



The HPC4Health Network

Building an Ontario-wide platform for human health data

HPC4Health: High-Performance Data and Computing for Health Care

Copyright © September 2018

PUBLISHED BY HPC4HEALTH

WWW.HPCFORHEALTH.CA

Executive Summary

Staggeringly brillint summary goes here.

Human Health Research in the Era of Genomics

NEXT-GENERATION GENOMICS AND ELECTRONIC MEDICAL RECORDS have become ubiquitous almost simultaneously, opening completely new windows onto the field of human health research.

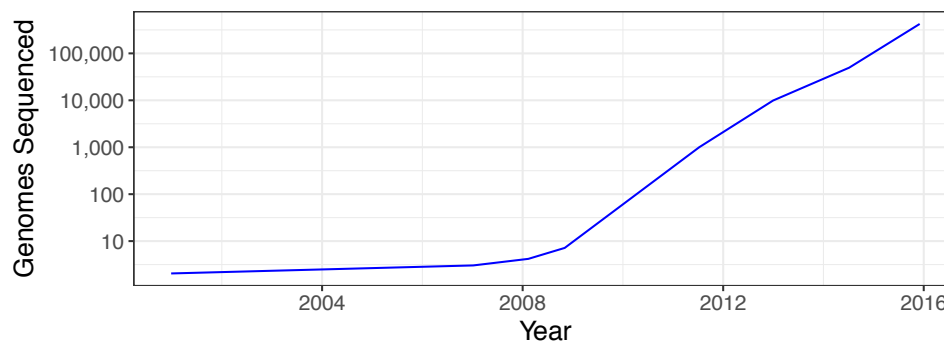


Figure 1: The cumulative number of human genomes sequenced over the past 15 years, data from [Stephens et al., 2015]. New technologies and data types have caused a inflection in the exponential rate of data growth which continues today.

In genomics, an exponentially-growing number of human genomes are being sequenced, making enormous amounts of evidence on the heredity of human disease states at previously inaccessible levels of detail (Fig 1). Data rates are only increasing as new devices become available, and new types of data are being generated; for instance, RNA sequencing allows us to go beyond simply sequencing “the” genome of an individual and instead measure the gene products being expressed in particular cells at particular times, giving us insight into not just predispositions but the precise disease state of cells over time.

Until now, the bulk of this genomic data has come from research projects — focused, generally short-term efforts to answer specific questions using genomic information. However, two changes in the practice of medicine are poised to radically increase the rate of genomic and human health data creation in Canada.

First, genomic medicine¹ is beginning to enter the standard of care, initially in the cases of hard-to-diagnose rare diseases or recurrent cancers, and starting to displace smaller and more limited genetic tests in other areas. The sheer scale of clinical medicine — in Canada, hospital spending alone is nearly seventy times the research funding budget of the CIHR — means that as adoption

¹ Genomic Medicine, as defined by the NIH's National Human Genome Research Institute: “An emerging medical discipline that involves using genomic information about an individual as part of their clinical care (*e.g.*, for diagnostic or therapeutic decision-making)”

² Electronic Medical Records, or EMRs, are digital version of paper chapters in hospitals or doctors offices; Electronic Health Records, or EHRs, are more integrated systems combining information across practices and institutions. EMRs are mature and growing in adoption rapidly, while EHR systems are still some time from being common.

increases, clinical genomic data creation will rapidly outpace that of research genomic data.

Second, information technology advances elsewhere in healthcare has led to the rapid adoption of Electronic Medical Record systems², describing a patient's condition, tests, and treatment in detail and at least partially in machine-readable form.

The rapid growth of genomic data volumes and the increasing depth and detail of clinical data present in EMRs offers enormous promise for human health research, with insights into both basic biology and to future treatments. The joint analyses of clinical and phenotypic health record data along with genomic information about the patient and their disease offers unprecedented opportunity for researchers to connect genetic predisposition, treatments, and outcomes, allowing the development of national, truly precision, medicine practices.

But making use of this data in an era of rapid growth raises multiple challenges. On the physical infrastructure side, simply making available the storage resources to capture and archive the onrush of data is a challenge, along with providing the computational power to perform increasingly sophisticated analyses. Architecting, building, and maintaining these systems, particularly tuned to the needs of health research, requires a specialized approach.

Human infrastructure is also required. Making productive use of the data means ensuring that the expertise exists and is available to for interpretation, and that those experts are continuously kept up to date on the new types of data and new techniques for analysis. This requires funding, ongoing training, and opportunities for professional recognition and growth of the experts, whether they be bioinformaticians, computational biologists, or systems administrators.

Finally, the unique challenges of dealing with health data means that the duty of care to patients to zealously protect the security and their privacy is paramount; sophisticated and enforceable policies around data governance and consent are required, along with international best practices around security and monitoring.

HPC4Health — a Made-in-Ontario Collaboration

The Tufte- \LaTeX document classes define a style similar to the style Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, tight integration of graphics with text, and well-set typography. This document aims to be at once a demonstration of the features of the Tufte- \LaTeX document classes and a style guide to their use.

