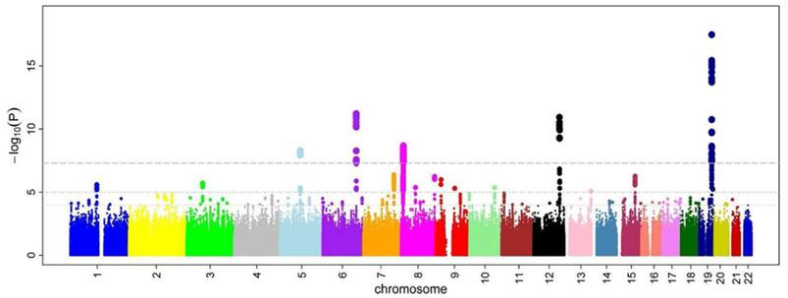
DEPTH: A Novel Algorithm for Feature Ranking with Application to Genome-Wide Association Studies

DEPTH Algorithm

- Novel algorithm for feature ranking
- Ultra-high dimensional data sets
  - e.g., genome-wide association studies, epigenome-wide association studies
- Which features correspond to signal/noise?

Conventional Analysis of GWAS data (1)

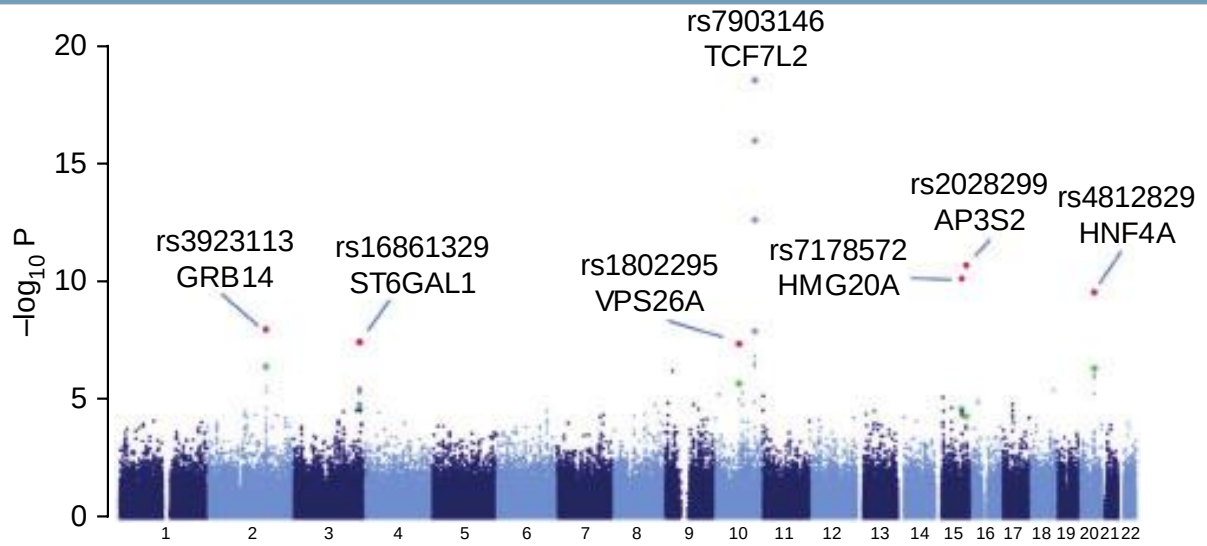
- Genome-wide Association Studies
  - Case-control setup
  - *N* samples, *p* markers (SNPs)
  - Allele frequency in cases vs. controls
  - Report effect size as an odds ratio



Conventional Analysis of GWAS data (2)

- Conventional approach to analysis
  - Test each marker independently
  - Calculate frequentist *p*-value for each marker
  - Adjust for multiple testing
  - All *p*-values less than threshold considered true associations
    - e.g., Genome-wide significance threshold

