Harnessing the graft-versus-leukaemia effect in acute myeloid leukaemia

Connor Sweeney

Green Templeton College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2021

Abstract

Allogeneic haematopoietic stem cell transplantation (allo-SCT) is the best-established cellular immunotherapy and is a common treatment for acute myeloid leukaemia (AML). Allo-SCT relies on donor T cells identifying AML cells as foreign and eradicating them, which is known as graft-versus-leukaemia (GvL). Donor T cells recognise AML through interactions between their T cell receptor (TCR) and peptide antigens bound by human leukocyte antigen (HLA) on the surface of AML cells. Despite being administered for over 50 years, our understanding of the AML antigens that elicit a donor T cell response after allo-SCT is very limited.

Identification of antigens that mediate GvL may directly lead to improvements in selection of allo-SCT donors for treating AML patients. A donor whose cells can produce a specific alloimmune response against the recipient's AML cells would reduce relapse. Furthermore, antigens represent potential targets for immunotherapy.

This study establishes a novel workflow to identify and characterise GvL antigens and the donor-derived T cells that respond to them. GvL antigens are likely to result from germline DNA variants that are present in a patient but absent in their allo-SCT donor. Exome and RNA sequencing were used to identify mismatched DNA variants between patient and donor that are expressed in AML. Candidate GvL antigens were identified using: (1) *in silico* HLA-binding prediction, and (2) mass spectrometry of peptides eluted from HLA on AML cells.

Predicted GvL peptide antigens were functionally screened, which led to the identification of 24 validated antigens from 9 patients. Of the antigen-specific T cell responses characterised, 86% stimulate $CD4^+$ donor T cells. For a GvL antigen derived from the *PADI4* gene, responses are HLA-DP restricted and donor T cells are $CD4^+$ with a cytotoxic T helper 1 phenotype.

As a proof of concept, a T cell therapy was developed targeting the PADI4 antigen. Antigen-specific TCR sequences were identified and inserted into human T cells. These were then used to assess T cell-mediated killing of AML cells *in vitro* and *in vivo*, and also to investigate the mechanism of PADI4 antigen presentation by HLA.

Word Count: 41,275

Harnessing the graft-versus-leukaemia effect in acute myeloid leukaemia



Connor Sweeney
Green Templeton College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2021

Acknowledgements

I would like to thank my supervisor, Professor Paresh Vyas, for his scientific guidance, mentorship and support throughout my DPhil. Current and former members of the lab have been fantastic in helping with scientific ideas, giving advice on experimental techniques, assisting with experiments and giving feedback on presentations and my thesis. In particular, I would like to thank Lynn Quek, Bilyana Stoilova, Catherine Garnett, Tom Jackson, Grigore Gafencu, Batchimeg Usukhbayar, Pengjun Xi, Marlen Metzner, Vasiliki Symeonidou and Susann Rahmig. A special thanks to Gerda Mickute, who joined the lab in 2020. The work in Chapter 6 was completed together with Gerda and I am confident that she will be successful in building on the project.

I would also like to sincerely thank Professor Persephone Borrow, whose advice and guidance has been absolutely essential in planning immunology assays and interpreting the results of experiments. Members of her lab have been very generous with their time in providing advice and teaching experimental techniques. Specifically, I would like to thank Ane Ogbe, Tom Partridge and Wayne Paes. I would also would like to acknowledge that immunopeptidomic analysis of AML samples by mass spectrometry was performed by Annalisa Nicastri, Nicola Ternette and Silvia Salatino.

Thanks also to the members of my Thesis Committee, Professors Adam Mead, Paul Klenerman and Tudor Fulga for their guidance during my DPhil. I would like to thank my Green Templeton College Advisor, Professor Caroline Pond, for her support throughout my DPhil and for her feedback on my thesis.

I would also like to acknowledge the Wellcome Trust for funding my DPhil and Bristol Myers Squibb for funding subsequent work to build on this project. I hope that the funding will bring us closer to translating the scientific findings into clinical practice.

Most importantly, I would like to thank my family. My wife Emma and daughter Aoife have had to be extremely patient and understanding of my work commitments. They have provided me with much needed grounding, which has been essential during the ups and downs of my DPhil.

Finally, I would like to thank all the patients who have donated samples for research. Without their generosity, none of this work would have been possible.

Abstract

Allogeneic haematopoietic stem cell transplantation (allo-SCT) is the best-established cellular immunotherapy and is a common treatment for acute myeloid leukaemia (AML). Allo-SCT relies on donor T cells identifying AML cells as foreign and eradicating them, which is known as graft-versus-leukaemia (GvL). Donor T cells recognise AML through interactions between their T cell receptor (TCR) and peptide antigens bound by human leukocyte antigen (HLA) on the surface of AML cells. Despite being administered for over 50 years, our understanding of the AML antigens that elicit a donor T cell response after allo-SCT is very limited.

Identification of antigens that mediate GvL may directly lead to improvements in selection of allo-SCT donors for treating AML patients. A donor whose cells can produce a specific alloimmune response against the recipient's AML cells would reduce relapse. Furthermore, antigens represent potential targets for immunotherapy.

This study establishes a novel workflow to identify and characterise GvL antigens and the donor-derived T cells that respond to them. GvL antigens are likely to result from germline DNA variants that are present in a patient but absent in their allo-SCT donor. Exome and RNA sequencing were used to identify mismatched DNA variants between patient and donor that are expressed in AML. Candidate GvL antigens were identified using: (1) *in silico* HLA-binding prediction, and (2) mass spectrometry of peptides eluted from HLA on AML cells.

Predicted GvL peptide antigens were functionally screened, which led to the identification of 24 validated antigens from 9 patients. Of the antigen-specific T cell responses characterised, 86% stimulate $CD4^+$ donor T cells. For a GvL antigen derived from the *PADI4* gene, responses are HLA-DP restricted and donor T cells are $CD4^+$ with a cytotoxic T helper 1 phenotype.

