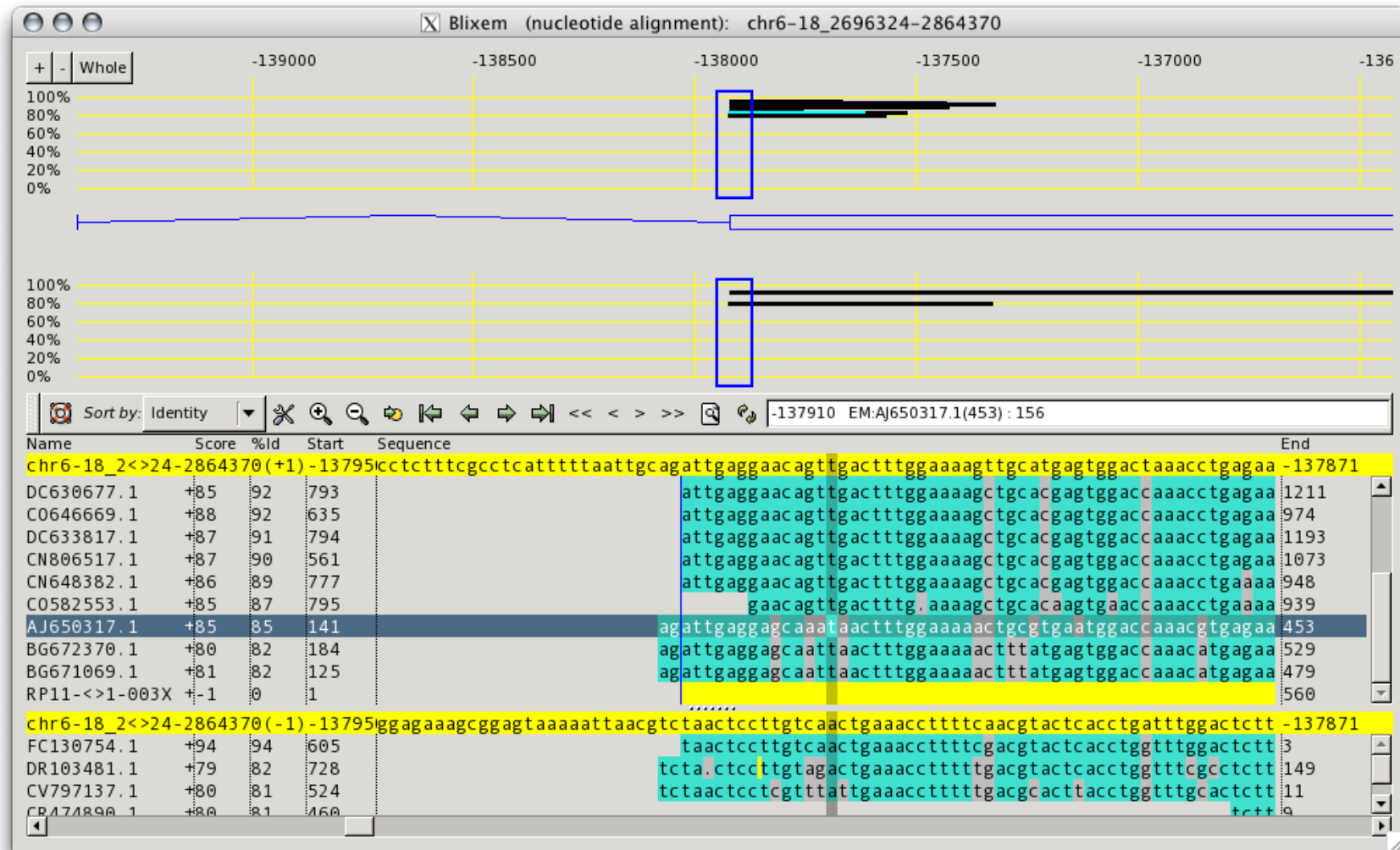


But, how to build them?

Havana Manual Annotation Tool – ZMap



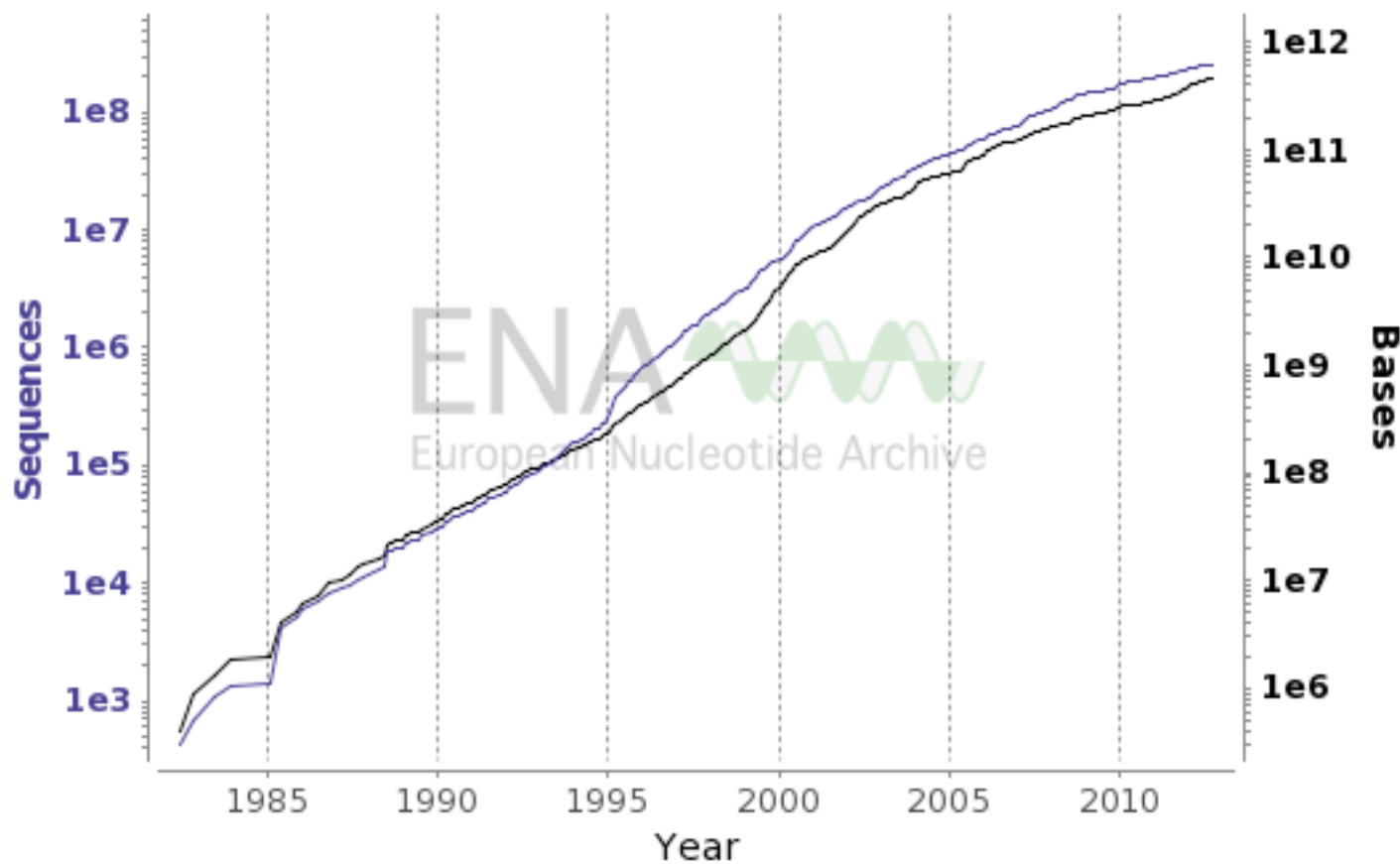
Havana Manual Annotation Tool – Blixem



Is there a Problem? Well, the Growth Rate ...

EMBL-Bank Growth

15-Oct-2012



— Sequences (255.1 millions) — Bases (466.6 billions)

<http://www.ebi.ac.uk/ena/about/statistics>

Ensembl Project – Aims

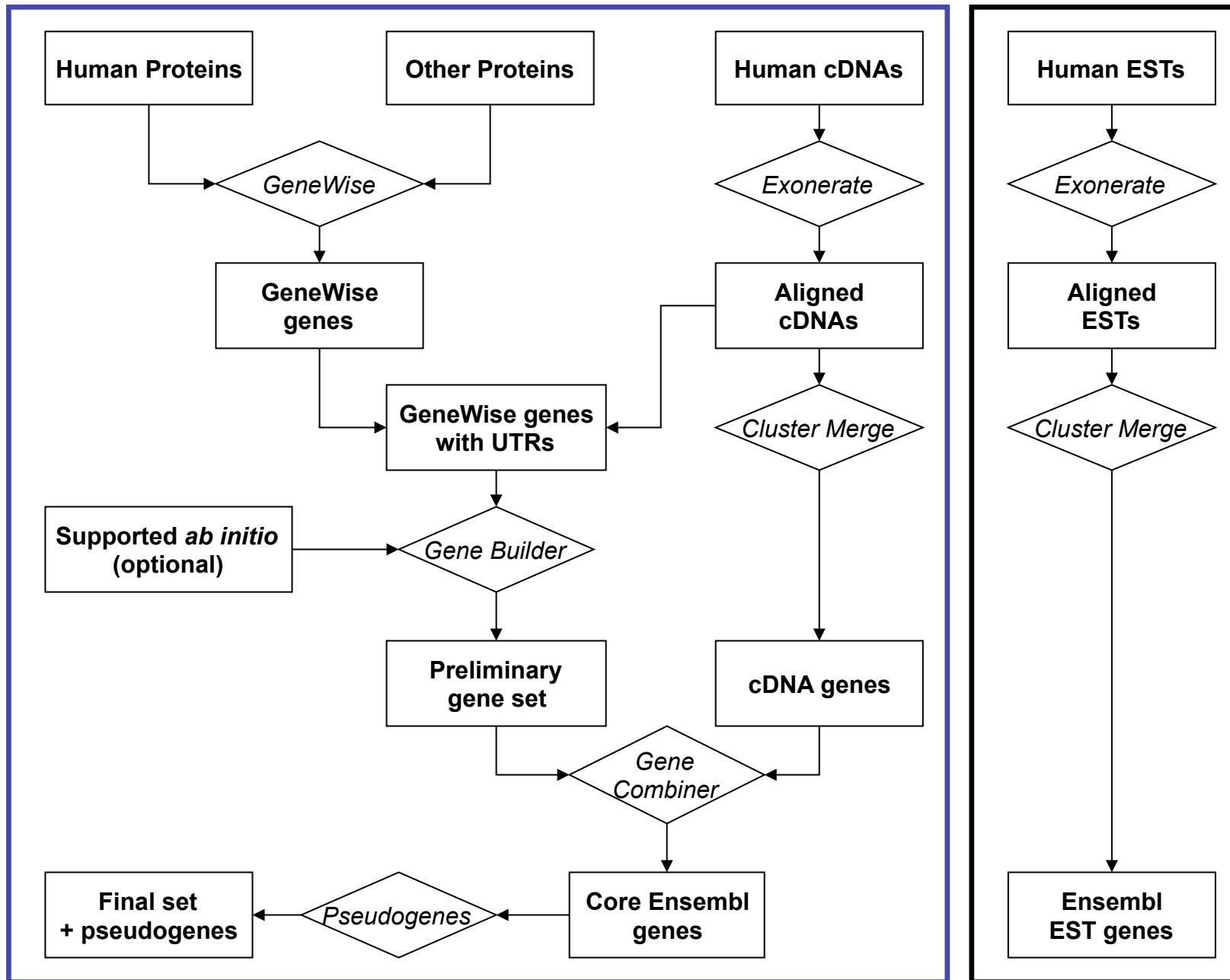
- Funded to provide metazoan genomes to the world
- Aims to provide automated genome annotation system
- Integrate with other resources of biological information
- Graphical representation of complex data sets
- Large-scale data export tools
- Openness - all software, data and results freely available

