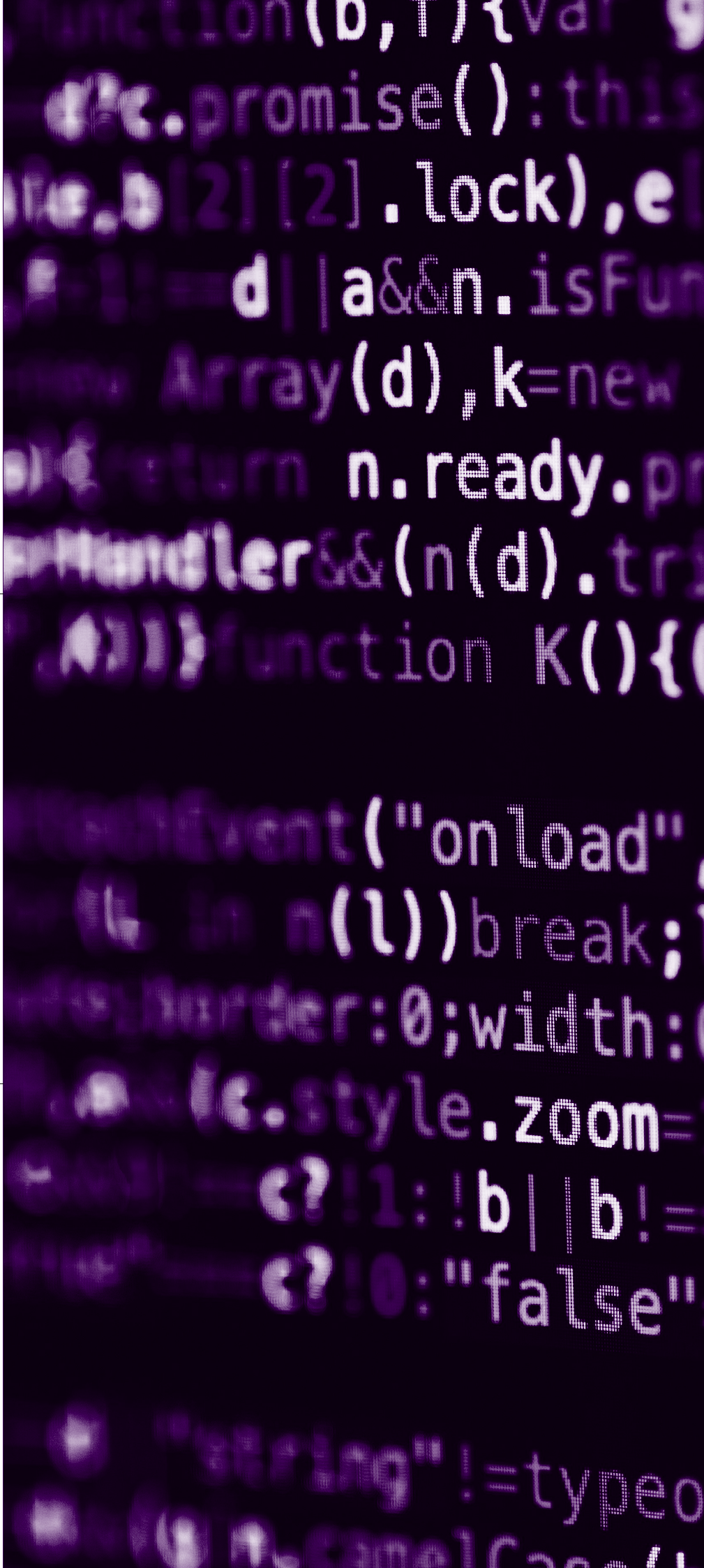


COMPUTATIONAL BIOLOGY FACILITY



CBF.LIVERPOOL.AC.UK
CBF@LIVERPOOL.AC.UK



The CBF





Data Driven Success

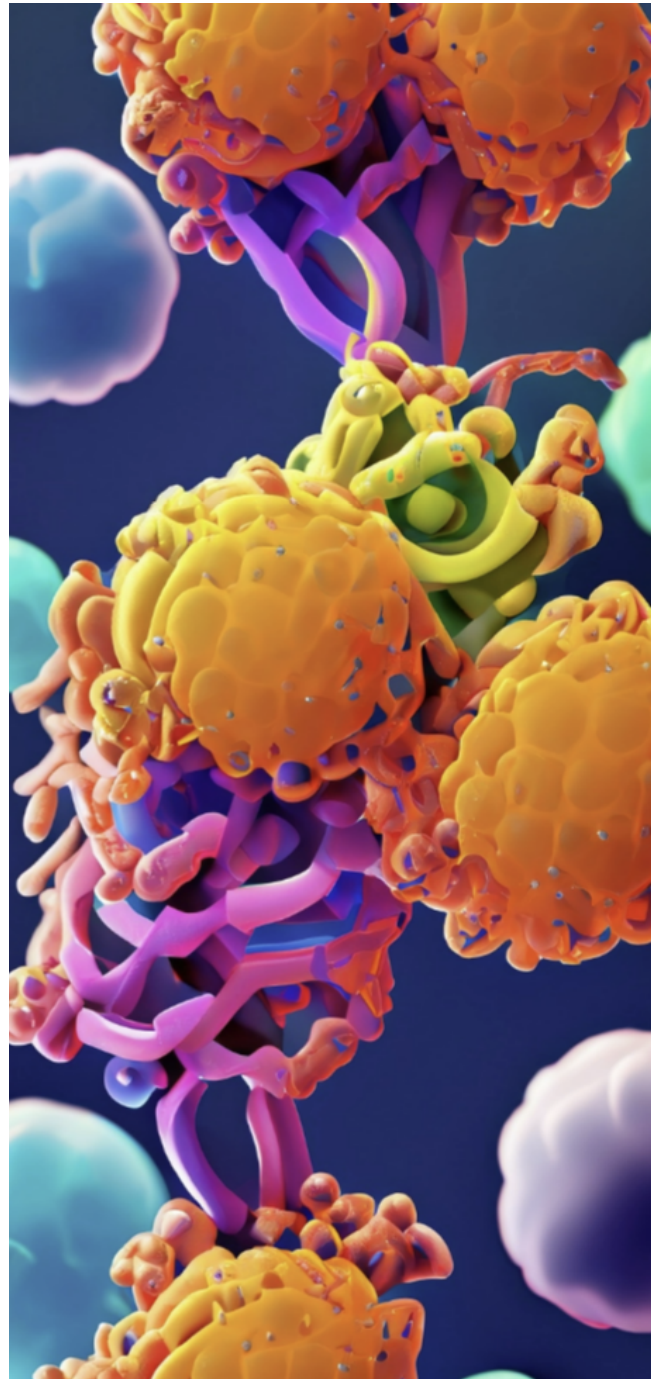
The Computational Biology Facility (CBF) is a shared research facility at the University of Liverpool. Our mission is to unlock the full potential of life sciences and clinical research using computational biology.