As a proof of concept, a T cell therapy was developed targeting the PADI4 antigen. Antigen-specific TCR sequences were identified and inserted into human T cells. These were then used to assess T cell-mediated killing of AML cells *in vitro* and *in vivo*, and also to investigate the mechanism of PADI4 antigen presentation by HLA.

Word Count: 41,275

Contents

List of Figures xiii

List of Tables xvii

List of Abbreviations xix

1 Introduction 1

1.1 Acute myeloid leukaemia . . . 2
1.2 Allogeneic stem cell transplantation for acute myeloid leukaemia . . 3
1.3 Historical context . . . 5
1.4 Donor selection . . . 7
1.5 Risk of relapse determines treatment in AML . . . 8
1.6 Evidence for a graft-versus-leukaemia effect . . . 10
1.6.1 Graft-versus-Host Disease and relapse . . . 11
1.6.2 T cell depletion and relapse . . . 14
1.6.3 Donor lymphocyte infusions reinstate a GvL effect . . . 17
1.6.4 Cure of patients treated with reduced intensity conditioning 18
1.7 Immune evasion mechanisms at post-transplant relapse . . . 19
1.7.1 Reduction in HLA presentation . . . 19
1.7.2 Expression of immune checkpoint molecules . . . 21
1.7.3 Immunosuppressive enzymes . . . 21
1.7.4 Alteration of pro- and anti-inflammatory cytokines . . . 22
1.8 Candidate GvL antigens . . . 23
1.8.1 Minor histocompatibility antigens . . . 27
1.8.2 Neoantigens . . . 29
1.8.3 Leukaemia-associated antigens . . . 30
1.8.4 NK cells and the GvL effect . . . 31
1.9 Identifying GvL antigens and T cells . . . 32
1.9.1 Identifying GvL antigens . . . 32
1.9.2 Identifying antigen-specific T cells . . . 34
1.10 Improving transplant outcomes in AML . . . 36
1.11 GvL antigens are potential targets for immunotherapy . . . 39
1.12 Aims . . . 41

2 Materials and Methods **43**
2.1 Collection and processing of human blood and bone marrow 44
2.2 Flow cytometry staining and sorting of leukaemia and T cells . . . 44
2.3 Whole exome sequencing 45
2.4 Bulk RNA sequencing 47
2.5 HLA binding prediction 48
2.6 Immunopeptidomic analysis 49
2.6.1 Antibody purification and resin preparation 49
2.6.2 HLA immunoprecipitation and elution of peptides 50
2.6.3 Liquid chromatography-tandem mass spectrometry 50
2.6.4 Data analysis 51
2.7 *Ex vivo* and cultured IFNγ ELISpot peptide screening 52
2.8 Intracellular cytokine staining after peptide stimulation 53
2.9 HLA allele restriction and epitope mapping 54
2.10 Isolation of antigen-specific T cells and sequencing their T cell receptors (TCRs) 55
2.11 Processing and analysis of single cell TCR-Seq and RNA-Seq data . 56
2.12 Bulk TCR sequencing 57
2.13 PADI4 genotyping 58
2.14 Assessing PADI4 expression 58
2.14.1 RT qPCR for *PADI4* gene expression 58
2.14.2 Western blotting for PADI4 59
2.15 Virus production and transduction 59
2.15.1 TCR cloning 59
2.15.2 Retrovirus production 60
2.15.3 Retroviral transduction of T cells 60
2.15.4 Lentivirus production 61
2.15.5 Lentiviral transduction of T cells 61
2.16 *In vitro* T cell activation and killing assays 62
2.16.1 T cell activation assays 62
2.16.2 Generating monocyte-derived dendritic cells 63
2.16.3 Killing assays 63
2.17 *In vivo* killing of AML 64

3 Prediction of graft-versus-leukaemia antigens **65**
3.1 Multi-omic approach to antigen prediction 65
3.1.1 Experimental overview 65
3.1.2 Patient cohort 68
3.1.3 Cell sorting for whole exome sequencing and RNA-Seq . . . 69

3.2 Identifying mismatched exonic variants . . . 72
3.2.1 Germline and somatic variant calling . . . 72
3.2.2 Germline mismatched variants . . . 79
3.2.3 Somatic variants . . . 79
3.2.4 Population frequency of mismatched variants . . . 83
3.3 RNA expression of variants . . . 84
3.3.1 Gene expression in AML . . . 84
3.3.2 Gene expression in normal tissues . . . 85
3.4 HLA binding prediction . . . 86
3.5 Immunopeptidomic analysis of AML . . . 89
3.5.1 Approach to identify HLA-bound peptides using tandem mass spectrometry . . . 89
3.5.2 Detection of HLA-bound peptides . . . 90
3.5.3 Comparison of peptide and RNA expression . . . 93
3.5.4 Comparing HLA-binding prediction to mass spectrometry detection . . . 95
3.5.5 Candidate alloreactive peptides identified by mass spectrometry 96
3.6 Conclusions: Antigen prediction . . . 98

4 Identification and characterisation of a GvL antigen 101
4.1 Approach to screening predicted GvL antigens . . . 102
4.1.1 Screening using IFNγ ELISpot . . . 102
4.1.2 Patient selection . . . 102
4.1.3 Compiling a peptide library for screening . . . 104
4.2 IFNγ ELISpot screening . . . 105
4.2.1 *Ex vivo* screening . . . 105
4.2.2 *In vitro* culture with peptide increases sensitivity of IFNγ ELISpot . . . 108
4.3 Characterisation of T cell responses to the PADI4 antigen . . . 111
4.3.1 Epitope mapping and functional titration . . . 111
4.3.2 Antigen-specific T cells are $CD4^+$ . . . 114
4.3.3 T cell responses are HLA-DP restricted . . . 117
4.4 Clinical utility of the PADI4 antigen . . . 120
4.5 PADI4 expression in human tissues . . . 120
4.5.1 The function of PADI4 . . . 120
4.5.2 Expression in normal tissue . . . 121
4.5.3 Expression in cancer . . . 125
4.6 Phenotyping antigen-specific T cells using single cell RNA sequencing 128
4.6.1 Isolating antigen-specific T cells . . . 128

