

# ***Setting Priorities in the CRI Bioinformatics Core***

Matt Eldridge

Head of CRI Bioinformatics Core

# CRI Bioinformatics Core

---

## CRUK Cambridge Research Institute

- One of 5 core-funded Cancer Research UK research institutes
- 19 research groups, 7 core facilities
  - Basic research in cancer biology
  - Clinical research on specific cancers
  - Population-based studies in screening and prevention

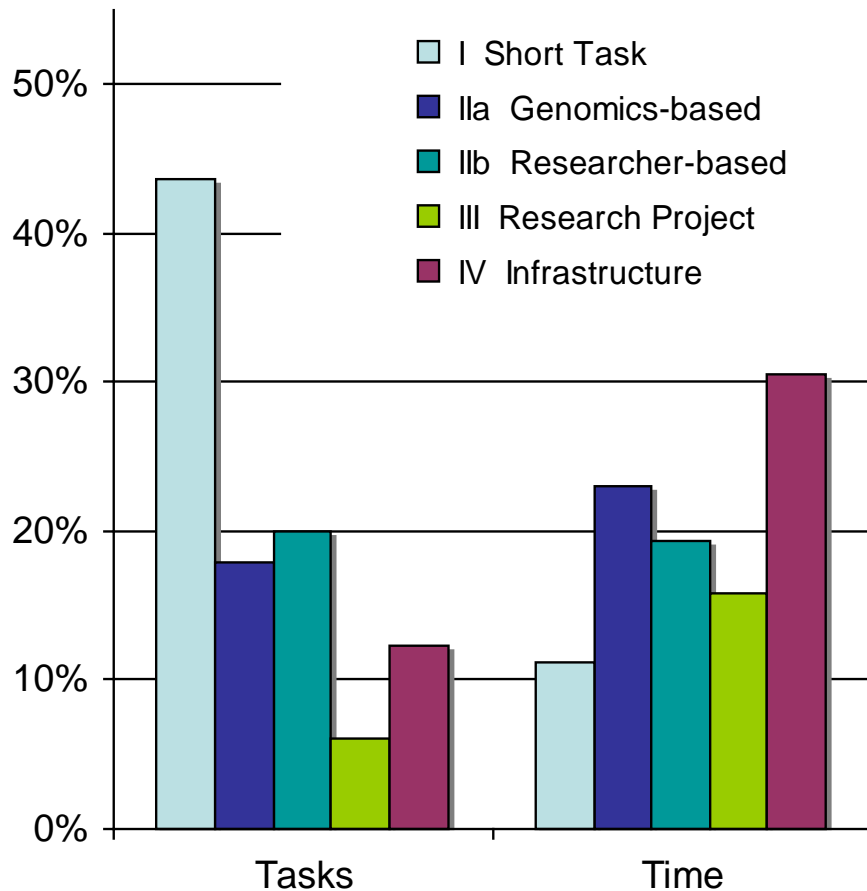
## Bioinformatics Core

- Primary focus on high-throughput genomics platforms
  - **Microarrays** – Illumina & Agilent
  - **High-throughput sequencing** – 3 Illumina GALLs
- 8 in team
  - 1 statistician, 3 microarray analysts, 2 HTS specialists, 2 software/db dev
- No cross-charging
- Monitor usage by research group/project

# CRI Bioinformatics Core Activities

- Statistical support
- Experimental design
- Primary data analysis
  - microarray QC, spatial defect removal
  - Illumina GA pipeline
- Secondary/downstream analysis
  - Differential expression
  - ChIP-seq peak calling
  - Structural variation, genomic rearrangements
  - SNP and CN analysis
  - microRNA profiling
  - GO enrichment, GSEA
- Training
  - motif finding
  - functional/network analysis
  - microarray analysis
- Data management
  - Solexa storage
  - Microarray database
- Software development
  - Bioconductor packages
    - Beadarray
    - Illumina annotation packages
  - Solexa & microarray LIMS
- Bioinformatics tools
  - Ensembl, Galaxy, Cytoscape

# Types of Projects



- **Class I – Short tasks**
  - Typically few hours, well defined output
- **Class IIa – Genomics-based**
  - Analysis of data from Genomics Core
  - Initiated in experimental design meeting
  - Defined output (Sweave report, DE gene list, Cytoscape session)
- **Class IIb – Researcher-based**
  - Follow-on from previous class Iia project or meta-analysis on existing data
  - Initiated by contact from researcher
  - Pre-agreed output and timescale
- **Class III – Research projects**
  - Collaborative, open-ended
  - Poorly defined output
- **Class IV - Infrastructure**

# Workload

---

- Microarrays
  - ~5 projects per month
  - ~30 arrays/samples per project on average
- High-throughput sequencing
  - Primary analysis for 3 GALLs
  - 2 x 50bp SE runs per GALL per week
  - Secondary analysis for 1/3 data
- 6 – 12 projects per person at any given time

# Managing Workload

- Define & refine process
  - Set expectations
  - Define scope/deliverables
- Project/issue tracking system – Redmine
- Deliver data/output in stages
  - e.g. BED/WIG track within 3–5 days, peaks within 1–2 weeks, downstream analysis results later still
- Standardize and automate data analysis pipelines
- Train researchers to carry out downstream analysis tasks for themselves – classroom, individual, wiki
  - Online functional analysis tools – DAVID, GeneTrail, etc.
  - Motif analysis
  - Cytoscape
  - Galaxy for operating on genomic interval/feature data
  - IGB & IGV browsers for data visualization