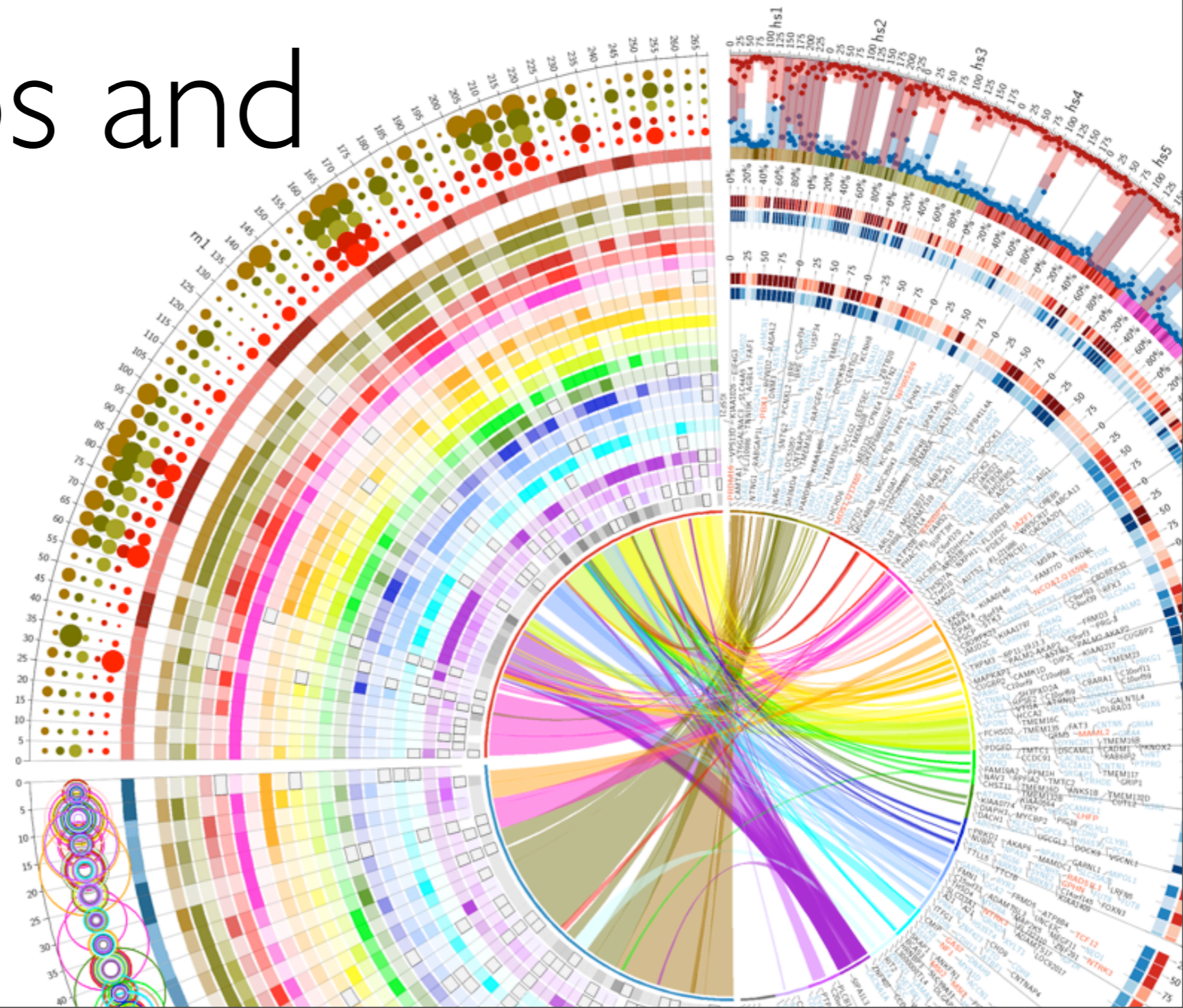


# Circular visualization of relationships and dense data



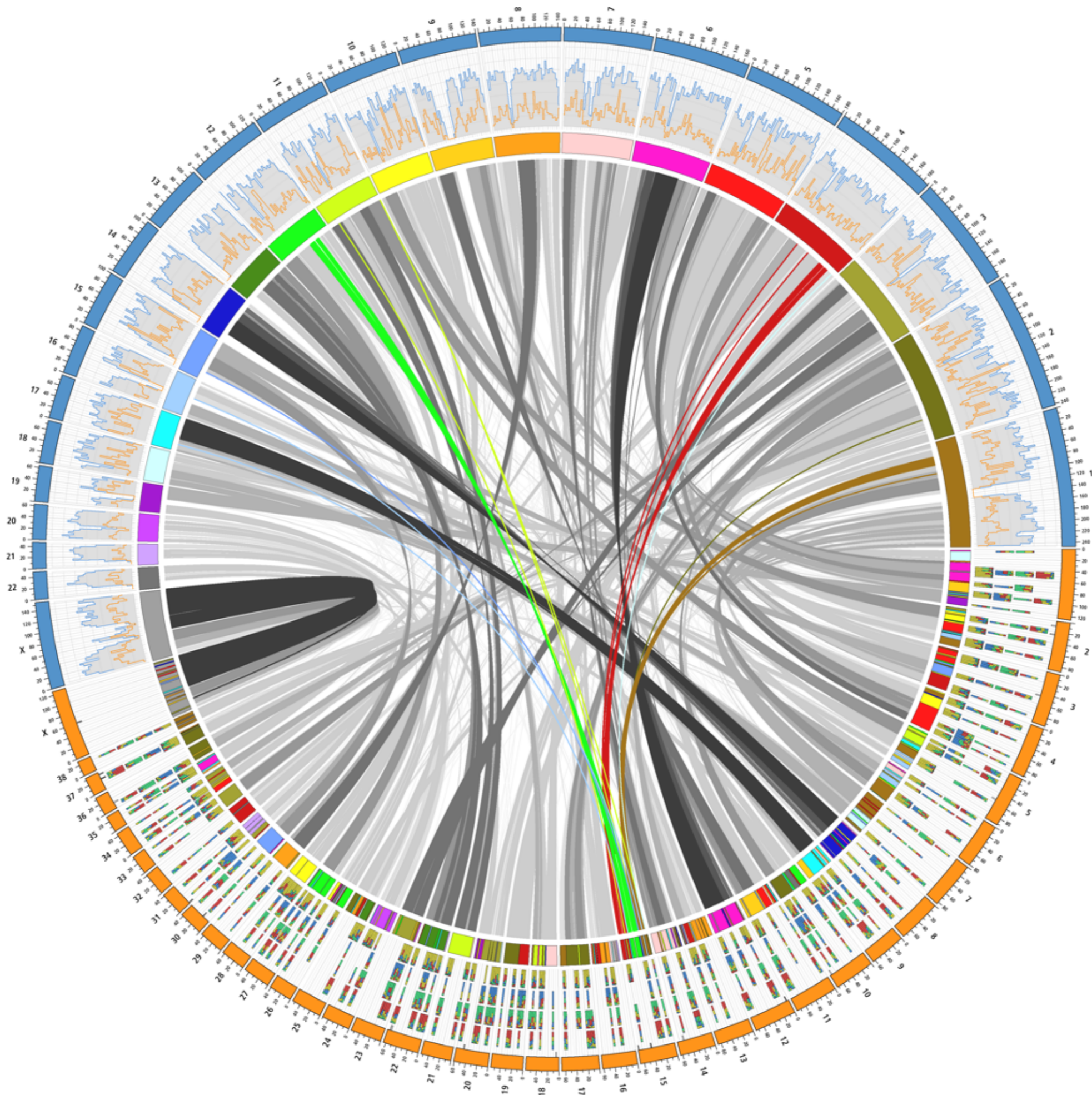
Andras Marczell, B.Sc.

