Data integration and computational systems approaches to target discovery and pharmacokinetic modelling

Kenji Mizuguchi

http://mizuguchilab.org
kenji@nibiohn.go.jp
Data integration for pharmacokinetic modelling

Target Discovery
Toxicity-related attrition

Primary Reasons for Drug Withdrawals from US, European or Asian Markets (Toxicity) 1998-2008

- Cardiac toxicity
- Hepatic toxicity
- Muscle
- Gastric
- Psychiatric
- Others
- 11
- 4
- 17
- 15

J.S. MacDonald, R.T. Robertson

Pharmacokinetics; PK

Basis for understanding toxicity, as well as poor efficacy

Kenji Mizuguchi (NIBIOHN)

- Integrated databases
- Multi-scale modelling (prediction) systems

Teruki Homma (RIKEN)
Hiroshi Yamada (NIBIOHN)

funded by
Drug Discovery Informatics System

Pharmacokinetics

Collaboration with pharmaceutical companies

DB integration

Cardiac toxicity

Hepatic toxicity

Prediction system

Portal site

Compound prioritization at screening

Molecular design during optimization

DB integration
ADMET

Absorption and Metabolism
(Therapeutic drug monitoring)

- A drug which has a well absorption and faster metabolism
- A drug which has a poor absorption and slower metabolism

Distribution
The compound needs to be carried to its effector site, most often via the bloodstream.

Excretion
The kidney is the most important site and it is where products are excreted through urine.

Absorption

Distribution

Metabolism

Toxicity

Excretion

Blood concentration

Time
Collecting pharmacokinetic parameters

Non-curated experimental data

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Value</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Comp. A</td>
<td>66000</td>
<td>mL/min/g</td>
<td>Intrinsic clearance in human liver microsomes</td>
</tr>
<tr>
<td>2</td>
<td>Comp. B</td>
<td>135</td>
<td>mL/min/g</td>
<td>Drug metabolism in human liver microsomes assessed as intrinsic clearance</td>
</tr>
<tr>
<td>3</td>
<td>Comp. C</td>
<td>58.5</td>
<td>mL/min/kg</td>
<td>Intrinsic clearance in human microsome in presence of NADPH</td>
</tr>
<tr>
<td>4</td>
<td>Comp. C</td>
<td>16.2</td>
<td>mL/min/kg</td>
<td>Intrinsic clearance in human microsome</td>
</tr>
</tbody>
</table>

Check the original publications

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Value</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Comp. A</td>
<td></td>
<td></td>
<td>Intrinsic clearance in human liver microsomes</td>
</tr>
<tr>
<td>2</td>
<td>Comp. B</td>
<td>135</td>
<td>μL/min/mg</td>
<td>Drug metabolism in human liver microsomes assessed as intrinsic clearance</td>
</tr>
<tr>
<td>3</td>
<td>Comp. C</td>
<td>58.5</td>
<td>μL/min/mg</td>
<td>Intrinsic clearance in human microsome in presence of NADPH</td>
</tr>
<tr>
<td>4</td>
<td>Comp. C</td>
<td>16.2</td>
<td>μL/min/mg</td>
<td>Intrinsic clearance in human microsome</td>
</tr>
</tbody>
</table>

Import the curated data to database

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Value</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Comp. A</td>
<td>66</td>
<td>mL/min/mg</td>
<td>Intrinsic clearance in human liver microsomes</td>
</tr>
<tr>
<td>2</td>
<td>Comp. B</td>
<td>135</td>
<td>μL/min/mg</td>
<td>Drug metabolism in human liver microsomes assessed as intrinsic clearance</td>
</tr>
<tr>
<td>3</td>
<td>Comp. C</td>
<td>58.5</td>
<td>μL/min/mg</td>
<td>Intrinsic clearance in human microsome in presence of NADPH</td>
</tr>
</tbody>
</table>

Esaki et al. Mol Inform (2018)
Data quality influences prediction accuracy

Non-curated data

Calculate Descriptors from Structure
- Atom count
- MW
- logP
- PSA

Accuracy: 0.75
Kappa: 0.56

Accuracy: 0.67
Kappa: 0.47

Curated data

Build prediction models & Cross validation

External test set

Evaluate using test set

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable</td>
<td>Unstable</td>
</tr>
<tr>
<td>Stable</td>
<td>2716</td>
</tr>
<tr>
<td>Unstable</td>
<td>158</td>
</tr>
</tbody>
</table>

Accuracy: 0.77
Kappa: 0.59

Accuracy: 0.74
Kappa: 0.54
Data integration for pharmacokinetic modelling

Target Discovery
TargetMine
Data warehouse for drug target prioritization
http://targetmine.mizuguchilab.org

- Integrate a wide range of data sources.
- Enable complicated searches that are difficult to achieve with existing tools.
- Custom annotations and in-house data.

Powered by

InterMine
at the Cambridge Systems Biology Centre

Data sources in TargetMine

- micro RNA & target associations
- Transcription factor target associations
- Pathways & interactions
- Genes & proteins
- Domains & 3D structures
- Chemical compounds & interactions

INTEGRATION

Homo sapiens, Rattus norvegicus, Mus musculus
Obtaining structured clinical data

Questionnaire

Clinical laboratory tests

EMR

Pulmonary Function Tests

Integrated DB

X-ray, CT

With Osaka University Respiratory Medicine, Medical Informatics
Important features as distributed word representations have been proposed in the literature to improve the performance of linear models on representations obtained via AEs. Tran et al. have used this approach for representing EHR concepts. They formulated a modified restricted RBM which uses a structured approach for representing clinical concepts. The sequential ordering of medical codes to derive distributed code embeddings. Skip-gram is a popular technique for learning models. Similarly, E. Choi et al. have applied deep unsupervised representation learning techniques to derive EHR concept vectors that capture the latent similarities and natural clusters between medical concepts.

A recent study by Choi et al. has shown that using skip-gram in the context of clinical data can produce distributed vector representations. Aside from NLP-inspired methods, other techniques such as generative adversarial networks (GANs), deep belief networks (DBNs), and autoencoders (AEs) have been used for representing EHR concepts. These models can be trained on raw features and then used to derive patient representations.

From clinical data to drug targets

Omics data

Patient vectors

Images

Clustering and learning

Thousands of Sparse Codes

Key proteins/genes

With AIRC
Conclusions

- General themes for rational drug discovery and development
- Data integration is key
- Need to automate curation and knowledge extraction