Quantitative prediction of skin sensitisation potency based on structural alert spaces

vICGM, April 2016

Martyn Chilton
Scientist
martyn.chilton@lhasalimited.org

Overview

- Background
- Lhasa EC3 dataset
  - Data gathering and curation
  - Composition
- EC3 model
  - Methodology
  - Performance
  - Limitations
  - Demonstration
- Conclusions

Background: Derek Nexus and skin sensitisation

- Derek Nexus has 88 alerts for skin sensitisation
  - Based on assay data from mice, guinea pigs and human
- Currently we make qualitative predictions
  - Hazard identification
- We also want to be able to quantitatively estimate skin sensitisation potency
  - To aid in risk assessment
  - Desirable for ethical and regulatory reasons
  - Requires skin sensitisation potency data

Background: The LLNA

- The murine Local Lymph Node Assay (LLNA) is the gold standard assay for predicting skin sensitisation
- Measures the proliferation of T-lymphocytes in the lymph nodes
  - One of the key events in the skin sensitisation Adverse Outcome Pathway (AOP)
- Provides a measure of potency through an EC3 value
  - Estimated concentration of a compound that causes a 3-fold increase in lymphocyte proliferation compared with controls

OECD (2012), *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence*, Series on Testing and Assessment, No. 168, ENV/JM/MONO(2012)10/PART1

Background: The LLNA

- EC3 values have been shown to correlate with human skin sensitisation potential

ICCVAM. 2011. ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. NIH Publication No. 11- 7709. Research Triangle Park, NC: National Institute of Environmental Health Sciences

Background: The LLNA

- EC3 values have been shown to correlate with human skin sensitisation potential
- Sensitisers can be assigned to one of four ECETOC potency categories:

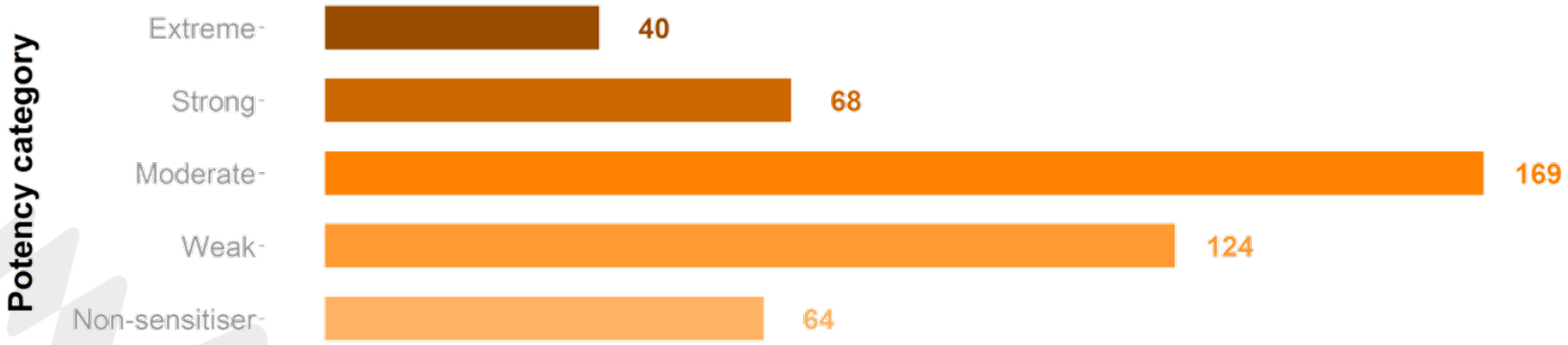
Kimber *et al.*, *Food Chem. Toxicol.* **2003**, *41*, 1799-1809

Lhasa EC3 dataset: Data gathering and curation

- We gathered as much publicly available EC3 data as possible
- The data was curated to ensure it was of high quality
  - Original experimental reports were located and examined
  - Unsuitable/unreliable data were not included in the final dataset
- When more than one LLNA study was found for the same compound the median EC3 value was taken

Lhasa EC3 dataset: Composition

- Data from 1051 LLNA studies were collected, resulting in a dataset containing 664 unique compounds
- Of these, 465 fire only one alert in Derek Nexus
  - These compounds span a good range of EC3 values
  - They include some non-sensitisers that fire a Derek alert



