Comparative Genomics

Even though medical genetics is focused on detection of relationship between the structural and functional aspects of human genome and the pathophysiology of human disease, it is natural that due to ethical or logistical reasons it is not always possible to test the experimental hypothesis directly in human subjects. Therefore majority of the current fundamental genetic laws was discovered in model systems. There is a wide variety of utilized model systems ranging from one-cell organisms to higher mammals. *Caenorhabditis elegans* or *Drosophila melanogaster* as well as the representative of plants, *Arabidopsis thaliana*, represent the classical multicellular models. Here, we will focus on two mammal models, mouse (*Mus musculus*) and rat (*Rattus norvegicus*).

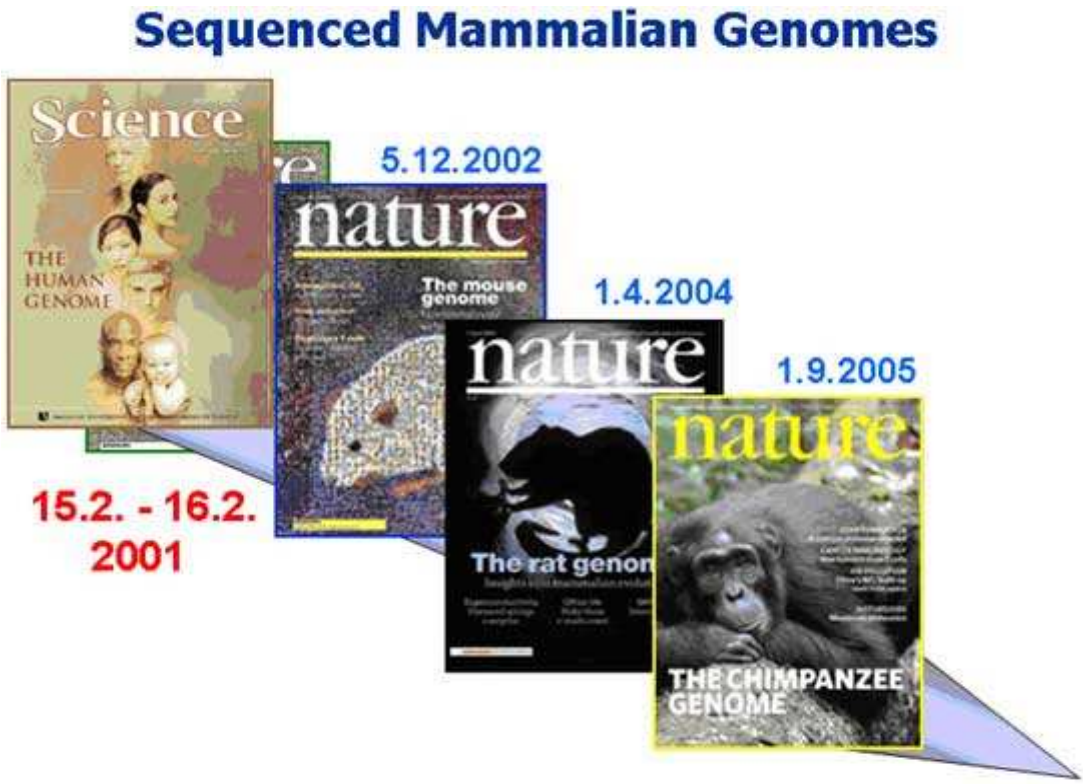

Figure 1. Up-to-now sequenced mammalian genomes

Comparative genomics uses the availability of human and model organism genomes to find the relevant information via their analytical assessment on global level. Using the tools of comparative genomics, it is feasible to predict not yet annotated human genes, though, their existence has to be afterwards verified by laboratory methods. For example, the gene for apolipoprotein A5 (APOA5) was discovered recently thanks to comparative sequencing of region of human chromosome 11 syntenic to mouse and rat regions, where this gene was primarily identified. The whole genome analysis shows that the large conserved sequences in the human genome can be also found within the genomes of one or more groups of model organisms. Comparative genomics does not focus exclusively on new gene discovery, but other functional features of the genome can be addressed, such as DNA motifs (binding sites) of transcription factors. Two most commonly used bioinformatic tools for such studies are VISTA (http://www-gsd.lbl.gov/vista) and PIPMAKER (http://bio.cse.psu.edu/pipmaker). Using the internet interfaces of such applications, large chunks of DNA sequences (acquired either from sequencing or biological databases) can be analyzed for the presence of conserved

regions in other species. VISTA even contains the pre-made comparisons of complete mouse and human genomes. These applications can be used in both directions for identification of genes, regulatory sequences and as a basis for understanding the mechanisms and history of the genome evolution.

