

# Hauptseminar Bioinformatik – Informationen und Themen

Aktuelle Methoden für Next-Generation-  
Sequencing

Prof. Dr. Caroline Friedel  
WS 2016



# Organisatorisches



- Zeit und Ort: Montag, 10-12, Amalienstr. 17, A105
- Vortragsdauer: 50-60 min + Diskussion
- Anwesenheitspflicht
- 2 Vorbesprechungen mit Betreuer:
  - 3 Wochen vor Vortrag erster Entwurf des Vortrags
  - 1 Woche vor Vortrag fertiger Vortrag
- Termine selbständig ausmachen !!
- Wenn 3 Wochen vorher kein Termin für die erste Vorbesprechung ausgemacht wurde  $\Rightarrow$  Vortrag fällt aus
- Wenn 1 Woche vorher kein fertiger Vortrag vorgelegt wurde  $\Rightarrow$  Vortrag fällt aus



- Scheinvoraussetzungen:
  - Vortrag
  - Teilnahme an Diskussion:
    - Mindestens 6 sinnvolle Fragen über das ganze Seminar gerechnet
  - Ausarbeitung (Deadline: 2 Wochen nach Vortrag)
- Siehe auch die allgemeine Hinweise von Prof. Dr. Volker Heun zu Seminaren (soweit nicht explizit anders angegeben):
  - [https://www.bio.ifi.lmu.de/webfm\\_send/3767](https://www.bio.ifi.lmu.de/webfm_send/3767)

# Themen und Termine



Datum	Thema
17.10.16	Bowtie and Bowtie2
24.10.16	<i>Voraussichtlich kein Vortrag</i>
31.10.16	BWA
7.11.16	Genome assembly with de Bruijn graphs
14.11.16	SGA
21.11.16	STAR
28.11.16	TopHat
5.12.16	ContextMap
12.12.16	Transcript reconstruction
19.12.16	<i>Voraussichtlich kein Vortrag</i>
16.1.17	Alternative Splicing
23.1.17	Alternative polyadenylation
30.1.17	circRNAs



- Aus der Literatur die wichtigsten Inhalte herausarbeiten
  - Eigenes Literaturstudium (und weitere Literatursuche!)
  - Verstehen der Literatur (inklusive Sekundärliteratur)
  - Aufbereiten für den Vortrag
- Die Themen gut verständlich und anschaulich darstellen
  - Überblick herausarbeiten
  - Verständnis vermitteln
  - Beispiele!
- Auch aus anderen Vorträgen lernen (daher Fragen stellen!)



- Schriftgröße nicht zu klein (Arial, 24 pt)
- Farben sinnvoll, aber sparsam und einheitlich einsetzen
- Satzfragmente, keine vollständigen Sätze
- Besser: Bilder statt Worte
- Folien nicht überladen
- Beispiele
- Wesentliches deutlich erklären
- mit Bildern/Zeichnungen/Skizzen unterstützen
- klare Strukturierung (keine Verweise auf später)
- freie Rede!



- Formale Gliederung in 3 Teile:
  - Einleitung
  - Hauptteil
  - Schluss
- Anfang und Schluss
  - Problemstellung motivieren
  - Zusammenfassung (Take-Home-Message)
- Hauptteil: oft inhaltliche Unterteilung in 3 Teile:
  - Biologischer/Allgemeiner Hintergrund
  - Methode
  - Ergebnisse



- ca. 12 Seiten
- Inhalt in eigenen Worten wiedergeben!
- Hier wenig Bilder
- Hauptziel: Einüben des Schreibens
- Klare Gedankengänge
- Korrekte Form (Absätze, Blocksatz, Font)
- Korrektes Zitieren
- Abgabe 2 Wochen nach dem Vortrag

# Themen



- Literaturangaben immer nur Startpunkt für die Literaturrecherche
- Ziel ist die Methoden verständlich und ausführlich genug darzustellen
- Aber auch nicht zu ausführlich und mit unnötigen Details
- In 50-60 Minuten
- Falls Vortrag zu lang: Kürzen und auf Wesentliches kondensieren
- Falls Vortrag zu kurz: Mehr Details zu Methoden bzw. zusätzliches Material
- Nicht einfach nur den Artikel präsentieren!

# Read Alignment



## 1. Bowtie and Bowtie2:

- *Langmead, et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25*
- *Langmead and Salzberg. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods, 9:357-359*

## 2. BWA

- *Li and Durbin. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60*
- *Li and Durbin. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics, 26: 589-595*



## 3. Genome assembly with de Bruijn graphs

- *Compeau et al. (2011) How to apply de Bruijn graphs to genome assembly. Nature Biotechnol. 29, 987–991*
- *Zerbino, D.R. & Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.*

## 4. String Graph Assembler (SGA)

- *Simpson and Durbin (2012) Efficient de novo assembly of large genomes using compressed data structures, Genome Res. 22: 549-556*
- *Simpson, J.T., (2012) Efficient sequence assembly and variant calling using compressed data structures, PhD thesis, chapter 2*



## 5. STAR

- *Dobin et al. (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29: 15-21*

## 6. TopHat

- *Trapnell, et al. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25: 1105-1111.*
- *Kim, et al. (2011) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 14:R36*

## 7. ContextMap

- *Bonfert et al. (2015) ContextMap 2: fast and accurate context-based RNA-seq mapping. BMC Bioinformatics, 16:122*
- *Bonfert, et al. (2013) Mining RNA-Seq Data for Infections and Contaminations. PloS one, vol 8, pp. e73071.*

# Transcript reconstruction and alternative splicing



## 8. Transcript reconstruction

- *Trapnell, C. et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515*
- *Guttman et al. (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nature Biotechnol. 28,503–510*

## 9. Alternative Splicing:

- *Shen, S., et al. (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. PNAS, 111:E5593-E5601.*



## 10. Alternative polyadenylation

- *Xia, Z., et al. (2014) Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nature Communications 5, 5274.*

## 11. circRNAs

- *Gao et al. (2015) CIRC: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biology, 16:4*
- *Cheng et al. (2016) Specific identification and quantification of circular RNAs from sequencing data. Bioinformatics 32:1094-6*