Ensembl Project – Background

- Started in 1999 by Michele Clamp and Ewan Birney
- Joint project between
 - European Bioinformatics Institute (EBI)
 - Wellcome Trust Sanger Institute (WTSI)
- Group of ca 45 people led by **Paul Flicek** (EBI) and **Steve Searle** (WTSI)
- Ensembl strategy committee: **Ewan Birney** (EBI), **Tim Hubbard** (WTSI) and **Richard Durbin** (WTSI)
- Mainly Wellcome Trust funded
- Additional EMBL, EU, BBSRC and NIH-NIAID funding

The Ensembl System – Overview

1. Scalable database system for storing genomic data and annotation
2. Automatic annotation method based on a rule set of heuristics
Ensembl Genome Analysis and Annotation Pipeline
3. Public web interface for genome annotation display
Ensembl Genome Browser



Ensembl Uses Computational Algorithms – Exonerate

Query: SMAD2_HUMAN Q15796 Mothers against decapentaplegic homolog 2 (SMAD 2).
 Target: AC120349 AC120349.5 Homo sapiens chromosome 18, clone RP11-380M21, complete sequence.

```

1 : MetSerSerIleLeuProPheThrProProValValLysArgLeuLeuGlyTrpLysL :      20
  |||
MetSerSerIleLeuProPheThrProProValValLysArgLeuLeuGlyTrpLysL
62055 : ATGTCGTCATCTTGCCATTCACGCCGCCAGTTGTGAAGAGACTGCTGGGATGGAAGA : 62112

21 : vsSerAlaGlyGlySerGlyGlyAlaGlyGlyGlyGluGlnAsnGlyGlnGluGluLy :      39
  |||

```

```

79 : ro{Se} >>>> Target Intron 1 >>>> {r}ThrCysSerGluIle
  ||{|}|++          25956 bp          ++{|}|
  ro{Se}              {r}ThrCysSerGluIle
62287 : CA{AG}gt.....ag{C}ACTTGCTCTGAAATT

```

```

59 : AspGluLeuGluLysAlaIleThrThrGlnAsnCysAsnThrLysCysValThrIleP :      78
  |||
AspGluLeuGluLysAlaIleThrThrGlnAsnCysAsnThrLysCysValThrIleP
62227 : GATGAGCTTGAGAAAGCCATCACCCTCAAACTGTAATACTAAATGTGTTACCATAC : 62286

```

```

79 : ro{Se} >>>> Target Intron 1 >>>> {r}ThrCysSerGluIleTrpGl :      86
  ||{|}|++          25956 bp          ++{|}|
  ro{Se}              {r}ThrCysSerGluIleTrpGl
62287 : CA{AG}gt.....ag{C}ACTTGCTCTGAAATTGGGG : 88266

```

```

87 : yLeuSerThrProAsnThrIleAspGlnTrpAspThrThrGlyLeuTyrSerPheSer :     105
  |||
yLeuSerThrProAsnThrIleAspGlnTrpAspThrThrGlyLeuTyrSerPheSer
88267 : ACTGAGTACACCAAATACGATAGATCAGTGGGATACAACAGGCCTTTACAGCTTCTCT : 88323

```

```

106 : GluGlnThr{Ar} >>>> Target Intron 2 >>>> {g}SerLeuAspGlyA :     114
  |||||{|}|++          1038 bp          ++{|}|
  GluGlnThr{Ar}              {g}SerLeuAspGlyA
88324 : GAACAAACC{AG}gt.....ag{G}TCTCTTGATGGTC : 89388

```

Biological Evidence

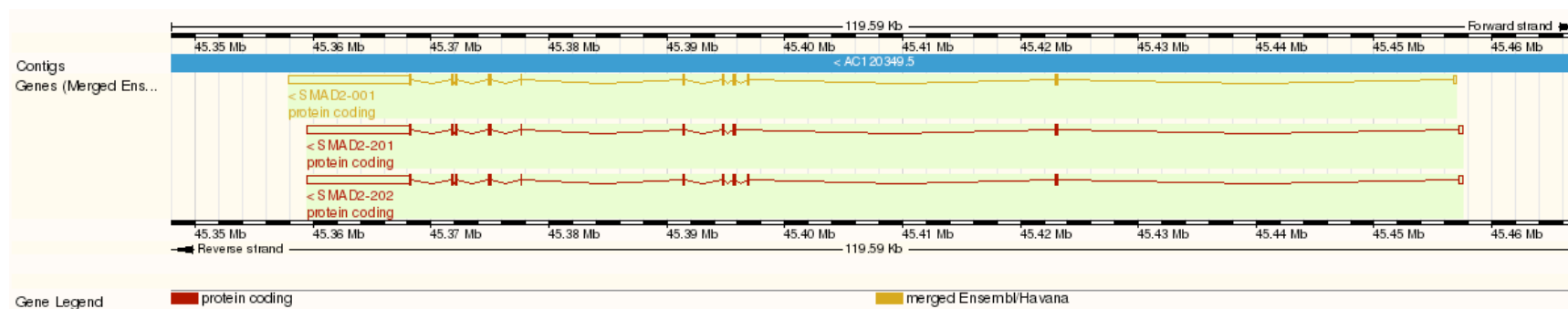
All Ensembl gene predictions are strictly based on **experimental evidence**

- **UniProtKB/Swiss-Prot**
A manually curated database and therefore of highest accuracy
- **NCBI RefSeq**
A partially manually curated database
- **UniProtKB/TrEMBL**
Automatically annotated translations of ENA coding sequence (CDS) features
- **European Nucleotide Archive (ENA)**
Comprehensive set of nucleotide sequences owned by original submitters

International Nucleotide Sequence Database Collaboration (INSDC)
ENA (EBI, UK), GenBank (NCBI, USA) and DDBJ (NIG, Japan)

Ensembl-Havana Merged Gene Set

- **Automatically annotated Ensembl** transcript sets for human, mouse, zebrafish and pig are merged into **manually curated Havana** transcript sets.
- Exon boundary **and** coding sequence (CDS) have to overlap
- GENCODE gene set for human <http://www.genencodegenes.org/>
- Comparison provided colour-coded:
 - Ensembl-Havana merged: **gold**
 - Ensembl or Havana only: **red**