Even databases of model organism genomes have the comparative genomic modules incorporated - as an example see the virtual comparative map VCmap (http://rgd.mcw.edu/VCMAP) on Rat Genome Database server. It is possible to generate a comparative map of a given part of the genome (human, rat and mouse) using algorithms for prediction of syntenic regions with the combination of knowledge of available genome maps and sequences.

It was comparative genomics that showed that rat and mouse are, at least at the level of DNA sequences, closer to the humans and primates than other mammalian models (pig, dog). Rat and mouse models offer wide variety of advantages for the experimental investigation of genetically determined traits - selective mating, production of sufficient number of offspring for segregation analyses (to produce one generation takes a few months), the possibility of standardization and specific manipulation of environmental conditions (e.g. diet administration with specific content, environmental temperature, day/light regime…). Often inbred animal strains are used, these animals are actually identical within one strain and gender (which equals to genetic (not epigenetic!) identity of monozygotic twins), and this allows us to avoid the problematic factor of genetic heterogeneity of human cohorts. Nowadays the main advantage of animal model use is the targeted genome modification in the conditions of living organism and the follow-up of changes caused by such modification. One of the disadvantages is the fact that the findings from model organism cannot be directly used within the human context, but meticulous validation is needed. Apart from inherent differences in human and rat (mouse) physiology, it is necessary to keep in mind that every inbred model of investigated trait represents only one specific fixed combination of alleles that lead to the trait (disease) manifestation. The aim to cover high number of different allelic combinations is pursued by large international consortia, currently starting or running projects of broadly based schemes of breeding of many available model strains. The projected result of these efforts are panels of hundreds of recombinant inbred strains that will cover a great portion of genetic variability available in parental strains.

The most commonly used types of experimental strains of rat and mouse and actual examples of their utilization in medical genetics are covered in the e-learning course "Introduction to experimental genetics", here we provide only a short summary:

Congenic strains are special type of strains where a specific and defined part of the genome from one inbred strain (strain A) was introgressed on the genetic background of second inbred strain (strain B). The congenic strain is an inbred strain, so it is homozygous throughout the whole genome, the only difference between strain B and congenic strain B.AX (where X is a specific number of the chromosome) is just the "differential" segment of chromosome X. If strain B and congenic strain B.AX differ in any phenotype (body weight, glycemia, or sensitivity to teratogen), it can be deducted that within the differential segment there is a gene or there are genes responsible for genetic determination of followed trait.

Consomic strains are special type of congenic strains. The differential segment should be the whole chromosome.

Recombinant inbred strains are a robust model for segregation and linkage analysis. By breeding the parental generation of two inbred strains the first filial F1 generation was obtained. F1 animals are identical, heterozygous throughout the whole genome. By breeding the F1 animals together, the second F2 filial generation is obtained, but within the genomes of F2 animals the alleles of both parental strains randomly segregate together with predispositions for phenotypic traits. After that we randomly select a couple (male and female) from F2 population and these two became the originating couple for the future recombinant inbred strain. The inbreeding is achieved by repetitive breeding of brother x sister for more than 20 generations. The strains are inbred and each strain carries in its genome a specific combination of alleles of parental strains. One of the most widely used set of recombinant inbred strains of rat, set HXB and BXH was developed at the Institute of Biology and Medical Genetics of the First Faculty of Medicine in collaboration with the Institute of Physiology of the Academy of Sciences of Czech Republic in 80s of the 20th century.

By insertion of a specific gene within a vector into the genome of recipient, **transgenic strains** are generated. If the gene is inserted together with tissue-specific promoter, then it is expressed only in the given tissue). The inserted gene does not need to come from the same species, namely mouse transgenic models expressing human genes are common (dubbed "humanized mice").

The "knock-out" and "knock-down" models are now available only in mouse. In these models the specific gene is either turned off fully ("knock-out") or its function and expression is just limited using RNA interference ("knock-down", see glossary).