Manhattan Plot and Skyline

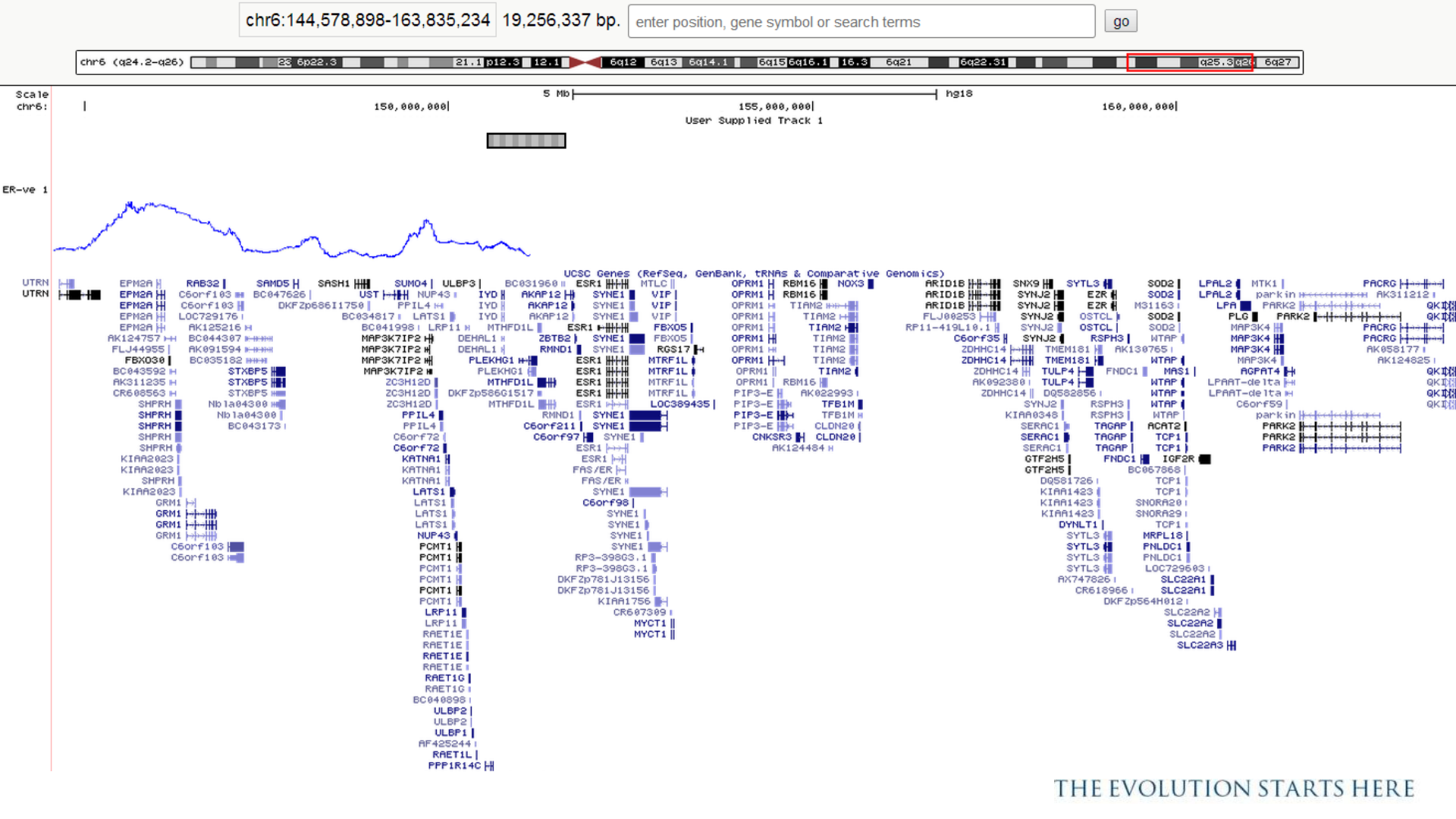


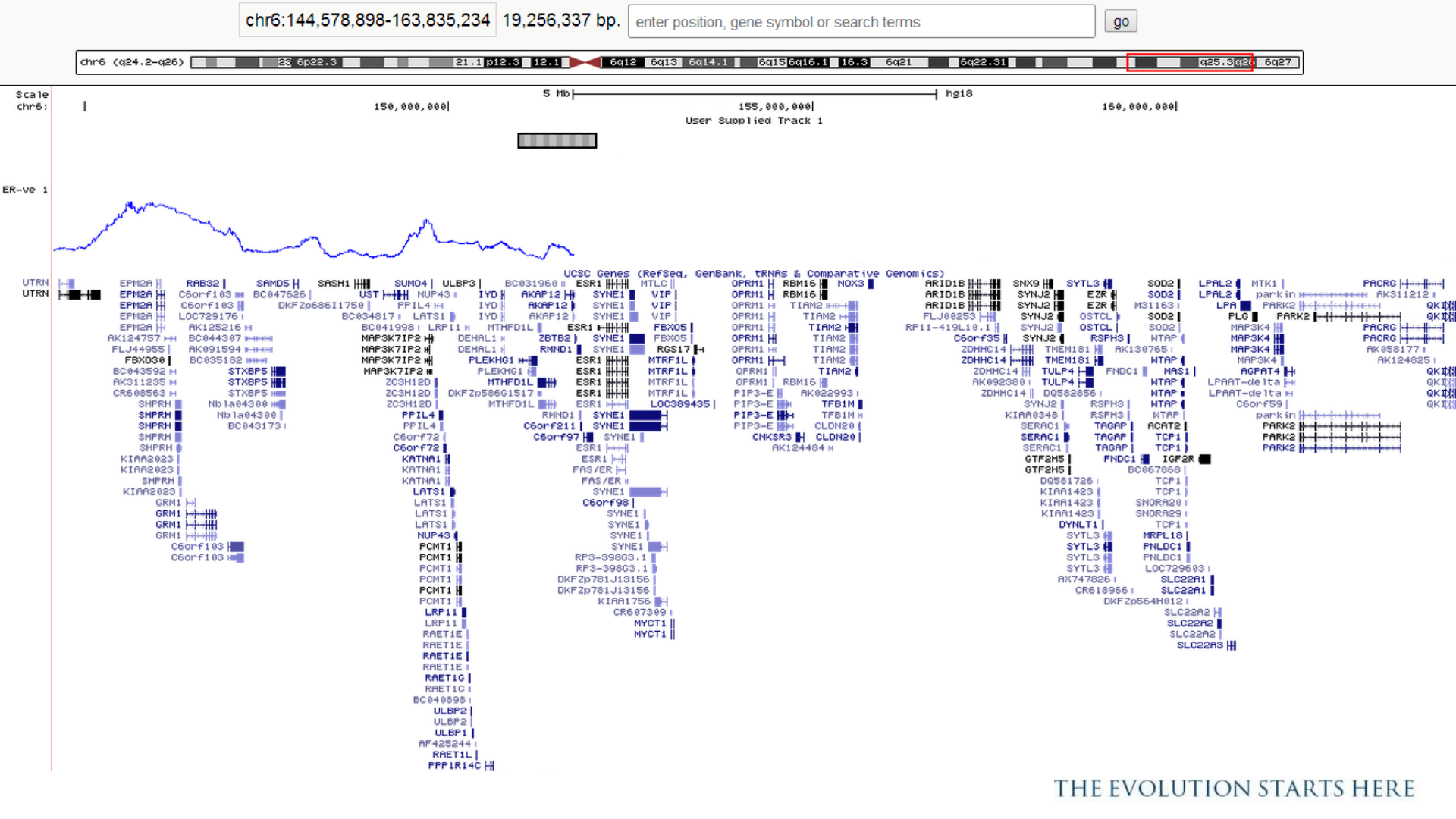
Conventional Analysis of GWAS data (3)

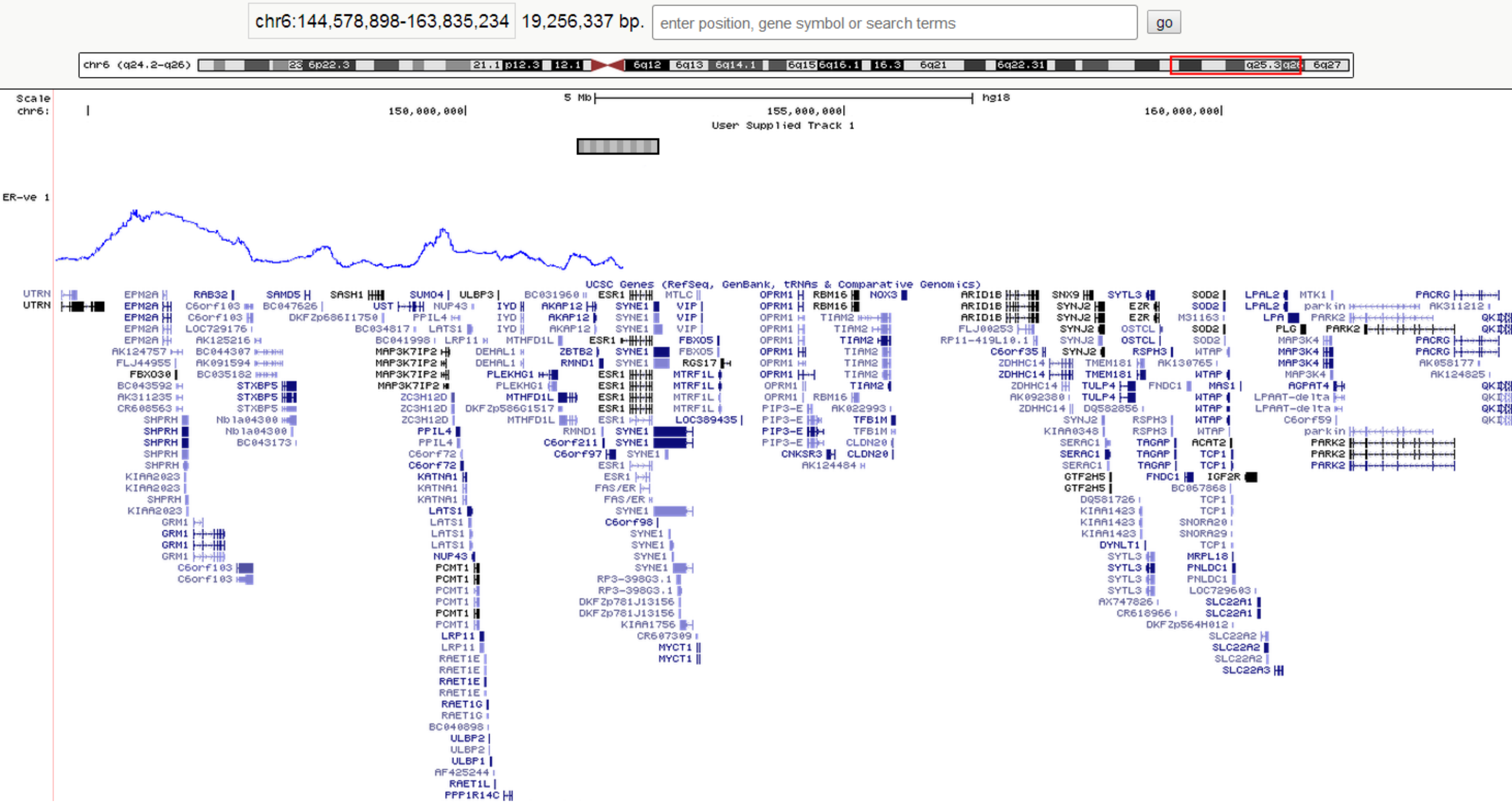
- Difficult statistical problem
  - Large number of markers
  - Correlated data
  - Disease causing variants not measured
  - Conventional analysis minimises family-wise error rate
    - Highly conservative publications

DEPTH (1)