# Využitie

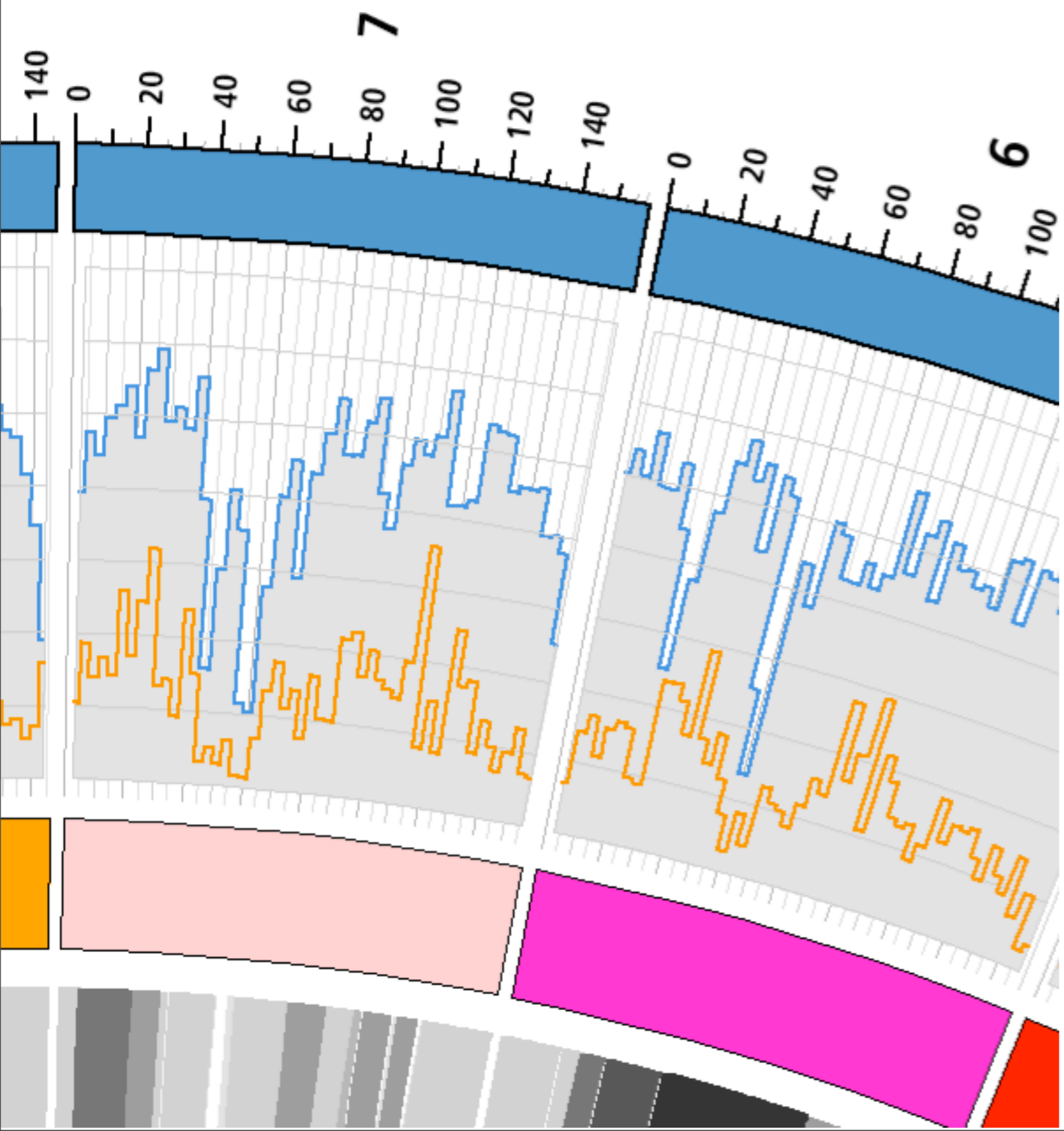
- Vizualizácia mnohodomenzionálných dát ( $>3$ )
- Prírodné kontinuálne zobrazovanie
- Zobrazenie relácií