Merge Criteria – Coding versus Coding

Perfect match

Ensembl 
HAVANA 



Coding match 1

Ensembl 
HAVANA 



Coding match 2

Ensembl 
HAVANA 



Coding match 3

Ensembl 
HAVANA 

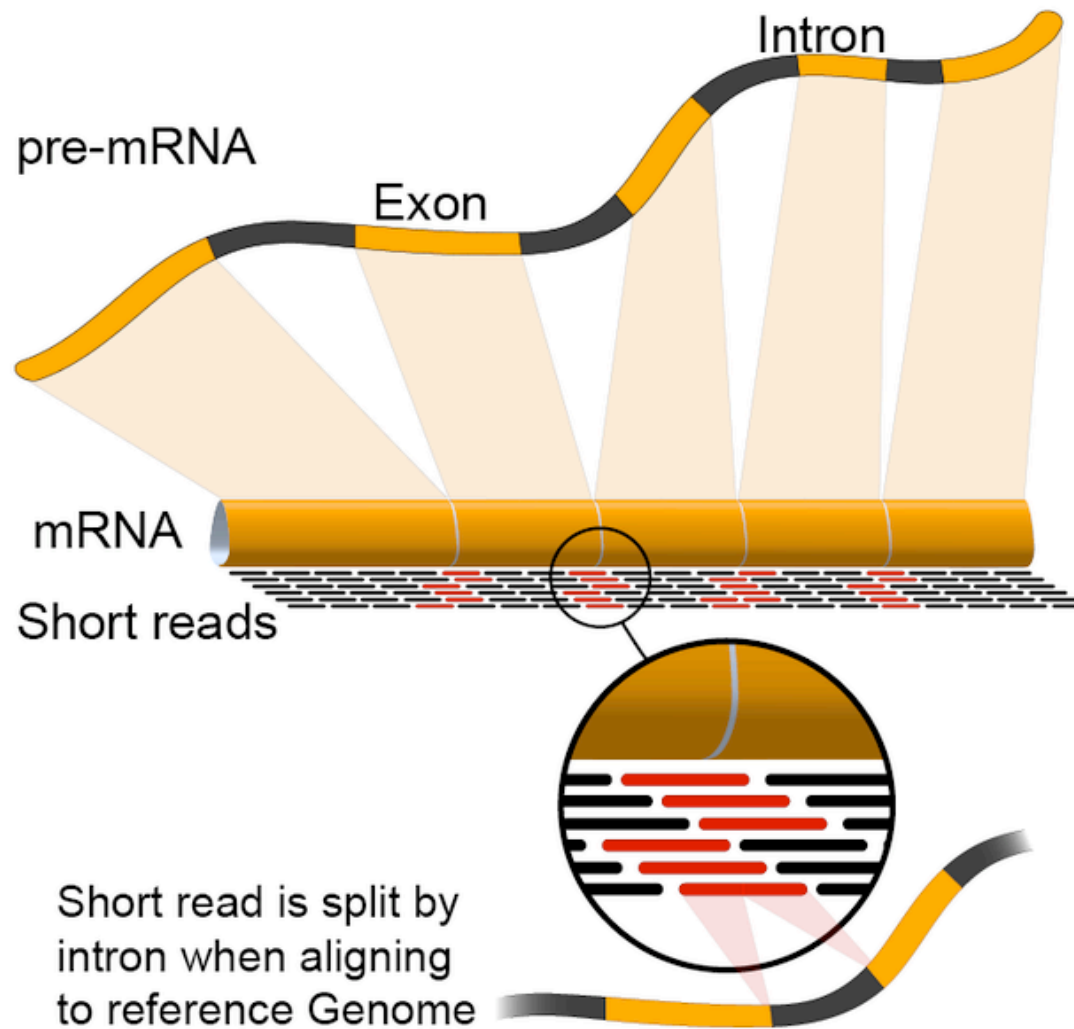


No match

Ensembl 
HAVANA 



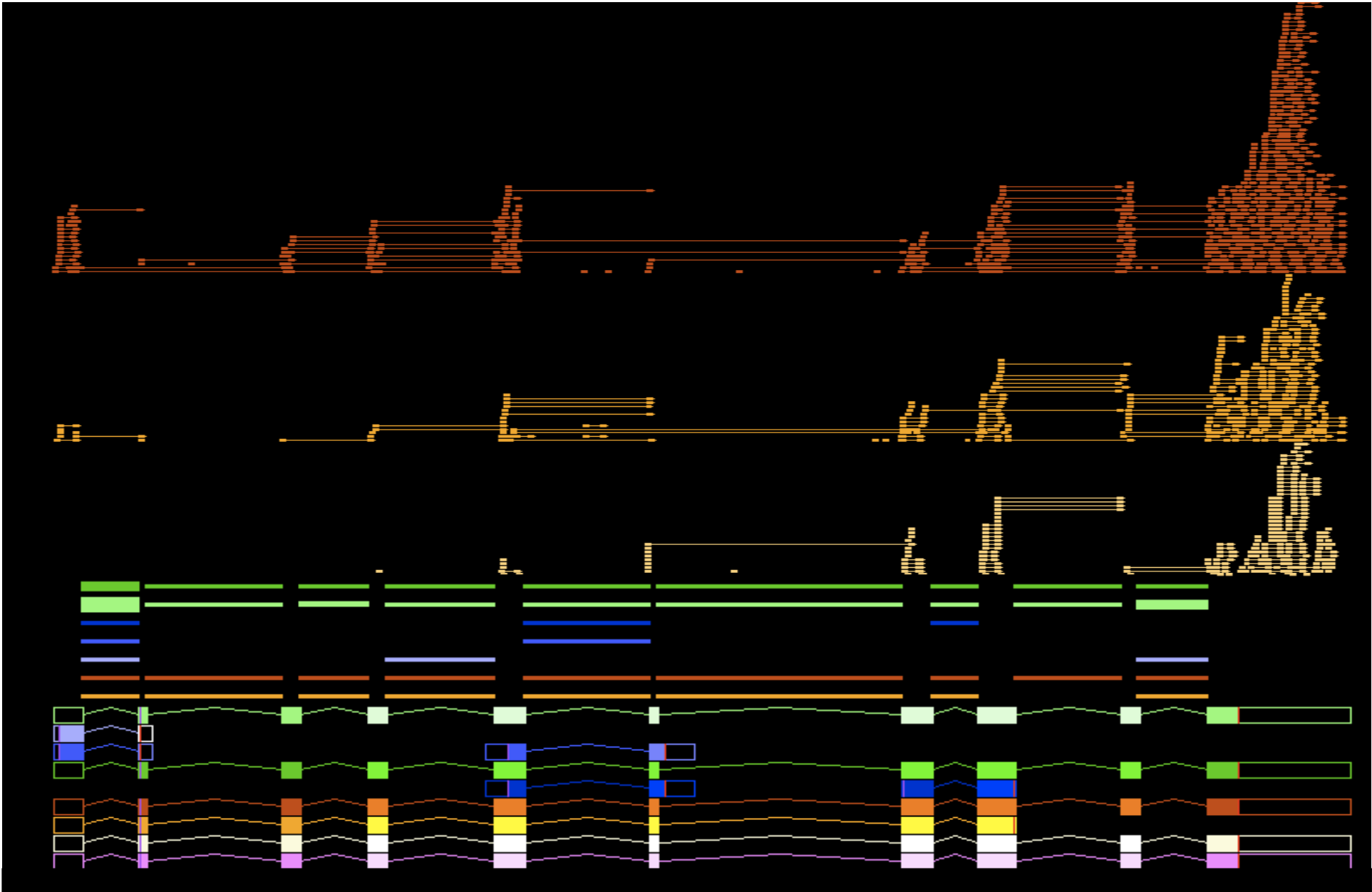
New Developments – RNA-Seq-based Annotation



<http://en.wikipedia.org/wiki/File:RNA-Seq-alignment.png>

RNA-Seq Technology

- Whole Transcriptome Shotgun Sequencing (WTSS)
- High-throughput, deep sequencing of cDNA libraries
- Challenges:
 - Tens of millions of short reads, ideally ~100 bp, paired ends
 - Hundreds of Gigabyte to few Terabyte of data
- Options for Ensembl:
 - Build a **complete** gene set from scratch
 - **Add novel** models into an existing gene set
 - **Improve existing gene sets** by extending transcript models