4.6.2 Antigen-specific T cells are CD4$^+$ T helper 1 cells with a cytotoxic phenotype . . . 131
4.6.3 Pathway analysis . . . 136
4.7 Identifying antigen-specific TCR sequences . . . 139
4.8 Antigen-specific T cells are not CD8$^+$ . . . 142
4.9 *Ex vivo* post-transplant T cell receptor repertoires . . . 145
4.10 Conclusions: PADI4 antigen identification and characterisation . . . 147

5 Applying the antigen-identification workflow to characterise further GvL responses 149
5.1 Characterising GvL responses in a cohort of transplant recipients . 149
5.1.1 Aims . . . 149
5.1.2 Clinical evidence for a GvL response . . . 151
5.1.3 Functional screening of predicted GvL antigens . . . 153
5.2 Patient OX1149 . . . 154
5.2.1 Clinical features . . . 154
5.2.2 Screening for alloreactive peptides . . . 155
5.2.3 Characterising T cell responses to peptide 190 . . . 160
5.2.4 Characterising T cell responses to peptide 290 . . . 162
5.3 Patient OX289 . . . 163
5.3.1 Clinical features . . . 163
5.3.2 Screening for alloreactive peptides . . . 164
5.3.3 Immunophenotyping of antigen-specific T cell responses . . . 176
5.3.4 HLA restriction of peptide responses . . . 178
5.4 Summary of alloreactive peptides . . . 180
5.5 Predicting antigen binding to common HLA alleles . . . 183
5.6 Expression of antigens in normal tissues . . . 185
5.7 Identifying antigen-specific TCR sequences . . . 188
5.8 Functional screening of peptides identified by mass spectrometry . . 191
5.8.1 *Ex vivo* testing . . . 191
5.8.2 Testing following *in vitro* culture . . . 191
5.9 Conclusions: Identifying and characterising alloreactive antigens . . 193

6 Targeting the PADI4 antigen using adoptive T cell therapy 195
6.1 Targeting the PADI4 antigen with T cell therapy . . . 195
6.2 TCR transfer using viral vectors . . . 196
6.2.1 Retroviral TCR transfer . . . 196
6.2.2 Lentiviral TCR transfer . . . 198
6.3 Identifying target cells for functional assays . . . 204
6.4 T cell activation by AML . . . 210

6.5 *In vitro* killing of AML by targeting the PADI4 antigen 213
6.6 Assessing the mechanism of PADI4 antigen presentation 215
6.7 *In vivo* killing of AML by targeting the PADI4 antigen 219
6.8 Conclusions: Targeting the PADI4 antigen using adoptive T cell therapy 221

7 Discussion **223**
7.1 Clinical importance of identifying of GvL antigens 223
7.2 Prediction of GvL antigens 224
7.3 Identification and characterisation of the PADI4 antigen 226
7.3.1 Functional screening of predicted HLA-binding peptides . . 226
7.3.2 Phenotype of PADI4 antigen-specific T cells 229
7.4 The antigenic landscape following allogeneic stem cell transplantation 230
7.4.1 Tissue distribution of antigen expression 230
7.4.2 HLA restriction of antigens 231
7.4.3 Tracking antigen-specific T cell responses post-transplant . . 232
7.4.4 Clinical impact of alloreactive antigens following allogeneic stem cell transplantation 232
7.5 Targeting GvL antigens using adoptive T cell therapy 233
7.5.1 Targeting the PADI4 antigen 233
7.5.2 Clinical implications of TCR therapy 235

Appendices

A Appendices **241**

B Appendices **245**

References **251**

List of Figures

1.1 Allogeneic haematopoietic stem cell transplantation. 4
1.2 GvL and GvHD T cell responses. 12
1.3 Possible mechanisms of post-transplant immune evasion. 20
1.4 HLA presentation of antigen to T cells. 24
1.5 MHC class I and II antigen presentation pathways. 26
1.6 Fluorescent HLA-peptide tetramer. 35
1.7 Methods of improving transplant outcomes in AML. 37

3.1 Overview of exome-guided antigen prediction. 67
3.2 Reverse immunology approach to antigen prediction. 68
3.3 Flow cytometry sorting of AML and T cells for exome sequencing and RNA-Seq. 71
3.4 Overview of germline and somatic variant calling pipelines. 72
3.5 Filtering of germline variants. 74
3.6 Germline DNA libraries for exome sequencing. 75
3.7 Exome sequencing coverage and VAF of germline variants. 76
3.8 Identifying mismatched germline variants between patient and donor. 77
3.9 Number of patient-donor mismatched germline variants. 78
3.10 Annotation of nucleotide variants. 79
3.11 Mismatched germline variants between patient and donor. 80
3.12 Number of somatic variants in AML. 80
3.13 Somatic variants in recurrently mutated genes. 82
3.14 Allele frequency of germline variants. 83
3.15 Variants detected across multiple patients. 84
3.16 RNA expression of variants. 85
3.17 Normal tissue expression of genes containing patient-donor mismatched variants. 86
3.18 HLA binding prediction. 89
3.19 Tandem mass spectrometry of peptides eluted from HLA molecules on AML. 90
3.20 HLA molecules were immunoprecipitated, peptides eluted and then analysed by mass spectrometry. 91
3.21 Length of peptides eluted from HLA molecules. 92
3.22 Amino acid representation by position along HLA class I peptide sequences. 93

3.23 Genes represented in the immunopeptidome. 94
3.24 Immunopeptidomic analysis has limited sensitivity to detect patient-donor mismatched peptides. 95
3.25 In silico binding prediction of HLA class I peptides detected by LC-MS/MS. 96
3.26 In silico binding prediction of HLA class II peptides detected by LC-MS/MS. 97
3.27 A total of 31 patient-donor mismatched peptides were detected by LC-MS/MS. 98