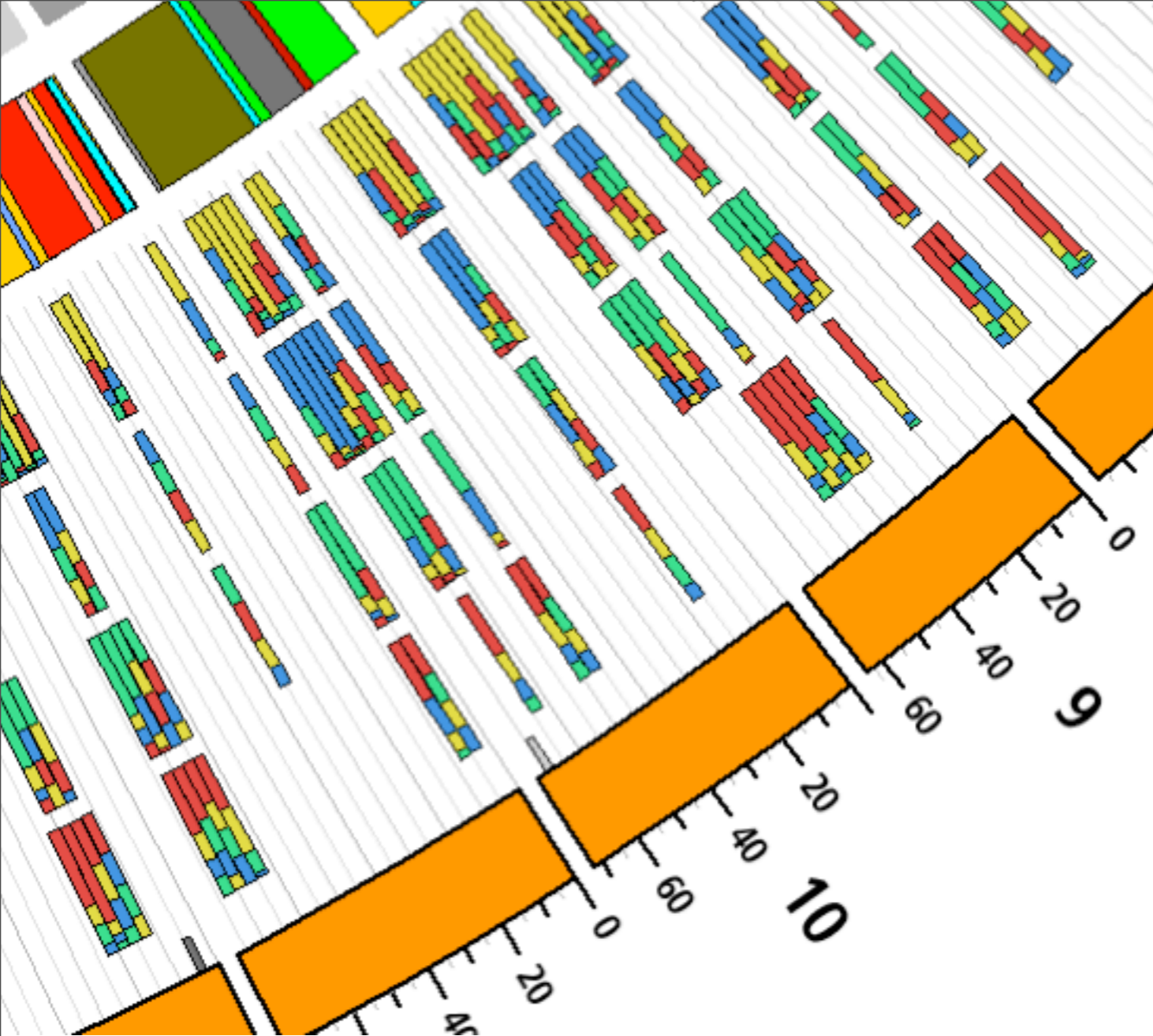






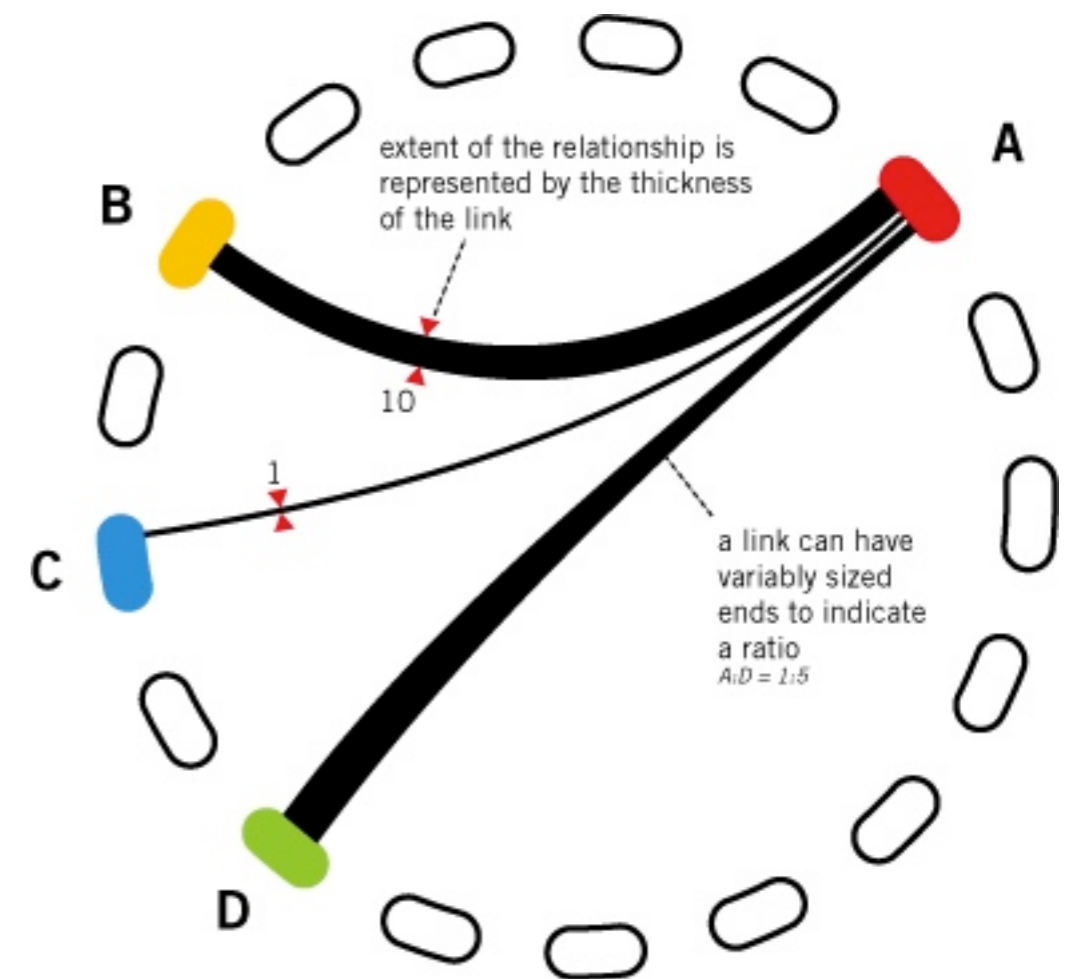
