Protein abundances from mass spectrometer were quantified by
MaxQuant 1.6.0.16 (ref: Cox, 2008) using 2-plex labeling (`Arg10` and `Lys8`).

The search was run against human UP000005640 proteome (version GCA_000001405.26)
from UniProt (ref: The UniProt Consortium, 2017).
Search parameters were set up to allow variable oxidation on methionines
and acetylation on N-termini as well as fixed carbamidomethylation on cysteines.
Tryptic peptides with up to 2 missed cleavages were considered for analysis
and tolerance settings were set to Orbitrap instrument.
Co-fragmented peptide identification and matching between runs has been enabled.
For protein quantification, minimum ratio count was set to 1
and remaining settings left default.

The normalised H/L ratios in `proteinGroups.txt` file output by MaxQuant
have been transformed to log2 scale for further processing.
Each protein in the dataset containing Ankyrin repeat-containing domain
has been identified using information from InterPro superfamily IPR036770 (ref:Finn, 2017).

Twenty six proteins that were marked as either "Potential contaminant",
"Reverse" or "Only identified by site" by MaxQuant were dropped from the dataset
(i.e. 1.81% of data). Further 76 proteins (5.39% of the remaining data),
have been dropped because they had only forward, or reverse ratio, but not complete pair.

Out of the remaining data points we were left with 12 ANK-repeat containing proteins:
BCORL1 (2 distinct protein groups), BCOR, RFXANK, MPHOSPH8, EHMT1, EHMT2,
TONSL, ANKRD32, ANKRD11, BARD1 and NFKBIL1.

Four protein families were then annotated manually:
(1) BRCA1 family - BRCA1, FAM175A, BARD1, BRE, UIMC1, FAM175A, BRCC3, BABAM1;
(2) RAD18/SLF family - RAD18, ANKRD32, FAM178A;
(3) TONSL family - TONSL;
(4) ORC family - ORC1, ORC2, ORC3, ORC4, ORC5, LRWD1

Proteins were labelled their gene names, except in the following cases
where we substituted the following names to their synonyms for visualisation purposes:

```
(1) BRCA1 members:

  BRE: BRCC45,
  UIMC1 RAP80,
  FAM175A: ABRAXAS1
  BRCC3: BRCC36
  BABAM1: MERIT40
```

```
(2) RAD18/SLF members:

  RAD18: RAD18/RNF73
  ANKRD32: SLF1/ANKRD32
  FAM178A: SLF2/FAM178A

(3) Histones:

  HIST1H2AD;HIST1H2AJ;HIST1H2AH;H2AFJ;HIST1H2AG;HIST2H2AC;HIST2H2AA3;H2AFX: H2A¹
  HIST3H2A;HIST1H2AB;HIST1H2AC: H2A²
  H2AFV;H2AFZ: H2A³
  HIST1H3A;HIST2H3A;HIST3H3;HIST2H3PS2: H3
  HIST1H4A: H4
  HIST1H2BK;H2BFS: H2B¹
  HIST1H2BC: H2B²
```

This data was then plotted on a scatter plot, with log2 forward ratio on x,
and reverse ratio on the y axis.
Labels for BRCA1, RAD18, TONSL, ORC groups were automatically drawn and colour-coded.
Proteins containing ANK repeat domain were highlighted in bold,
as well as with an asterisk near their label.

Additional proteins that have either reverse or forward ratio
above the threshold of 1.4 were also labelled along with
LSM7, PAX6, TLX3, H2A², DNMT1, CPSF7, TTN, INO80E, E2F6, and FIZ1 which were
included to the plot manually,
as either they clustered visually close to proteins of interest, or were significant outliers.

Two versions of the plot were generated.
In the first version, the axes are cropped to the log2 range (-4.5, 4.5), which
excluded the significant outliers SPTA1 (x=5.0295, y=-0.0470),
H4 (x=-7.7283, y=-5.0210), H3 (x=-6.7843, y=-6.8198) and STAG2 (x=3.5792, y=7.8352)

The second version (intended for supplementary reference),
was generated without the cropping of the axes.

We also provide an .xlsx file with the processed dataset.
Raw data files (both the output from mass spectrometer and MaxQuant)
will be uploaded to PRIDE.

References

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized

p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology, 26(12), 1367–72. http://doi.org/10.1038/nbt.1511

Robert D. Finn, Teresa K. Attwood, Patricia C. Babbitt, Alex Bateman, Peer Bork, Alan J. Bridge, Hsin-Yu Chang, Zsuzsanna Dosztányi, Sara El-Gebali, Matthew Fraser, Julian Gough, David Haft, Gemma L. Holliday, Hongzhan Huang, Xiaosong Huang, Ivica Letunic, Rodrigo Lopez, Shennan Lu, Aron Marchler-Bauer, Huaiyu Mi, Jaina Mistry, Darren A. Natale, Marco Necci, Gift Nuka, Christine A. Orengo, Youngmi Park, Sebastien Pesseat, Damiano Piovesan, Simon C. Potter, Neil D. Rawlings, Nicole Redaschi, Lorna Richardson, Catherine Rivoire, Amaia Sangrador-Vegas, Christian Sigrist, Ian Sillitoe, Ben Smithers, Silvano Squizzato, Granger Sutton, Narmada Thanki, Paul D Thomas, Silvio C. E. Tosatto, Cathy H. Wu, Ioannis Xenarios, Lai-Su Yeh, Siew-Yit Young and Alex L. Mitchell (2017). InterPro in 2017 — beyond protein family and domain annotations. Nucleic Acids Research, Jan 2017; doi: 10.1093/nar/gkw1107

UniProt: the universal protein knowledgebase, Nucleic Acids Research, Volume 45, Issue D1, 4 January 2017, Pages D158–D169, https://doi.org/10.1093/nar/gkw1099