 

# Practical session 1: Databases

In this practical session you will use several online-available data portals to find information on genes, transcripts, proteins, their function, biological pathways and nanomaterials.

Read the practical assignments carefully – you can write your answers below the questions. If you still have questions you can ask the instructors.

*Assignment 1: Introduction to Ensembl*

First you are going to learn how to use the Ensembl database. Together with the instructor you will walk through the Ensembl tutorial, see Tutorial\_Ensembl-walk\_through.pdf

Ensembl provides genes and other annotations such as regulatory regions, conserved base pairs across species, and sequence variations. The Ensembl gene set is based on mRNA evidence in the NCBI RefSeq databases. Website: [***www.ensembl.org***](http://www.ensembl.org)

**!** Please note that if for some reason the ensembl website is slow you can choose a mirror in a different continent (see image below)!



*Assignment 2: Ensembl – do it yourself!*

You will learn how to obtain information on a specific gene in Ensembl ([www.ensembl.org](http://www.ensembl.org)). For this a member of the human superoxide dismutase family, the superoxide dismutase 1 (SOD1) was selected.

Superoxide dismutase (SOD) is an enzyme which plays an important role in handling oxidative stress in cells. It catalyzes the reaction of superoxide (O2-) to oxygen (O2) or hydrogen peroxide (H2O2). Superoxides are usually produced as a by-product of the oxygen metabolism but they are also produced in case of cellular stress – caused e.g. by nanoparticles. Superoxide can cause severe damage to cellular molecules leading to cell death. There are 3 main SOD subclasses known: SOD1, an enzyme which occurs in the cytoplasm, SOD2 which occurs in the mitochondria, and SOD3 which is released to the extracellular space.

🖰 Find the human SOD1 gene in Ensembl and go to the Gene tab.

1. As explained in today’s lecture, in Ensembl all the genes have a unique identifier. **What is the Ensemble ID of the SOD1 gene?**
2. **On which chromosome and which strand the gene is located?**
3. Genes that are located on the same chromosome are linked genes. This means that they have the tendency to be inherited together. Therefore, it is of interest to know which protein coding genes are in close proximity. **What is the next known protein coding gene on the sequence to the right of the SOD1 gene? Which strand?** (Hint: go to the Location tab)
4. Genes can have more than one transcript due to alternative splicing. These transcripts can give rise to proteins which differ in their sequences and often in their activities or can cause changes in regulatory elements such as translation enhancers or RNA stability domains, which may have a dramatic effect on the level of protein expression. **How many transcripts have been predicted for this gene?**
5. **How many of these transcripts are protein coding?**
6. **How many base pairs has the longest protein coding transcript and what is the length of the protein (the number of amino acids) it encodes?**

*Assignment 3: NCBI*

In assignment 2 you used the Ensembl website to gather information on the human SOD1 gene. In this exercise you will use several NBCI-Entrez databases to find more information about this gene. NCBI offers a training and tutorial page at: <http://www.ncbi.nlm.nih.gov/guide/training-tutorials/>

Start with opening the **NCBI Gene page** for human SOD1 and answer the following questions. (NCBI homepage <http://www.ncbi.nlm.nih.gov/>, NCBI Gene page <http://www.ncbi.nlm.nih.gov/gene>). Hint: apart from information directly on this page, you may also want to use links to other web pages (for example at the right side of the page).

1. NCBI gene also uses specific identifiers. **What is the NCBI gene ID of the human SOD1 gene?**

**!** Hint: Background information on the *Gene database* is available at <http://www.ncbi.nlm.nih.gov/books/NBK21085/> !

1. Human genes are given an official name by the HUGO Gene Nomenclature Committee (HGNC). However, one gene can have other names which are not official. **Give two alternative names (not IDs) of the human SOD1 gene.**
2. In the NCBI gene entry the NCBI reference sequences are given. **What is the RefSeq ID of the mRNA of this gene and of the protein?**
3. **What are the NCBI Gene IDs of the homologues of human SOD1 in mouse and rat?** **How/where did you find this?**

**!** Hint: You find this information looking for HomoloGene (under the box on the right side Related Information) or select Orthologues: all in the summary.

*Assignment 4: Find genes by function using Gene Ontology*

Apoptosis is controlled cell death – in contrast to necrosis, which is uncontrolled. Apoptosis can be triggered by various events, like oxidative stress induced by nanoparticles.

In this assignment you will use the **gene ontology** using AmiGO2 to find apoptosis related genes: <http://amigo.geneontology.org/amigo>

Go to the homepage of AmiGO2 and search for apoptosis. Choose “Genes and gene products”, and select organism: *Homo sapiens* (+).

1. How many genes and gene products are annotated with apoptosis?
2. Name 3 genes which are directly annotated (“Direct annotation”) with apoptotic processes.
3. Look up the gene CASP2 in Ensembl and NCBI. What are the identifiers?
4. Give 3 biological processes this gene is involved in.

*Assignment 5: Biological pathways WikiPathways*

Go to WikPathways [www.wikipathways.org](http://www.wikipathways.org) and look up oxidative stress.

1. Look for SOD2 in the pathway. What is the Ensembl identifier for this gene?

! Hint: klick on the gene symbol and look at the pop-up.

1. How does reactive oxygen species influence expression of SOD2?
2. Which ontology annotation is given for this pathway?
3. Which paper is given as a reference for this pathway?

*Assignment 6: GEO*

Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/gds> is a curated repository for gene expression data.

1. How many studies are available for “silver nanoparticles”? How many of them are done with human cells?

Have a look at the study with the number **GSE62253** and have a look at the experimental setup:

1. Which cells are used for this experiment?
2. What type of microarray was used?
3. What are the experimental groups? (Hint: there are 4 groups with each 3 samples)
	1. Which one is the control?
	2. How much silver nanoparticles are used?
	3. Why do you think did they use AgNO3?

*Assingnment 7: Nanoparticle database eNanoMapper*

eNanoMapper is a database for nanomaterial specific research. Information on specific studies can be found by using either nanoparticle name, identifier, composition, or free text search.

1. Use “Search” “nanomaterials by name” and enter Ag (chem. Symbol for silver). How many studies are available?
2. Klick on the first Substance name “Ag NM-300k\_65nm\_0.1\_Bath”. Which experimental data is available? (Hint: klick on P-Chem for physico-chemical, and Tox for biological/toxicological data)
3. Can you find the sizes of this nanoparticle used? (Primary size and in situ)