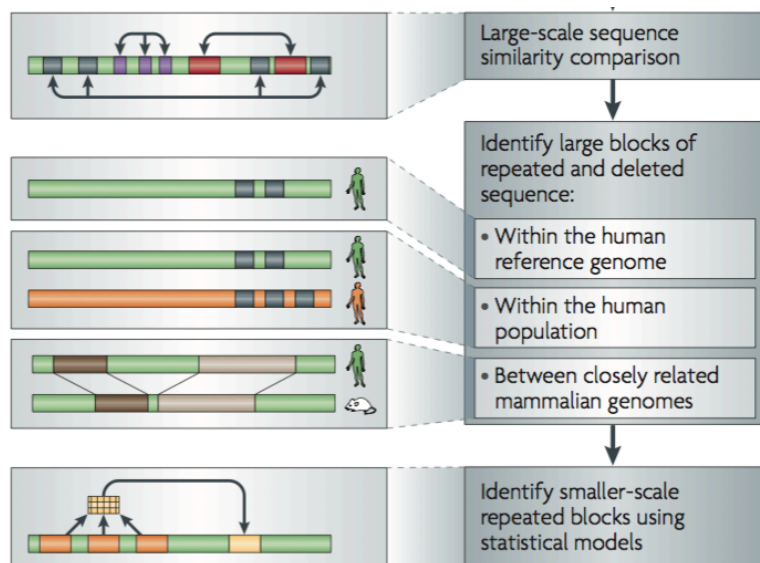# Non-coding Annotations: Overview
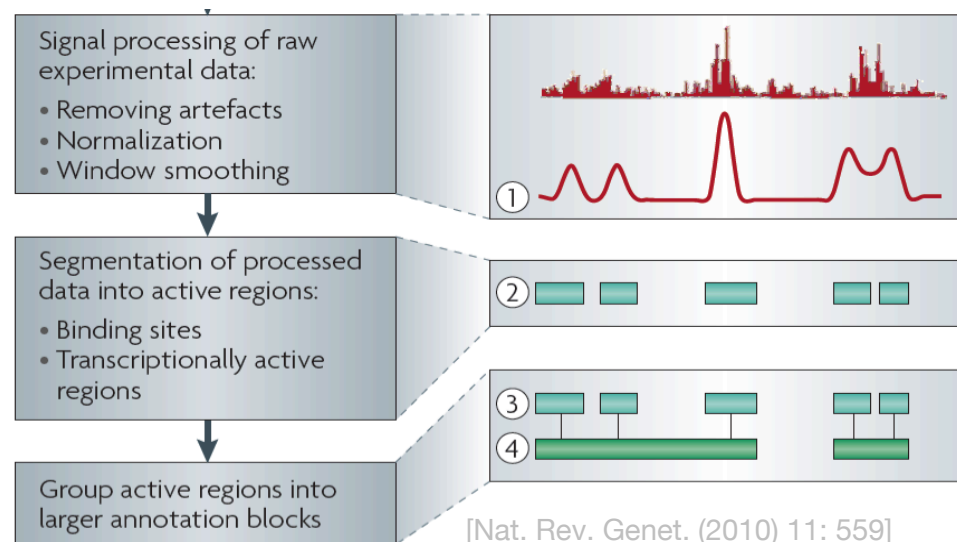
There are several collections of information "tracks" related to non-coding features

Sequence features, incl. **Conservation**

Functional Genomics
Chip-seq (Epigenome & seq. specific TF)
and ncRNA & un-annotated transcription

Large-scale sequence similarity comparison

Identify large blocks of repeated and deleted sequence:
- Within the human reference genome
- Within the human population
- Between closely related mammalian genomes

Identify smaller-scale repeated blocks using statistical models

Signal processing of raw experimental data:
- Removing artefacts
- Normalization
- Window smoothing

Segmentation of processed data into active regions:
- Binding sites
- Transcriptionally active regions

Group active regions into larger annotation blocks

[Nat. Rev. Genet. (2010) 11: 559]

[Mark]

Functional Genomics Annotations

A) PEAKS
1. DNase peaks at the UCSC genome browser {on many cell lines}
2. The regulation track at the UCSC genome browser, with compilation of TF ChIP-seq peaks from uniform processing (individual peaks are annotated with TF and cell line)
3. Blacklist Regions

B) RNA BASICS [~Regulatory]
4. A matrix of expression data of known genes (or exons) for protein-coding genes & known ncRNAs [Gencode version (V19)] {on many cell lines}
5. Novel RNA contigs track, i.e., possible novel transcripts. "Transcriptionally Active Regions (TARs)"
6. Novel junctions

C) PROMOTERS
Annotated GENCODE TSSes (TSSes with FANTOM CAGE support ?)

D) ENHANCERS (Supervised)
- Yip et al., Ren et al. &c

E) UNSUPERVISED SEGMENTATIONS, INCLUDING ENHANCERS
- ChromHMM, SegWay, HiHMM....

F) HOT/LOT REGIONS

G) CONNECTIVITY
7. Enhancer-target gene connection
8. TF-target network connectivity.
9. TADs: Topologically Associated Domains.

H) OTHER [~Regulatory]
10. List of Allelic SNPs & Regions

[Mark]