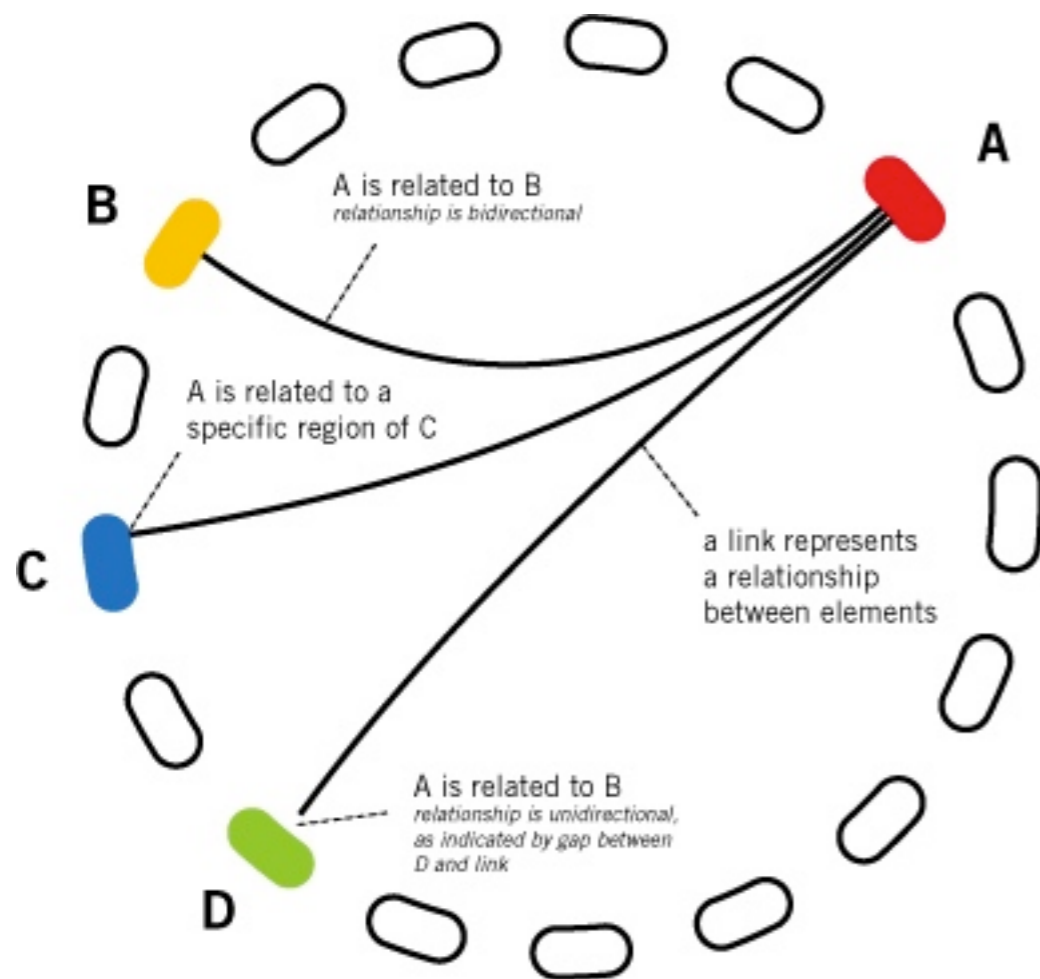
# Conservation info