We work as scientific partners and as service providers offering tailor-made solutions across a wide range of bioinformatics, statistics and functional interpretation of data.

The CBF is led by sector leading academics and industry experts managing a team of experienced data scientists and software engineers to deliver a wide portfolio of projects. Operating on a cost-recovery basis, we emphasise delivering exceptional value for your investment. Whether you are seeking specialised support or comprehensive expertise, the CBF is your partner in elevating your scientific discoveries.

CONTACT US:

- [CBF.LIVERPOOL.AC.UK](https://www.cbf.liverpool.ac.uk) 
- CBF@LIVERPOOL.AC.UK 
- [@LIVUNICBF](https://www.liverpool.ac.uk/cbf) 
- [CBF LINKEDIN](https://www.linkedin.com/company/cbf-liverpool) 



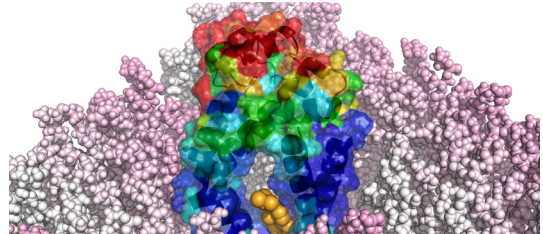
SERVICES



TRAINING & WORKSHOPS

Online, in-person or hybrid.

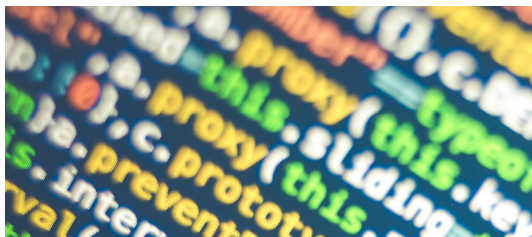
For every level – from beginner to expert.



PROTEIN BIOINFORMATICS

Protein structure modelling, functional inference and re-design.

Utilising deep learning models such as Alphafold2.



SOFTWARE AND DATABASES

Bespoke software solutions.

Database design and implementation.

High performance computing.

Using Large Language Models (LLMs) and AI for research.



DATA ANALYSIS

Omics and multi-omics analyses and integration.

Statistical analysis and modelling.
Machine learning.

Development of custom bioinformatics pipelines.

Pathway analysis and biomarker discovery.

Study design and power analysis.

Support with grant applications.

EXPERIMENTAL DESIGN

EXPERIMENTAL DESIGN IS THE FOUNDATION OF SCIENTIFIC PROGRESS, GUIDING RESEARCHERS TOWARDS MEANINGFUL DISCOVERIES. AT OUR FACILITY, WE UNDERSTAND THE IMPORTANCE OF ROBUST DESIGN AND OFFER TAILORED SUPPORT SERVICES TO PROPEL YOUR PROJECTS FORWARD AT EVERY STEP. TAKE A CLOSER LOOK AT HOW WE CAN ENHANCE YOUR RESEARCH JOURNEY:

POWER CALCULATIONS

- Unlock funding opportunities and ensure project success with tailored power calculations.
- Our expertise in omics technologies ensures precision in optimising sample sizes for statistical analysis, bringing your project closer to achieving its goals.

RANDOMISATION TECHNIQUES

- Minimise bias and enhance the reliability of your results with our rigorous randomisation protocols.
- We'll guide you through implementing strategies to reduce confounding variables and maximise the impact of your research.

BATCH ASSESSMENT AND QUALITY CONTROL

- Safeguard the integrity of your data with our expert advice on batch assessment and mitigation of unwanted effects.
- Ensure the quality and reliability of your results by incorporating suitable quality control samples and standards into your experimental design.

DATA ANALYSIS PLAN

- Develop a clear roadmap for data analysis, key for large trials and consortia applications.
- Receive personalised assistance in selecting appropriate statistical methods and software tools for your unique research goals.

MASTERCLASSES IN EXPERIMENTAL DESIGN

- Empower yourself and your team with our exclusive masterclasses in experimental design considerations.

PARTNER WITH US AND UNLOCK THE FULL POTENTIAL OF YOUR RESEARCH ENDEAVORS.

UNLOCK THE FULL POTENTIAL OF YOUR RESEARCH

BIOMARKER DISCOVERY

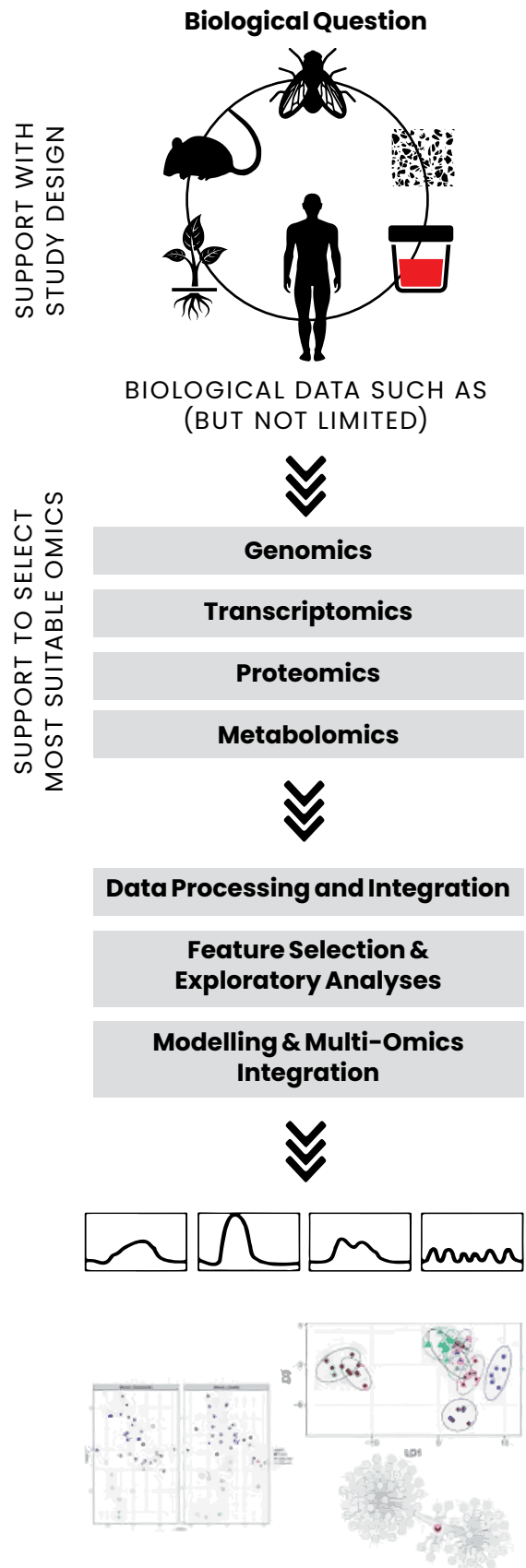
Biomarkers play a critical role in improving diagnostics, drug development, predicting disease states and response to therapy.