0.206Mb

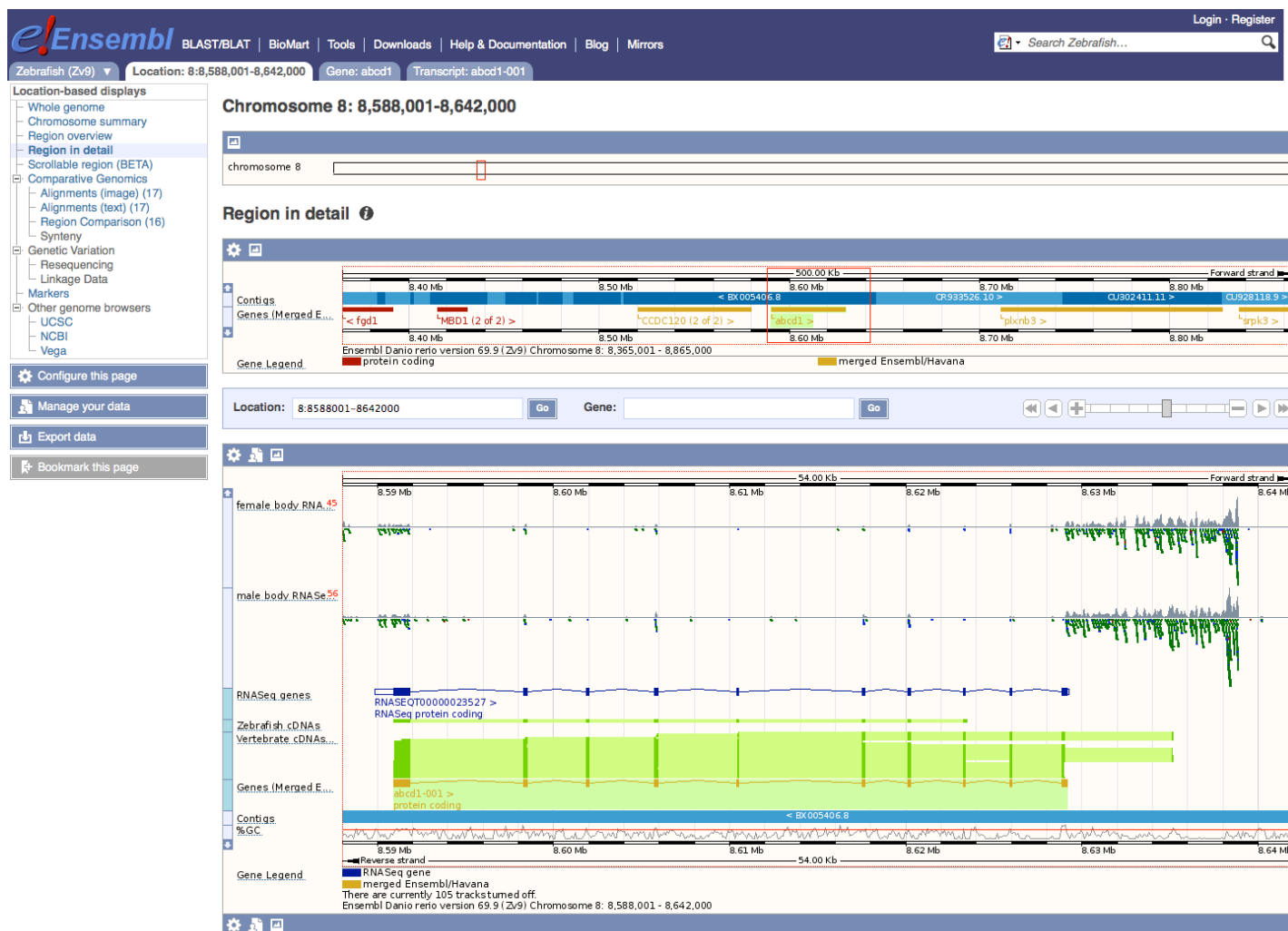
0.208Mb

0.21Mb

0.212Mb

0.214Mb

Danio rerio abcd1 – RNA-Seq Transcripts

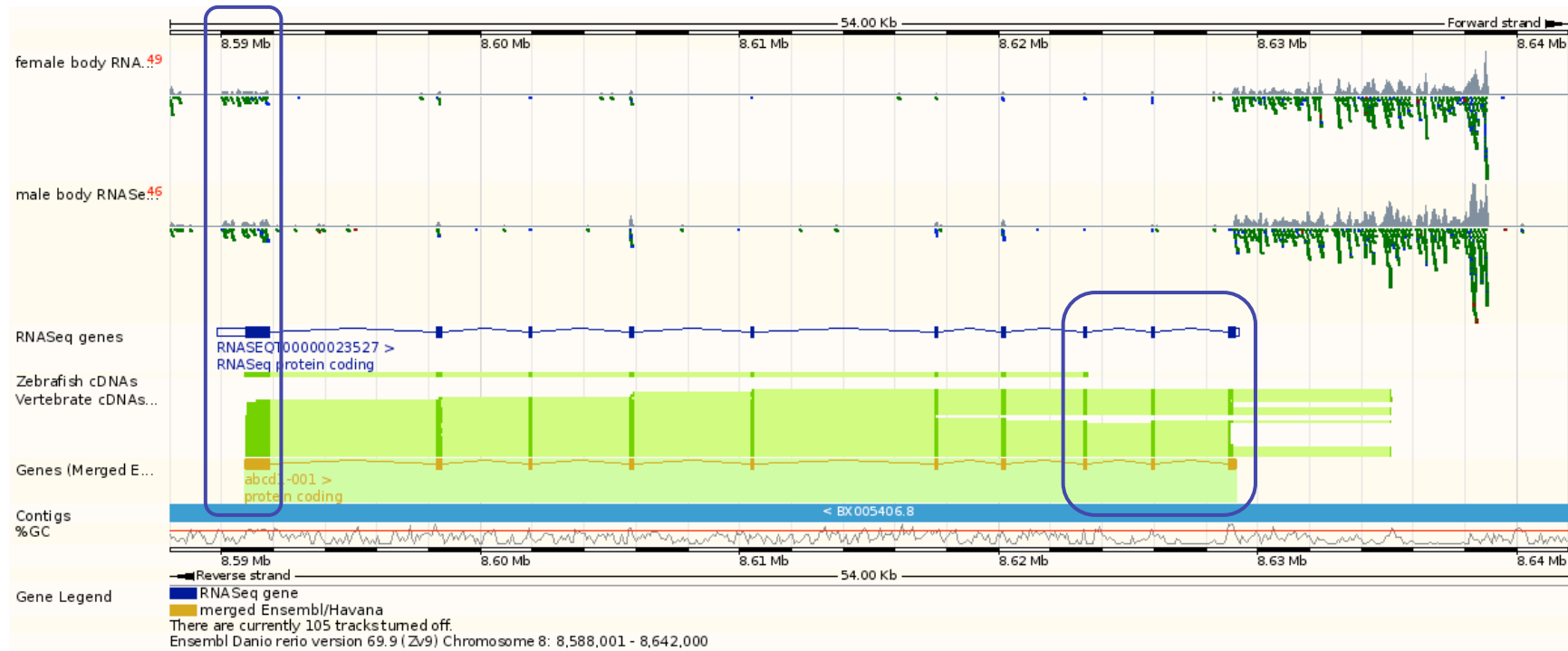


Ensembl release 69 - October 2012 © WTSI / EBI

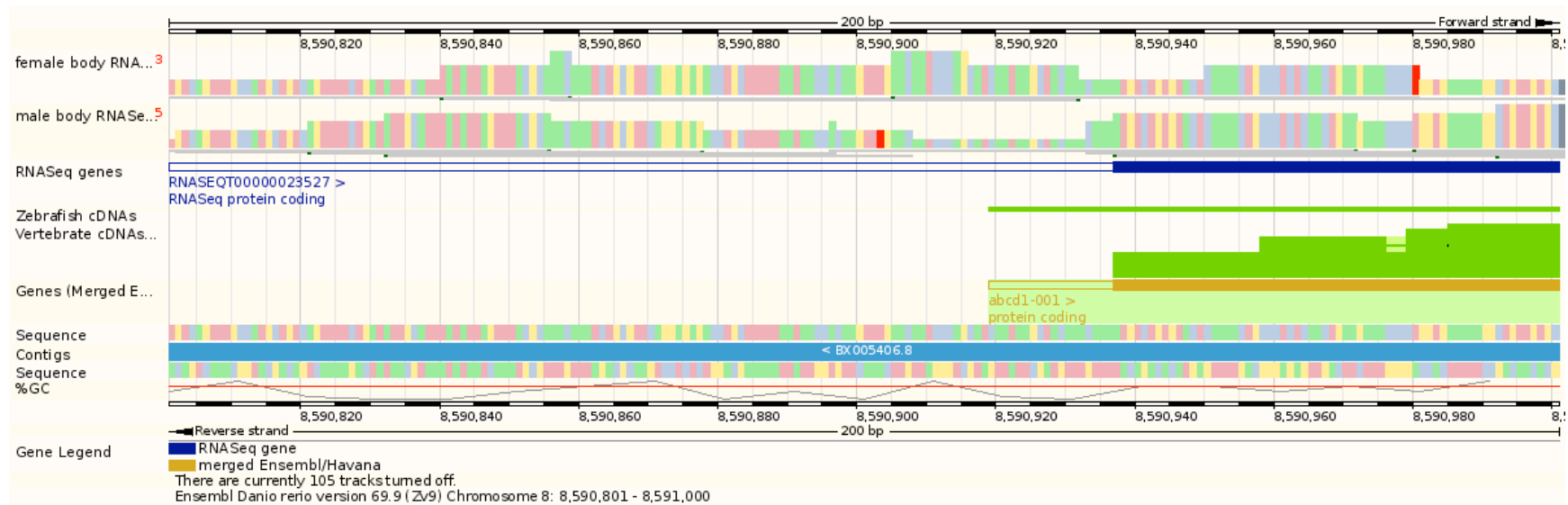
[About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

[Permanent link](#) - [View in archive site](#)

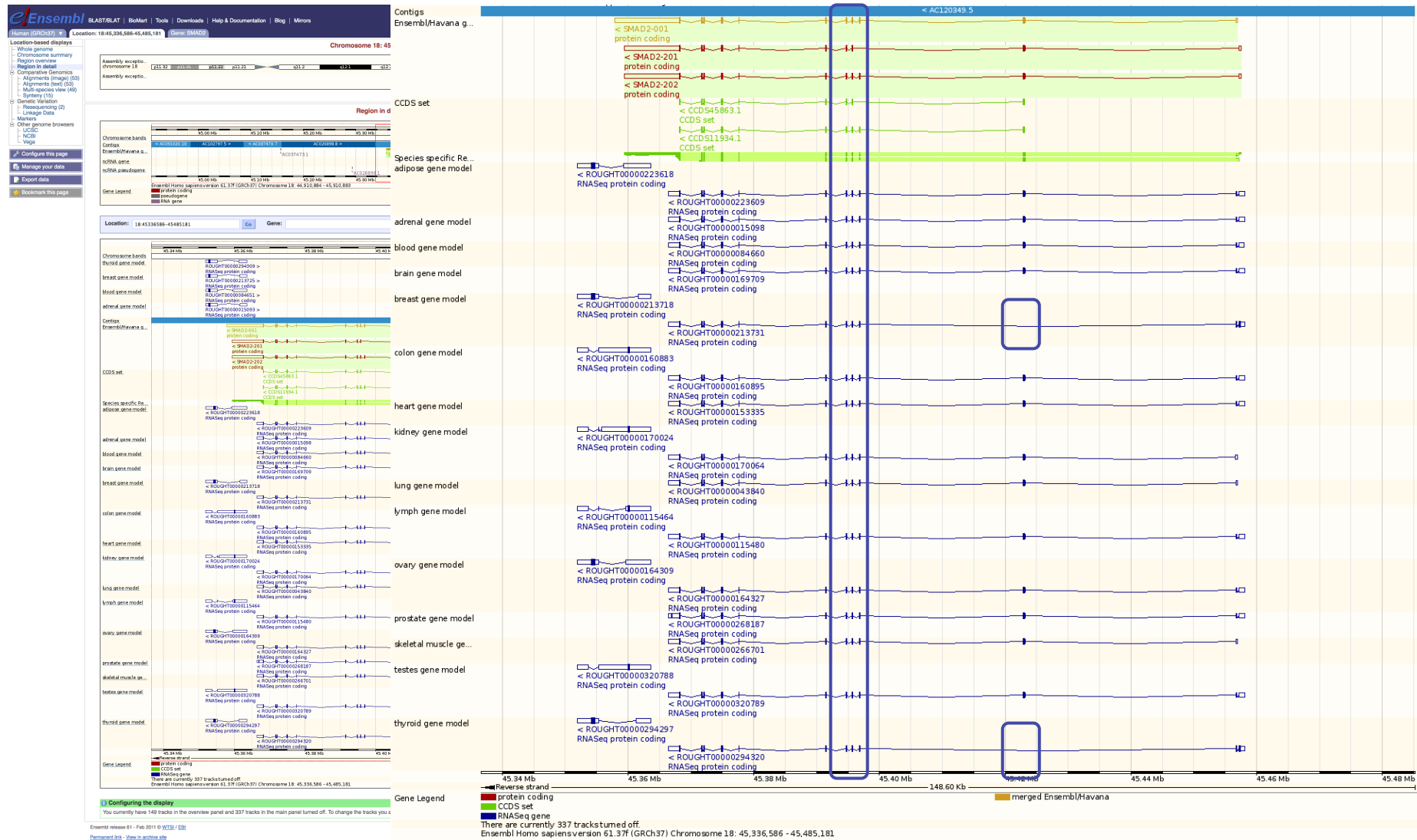
Danio rerio abcd1 – RNA-Seq UTR Extension



