DR XOSÉ M FERNÁNDEZ
Chief Data Officer
Curie Institute, Paris – France

Harnessing the Complexity of Cancer

The increasing burden due to cancer poses a threat to human societies following the adoption of cancer-associated lifestyle choices including smoking, physical inactivity, and ‘‘westernised’’ diets alongside growing aging populations.

Cancer is not a homogeneous disease, the term is misleading into thinking of a unique ailment, but we face a complex disease that mutates and changes rapidly before we can claim significant breakthroughs. However, remarkable progress were made in developing a suitable toolset to study cancer. We have access to scientific technologies and computational resources unthinkable for earlier generations, including gene editing and artificial intelligence. Technological innovations have been the fundamental drivers of economic growth for more than 250 years. Yet, failure rates at each stage of the “bench to bedside” development pathway are far too high.

TOOLS FOR THE FUTURE TODAY

We are living a digital revolution driven not only by the abundance of data, but also by our capability to collect, store and analyse the information (the vital principle running our world, transforming every branch of knowledge). Within this context, we are deploying an infrastructure to harness the data deluge at the Institut Curie. Each of our patients are systematically screened for a myriad of molecular information at the clinic, generating terabytes of data per patient from which decisions must be taken regarding the best therapeutic options available. Integrating such data requires computational methods that require bespoke procedures, in particular in those cases when useful domain knowledge is already available.

When patients are treated, their care is informed not only by their own health data – their medical history, test results, imaging and so on – but also by the health data of thousands of other people. Besides, precision medicine relies upon comprehensive data on patient treatment and outcomes, both for analysis leading to improved models that provide the basis for enhanced treatment, and for direct use in clinical decision-making. In fact, it is data from previous patients that will probably play the biggest role in making a patient well again, for it gives our treating teams the essential insights and knowledge on which to base their care. Data warehouses combine and standardise multiple databases, which owing to the sensitive nature of the data stored are understandably constrained by many legal and ethical considerations. A clear example is **ConSoRe** developed within the Unicancer excellence network and a cornerstone of our data strategy.

ConSoRe annotates and normalises our medical records relying on key medical references to structure all the patient files (+3000 new documents are added to our medical records on a daily basis). The system can also infer the patient disease history (with new modules being developed with machine learning approaches improving inference in ConSoRe, what should allow us to better assess the quality of the underlying data).

Innovative technologies are only widely adopted in medicine when they show genuine clinical utility. Alongside addressing an important clinical problem, its deployment must also result in significantly improved outcomes for patients, it is not too complex or resource intensive to implement and use, and should have the potential for widespread adoption and diffusion. This is why we are continuously benchmarking our progress with state-of-the-art approaches. Here, **ESME** (Unicancer's programme gathering from more than 30,000 patients) can be used as a baseline to assess ConSoRe when compared to a team of manual curators.

Essential in cancer research is to relate a detailed and unique description of a cancer to useful properties such as response to therapy or risk of relapse. We are currently investigating the potential of representation learning for cancer genomics, exploiting the hierarchical and multi-scale nature of the data available from altered cancer genomes (not always a whole cancer was sequenced) at the molecular tumour board. Focusing on data representation, storage, and analysis, we are building a digital ecosystem integrating new and existing technologies and data.

Dr Xosé M Fernándezwill be speaking about using big data to harness the complexity of cancer at 11:30am on Friday 12 October 2018 during day 3 of the World Hospital Congress.