Biomarker discovery is challenging and requires a multidisciplinary approach in which clinicians, laboratory-based researchers and data scientists must collaborate.

From classical statistical approaches to novel machine learning pipelines for composite biomarker discovery, we have the expertise.

We can help you make the most of your experimental design and data analyses, or design a way to exploit public data repositories to answer your biological question. We will guide you from experimental design to validation:

- **Study design** – advice on the best biological material to use, most appropriate cohorts, most suitable omics and help with power calculations and experimental design.
- **Data processing** – including data integration strategies, and normalisation and batch correction techniques.
- **Exploratory analyses** – including data visualisation techniques to assess data variance, effect of covariates and sources of variation.
- **Variable selection** – using machine learning, we have expertise building custom state-of-the-art pipelines exploiting algorithms such as random forests and regularised regression to refine large candidate lists.
- **Modelling and multi-omics integration** – integrating different omics datasets and creating predictive models to unveil comprehensive biomarker signatures.
- **Validation** – using model validation, recalibration techniques and statistical analyses to validate biomarkers.

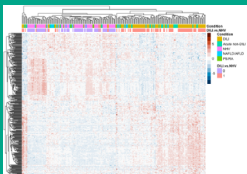


CASE STUDIES

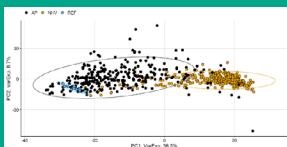
Case study: TransBioLine

We are leading the computational biology aspect of the TransBioLine international consortium, which aims to explore and qualify new biomarkers for drug development and disease diagnostics and overall improvement of drug safety (Schofield, A.L., et al., 2021).

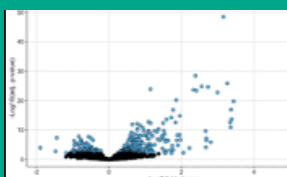
Circulating miRNAs are potential biomarker candidates, with several advantages over other types of biomarkers. Circulating miRNA profiles generate large and complex datasets. We develop computational models that use machine learning to predict drug-induced organ injury and disease, and have supported the consortium validating these models.



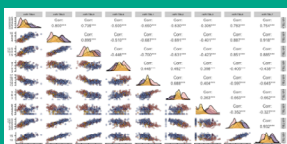
miRNAseq data
pre-processing



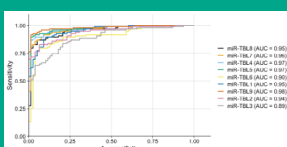
Quality control and
data integration



Differential expression



Multivariate candidate
selection



Validation analyses

Case study: Diagnostic meta-analysis for differentiation of viral and bacterial biomarkers

We partner with the Defence Science and Technology Laboratory (DSTL) on different projects. In Myall A., et al, 2020 & 2021, we used AI to analyse data from various sources to identify biomarkers to differentiate between healthy, bacterial and viral diseases in the blood.

Case studies: Biomarker discovery for human diseases

We have partnerships with healthcare services in the North West region, including Royal Liverpool University Hospital, the Liverpool Women's Hospital, Alder Hey Children's Hospital and the Manchester University NHS Foundation Trust. Some projects include:

- Working with the Experimental Arthritis Treatment Centre for Children to identify markers to improve diagnostics and treatment of conditions such as psoriasis arthritis, juvenile lupus and arthritis, and to better differentiate paediatric autoimmune and Inflammatory disorders (Carlsson E., et al, 2021).
- Investigating women's conditions – such as developing a composite biomarker candidate to predict adverse pregnancy outcomes (Hill C., et al., 2024), markers for conditions such as adenomyosis or fibroids, and improving understanding of recurrent implantation failure (Tempest, N. et al, 2023).
- Improving cancer diagnostics – such as for renal cell carcinoma (Cooley, L.S. et al., 2021) pancreatic cancer (Oldfield, L. et al., 2023), head and neck malignancies and morbidities (Brooker, R., et al. 2021) or breast cancer therapy resistance (Sarmiento-Castro, A. et al, 2020).
- Biomarker discovery for diagnosis and treatment of osteoarthritis (Timur, U.T., et al, 2021; Salerno, A., et al., 2020).

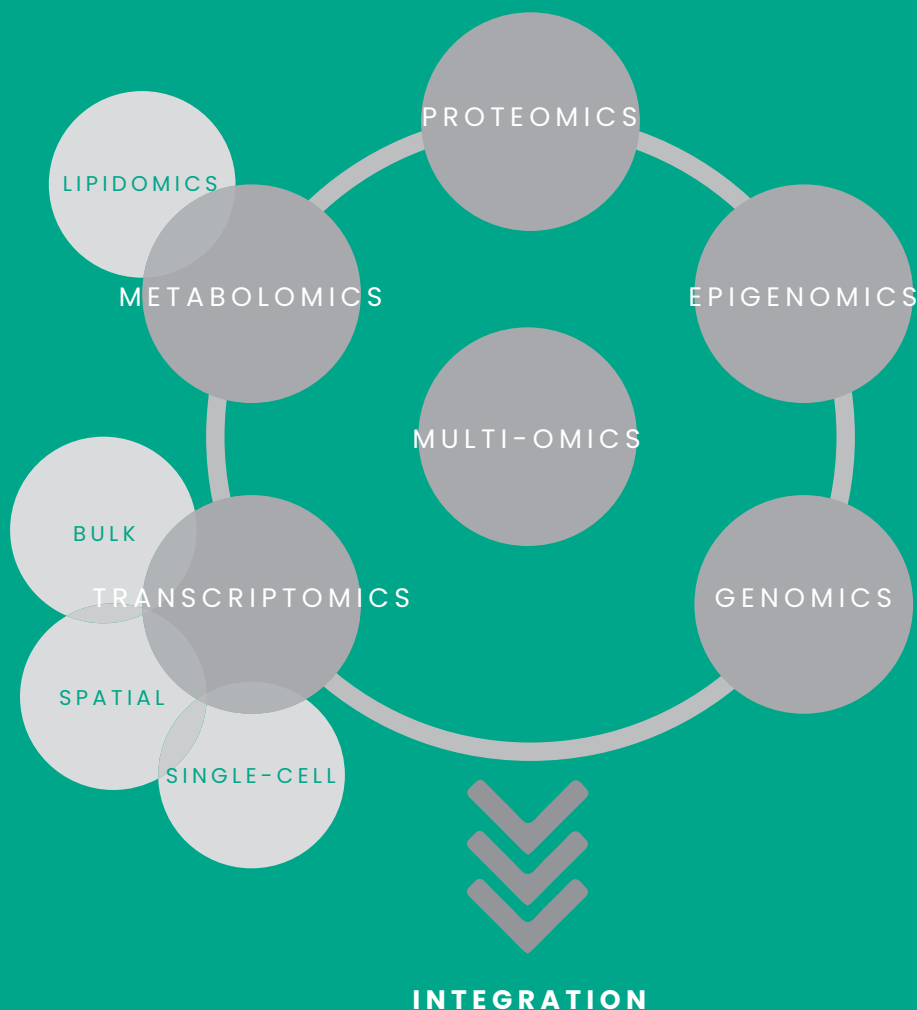
OMICS AND MULTI-OMICS

Analyses and integration

Life sciences are complicated, but we know how to let the data tell the story.