# Biomarkery

# Vizualizácia relácií



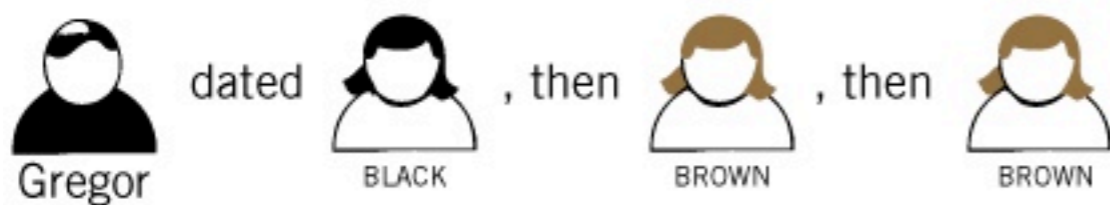
Relácie: uni- a bidirekcionálne

Pomer relácií indikované šírkou pásku

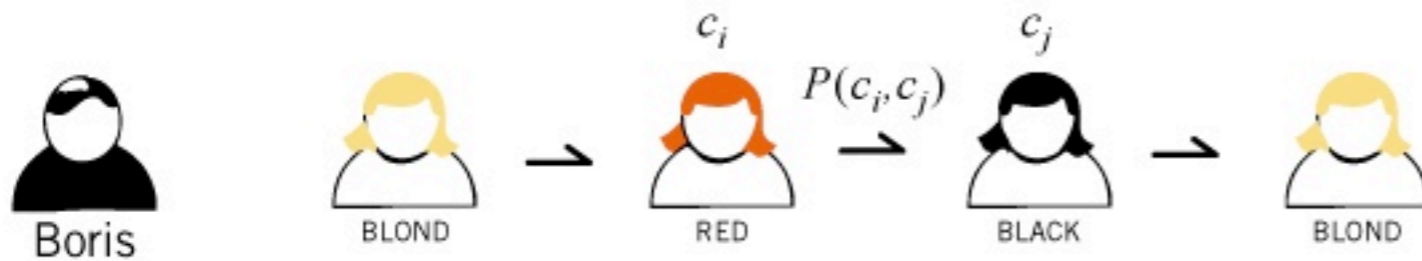
# Príklady

## HAIR COLOR PREFERENCE IN RELATIONSHIPS

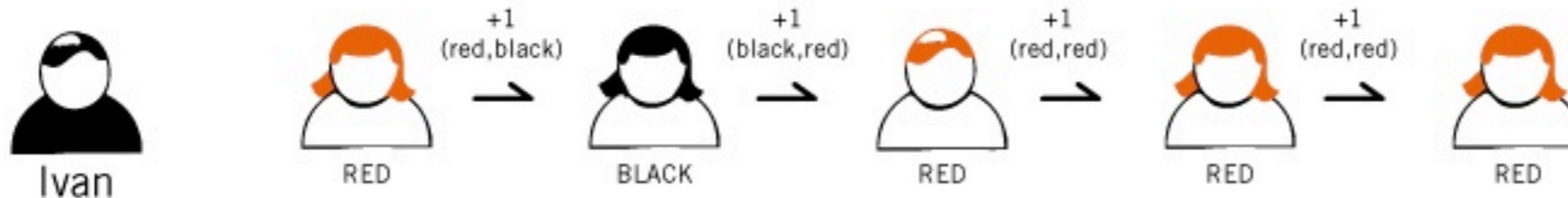
### Progression of Partners' Hair Color



Gregor's first partner had **black** hair.  
His next partner has **brown** hair.  
His third partner had **brown** hair.



By collecting information about history of many relationships, we can determine the average transition probability, which is the chance of dating someone with hair color  $c_j$  after dating someone with hair color  $c_i$ .



This transition probability can be stored in a table, with  $c_i$  and  $c_j$  being the rows and columns and the cell at  $(c_i, c_j)$  being the total number of observed transitions from  $c_i$  to  $c_j$ .

Ivan's contribution to the transition table.

COLOR	TO			
	BLACK	BROWN	BLOND	RED
FROM BLACK				+1
FROM BROWN				
FROM BLONDE				
FROM RED	+1			+2

Transition table for relationships of Gregor, Boris and Ivan.

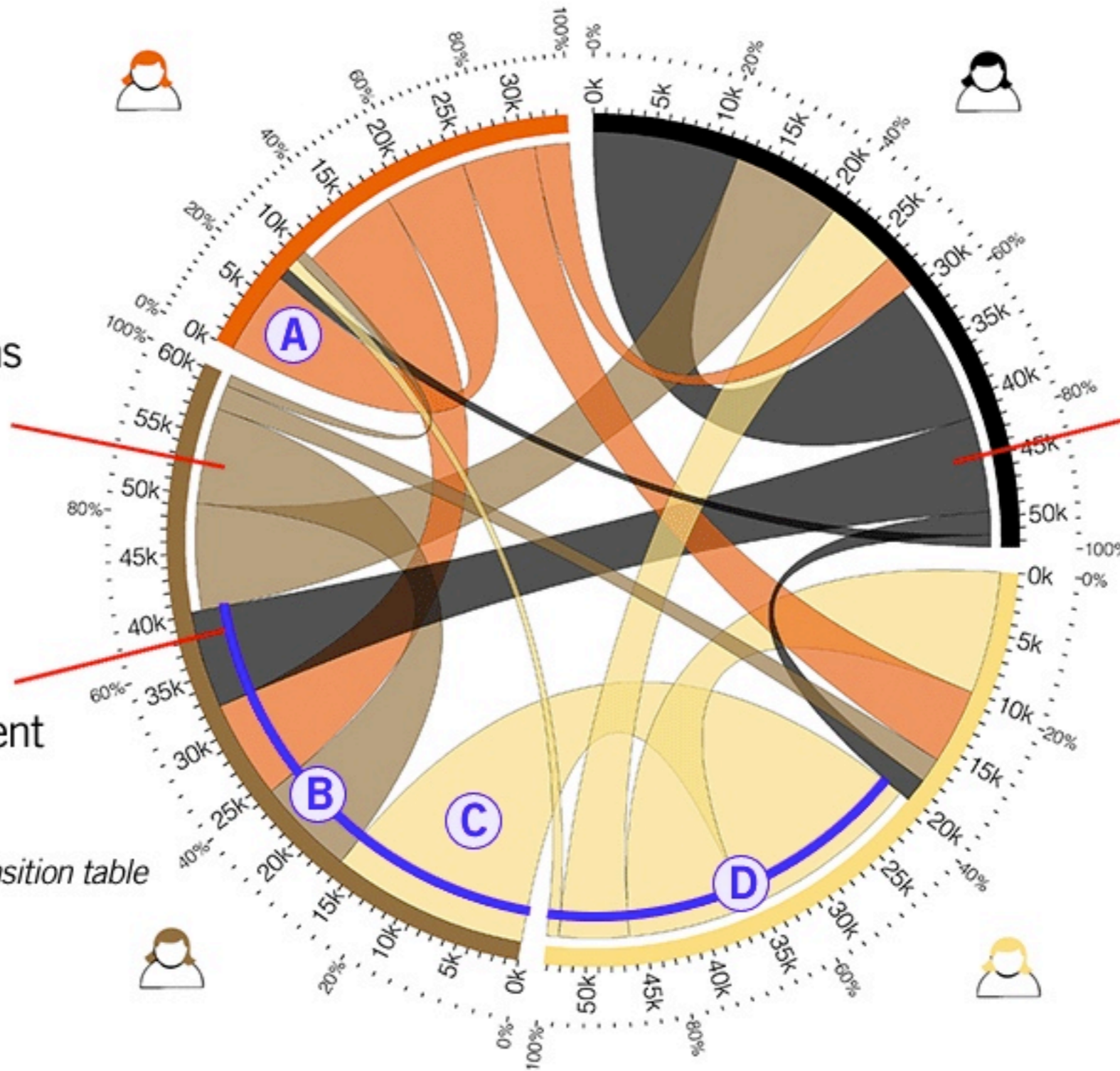
COLOR	TO			
	BLACK	BROWN	BLOND	RED
FROM BLACK		1	1	2
FROM BROWN		1		
FROM BLONDE				1
FROM RED	2			2



each ribbon  
represents transitions  
between colors

ribbon starts  
at the color segment  
of the of the  
disposed partner  
*color is a row in the transition table*

