Accessible AI for Small Molecule Discovery Research

Considerations and Value Drivers

Tamsin Mansley*, Benedict Irwin, Peter Hunt, Samar Mahmoud, Bailey Montefiore and Matthew Segall

Agenda

- Very brief introduction to Optibrium
- Applying AI in drug discovery
  - AI for different organisation types
  - Introduction to Cerella™
- Case studies
  - Practical application to heterogeneous drug discovery project data
  - Large-scale application to global pharma data
  - Combination with generative methods
- Conclusions

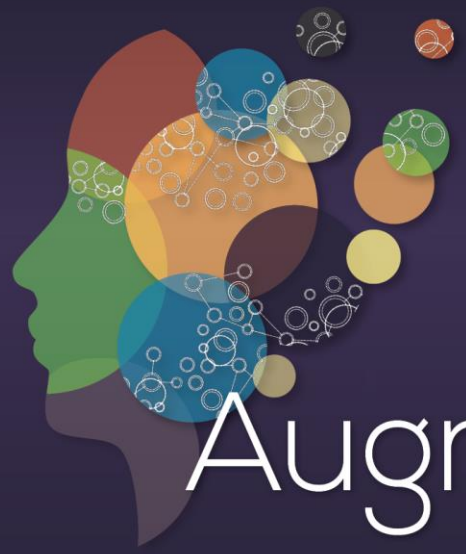
Introduction to Optibrium

www.optibrium.com

Augmented Chemistry™

- Optibrium creates elegant software solutions for small molecule design, optimisation and data analysis
  - R&D of novel technologies to guide decisions and improve efficiency in drug discovery
- Global customer base from top-ten pharma to small biotech and academia
  - >170 customers worldwide
  - Adoption in other chemistry fields, e.g. animal health, agrochemicals, etc.

optibrium™



Augmented Chemistry®

... enhance your expertise with
AI to make better decisions

Applying AI to Drug Discovery

optibrium™

AI for Different Types of Drug Discovery Organisation

AI's potential: Better decisions, faster progress, high-quality candidates

- **Large Pharma**
  - Extract maximum value from large, global, cross-project data sets
  - Frequently dedicate big teams to data science and AI research
    - Often struggle to get buy-in and adoption by users

- **Biotech**
  - Resource constraints limit the amount of data that can be generated
  - Need a turnkey solution to leverage limited data effectively
    - Caveat: Need *enough* data – not a 'magic bullet'

AI for Different Types of Drug Discovery Organisation

AI's potential: Better decisions, faster progress, high-quality candidates

- **Academic groups and consortia**
  - Heterogeneous data from multiple collaborations and public data sets
  - Data is often highly variable – it is difficult to get a consistent picture

- **Consistent themes**
  - Predicting intractable, hard-to-model endpoints, to make best use of valuable resources
    - For example, PK properties, and expensive, downstream experimental endpoints
  - We know we are missing opportunities because of missing, uncertain or inaccurate data and we want to find them

AI for Different Types of Drug Discovery Organisation

AI's potential: Better decisions, faster progress, high-quality candidates

- **VC or Investor carrying out Due Diligence**
  - Investments and acquisitions are $ multi-million bets made with very little information
  - More information about a project provides a better understanding about its chance of success
    - Is the data consistent?
    - Will it be possible to improve on the lead compound?
    - Are there already some hidden gems?
  - A **rigorous, data-driven** analysis can help to quantify the risk