EC3 model: Initial considerations

- We would like to make use of existing knowledge captured in Derek's alerts for skin sensitisation
  - Each alert space corresponds to a group of chemicals which are believed to react with skin proteins through the same mechanism
- Any model built needs to be transparent and interpretable
- The methodology must be scientifically defensible

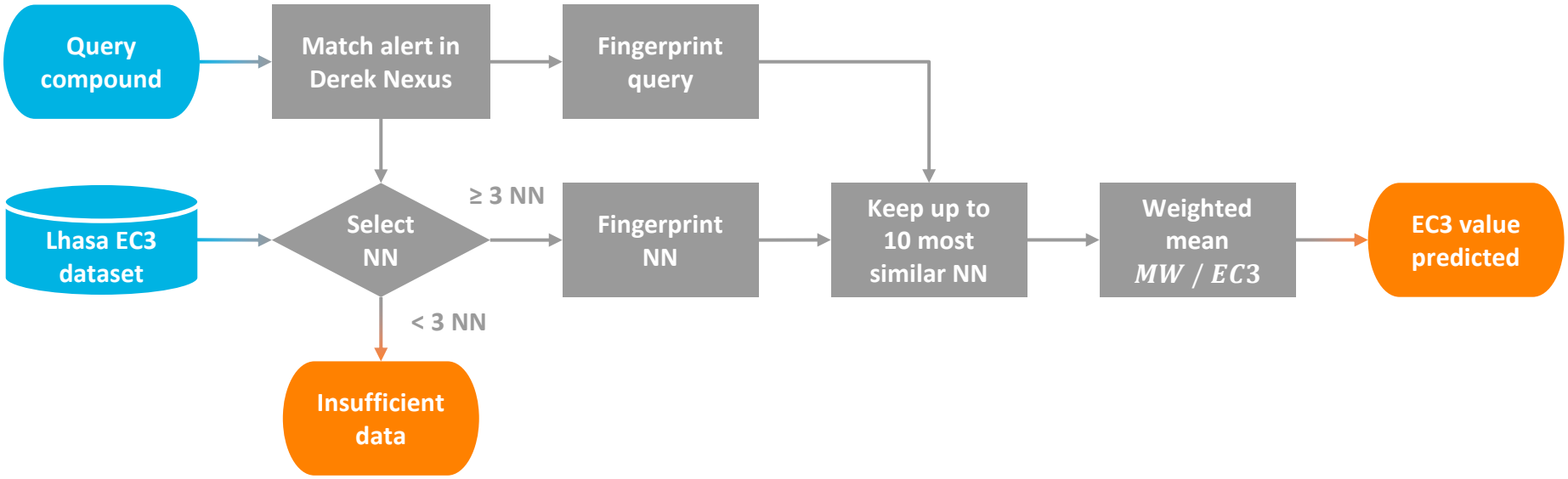
EC3 model: Possible methodologies

- Regression models for different structural alerts
  - Some success, but not very interpretable
- Average EC3 values for each structural alert
  - Worked well for some alerts, but not others
- Finding nearest neighbours from within an alert space
  - Provided transparent and interpretable predictions