Omics data provides detailed insight into the biological functions of organisms, however it can be challenging to extract this information. We have the expertise to navigate challenges of technical variance, missing values, data normalisations and transformations. Our capabilities include single cell and spatial technologies - regardless of the type of -omics data, technology, platform and study design we can unlock the full potential of your data.

As well as supporting researchers and clinicians, we also push the boundaries of the omics field by developing novel methodologies (Daly, L.A. et al. 2023, del Carratore, F., et al, 2023 ,Davies, H.A., et al, 2021), developing sector-leading standards (Hoffmann, N. et al, 2019), databases (Gonzalez-Galarza, F.F., et al. 2020, Soul, J. et al., 2021) software to improve analyses (Del Carratore, F. et al., 2019, Soul, J. et al., 2019) and visualisations (McCabe et al., 2021).



CASE STUDIES

We can support with the complexities of multi-omics projects, assisting with experimental design, methodology and software choices. The CBF has extensive experience with multi-omics integration, knowing how best to utilise pipelines and methodologies, such as network-based approaches, to get the most out of your data. Some examples of our published multi-omics work includes:

- Time course multi-omics Integration of sncRNA and proteomics in osteoarthritis ([Johnson, E.J., et al., 2023](#), [Clarke, E.J., et al, 2022](#)).
- Integration of DNA methylation and mutational profiles to Improve understanding of mouth malignancies ([Green, et al, 2023](#)).
- Integrating metabolomics and transcriptomics data of microorganisms to improve understanding of antibiotic production ([del Carratore, F. et al, 2021](#)).
- Integration of genomics and transcriptomics to elucidate changes In Insulin sensitivity ([Takeshita, L.Y., et al, 2021](#)).
- Integrating transcriptomics, metabolomics and experimental data to improve adverse outcome pathways in ecotoxicology ([Brockmeier, E. K. et al, 2022](#), [Basili, D. et al, 2018](#)).
- Exploiting co-expression networks with functional genomic data to identify novel regulatory mechanisms controlling metabolic adaptation ([Clarke et al. 2017](#)), and identify novel mechanisms in cancer ([Trevino, V. et al, 2016](#); [Clarke, K. et al., 2015](#)).

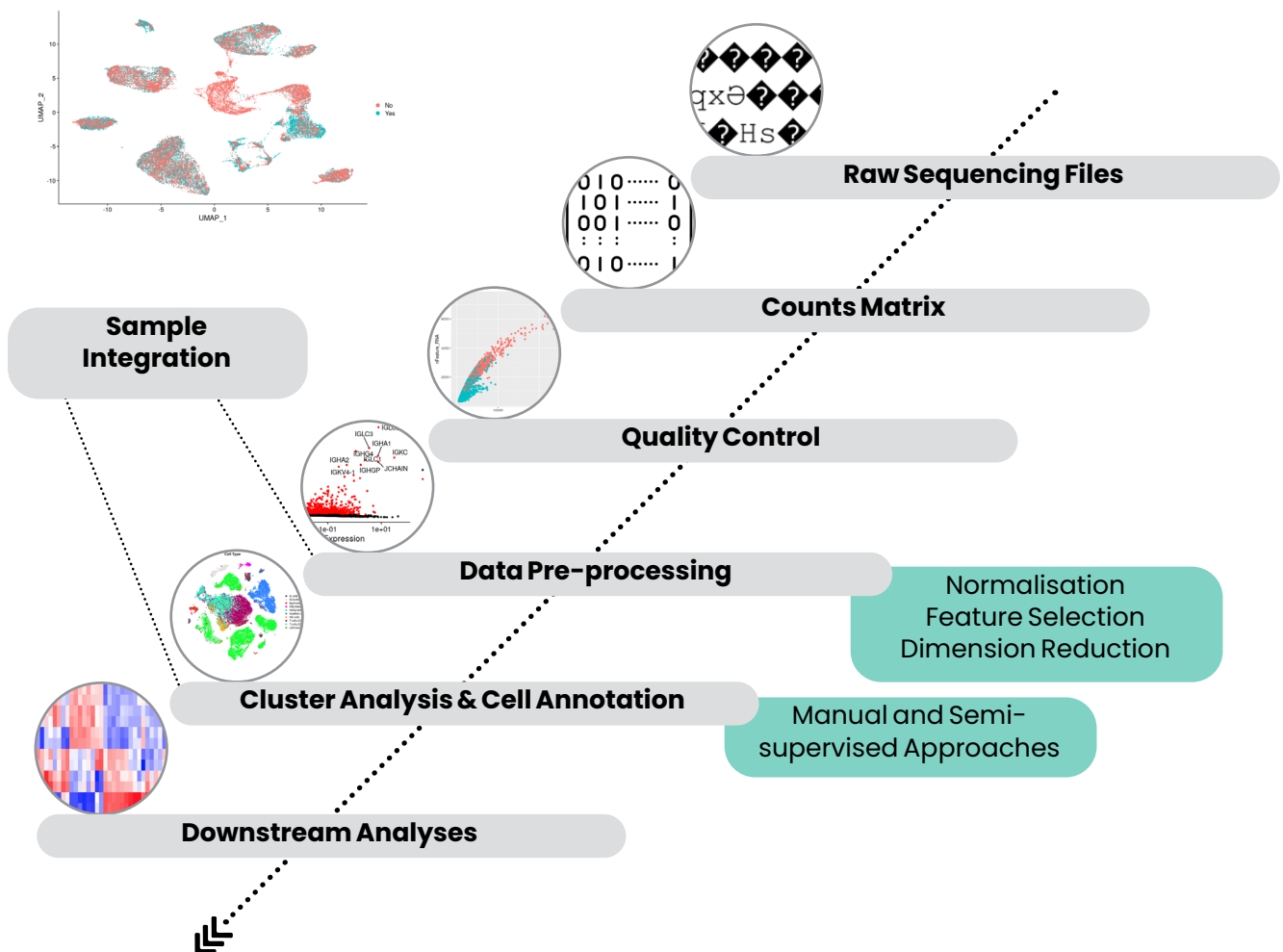
The Vaginal Microbiome and Post-Menopausal Health (VAN), in collaboration with Liverpool Women's Hospital, investigates the interplay between the vaginal microbiome and metabolome in vaginal atrophy. We are evaluating the differences between women who do and do not experience the condition, and aiming to identify metabolites that may be driving it.



SINGLE-CELL SEQUENCING ANALYSES

Single-cell sequencing has emerged as a revolutionary tool in understanding complex tissues and uncovering cellular heterogeneity.

The CBF has the expertise and established pipelines for single cell data that can answer the biological questions you want asking.



