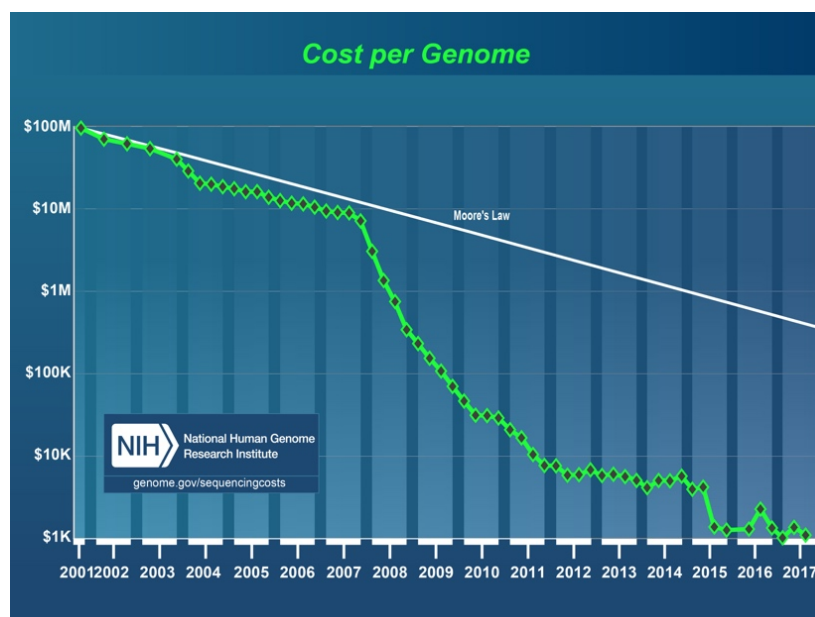


# Heterogenous Computing in Genomics

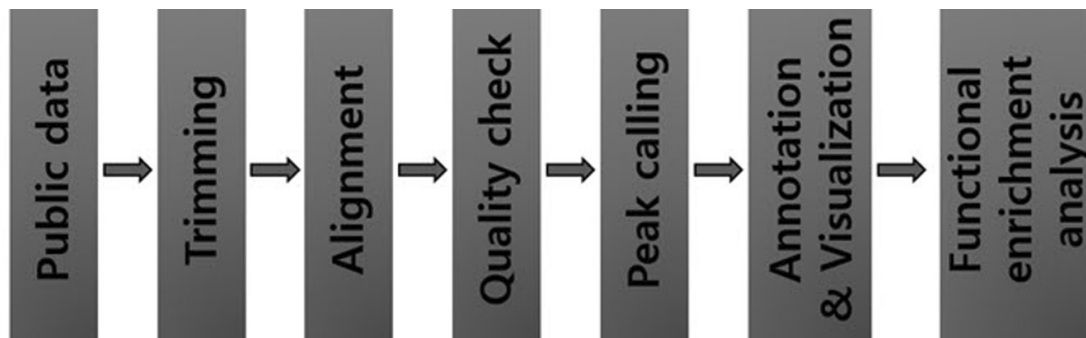
## Hintergrund / Motivation

In der Halbleiterindustrie zählt das Moore'sche Gesetz zu einer der wohl bekanntesten Metriken, mit welcher Geschwindigkeit sich die Transistorzahl und damit die Rechenleistung in Prozessoren weiterentwickelt. Aber nicht nur in der Prozessorfertigung ist das Moore'sche Gesetz eine wichtige Messlatte: In der Biotechnologie-Branche hat das Gesetz lange die Fortschritte in der Genom-Sequenzierungstechnik akkurat beschrieben. Seit 2008 jedoch wurde das von Moore prognostizierte Entwicklungstempo mehrfach überboten, wodurch die Kosten für die Sequenzierung des menschlichen Genoms im Jahr 2017 schließlich die Schallmauer von 1000 USD erreichen konnten [1]. Durch die immer weiter sinkenden Sequenzierungskosten rückt die Vision von flächendeckend verfügbaren, patientennahen Laboren in greifbare Reichweite. Während die Technologien zur Datenerhebung solche rasanten Fortschritte erzielt haben, stagniert die Entwicklung in der Prozessortechnik [2]. Nur durch die Verwendung von heterogenen Rechnerarchitekturen kann die Rechenleistung in gewohntem Maße gesteigert werden, um so die Verarbeitung der stets wachsenden Datenmengen zu bewerkstelligen.



## Beschreibung

Viele Verarbeitungspipelines von Sequenzierungsdaten basieren nach wie vor auf unzureichend parallelisierten Werkzeugen und nutzen damit die Ressourcen moderner Prozessor- und Beschleunigerhardware nicht effizient aus. Ziel dieses Masterprojektes ist es, aus der Landschaft der häufig verwendeten Werkzeuge zur Verarbeitung von Sequenzierungsdaten einen interessanten Kandidaten auszuwählen und diesen unter Zuhilfenahme neuester Beschleunigertechnologien zu parallelisieren. Teilnehmer können abhängig von Interesse und Kenntnisstand auf GPU- und FPGA-Beschleuniger zurückgreifen, allerdings sind auch CPU-basierte Parallelisierungsstrategien nicht auszuschließen.



Die effiziente Nutzung von Parallelhardware eröffnet eine Reihe interessanter Fragestellungen, die im Rahmen des Masterprojekts verfolgt werden können. So ist nicht nur die schnellere Verarbeitung auf Höchstleistungshardware ein mögliches Optimierungsziel, sondern auch die Bewältigung von aufwändigen Sequenzierungspipelines (Beispiel siehe Abbildung [3]) auf kostengünstiger, energieeffizienter Hardware kann als mögliches Projektziel verfolgt werden.

Für GPU-basierte Beschleunigung werden wir auf OpenCL (C/C++) und ggf. Aparapi (Java) als bevorzugte Programmiermodelle setzen. Im Falle von FPGA-basierten Ansätzen verwenden wir das CAPI SNAP Framework, welches die Implementierung von Applikationslogik mittels High-Level Synthesis C/C++ Dialekten vereinfacht. Sollten sich die Teilnehmer für eine CPU-basierte Implementierungs-Strategie entscheiden, wird vermutlich OpenMP oder das Intel Thread Building Blocks Framework eingesetzt werden.

## Kontakt

Bei weiteren Fragen stehen wir gerne zur Verfügung. Projektverantwortlich sind Max Plauth, Felix Eberhardt sowie Andreas Polze.

[1] <https://www.genome.gov/sequencingcostsdata/>

[2] <http://www.nature.com/news/the-chips-are-down-for-moore-s-law-1.19338>

[3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5389943/>

