

NGS research and service at the CBU

Kjell Petersen, Inge Jonassen

Computational Biology Unit, Uni Computing, Uni Research AS, Bergen, Norway

<http://www.bioinfo.no/>

<http://www.uni.no/computing/units/cbu>

CBU consists of seven research groups and one service group, specialising in different aspects of computational biology. A common denominator in many of our projects are high throughput data sets, with Next Generation Sequencing as a prominent data providing technology. CBU also coordinates the Norwegian Bioinformatics platform that offers both helpdesk support and training to scientists in the field of functional genomics.

A natural research focus in Bergen is marine genomics. The recently published genome of cod [1] were accomplished with CBU as an active partner in the bioinformatics work, in particular the assembly of the 454 reads. Through this and other projects special competence on analysis of high-throughput sequencing (in particular, 454) data has been built, as documented in the work to realise the FlowSim tool [2].

Metagenomics on samples from extreme environments along the mid-Atlantic ridge is another field of high interest in Bergen, due to the Centre of Geo-biology (a National Centre of Excellence) situated next to CBU. Through this collaboration, new approaches to handle amplicon sequencing datasets from 454 have been developed, and implemented in the AmpliconNoise software tool [3].

Through our role in the National bioinformatics helpdesk and our close collaboration with the Norwegian Microarray Consortium, we have extensive experience in designing experiments and analysing gene expression data from high throughput datasets. Both analysis of data in research projects and training that we provide through the Bioinformatics platform and NMC have successfully been based on the J-Express analysis software suite [4].

In addition to algorithms and tools, a suitable infrastructure for step-by-step analysis of your workflow, as well as sharing of data, results and methods across disciplines in a project group, is vital for proper utilization of your data. This is the aim of the eSysbio project, and components of the system are currently in use to implement the StoreBioinfo portal for providing high capacity storage for Life Science data sets in NorStore storage resource (national e-infrastructure).

Based on the total research experience and expertise of CBU and on the analysis and e-infrastructure competence built in the national network operated over the 9 previous years, we have coordinated an application for establishing a Norwegian node of the ELIXIR pan-European infrastructure network for bioinformatics.

References

1. Star et al. The genome sequence of Atlantic cod reveals a unique immune system, *Nature*, **2011**, 477, 207-210.
2. Balzer et al. Characteristics of 454 pyrosequencing data, *Bioinformatics*, **2010**, 26, i420-i425.
3. Quince et al. Removing noise from pyrosequenced amplicons, *BMC Bioinformatics*, **2011**, 12, 38.
4. Stavrum et al. Analysis of gene-expression data using J-Express, *Curr Protoc Bioinfo*, **2008**, Chapter 7, Unit 7.3.

Relevant Web sites

5. <http://www.bioinfo.no/>
6. <http://www.microarray.no/>
7. <http://jexpress.bioinfo.no/>
8. <http://www.esysbio.org/>
9. <http://storebioinfo.norstore.no/>