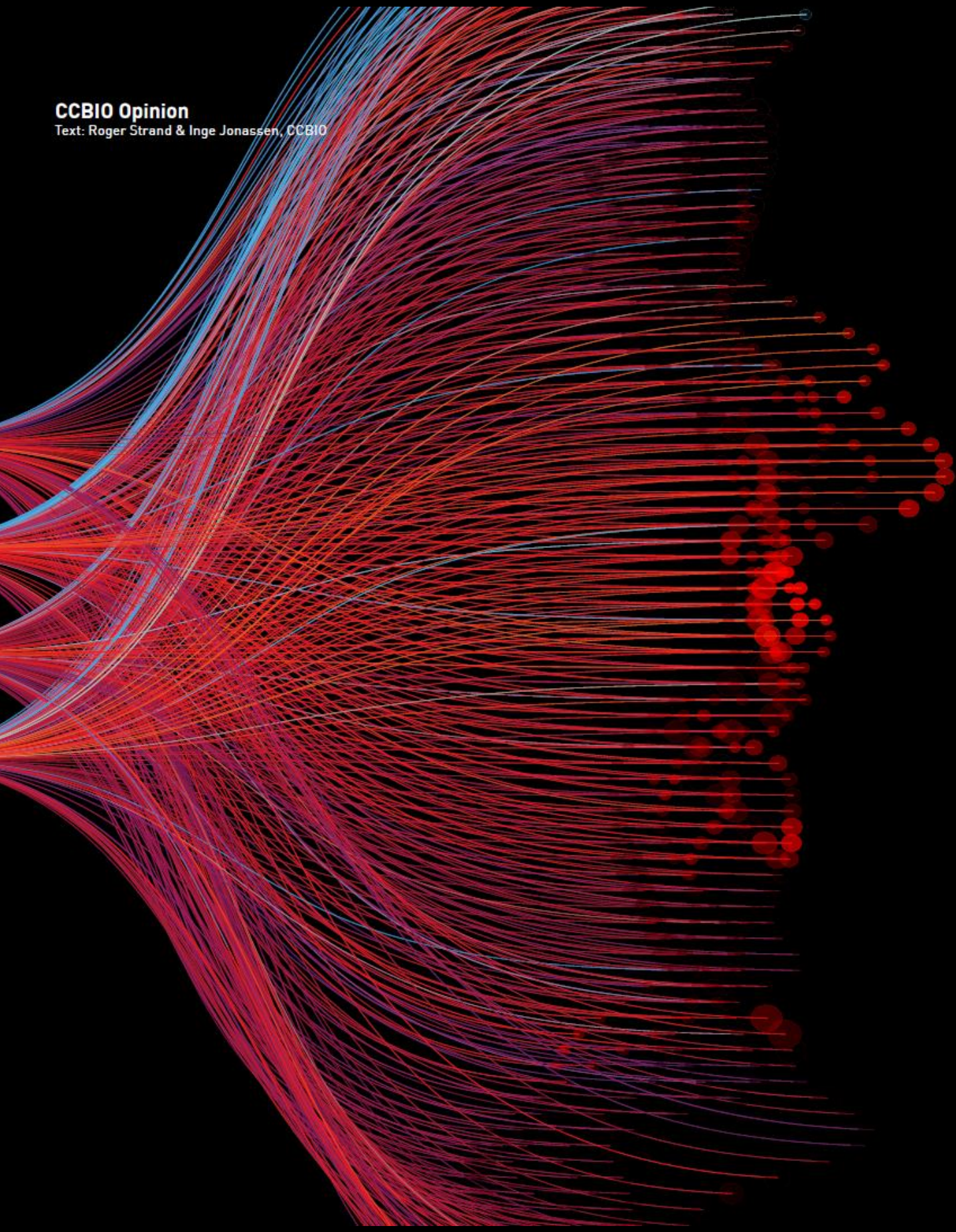

CCBIO Opinion

Text: Roger Strand & Inge Jonassen, CCBIO

Too Much Data? The Gastro Perspective

In popular culture, "TMI" denotes "Too Much Information". You ask someone "How are you today?" and he replies "Good, much better than last week when I was so constipated. This morning I was quite flatulent but it seems to have passed." In almost any other context than that of gastroenterology, this is Too Much Information. A simple "I'm fine, thank you" would do well.