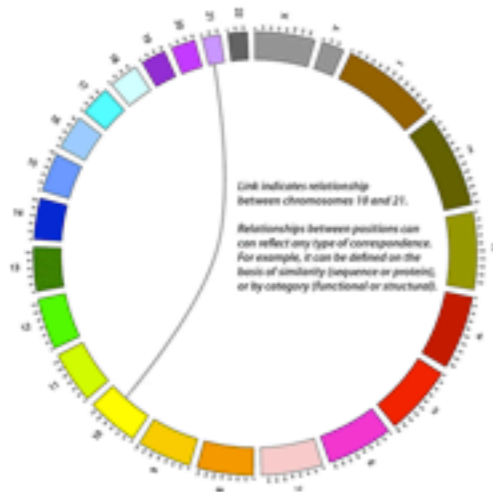
ribbon ends  
slightly before  
the color segment  
of the new partner  
*color is a column  
in the transition table*



# Cross-Species gene conservation

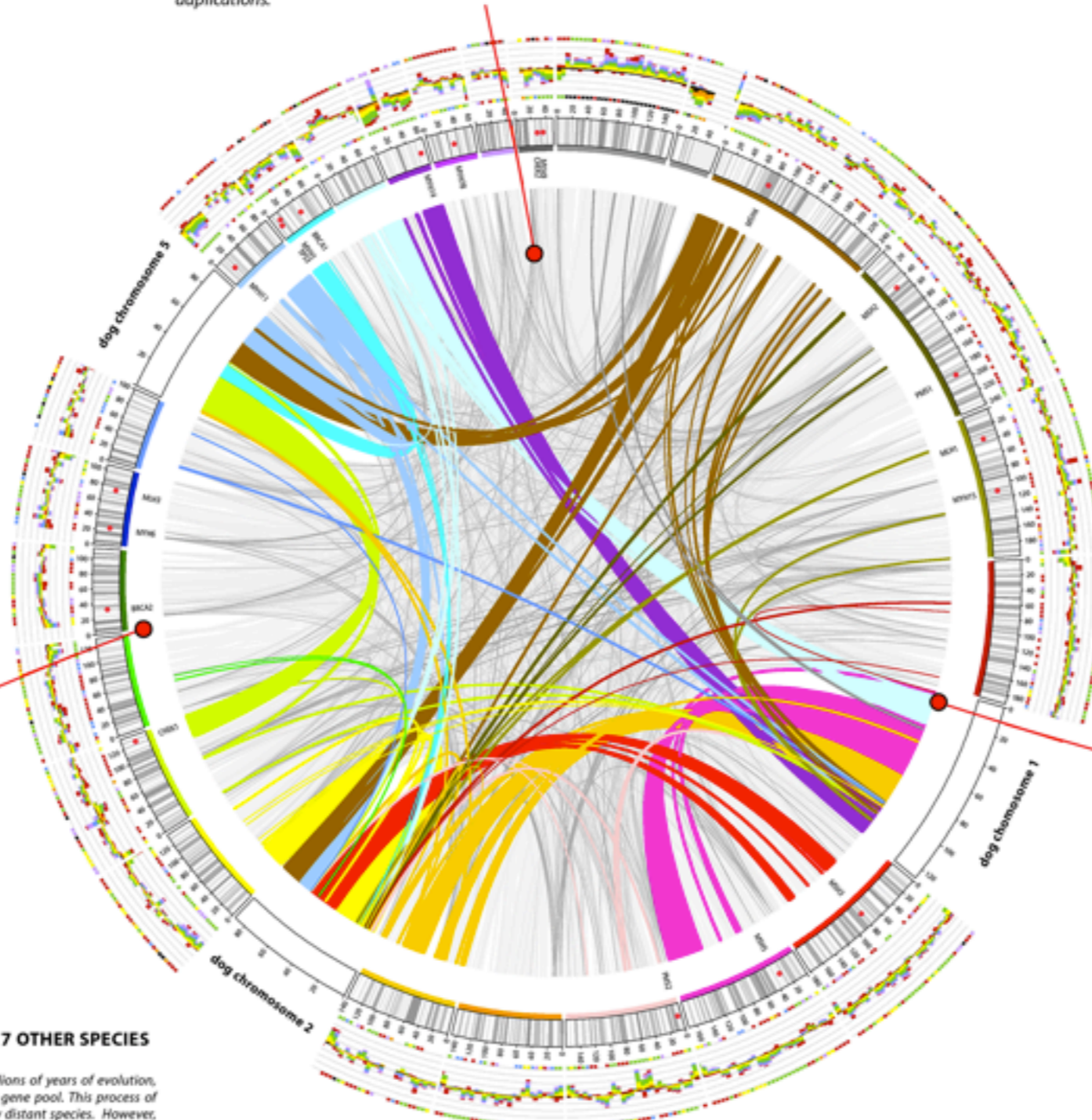
## CIRCULAR COMPOSITION

There are 23 pairs of chromosomes in the human genome. For visual clarity, these chromosomes are circularly arranged. Relationships between genomic regions can be clearly demonstrated by links within the circle.



The human genome contains many regions similar to one another. Some regions of the genome appear in varying copy number and similarity, in tandem or scattered throughout the genome. In total, about 50% of the human genome is considered to be repeated in this fashion.

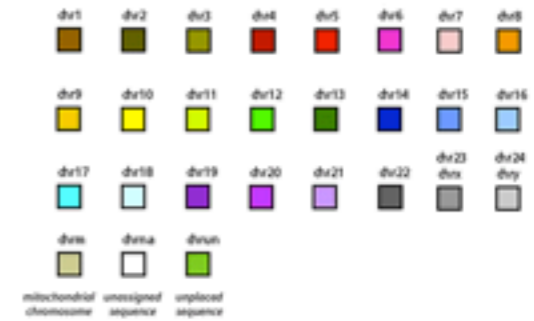
One class of repeats are **segmental duplications**, which are regions of the genome at least 1,000 bp in size which have a corresponding copy elsewhere with more than 90% sequence identity. The grey links connect all segmental duplications larger than 2,000 bp within the genome. Darker links indicate larger duplications.