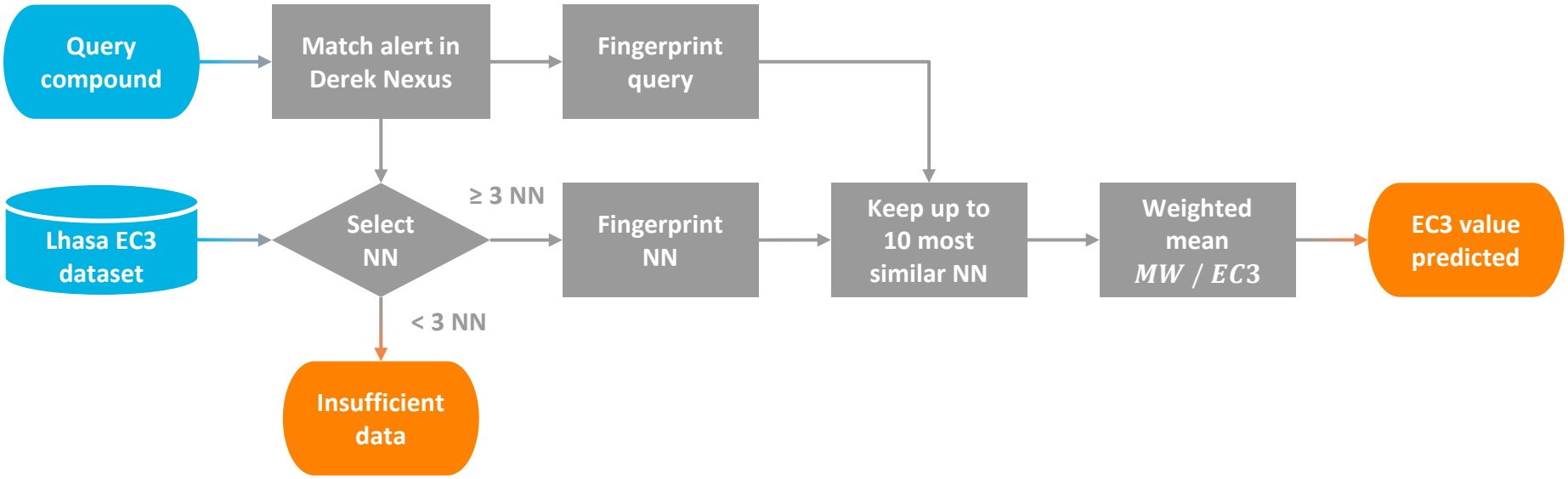
EC3 model: Possible methodologies

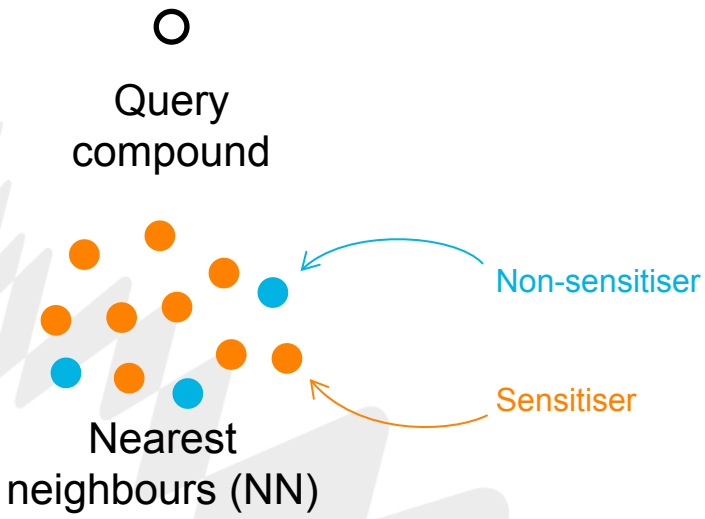
- Regression models for different structural alerts
  - Some success, but not very interpretable
- Average EC3 values for each structural alert
  - Worked well for some alerts, but not others
- **Finding nearest neighbours from within an alert space**
  - Provided transparent and interpretable predictions

