

Data S1. Lemur lambda light chain and surrogate light chains found by SPLASH, related to Figure 4.

As detailed in Methods, we attempted to identify light chain variable regions in lemur B cells where the BASIC pipeline could not. We were successful in identifying a full variable region by extension (using `grep` of raw reads) of an initial seed consensus sequence found by SPLASH. Below we give the sequence identified, and the NCBI IgBlast report for it. IgBlast uses human Ig reference sequences (lemur Ig regions are not well annotated) so it is uncertain which differences are due to lemur and which to hypermutation, however there appears to be a high mutational load in this variable region, which may be why BASIC could not identify it.

This is the cell that had a full V-region:

cell # MAA001400_B109012_I1_S193

```
>cons1-MAA001475_B112525_O18_S354_R1_001  
GGCCTTGGGCTGACCTAGGACGGTGAGCTGGGTCCCTCTGCCGAAGACAAACATCGACTGAGGCTCAGACCAA
```

```
>cell_1_lambda_VJ_assembled  
GTGAGTCCCCAGGAACCAGAGCTCACAGGAGCCTCCACCATGGCCAGGGCTCCTGTCCTC  
GTCCCTCTCCTCGCTCTCTGCTCAGGGTCTGGGCACAGTCTGGACTCACCCAGGAAGCC  
TTGGTGTGAGGGTCTGTGGGACACAAGGTCACCCTGTCCTGCGCTGGACACAGCAACAGT  
GTTGGTTTCAATTTGGGGTGGACTGGTGCCAGCAGGTTTCTGGTGGTGCCCCAAAACCTGTG  
ATGCTCGGGACAACCTCGGCCCTCAGGGATCCCCGATCGCTTCTCCGGCTCCAGGTCTGGG  
AACACGGCCTCTTTGACCATCTCGGACCTCCCGGACCTCCAGCCGGAGGACGAGGCTGAC  
TATTACTGTTCAACTTGGGACAGAACCTGCGTGCTCATGTGTTTCGGCGGTGGGACCAAG  
GTGATCGTCTAGGTGAGCCCAAGGCCGCCCTCGGTACGCTGTTCCCGCCCTCCTC
```

Protein translation:

```
>cell_1_lambda_VJ_assembled  
MARAPVLVPLLALCSGWAQSGLTQEALVSGSVGHKVTLSACAGHSNSVGSFGVDWCQQVP  
GGAPKTVMLGTTTRPSGIPDRFSGSRSGNTASLTISDLPLQPEDEADYYCSTWDRTLRAH  
VFGGGTKVIVVGQPKAAPSVTLFPPS
```

Below is the NCBI IgBlast report for the full V-region. Coding differences from germline gene are shown in **magenta**.

Top V gene match	Top J gene match	Chain type	stop codon	V-J frame	Productive	Strand	V frame shift
IGLV2-14*05	IGLJ6*01	VL	No	In-frame	Yes	+	No

V region end	V-J junction	J region start
TACTG	TTCAACTTGGGACAGAACCCTGCGTGCTC	ATGTG

```

<-----FR1-IMGT-----><-----CDR1
  Q S G L T Q E A L V S G S V G H K V T L S C A G H S N S V
V 69.1% (192/278) Query_1 97 CAGTCGGACTCACCAGGAAGCCTTGGTGTGAGGGTCTGTGGACACAAGTCCACCTGTCTCGCGTGGACACAGCA---ACAGTGT 183
  .....CC..G..T...CCT...CC.....T.....CCT.....GTC.A.....A.C.....A.....AC.....GTG..GT..G. 90
  Q S A L T Q P A S V S G S P G Q S I T I S C T G T S S D V G
V 68.7% (191/278) IGLV2-18*01 1 .....CC..G..T...CCTC...CC.....C.....CCT.....GTCA.....A.C.....A.....AC.....---GTGAC... 87
V 68.7% (191/278) IGLV2-18*02 1 .....CC..G..T...CCTC...CC.....C.....CCT.....GTCA.....A.C.....A.....AC.....---GTGAC... 87

-----IMGT-----><-----FR2-IMGT-----><CDR2-IM><-----
  G S F G V D W C Q Q V P G G A P K T V M L G T T R P S G
V 69.1% (192/278) Query_1 184 GGTCATTTG--GGTGGACTGGTCCAGCAGGTTCTCTGGTGGTCCCCAAAACGTGATGCTCGGG--ACAACCTCGGCCCTCAGGG 267
  ...ATAACT--AT..CTC...A...A...CAC..A..CAAA.....CTCA...TTAT.A.GTC.GT.A..... 177
  G Y N Y V S W Y Q Q H P G K A P K L M I Y E V S N R P S G
V 68.7% (191/278) IGLV2-18*01 88 ...AGT.A.AACC.T..CTC...A.....CCC..A..CACA.....CTCA...TTAT.A.GTC.GT.A..... 177
V 68.7% (191/278) IGLV2-18*02 88 ...AGT.A.AACC.T..CTC...A.....CCC..A..CACA.....CTCA...TTAT.A.GTC.GT.A..... 177

-----FR3-IMGT-----
  I P D R F S G S R S G N T A S L T I S D L P D L Q P E D E A
V 69.1% (192/278) Query_1 268 ATCCCCGATCGCTTCTCCGGTCCAGTCTGGGAACACGGCCTCTTTGACCATCTCGGACCTCCCGACCTCCAGCCGGAGGACGAGGT 357
  G...T.....T.....T.....A.....C.....CC.....-----T.GG.....G.T..... 258
  V P D R F S G S K S G N T A S L T I S G L Q A E D E A
V 68.7% (191/278) IGLV2-18*01 178 G...T.....T..G...A.....C.....CC.....-----T.GG.....G.T..... 258
V 68.7% (191/278) IGLV2-18*02 178 G...T.....T..G...A.....C.....CC.....-----T.GG.....G.T..... 258

-----><-----CDR3-IMGT-----><-----FR4-IMGT----->
  D Y Y C S T W D R T L R A H V F G G G T K V I V V
V 69.1% (192/278) Query_1 358 GACTATACTGTCAACTTGGGACAGAACCCTGCGTGCATGTGTGCGCGGTGGGACCAAGGTGATCGTCGTAG 433
  ..T.....-----TGGTGGTGGGACCAAGGTGATCGTCGTAG 269
  D Y Y
V 68.7% (191/278) IGLV2-18*01 259 ..T.....----- 269
V 68.7% (191/278) IGLV2-18*02 259 ..T.....----- 269
J 90.6% (29/32) IGLJ6*01 3 -----A...C.....C..... 34
J 88.2% (30/34) IGLJ3*02 5 -----A.....C.....C..... 38
J 87.1% (27/31) IGLJ2*01 8 -----A.....C.....C..... 38

