Sandro Andreotti
Bioinformatics Solution Center

Freie Universität Berlin

Projektmanagement (Softwarepraktikum)

Thema: Workflows

Typisches Szenario in der Praxis

Benötigt: Auswertung biologischer Massendaten z.B.
- NGS
- Massenspektrometrie
- Mikroskopie

(Leider) selten!

Es existiert ein Tool für die gesamte Analyse!

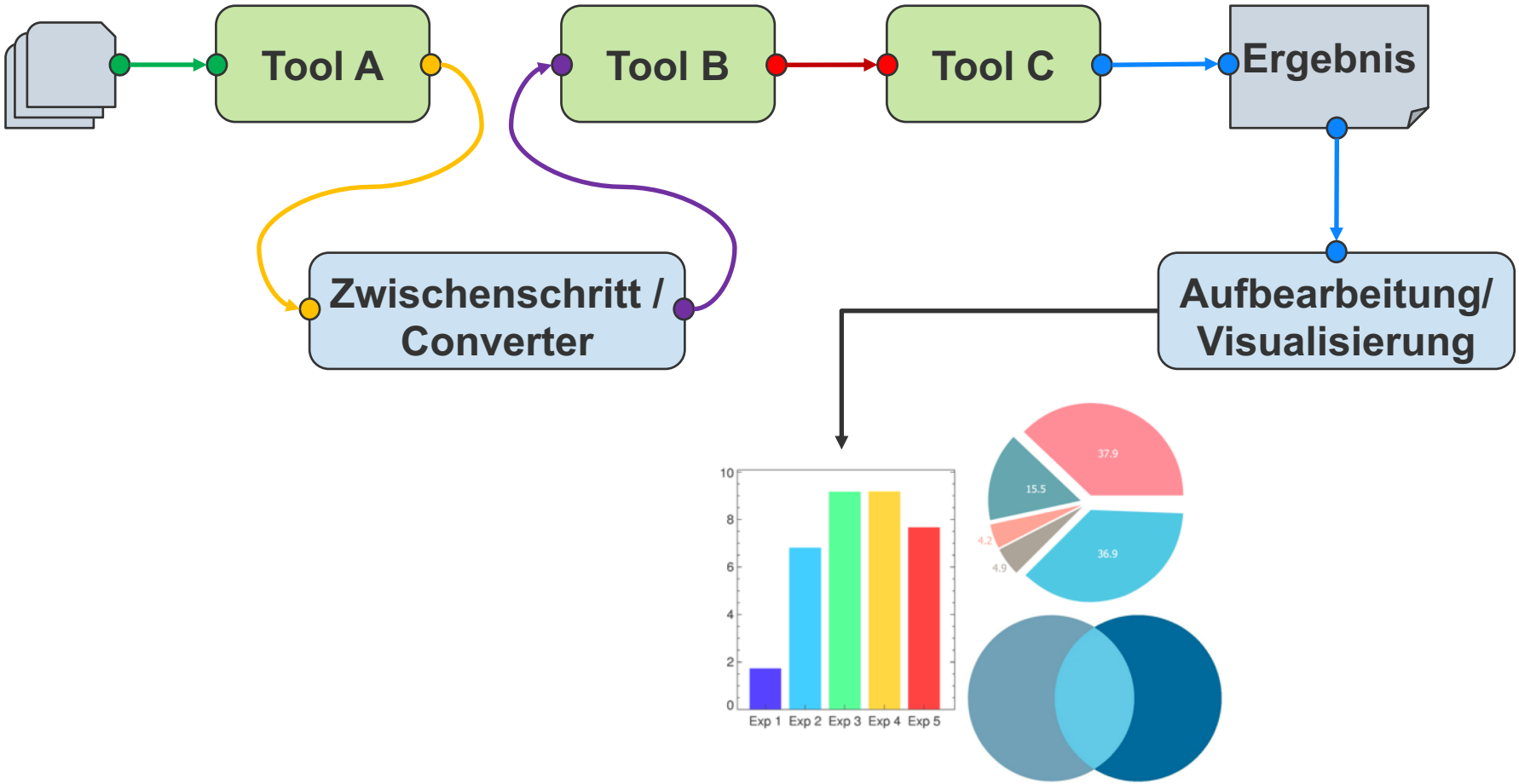
(Zum Glück) selten!

Es existiert noch gar keine Software! Man muss alles selbst entwickeln!