EC3 model: Alert-based nearest neighbours

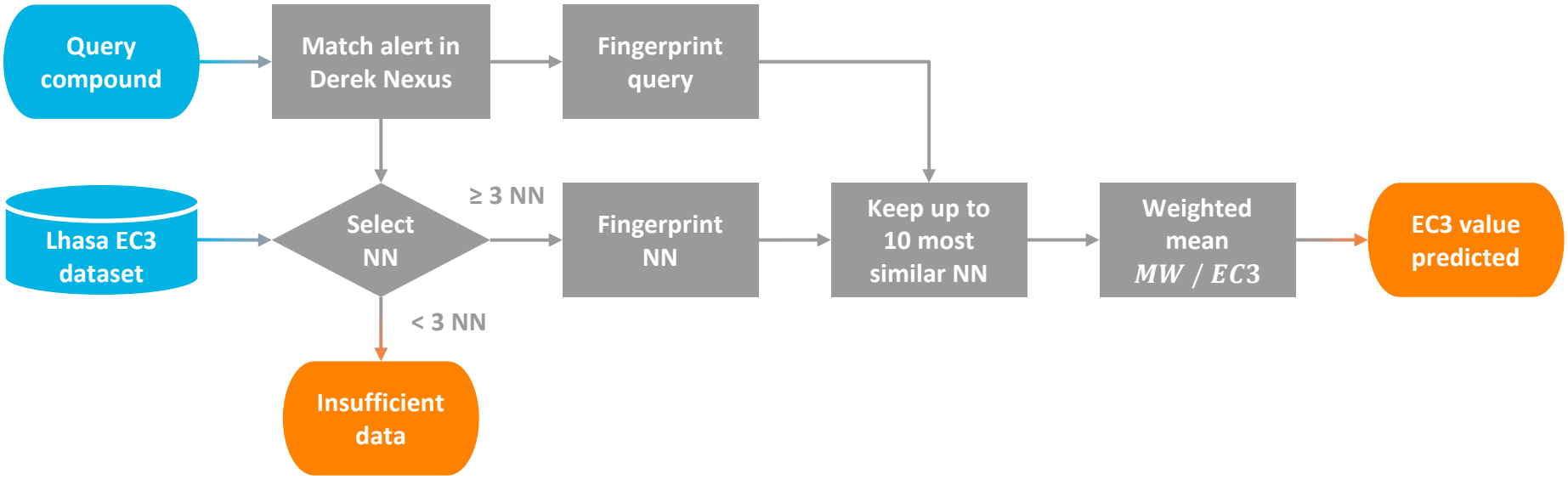


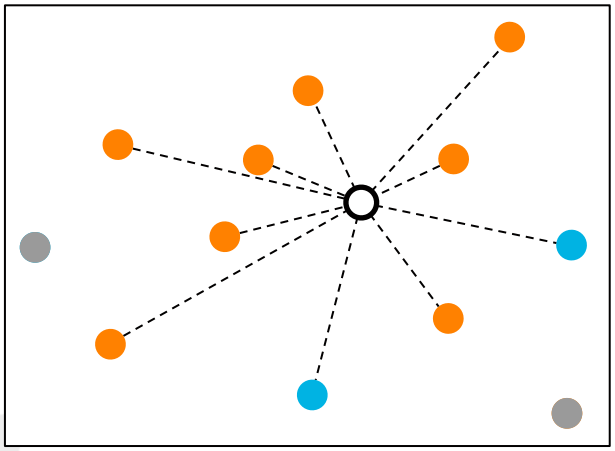
EC3 model: Alert-based nearest neighbours





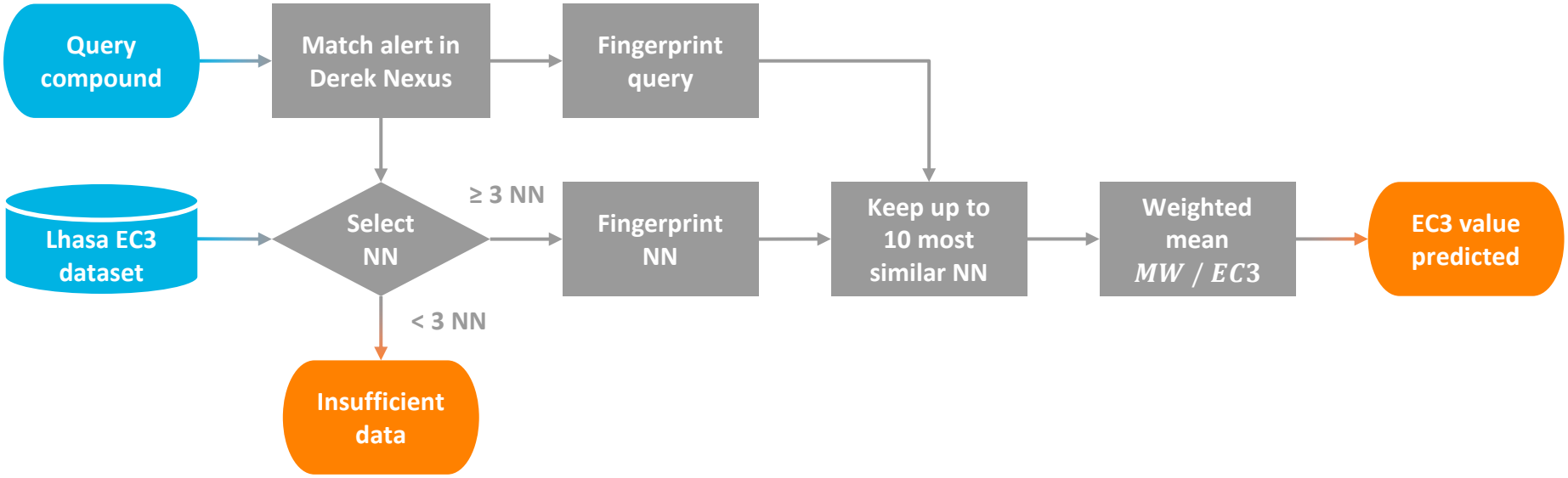
EC3 model: Alert-based nearest neighbours





