

Übung Galaxy

Galaxy ist ein Tool, das verschiedene bioinformatische Tools online zur Verfügung stellt. Es gibt verschiedene Institute, die eigene Varianten von Galaxy für spezielle Fragestellungen zur Verfügung stellen.

In unserem Fall werden wir die Hauptgalaxy-Seite <https://usegalaxy.org> verwenden

1) Galaxy Account anlegen

Um Die Übung durchführen zu können, ist es notwendig einen Account anzulegen. Klicken Sie bei User (ganz oben) auf Register und geben Sie E-mail adresse und Passwort an. Sie bekommen ein Bestätigungs E-mail mit einem Link zur Aktivierung Ihres Accounts. Melden Sie sich danach an.

2) Daten in Galaxy laden

Laden Sie Daten in Galaxy indem sie unter “Get data” (oben links) die Option EBI SRA auswählen. Von dort kommen Sie zur Homepage des europäischen Sequence read archives. Suchen Sie die Daten ERR1016568. Wählen Sie den Run ERR1016568 aus und suchen Sie den Link FastQ Files (Galaxy). Die Daten werden dann automatisch in Galaxy geladen und erscheinen im rechten Panel (History). Sobald der Abschnitt grün leuchtet, sind die Daten fertig. Bei Klick auf das Auge können die Rohdaten betrachtet werden, Klicken Sie auf den Stift um Attribute zu verändern und ändern sie im Tab Datatype den Datentyp auf fastqsanger.gz.

2.1) Rohdaten Qualität

Verschaffen Sie sich einen Überblick über die Qualität der Rohdaten über die **“NGS:QC and Manipulation”** und dann **“FastQC:Read QC”** auf der linken Seite in der Tool-Sektion.

Wählen Sie bei **“Short Read data from your current history”** die hochgeladene Datei aus und klicken Sie auf **“Execute”**. Warten Sie bis das Programm fertig ist und klicken Sie dann auf das Auge. Sie kommen nun auf die Zusammenfassung von FastQC.

Was fällt Ihnen auf?

Lesen Sie die Beschreibung der Studie (<http://www.ebi.ac.uk/ena/data/view/PRJEB10828>) im EBI SRA durch.

2.2) Alignment

Das Alignment wird im öffentlichen Server blockiert, da es zuviele Ressourcen in Anspruch nehmen würde. Theoretisch würde es so gehen:

Bei **“NGS: Mapping”**: Klicken Sie auf **“Map with BWA -MEM”**. Sie können die Standardeinstellungen beibehalten ändern Sie nur das Referenzgenom auf **“Homo sapiens**

(hg19 with mtDNA replaced with rCRS...) und wählen Sie die hochgeladene FastQ-Datei aus. Bei Single or Paired-end reads wählen Sie **“Single”**. Sonst müssen Sie keine Settings verändern und klicken Sie auf **“Execute”**.

2.3) Alignment laden

Laden Sie Alignment Daten in BAM-Format in Galaxy indem sie unter “Get data” (oben links) die Option EBI SRA auswählen. Von dort kommen Sie zur Homepage des europäischen Sequence read archives. Suchen Sie die Daten ERR1016568. Wählen Sie den Run ERR1016568 aus und suchen Sie den Link Submitted Files (Galaxy). Verschaffen Sie sich wieder einen Überblick über die Qualität mittels FastQC (siehe 2.1). Verändern Sie bei den Attributen die passende Datenbank “Homo sapiens b37 (hg_g1k_v37)”.

2.4) Alignmentstatistiken

Unter NGS:SAMtools finden Sie den Punkt ldxStats. Damit kann aus einer BAM-Datei eine Tabelle erstellt, werden, aus der man lesen kann, wieviele Reads wo hin aligned werden konnten.

2.5) SNP calling

Bei **“NGS: SAM Tools”** wählen sie **“Generate Pileup”**. Wählen Sie die BAM Datei (sollte bereits ausgewählt sein) und wählen Sie Human (hg_g1k_v37) als Referenzgenom aus. Klicken Sie auf **“Execute”** und warten Sie bis das Programm fertig ist. Sehen Sie sich den Output über das Augensymbol an. Pileup erstellt einen Eintrag pro Position, egal ob Abweichungen gefunden werden oder nicht.

2.6) SNP filtering

Bei **“NGS: SAM Tools”** wählen sie **“Filter Pileup”**. Wählen Sie die Pileup Datei (sollte bereits ausgewählt sein; ansonsten muss der Datentyp der Pileup Datei kontrolliert und angepasst werden) . Dies filtert Positionen die sich von der Referenz unterscheiden, eine gewisse Qualität aufzeigen und zumindest von drei Reads gecovered werden. Klicken Sie auf **“Execute”** und warten Sie bis das Programm fertig ist. Sehen Sie sich nun die gefundenen Varianten an.