```

Here are the BLAST matches of consensus from two cells to the unique region of IGLL1/5, one of the surrogate light chains. This suggests that these two cells have not yet rearranged their light chain.

cell # MAA001475_B112525_C4_S52

```
>cons1-MAA001475_B112525_C4_S52_R1_001
GGCCTTGGGCTGACCTGCTCCACGGGATCCGCGGCACTGGACCGGCTGCTTCCCGCCCGGGCTCCGGGGCC
```

Consensus sequence above is the original strand, but BLAST was done with its reverse-complement, the sense strand (the Query below). The alignment indicates a 1-nt insertion in the consensus relative to IGLL5 (which would cause a frameshift).

```
PREDICTED: Microcebus murinus immunoglobulin lambda-like polypeptide 5 (LOC105882024),
transcript variant X2, mRNA
Sequence ID: XM_012784097.2 Length: 821 Number of Matches: 1 Range 1: 193 to 250
Score Expect Identities Gaps Strand
91.6 bits(49) 1e-14 56/59(95%) 1/59(1%) Plus/Plus
      G P G A R A G S S R S S A A D P V E Q
      A P E P G R E A A G P V P R I P W S R
Query 1 GGCCCCGAGCCCCGGCGGGAAGCAGCCGGTCCAGTCCGCGGATCCCGTGGAGCAGGT 59
      |||||
Sbjct 193 GGCCCTGGAGCCCCGGCGGGAAGCAGCCGGTCCAG-CCCAGCGGATCCCGTGGAGCAGGT 250
      G P G A R A G S S R S S P R I P W S R
```

Protein comparison of the mouse lemur IGLL5 above to human IGLL5 (where domains are identified) shows that the above region clearly lies in the IGLL5 N-terminal unique region. The matching part above is shown in blue.

```
Query 12 LRLGKGQVGCDA PK--GPGPRLRWPLLLLGLAVGTHGFLSSTEAPRSRAPGPGARAGSSR 69
      +R  QVGC+ P+  GPGPR RWPLLLLGLA+  HG L  AP+S  P  PGA  GSSR
Sbjct 1  MRPKTQVGCETPEELGPGPRQRWPLLLLGLAMVAHGLLRPMVAPQSGDPPGASVGS SR 60

Query 70 SSPRIPWSRFLLQPSPRGAGARCWPRGFWSEPQSLWYIFGRGTQLTILGQPKAAPSVTLF 129
      SS R  W R LLQSP+ A  RCWPRGFWSEPQSL Y+FG GT+++T+LGQPKA P+VTLF
Sbjct 61 SSLRSLWGRLLLQPSPRADPRCWPGRGFWSEPQSLCYVFGTGTKVTVLGQPKANPTVTLF 120
                                 J-region             C-region

Query 130 PPSSEELQANKATLVCLMSDFYPGA VSVAWKADGSAVTQGVETTQASKQSNKYAASSYL 189
      PPSSEELQANKATLVCL+SDFYPGA V+VAWKADGS V  GVETT+ SKQSN KYAASSYL
Sbjct 121 PPSSEELQANKATLVCLISDFYGA VTVAWKADGSPVKAGVETTKPSKOSNNKYAASSYL 180

Query 190 SLSPAQWKAGGRFSCQVTHEGSTVEKTVAPA ECA 223
      SL+P QWK+  +SCQVTHEGSTVEKTVAP EC+
Sbjct 181 SLTPEOWKSHRSYSCOVTHEGSTVEKTVAPTECS 214
```