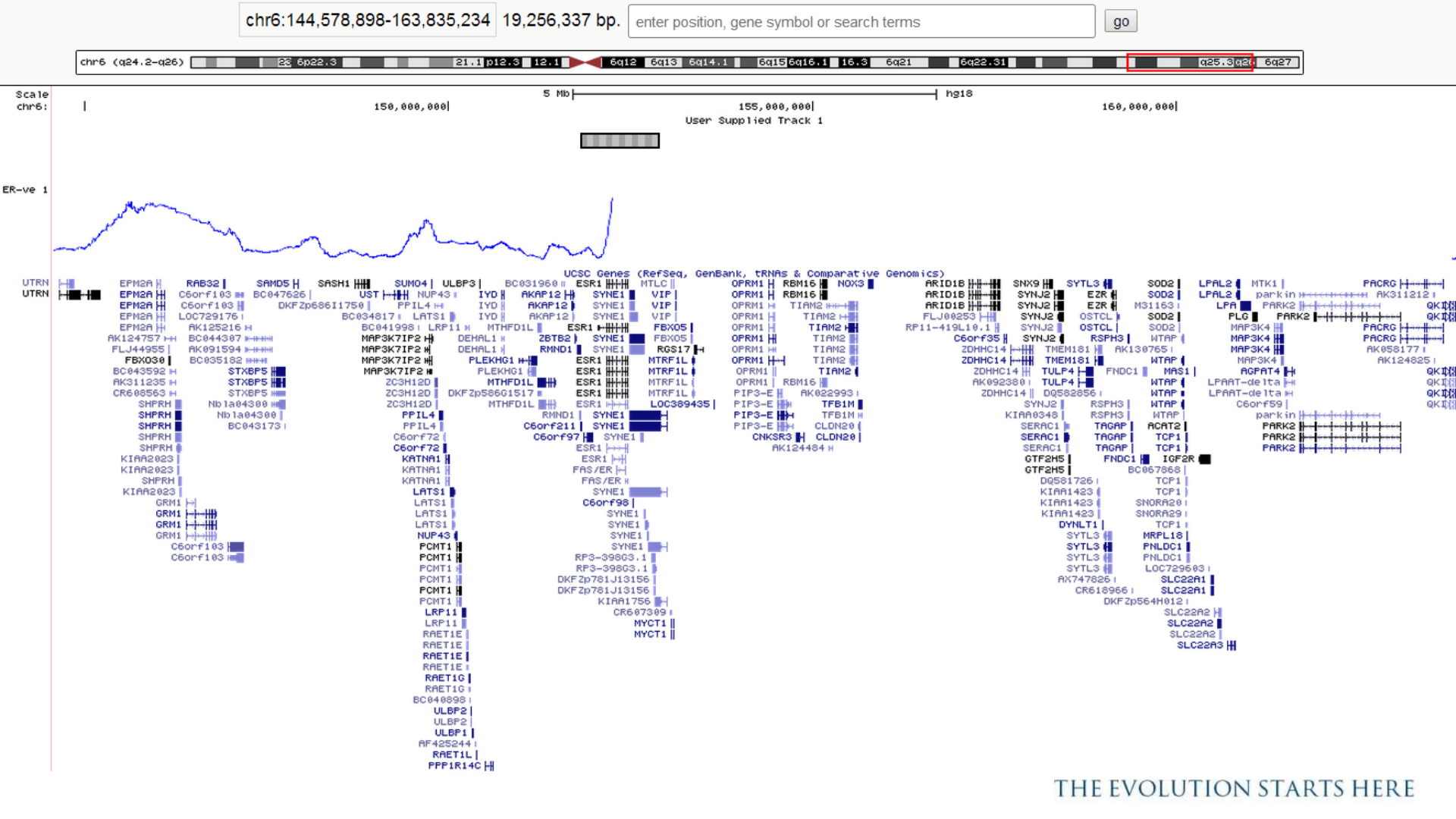
- Our strategy: DEPTH
  - NHMRC Project grant
  - Designed to run in a parallel environment
  - Exploits data correlation structure
  - Examine all markers, or subset of markers
    - e.g., all markers in a gene or pathway of genes