Danio rerio abcd1 – RNA-Seq Read Alignments

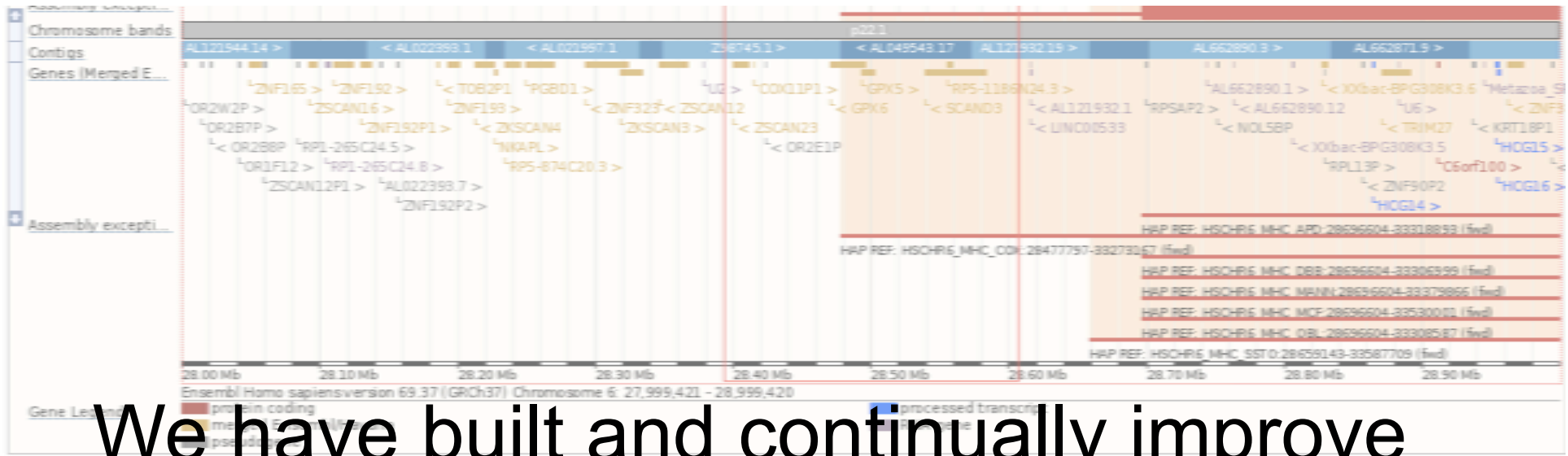


Homo sapiens SMAD2 – Illumina® Body Map 2.0

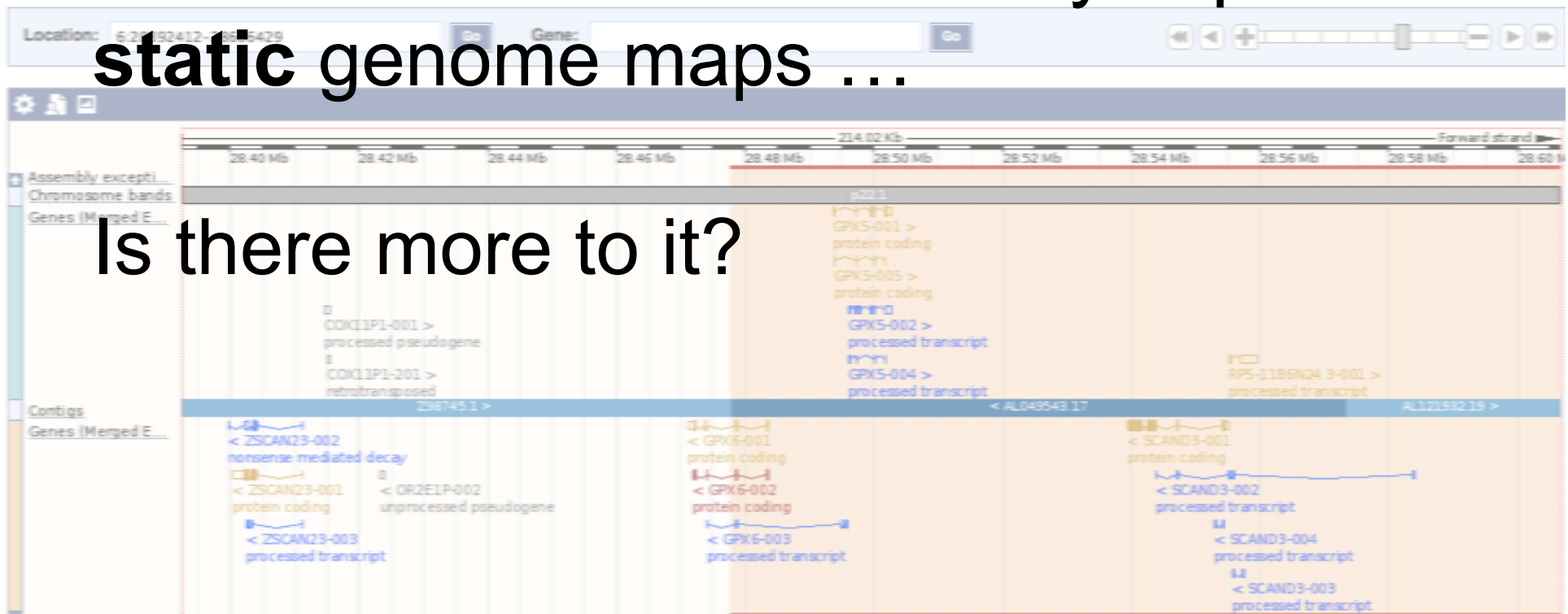


Homo sapiens MLL – Illumina® Body Map 2.0



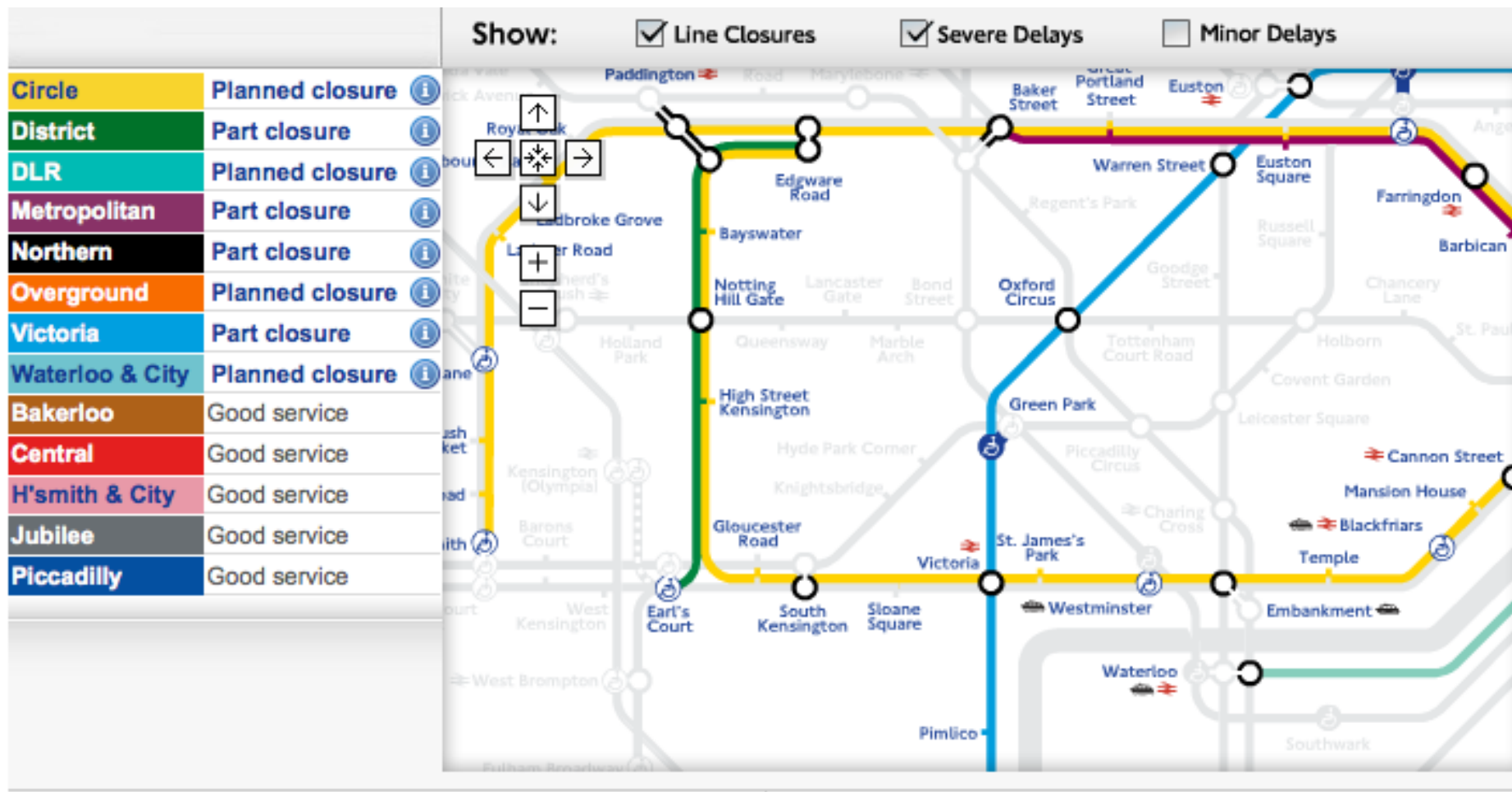


We have built and continually improve
static genome maps ...



Is there more to it?

Service Updates – A Dynamic Map



<http://www.tfl.gov.uk/tfl/livetravelnews/realtime/tube/default.html>

Epigenetics

- Epigenetics signifies stable and heritable changes in gene expression without changes in the genetic code.

- Molecular mechanisms

- DNA methylation

- In vertebrates typically occurs at CpG sites
 - Conversion of cytosine to 5-methylcytosine
 - DNA (cytosine-5-)-methyltransferases (DNMT)

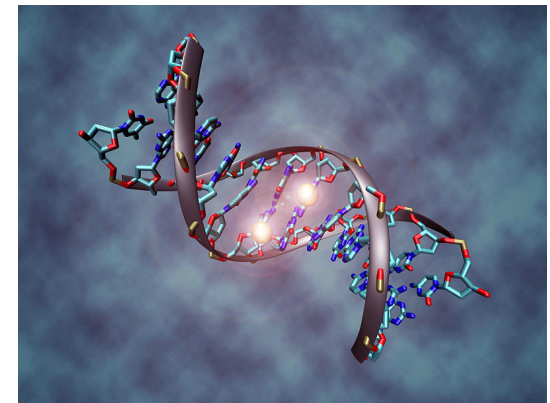
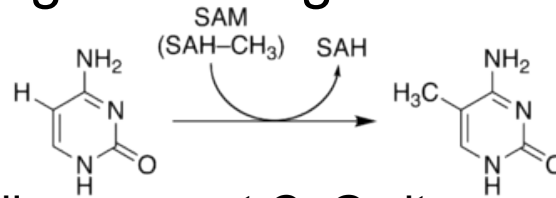
- Histone modifications (post-translational)

- Acetylation

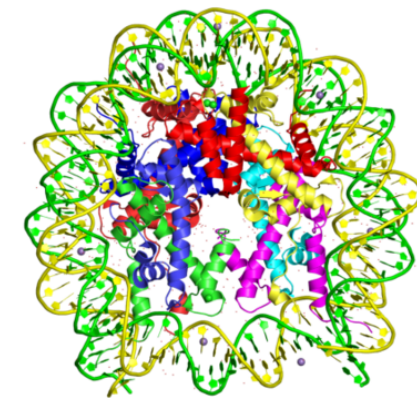
- Histone acetyltransferases (HAT)
 - Histone deacetylases (HDAC)

- Methylation (mono-, di-, tri-)

- Histone methyltransferases (HMT)
 - Histone demethylases



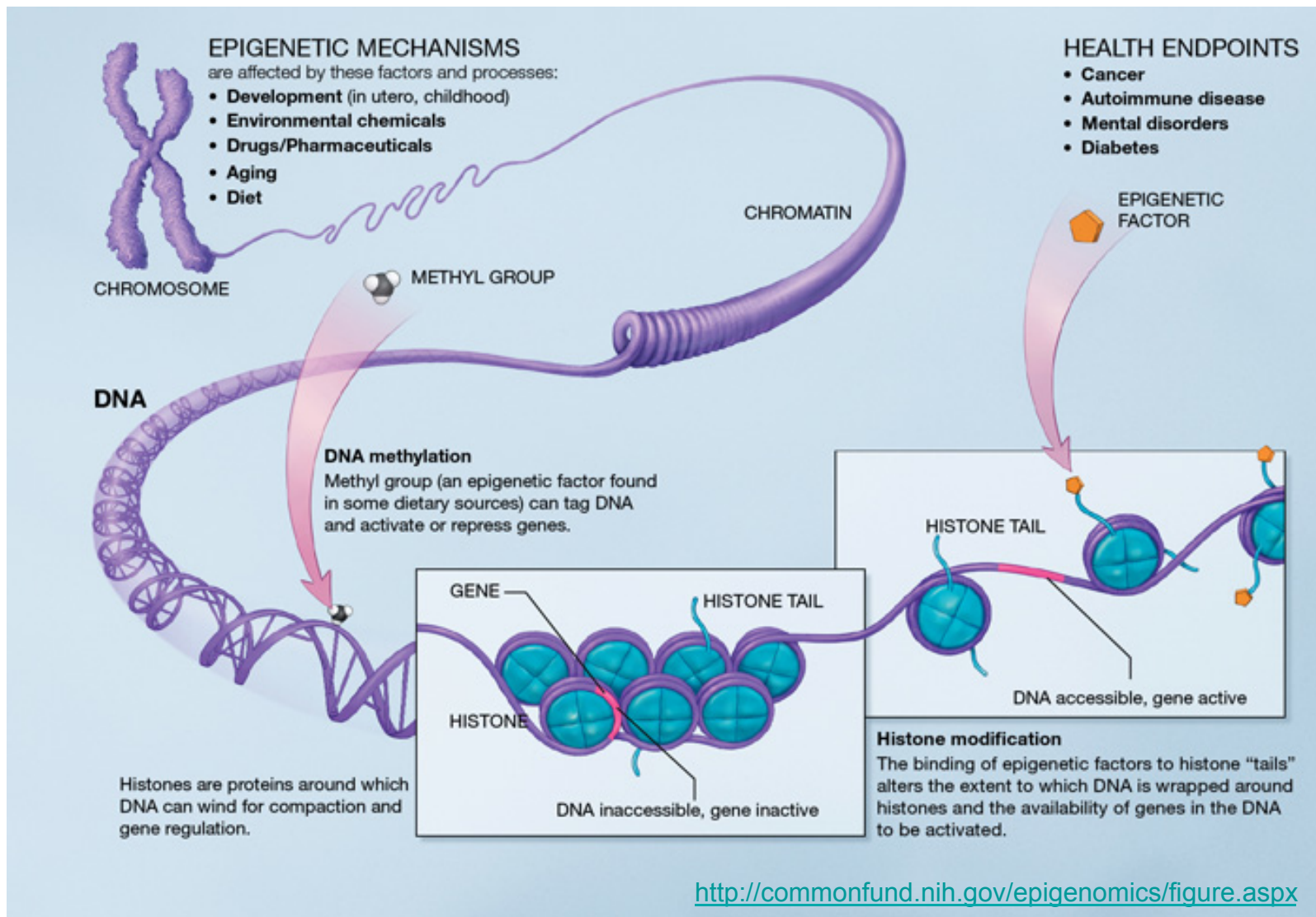
2cu5



PDB

2cu5

Epigenetic Mechanisms



<http://commonfund.nih.gov/epigenomics/figure.aspx>

The ENCODE (ENCyclopedia Of DNA Elements) Project



[Science Volume 306, Issue 5696](#)
(22 October 2004)