4.1 IFNγ ELISpot. 103
4.2 Producing overlapping peptides for screening. 105
4.3 Ex vivo screening of peptide pools. 107
4.4 Approach to screening predicted GvL peptides for T cell responses. 108
4.5 Cultured IFNg ELISpot screening of peptide pools. 110
4.6 OX866 6 years post-transplant. 110
4.7 Absence of T cell responses to the PADI4 reference peptide. 111
4.8 PADI4 antigen epitope mapping. 113
4.9 Functional titration of peptide responses. 114
4.10 Immunophenotyping of PADI4 T cell responses. 116
4.11 Cytokine responses. 117
4.12 Identifying HLA restriction using HLA-blocking antibodies. 118
4.13 Investigating HLA class I-restricted responses using single HLA-transfected .221 cells. 119
4.14 Expression of PADI4 in normal tissues. 123
4.15 PADI4 expression in normal haematopoietic cells. 124
4.16 Expression of PADI4 in AML and other cancers. 126
4.17 PADI4 expression in haematological malignancies. 127
4.18 Isolation of peptide-responsive T cells for scRNA-Seq. 129
4.19 Sorting antigen-specific T cells for scRNA-Seq. 130
4.20 $CD4^+$ T cells cluster according to activation by PADI4 peptide. . . 132
4.21 Differentially expressed genes between peptide-responsive and control T cells. 133
4.22 Markers of T cell activation. 134
4.23 Expression of markers of chronic activation / exhaustion. 135
4.24 Activated CD4+ are Th1 cells with a cytotoxic phenotype. 136
4.25 GSEA of differentially expressed genes between Activated and Control cells. 138
4.26 $CD4^+$ clones in Activated and Control repertoires. 139

4.27 Expanded Clones 1 and 2 have distinct TCRs and transcriptional profiles. . . . 140
4.28 Differentially expressed genes between expanded Clones 1 and 2. . . . 142
4.29 Clustering of CD4 and CD8 T cells. . . . 143
4.30 CD8 clonotypes in IFNg+ and IFNg- repertoires. . . . 144
4.31 Overlap between Activated and Control repertoires. . . . 144
4.32 *Ex vivo* post-transplant T cell repertoires. . . . 146

5.1 Approach to screening predicted GvL peptides for T cell responses. 154
5.2 OX1149 pooled peptide screening by IFNγ ELISpot. . . . 156
5.3 OX1149 Pool 7 IFNγ ELISpot screening. . . . 157
5.4 Peptide 190 reference control peptide. . . . 158
5.5 OX1149 Pool 10 IFNγ ELISpot screening. . . . 159
5.6 Peptide 290 reference control peptide elicits responses. . . . 160
5.7 Peptide 190 stimulates $CD4^+$ T cells. . . . 161
5.8 T cell responses to peptide 190 are HLA-DR restricted. . . . 162
5.9 Peptide 290 stimulates $CD4^+$ T cells. . . . 163
5.10 T cell responses to peptide 290 are HLA-DR restricted. . . . 163
5.11 OX289 pooled peptide screening by IFNγ ELISpot. . . . 165
5.12 OX289 Pool 1 IFNγ ELISpot screening. . . . 166
5.13 OX289 Pool 2 IFNγ ELISpot screening. . . . 167
5.14 OX289 Pool 3 IFNγ ELISpot screening. . . . 170
5.15 OX289 Pool 7 IFNγ ELISpot screening. . . . 171
5.16 OX289 Pool 6 IFNγ ELISpot screening. . . . 172
5.17 OX289 Pool 8 IFNγ ELISpot screening. . . . 173
5.18 OX289 Pool 10 IFNγ ELISpot screening. . . . 174
5.19 OX289 Pool 11 IFNγ ELISpot screening. . . . 175
5.20 Immunophenotyping of $CD4^+$ T cell responses. . . . 177
5.21 Immunophenotyping of $CD8^+$ T cell responses. . . . 178
5.22 Assessing HLA restriction of $CD4^+$ T cell responses. . . . 179
5.23 Assessing HLA restriction of $CD8^+$ T cell responses. . . . 180
5.24 Predicted binding of GvL peptide antigens to common HLA class I molecules. . . . 184
5.25 Predicted binding of GvL peptide antigens to common HLA class II molecules. . . . 185
5.26 Expression of GvL antigens in normal tissues. . . . 187
5.27 Expression of GvL antigens in tissues affected by acute GvHD. . . . 188
5.28 OX1149 $CD4^+$ TCR repertoires following peptide stimulation. . . . 189
5.29 OX289 $CD4^+$ TCR repertoires following peptide stimulation. . . . 190

5.30 *Ex vivo* IFNγ ELISpot testing of peptides identified by mass spectrometry. . . . 192
5.31 IFNγ ELISpot testing of peptides identified by mass spectrometry, following *in vitro* culture. . . . 193

6.1 pMP71-TCR retroviral vector. . . . 197
6.2 Retroviral transduction of T cells. . . . 198
6.3 pHRsin-TCR lentiviral vector. . . . 199
6.4 Lentiviral transduction of T cells. . . . 199
6.5 Serial transduction efficiencies. . . . 200
6.6 $CD4^+$ and $CD8^+$ enriched T cell cultures. . . . 201
6.7 Clone 2.1 exhibits PADI4 antigen-specific responses. . . . 203
6.8 HLA-DP is expressed on cultured T cells. . . . 203
6.9 PADI4 genotyping. . . . 205
6.10 PADI4 protein expression. . . . 206
6.11 PADI4 RNA expression. . . . 207
6.12 HLA-DP is expressed on AML cells. . . . 208
6.13 Summary of PADI4 genotyping and HLA-DP typing of AML samples. 209
6.14 Summary of PADI4 and HLA-DP expression by AML samples. . . . 210
6.16 PADI4 Clone 2.1 is activated by AML that has been pulsed with PADI4 peptide. . . . 213
6.17 PADI4 Clone 2.1 kills AML cells that have been pulsed with PADI4 peptide. . . . 214
6.18 PADI4 Clone 2.1 kills AML cells only when pulsed with PADI4 peptide. 215
6.19 Testing extracellular antigen presentation by monocyte-derived dendritic cells 216
6.20 MDDCs present exogenous PADI4 antigen. . . . 217
6.21 Testing antigen presentation by AML cells. . . . 218
6.22 AML activates T cells. . . . 219
6.23 Preliminary testing *in vivo* killing of AML by PADI4 T cell Clone 2.1. 221