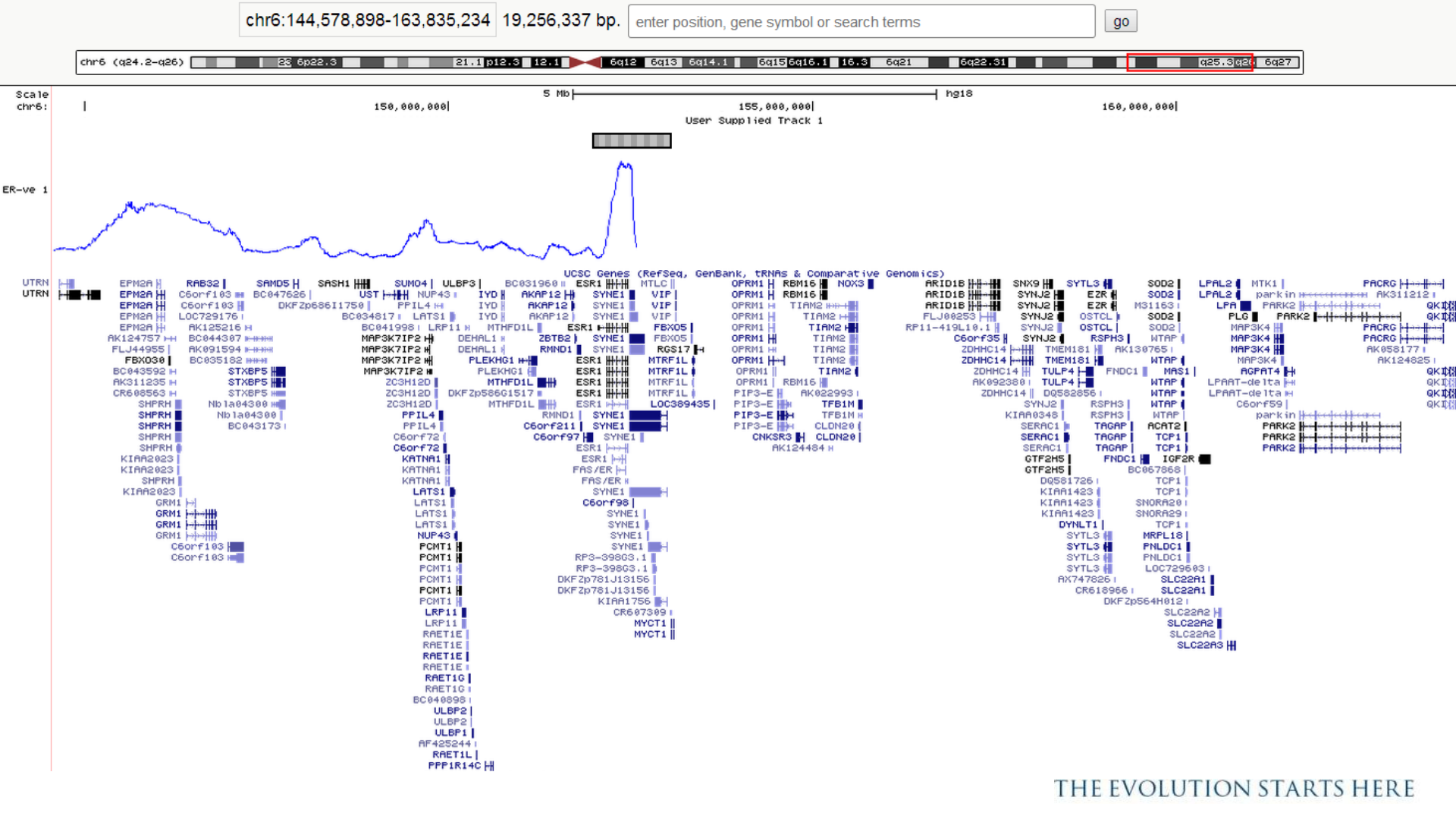

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

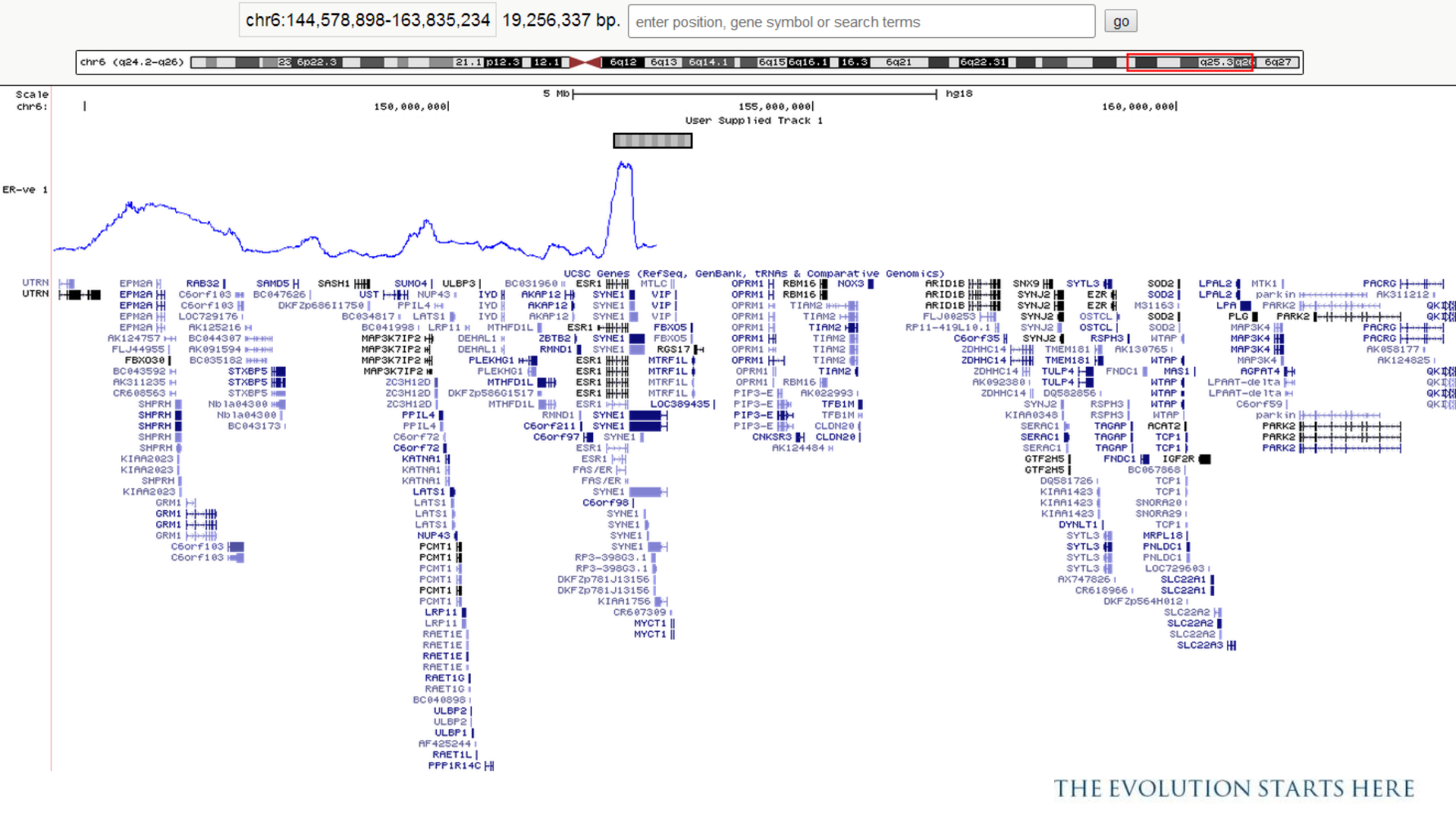

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

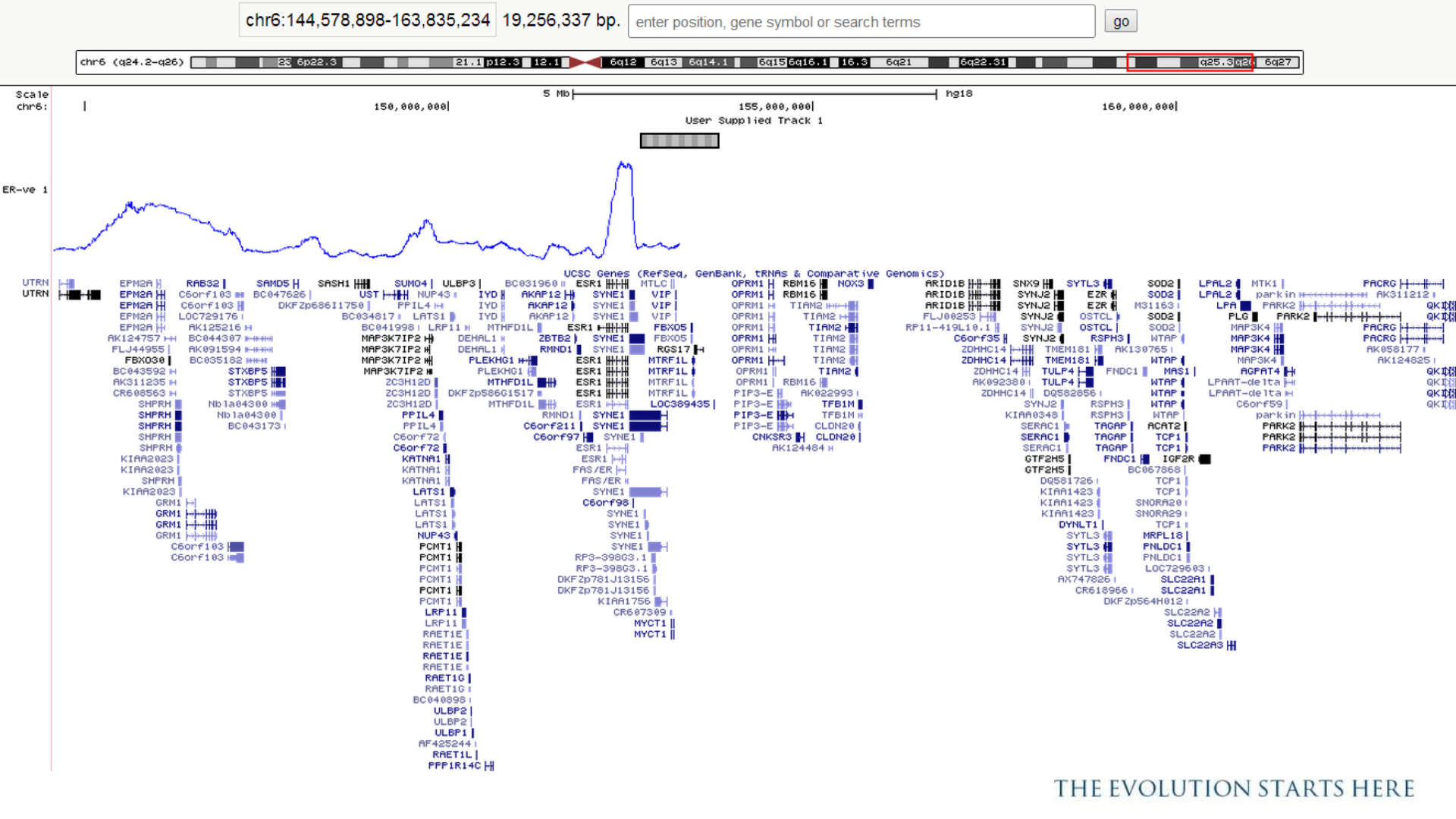


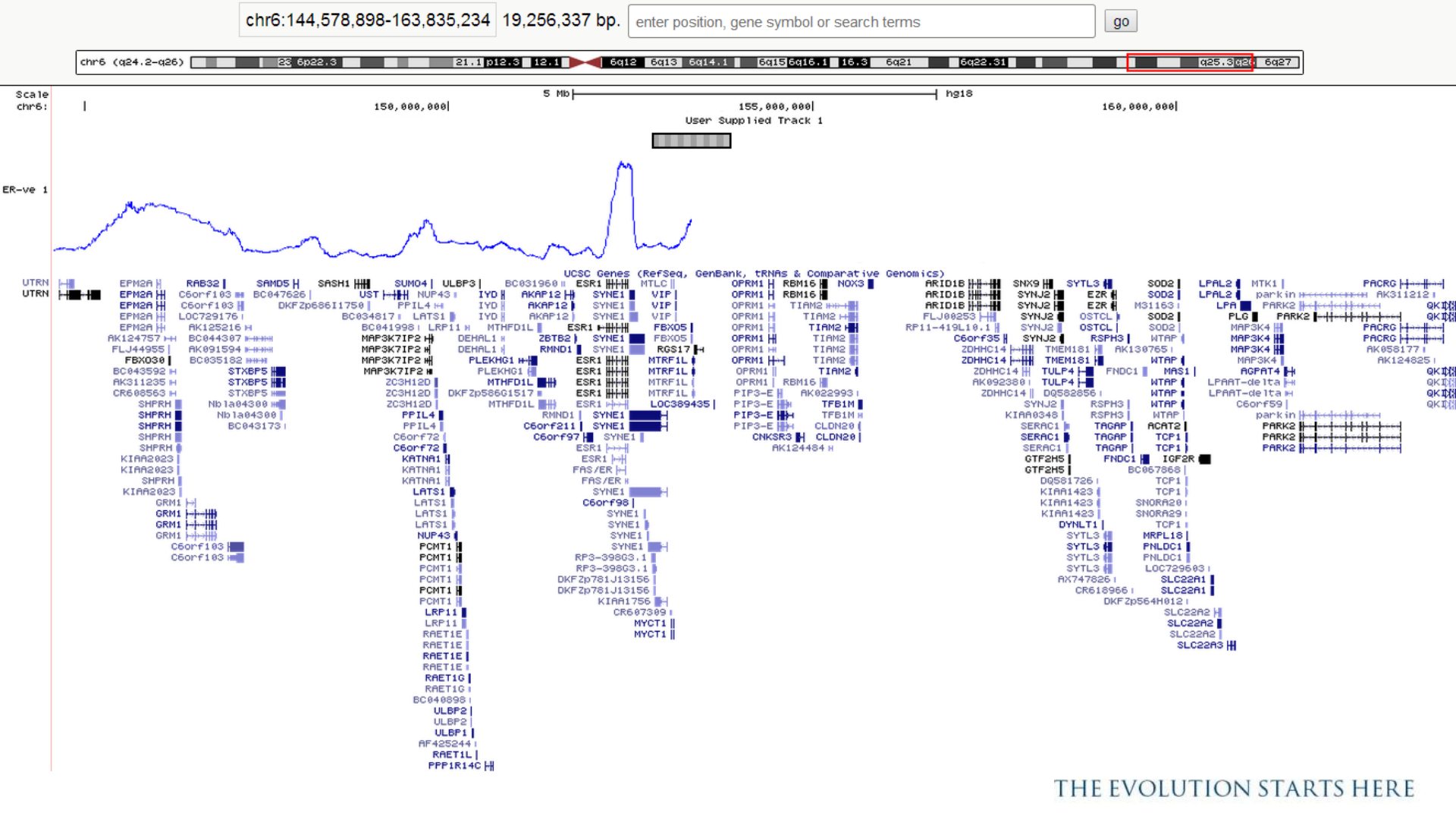
THE EVOLUTION STARTS HERE