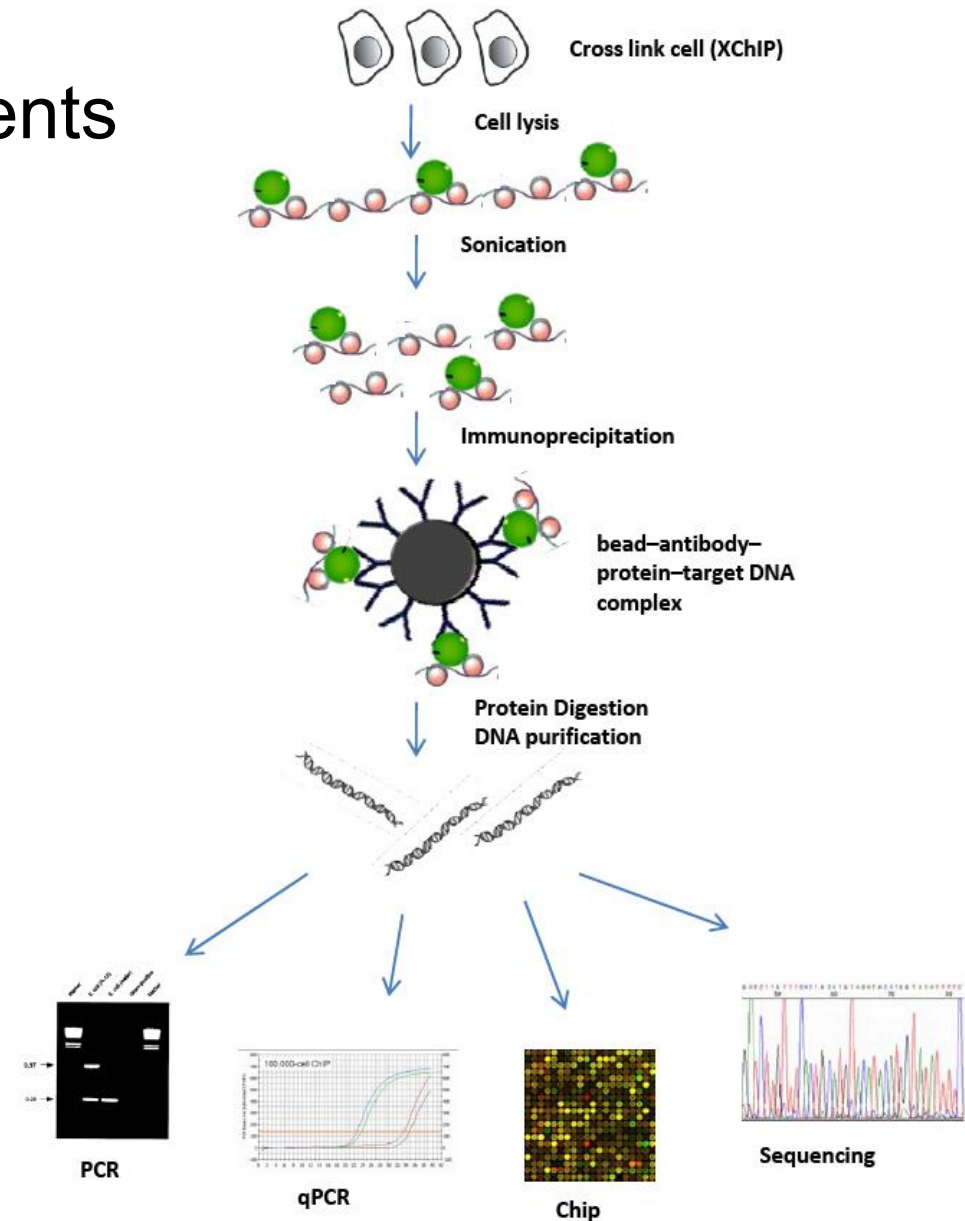
[Nature Volume 447 Number 7146](#)
(14 June 2007)



[Nature Volume 489 Number 7414](#)
(6 September 2012)

Genome-Wide Experiments

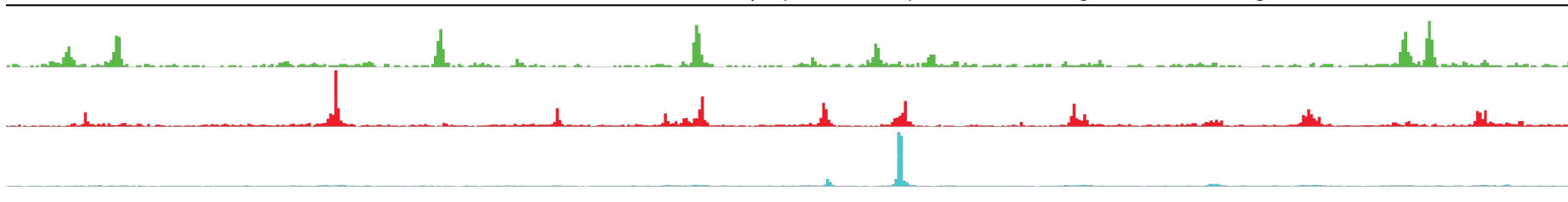
- Chromatin Immunoprecipitation (ChIP)
- Pull-down of transcription factors bound to DNA
- Pull-down of modified histones bound to DNA
- Initially, ChIP on CHIP
 - Hybridisation with tiling array (chip)
- Now, mainly ChIP-Seq
 - Massively parallel sequencing



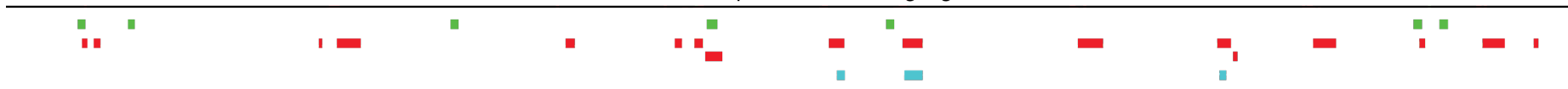
http://en.wikipedia.org/wiki/Chromatin_immunoprecipitation

Ensembl Regulatory Build – Schema

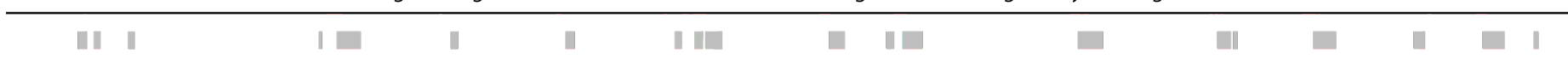
Download available data from SRA/ArrayExpress and map reads to current genome build using bwa.



Peak call with SWEMBL, reconcile replicates for binding regions (DNase 1 and TFs).



Single linkage cluster across all data and cell lines to give multicell regulatory core regions.



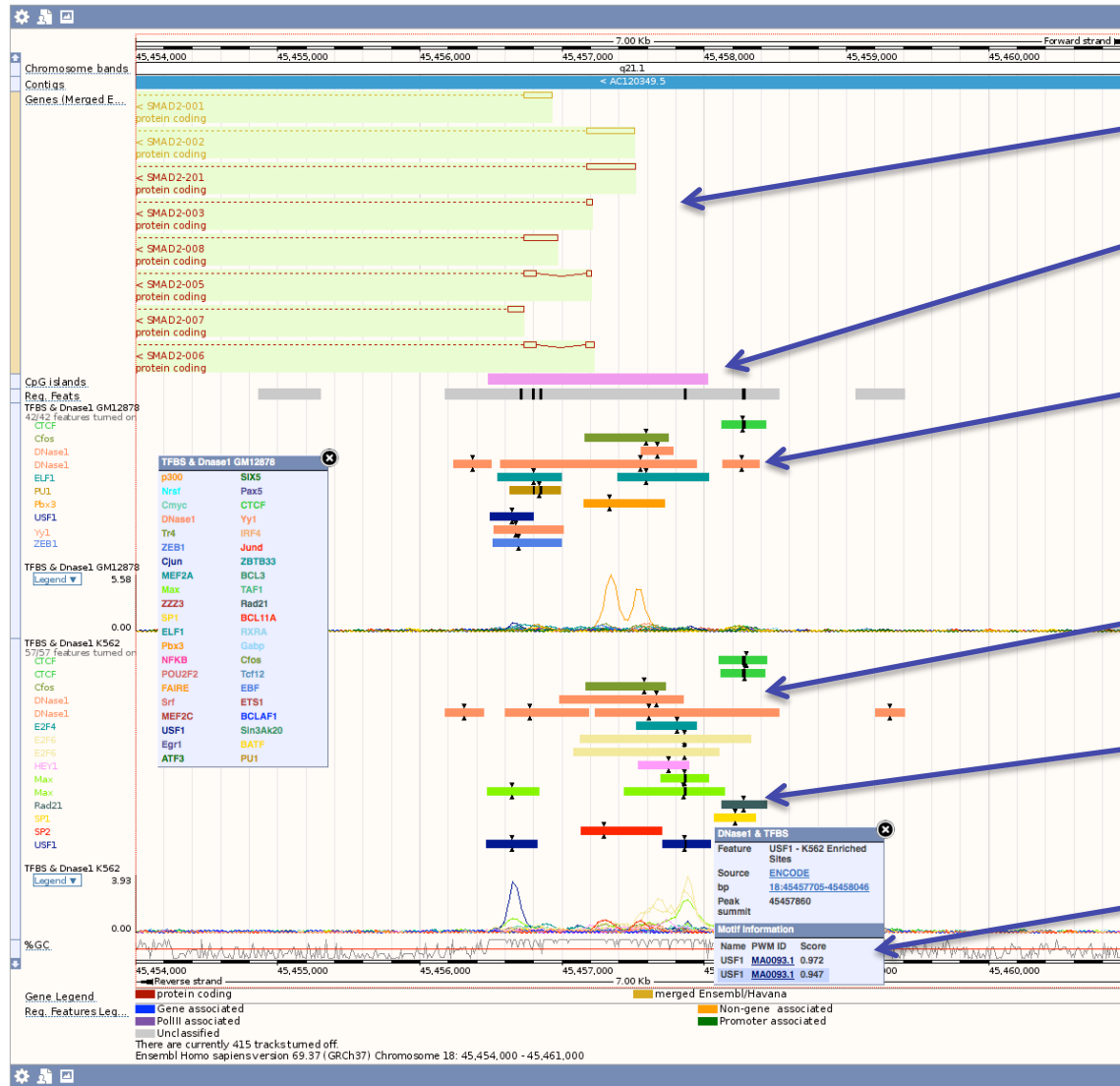
Overlap with supporting data within cell lines e.g. histone modifications, to define supporting bounds.