Challenges of Using Data in Drug Discovery

Augmented Chemistry®

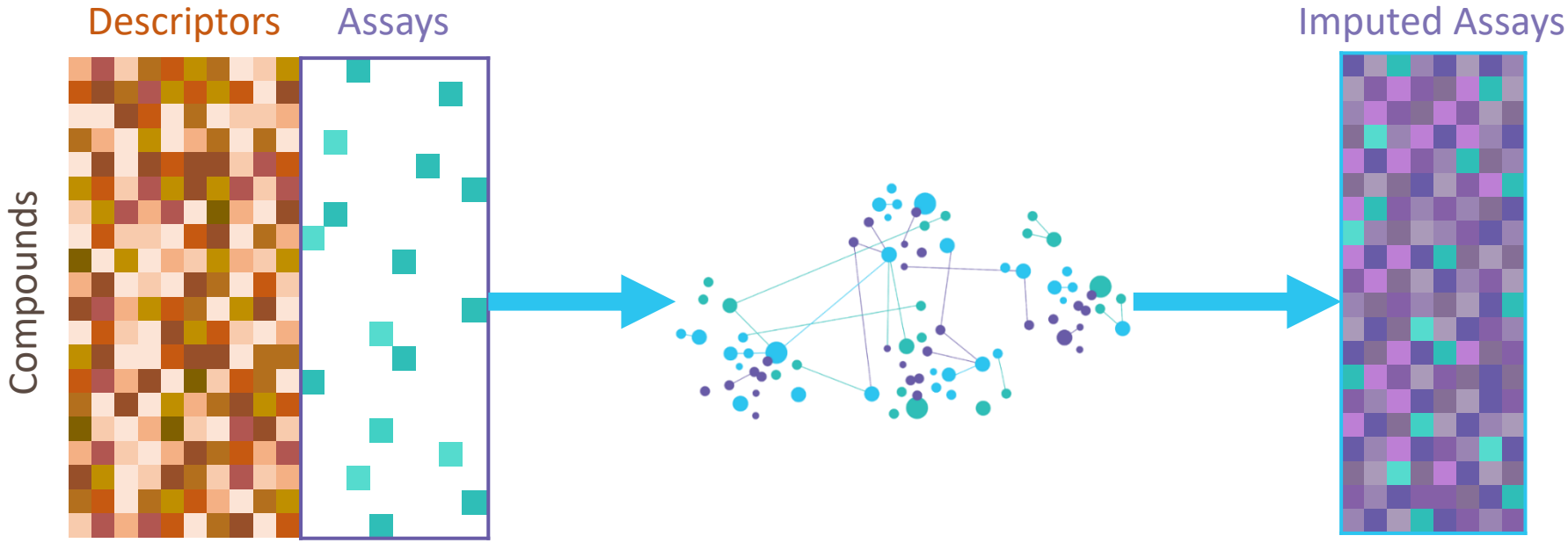
- It’s impossible to measure all of my compounds in all of my assays, how do I make the most of the data I have?

- I know there is variability in my experiments, how do I avoid being led astray by artefacts and errors?
- What are the most valuable experiments to run? What data will give me the most information with which to make decisions?
- How can I use the limited data I have to make better predictions for new compound designs, and choose the best ones for synthesis?

Alchemite™ Deep Learning Imputation
Optibrium's exclusive partnership with Intellegens

- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models

- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds

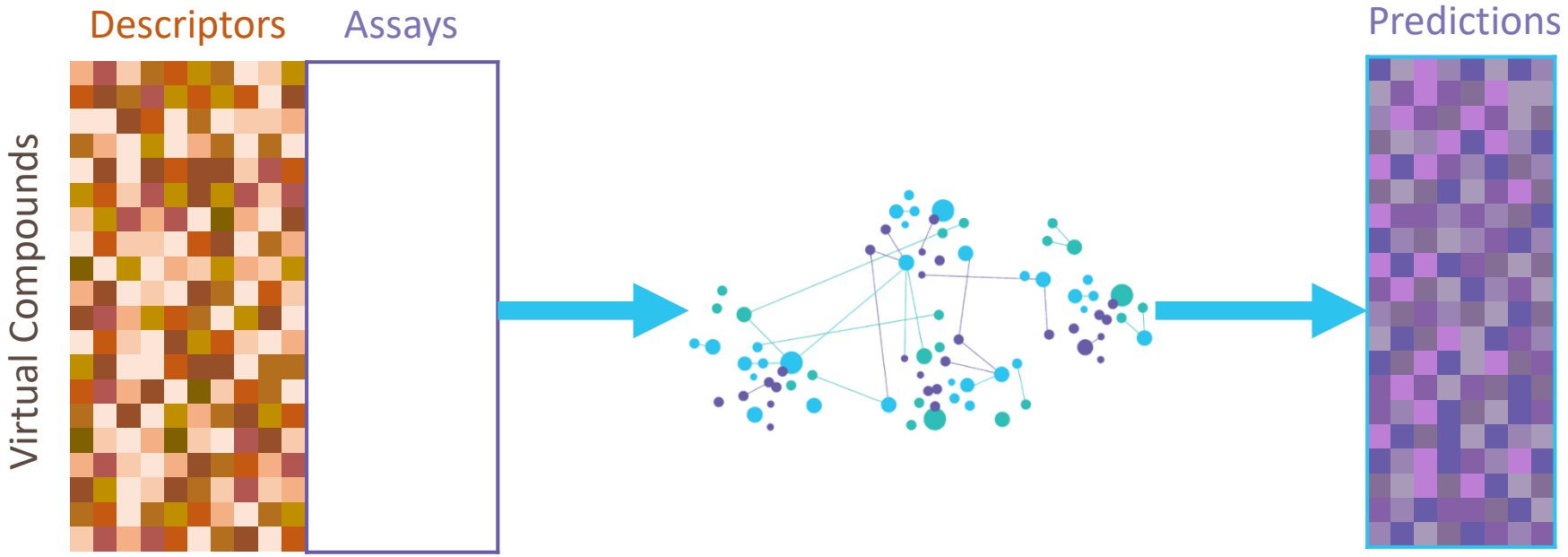


Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Alchemite™ Deep Learning Imputation

Optibrium's exclusive partnership with Intellegens

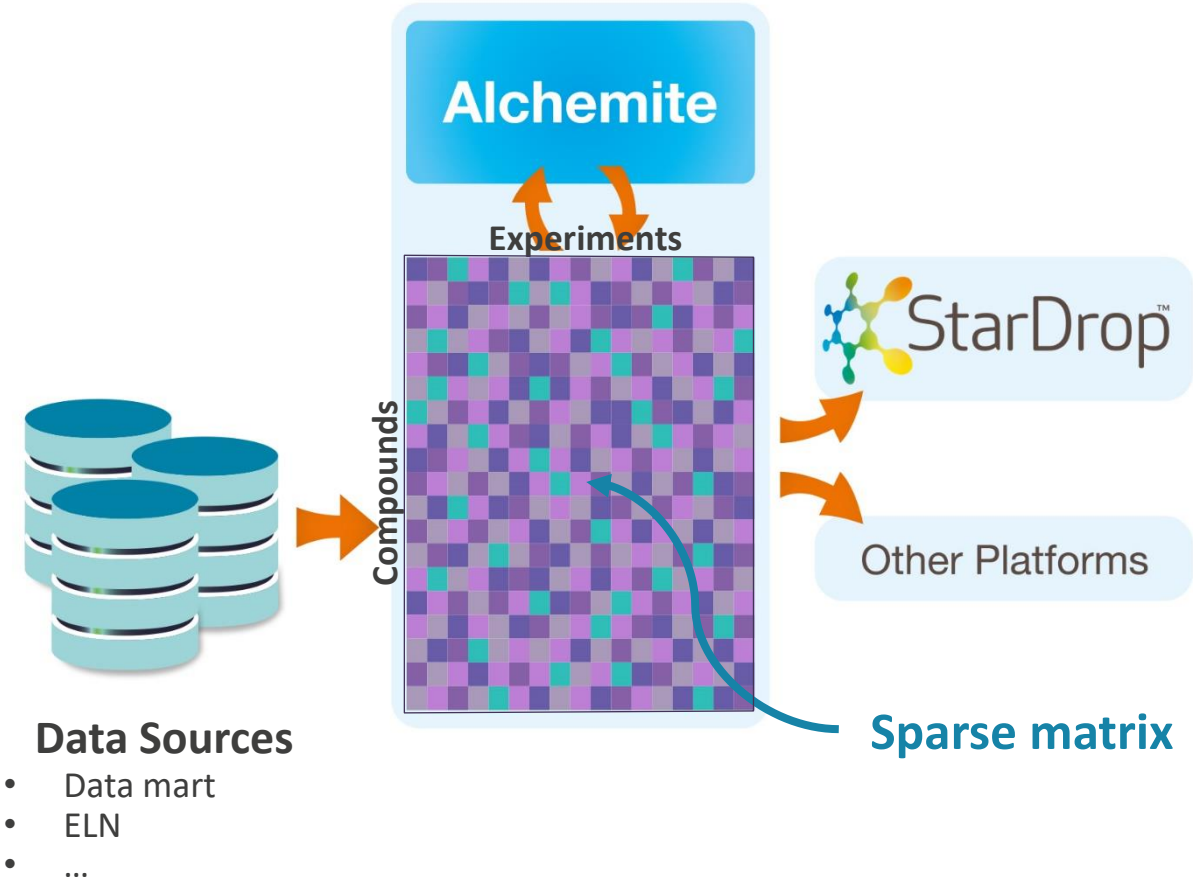
- Learns directly from relationships between experimental endpoints as well as SAR
  - Makes better use of sparse and noisy experimental data than conventional QSAR models
- 'Fills in' the gaps in your data and makes predictions for 'virtual' compounds
  - Generates more accurate predictions to target high-quality compounds