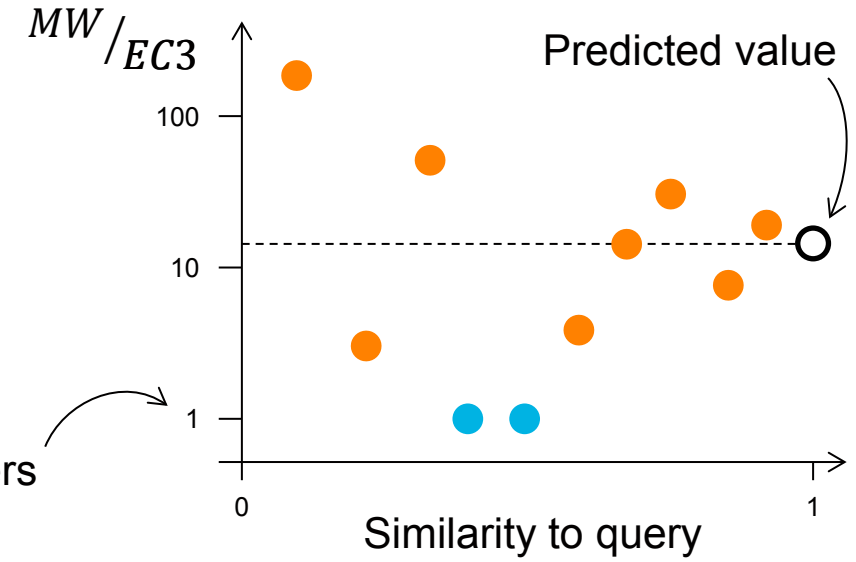
Chemical space

EC3 model: Alert-based nearest neighbours



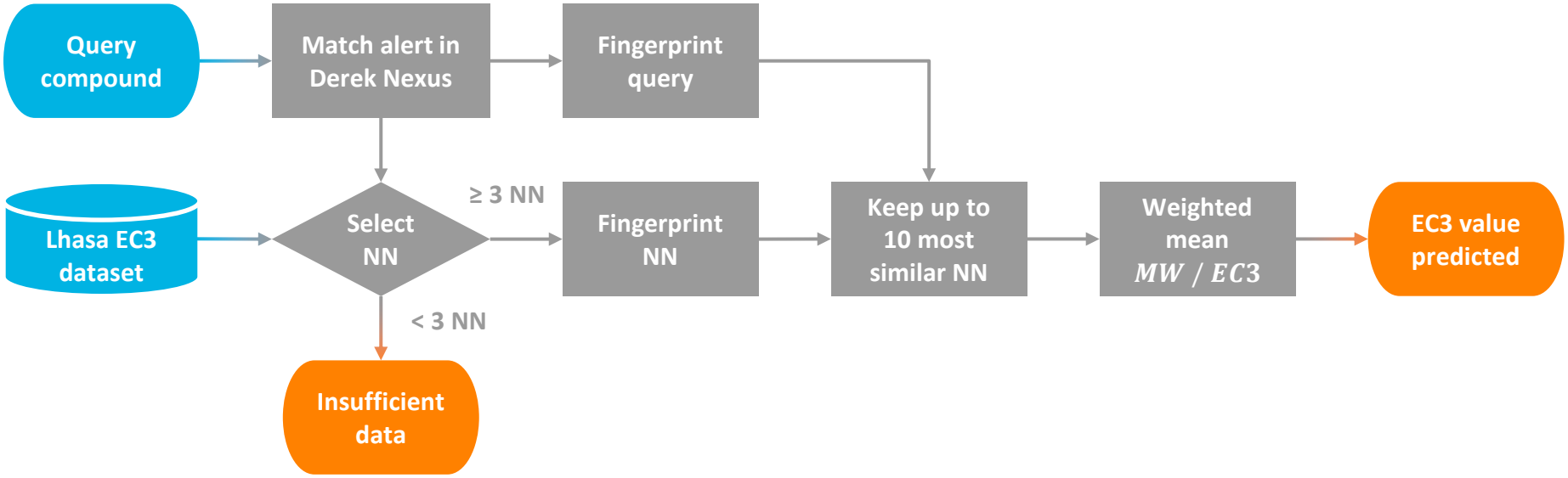
$$\frac{MW_q}{EC3_q} = \frac{\sum_{n=1}^{N}\left(\frac{MW_n}{EC3_n}\right)T_{q,n}}{\sum_{n=1}^{N}T_{q,n}}$$

