

Good Science, Honest Science

PRBB RECORD KEEPING AND DATA MANAGEMENT CAMPAIGN

Get it right, keep it clean, make a record

Workshop title: “*In silico* data tsunami: will you survive?”

Wed 27th January 12:00 - 15:00 Room Charles Darwin

Aim of the workshop: To informally discuss current challenges and potential solutions in the data management and traceability of computational data. A series of short case-study presentations by researchers at the different centres of the PRBB will be followed by a session of group work and general discussion on three or four specific challenges when dealing with *in silico* data, such as management of big data, recording, reproducibility and reliability of the results. The workshop aims to be informal and participative, and it should be of interest to all researchers in computational biology.

Chairs: Cedric Notredame (CRG) and Ivo Gut (CNAG-CRG)

Agenda:

12:00 - 12:05 Welcome

Chair: Cedric Notredame (CRG)

12:05 - 13:30 Presentation of short case studies (15' each)

Title: Who's afraid of metadata?

Speaker: Anna Vlasova (Computational Biology of RNA Processing Group, CRG)

Abstract: The high multidimensionality of data in big projects, such as ENCODE and GTEx, is overwhelming to most users. We will show how proper data description and management allows users to navigate in the data space and helps to gain a “big picture” understanding of the data, for example reveal patterns and associations - be it technical, such as “batch effects”, or truly biological - that would otherwise stay hidden in it.

Title: Do computers change their minds?

Speaker: Paolo Di Tommaso (Comparative Bioinformatics Group, CRG)

Abstract: The growing complexity of computational biology data analysis applications is posing a serious challenge on the ability to replicate In-silico experiments. This session will discuss some of the obstacles involved in creating reproducible computational research as well as the best practices to overcome them.

Title: **Doing medicine with computers**

Speaker: Márton Münz (Biomedical Genomics Group, UPF)

Abstract: *What are the specific challenges in bioinformatics when the goal is to develop in-silico tools to be used in the clinical setting? How can we make sure that our data processing pipeline is reliable enough for being applied in medical practice?*

13:30 - 14:30 Pica pica and discussion
Work in groups

14:30 – 15:00 Presentation of group work and general discussion. Wrapping up.
Chair: Ivo Gut, director of CNAG-CRG