Strategic plan for Cancer Systems Biomedicine at the Institute for Cancer Research, Oslo University Hospital – Report of the working group

Working group from the Institute for Cancer Research (ICR), Oslo University Hospital (OUH):

Tero Aittokallio, Kjetil Boye, Kjersti Flatmark, Eivind Hovig, June Myklebust, Sigve Nakken, Rolf Skotheim, Leonardo A. Meza-Zepeda (leader) and Ragnhild A. Lothe (co-lead).

The group received advice from Arnoldo Frigressi, leader of Oslo Centre for Biostatistics & Epidemiology, which was partly taken into account and helped formulate the current report.

The five-point mandate is in short dealt with below.

Cancer systems biomedicine is here defined as computational modelling of networks and processes within the cell, tissue, organ and individual. In particular relevant are multi-omics technologies and functional approaches applied to model systems and patient samples. At the level of cohorts, machine learning modelling of multilevel data is expected to support the next generation of cancer medicine.

The response of the working group to each of the mandate points

We need to strengthen the competence at the Institute for Cancer Research (ICR) in advanced bioinformatics, biostatistics and machine learning, as well as in pharmacogenomics, and for the period of the next three years we suggest the following.

- What needs (methods, scaling) do you think the research at the Institute for Cancer Research and associated projects will have for bioinformatics, computational biology, biostatistics, and systems biology/systems biomedicine in the future?

- Data science expertise should be prioritized in the recruitment of postdocs and young investigators to the ICR. Appropriate agreements with relevant competence hubs at the University of Oslo (UiO) should be made to facilitate recruited researchers to be integrated into these environments and facilitate the use of stronger methods in biostatistics and machine learning methodologies. The Faculty of Medicine and Faculty of Mathematics and Natural Sciences at UiO have several relevant competence hubs, including the Centre for Bioinformatics, dScience, Simula, and Oslo Centre for Biostatistics & Epidemiology (OCBE).
- Machine learning with multi-modal large-scale data integration (e.g. molecular, clinical and imaging) is a research area of increasing interest, and it is expected to be implemented in diagnostics and clinics in the near future. Digital pathology-based approaches for quantitative and spatial assessments, based on both imaging and machine learning, is also part of this context, with the potential of combining the outer visual and the inner molecular layers into more holistic models. The ICR, and other units within OUH, with complementary competencies and well-established multidisciplinary networks, are well-situated to take on this challenge.

- How do you maintain and further develop a strong professional environment that is above the critical mass, produces front-line research, develops new methods, and trains expertise?

- Data science should be strengthened at ICR in the existing research groups, and the ICR should use the group leader position soon to be available to establish a group in data science within an existing department to ensure close proximity to biomedical know-how. The position should include a start-up package to attract the best candidates.

- Core facilities should be strengthened to support cancer systems biomedicine. This can be achieved by shared positions of core facility and research personnel, and increased interaction with research and diagnostic environments to provide competence and relevant services.
- New technologies and analytical tools for data processing and more advanced downstream analyses are often developed in the research groups. Once shown to be useful and reproducible, and through interaction with core facilities, these could be implemented as new services.
- ICR should further develop the links with ELIXIR Norway, and work towards establishing a node of this network at ICR with a specific focus towards precision cancer medicine.
- We agree with the SAB's suggestion that data science at ICR will benefit from cross-group and department integration. We propose to launch a grand challenge research project across departments to integrate and strengthen different experimental and computational research levels and support collaboration on a joint project. The theme and leader should be decided based on an open internal call with some budget. By creating a common flagship project, we will also aid in the training and development of young experimental and computational scientists. Such a common project could also be the basis for further development towards a hospital priority area and form the basis for external grant applications.

- What is the best organization of such a professional environment for research, development, and service?

