Grant Impact Report

Name	Daniel Hodson
Job title	Honorary Consultant Haematologist
Organisation	Wellcome-MRC Cambridge Stem Cell Institute and Department of Haematology, University of Cambridge
Co-investigators (if applicable)	
Grant awarded	Early Stage
Year awarded	2021
Date started	1/1/2022
Date completed	31/12/2022
Total amount expended (£)	£19,628

This is the BSH grant impact report form. Please enter the full grant details above, and fill out the form below. The form should be completed electronically and sent to grants@b-s-h.org.uk. Please note that the report can only be accepted if all sections have been completed in full.

In addition: Please include a recent photo of yourself.

Your grant report and photo may be published in our communications materials, including our website and social media platforms.

To see previously published grant impact profiles, please visit our website.

British Society for Haematology
Listening • Learning • Leading

1. Please summarise what the grant enabled you to achieve; what would not have been possible without the funding? (Up to 500 words)

This Early Stage Grant allowed me to develop a set of unexpected but exciting findings that had arisen in the lab related to the identification of large numbers of potential "micropeptides" translated from previously unrecognised small open reading frames (smORFs) within the human genome.

Specifically, this grant allowed me to retain within my lab an extremely talented junior clinical academic, Dr Joanna Krupka by bridging her salary between grants. This was essential to keep the momentum of this project and to retain an excellent and skilled bioinformatician within my research group.

This project was joint funded by the British Society of Haematology, along with the Addenbrooke's Charitable Trust and also the RoseTrees Trust. Together these grants provided 12 months of funding for Dr Krupka. During this time, she completed the extremely complex computational analysis to prioritise the most promosing smORFs. This allowed us to design and construct a targeted CRIPSR library, which we deployed across five lymphoma cell lines and in non-malignant human B cells grown ex vivo. These experiments have provided compelling evidence for the biological functionality of large numbers of previously unknown micropeptides within human lymphoid cells.

2. Briefly describe the aims and intended outcomes of this project. Please clearly indicate if there is any sensitive information in this report that should remain confidential for now. (Up to 300 words)

Micropeptides are small protein molecules generated within cells. They are encoded within the genome by small "open reading frames" (smORFs). A relatively small number of translated micropeptides have already been described in the literature but until recently, technical limitations have precluded the experimental identification of these micropeptides or their smORFs. New technology, including the so-called ribosome foot printing is beginning to reveal the extent of translated smORFs within the genome and biological importance of the micropeptides they encode. Our own ribosome foot printing data performed as part of a separate project unexpectedly revealed thousands of translated smORF encoding potential micropeptides in normal and cancerous B cells.

The aims of this proposal were to filter our large list of potential non-canonical ORFs using advanced bioinformatic techniques applied to ribosome profiling, RNA seq and mass spec proteomic data. Based on this analysis we planned to design a customised CRISPR library to interrogate the B cell micropeptidome to screen for micropeptides with evidence of functional activity in human lymphoid cells. These analyses and tools would then form the basis of future, larger funding applications.

3. **Describe the key outcomes to date, including whether this grant has resulted in further research. Please summarise your conclusions. (Up to 600 words)**

- Large proteomic datasets were downloaded and analysed to screen for smORF with protein level evidence that they encode micropeptides. This peptide data was then integrated with smORF prediction scores, conservation across species and evidence of domain structure to rank potential smORFs into those most likely to encode functional peptides.
- Using the above list we designed, cloned and verified a custom CRISPR library designed to target the top 1600 small ORFs.
- Applied CRISPR library to lymphoid cell lines and normal human B cells ex vivo to identify 120 smORFs that appear to be essential for the survival of all lines after CRISPR knockout, suggesting these encode functional micropeptides.
- Validated findings from the CRISPR screen using individual gRNAs in 26 top scoring smORFs.
- Applied a bioinformatic method of translational co-regulation (Parsimonious Gene Correlation Network Analysis [PGCNA]) to predict function of these micropeptides.