Whitehead *et al.* J. Chem Inf. Model. (2019) **59**(3) pp. 1197-1204, Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857

Cerella
Enabling active learning in drug discovery

- **Automatically** updates and prepares experimental data for model building
  - Connects seamlessly to data repositories
  - Applies cleaning, business rules and transformations to data for best model performance
- **Automatically** updates Alchemite models as new data become available
  - Always work with results based on the latest information
  - Remove the burden of manually building and updating models
- Manage ‘massive matrix’ of imputed results for easy access
  - May contain O(10^{10}) data points!
- Provide seamless access to results
  - Using StarDrop™ or any platform via a RESTful API



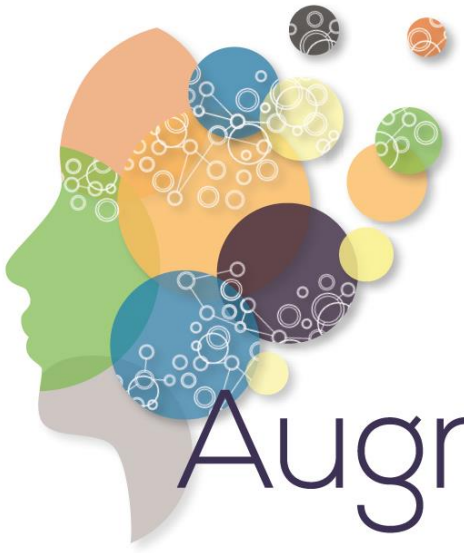
Watch our webinar at http://bit.ly/cerella_active

Deep Learning with Cerella
Reducing the time and cost of discovery cycles

Cerella™
an Augmented Chemistry® software product

- Proactively **highlight high-quality compounds** by more accurately ‘filling in’ sparse data (imputation)
- **Increase confidence** in decision making, identify **hidden opportunities,** flag outliers and false negatives
- Translate AI insights into planning of experiments and **focus on the most valuable measurements**
- Gain more value from your compound data, **accurately predicting complex endpoints**, intractable with conventional QSAR modelling
- Add value even for **individual project datasets**, while scaling to **global compound data** repositories