B.1 Nucleotide sequences of T cell receptors for PADI4 antigens 250

List of Tables

1.1 European LeukemiaNet (ELN) recommendations for allogeneic stem cell transplantation in patients with AML in first complete remission. 9
1.2 Methods of T cell depletion. 15

3.1 Purity of sorted AML and T cells. 70
3.2 Somatic variants detected by exome sequencing. 81

4.1 Peptides used for epitope mapping. 112

5.1 Clinical evidence for a GvL response. 152
5.2 Summary of antigens and their T cell responses. 181

A.1 Baseline characteristics of AML patients. 242
A.2 Treatment summary for patient cohort. 243
A.3 GvHD, immunosuppression and treatment with donor lymphocyte infusions. 244

B.1 Antibodies and other stains used in flow cytometry assays. 246
B.2 Summary of functionally validated alloreactive antigens. 247
B.3 HLA types of the patients studied. 248

List of Abbreviations

5' RACE . . . 5' Rapid amplification of cDNA ends

ACN . . . . . Acetonitrile

AEC 3-amino-9-ethylcarbazole

ALL . . . . . Acute lymphoblastic leukaemia

Allo-SCT . . . Allogeneic haematopoietic stem cell transplantation

AML . . . . . Acute myeloid leukaemia

APC . . . . . Antigen presenting cell

ATG . . . . . Anti-thymocyte globulin

β2M β2-microglobulin

BCA . . . . . Bicinchoninic acid

bp . . . . . Base pair

BM . . . . . Bone marrow

BWA-MEM . Burrows-Wheeler aligner maximal exact match

CAR-T . . . . Chimeric antigen receptor T cell

cDNA . . . . . Complementary DNA

CDR3 . . . . . Complementarity-determining region 3

CEF . . . . . Cytomegalovirus / Epstein-Barr virus / Influenza peptide pool

CIPR . . . . . Cluster Identity Predictor

CLIP . . . . . Class II-associated invariant chain peptide

CLL . . . . . Chronic lymphocytic leukaemia

CML . . . . . Chronic myeloid leukaemia

CMP . . . . . Common myeloid progenitor

CMV . . . . . Cytomegalovirus

CPM . . . . . Counts per million

CTLA-4 . . . . Cytotoxic T-lymphocyte-associated protein 4

dbSNP . . . . Database of Single Nucleotide Polymorphisms

DC . . . . . Dendritic cell

DICE . . . . . Database of Immune Cell Expression

DLI . . . . . Donor lymphocyte infusion

DMSO . . . . . Dimethyl sulfoxide

DTT 1,4-dithiothreitol

EBMT . . . . . European Society for Blood and Marrow Transplantation

EBV . . . . . Epstein-Barr virus

EDTA . . . . . Ethylenediaminetetraacetic

ELISpot . . . . Enzyme-linked immune absorbent spot

ELN . . . . . European LeukemiaNet

ER . . . . . Endoplasmic reticulum

FACS . . . . . Fluorescence-activated cell sorting

FBS . . . . . Fetal bovine serum

FC . . . . . Fold change

FDR . . . . . False discovery rate

FLT3-ITD . . Fms-related tyrosine kinase 3 internal tandem duplication

FMO . . . . . Fluorescence-minus-one

G-CSF . . . . Granulocyte colony-stimulating factor

GEM-RT . . . Gel bead-in-emulsion reverse transcription

GM-CSF . . . Granulocyte-macrophage colony-stimulating factor

GMP . . . . . Granulocyte-monocyte progenitor

GSEA . . . . . Gene set enrichment analysis

GvL . . . . . Graft-versus-leukaemia

GvHD . . . . . Graft-versus-host disease

GWAS . . . . . Genome-wide association study

HCD . . . . . Higher energy collision dissociation

HLA . . . . . Human lymphocyte antigen

HPLC . . . . . High performance liquid chromatography

HR . . . . . Hazard ratio

HRP . . . . . Horseradish peroxidase

HSC . . . . . Haematopoietic stem cell

IBMTR . . . . International Bone Marrow Transplant Registry

ICS . . . . . Intracellular cytokine staining

IDH . . . . . Isocitrate dehydrogenase

IFNγ . . . . . Interferon gamma

IGV . . . . . Integrative Genomics Viewer

IL-2/4/7/10/15 Interleukin-2/4/7/10/15

IMDM . . . . Iscove's Modified Dulbecco's Medium

KIR . . . . . Killer cell immunoglobulin-like receptor

LAA . . . . . Leukaemia-associated antigen

LC-MS/MS . Liquid chromatography - Tandem mass spectrometry

LDS . . . . . Laemmli sodium dodecyl sulphate

LTR . . . . . Long terminal repeat

LV . . . . . Lentivirus

MAIT . . . . . Mucosal-associated invariant T cell

MDDC . . . . Monocyte-derived dendritic cell

MDS . . . . . Myelodysplastic syndrome

MEP . . . . . Megakaryocyte-erythrocyte progenitor

MHC . . . . . Major histocompatibility complex

miHA . . . . . Minor histocompatibility antigen

MIIC . . . . . MHC class II compartment

MNC . . . . . Mononuclear cell

MPP . . . . . Multipotent progenitor

MPSV . . . . . Myeloproliferative sarcoma virus

MRD . . . . . Measurable residual disease

mTCRβ . . . . Murine T cell receptor β

MUD . . . . . Matched unrelated donor

MS/MS . . . . Tandem mass spectrometry

NHS . . . . . National Health Service

NK cell . . . . Natural killer cell

OS . . . . . Overall survival

PAS . . . . . Protein A sepharose

PB . . . . . Peripheral blood

PBMC . . . . Peripheral blood mononuclear cell

PBSC . . . . . Peripheral blood stem cell

PBS . . . . . Phosphate-buffered saline

PCR . . . . . Polymerase chain reaction

PD-1 . . . . . Programmed cell death-1

PHA . . . . . Phytohaemagglutinin

PRMT . . . . Protein arginine methyltransferase

PVDF . . . . . Polyvinylidene difluoride

RAB-10 . . . . RPMI-1640 media with 10% heat-inactivated human AB serum

RIC . . . . . Reduced intensity conditioning

RIPA . . . . . Radioimmunoprecipitation assay