All project aims have been achieved. Most importantly we have now moved from a very large list (>43K) of potentially translated smORFs to a focused list of 20 high-confidence smORFs that we believe encode functional micropeptides that appear to be essential for the survival of lymphoid cells. These have all been individually validated. The PGCNA analysis suggest that these micropeptides may participate in various core cellular processes, including cell cycle regulation or metabolism. Therefore, the current work has confirmed the significance of our previous RiboSeq findings and laid the ground for further, more focused, experiments.

By performing individual smORF validation of the top 20 CRISPR screen hits and by incorporating the computational PGCNA analysis we are now further ahead than we expected to be at the end of this grant. We hypothesise that identified micropeptides may compose an entirely new family of regulatory molecules and therefore potential therapeutic targets.

A selection of the key data highlights from this project are presented as an appendix to this report.

4. List published papers, oral and/or poster presentations as a result of this grant.
Include manuscripts in preparation or in submission / under review, prefaced by an asterisk.

Dr Joanna Krupka. *In search of lost ORFs: systematic map of noncanonical Open Reading Frames essential for lymphoid cells.* Poster presentation at 26th Annual Meeting of the RNA SOCIETY - On-line. **Recognition for excellence in RNA research and as a recipient of an RNA SOCIETY Poster Award at RNA 2021**

Dr Joanna Krupka. *In search of lost ORFs: systematic map of noncanonical Open Reading Frames essential for lymphoid cells.* Oral presentation at Wellcome-MRC Cambridge Stem Cell Institute Retreat June 2022. This presentation won the **prize for best student presentation**.

Dr Joanna Krupka. *In search of lost ORFs: ultra-sensitive map of noncanonical Open Reading Frames essential for lymphoid cells.* Oral presentation at Cambridge Lymphoma Biology International Symposium September 2022

Dr Joanna Krupka. *An ORF-full big Universe of small peptides*. Oral presentation at UK-Germany Lymphoma Meeting October 2022.

5. Did any patent applications arise from this work? (If yes, please detail. Up to 200 words)

No

6. Were you successful in any further grant applications as a result of this work? (If yes, please detail. Up to 200 words)

This work contributed to the successful award of an Academic Clinical Fellowship to Dr Joanna Krupka (awarded January 20223). The project will be taken forward as the research component of this fellowship.

This work also provides the pilot data currently being used to prepare intermediate fellowship applications by Dr Krupka to CRUK and Wellcome.

7. Did new collaborations arise from this work? (If yes, please detail. Up to 400 words)

The project has led to two new collaborations:

The first is with **Prof Ryan Morin** (Simon Fraser University Vancouver and British Columbia Cancer), who will use his extensive whole genome sequencing data set to ask if our predicted small ORFs are associated with recurrent mutation in lymphoma, or alternatively if they appear protected from the background mutation seen in non- coding regions that might disrupt the expression of a coding gene. Either finding would allow us to prioritise small ORFs that might play a role in lymphoma, either promoting or suppressing lymphoma development.

The second is with **Dr Inaki Subero** (IDIBAPS, Barcelona) who will correlate our predicted small ORF regions with his extensive database of epigenetic marks in a range of lymphoid malignancies.

8. What was the funding amount you received and how was it actually spent? (detail item/activity and amount spent in pounds)

Salary for Dr Joanna Krupka (Clinical postdoctoral fellow)

£15,540

Consumable costs including purchase of oligo library to clone the micropeptide focused CRISPR library