Position of genes implicated in hereditary cancer. For example, certain mutations in BRCA1 and BRCA2 genes are associated with increased risk of breast cancer.

## HUMAN CHROMOSOME COLOR CODE

Each human chromosome is coded according to this conventional color scheme. For illustrations without explicit linkages between related regions, a standard color encoding provides easily interpretable visualizations. Typically, such images use color to associate a genomic location with a corresponding target chromosome.



Each position on dog a chromosome links to the most similar region on the human genome. The color of link is coded by the target human chromosome.

Synteny between dog and human has a block structure. Large contiguous regions of dog chromosome 1 have high degree of similarity with regions of human chromosomes 6, 9, 18 and 19. This block structure is seen in most syntenic relationships and reflects the underlying mechanism of recombination of genetic material.

## SEQUENCE CONSERVATION BETWEEN HUMAN AND 7 OTHER SPECIES

As organisms were subject to selective environmental pressures during millions of years of evolution, their genomes slowly changed as beneficial mutations accumulated in the gene pool. This process of genomic drift is responsible for the variation in both related and genetically distant species. However, because many fundamental biochemical processes (metabolism, signal transduction, DNA replication and repair) are evolutionarily stable, genomic regions that code the essential building blocks of life are not subject to this drift. Consequently, even species as different as the chimp and zebrafish contain some level of genomic similarity - regions of their respective genomes with a high degree of sequence identity.

Ongoing sequence projects of a variety of mammals and other organisms help establish the degree of conservation between various species and directly measure the extent of evolutionary drift between them. In turn, this helps define the phylogenetic distance between organisms. This kind of study is the focus of comparative genomics.

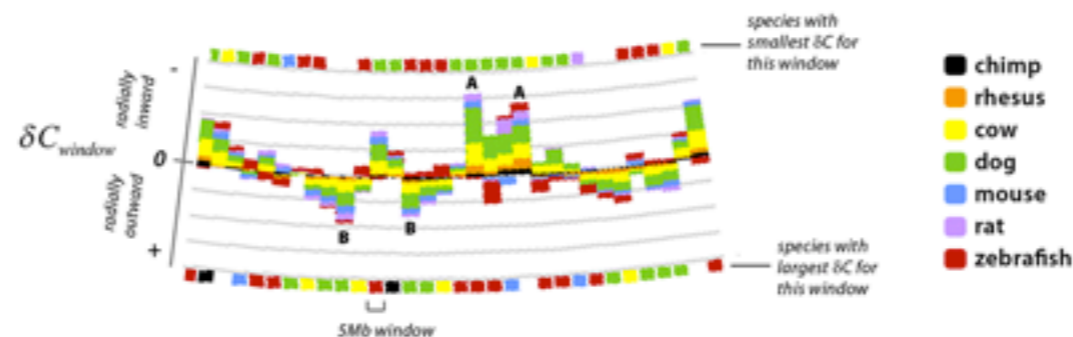
Comparative genomics is not only useful in understanding the fundamental process and rates of evolution, however. The results from comparative studies permit the basic research of laboratory animals inform the direction of applied research in health and medicine. For example, by examining the interplay of genes in the onset and progression of cancer in a laboratory mouse, correspondence with cancer in humans can be found.

## ILLUSTRATING CONSERVATION

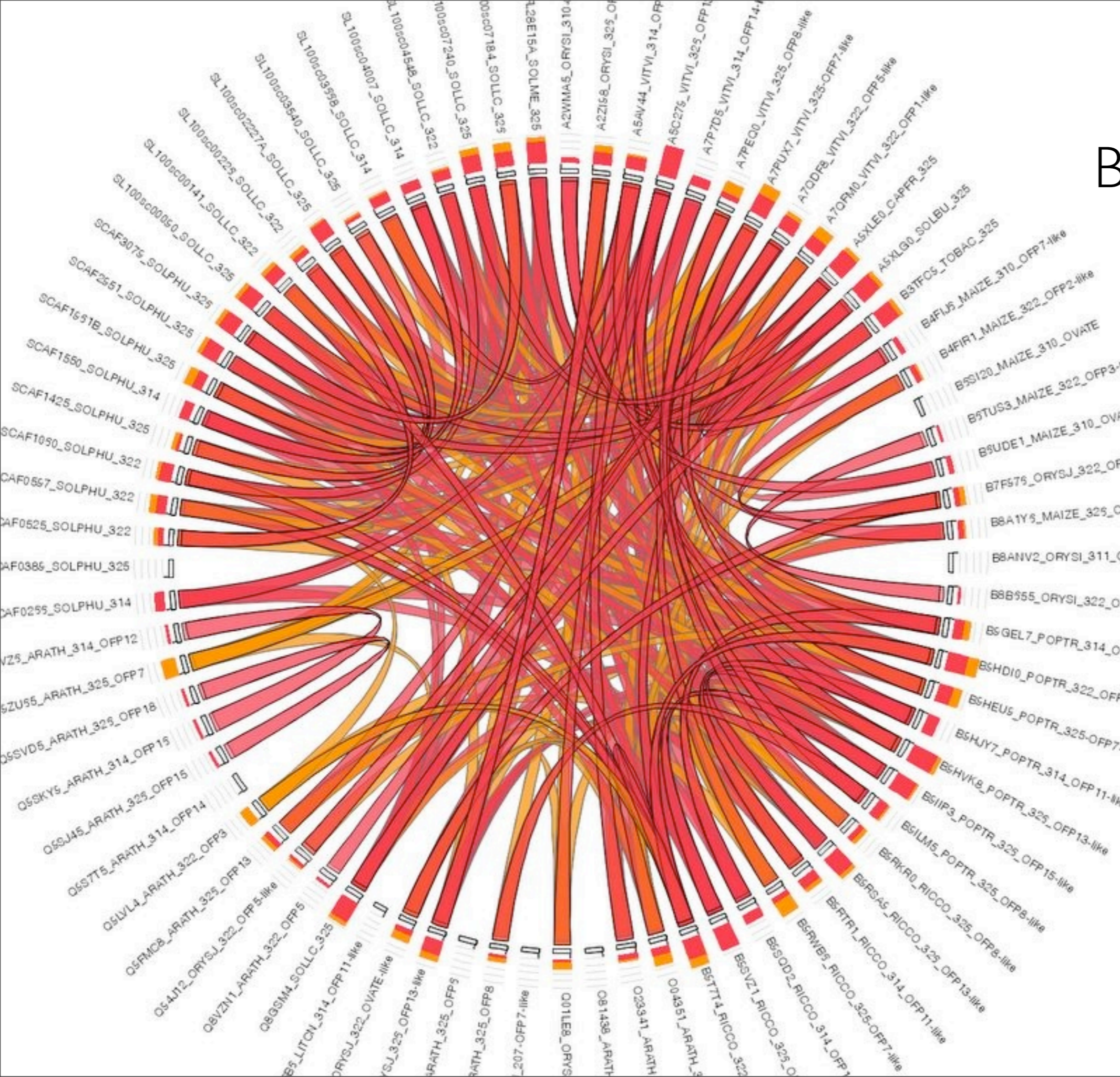
For each 5Mb window in the human genome, conservation is shown for 7 different species: chimp (*Pan troglodytes*), rhesus (*Rhesus macaque*), cow (*Bos taurus*), dog (*Canis familiaris*), mouse (*Mus musculus*), rat (*Rattus norvegicus*) and zebrafish (*Danio rerio*). For a given species, conservation is shown as the relative difference between average conservation within the window and the genome.