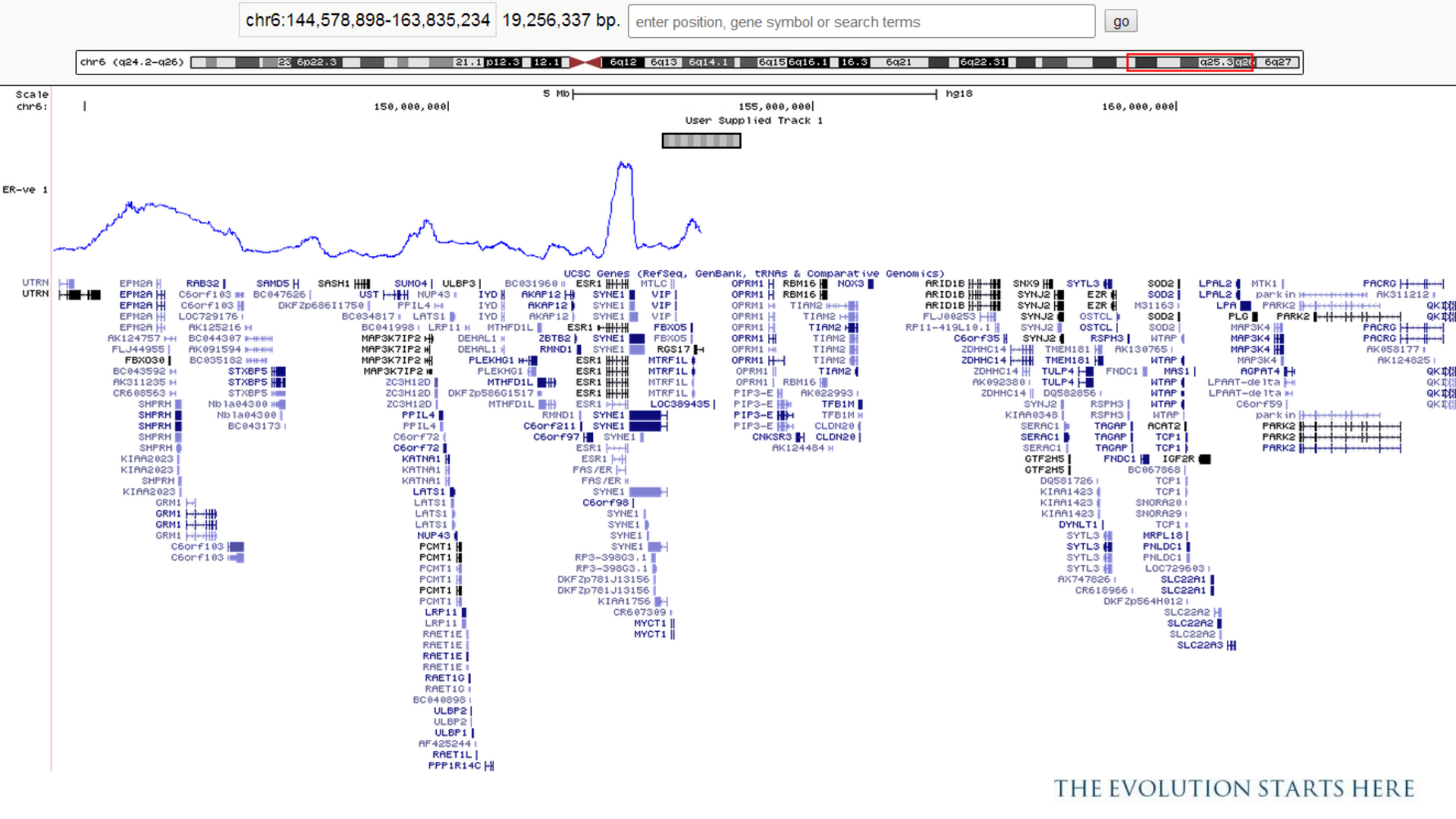

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

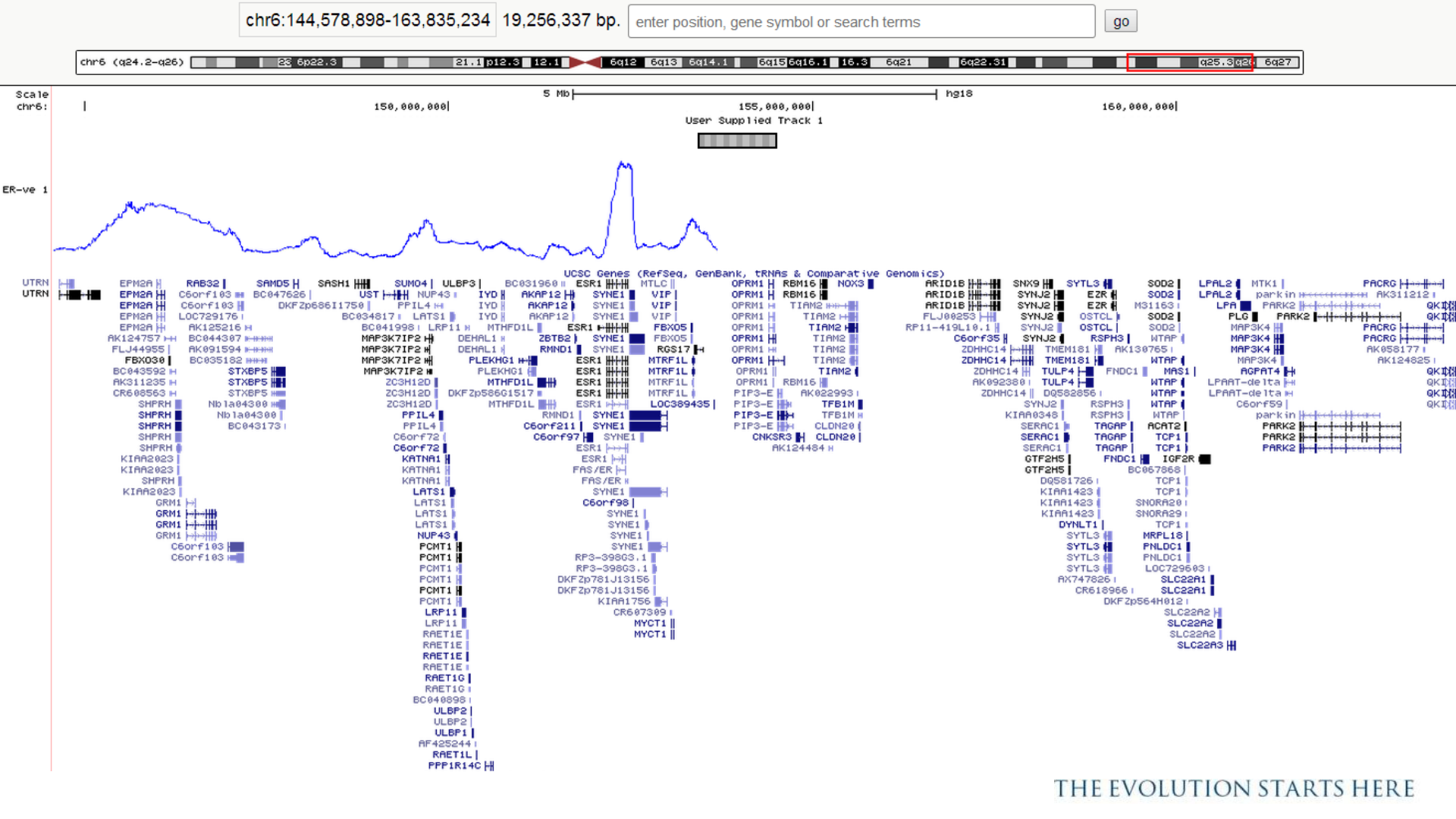

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

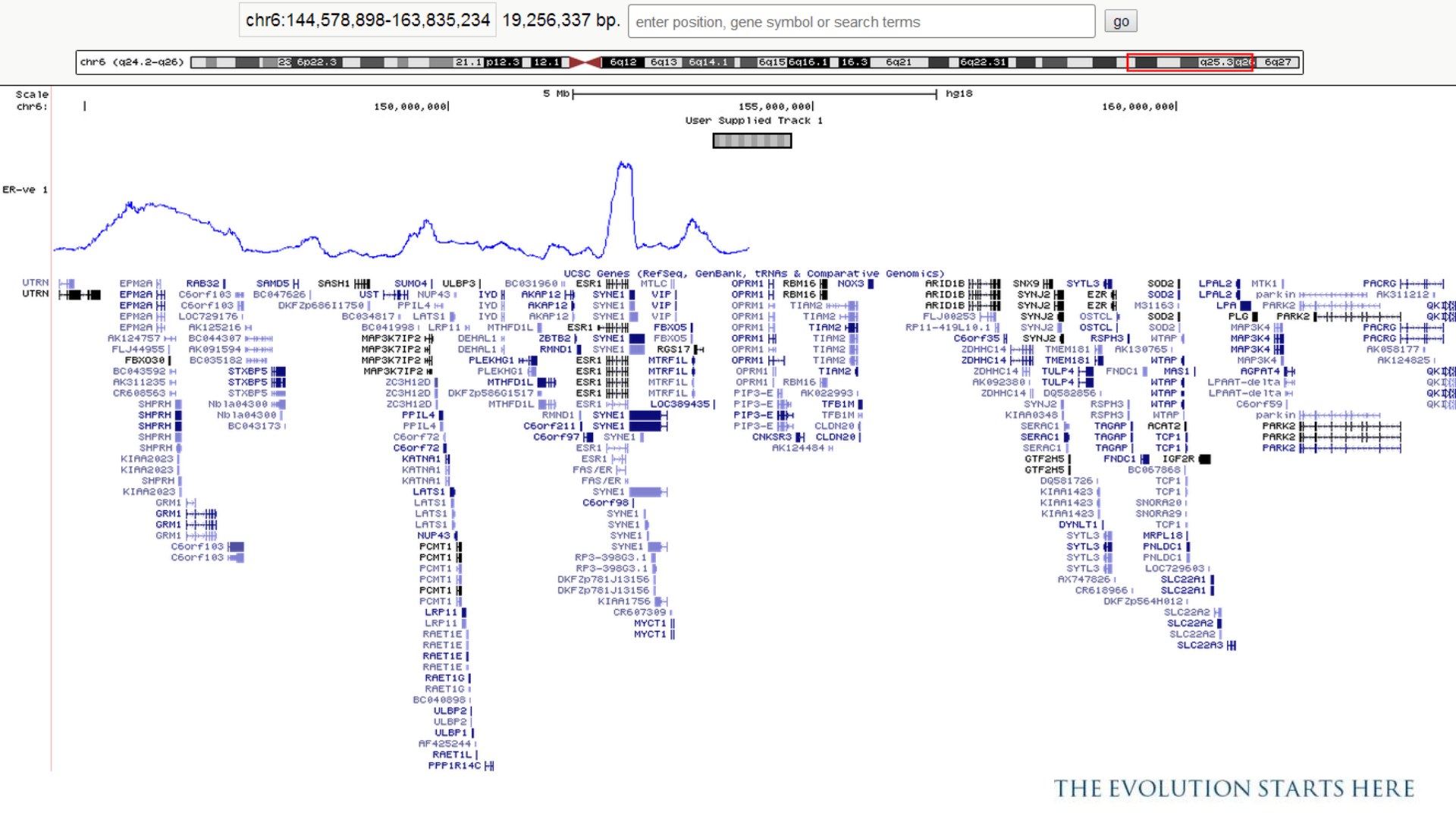

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