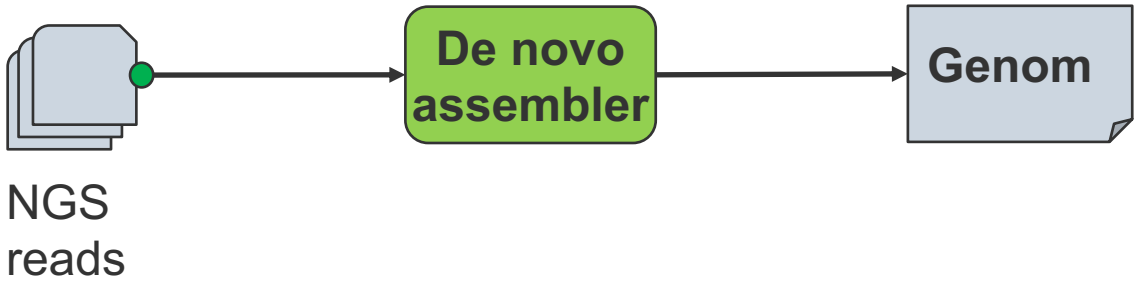
Häufig!

Es existieren Tools für einzelne Schritte der Analyse!

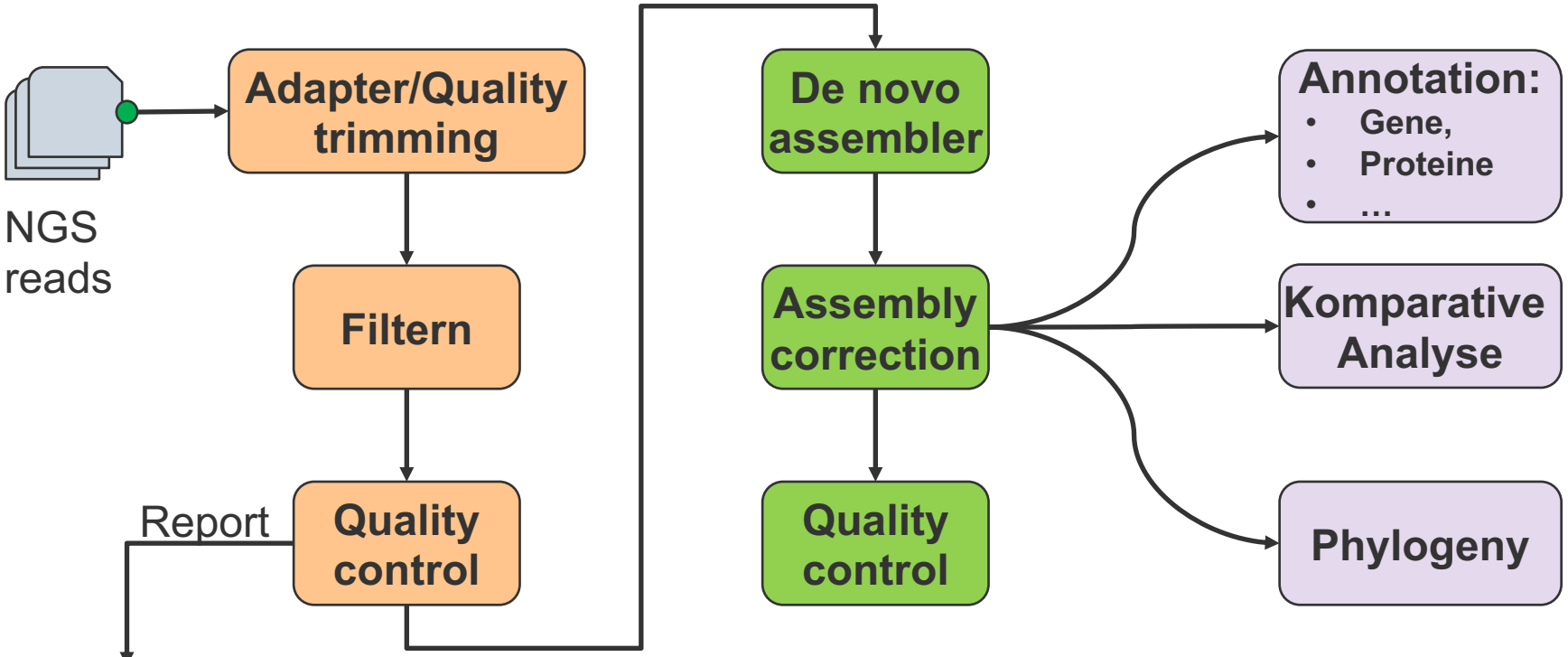
Einzelne Tools → Analyse Pipeline

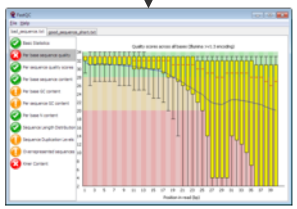


Beispiel: Genom Assemblierung



Beispiel: Genom Assemblierung

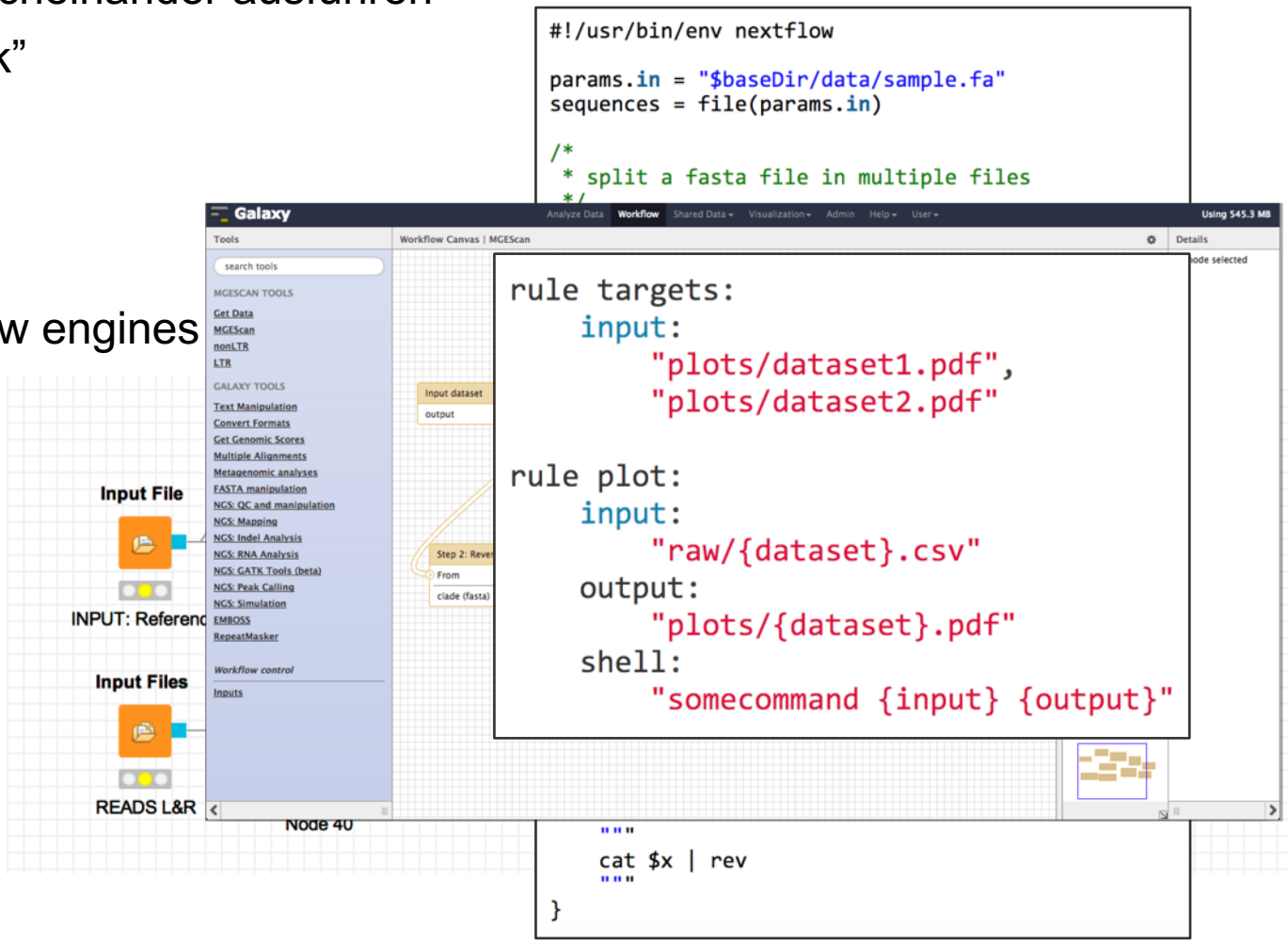




From: FastQC

Realisierung

1. Von Hand Tools nacheinander ausführen
2. Eigenes “framework”
   - Batch script
   - Python script
   - ...
3. Generische workflow engines
   - Script basiert:
     - Snakemake
     - Nextflow
   - GUI basiert:
     - Galaxy
     - KNIME



Ablauf des Praktikums

1. Tutorial (eine Woche) zur Einarbeitung in KNIME und Snakemake
2. Literaturrecherche zur Auswahl einer aktuellen, relevanten und relativ komplexeren Analyse in der Bioinformatik
3. Weitergehende Recherche nach bestehenden '*state of the art*' Workflows
4. Implementierung, Erweiterung, Verbesserung der Workflows in Snakemake und KNIME
5. Evaluierung auf realen Daten

Zeitplan	
Vorbesprechung	**25.2.**
Tutorial	**11.3. – 15.3.**
Recherche zu Analyse und Workflows	16.3. – 27.4.
Bearbeitung	bis 3.5.

Quantitative Aufteilung: (in %)
Praktische Programmierarbeit: 50%
Soft Skills: 50%

Verwendete Programmiersprache(n):
R, Python oder andere Skriptsprache

Schwierigkeitsgrad
A Programmieren ★★★★
B Biologie/Chemie ★
C Projektmanagement ★★★

Erforderliche Vorkenntnisse:
R, Python

Weitere Informationen