



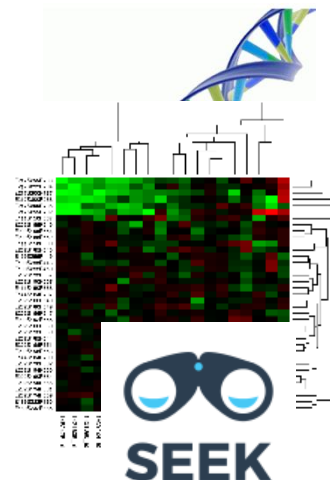
Vergabe einer Masterarbeit (Informatik/MPM): Verknüpfung von Forschungsplattformen und deren Data Provenance

Hintergrund / Problemstellung:

Im Rahmen früherer Projekte wurden und werden die Forschungsplattformen tranSMART und cBioPortal am Uniklinikum etabliert. Diese bieten ein ausgefeiltes Datenmodell zur Integration von klinischen und omics-Daten sowie rudimentäre Analyse-Methoden derselben, z.B. für die Suche nach Biomarkern.

In Projekten von Grundlagenforschern der Bioinformatik oder Systemmedizin werden ebenfalls Omics-Daten generiert und in Forschungsplattform wie SEEK gespeichert oder mittels Galaxy prozessiert.

Bislang existieren die Forschungsplattformen tranSMART [1] & cBioPortal [2] sowie SEEK [3] & Galaxy[4] nur losgelöst nebeneinander; wie ein Zusammenspiel aussehen könnte ist unbekannt. Ebenso, inwiefern sich SEEK&Galaxy zur Versionierung („Data Provenance“) und Langzeitarchivierung eignen, wofür es seitens der Universität Göttingen vielversprechende Indizien ([Link](#)) gibt.



Ziele dieser Arbeit:

- Verknüpfung von tranSMART, cBioPortal, SEEK und Galaxy (soweit sinnig&möglich)
- Konzept für Data Provenance
- Kongresspublikation darüber, z.B. in *Studies in Health Technology and Informatics*

Aufgaben / Fragestellungen (Vorschlag):

- Einarbeitung
 - Struktur / Natur der Omics-Daten der o.g. Plattformen sowie die damit durchgeführten (Bioinformatik)-Analysen
 - Struktur, Features und Schnittstellen der o.g. Plattformen
- Konzeptionelle Phase
 - Kleine Anforderungsanalyse bzgl. Bedarfs seitens der beteiligten Forscher
 - Wie können die jeweiligen Programme über ihre Schnittstellen zu einer Omics-Pipeline miteinander verbunden werden (Vorbild siehe [5])?
- Praktische Phase
 - Prototypische Umsetzung
- Evaluation
 - technisch
 - tatsächlicher Nutzung/Nutzbarkeit

Zeitraumen:

- Beginn jederzeit

Anforderungen / Voraussetzungen:

- Grundkenntnisse der (Molekular-)Biologie hilfreich
- Studiengang: Informatik oder MPM mit deutlicher Affinität zum Installieren bzw. Parametrieren von Open-Source Programmen auf Linux-Rechnern.

Einstimmende Lektüre:

- [1] **tranSMART**: Scheufele, Elisabeth, et al. "tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform." AMIA Summits on Translational Science Proceedings 2014 (2014): 96. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4333702/>
- [2] **cBioPortal**: Gao, Jianjiog, et al. "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal." *Sci. Signal.* 6.269 (2013): pl1-pl1. <https://stke.sciencemag.org/content/6/269/pl1.abstract>
- [3] **SEEK**: Wolstencroft, Katherine, et al. "SEEK: a systems biology data and model management platform." *BMC systems biology* 9.1 (2015): 33. <http://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-015-0174-y>
- [4] **Galaxy**: Goecks, Jeremy, Anton Nekrutenko, and James Taylor. "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." *Genome biology* 11.8 (2010): 1. <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-8-r86>
- [5] **Verbindung dieser/ähnlicher Plattformen**: Satagopam, Venkata, et al. "Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases." *Big Data* 4.2 (2016): 97-108. <http://online.liebertpub.com/doi/full/10.1089/big.2015.0057>