PAN PROSTATE GENOMICS CONSORTIUM
October 2016

Lay Description
The PAN PROSTATE GENOMICS CONSORTIUM will collect and interrogate Whole Genome DNA Sequence and exome data generated around the world from over 1800 men with prostate cancer, including men from different clinical categories, and ethnicities.

What is the idea
THE PAN PROSTATE GENOMICS CONSORTIUM: Many groups around the world have generated Whole Genome DNA Sequence (WGS) data. To co-ordinate this work Professor Eeles held the first Pan Prostate Cancer workshop in London on 3 October 2014. The attending groups unanimously supported the idea that the accumulated data, should be collected and compared in a common format including common storage, re-analysis through a single pipeline, and investigation to achieve a variety of scientific goals. The data collected so far includes WGS and exomes from 1400 tumours and matching bloods, but this number is likely to expand considerably. The combined collection would include the following categories: (i) cancers from different ethnic groups: eg Caucasian, Asian, Black Caribbean; (ii) cancer from different stages of progression from normal, to organ confined disease, to metastases; (iii) early onset prostate cancer; (iv) prostate cancer from aggressive and indolent disease; and (v) prostate cancer patients managed by different treatments with information linked to detailed and ongoing clinical follow up data.

Rationale
This project is about providing breakthrough advances through analysis of a very large series of Whole Genome DNA data from prostate cancer contributed by many of the leading scientists and clinicians working in prostate cancer genomics. Key papers have already been published from the participating groups including papers in Nature, Lancet Oncology, Science, Elife and Nature Genetics. Many other publications have been submitted or are in preparation. Several members of this group are already participating in the Pan Cancer Analysis of Whole Genome Project (see below for details). Both in terms of its clinical behavior and in terms of the genetics within each cancer, prostate cancer is highly heterogeneous. There is therefore now a global need to carry out integrated large-scale analysis of all or most of the whole genome data from prostate cancer available worldwide.
Critically this project will provide a global framework for the future analysis of even larger series of genomes as costs of DNA sequencing, storage, and sequence analysis decrease.
In addition to WGS data it should, where available, be possible to add additional layers of data including transcriptome data and methylation data. However the proposed project focuses primarily on the analysis of WGS data.

What does the idea need to be successfully executed?
Our project is modelled on the Pan Cancer Analysis of Whole Genomes (PCAWG) Project that was set up by the International Cancer Genome Consortium (ICGC), and which is co-coordinated by partners of the pan-prostate initiative (namely, Peter Campbell and Jan Korbel). The PCAWG project has collected WGS data from ~2800 cancers representing many cancer types. To execute the Pan Prostate Cancer Project, we would need considerable storage space, computing power and personnel to support it: this is the core component of the study. We plan to use existing analysis pipelines currently under development in PCAWG to reanalyze the prostate cancer data.

Collaboration
Through the analysis of WGS data from over 1800 prostate cancers and matched normal blood samples this study will have unprecedented power to address many key issues relevant to the management of men who have or who are at risk of developing prostate cancer. This would not be achievable within a reasonable time scale by any individual group in isolation.

Synergy and Uniqueness
This project is a direct and natural extension to the ICGC Pan Cancer (PCAWG) Project focusing on the analysis of genomes from 2800 cancers of many different types. The PCAWG project includes 105 non-representative prostate cancer genomes and is underpowered to address several critical problems

relevant to prostate cancer clinical management. There is therefore an urgent need to take the next step and undertake a global project that will integrate all WGS data from prostate cancer.

Partners

There are a number of lead partners contributing to the initiative: Chris Hoven, Royal Melbourne Hospital, Australia; Robert Bristow & Paul Boutros, Princess Margaret Cancer Centre, Toronto, Canada; Steve Bova and Prof. Tapio Visakorpi, Tampere, Finland; Olivier Cussenot, Chirurgien des Hôpitaux chez Hopitaux Universitaires Paris, France; Ros Eeles and Colin Cooper, The Institute of Cancer Research, London. UK; Colin Cooper and Dan Brewer, University of East Anglia, and The Genome Analysis Centre, Norwich; UK, Andy Lynch and Charlie Massie, Cambridge Research Institute, UK; Guido Sauter and Thorsten Schlomm, Hamburg, Germany, Jan Korbel, EMBL Heidelberg, Germany; David Wedge, University of Oxford, UK; Max Loda, Dana Farber Boston, USA; Adam Kibel, Brigham and Women's Hospital Boston, USA; Chris Sander TCGA, USA and Chris Foster, HCA London, UK. These centres have generated and analyzed Whole Genome Sequence data either alone or in collaboration. All centres now wish to join together to form the Pan Prostate Cancer Group because of the added value and benefit this would have to patients with prostate cancer.

Process

The structure of the Pan Prostate Initiative is modelled on other successful consortia such as the PRACTICAL germline consortium as well as the PCAWG consortium.
There is a steering committee (SC) composed of the PIs of all groups.
There are 3 subcommittees (working groups, WGs) which report to the steering committee (Technical, Clinical and Pathology)

Data Analysis Centres: DACs

Three sites have been chosen to act as mirror sites for data analysis that will store and analyse all of the sequence data generated by the projects. These sites will be in EMBL Heidelberg, Germany, Sanger Centre, Cambridge, UK and VLSCI, Melbourne, Australia.