

Ausschreibung Masterarbeit:

Verbesserung des Mappings von RNASeq durch Genommaskierung

Junior research group transcriptome Bioinformatics

In RNASequencing Experimenten wird das Transkriptom zu einem gewissen Zeitpunkt in Form von relativ kurzen Stücken von RNA Molekülen, sogenannten reads untersucht. Ein wichtiger Schritt bei der Auswertung solcher Experimente ist das Mappen von diesen kurzen Sequenzen auf das Genom des Sequenzierten Organismus. Dabei beobachten wir immer wieder Sequenzstücke, die von verschiedenen Chromosomen zu kommen scheinen, das heisst ein Teil der Sequenz kommt von Chromosom A, während ein anderer von Chromosom B kommt. Ein anderes Problem kommt durch annotierte und nicht annotierte Pseudogene und Repeat Sequenzen zustande. Vor allem Pseudogene haben normalerweise keine Introns, sodass es richtig erscheint, ein Sequenzstueck auf das Pseudogen und nicht auf das eigentliche Ursprungsgen zu mappen.

Ziel der Arbeit ist es, die Gründe für Fehlmapping zu identifizieren und katalogisieren, sowie ein Maskiertes Genom zu erstellen, in dem die Probleme idealerweise nicht mehr auftreten.

Wir suchen eine Masterstudentin oder einen Masterstudenten, der oder die Kenntnisse in LINUX, bash, einer Skriptsprache (Perl oder python) sowie idealerweise in R besitzt.

Bei Interesse mailen Sie bitte Ihre Bewerbung an:
berni@bioinf.uni-leipzig.de

Master thesis:

Better RNASeq mapping through masking of genome

Junior research group transcriptome Bioinformatics

In RNA sequencing experiments, the transcriptome is analyzed via relatively short RNA molecules at a certain point in time. An important step in the analysis of the resulting data is the mapping of this short sequences (reads) back onto the genome of the sequenced organism. In mapping, we often observe sequences that partially map onto different chromosomes. Another Problem are annotated and not annotated pseudogenes and repeats. These usually do not contain introns and are therefore mapped to instead of the actual gene of origin.

The aim of this work is to identify and analyze mismappings and generate a masked genome where these problems do not exist.

We are looking for a student with knowledge in LINUX, bash, a scripting language (Perl or python) and possibly R.

If you are interested in this work please contact:
berni@bioinf.uni-leipzig.de