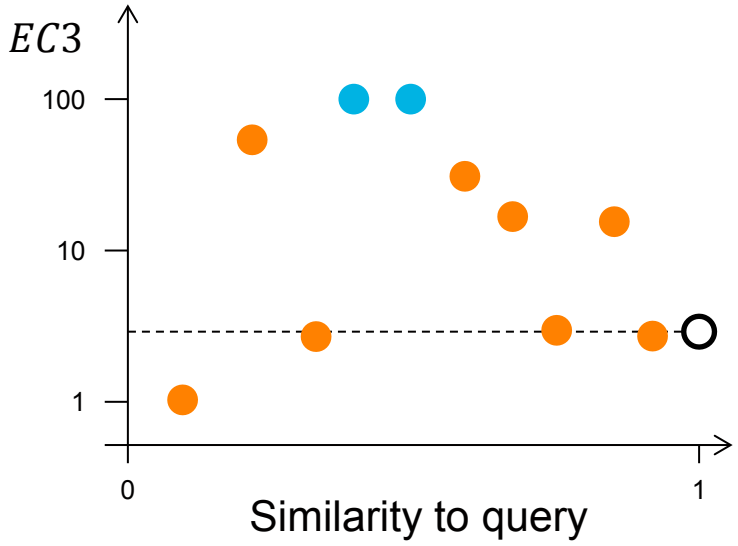
q	=	query compound
N	=	number of nearest neighbours
n	=	n^{th} nearest neighbour
$T_{q,n}$	=	Tanimoto index between q and n



A. Natsch *et al.*, *Toxicol. Sci.* **2015**, *143*, 319-332

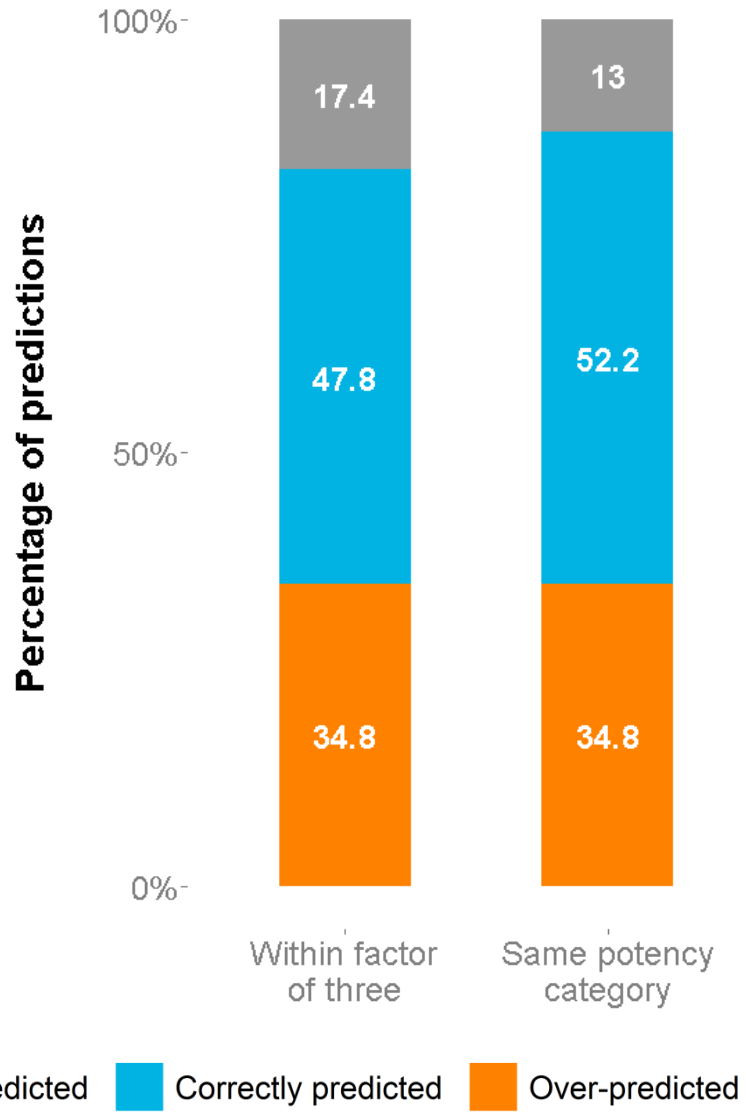
EC3 model: Alert-based nearest neighbours



