**UCSC Genome Browser exam exercise answers**

**USCS Genome Browser**: <http://genome.ucsc.edu/>

**USCS Gateway**: <http://genome.ucsc.edu/>  Genome Browser (or Upper bar  Genomes)

Ex. I – UCSC Genome Browser

1. Go to the genomic region chr6:45,296,054-45,518,819 of the Human assembly GRCh37/h19.

* Which genes do you see in this region? On which strand are they?

1. Enable **GC Percent** [dense] from **Mapping and Sequencing** tracks and **CpG Islands** [show] from **Regulation** tracks.

Now zoom into the region chr6:45,330,000-45,400,000.

* Does the GC composition reach a peak in correspondence of some important regulatory element?
* In which region do you notice the most regulatory activity? Does it involve a CpG island? (If yes, report *Coordinates*, *Chromosome band*, *Genomic size* of the first (along the genome) of them)

Ex. II – BLAT {*Upper bar*  *Tools*  *Blat*}

BLAT is a fast alignment tool, useful when one wants to align long DNA sequence onto genomes, unless the similarity percentage is high (95%). It is also useful when one has a huge number (up to millions) of long reads to align onto a reference genome.

Let’s do a toy example.

Consider the following FASTA file:

> read1

ACCACATATTTTGCAAATTTTGCATGCTGAAACTTCTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA

> read2

ACCACATATTTTGCAAATTTTGCATGCTGATACTTCTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA

> read3

ACCACATATTTTGCAAATTTTGCATGCTGATACTACTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA

> read4

ACCACATATTTTGCAAATTTTGCATGCTGATACTACTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTAAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA

> read5

ACCACATATTTTGCAAATTTGCATGCTGAAACTTCTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA

> read6

ATGAATGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACATAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAAGGTAGATCTGAA

Use BLAT to map these sequences onto Human assembly GRCh37/h19.

* Does each read map in a single region?
* On the base of the alignment score and sequence identity, which genome region do the reads belong?
* Which gene this reads come from?