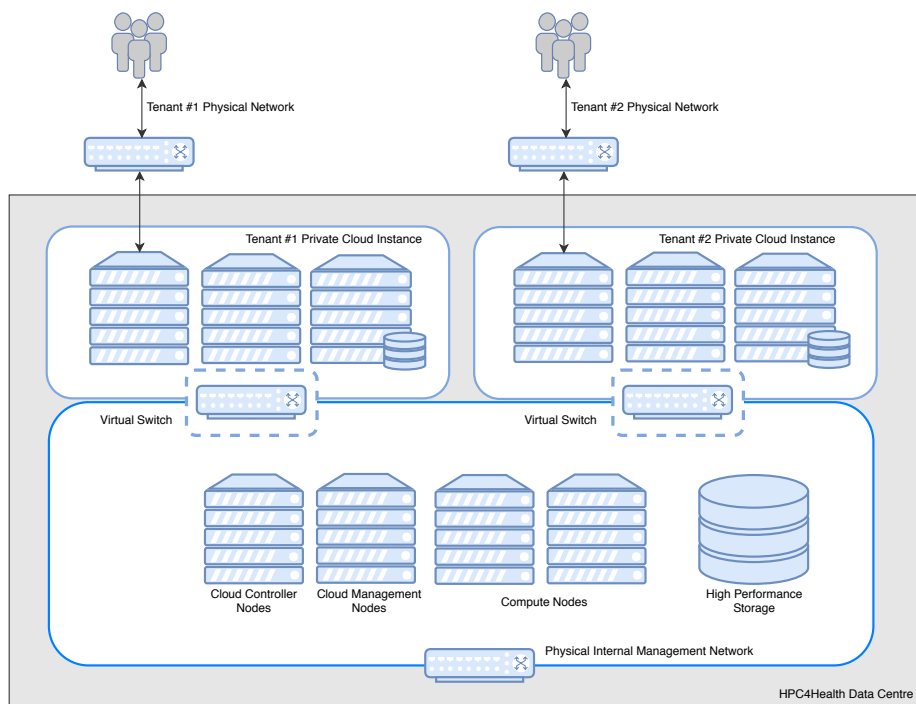


Figure 2: The HPC4Health Architecture

If you need more than two levels of section headings, you'll have to define them yourself at the moment; there are no pre-defined styles for anything below a `\subsection`. As Bringhurst points out in *The Elements of Typographic Style*, you should “use as many levels of headings as you need: no more, and no fewer.”

The Tufte- \LaTeX classes will emit an error if you try to use

`\subsubsection` and smaller headings.

IN HIS LATER BOOKS, Tufte starts each section with a bit of vertical space, a non-indented paragraph, and sets the first few words of the sentence in SMALL CAPS. To accomplish this using this style, use the `\newthought` command:

```
\newthought{In his later books}, Tufte starts...
```

Sidenotes

One of the most prominent and distinctive features of this style is the extensive use of sidenotes. There is a wide margin to provide ample room for sidenotes and small figures. Any `\footnotes` will automatically be converted to sidenotes.³ If you'd like to place ancillary information in the margin without the sidenote mark (the superscript number), you can use the `\marginnote` command.

The specification of the `\sidenote` command is:

```
\sidenote[⟨number⟩][⟨offset⟩]{Sidenote text.}
```

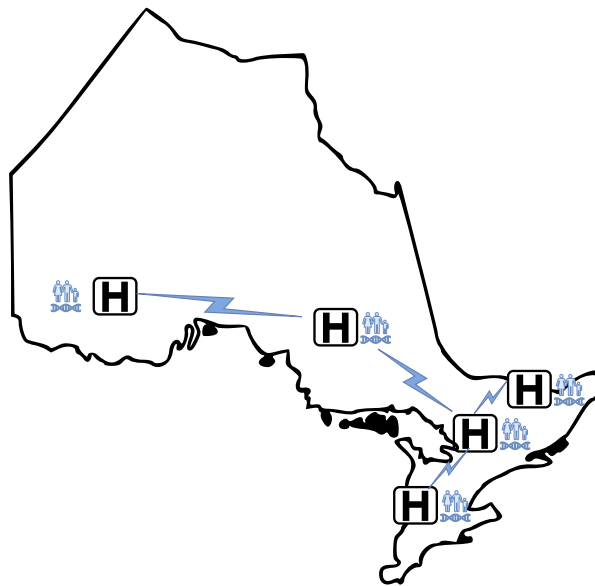
Both the `⟨number⟩` and `⟨offset⟩` arguments are optional. If you provide a `⟨number⟩` argument, then that number will be used as the sidenote number. It will change of the number of the current sidenote only and will not affect the numbering sequence of subsequent sidenotes.

³ This is a sidenote that was entered using the `\footnote` command.

This is a margin note. Notice that there isn't a number preceding the note, and there is no number in the main text where this note was written.

The H4H Network — Building on Strengths

Figure 3: The H4H Network



Bibliography

Zachary D Stephens, Skylar Y Lee, Faraz Faghri, Roy H Campbell, Chengxiang Zhai, Miles J Efron, Ravishankar Iyer, Michael C Schatz, Saurabh Sinha, and Gene E Robinson. Big data: astronomical or genetical? *PLoS biology*, 13(7): e1002195, 2015.

Index

`\footnote`, [8](#)

`\marginnote`, [8](#)

`\newthought`, [8](#)

`\sidenote`, [8](#)