Contents

Timetable

Day 1:

08:30 - 09:00 Registration and coffee (Conference Centre)
09:00 - 09:45 Participant Talks (Tennis Court Suite)
09:45 - 10:30 Introductory talk
10:30 - 11:30 Sanger Tour
11:30 - 13:00 Module 1: Introduction to Ensembl
13:00 - 14:00 Lunch
14:00 - 15:30 Module 2: Introduction to UCSC and other browsers
15:30 - 16:00 Coffee
16:00 - 17:30 Module 3: Comparative Genomics

18:00 -19:00 Welcome drinks (Hinxton Hall)
19:00 - 20:00 Dinner

Day 2:

09:00 - 10:30 Module 4: Working with Encode data
10:30 - 11:00 Coffee
11:00 - 12:30 Module 5: More complex genome browsing
12:30 - 13:30 Lunch
13:30 - 15:00 Module 6: Genomic Variation
15:00 - 15:30 Coffee
15:30 - 16:30 Talk: Next Gen Sequencing
16:30 – 17:30 Own research/ask the instructors

19:00 – 20:00 Dinner (Hinxton Hall)

Day 3:

09:00 - 10:30 Module 7: Variation and Disease
10:30 - 11:00 Coffee
11:00 - 12:30 Module 8: Proteins, Complexes and Pathways
12:30 - 13:30 Lunch
13:30 - 15:00 Module 9: ncRNA
15:00 - 15:30 Coffee
15:30 - 16:30 Talk: Metagenomics
16:30 - 17:00 Own research / ask the instructors
End of Workshop

Biographies of Instructors

Matthew Clark: Since July 2013 I have been running the Plant and Microbial Genomics group at The Genome Analysis Centre (TGAC), a BBSRC funded institute on the Norwich Research Park campus in the UK, where my team uses genomics technology to study for example plant:pathogen interactions. Between December 2010 and July 2013 I was the Sequencing Technology Development team leader, focusing on testing and developing new sequencing technologies.

Before joining TGAC, I worked at the Wellcome Trust Sanger Centre (2003-10) using gene expression profiling to analyse zebrafish mutants, and designing and executing a novel genetic mapping strategy combining next generation sequencing and SNP array technologies. The new SNP map underpins the published zebrafish genome sequence (Nature 2013). I received my PhD from the Max Planck Institute for Molecular Genetics in Berlin, for work identifying Zebrafish genes and their expression by constructing arrayed cDNA libraries of different developmental stages and analyzing them using array technolgies and, in collaboration with WashU, EST sequencing.

Rob Finn leads the Protein Families team at the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI). This team is responsible for two of the most widely used resources for annotating proteins – InterPro and Pfam. The Protein Famililies team is also responsible for other important databases such as MEROPS and Rfam (RNA families). Rob also has a keen research interest in the understanding of environmental community structures, and is responsible for the EBI metagenomics portal.

Prior to joining EMBL-EBI, Rob worked at Janelia Research Campus in the US, where he led a group that designed fast, web-based protein sequence homology searches - the HMMER webservers. Between 2001-2010, he was the project leader for Pfam at the Wellcome Trust Sanger Institute in the UK. Rob's academic background is in microbiology and he holds a PhD in biochemistry from Imperial College, London, UK.

Arox Kamng'Onga: College of Medicine, Malawi

Currently, I am involved in two projects: (i) Investigating the ability of pneumococcal serotype 6B variants to colonise the human nasopharynx and cause invasive disease, using mouse models and (ii) Investigating the association between the Vaginal microbiome and Neonatal Sepsis. The human host-pathogen interaction has resulted in evolutionary pressures that have tailored the response of species or individuals to pathogens and has been associated with changing patterns of immunity. Understanding genomes and identifying sites of genetic conflict, may lead to a better understanding of disease susceptibility patterns in individuals and to the development of effective disease control measures.

Jane Loveland has a background in virus research, plant biochemisty and molecular biology. Whilst working in the photosynthesis group at IACR Rothamsted she completed her PhD (Lancaster University), investigating the biochemistry and molecular biological aspects of inhibitors of Rubisco (CA1Pase). In 2000 Jane left bench science and moved into bioinformatics, working as a computational molecular biology specialist for the BBSRC, providing teaching and support for molecular biology software packages for 8 BBSRC Institutes. Jane joined the Sanger Institute in 2002 as part of the HAVANA group working on manual genome sequence annotation of human, mouse and zebrafish and latterly the swine and rat genomes. Jane is part of the HAVANA management team and has subsequently been involved in a number of major international collaborations, including generating the GENCODE annotation for the ENCODE project, collaborating with NCBI /UCSC and Ensembl to produce the CCDS annotation and as a co-ordinator for the Immune Response Annotation Group for the swine community. Jane has been the co-ordinator of the Open Door Workshops since 2002, and presents human, mouse and zebrafish workshops around the world.

Bert Overduin: I have a M.Sc. degree in Biology from Leiden University, The Netherlands, and a Ph.D. from the VU University Amsterdam, where I worked on the isolation of a fungal disease resistance gene from tomato by means of transposon tagging. After a postdoc at the University of California Davis, which focused on disease-associated programmed cell death (apoptosis) in plants, I moved back to the Netherlands and worked for six years as an IT consultant for Capgemini, mostly in the areas of web development, functional application management and business acceptance testing. The past nine years I have been working at the EMBL - European Bioinformatics Institute (EBI) in Cambridge, United Kingdom, as a member of the Ensembl Helpdesk and Outreach team. During this time I have given over 300 bioinformatics workshops in 26 countries on five continents, reaching more than 6000 people. Since 1 May 2014 I have been working as a Training and Outreach Bioinformatician at Edinburgh Genomics, (http://genomics.ed.ac.uk), the high throughput genomics facility of The University of Edinburgh.

Emily Perry: I am the Ensembl Outreach Project Leader: I am responsible for a small team that teaches workshops, creates online training materials, answers helpdesk queries, usability tests new features and manages social media and promotion, as well as taking part in strategic decision making for the Ensembl project. Before working at the EBI, I did my PhD in molecular biology at the MRC Human Genetics Unit in Edinburgh, then worked for the University of Edinburgh's SCI-FUN group, touring Scottish secondary schools with an interactive science roadshow. I have been at the EBI since September 2012 as an Ensembl Outreach Officer, and was promoted to Outreach Project Leader in March 2015.

Glossary of Abbreviations and Terms

BAC	Bacterial artificial chromosome, a large insert bacterial clone used for mapping
BLAST	Basic local alignment search tool.
Build	The version of the human genome assembly.
cDNA	complementary DNA, stretch of DNA that has been reverse transcribed from RNA.
Contig	A stretch of contiguous (continuous) something. Confusingly, used in many different contexts. Most often, it means a stretch of continous DNA sequence. But can also be used to mean a continuous assembly of fragments, or of clones, without implying that their sequence is known
EBI	European Bioinformatics Institute, Hinxton. An outstation of the European Molecular biology Laboratory.
ECR	Evolutionary Conserved Region
EMBL	European Molecular Biology Laboratory, the name of the European DNA database
EST	Expressed sequence tag *(see appendix 2 for more details)*
mRNA	messenger RNA, processed RNA molecule to be translated to form protein
NCBI	National Centre for Biotechnology Information. Part of the U.S. National Library of Medicine (NLM), National Institutes of Health (NIH).
PAC	P1 artificial chromosomes, a large insert bacterial clone used for mapping
PFAM	Protein family, a searchable database of protein domains
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
STR	Simple tandem repeats
STS	Sequence tagged site
TIGR	The Institute of Genome Research
TrEMBL	Translated EMBL