Medicines Discovery Catapult (MDC)

Rafael C Jimenez

Accessing biomedical data for data analysis

B4BB, SME Forum
2022-07-06
MDC

Established in 2015
Based in Alderley Park, Cheshire - the UK’s largest single site life sciences campus
Part of the Catapult network – established by Innovate UK
Independent, not-for-profit organisation
Team of > 130 specialists with extensive industry experience

To reshape drug discovery for the benefit of patients

Supporting the drug discovery industry
SMEs and innovators to deliver growth for the drug discovery economy

Leveraging UK infrastructure
Connecting the community

Tackling industry-led challenges
That limit today’s discovery process

Industrialising and driving adoption
Of new scientific tools and modern techniques for discovering medicines
MDC capabilities

- Biomarkers
- Drug Delivery Biodistribution
- Biomedical informatics
- Cell model analysis and validation
- Virtual R&D and Syndicates
MDC Informatics

Analyse and interpret complex data to help you to make decisions about developing new medicines

Multidisciplinary expertise
In-house developed assets
Integration of third party data
Specialised software
Data infrastructure
Team and expertise

- **Chem-informatics**: Virtual screening, characterisation of drug properties, QSAR modeling
- **Metabolome Informatics**: Spectral deconvolution, MS database searches, metabolite identification & quantification
- **Bioinformatics**:
  - **Clinical Informatics**: Precision medicine, EHR, eCRFs, RWD, digital biomarkers, clinical study data analysis
  - **Genome Informatics**: Genome variation, functional genomics, genome editing, genome alignments, annotation
  - **Proteome Informatics**: Homology, alignments, variation, reactions, structures, domains, PTMs, identification
- **Bioimaging Informatics**: Pre-clinical (eg. X-ray CT, PET, MRI, ultrasound), microscopy, MS imaging, in-vivo imaging
- **Systems Biology and Multi-Omics Integration**: Molecular networks, Pathways, In silico biological models, multimodal data analysis
- **Data Science**: Machine learning, Text mining, NLP, Image recognition, Computer vision
- **Data Infrastructure**: Data management, FAIR data, Software development, Workflows, Infrastructure services
Informatics capabilities

**assets**
- **CartaPharma**: Resource integrating ext & int data assets
- **CriSPy**: DB with insights about drug discovery UK companies
- **Aurac**: Plugin enriching and visualising biochem entities
- **AssayNet**: Data mining resource of assays, drugs & targets
- **Ares**: Virtual screening pipeline using structural methods
- **DeepADMET**: Data mining resource of ADME properties
- **DREA.M.ai**: DB ranking drugs based on its properties
- **Apollo**: DB of antimicrobials & their gold-standard annotations
- **MDC-PPDMS**: Database of druggable domains
- **DSP ATLAS**: Tool to manage & analyse spatial expression data

**capabilities**
- **DARTS**: DB of real drugs and biological targets
- **Asclepius**: DB of Clinical Trials & Drug-Target-Disease annotations
- **VESPA**: Tool to predict the impact of mutations in proteins
- **Narcissus**: Tool to handle metadata & support analysis workflows
- **Polyphemus**: DB with compounds from vendors and external DBs
- **Pan**: Enriched library of chemical structures using FBLD
- **Dionysus**: Dataset of drugged and druggable genes
- **DocLib**: Software components for text mining
- **NER**: Tool for named entity recognition for biochem
- **AMBER+**: Toolkit to use and integrate AMBER

**insights**
- **Disease mechanisms**
- **Biomarkers**
- **Targets**
- **Druggability**
- **Drugs**
- **Efficacy**

**Central Nervous System | Infectious Diseases | Oncology | Complex medicines**
Workshop expectations

Learn more about **health data resources** and solutions to access, integrate, and analyse health data

Explore the **opportunity** for public funded service providers to **work as a collective** to support SMEs
Data to insights

SME

Question

Search
Interpretation
Access
Processing
Analysis

Insights

DATA Provider
Data to insights

- No capacity, tools or knowledge to find, access, process, integrate, analyse and interpret 3rd party data

**issues**

- No capacity to help SMEs with their technical requirements
- Limited number of tools and reference data to support data analysis
- Limited data analysis support or domain knowledge to help with data analysis
Work as a collective

- Quicker access to insights
- Focus on the science
- No need to worry about legal or technical related issues

SME

- Synergetic capacity to support SMEs
- Easier adoption of common solutions
- Better integration of public funded initiatives
- Not sharing data but insights
Questions

- What are the **needs** (and bottlenecks) of SMEs to **exploit** cohort data?
- What practical **solutions** could be easily implemented **across organisations** to support SMEs?
- What **agreements between organisations** could help to better support SMEs?
Thanks for your attention
Follow up proposal

Work as a collective to support SMEs to get insights from health data

- B4BB workshop: Exploring interest & solutions
- SME consultation: Learning from SMEs
- White paper: Proposing solutions
- Stakeholder survey: Validating solutions
- Follow up workshop: Working on implementation