Fachhochschule Nordwestschweiz
Hochschule für Life Sciences

Programmierung einer Auswertsoftware für RNA-seq Daten

Patrick Pfeifer

Life Science Technologies, Biomedizinische Informatik

KURZZUSAMMENFASSUNG

Die Vorliegende Arbeit beschreibt die Architektur und die Anwendungsmöglichkeiten einer neuen Software zur Analyse von RNA-seq Daten.

EINLEITUNG

Die Einführung neuartiger (eng. next-generation) Technologien der Gen-Sequenzierung ist ein wichtiger Meilenstein auf dem Weg zu einem besseren und tieferen Verständnis der Funktionsweise der Zellen.

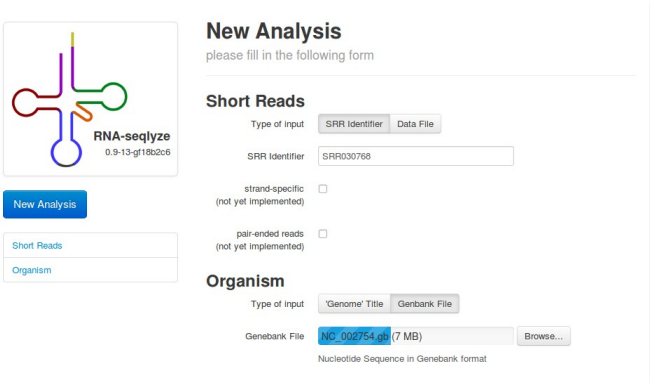
Bei der Durchführung von RNA-seq Experimenten fallen grosse Mengen von Daten in Form von relativ kurzen Sequenzabschnitten (< 200 Basen) an. Die Sequenzdaten werden dann an die NCBI und andere zentrale Sequenzarchive (SRA - Sequence Read Archive) übermittelt. Zum heutigen Tag sind bereits grosse Mengen an Daten gesammelt worden, die nun öffentlich online verfügbar sind.

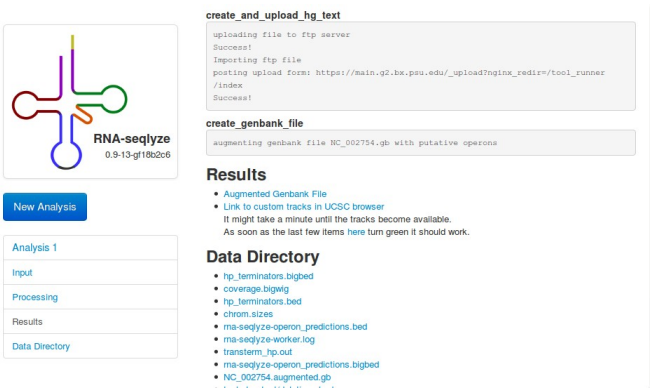
BESCHREIBUNG

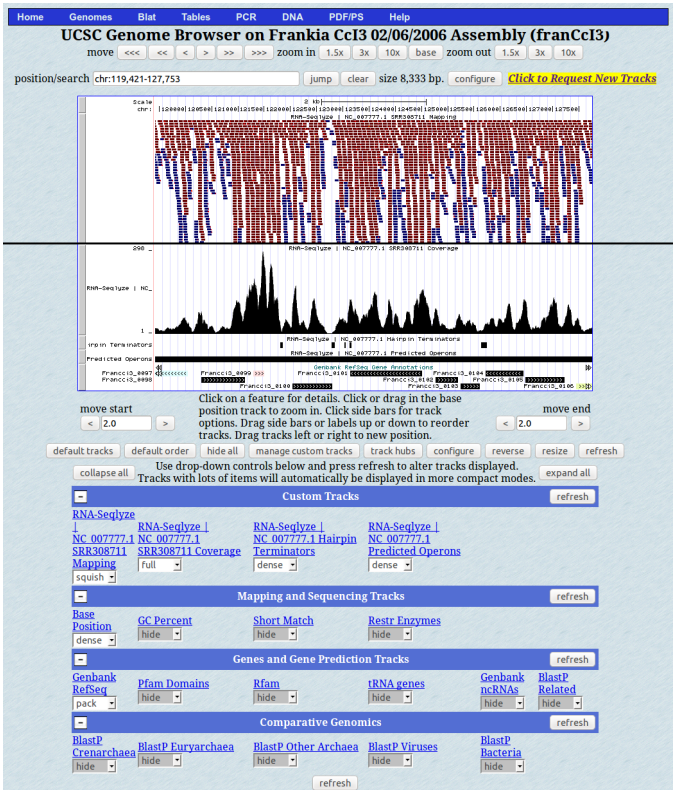
Die in dieser Diplomarbeit programmierte Web-Anwendung wurde erstellt, um die von RNA-seq Experimenten generierten Daten in einer bereits existierenden, populären, Visualisierungs-anwendung, dem *UCSC Archaea Browser* anzuzeigen. Durch die Verwendung der von dieser Anwendung unterstützten, bzw. definierten, Protokolle zur Formatierung und Darstellung der Daten, stehen dem Anwender viele Funktionen zur Verfügung, die sich mit einer gekapselten Desktop-Anwendung nur unter grossem Aufwand realisieren liessen. So können die Daten im *UCSC Archaea Browser* beispielsweise immer im Kontext mit den *aktuellen* RefSeq Gen-Annotationen betrachtet werden.

Diese Daten werden mit der bereits bekannten Genomsequenz verglichen um die Bereiche des Genoms zu finden, die transkribiert werden. Transkription wird nicht für alle Genombereiche erwartet, sondern nur dort, wo sich proteinkodierende Abschnitte bzw. Gene für RNA-Moleküle befinden. Eine Analyse der RNA-Seq Daten wird daher durch die Visualisierung der RNA-Seq Daten entscheidend unterstützt. Darüber hinaus kann die Annotation eines Genoms durch die Verfügbarkeit der Transkriptdaten verbessert werden.

Abbildungen: oben und mitte: Screenshots der Auswertsoftware;
unten: Screenshot des UCSC Browser mit
von der Auswertsoftware generierten custom Tracks







Begleittdozent/in: Prof. Dr. Georg Lipps

Experte: Dr. Sven Schuierer