



## DNA Microarray Data Analysis

IIRIS HOVATTA, KATJA KIMPPA, ANTTI LEHMUSSOLA, TOMI PASANEN,  
JANNA SAARELA, ILANA SAARIKKO, JUHA SAHARINEN, PEKKA TIIKKAINEN  
TEEMU TOIVANEN, MARTTI TOLVANEN, MAUNO VIHINEN AND GARRY WONG  
EDITORS JARNO TUIMALA AND M. MINNA LAINE

CSC

## List of Contributors

Iiris Hovatta  
National Public Health Institute  
Haartmaninkatu 8  
FI-00290 Helsinki  
Finland

Katja Kimppa  
PerkinElmer Life Sciences and Analytical Sciences  
- Wallac Oy  
P.O.Box 10  
FI-20101 Turku  
Finland

M. Minna Laine  
CSC, the Finnish IT center for science  
Keilaranta 14  
FI-02101 Espoo  
Finland

Antti Lehmussola  
Tampere University of Technology  
P.O.Box 553  
FI-33101 Tampere  
Finland

Tomi Pasanen  
University of Helsinki  
P.O.Box 68  
FI-00014 University of Helsinki  
Finland

Janna Saarela  
Biomedicum Biochip Center  
Haartmaninkatu 8  
FI-00290 Helsinki  
Finland

Ilana Saarikko  
University of Helsinki  
P.O.Box 68  
FI-00014 University of Helsinki  
Finland

Juha Saharinen  
National Public Health Institute  
Haartmaninkatu 8  
FI-00290 Helsinki  
Finland

Pekka Tiikkainen  
VTT  
P.O.Box 106  
FI-20521 Turku  
Finland

Teemu Toivanen  
Centre for Biotechnology  
Tykistökatu 6  
FI-20521 Turku  
Finland

Martti Tolvanen  
Institute of Medical Technology  
Biokatu 8  
FI-33520 Tampere  
Finland

Jarno Tuimala  
CSC, the Finnish IT center for science  
Keilaranta 14  
FI-02101 Espoo  
Finland

Mauno Vihinen  
Institute of Medical Technology  
Biokatu 8  
FI-33520 Tampere  
Finland

Garry Wong  
A. I. Virtanen -institute  
University of Kuopio  
FI-70211 Kuopio  
Finland

# 2 Web extra: Biological sequence annotations

Juha Saharinen

## 2.1 Using Ensembl server to batch process of annotations

Here are two small examples using these alternative interfaces to the Ensembl data. In example 1, 10 probe sets from Affymetrix U133A chip are queried from Ensembl *Homo sapiens* database and their Ensembl GeneStableIDs, and gene's sequence locations are returned. Example 2 shows a small Perl script, using EnsemblAPI, to retrieve the genomic sequence from a given organism, chromosome and physical position.

### 2.1.1 Example 1

```
bbu-juhaad@Optonix:~> mysql -h ensembl.db.ensembl.org -u anonymous homo_sapiens_core_27_35a
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1217078 to server version: 4.0.18-standard-log

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> select xref.dbprimary_acc, gene_stable_id.stable_id, seq_region.name, gene.seq_region_start, gene.seq_region_end,
gene.seq_region_strand
from xref
join external_db on external_db.external_db_id = xref.external_db_id
join object_xref on object_xref.xref_id = xref.xref_id
join translation on translation.translation_id = object_xref.ensembl_id
join transcript on transcript.transcript_id = translation.transcript_id
join transcript_stable_id on transcript_stable_id.transcript_id = transcript.transcript_id
join gene on gene.gene_id = transcript.gene_id
join gene_stable_id on gene_stable_id.gene_id = gene.gene_id
join seq_region on seq_region.seq_region_id = gene.seq_region_id
where (external_db.db_name = 'AFFY_HG_U133A')
limit 10;
```

dbprimary_acc	stable_id	name	seq_region_start	seq_region_end	seq_region_strand
209137_s_at	ENSG00000102034	X	128924393	128970188	-1
209136_s_at	ENSG00000102034	X	128924393	128970188	-1
202264_s_at	ENSG00000196586	6	76515703	76682665	1
202264_s_at	ENSG00000118260	2	208220192	208289067	1
AFFX-HSAC07/X00351_5_at	ENSG00000196586	6	76515703	76682665	1
AFFX-HSAC07/X00351_5_at	ENSG00000166261	11	123100699	123117573	-1
217991_x_at	ENSG00000166261	11	123100699	123117573	-1
209137_s_at	ENSG00000125447	17	70744290	70769299	-1

```
| 209136_s_at | ENSG00000125447 | 17 | 70744290 | 70769299 | -1 |
| 217988_at | ENSG00000163251 | 2 | 208456223 | 208459624 | -1 |
+-----+-----+-----+-----+-----+-----+
10 rows in set (0.11 sec)
```

```
mysql>
mysql> quit
Bye
```

### 2.1.2 Example 2

```
#!/usr/bin/perl

use Bio::Ensembl::DBSQL::DBAdaptor;

sub main
{
    $db = Bio::Ensembl::DBSQL::DBAdaptor->new
        (-host => 'ensemldb.ensembl.org',
         -dbname => $ARGV[0],
         -user => 'anonymous');

    # get the slice adaptor and fetch a slice on a given region
    $slice_adaptor = $db->get_SliceAdaptor();
    $slice = $slice_adaptor->fetch_by_region('chromosome', $ARGV[1], $ARGV[2], $ARGV[3]);

    # print out the sequence from this region
    print $slice->seq();
}

#start program
{
    if ($#ARGV != 3) {
        print "\nGenoFetcher - Juha Saharinen, Biomedicum Bioinformatics Unit & KTL, 2004\n\n";
        print "Use of GenoFetcher is \n";
        print "genofetcher <database> <chromosome> <chromosome_start> <chromosome_end>\n\n";
        print "Like genofetcher homo_sapiens_core_27_35a 12 1000000 1001000\n\n";
        print "This will return slice of human chromosome 12 from 1.000.000 to 1.001.000\n\n";
    }
    else {
        main;
    }
}
```