

## From cutadapt to sequencetools (sqt): a versatile toolset for sequencing projects

**Marcel Martin, Sven Rahmann**

Bioinformatics, Computer Science XI, TU Dortmund, 44221 Dortmund, Germany

Genome Informatics, Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany

<http://www.rahmannlab.de/>

We are developing a suite of scriptable tools for both small and large typical tasks arising in high-throughput sequencing projects. Following the nix philosophy, each tool has a specific task, and power and flexibility come from the ability to combine these tools in various ways.

As an example, we present cutadapt in details: When small RNA is sequenced on current sequencing machines, the resulting reads are usually longer than the RNA and therefore contain parts of the 3' adapter. That adapter must be found and removed error-tolerantly from each read before read mapping. Previous solutions are either hard to use or do not offer required features, in particular support for color space data. As an easy to use alternative, we developed the command-line tool cutadapt, which supports 454, Illumina and SOLiD (color space) data, offers two adapter trimming algorithms, and has other useful features.

This, and other tools, are presently organized in a toolset that will be available under the name sqt. We will briefly outline the design idea of this set of tools and report on the current state of development.

### References

1. Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads EMBnet.journal 17(1), May 2011.

### Relevant Web sites

2. Cutadapt, including its MIT-licensed source code, is available at <http://code.google.com/p/cutadapt/>
3. Sqt website: <http://code.google.com/p/sqt/>