Single Cell Data Science

4 – 8 June 2018 @ Oort

Aims and Description: Single cell technologies seek to turn cell-to-cell heterogeneity at the level of DNA or RNA from a problem into a strong advantage; we are probably only beginning to appreciate the large variety of single-cell sequencing applications for fundamental and clinical research, documented by decisive advances in stem cell development, cancer progression and immune system formation. Most recent experimental advances allow to sequence the DNA or RNA of tens of thousands (and not only of hundreds) of single cells in a massively parallel manner. *Single Cell Data Science*, as a new era capturing these developments, is characterized by the urgent need for computer science driven solutions, and as such provides inspiration not only for biologists, but also for computer scientists, statisticians and modelers. The goal of this workshop was to survey the field and highlight urgent, but also inspiring computational and statistical challenges in the frame of a review paper or a white paper. This paper is supposed to arrange and clarify the technical challenges driving the field, and is to serve the participants as a reference compendium for future individual work.

Tangible Outcome: As an incentive for active participation, we have been writing a review / white paper that lists 20 such challenges, lists the promises they hold, when they are overcome, and describes them in sufficient technical detail. Writing this paper proceeded very favorably – the majority of participants contributed major efforts. At this point the paper counts already 45(!) pages, and is still growing. A tentative deadline for its finalization is end of July. We are aiming at publishing this paper in a high-impact journal.

Scientific Breakthroughs: Single cell data science deals with novel data structures, algorithms and, last but not least, novel statistical approaches. The combination of statistical biases and vastness of data provides it with a wholly new turn in the larger area of computational and statistical genomics. Many ideas were raised during the workshop. Overcoming the corresponding challenges clearly has the potential to lead to scientific breakthroughs. In any case, significant advances will be inevitable. While new inspiration is to be expected also for computer science, real breakthroughs, in terms of scientific relevance, are to be expected in particular in biomedical research, which depends on well-arranged "data infrastructure".

"**Aha" Moments:** We realized already before the workshop that the interest would be enormous. We ended up just collecting requests for participation, and approving them, rather than sending out additional invitations -- due to the enormous attention, there was no need for that. Several people from overseas paid on their own behalf to cover travel and accommodation expenses. Several leading researchers from Harvard or Cambridge, for example, decisively contributed. Everyone was enthusiastic and passionate to contribute. Overall, the "Aha" moments were the enormous publicity -- without that we needed to make particular efforts in that respect -- and the decided commitment of the participants to contribute to the workshop paper.

Format of the workshop: The "Lorentz format" was just absolutely perfect for this workshop – there could not have been a better platform for this. Ample room for discussions and work/writing sessions, also in combination with lots of free time to reflect and interact with others on a more private basis were key to success for this workshop!

Johannes Köster (Amsterdam, The Netherlands)
Alice McHardy (Braunschweig, Germany)
Ben Raphael (Princeton, USA)
Alexander Schönhuth (Amsterdam, The Netherlands)
Sohrab Shah (Vancouver, Canada)