Single-Cell Sequencing Course (Launching 2024)

This course will equip participants with the skills and knowledge needed to analyse their own single-cell seq data.

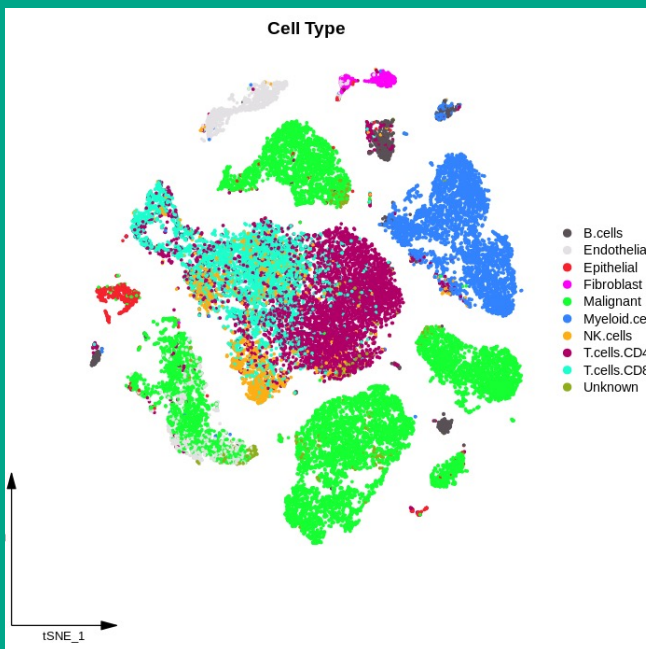


CASE STUDIES

Case Study: Bile Duct Cancer Analysis

The CBF applied our expertise and pipelines to investigate bile duct cancer.

We identified malignant cell types within the tumour micro-environment and their marker genes. Through the characterisation of the malignant cells this work allows for a more detailed understanding of the molecular mechanisms of the cancer compared to previous bulk-seq work.



Case Study: Identification of promising therapeutic approach for patients with metastatic pancreatic cancer

The CBF has undertaken the single-cell sequencing analyses of the recent publication in Nature Genetics by Astuti, Y. et al., 2024. In this team project, we identified that different types of macrophages play a key role in liver metastasis and provides a promising therapeutic target.



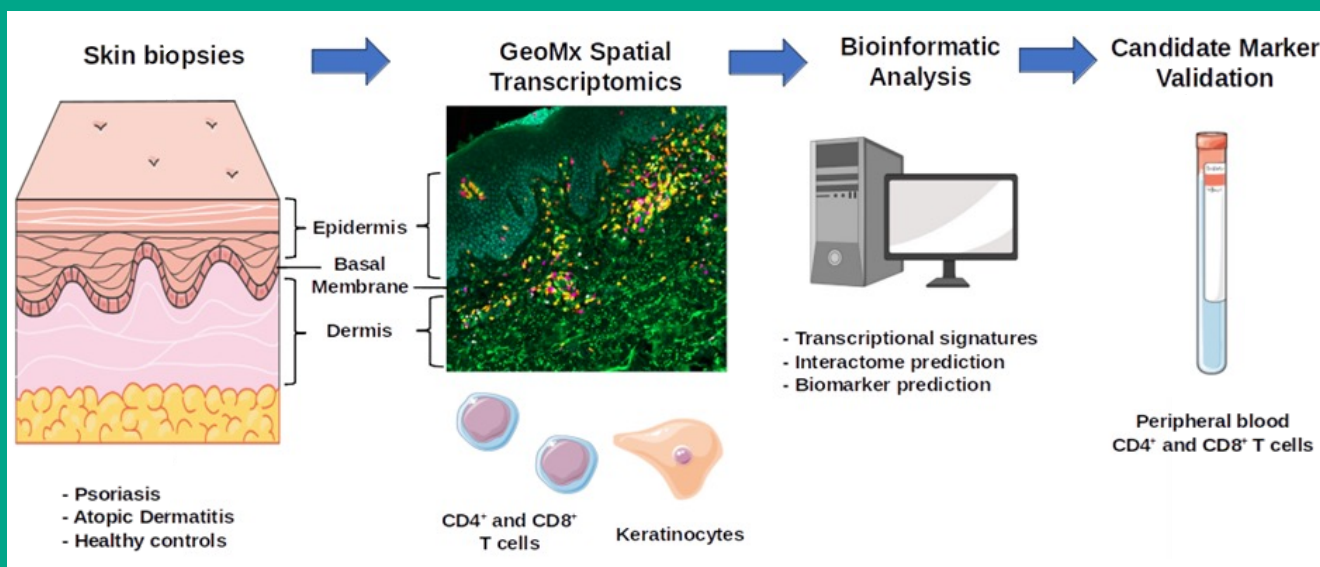
SPATIAL OMICS

Spatial omics allows mapping of the molecular landscape of tissues, providing crucial spatial context to the expression of biomolecules such as transcripts, proteins, and lipids. Analysing spatial omics data is challenging, requiring the need for complex statistical models.

The CBF has extensive experience in analysing a broad range of spatial omics experiments, from traditional mass spectrometry imaging to cutting-edge Nanostring GeoMX and 10X Visum spatial transcriptomics. We have established pipelines for specialised quality control, data normalisation, and differential expression analysis in order to identify altered genes and pathways between cell types and disease states in different tissues. We also offer bespoke analyses catered to each individual project, past examples have included ligand-receptor signalling to determine the crosstalk between cell types and deconvolution algorithms to estimate the proportion of pre-defined cell types in each sample.

Case Study: Identifying inflammation markers in paediatric patients

In a recent CBF collaboration with Alder Hey Children's Hospital, our analysis successfully identified a number of differentially expressed genes in T-cell populations in skin biopsies from patients with inflammatory skin conditions. These findings were subsequently validated in vitro and are about to be published.



Case Study: Applications in women's health

In close collaboration with a large team at the Liverpool Women's Hospital, the CBF has analysed spatial data to elucidate tissue-specific differences in conditions such as adenomyosis, endometriosis and fibroids.

We have also exploited the methodology to advance understanding of recurrent implantation failure (Tempest, N. et al, 2023), the findings which granted the lead clinical author the prestigious Laxmi Baxi Award by the Society for Reproductive Investigation.

DATABASES

To encourage reproducibility, experimental data should comply to the FAIR data principles: Findable, Accessible, Interoperable, and Reusable.

Storing your data in a database allows efficient data sharing, flexible interrogation, and data comparisons.

Behind every web application is a database. Combining the power of a database with a bespoke web application providing real time analysis can provide unique insights, and comparisons.