£4,088

Total Spend

£19,628

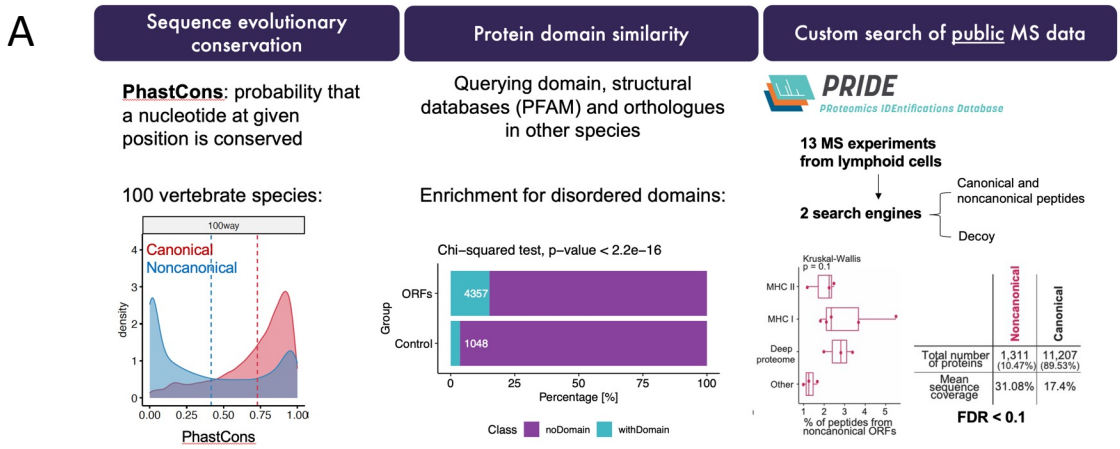
9. What are the future research priorities in this area?

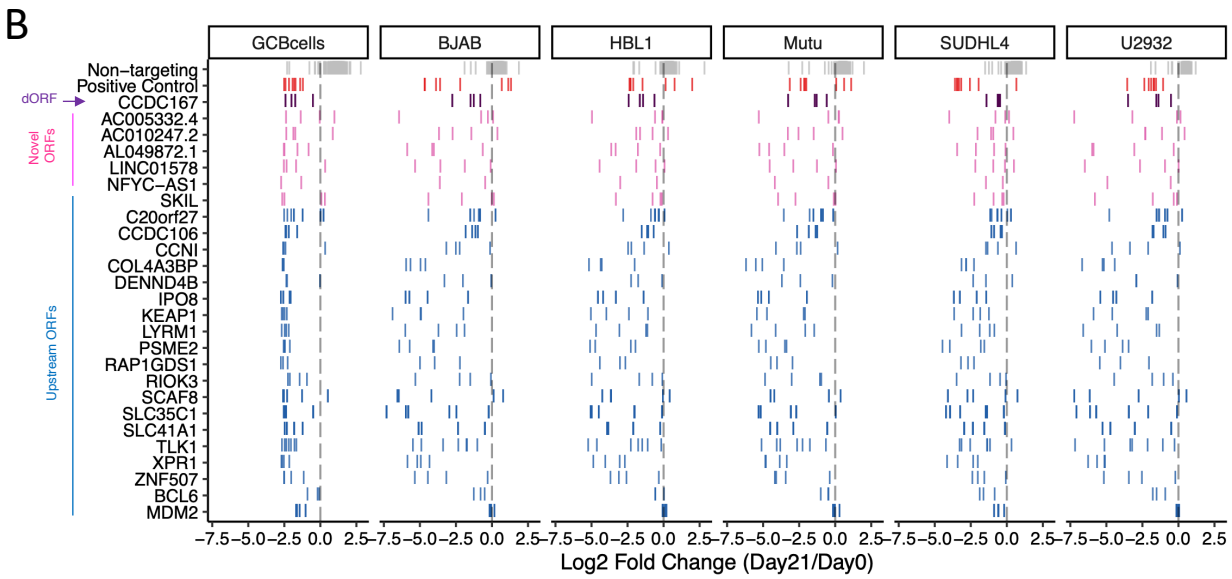
This research has revealed the unexpectedly large extent of translation that occurs in human lymphoid cells beyond what we consider canonical protein coding transcripts. We have identified thousands of potentially translated micropeptides and provided evidence that many of these encode a functional protein product that is in some way essential to the viability of lymphoid cells

