

Lecture 1 – Introduction & relational databases

Relational databases with MySQL

Database upgrade

- Download `bioinf_testdb.sql`
- Create a new database `bioinf_testdb`
`mysql> CREATE database bioinf_testdb;`
- Grant your user all rights on this database
- Create the tables and insert the data
`$ mysql bioinf_testdb < bioinf_testdb.sql`

Relational databases with MySQL

Exercises

- Find the gene named *MALAT1* in the gene table (bioinf_testdb)
 - In addition to the chromosomal position, return the size of the gene
- Find the known miRNA that is located the most distal on the p-arm of chromosome 2 (see figure below for hint)
- Find genes related to breast cancer (use description field)
- Return a list of genes located on chromosome Y in alphabetical order



1, 2, 3

A	T	G	G	C	T	A	T	A	C
1	2	3	4	5	6	7	8	9	10

Relational databases with MySQL

Exercises

- Return a list of the number of genes per biotype, most abundant biotypes first
- Return a list of the number of genes per status
- Combining the 2 previous results, which biotype is most known?
- Select only those biotypes that cover at least 1% of the human genome (hint: 'size' in previous exercise) and return this percentage.

Relational databases with MySQL

Excercises

- In the `bioinf_testdb` database
 - Examine the `gene`, `transcript` and `exon` table, how are they connected?
 - How many transcript does the PTEN gene have?
 - Return the position of the exons of transcript 221260
 - Return the transcripts of the *TP53* gene
 - Return their exons as well
 - Find the longest spliced transcript of *TP53* (taking into account the intron-exon structure)
 - How many exons does each transcript have?

Relational databases with MySQL

Exercises

- What is the name of the gene associated with transcript 260392?
- A mutation was found on chromosome 20, position 44002590. Which gene(s) overlap(s) with this position?
 - Select only those genes that have exons that overlap with this mutation. Which genes are they?
- A biotype column can be found in both the gene and transcript table. Are there transcripts that have a different biotype from the gene they're part of? What are their names?
 - Does the same go for status?
- Which chromosome has the most genes and how many are there?
- Which exon is the largest in the genome and how many base pairs are there?
- Which transcript has the most exons and how many are there?

Relational databases with MySQL

Rehearsal exercises

- Create a new database for your lab and include following data
 - All trainings
 - Subject, duration
 - All lab members
 - Name, lastname, birth_date, training
 - All equipment
 - Name, manufacturer, purchase_date
 - All experiments
 - Name, performed_by, equipment_used, date
 - All results
 - Directory, experiment, status
- Fill with some data

Relational databases with MySQL

Rehearsal exercises

- Pay attention to
 - Semi-colon after each statement
 - Order of table creation
 - Primary key in each table
 - Parameters of create statement between brackets
 - Correct usage of constraints (foreign keys, ...) and column types

THINK BEFORE YOU CREATE

Relational databases with MySQL

Rehearsal exercises (part 2)

- In your newly created database, search for
 - All experiment equipment purchased after 1st of January 1985
 - The number of experiments each lab member conducted
 - A list with all equipment used in a successful experiment
 - A list with all lab members that failed an experiment
 - Who followed wich trainings?
 - Number of participants per training