

1 **Supplementary Information to:**

2 Template-based assembly of proteomic short reads for *de novo* antibody sequencing and
3 repertoire profiling

4

5 **Authors:**

6 Douwe Schulte¹, Weiwei Peng¹, Joost Snijder^{1*}

7

8 * corresponding author: j.snijder@uu.nl

9

10 **Affiliations:**

11 ¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research
12 and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH
13 Utrecht, The Netherlands

14

15 **Contents**

16 Legend of Supplementary Data with Stitch HTML reports

17 CDRH3 reconstructions of Herceptin and anti-FLAG-M2

18 Details of sequence variations in F59 Fab

19 Description of multiple myeloma light chain sequences.

20

21

22

23

24

25 **Supplementary Data.**

26 All reported Stitch analyses are provided with the complete output reports. The supplementary
27 data can be unzipped to browse the interactive HTML reports. The corresponding analysis
28 parameters are provided under the 'Batch File' menu in each report.

29

A) Herceptin

```

template    ...DTAVYYCAR                YFDYWGQGTLTVSS...
de novo     DTAVYYCSRWGG      SRWGGDGFYAMDYWGQGTLTVSS

overlap     DTAVYYCSRWGG
            SRWGGDGFYAMDYWGQGTLTVSS

ref         ...DTAVYYCSRWGGDGFYAMDYWGQGTLTVSS...

```

B) anti-FLAG-M2

```

template    ...DSAVYYCAR                YFDYWGQGTTLTVSS...
de novo     DSAVYYCAREKFGY      VYYCAREKFGYDYWGQGATLTVSS

overlap     DSAVYYCAREKFGY
            VYYCAREKFGYDYWGQGATLTVSS

ref         ...DSAVYYCAREKFGYDYWGQGATLTVSS...

```

31

32

Supplementary Figure S1. Detailed view of CDRH3 reconstruction by Stitch of the antibodies

33

shown in main text figure 2. Shown are the selected template sequences, aligned reads, found

34

overlap and known reference sequence.

35

A) F59 Heavy Chain from Refined+Decoy run

E^{PE}V^QL VESG^GG^L V^QP^GR^TSLR^LSCEA^SG^FT^FDDY
A^MH^WV^RQAP^GR^GL^EW^VS^GL^SW^SS^DN^LAY^AD^SV
K^GR^FT^LS^RD^NA^KN^SL^YL^QM^NS^LR^LD^DT^AF^YY^C
A^KD^VR^PY^YD^FW^YF^DW^GQ^GT^LV^TV^SS^AS^TK^GP

B) F59 Light Chain from Refined+Decoy run

E^SL^LT^QP^AS^VS^GS^PG^QS^LT^LS^CT^GT^SS^DV^GY^NY^VS
W^YQ^HP^GK^AP^KL^LL^SE^VS^DR^PS^GV^SS^RF^SG^SK^SG^NT
A^SL^TL^SG^LQ^AE^DE^AD^YF^CS^SY^TD^LS^TL^VV^FG^GG^TK^L

36

37

38

39

40

Supplementary Figure S2. Sequence logos of heavy (A) and light (B) chains of F59 Fab from fractionated serum sample from Refined+Decoy Stitch runs. Sequencing error are highlighted in red boxes.

A)

>Patient 1 (94.5% Δ -18.3 Da)
de novo | ESAL TQPR SVSGSPGQSVTISCTGTSSDVGGYNYVSWYQQHPGKAPKLMYDVKRPSGV PDRFSGSKGTTASLTISGLQAEDEADYYCCSYAGLDL FVLFGGGKLT V
reference | GPDL TQPR SVSGSPGQSVTISCTGTSSDVGGYNYVSWYQQHPGKAPKLMYDVKRPSGV PDRFSGSKGTTASLTISGLQAEDEADYYCCSYAGIDIFVLFGGGKLT V
*** * * *

>Patient 5 (A: 99.0% Δ -30.4 Da) & (B: 87.9%)
de novo | EIVLTQSPGTL S LSPGERATL SCRASQSVSSYLAWYQQKPGQAPRLLIYDASTRATGIPDRFSGSGS GDFLLTISSLEPEDFAMYYCQQYGRSPYTFGPGTKVDI
reference A | EIVLTQSPGTL S LSPGERATL SCRASQSVSSYLAWYQQKPGQAPRLLIYDASTRATGIPDRFSGSGS GADFLLLTISSLEPEDFAMYYCQQYGRSPYTFGPGTKVDI
reference B | EIVLSQSPD T L S LSPGERATL SCRAKSVSSNYVAVYQQKPGQAPRLLIYDAFTRATGIPDRFSGSGS ETDYTLTISTLEPEDFAMYYCQQYGRSPYTFGPGTKVDI
* * ** * * ** * *

>Patient 6 (100.0% Δ -0.5 Da)
de novo | DIQMTQSPSSL SASVGDRVITCRASEISSYVNWYQQKPGKAPKLLIYDASSLQSGVPPRFSGSASGTDFTLTISSLQPEDFATYYCQQSYSTPITFGQGTRLEI
reference | DIQMTQSPSSL SASVGDRVITCRASEISSYVNWYQQKPGKAPKLLIYDASSLQSGVPPRFSGSASGTDFTLTISSLQPEDFATYYCQQSYSTPITFGQGTRLEI