Case Studies



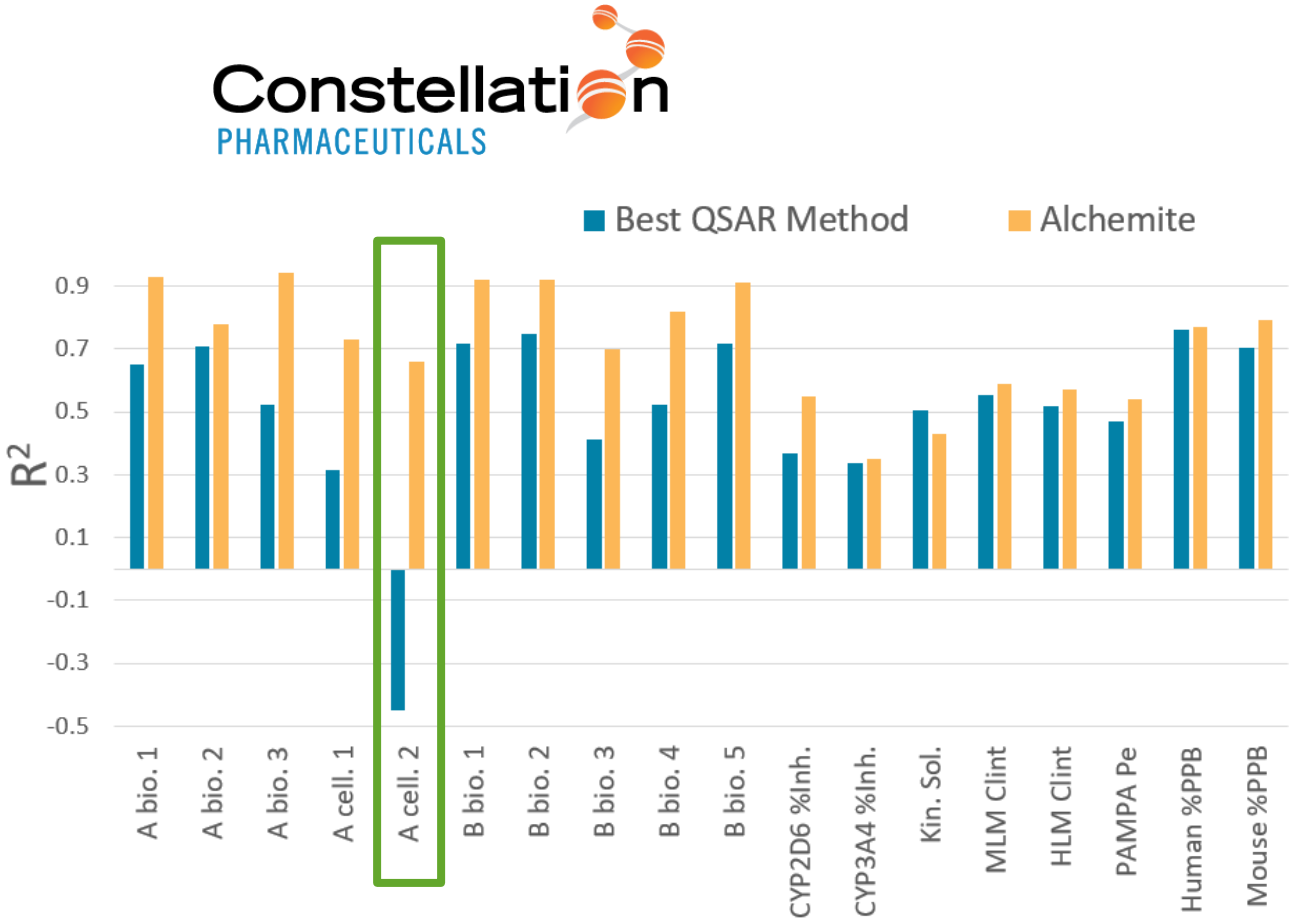
Augmented Chemistry®

Alchemite Application In Biotech

- Application to **heterogeneous** data across two projects
  - Target and phenotypic activities and ADME endpoints
  - 2453 compounds across 18 endpoints
- Significant improvement in accuracy

	Average R^2
Best QSAR	0.50
Alchemite™	**0.72**

- Example of value delivered:
  - Few false negatives among confidently-predicted inactives – could have saved >$600,000 in unnecessary synthesis



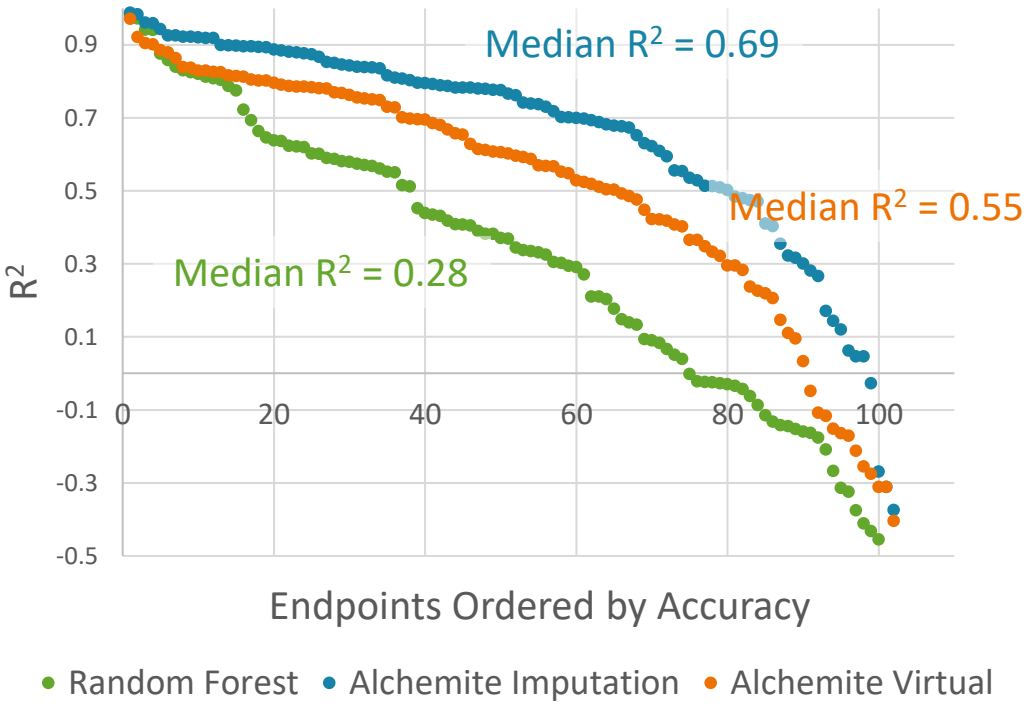
Irwin *et al.* J. Chem. Inf Model. (2020) **60**(6), pp. 2848–2857
Watch our webinar: http://bit.ly/practical_deeplearning

Alchemite Application in Big Pharma

- Application to large data set
  - **678,994** compounds
  - 1,116 experimental endpoints
  - 2% complete
- Covering a **full range** of drug discovery assays, incl
- Example of value delivered:
  - "…an extension of what medicinal chemists… do in a be possible for a person."