Apply annotation based on genomic distribution of supporting features.



SMAD2 Promoter – Transcription Factor Binding Sites



SMAD2 Transcripts

CpG Island

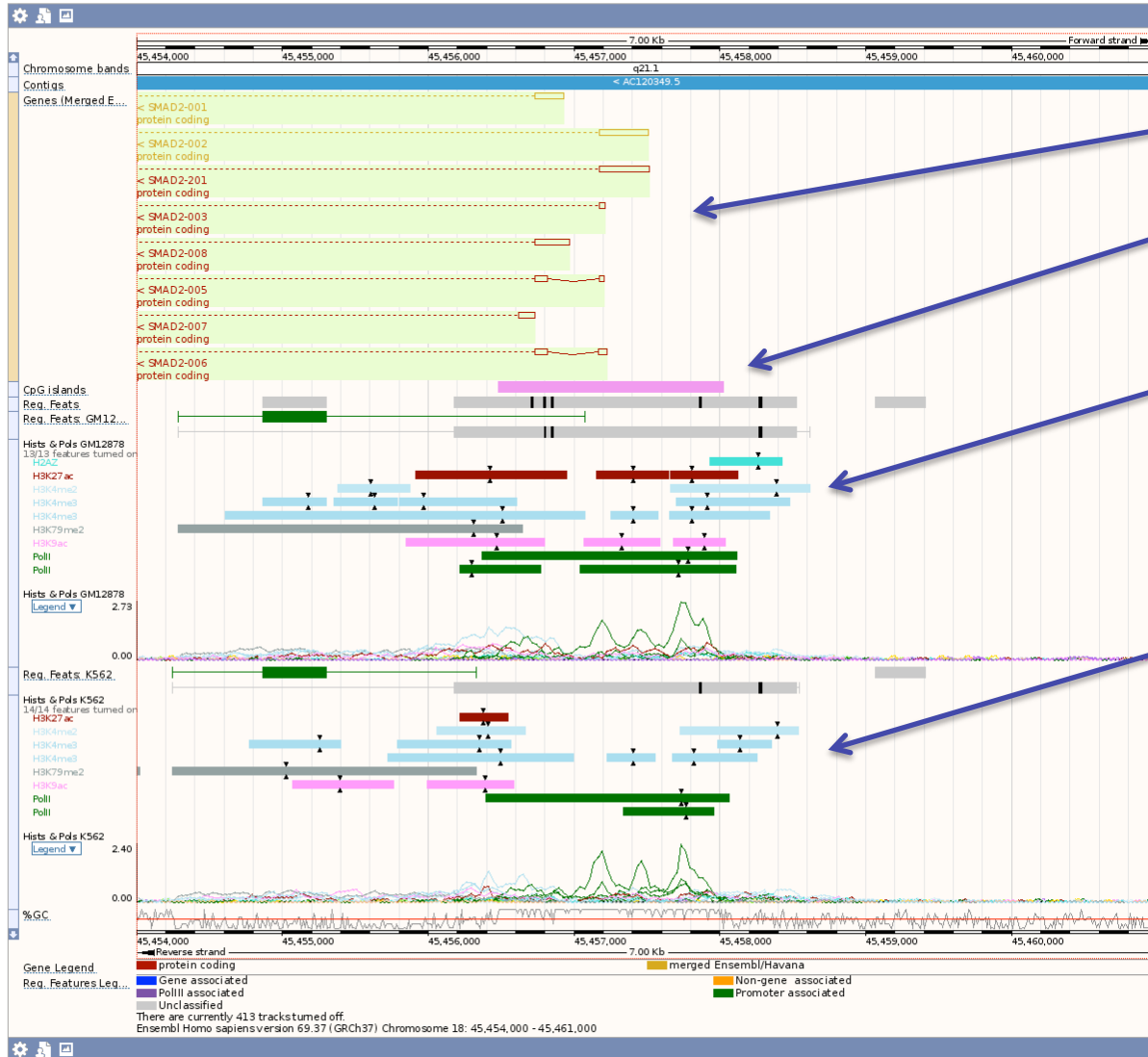
GM12878 Cell Line
Transcription Factor
Binding Sites