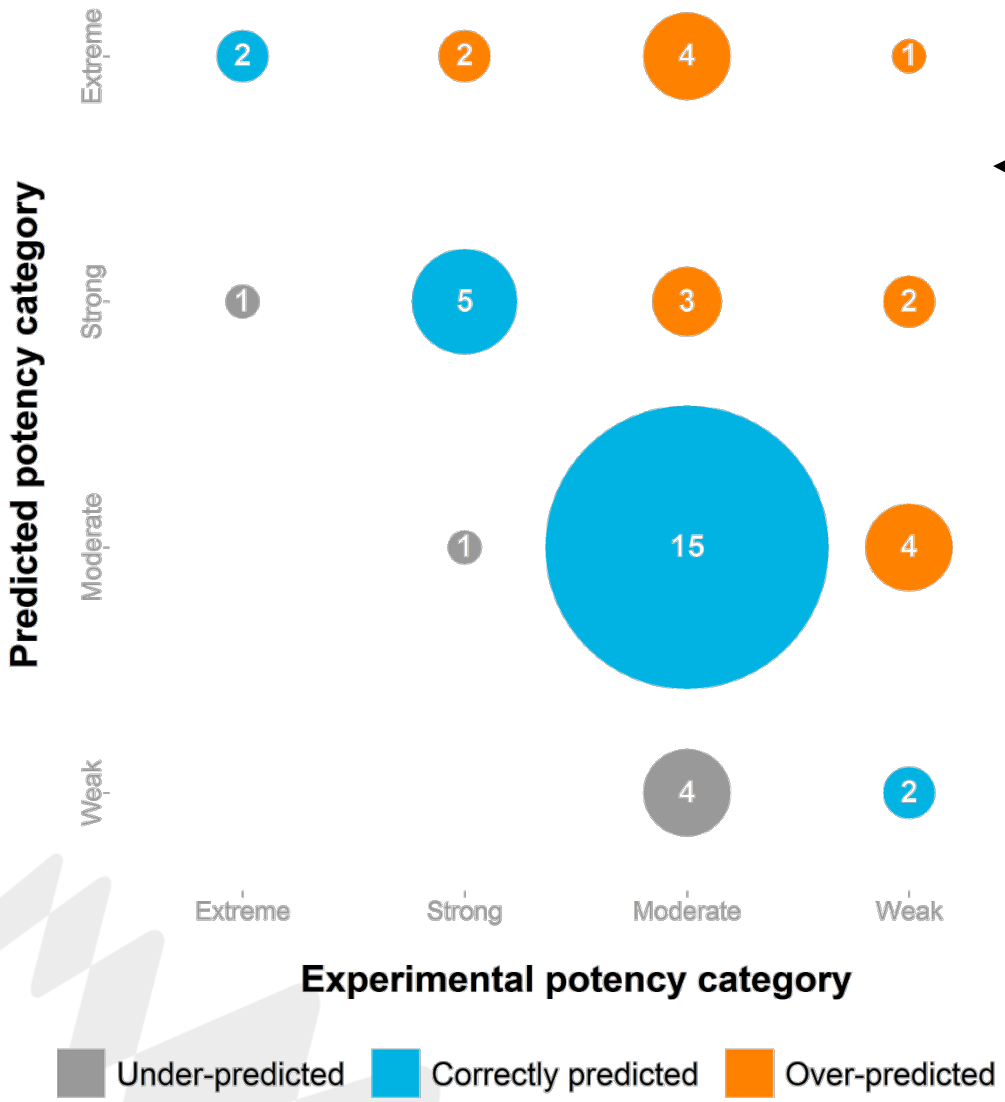


EC3 model: Performance

- The model was assessed using a validation set (*n* = 46)
- Predictions were judged as accurate according to two separate criteria:
  - Within a factor of 3 of the experimental EC3 value
  - Within the same ECETOC potency category as the experimental EC3 value