RNA-Seq . . . RNA sequencing

RPKM . . . . Reads per kilobase of transcript, per million mapped reads

RPMI-1640 . Roswell Park Memorial Institute medium

RT qPCR . . . Real-time quantitative polymerase chain reaction

RV . . . . . Retrovirus

scRNA-Seq . . Single cell RNA sequencing

scTCR-Seq . . Single cell T cell receptor sequencing

SD . . . . . Standard deviation

SDS-PAGE . . Sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SEM . . . . . Standard error of the mean

SFU . . . . . Spot forming units

SNP . . . . . Single nucleotide polymorphism

SNV . . . . . Single nucleotide variant

T-ALL . . . . . T cell acute lymphoblastic leukaemia

TAP . . . . . Transporter associated with antigen processing

TBI . . . . . Total body irradiation

TBS-T . . . . . Tris-buffered saline with 0.1% Tween-20

TCR . . . . . T cell receptor

TCR-Seq . . . T cell receptor sequencing

Th1/2/17 . . . T helper 1/2/17

TFA . . . . . Trifluoroacetic acid

TNFα . . . . . Tumour necrosis factor alpha

TPM . . . . . Transcripts per million

TRA . . . . . T cell receptor α chain

TRB . . . . . T cell receptor β chain

Treg . . . . . Regulatory T cell

UMAP . . . . Uniform manifold approximation and projection

UMI . . . . . Unique molecular identifier

UTR . . . . . Untranslated region

VAF . . . . . Variant allele frequency

VEP . . . . . Variant Effect Predictor

VOD . . . . . Veno-occlusive disease

WES . . . . . Whole exome sequencing

WT . . . . . Wild type

WT1 . . . . . Wilms' tumour protein

Chapter 1

Introduction

Contents

1.1 Acute myeloid leukaemia 2
1.2 Allogeneic stem cell transplantation for acute myeloid leukaemia 3
1.3 Historical context 5
1.4 Donor selection 7
1.5 Risk of relapse determines treatment in AML 8
1.6 Evidence for a graft-versus-leukaemia effect 10
1.6.1 Graft-versus-Host Disease and relapse 11
1.6.2 T cell depletion and relapse 14
1.6.3 Donor lymphocyte infusions reinstate a GvL effect . . . 17
1.6.4 Cure of patients treated with reduced intensity conditioning 18
1.7 Immune evasion mechanisms at post-transplant relapse 19
1.7.1 Reduction in HLA presentation 19
1.7.2 Expression of immune checkpoint molecules 21
1.7.3 Immunosuppressive enzymes 21
1.7.4 Alteration of pro- and anti-inflammatory cytokines . . . 22
1.8 Candidate GvL antigens 23
1.8.1 Minor histocompatibility antigens 27
1.8.2 Neoantigens 29
1.8.3 Leukaemia-associated antigens 30
1.8.4 NK cells and the GvL effect 31
1.9 Identifying GvL antigens and T cells 32
1.9.1 Identifying GvL antigens 32
1.9.2 Identifying antigen-specific T cells 34
1.10 Improving transplant outcomes in AML 36
1.11 GvL antigens are potential targets for immunotherapy 39
1.12 Aims 41

Much of the work in the Introduction (Chapter 1) has been adapted from its original publication by Frontiers in Oncology [1]. *I wrote all of the publication, with guidance and advice provided by Professor Paresh Vyas.*

1.1 Acute myeloid leukaemia

Acute myeloid leukaemia (AML) is the commonest aggressive leukaemia in adults, with an incidence of 3-5 cases per 100,000 population per year [2]. AML is a cancer that is characterised by infiltration of the bone marrow, blood and other tissues by highly proliferative, clonal cells of the haematopoietic system [3].

Normal blood cell production exhibits a cellular hierarchy of differentiation, which is maintained by haematopoietic stem cells. These cells have capacity for long-term self-renewal and their differentiation leads to the production of short-lived mature blood cells (red blood cells, white blood cells and platelets). Therefore, stem cells are required to sustain long-term haematopoiesis. AML cells are blocked in their differentiation, leading to an expansion of abnormal immature leukaemia cells in the bone marrow and a reduction in numbers of normal mature blood cells. Many of the clinical manifestations of AML result from failure of the bone marrow to produce normal blood cells, including fatigue, shortness of breath, infections, bleeding and bruising. A diagnosis of acute leukaemia is established following examination of a blood film and/or bone marrow aspirate smear under a light microscope, which reveals >20% abnormal immature leukaemic cells. AML is diagnosed by demonstrating the myeloid origin of these cells, usually by flow cytometry.

AML most commonly arises spontaneously (*de novo*) but can also result from either a preceding haematological malignancy or treatment with chemotherapy or radiotherapy. Irrespective of the aetiology, genetic aberrations are well described in AML and have implications for treatment and prognosis (Section 1.5). Genetic mutations are found in the vast majority AML and in some cases large chromosomal translocations or aneuploidy are observed [4].

1.2 Allogeneic stem cell transplantation for acute myeloid leukaemia

Allogeneic haematopoietic stem cell transplantation (allo-SCT) is the most established and commonly used cellular immunotherapy in cancer care. Allo-SCT is used to treat a range of haematological malignancies and AML is the most frequent disease indication, accounting for 36% of transplants in Europe [5]. Allo-SCT is routinely administered with curative intent to AML patients who are predicted to be at high risk of relapse.