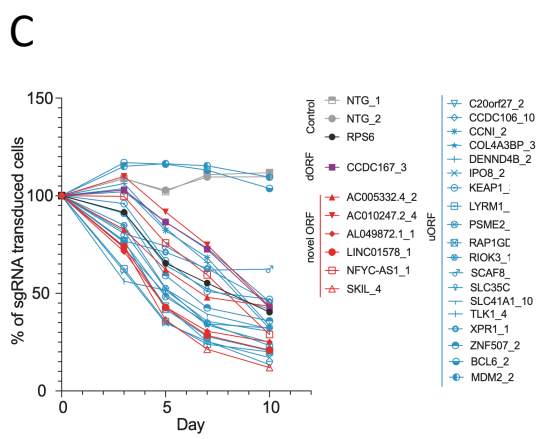
Since the field of micropeptides is extremely new, we began this project with a very large number of small ORFs each potentially encoding micropeptides, but with no clue as to the likely function or which to prioritise for further mechanistic investigation. Having narrowed the list of over 43 thousand small ORFs down to 20 top hits that we believe to encode functional micropeptides essential for lymphoid cells, the immediate questions arising are: 1) what is the function of these micropeptides, 2) which cellular processes do they contribute to, and 3) how are micropeptide-regulated functions altered during the transformation from normal to malignant lymphocyte.

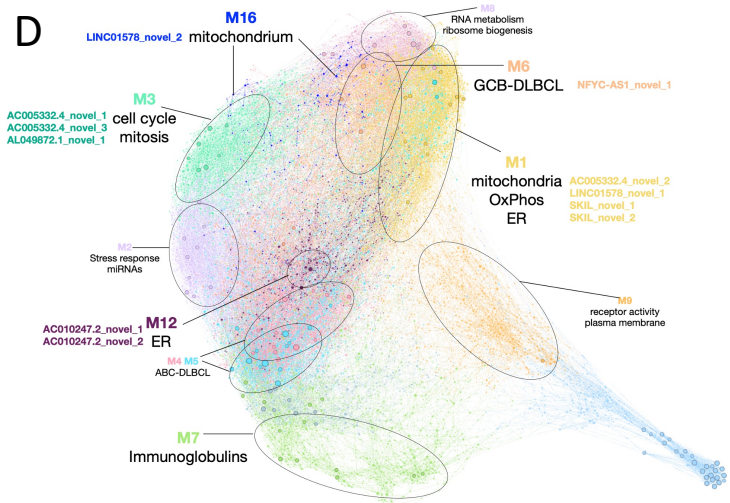
These questions will be taken forward and will form the basis of further fellowship application within the Hodson research group.

We hope ultimately to identify micropeptides that are selectively essential for lymphocytes or even more excitingly, selectively essential for lymphoid malignancy. This would provide represent a cancer-selective vulnerability that could then be exploited as a basis for therapy.









Summary of data generated in this project. **A).** Computational strategy used to prioritise translated smORFs most likely to encode functional micropeptides based on evolutionary conservation, domain structure and proteomic evidence from realignment of 13 public mass spectrometry data sets generated in lymphoid cells, either as total proteome or MHC I/II immunopeptidome. **B)** Summary of our CRISPR screen data using a custom-built library targeting 1,600 smORFs in six cell lines. Shown are the 26 smORFs with consistent dropout across all lines, compared to non-targeting controls (grey). Each coloured vertical line represents a single guide RNA. Many smORF guides show stronger dropout than positive control guides targeting MYC (red). **C)** Validation of CRISPR screen data by individual guide transduction compared to non targeting guides (NTG) and positive control guides (RPS6). **D)** Results of co-expression based network analysis (PGCNA) based on ribosome footprint density showing how some novel smORFs associate with core cellular processes, including cell cycle regulation or metabolism.