- Today, most of the bioinformatics and computational biology/medicine activities at the ICR are integrated within research groups. These environments with in-depth knowledge of biological systems and specific malignancies, and together with advanced data analysis, strengthen translational research due to its synergistic know-how and experience. This organisation also fosters the dynamic growth of multidisciplinary research and teams and is recommended to be continued and strengthened.
- In the short term, we recommend the establishment of a group in data science within the existing departments. This can be handled in parallel with the retirement of Eivind Hovig, who pursues research including both wet lab and bioinformatics.
- The ICR has today multiple core facilities organised under its own Department of Core Facilities with equal access for all users. All core facilities are integrated as part of strong research environments which improves services and facilitates the development of novel service offers. To further build competence at the Genomics and the Bioinformatics core facilities, the service responsible should have protected time (20%) to participate in active relevant research within the ICR with the aim of developing new service offers in demand. Active research in bioinformatics will primarily take place within research environments at ICR, and the interaction of core facilities with these environments will help develop the next generation of core facility services. The scientific responsibility at core facilities may be shared between 1-2 senior scientists that can provide wider competence, advice/contribution in grant applications and participate in service strategy. The current model has been a success to professionalise and increase the quality of services. To further build on this success, an additional long-term staff member in genomics/bioinformatics with a competitive salary should be a priority.
- The competence present at ICR core facilities and research environments often pilots new diagnostics and we suggest that dedicated funds from diagnostics environments and/or health authorities should support the R&D of novel diagnostics assays at the ICR, as well as the implementation process for standardisation and routine diagnostics.

- How does such a professional environment best interact with other activities in OUS and at UiO?

- ICR should start a dialogue with the relevant environments at UiO (some listed above) to continue and secure part-time positions at the UiO and OUH for senior researchers and group leaders within data science.
- ICR should not duplicate large-scale activities, neither technologies nor data analyses competencies already established at a scale in the hospital or UiO, with exception of cancer critical methods and analytical approaches (for example somatic DNA and RNA sequencing, where analyses and interpretation depend on competence in cancer biomedicine).

- Describe the needs for institutional data handling, storage, and access for data to be "findable, traceable and reusable".

- To develop cancer systems biomedicine, OUH and ICR must provide an infrastructure that allows the storage, handling, sharing and analysis of large amounts of research and clinical data. This infrastructure does not exist today, and the research portal (*forskningsportalen*) initiative led by HSØ does not provide basic functionalities to support this work. The IT infrastructure must be developed to support research projects of different sizes containing sensitive and non-sensitive data. Furthermore, the infrastructure should support the secure and structured storage of data and metadata, as well as the hosting of standardised analysis pipelines to support reproducible research, as well as data sharing across ICR. Standardisation of genomics and clinical data should be a focus to facilitate data integration. This is critical for grand challenge projects and other research that will involve several research groups and departments. Today, large projects depend on external services from the University of Oslo, like Services for sensitive data (TSD), or other ad-hoc or commercial services, which is not a sustainable solution.
- We agree with the SAB regarding the importance of a "data management officer" to ensure better standardisation, digitalization and quality control of data generated at ICR. Such a role will also secure compliance with FAIR principles. This activity can be hosted by the Bioinformatics Core facility in collaboration with current data infrastructure efforts at ELIXIR. Specific financing should be allocated to this activity by OUH, and we suggest that overheads paid by the research project can be used to cover this important activity.
- At a national level, the establishment of a national genome centre is under development, which will provide the infrastructure for the storage, sharing and use of genetic information for health care, quality assurance and research. This should be further investigated as to how and when this infrastructure may fulfil ICR needs. However, as computational infrastructure for big data storage and analyses is urgently needed at ICR, the development of its own infrastructure must not rely on this external opportunity, due to yet uncertain content, capacity, access, and timescale.
- The ICR has over the years generated a large number of high-quality multi-omics and functional datasets characterising large clinical cohorts and preclinical models. This wealth of information should become more visible through the ICR website. A dedicated section on our home page should include information on clinical cohorts, models and data generated by the ICR, to attract collaborations and sponsors, and increase the visibility of ICR. Using existing frameworks such as the cBioPortal, this will allow the comprehensive exploration of large cancer genomic data sets generated at ICR. This will further facilitate the reuse of data after publication by internal and external researchers, and prior to publication upon agreement. The development of such a platform can utilize existing federated EGA solutions, with support from ELIXIR Norway for a sustainable data repository that meets the FAIR and other data-sharing standards.