Patients assessed as medically fit are treated with cytotoxic chemotherapy and are stratified according to their risk of relapse, which is based on genetic features of their leukaemia and response to initial treatment. Approximately 70–80% of patients aged under 60, and 50% of older patients, achieve a remission with induction chemotherapy [6, 7]. Despite achieving a remission, without further treatment most patients would subsequently relapse, usually within 6 months, and relapse is associated with poor prognosis. Therefore, post-remission therapy aims to reduce relapse by eliminating residual leukaemia cells [8].

Patients at higher risk of relapse receive an allo-SCT, which remains the most effective anti-leukaemia treatment for the majority of AML patients with intermediate and poor risk disease [9]. Prior to transplantation a patient receives conditioning, which consists of high doses of chemotherapy, with or without radiotherapy (total body irradiation, TBI). Conditioning aims to kill cancerous cells but also reduces normal bone marrow haematopoietic and immune cells. Subsequent transplantation of haematopoietic stem/progenitor and immune cells from a healthy donor leads to reconstitution of normal haematopoietic and immune cells (Figure 1.1). Crucially, alloreactivity of donor T cells against the patient's leukaemia is responsible for the graft-versus- leukaemia (GvL) effect, which is a major mechanism for the curative effect of allo-SCT [10]. However, this must be balanced against alloreactivity against normal tissues, which manifests as graft-versus-host disease

(GvHD), a multi-system disorder that initially may affect the skin, gastrointestinal tract, liver and lungs, but can later affect almost any organ [11, 12].

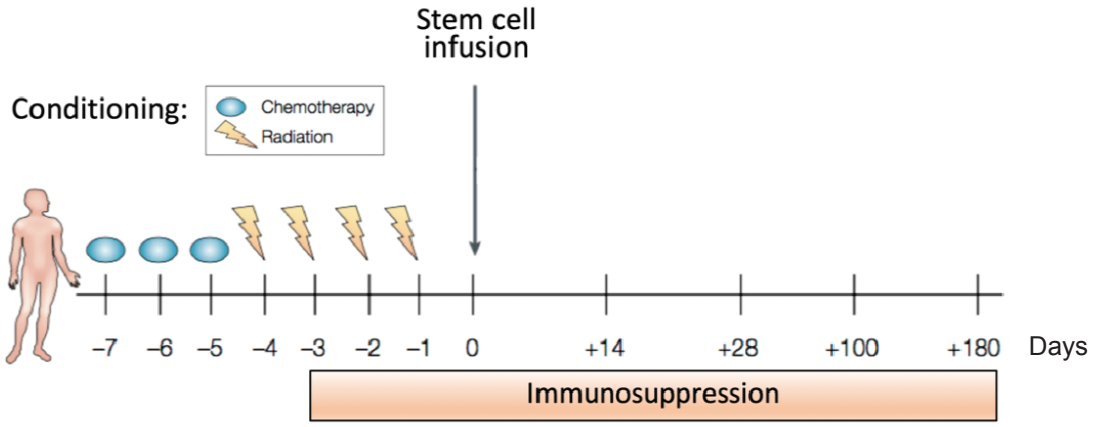


Figure 1.1: Allogeneic haematopoietic stem cell transplantation. Patients undergoing allo-SCT are treated with several days of chemotherapy, with or without total body irradiation, to eliminate their own haematopoietic cells. They then receive an infusion of haematopoietic stem and progenitor cells from a donor, which engraft and produce blood cells. Systemic immunosuppressive therapy starts prior to the transplant and continues for at least 3 months post-transplant, but can be longer if GvHD occurs. *Adapted from Barker and Wagner [13].*

Strategies to ameliorate the negative immunological effects include T cell depletion of grafts and treating recipients with immunosuppressive drugs [14]. However, these can also dampen the desired anti-tumour (GvL) responses, increasing the risk of relapse, and therefore a careful balance is required. After recovering from the toxicity of conditioning, if there is a concern about lack of a GvL response, for example due to loss of donor chimerism or evidence of molecular relapse, infusions of donor T cells can be administered as donor lymphocyte infusions (DLIs) in an attempt to produce a GvL response, with an associated risk of inducing GvHD [15, 16]. The overall survival (OS) of AML patients treated by allo-SCT remains modest at around 50% at 3 years, which is largely due to relapse and treatment-related mortality (GvHD and opportunistic infections), emphasising the need to improve upon current treatment approaches [17].

1.3 Historical context

When first developed over 50 years ago, the aim of allo-SCT treatment for human AML was to permit the delivery of high doses of chemotherapy that would otherwise be limited by toxicity to the haematopoietic system. A donor immune response against leukaemia was first suggested in 1956 by mouse transplantation experiments, in which murine leukaemia relapses were reduced following transplantation with allogeneic bone marrow (from a different mouse strain) compared with syngeneic marrow (same mouse strain) [18]. Some of the animals treated with allogeneic transplantation died from a "wasting syndrome" with diarrhoea that likely represented GvHD.

The first human allogeneic bone marrow transplantations were reported the following year by E. Donnall Thomas [19]. Six patients were treated with conditioning chemotherapy and radiotherapy, followed by an infusion of bone marrow from a healthy donor. Only two of the six patients showed evidence of engraftment and all died within 100 days of transplantation. Donors and patients were not matched for histocompatibility, as little was known about human histocompatibility antigens at this time.

The identification of human leukocyte antigens (HLA) and the development of methods to type these antigens enabled transplant programs to improve outcomes by patient-donor histocompatibility matching [20]. The HLA locus is on the short arm of chromosome 6 and is one of the most polymorphic regions of the human genome. The first HLA antigen (HLA-A2) was discovered in 1958 by Jean Dausset, who studied the sera from patients who had received multiple blood transfusions [21]. He found that sera from some of these patients agglutinated leukocytes from 11 of 19 individuals tested, but not leukocytes from the donor of the sera, suggesting the presence of an alloantigen [22]. The clinical importance of HLA matching in kidney transplantation was realised during the 1960s. Grafts between unrelated individuals are associated with a high likelihood of major incompatibility

due to the polymorphic nature of the HLA locus. However, within a family there are only 4 haplotypes (2 from each parent). In 1965, Terasaki and colleagues reported improved kidney graft survival from HLA-matched donors and by the early 1970s it was evident that kidneys transplanted between HLA-identical siblings produced the optimal outcomes [22].

