**Rationale for choosing light chains**

Three light chains chosen:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Kappa/Lambda | Vfamily | Vgene | Jfamily | Edited\_AA\_junction | Num\_AAs | pI | D |
| Kappa | IGKV3 | IGKV3-20 | IGKJ2 | QQYGSSPYT | 9 | 6.02 | 0 |
| Lambda | IGLV1 | IGLV1-44 | IGLJ2 | AAWDDSLNGVV | 11 | 3.49 | 2 |
| Kappa | IGKV3 | IGKV3-11 | IGKJ4 | QQRSNWPLT | 9 | 10.55 | 0 |

I have decided to first concentrate on the heavy chains. Therefore, I think that the best approach is to choose a very small number of light chains which are likely to pair with a wide range of heavy chains and represent a wide range of light chain properties. This way, if every heavy chain tested is paired with each of these three light chains, then the different heavy chains can be compared and we will know that any difference in binding is due to differences in the heavy chain, not the light chain it is paired with. Once I know more about the heavy chains, I can go back and introduce more variability in the light chains and see what effect that has on the heavy chain binding.

**Public light chains**

The CDR3 region of each of the three light chains above has been identified in the ([DeKosky et al., 2015](#_ENREF_1)) paper as being a public light chain (i.e. it is very commonly used in many people). All three of the chosen light chains contain precisely the same CDR3 nucleotide sequence as three public light chains identified in this paper. IGKV3-20,J2 was identified in this paper as the most frequently used public light chain.

I then chose the other two from the list of public light chains, but I tried to find ones that had opposite properties so they would hopefully be as different as possible.

I also checked that VK3-20, VL1-44 and VK3-11 have been found to bind to the heavy chains containing the VH3 genes ([DeKosky et al., 2015](#_ENREF_1)) (supplemental material).

**Kappa/lambda, isoelectric point and aspartic acid usage**

It has been shown in studies using mice ([Li et al., 2001](#_ENREF_3)), and recently humans ([Kalinina et al., 2014](#_ENREF_2)), that the presence of aspartic acid (D) in the CDR regions and a low isoelectric point are common characteristics of “editor light chains” (i.e. light chains that are able to change the binding specificity of DNA-specific IgH chains). However, Wardemann did not find evidence of this in her 2004 paper, but she did find that lambda light chains were better at rescuing autoreactive heavy chains than kappa light chains ([Wardemann et al., 2004](#_ENREF_4)).

Therefore, I chose a lambda light chain that had a low pI and two aspartic acids in the CDR3 region (IGLV1-44,J2) and also a kappa light chain that had a high pI, no aspartic acids in the CDR3 region and an arginine in the CDR3 region (IGKV3-11,J4).

Additionally, in the ([Wardemann et al., 2004](#_ENREF_4)) paper, KV3-20, KV3-11 and LV1-44 were all tested as editor light chains (and the CDR3 regions were very similar to those above). KV3-20 and KV3-11 sometimes rescued heavy chains, but mostly not, whereas LV1-44 often did rescue heavy chains. None of the light chains tested in this paper ever rescued any heavy chains.

**Question:**

In the literature, the pI of all three CDR regions is often taken into account, whereas the pI in our dataset is just for the light chain CDR3 region. Does this matter?

**Chosen public light chain values in comparison to the light chain dataset as a whole:**











**Unexpected:**

From the literature, I was expecting low pI and D usage to be preferred (as they are associated with rescuing autoreactive heavy chains), so I was expecting, in our dataset, to see lower pI and higher D frequency in transitional and naïve cells compared to immature. Surprisingly, these two properties were among the only ones to be significantly different between the cell types, and it was in the opposite direction to what was expected (i.e. transitional and naïve cells had higher mean pI and lower mean D frequency than immature cells). Could this perhaps indicate that editor light chains are found a lot in the immature repertoire as they are trying to rescue autoreactive heavy chains, but many of them fail and therefore do not make it into the periphery? So maybe in the case of the light chains it may not be as simple as looking for the features that appear to be selected against?

**References**

DEKOSKY, B. J., KOJIMA, T., RODIN, A., CHARAB, W., IPPOLITO, G. C., ELLINGTON, A. D. & GEORGIOU, G. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nature Medicine,* 21**,** 86-91.

KALININA, O., WANG, Y., SIA, K., RADIC, M., CAZENAVE, P. A. & WEIGERT, M. 2014. Light chain editors of anti-DNA receptors in human B cells. *Journal of Experimental Medicine,* 211**,** 357-364.

LI, H., JIANG, Y. F., PRAK, E. L., RADIC, M. & WEIGERT, M. 2001. Editors and editing of anti-DNA receptors. *Immunity,* 15**,** 947-957.

WARDEMANN, H., HAMMERSEN, J. & NUSSENZWEIG, M. C. 2004. Human autoantibody silencing by immunoglobulin light chains. *Journal of Experimental Medicine,* 200**,** 191-199.