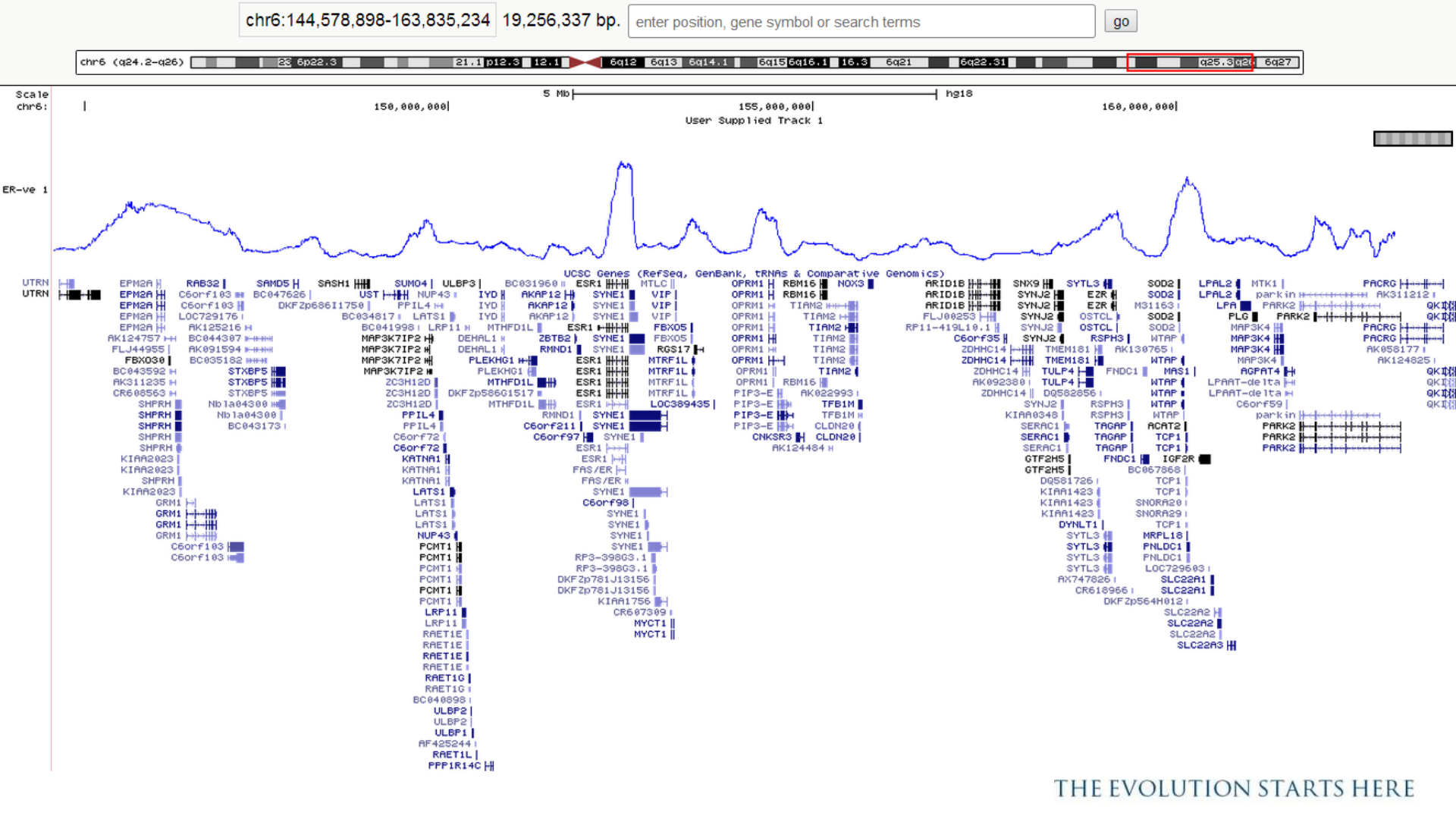

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE



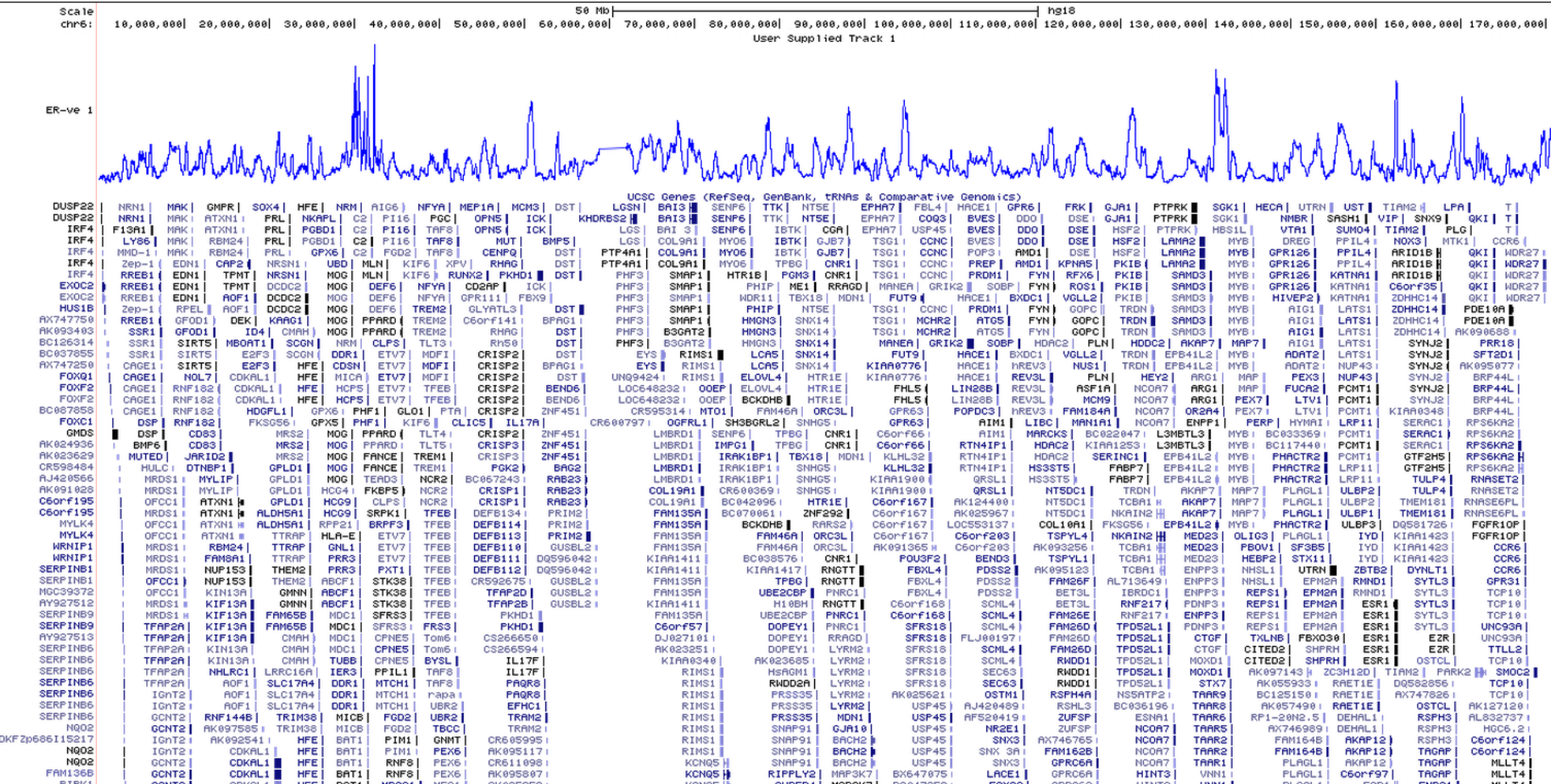

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE



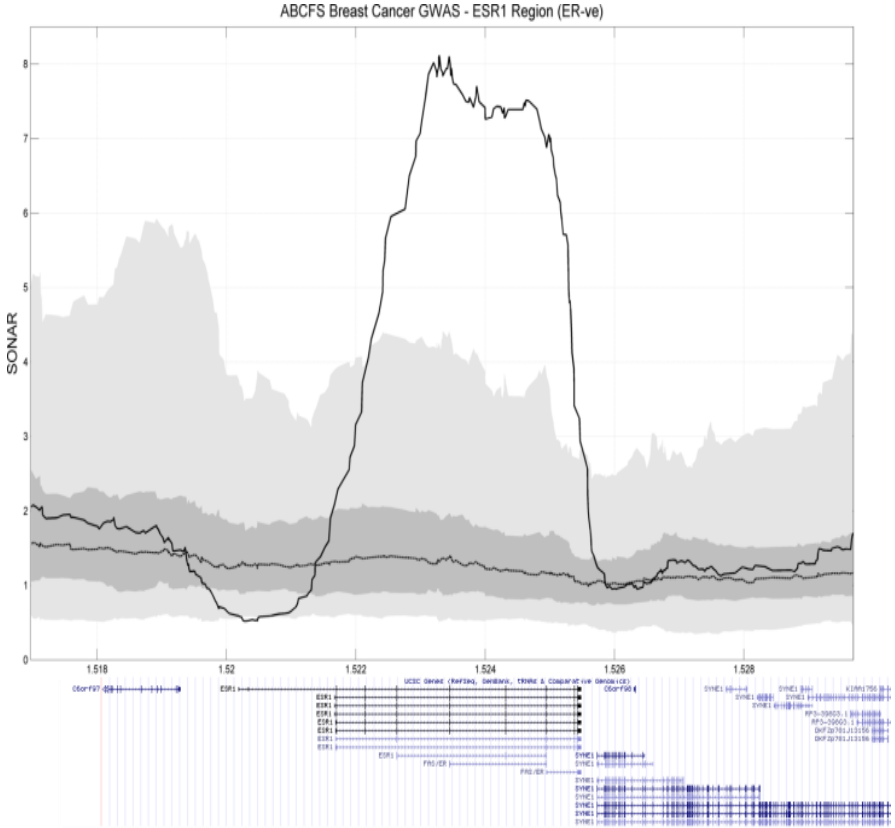

chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE


chr6:144,578,898-163,835,234
19,256,337 bp.
enter position, gene symbol or search terms
go
chr6 (q24.2-q26)
Scale
chr6:
5 Mb
hg18
150,000,000
155,000,000
160,000,000
User Supplied Track 1
ER-ve 1
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
UTRN
UTRN
THE EVOLUTION STARTS HERE