In contemporary science, the equivalent of constipation and flatulence is caused by the uncontained acquisition of Too Much Data. A Nature editorial (2016) reported a future estimate of 0.4 terabytes (TB) of clinical data, 6 TB of genomics data and 1,100 TB "additional data" per patient. 1,100 TB of data is 100 times more than the total amount of textual information in the US Library of Congress. In CCBIO, as in most of biomedicine, acquisition of big data has become a daily routine. For instance, CCBIO's research was recently strengthened with the purchase of the Hyperion Imaging System, which takes immunohistochemistry to a new level, simultaneously measuring 35 protein markers across tissue sections by using imaging mass cytometry. Such advances give rise to hopes of scientific breakthroughs but also very practical questions about what to do with the data.

When the problems of big data hit in, one typical response is to call for help from bioinformatics, as if the bioinformaticians were a sort of Lactobacillus of the sciento-gastric ecosystem. They are not. Bioinformatics plays a key role in modern bioscience but it does not offer technical fixes and silver bullets to the challenges of big data. Unlike the digestive system, the purpose of scientific research is not to degrade its inputs but to refine and convert them: From information to knowledge; from knowledge to understanding. The secret ingredients in this process are ideas, models and thought. While the experienced bioinformatician may help expanding the repertoire of models and choosing and shaping the adequate approaches for data analysis and presentation, she or he cannot replace the need for clear scientific ideas about the purpose of the study, the hypotheses to be generated and possibly tested, the explanations to be sought, or the variables to be modelled and predicted. This predicament was anticipated already in 1973 by the philosopher Karl Popper, who warned that "too many dollars may chase too few ideas" and that "big science may destroy great science, and the publication explosion may kill ideas." (Popper 1975).

Yet, it is too late for nostalgic returns to single variables and data-low diets. In Norway, a new Centre for Digital Life was formed in 2016. The mission of this nation-wide network organization is to transform Norwegian biotechnology and life science by developing deeper interactions between bioscience on one hand and bioinformatics, engineering sciences, mathematics and exact sciences on the other. "Deeper interactions" means going beyond ordinary divisions of labour and facilitating mutual learning processes between the biological and the computational. Those who know about the biology need to learn more about the numbers, and the number crunchers need to learn about the biological systems being modelled. Indeed, the vision of Centre for Digital Life is that Norwegian biotechnology becomes transdisciplinary, in the sense that the collaborations across disciplines become so deep and so tight that the names and boundaries of disciplines become obsolete. Otherwise, the proper data digestion into real knowledge and understanding seems difficult to achieve.

CCBIO entered the Centre for Digital Life Norway as a partner in 2018. The collaborations between CCBIO bioscientists and distinguished bioinformatics groups, within as well as outside our own country, are multiplying and getting deeper and closer. CCBIO being a Centre of Excellence, we keep in mind Popper's warning that "the publication explosion may kill ideas." As we enter the next years of CCBIO, the focus should be less gastroenterology than gastronomy: Combine the use of Hyperion and our other sophisticated tools with digital and computational approaches and, importantly, time to think, in order to create data-rich and idea-rich knowledge that nourishes, inspires and enlightens. ••

REFERENCES: *Nature, 2016, "The power of big data must be harnessed for medical progress," 539:467-468. Popper, K. 1975, "The Rationality of Scientific Revolutions" (Herbert Spencer Lecture), in R. Harré (ed.), Problems of Scientific Revolution, Oxford: Oxford University Press, pp. 72–101.*