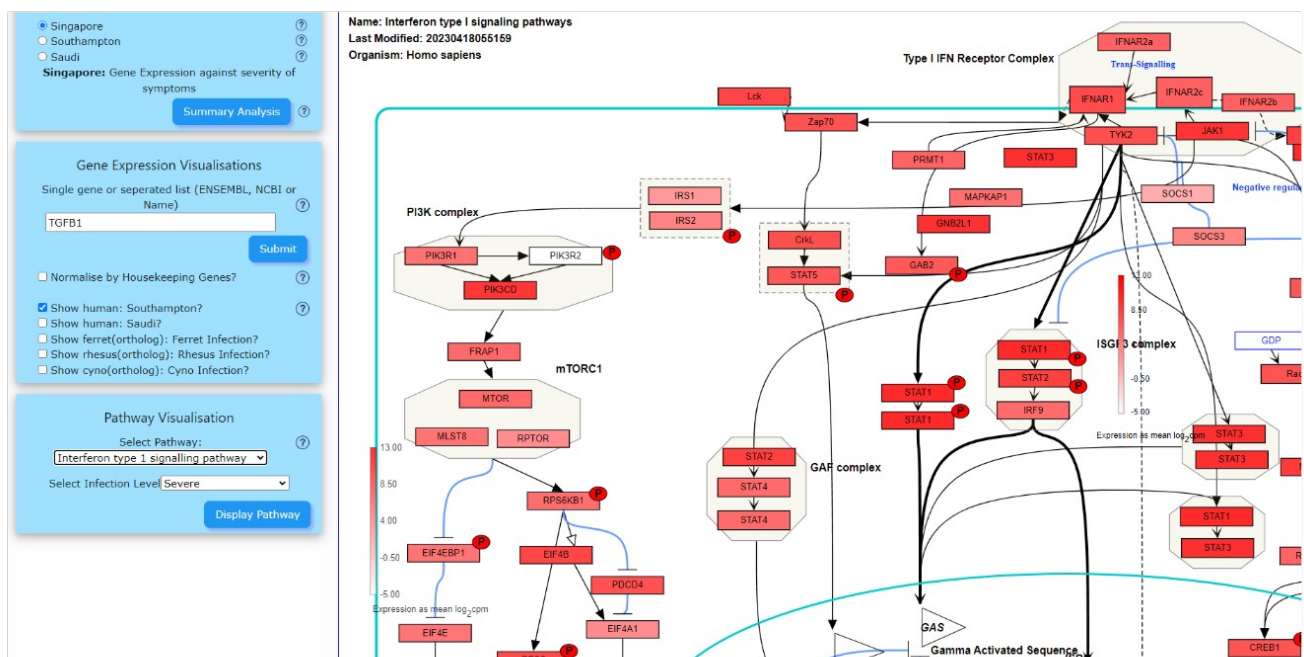


We are able to:

- Design a custom database schema.
- Specify, create and install any required software.
- Build and deploy a website, including regular backups and registration of domain names.
- Maintain the website and software for an extended period.

Case Study: CovVis – joined up COVID-19 analysis

Using RNAseq data from multiple human and animal COVID-19 investigations, we created a data exploration and visualisation platform. Integrating data from experiments from several research groups to increase the accessibility of the data and help determine genes of particular interest and whether animal models reflect the responses observed in humans.

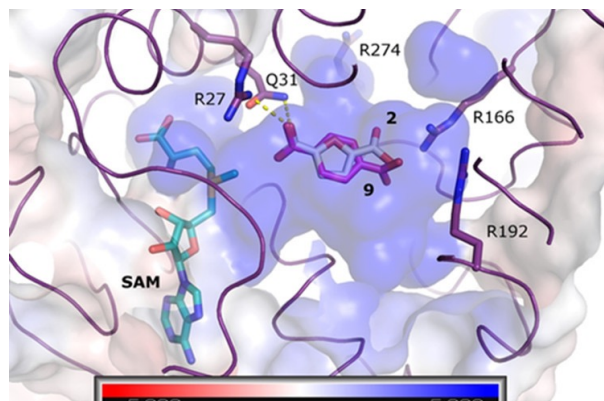


PROTEIN BIOINFORMATICS

The CBF offers a range of services across protein sequence and structure bioinformatics to help predict, understand and redesign protein activities. Our team has expertise in protein modelling of all kinds, using the most cutting edge technologies. These can be used for biomedical and biotechnological applications such as structure-based SNP interpretation, genome mining, protein design, and ligand-receptor binding.

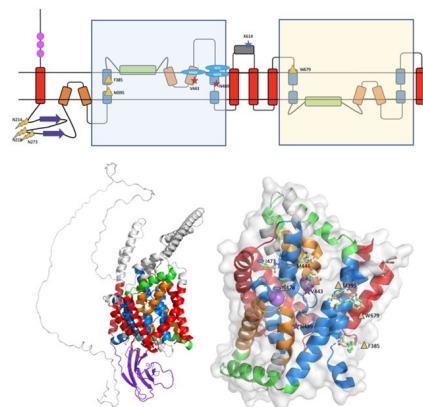
Case Study: Protein redesign for commercially valuable enzyme activities

Synthesis of bioplastics from cellulose waste offers a route to replace current polymers with biodegradable alternatives. In collaboration with the Department of Chemistry, structure-based analyses of AlphaFold 2 models and small molecule docking were used to suggest mutations to reshape the natural catalytic site of a fungal protein, successfully enhancing activity. ([Ward L.C., et al 2023](#))



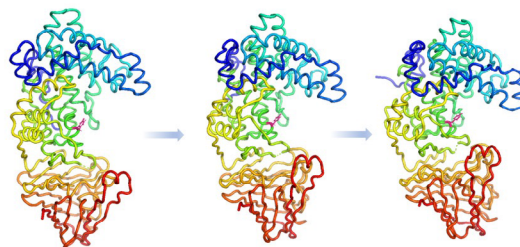
Case Study: Function annotation of a human membrane transporter

Assigning function to cryptic proteins is of major importance. Work by bioinformatic specialists in the CBF used AlphaFold 2 and structure-based function annotation methods to address the Oca2 protein associated with human albinism. It revealed an unsuspected GOLD domain in purple, linked to subcellular transport; overturned the prevailing view of its topology; and enabled mapping of known damaging mutations ([Mesdaghi S., et al. 2023](#)).



Case Study: Determining the dynamics properties of an enzyme

Post-AlphaFold 2, consideration of the dynamic properties of macromolecules is increasingly important as high quality structures become more readily accessible. Molecular Dynamics simulations of a new crystal structure were carried out to facilitate the prediction of how domain motions bring together residues forming the catalytic site.



SOFTWARE

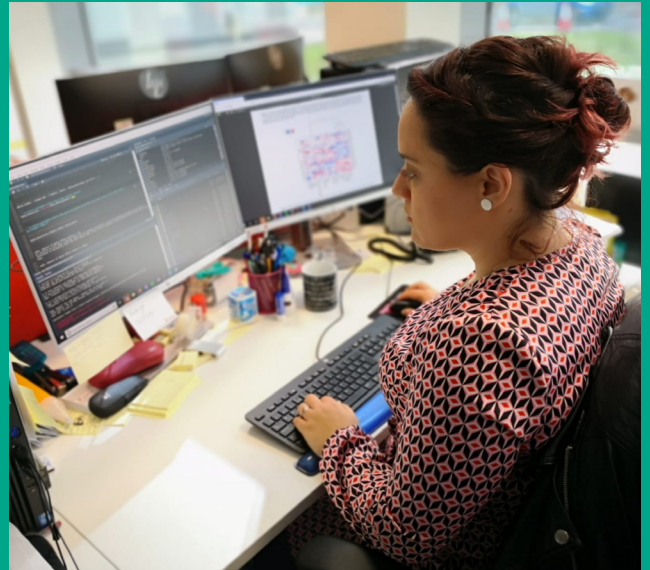
We specialise in developing tailored software solutions that allow user-friendly interaction with tools and pipelines. Whether you're working with existing pipelines or embarking on new developments, our team is well-equipped to assist.