>Patient 7 (94.3% Δ 60.8 Da)
de novo | DIQMTQSPSSL SASVGDRVITICQASQDI AKYLNWYQQKPGKPPKLLIYDTSNLETGVP SRFSGSGS GDTFTFTINSLQPEDIATYYCQQYDDFPLTFGPGTKVDI
reference | DIQMTQSPSSL SASVGDRVITICQASQDL AKYLNWYQQKPGKPPKLLIYDTSNLETGVP SRFSGSGS NG-GGTDFTFTINSLQPEDLATYYCQQYDDFPLTFGPGTKVDI
* **** *

>Patient 8 (99.1% Δ -0.5 Da)
de novo | DIQMTQSPSSL SASVGDRVITICRASQSSISLAWYQQKPGKAPKLLIYDASSLETGVP SRFSGSGS GTEFTLTISSLQPDFFATYYCQHYNYSYSLTFGGGKVEI
reference | DIQMTQSPSSL SASVGDRVITICRASQSSISLAWYQQKPGKAPKLLIYDASSLETGVP SRFSGSGS GTEFTLTISSLQPDFFATYYCQHYNYSYSLTFGGGKVEI
*

>Patient 13 (92.8% Δ -12.5 Da)
de novo | EPAL TQPPSVSGAPGQRVITISCTGSSNIGAGWDVHWYQQLPGTVPKLLIYADRNRPSGVPERFSGSKSGTDAALAIAGLQAEDEADYYCQSYDSALSGFYVFGTGKVLV
reference | EAPL TQPPSVSGAPGQRVITISCTGSSNLGAGWDVHWYQQLPGTVPKLLIYADRNRPSGVPERFSGSKSGTSA TVAIAGLQAEDEADYYCQSYDSALSGFYVFGTGKVI V
*** * * * *

>Patient 15 (99.1% Δ -0.8 Da)
de novo | DIQMTQSPSSL SASVGDAVITICRASQSNVWLA WYQQKPGKPPKLLIYEASNLESGVPSRFSGSGS GTEFTLTISSLQPDFFATYYCQQYNSYPYTFGGGAKLEI
reference | DIQMTQSPSSL SASVGDAVITICRASQSNVWLA WYQQKPGKPPKLLIYEASNLESGVPSRFSGSGS GTEFTLTISSLQPDFFATYYCQQYNSYPYTFGGGAKLEI
*

>Patient 18 (99.1% Δ -0.7 Da)
de novo | DIQMTQSPSSL SASVGDRVITICQASRDISNYLNWYQQKPGKAPMLLIYAASNLTGVP SRFSGSGS GDTFTFTINSLQPEDIATYYCQQYGNLPLTFGGGKVEI
reference | DIQMTQSPSSL SASVGDRVITICQASRDISNYLNWYQQKPGKAPMLLIYAASNLTGVP SRFSGSGS GDTFTFTINSLQPEDLATYYCQQYGNLPLTFGGGKVEI
*

>Patient 19 (99.1% Δ -0.8 Da)
de novo | DIQMTQSPSSL SASVGDRVITICQASQDIGNYLNWYQQKPGKAPRLLIYDASDLEEGVPSRFSGSGS GDTFTFTINSLQPEDFATYYCQQYHTLPPLTFGGGKVDV
reference | DIQMTQSPSSL SASVGDRVITICQASQDLGNYLNWYQQKPGKAPRLLIYDASDLEEGVPSRFSGSGS GDTFTFTINSLQPEDFATYYCQQYHTLPPLTFGGGKVDV
*

>Patient 20 (99.1% Δ -1.8 Da)
de novo | DIQMTQSPSSL STSVGDRVITICRASQSIRTWLA WYQQKPGKAPKLLIYKASTLETGVP SRFSGSGS GTFMFTLTISSLQPEDFATYYCQQYNDYSGTFGGGKLEI
reference | DIQMTQSPSSL STSVGDRVITICRASQSIRTWLA WYQQKPGKAPKLLIYKASTLETGVP SRFSGSGS GTFMFTLTISSLQPEDFATYYCQQYNDYSGTFGGGKLEI
*

B)

EIVLTQSPGTL S LSPGERATL SCRASQSVSSYAWYQQKPGQAPRLLIYDASTRATGIPDRFSGSGS GADFLLLTISSLEPEDFAMYYCQQYGRSPYTFGPGTKVDL K

41

42

43

Supplementary Figure S3. A) Stitch reconstruction of light chains from multiple myeloma patient urine (Chamot-Rooke and colleagues, ref 43). CDRs are annotated, sequence conflicts

44 highlighted by an asterisk (*) and sequence identity in parentheses, along with the mass error of
45 the predicted Stitch sequence compared to the experimentally determined mass of the intact light
46 chain. B) Sequence logo of P5 light chain from Stitch analysis, with highlighted variants in red
47 boxes.