Characterisation of the immune repertoire of a humanised transgenic mouse through immunophenotyping and high-throughput sequencing

**Supplementary Information**

Richardson E1,2\*, Binter Š1\*, Kosmac M1, Ghraichy M3,4, von Niederhausern V3,4, Kovaltsuk A2, Galson J5, Trück J3,4, Kelly DF6, Deane CM2, Kellam P1, 7, Watson SJ1

1 Kymab, a Sanofi Company, Babraham Research Campus, Cambridge, UK

2 Department of Statistics, University of Oxford, Oxford, UK

3 Division of Immunology, University Children's Hospital, University of Zurich, Zurich, Switzerland

4 Children's Research Center, University of Zurich, Zurich, Switzerland

5 Alchemab Therapeutics Ltd, Kings Cross, London, UK

6 Department of Paediatrics, University of Oxford, Oxford, UK

7 Department of Infectious Diseases, Faculty of Medicine, Imperial College London, UK

\* Contributed equally to the study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **IGHV subgroup** | **Dominant H1 canonical form** | **Dominant H2 canonical form** | **% in dominant H1 canonical form** | **% in dominant H2 canonical form** |
| Human | IGHV1 | H1-8-A | H2-8-A | 60.5 | 100.0 |
| Human | IGHV2 | H1-10-A | H2-7-A | 100.0 | 100.0 |
| Human | IGHV3 | H1-8-A | H2-8-B | 100.0 | 86.7 |
| Human | IGHV4 | H1-8-A | H2-7-A | 63.7 | 100.0 |
| Human | IGHV5 | H1-8-C | H2-8-A | 100.0 | 100.0 |
| Human | IGHV6 | H1-10-B | None | 100.0 | 100.0 |
| Human | IGHV7 | H1-8-C | H2-8-A | 99.9 | 100.0 |
| Kymouse | IGHV1 | H1-8-C | H2-8-A | 66.3 | 100.0 |
| Kymouse | IGHV2 | H1-10-A | H2-7-A | 100.0 | 100.0 |
| Kymouse | IGHV3 | H1-8-A | H2-8-B | 100.0 | 78.9 |
| Kymouse | IGHV4 | H1-8-A | H2-7-A | 46.2 | 100.0 |
| Kymouse | IGHV5 | H1-8-C | H2-8-A | 66.3 | 100.0 |
| Kymouse | IGHV6 | H1-10-B | None | 100.0 | 100.0 |
| Kymouse | IGHV7 | H1-8-C | H2-8-A | 100.0 | 100.0 |
| Mouse | IGHV1 | H1-8-A | H2-8-A | 49.9 | 96.7 |
| Mouse | IGHV2 | H1-8-A | H2-7-A | 99.9 | 100.0 |
| Mouse | IGHV3 | H1-9-A | H2-7-A | 90.2 | 100.0 |
| Mouse | IGHV4 | H1-8-A | H2-8-B | 81.4 | 100.0 |
| Mouse | IGHV5 | H1-8-A | H2-8-B | 98.1 | 100.0 |
| Mouse | IGHV6 | H1-8-A | H2-10-A | 99.9 | 100.0 |
| Mouse | IGHV7 | H1-8-A | H2-10-A | 100.0 | 100.0 |
| Mouse | IGHV8 | H1-10-A | H2-7-A | 100.0 | 94.3 |
| Mouse | IGHV9 | H1-8-A | H2-8-A | 100.0 | 100.0 |

**Supplementary File 1a**: dominant canonical forms per IGHV subgroup. Predicting the canonical form based solely on the dominant form observed for the IGHV subgroup would result in accuracy between 46.2 and 100%. Differences can be seen between the Kymouse and human in IGHV1, in which the majority of human sequences are H1-8-A vs. H1-8-C in the Kymouse.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CDRH3 length (A.A.) | Number of unique CDRH3s | Number of clusters with 2Å cut-off | Number of clusters with 1Å cut-off | Number of clusters with 0.6Å cut-off |
| 4 | 25 | 2 | 2 | 5 |
| 5 | 43 | 2 | 2 | 6 |
| 6 | 129 | 2 | 5 | 12 |
| 7 | 309 | 2 | 7 | 26 |
| 8 | 640 | 4 | 13 | 60 |
| 9 | 1304 | 4 | 24 | 115 |
| 10 | 4348 | 6 | 30 | 175 |
| 11 | 6373 | 6 | 37 | 264 |
| 12 | 6799 | 6 | 69 | 448 |
| 13 | 6828 | 12 | 111 | 894 |
| 14 | 6040 | 16 | 188 | 1275 |
| 15 | 4790 | 19 | 220 | 1460 |
| 16 | 3769 | 23 | 289 | 1806 |
| **Total** | **41,397** | **104** | **997** | **6546** |
| Metric |  | Metric at 2Å | Metric at  1Å | Metric at  0.6Å |
| Kymouse and humans monophyletic | - | True | True | True |

**Supplementary File 1b:** information about the number of CDRH3 structural clusters produced with different thresholds under the select greedy clustering algorithm. 0.6Å was the threshold selected by Kovaltsuk and colleagues (Kovaltsuk et al, 2020) in the original SAAB+ application.