EC3 model: Performance



When the model is wrong, it tends to over-predict rather than under-predict the potency

EC3 model: Limitations

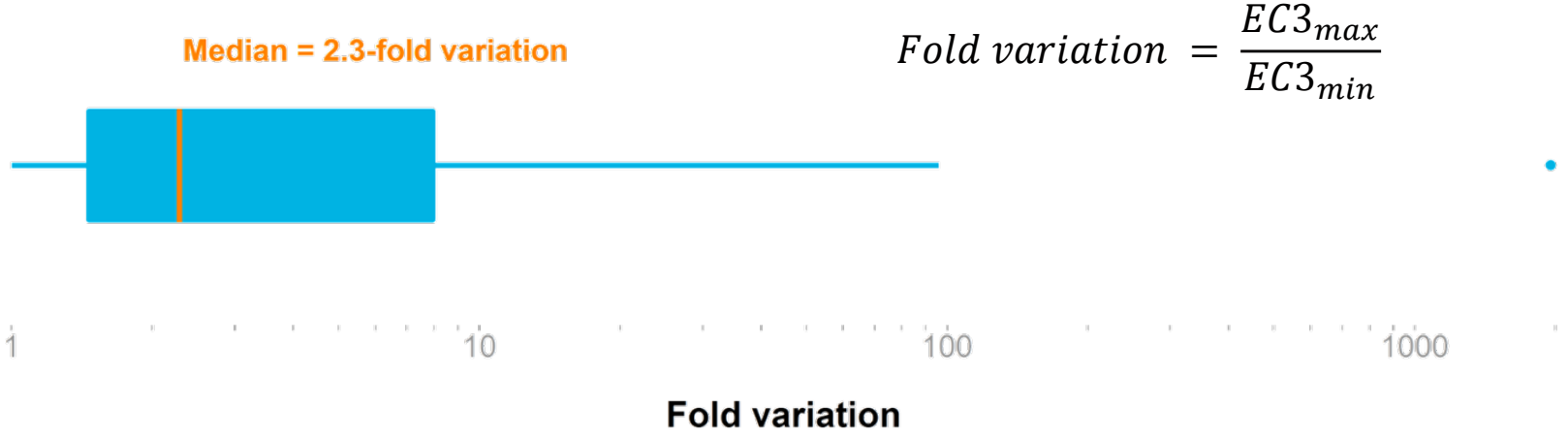
1. Coverage

- Directly linked to the size of the Lhasa EC3 dataset
  - This depends on the amount of publicly available LLNA data
- The EC3 model covers 39 of the skin sensitisation alerts within Derek Nexus
- Currently there are 49 alerts with fewer than three compounds in our dataset
  - Potential validation compounds: ~80% coverage
  - Do you have data you could share?

EC3 model: Limitations

2. Variability in LLNA data

- EC3 values can vary between different assay runs
  - This can be seen in the 87 compounds in the Lhasa EC3 dataset with multiple EC3 values



- This will affect the overall accuracy of the model

Conclusions

- We have developed an EC3 model which makes quantitative predictions of skin sensitisation potency
  - Built upon high quality, publicly available LLNA data
- Predictions are made by finding nearest neighbours to the query compound within defined structural alert spaces
  - Makes use of existing knowledge found in Derek Nexus alerts
- The model performs well against a validation set, both in terms of predicting EC3 values and potency categories
- Provides transparent and interpretable predictions

Acknowledgements

- Steve Canipa
- Donna Macmillan
- Jeff Plante
- Jonathan Vessey

Thank you for your attention

Any questions?

Lhasa Limited

shared **knowledge** • shared **progress**

Lhasa Limited Registered Office
Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS
Registered Charity (290866)

+44 (0)113 394 6020
info@lhasalimited.org
www.lhasalimited.org

Company Registration Number 01765239. Registered in England and Wales. VAT Registration Number GB 396 8737 77.

ISO 9001: CERTIFIED