$$\delta C_{window} = \frac{avg_{window} C - avg_{genome} C}{avg_{genome} C}$$

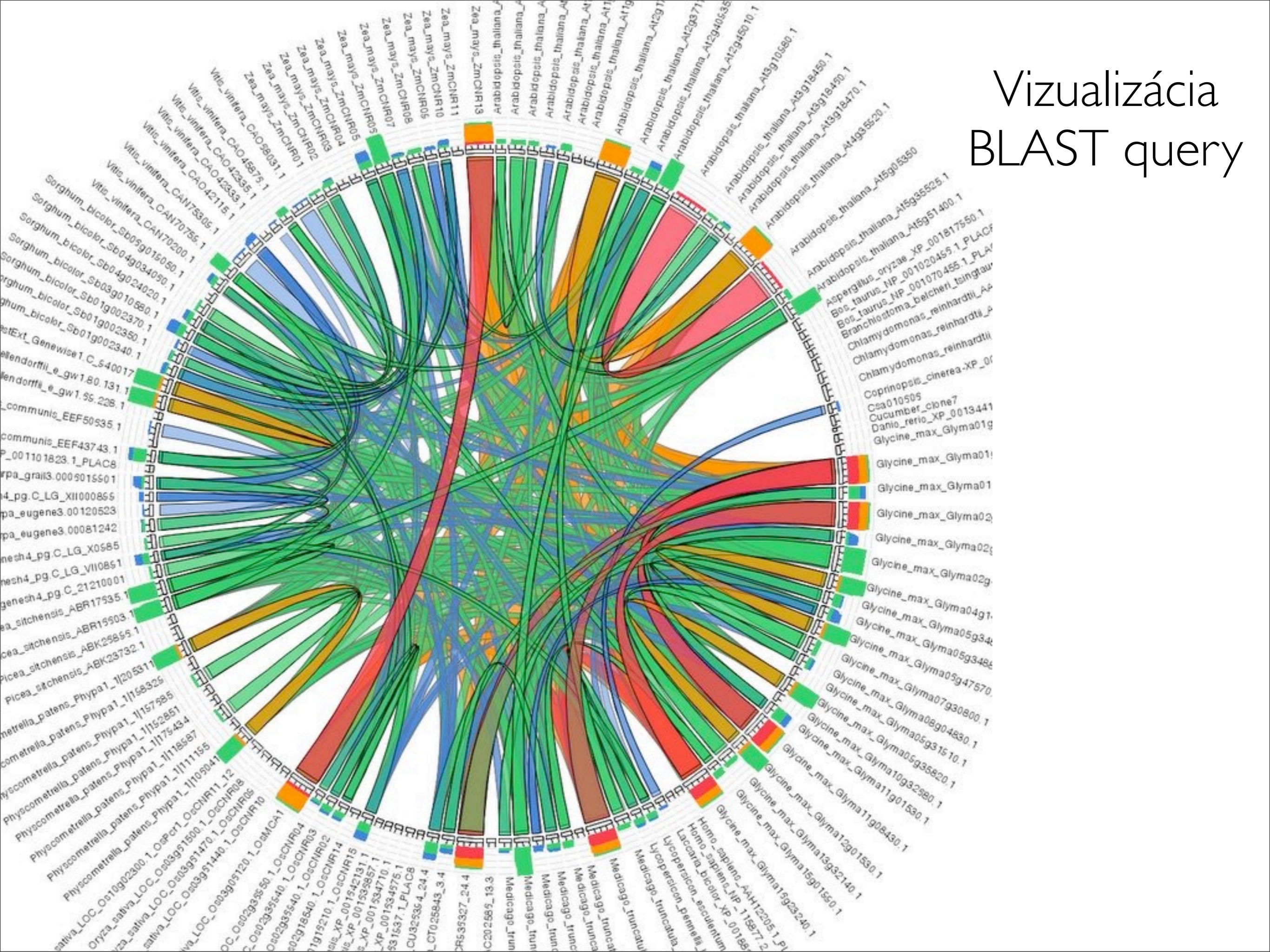
Some regions of the human genome have drifted further (A) and have negative relative conservation differences for each species. Other regions (B) show large positive conservation differences, indicating that this region is preferentially conserved relative to the whole genome.



# Vizualizácia BLAST query



# Vizualizácia BLAST query



# References

- Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
- Roux-Rouquié, M. & Soto, M. *Lecture Notes in Computer Science.* **3380**, 28–43 (Springer Berlin Heidelberg: Berlin, Heidelberg, 2005).
- Symons, S. MGv: A Generic Graph Viewer for Comparative Omics Data. *Bioinformatics* (2011).