K562 Cell Line
Transcription Factor
Binding Sites

K562 SP1 site

K562 USF1
PWM JASPAR links

SMAD2 Promoter – Histone Modifications



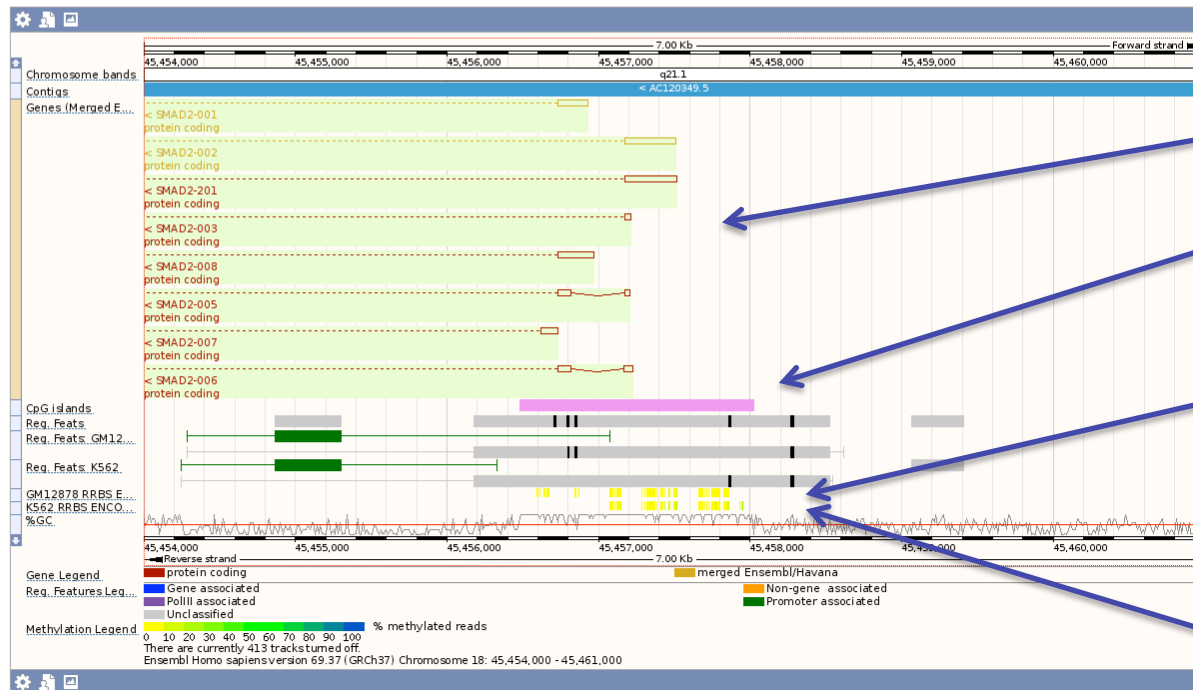
SMAD2 Transcripts

CpG Island

GM12878 Cell Line
Histone Modifications

K562 Cell Line
Histone Modifications

SMAD2 Promoter – DNA Methylation



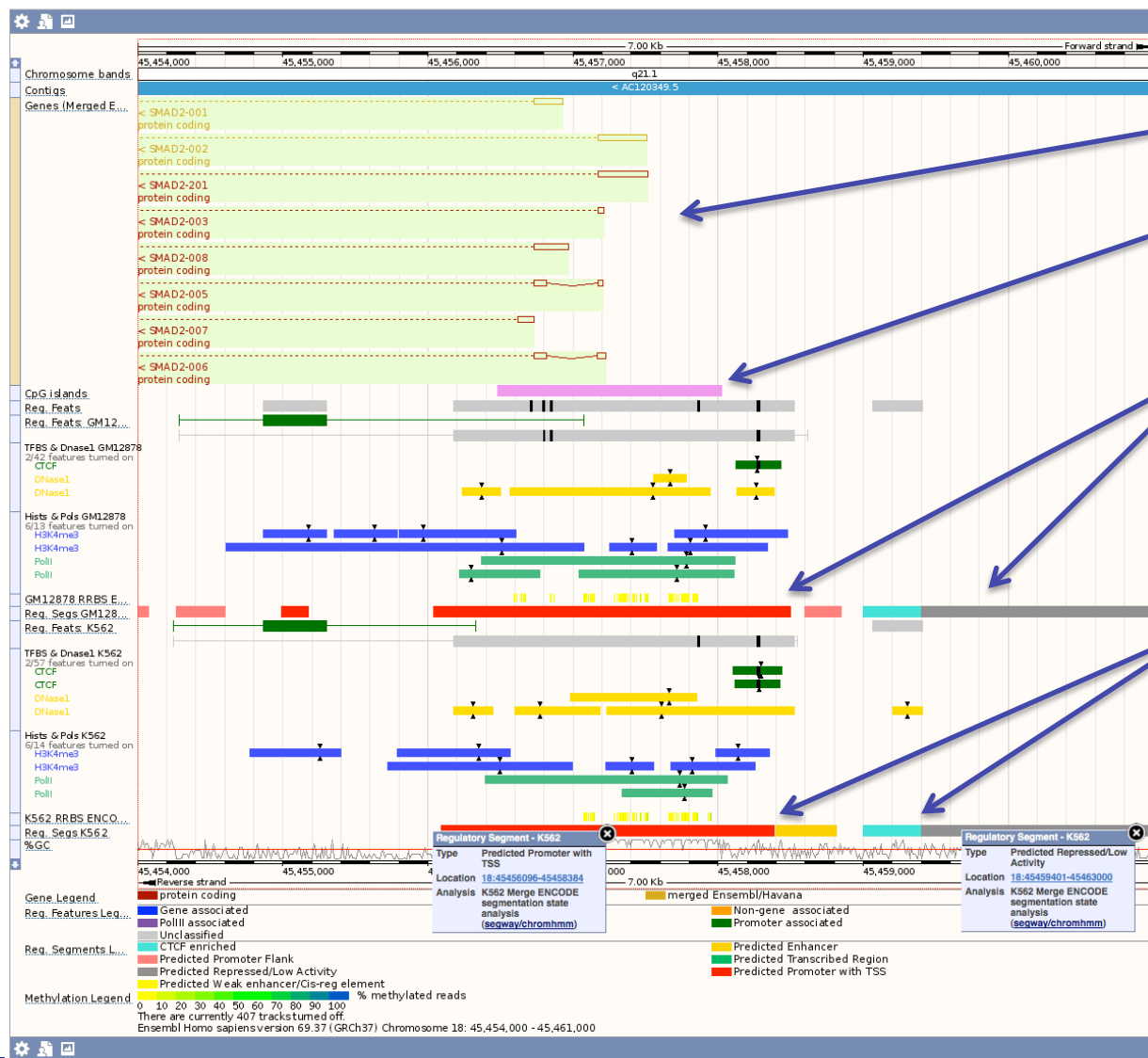
SMAD2 Transcripts

CpG Island

GM12878 Cell Line
DNA Methylation

K562 Cell Line
DNA Methylation

SMAD2 Promoter – Genome Segmentation



SMAD2 Transcripts

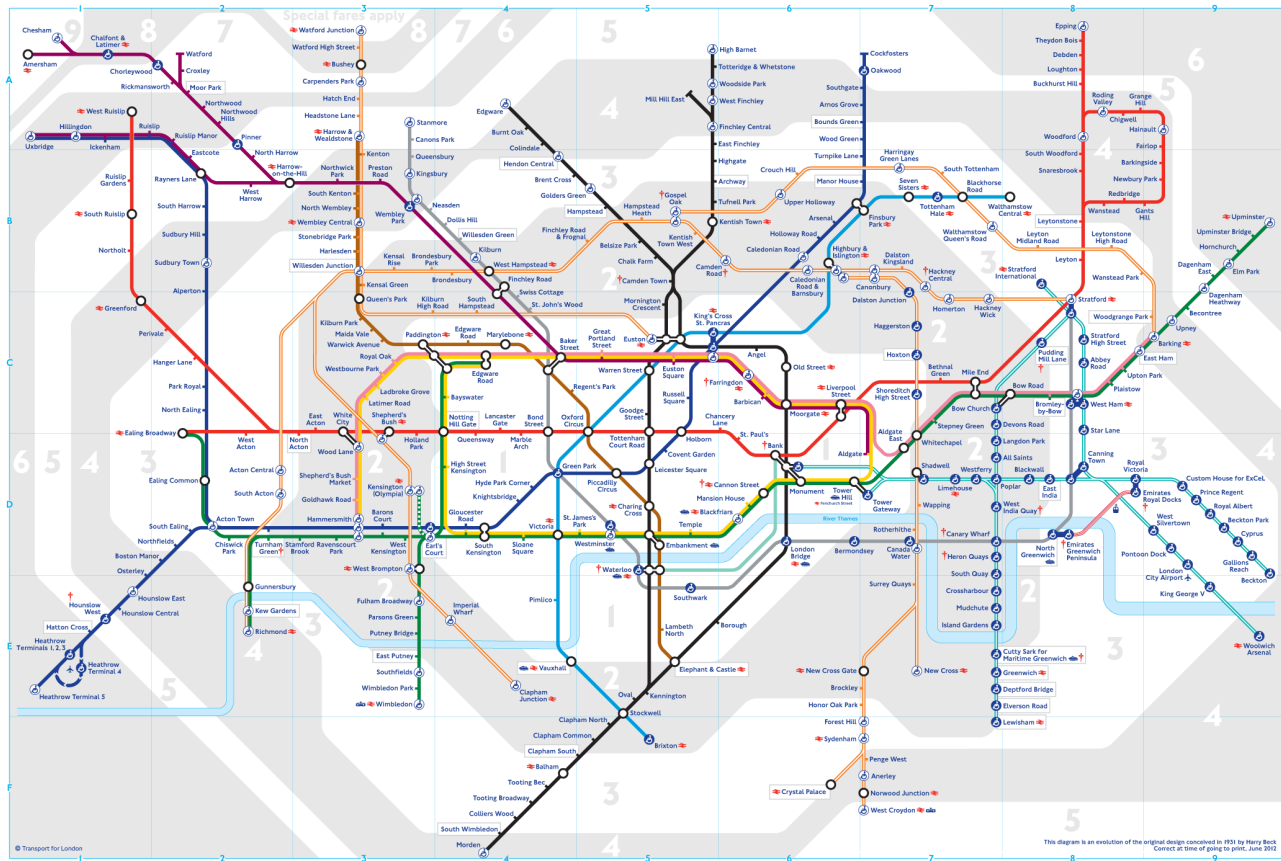
CpG Island

GM12878 Cell Line
Genome Segmentation

K562 Cell Line
Genome Segmentation

Putting It Back Together ...

Tube map



Check before you travel

Bank	Waterloo & City line open 0621-2148 Mondays to Fridays and 0802-1837 Saturdays. Closed Sundays and Public Holidays. During the London 2012 Games the line will operate 0602-0100 Mondays to Fridays and 0802-0100 Saturdays and Sundays.
Camden Road	Step-free from July 2012
Camden Town	Sunday 1300-1730 open for interchange and exit only
Canary Wharf	Step-free interchange between Underground, Canary Wharf DLR and street level
Cannon Street	Open until 2100 Mondays to Fridays and 0730-1930 Saturdays. Closed Sundays. During the London 2012 Games the station will be operating the same opening and closing times as other LU stations.
Cuttly Sark for Maritime Greenwich	Station is due to be closed at certain times and dates during the London 2012 Games. Check before you travel
Emirates Greenwich Peninsula	Opening summer 2012. Special fares apply
Emirates Royal Docks	Opening summer 2012. Special fares apply
Farringdon	Step-free from July 2012
Gospel Oak	Step-free from July 2012
Hackney Central	Step-free from July 2012
Heron Quays	Step-free interchange between Heron Quays and Canary Wharf Underground station at street level
Hounslow West	Step-free access for wheelchair users only
Padding Mill Line	Station closed Saturday 14 July until Wednesday 15 September inclusive
Turnham Green	Served by Piccadilly line trains until 0650 Monday to Saturday, 0745 Sunday and after 2230 every evening. At other times use District line
Waterloo	Waterloo & City line open 0615-2141 Mondays to Fridays and 0802-1837 Saturdays. Closed Sundays and Public Holidays. During the London 2012 Games the line will operate 0615-0100 Monday to Fridays and 0800-0100 Saturdays and Sundays.
West India Quay	Not served by DLR trains from Bank towards Lewisham before 1900 on Mondays to Fridays and at any time between 16 July and 9 September 2012. Please check the Step-free Tube Guide for more detailed information.

Key to lines

- Bakerloo
- Central
- Circle
- District
- Greenwich
- Hammersmith & City
- Jubilee
- Metropolitan
- Northern
- Piccadilly
- Victoria
- Waterloo & City
- DLR
- London Overground
- Emirates Air Line

MAYOR OF LONDON

Website tfl.gov.uk

24 hour travel information 0843 222 1234*

*You pay no more than 5p per minute if calling from a BT landline. There may be a connection charge. Charges from mobiles or other landline providers may vary.

Travel information at stations Help points

Transport for London UNDERGROUND



Ensembl Project – Team

*D84–D90 Nucleic Acids Research, 2012, Vol. 40, Database issue
doi:10.1093/nar/gkr991*

Published online 15 November 2011

Ensembl 2012

Paul Flicek^{1,2,*}, M. Ridwan Amode², Daniel Barrell², Kathryn Beal¹, Simon Brent², Denise Carvalho-Silva¹, Peter Clapham², Guy Coates², Susan Fairley², Stephen Fitzgerald¹, Laurent Gil¹, Leo Gordon¹, Maurice Hendrix², Thibaut Hourlier², Nathan Johnson¹, Andreas K. Kähäri¹, Damian Keefe¹, Stephen Keenan¹, Rhoda Kinsella¹, Monika Komorowska¹, Gautier Koscielny¹, Eugene Kulesha¹, Pontus Larsson¹, Ian Longden¹, William McLaren¹, Matthieu Muffato¹, Bert Overduin¹, Miguel Pignatelli¹, Bethan Pritchard², Harpreet Singh Riat², Graham R. S. Ritchie¹, Magali Ruffier², Michael Schuster¹, Daniel Sobral¹, Y. Amy Tang², Kieron Taylor¹, Stephen Trevanion², Jana Vandrovcova¹, Simon White², Mark Wilson², Steven P. Wilder¹, Bronwen L. Aken², Ewan Birney¹, Fiona Cunningham¹, Ian Dunham¹, Richard Durbin², Xosé M. Fernández-Suarez¹, Jennifer Harrow², Javier Herrero¹, Tim J. P. Hubbard², Anne Parker², Glenn Proctor¹, Giulietta Spudich¹, Jan Vogel², Andy Yates¹, Amonida Zadissa² and Stephen M. J. Searle²

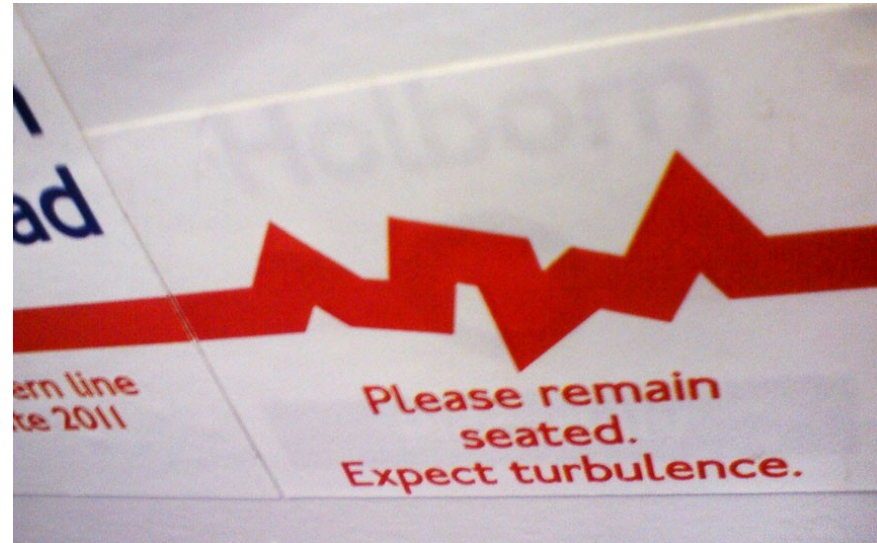
¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton Cambridge CB10 1SD, UK and

²Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Q&A

QUESTIONS

ANSWERS



<http://www.telegraph.co.uk/news/picturegalleries/uknews/9601453/Spoof-signs-brighten-up-the-tube-guerrilla-stickers-on-London-Underground.html?frame=2366141>