Takeda | ONCOLOGY

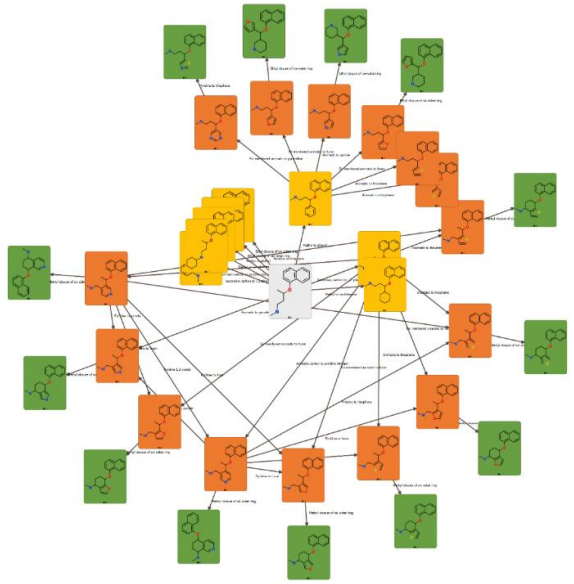
Prospective Prediction of Project Target Activities

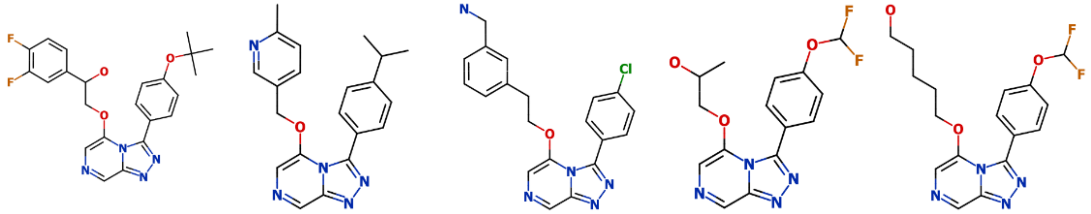


Irwin *et al.* App. AI Lett. (2021) DOI: 10.1002/ail2.31
Watch our webinar: http://bit.ly/largescale_imputation

Collaboration with Open Source Malaria (OSM)
Combination with generative methods

- Application to **sparse antimalarial activity data**
  - Targeting novel MoA – *Pf*ATP4
  - Alchemite generated one of the top-ranked models in the OSM competition
- New compound ideas were generated using the Nova™ module in StarDrop™
  - Prioritised with Alchemite model
  - Good activity profile and properties
- A **confidently** predicted compound was synthesised and tested by OSM
  - **Only confirmed active** of those proposed by four organisations
- Example of value delivered:
  - "[The Optibrium compound] is something that I am… sure none of the humans involved in the SAR over the years would have bothered making…"





Optibrium/ Intellegens	Davy Guan	Exscientia	Molomics	Molomics
0.65 µM	>25 µM	11 µM	>25 µM	14 µM

Irwin *et al.* Int. Pharm. Ind. (2020) **12**(2) pp. 28-31
Watch our webinar http://bit.ly/ai_antimalarials

Conclusions

- Organisations of different types and sizes have the same goals
  - Make better use of valuable resources
  - Avoid missed opportunities
  - Make go/no-go decisions with confidence and quickly identify the next drug candidate
- They are looking at the same problems through different lenses
  - Too much data vs too little
  - Sparse, noisy, heterogeneous data
  - Budgetary/resource constraints
- AI approaches, like **Cerella™**, provide insights by using all the available data.
  - Enables confident decision-making by highlighting interesting data and patterns
  - Significantly reduces the time and cost of discovery cycles

Acknowledgements

- Partners at Intellegens Limited
  - Ben Pellegrini
  - Gareth Conduit
  - Tom Whitehead
- Colleagues at Optibrium
- 18+ customer organisations who have participated in projects leading to additional Case Studies
- Session organisers – Marti Head and Kate Hardy

For more information visit: optibrium.com/cerella