**NGSANE: A Lightweight Production Informatics Framework for High Throughput Data Analysis**

Fabian A. Buske

Cancer Epigenetics Program, Cancer Research Division, Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney 2010, Australia

Unprecedented computational capabilities and high-throughput data collection methods promise a new era of personalised, evidence-based healthcare, utilising individual genomic profiles to tailor health management as demonstrated by recent successes in rare genetic disorders [[1](#_ENREF_1), [2](#_ENREF_2)] or stratified cancer treatments [[3](#_ENREF_3)]. However, processing genomic information at a scale relevant for the health-system remains challenging due to high demands on data reproducibility and data provenance. Furthermore, the necessary computational requirements requires a large investment associated with compute hardware and IT personnel, which is a barrier to entry for small laboratories and difficult to maintain at peak times for larger institutes. This hampers the creation of time-reliable production informatics environments for clinical genomics. Commercial cloud computing frameworks, like Amazon Web Services (AWS) provide an economical alternative to in-house compute clusters as they allow outsourcing of computation to third-party providers, while retaining the software and compute flexibility.

To cater for this resource-hungry, fast pace yet sensitive environment of personalized medicine, we developed NGSANE [[4](#_ENREF_4)], a Linux-based, HPC-enabled framework that minimises overhead for set up and processing of new projects yet maintains full flexibility of custom scripting and data provenance when processing raw sequencing data either on a local cluster or Amazon’s Elastic Compute Cloud (EC2).

1. Bainbridge, M.N., et al., *Whole-genome sequencing for optimized patient management.* Sci Transl Med, 2011. **3**(87): p. 87re3-87re3.

2. Talkowski, M.E., et al., *Clinical diagnosis by whole-genome sequencing of a prenatal sample.* N Engl J Med, 2012. **367**(23): p. 2226-32.

3. Pellatt, A.J., et al., *Genetic and lifestyle influence on telomere length and subsequent risk of colon cancer in a case control study.* Int J Mol Epidemiol Genet, 2012. **3**(3): p. 184-194.

4. Buske, F.A., et al., *NGSANE: A Lightweight Production Informatics Framework for High Throughput Data Analysis.* Bioinformatics, accepted.