Our expertise extends to the creation of websites, intuitive dashboards, high-performance computing services, locally installable solutions, and mobile applications.

We develop websites, dashboards, locally installable software and mobile applications, often utilising AI, so that you and your collaborators can automate workflows and pipelines. This ensures reproducibility and saves time so you can focus on the tasks you care about.

As well as this, we can develop more powerful tools utilising our high-performance computing resources.

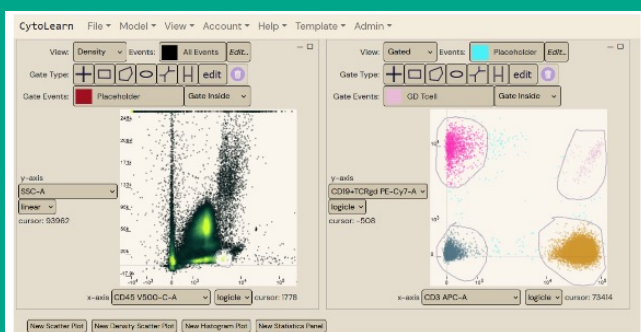
We will work with you to work out the most practical route, whatever your aims are.



Case study: CytoLearn

CytoLearn is web-based flow cytometry analysis software developed in collaboration with the Haemato-Oncology Diagnostic Service (HODS), based at the Royal Liverpool University Hospital.

Users can upload flow cytometry data files to the user-friendly interface, where the CytoLearn algorithms predict how the data should be analysed. If users do not agree with CytoLearn's output, they can interact with the plotted data to update CytoLearn's knowledge, improving future predictions.



Case study: PubLLican

PubLLican is a prototype Artificial Intelligence application to extract structured data from biomedical publications.

Using either open source Large Language Models, such as Meta's Llama, or commercial products such as ChatGPT it extracts the text from a publication pdf file then uses the AI to summarise the results of the paper and to extract, in computer-readable formats, any relevant species, genes, and Gene Ontology terms from the paper.

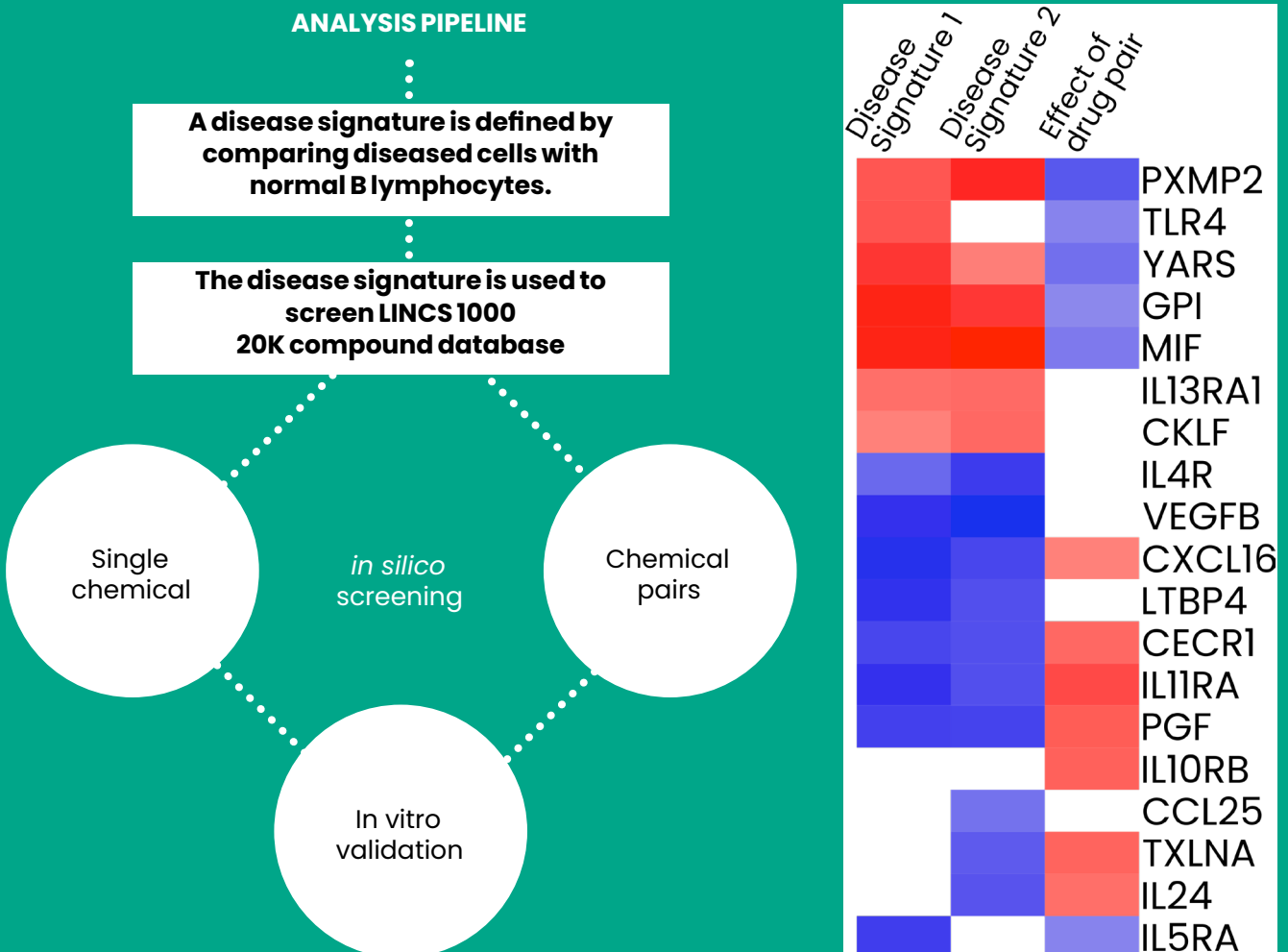
These are cross-checked against our existing database and evaluated by comparing with expert annotations. We have been awarded funding to develop this prototype into a full application.

DRUG REPOSITIONING

Drug repositioning aims to identify new applications for drugs that have already been proven safe and can be commercialised. The screening process can be made more efficient by using *in silico* methods that exploit existing datasets that represent the response of cells to a large collection of drugs.