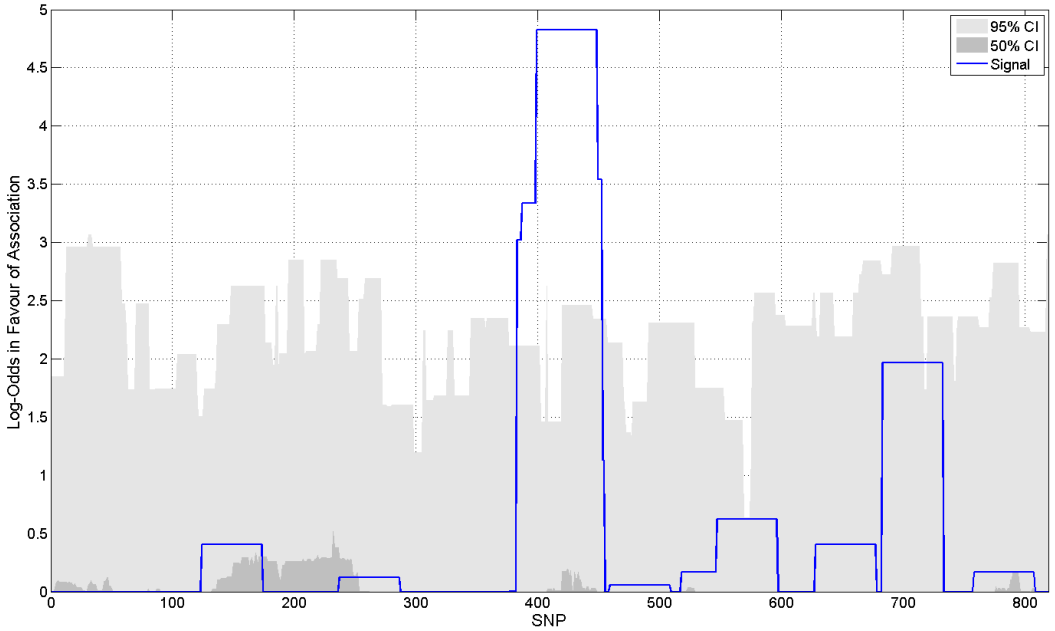
chr6:1-170,556,128 170,556,128 bp. enter position, gene symbol or search terms go



Is this peak “statistically significant”?



A non-parametric measure of signal



DEPTH and CCFR Colorectal Cancer GWAS data (1)

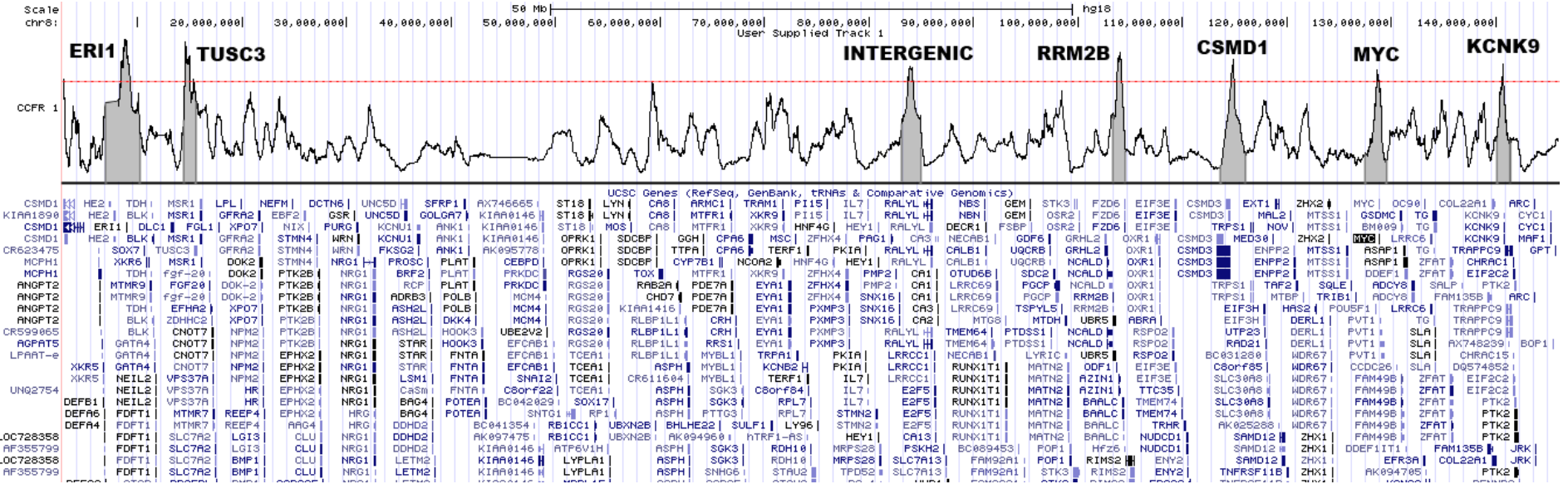
- CCFR Phase I GWAS data set
  - 1,179 cases and 998 controls
  - 2,121,264 markers
  - Standard QC
  - DEPTH genome-wide analysis (NHMRC Project Grant)

DEPTH and CCFR Colorectal Cancer GWAS data (2)

- DEPTH identified ~60 genomic regions associated with risk of CRC

- The genomic regions:
  - Some already known (e.g., *ERCC6*, *SMAD7*, *DCC*)
  - Many novel regions (e.g., *TUSC3, VIM, LIMA1*)

DEPTH and CCFR Colorectal Cancer GWAS data (3)

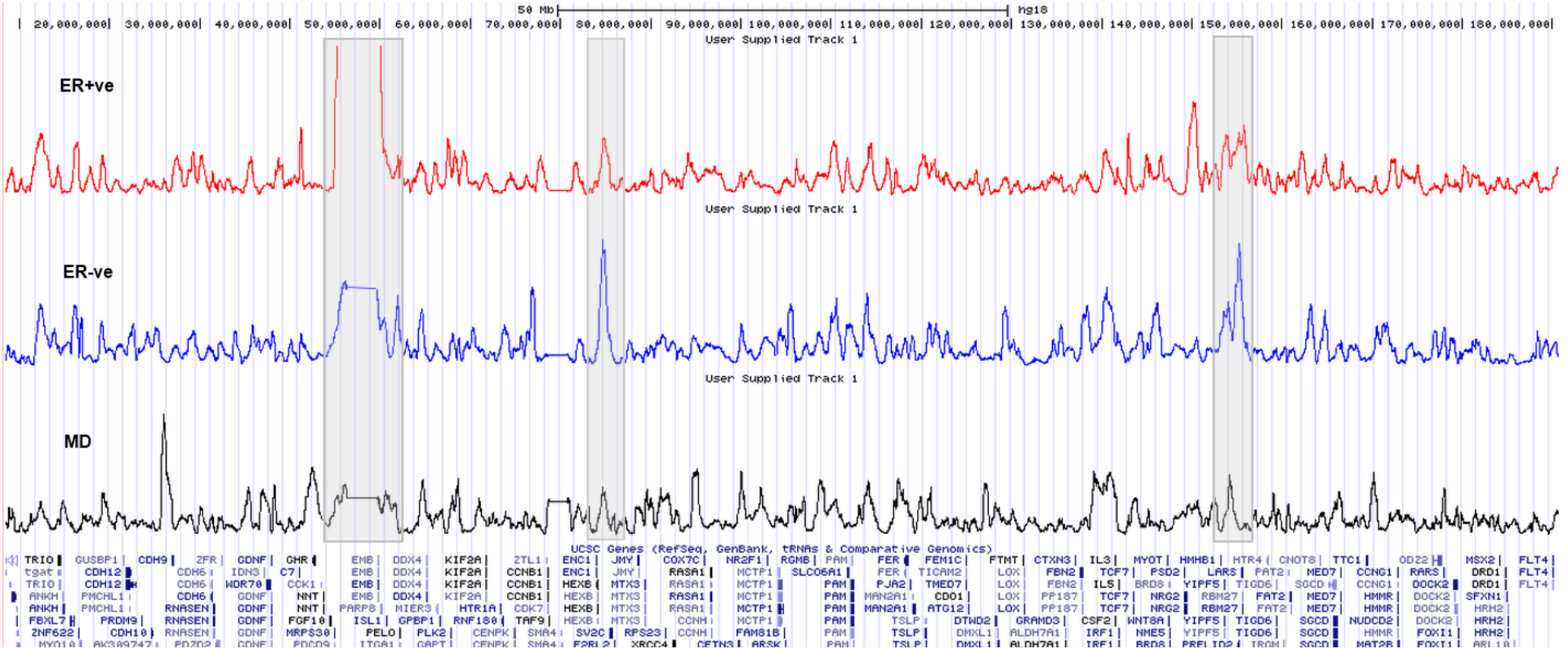


Future work (1)

- DEPTH sub-analyses
  - Proximal colon versus distal colon and rectum
  - Mismatch repair deficient versus proficient tumours
  - Case-case analysis (stratified by age)

- DEPTH analysis of combined CRC GWAS and EWAS data

Future work – Replication (2)




Thank You