



Aalto University
School of Science

Explaining mixture models through semantic pattern mining and matrix visualization

Prem Raj Adhikari¹, Anže Vavpetič², Jan Kralj², Nada Lavrač²,
Jaakko Hollmén¹

¹Department of Information and Computer Science
Helsinki Institute for Information Technology (HIIT)
Aalto University School of Science, Finland

²Department of Knowledge Technologies
Jožef Stefan Institute, Ljubljana, Slovenia

Story of a Statistician

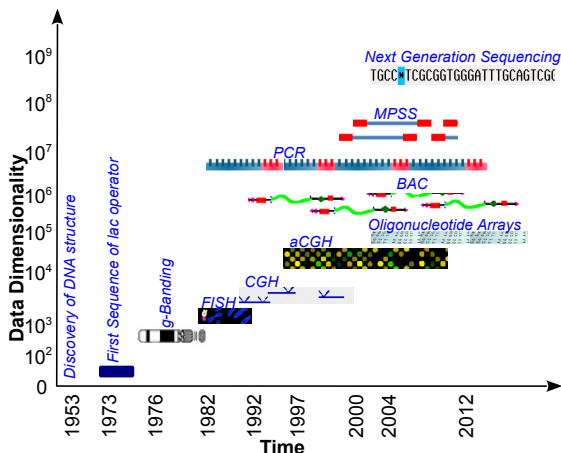


Importance of Using More Samples

- ▶ We expect that a sample of size 10 to be more informative than a sample of size 1
- ▶ A sample size of 100 more informative than 10, etc.
- ▶ How much better?
- ▶ The answer is **The Square Root Law**

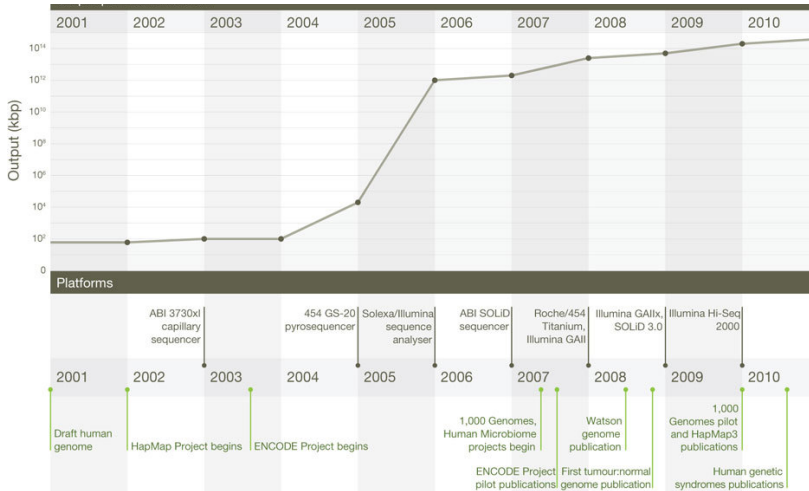
Accuracy of information = $\sqrt{\text{volume of information}}$

The Multiresolution Data in Biology



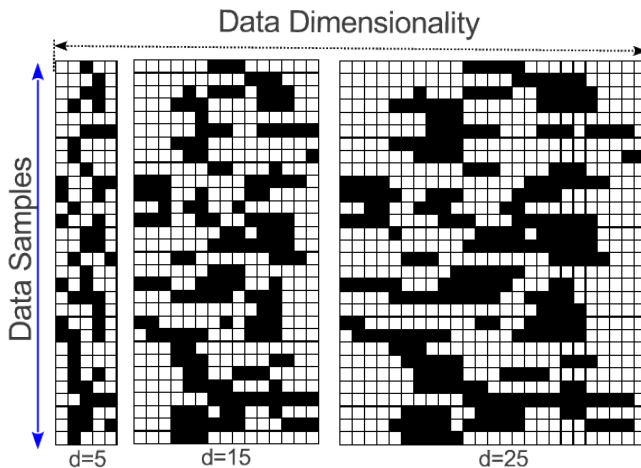
- Multiresolution data is everywhere: biology, computer vision, telecoms ...
- Older Generation Technology \Rightarrow Data in Coarse Resolution
- Newer Generation Technology \Rightarrow Data in Fine Resolution

The Multiresolution Data in Biology



E. R. Mardis, A decade's perspective on DNA sequencing technology, *Nature*, 2011

DNA Copy Number Amplification Dataset



How to analyze data in multiple resolutions in a single analysis?