We have experience exploiting a variety of tools and databases for enquiry. For example, LINCS1000, where data of 20,000 drug treatments representing a wide range of pharmaceuticals and chemicals is accessible.

We have the expertise and resources to match a disease-specific expression signature with databases to search for new drugs as therapeutics. We have applied this technology in collaboration with the Liverpool Women's Hospital, identifying drugs to help in Recurrent Implantation Failure and Adenomyosis.



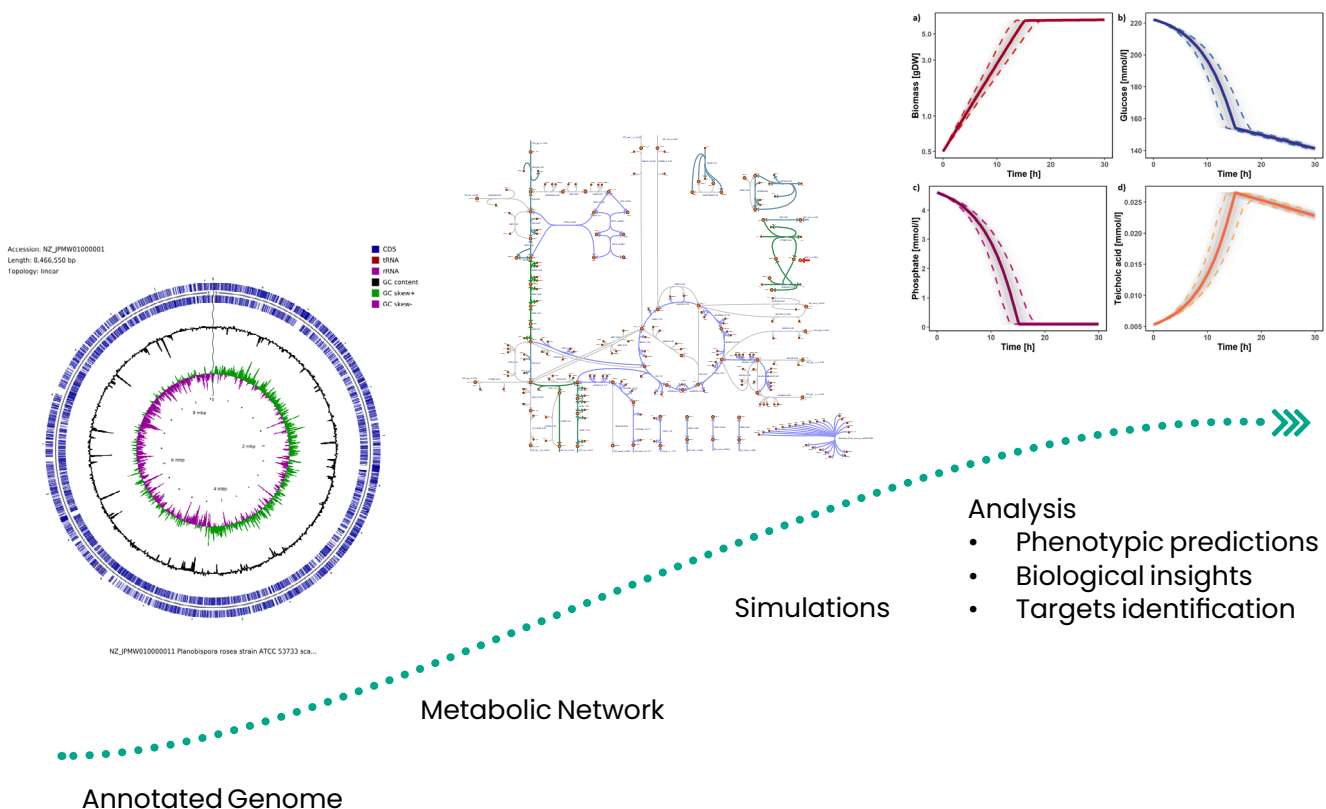
GENOME-SCALE METABOLIC MODELS

Reconstruction and analysis of Genome-scale metabolic models Unlocking the Secrets of Life's Building Blocks

Discovering the inner workings of organisms is like deciphering a complex puzzle. That's where Genome-scale Metabolic Models (GSMMs) come in. These models are like virtual blueprints, connecting genes, enzymes, and metabolites to reveal how living things transform nutrients into growth.

Why does this matter? Because GSMMs have endless possibilities. They decode biological data, pinpoint genetic targets for tweaking, unveil potential drug targets, and even simulate how organisms interact at a molecular level.

At the CBF we have ample experience exploiting GMMs with powerful tools like dynamic Flux Balance Analysis (dFBA). This allows to put the models in motion, showing how metabolism changes in response to changes in the environment or genes and it is a key tool for biotechnological research and bioproduction improvement.



Case study: Unlocking antibiotic production in *Planobispora rosea*

Del Carratore et al. (2021) reconstructed the very first GSMM for *Planobispora rosea* using transcriptomics and metabolomics data for its refinement. Through a dFBA analysis, such a model was used to elucidate the limiting growing factors of the microorganisms, key for biotechnological production and to gain understanding on phosphate regulation and its role in antibiotic production.

COURSES, EVENTS AND OUTREACH

Throughout the year we run courses and events so that you can learn more about performing your own data analysis. We can also provide personalised support tailored for your specific needs.

We have over ten years' experience in the field, we run a series of bespoke courses and events annually, sharing the skills and knowledge from within the CBF. We have a varied portfolio of courses, both run online and in-person. Our courses have a combination of short lectures and hands-on-exercises with targeted 1-to-1 support throughout. Over the years we have trained academic, industrial, clinical and government researchers and are official course provider for the BBSRC DTP3.

Our courses are in constant development to meet ever-advancing research needs. We also run personalised training programmes for small groups or one-to-one sessions and drop-in coding clinics.

Online, hybrid or in-person opportunities

Several day courses

Drop-in sessions and online support groups

Conferences and workshops

Please see our website for the event calendar or contact us with any requests for training courses or personalised training. See below some of the feedback from previous attendees:

"Very practical focused which is great, knowledgeable trainers who are willing to help. Very good curve in the complexity of the content of the course. Handbooks are also explained clearly so I can easily understand each term in R code line. Best programming course I have ever attended, even better than my undergraduate programming modules."

"Took the intermediate course at the beginning of lockdown and I can honestly say I am using R every day now (and having so much fun!)-recommend to anyone who has data to analyse."

"Can't recommend these courses and the support team enough! Fab accessible content that got this (former) R-phobic coding and growing in confidence."

The CBF is committed to bridging the gap for underrepresented communities in the profession. We continuously engage in a variety of activities, including International Women's Day Scholarships and Widening Participation events, but we are always keen to engage in more. Please contact us with any ideas.



CONTACT US:

- [CBF.LIVERPOOL.AC.UK](https://www.liverpool.ac.uk/cbf/) 
- CBF@LIVERPOOL.AC.UK 
- [@LIVUNICBF](https://twitter.com/LIVUNICBF) 
- [CBF LINKEDIN](#) 