In 1965, Mathé reported histocompatibility testing of bone marrow donors for a patient with acute lymphoblastic leukaemia (ALL) [23]. The patient was given grafts from six family members and developed a "secondary syndrome," later to be described as GvHD, involving the skin, gastrointestinal tract, and liver. Six months after the transplant, further bone marrow was administered with the aim of enhancing the immune response against leukaemia. In order to choose the donor least likely to reactivate a secondary syndrome, histocompatibility testing was performed by giving the patient skin grafts from each donor. The donor was selected whose skin graft was not rejected, but this still led to a "secondary syndrome" that responded to immunosuppressive treatment with corticosteroids [24].

The development of techniques to type HLA in patients and donors in the mid/late 1960s enabled E. Donnall Thomas to open an allogeneic bone marrow transplantation program in Seattle using HLA-matched donors for patients with acute leukaemia. In 1977, they reported 100 transplantations for relapsed and refractory acute leukaemia, using chemotherapy and radiation therapy conditioning administered prior to transplantation [25]. HLA matching was performed using mixed leukocyte culture and all transplants used HLA-identical related donors. Only 13 patients were free from disease after 1–4.5 years follow-up in this case series. However, administering a transplant earlier in the course of AML treatment resulted in a cure rate of 50% in patients transplanted in first remission [26]. Crucially, Thomas appreciated that the donor immune system likely plays a key role in eliminating residual leukaemia cells. Although survival was reduced in patients with severe GvHD, most patients did not die of relapse. Furthermore, patients with GvHD relapsed later than those without GvHD [26, 27]. In 1990, E. Donnall

Thomas won a Nobel Prize for his discoveries concerning cell transplantation in the treatment of human disease.

1.4 Donor selection

Despite being the preferred donor, a HLA-matched sibling is available for only 30% of patients. The use of HLA-matched unrelated donors for stem cell transplantation drastically increased the number of patients who could be treated [28]. In the years after the first unrelated donor transplant, donor registries were established, but due to large variation in ethnic representation, the chance of finding a full HLA match varied widely according to ethnicity [29]. Therefore, international collaboration has been essential for the establishment of a global donor registry. The International Bone Marrow Transplant Registry (IBMTR) was founded in 1972 to record transplant outcome data. In 1974, the European Society for Blood and Marrow Transplantation (EBMT) was formed and in 1988, Bone Marrow Donors Worldwide was established. There are currently over 34 million registered donors worldwide and almost 800,000 units of cord blood, collected at births from umbilical cords [30]. Although a donor is found for most patients, a high resolution HLA-matched donor may not be available for all, in particular for those who are not of white European descent, as the ethnic diversity of patients is not reflected in donor registries [31].

In the absence of a HLA-matched donor, the use of alternative donor allo-SCTs has increased the pool of available donors. This involves either a HLA-mismatched unrelated donor, haploidentical related donor or an umbilical cord blood transplant. Mismatched allo-SCTs generally have a mismatch at 1 or 2 HLA alleles (HLA-A, -B, -C, -DRB1, or -DQB1), as a greater degree of mismatch is associated with unacceptably high risks of GvHD and associated mortality [20]. Haploidentical related donor transplants are from siblings, parents or children of the patient who share at least 50% of HLA alleles. Due to bidirectional alloreactive T cell responses, the first haploidentical transplants were associated with high mortality from GvHD and also a host-versus-graft reaction that resulted in rejection of transplanted cells

[32]. Improvement in T cell depletion methods have reduced GvHD and graft failure following haploidentical allo-SCT. The use of haploidentical transplants has been increasing in recent years and EBMT reported that 14.7% of allo-SCTs in 2017 used haploidentical donors [33]. Umbilical cord blood transplantation is an alternative source of haematopoietic stem cells for transplantation. Naïvety of immune cells in cord blood permits less stringent HLA matching. However, due to the relatively low number of stem cells in a unit, adult transplants require either two cord units or *ex vivo* expansion of stem cells. As the use of haploidentical donors has increased, cord blood transplants have reduced and 2% of allo-SCTs reported by EBMT in 2017 used cord blood donations [33].

1.5 Risk of relapse determines treatment in AML

Although allo-SCT reduces relapse, non-relapse mortality due to complications of the transplant including GvHD and infection may counterbalance this beneficial effect in many patients. Therefore, when deciding which individuals would benefit from allo-SCT, there must be a patient-specific evaluation. The European LeukemiaNet (ELN) AML Working Party proposes a dynamic risk assessment that integrates the cytogenetic and molecular genetic features of AML at diagnosis with the patient's response to induction therapy to estimate the risk of relapse after consolidation treatment with either allo-SCT or chemotherapy. This relapse risk is balanced against the non-relapse mortality from allo-SCT that is estimated using the patient's co-morbidities with the haematopoietic cell transplantation comorbidity index, HCT-CI (Table 1.1) [34]. The ELN suggest that if, based on an individual's risk assessment, the disease-free survival is predicted to improve by at least 10%, allo-SCT should be recommended. In the absence of significant co-morbidities, this translates to intermediate and poor risk patients.

The ELN genetic risk stratification has been refined in recent years, reflecting a deeper understanding of the genomic landscape of AML, and mutations in genes such as *RUNX1*, *ASXL1*, and *TP53* now contribute to the adverse risk category [4, 36]. Assessment of post-treatment measurable residual disease (MRD)

AML risk group	Risk assessment	Risk of relapse following consolidation treatment		Nonrelapse mortality risk that would indicate allo-SCT as consolidation treatment	
		Chemotherapy* (%)	Allo-SCT (%)	HCT-CI score	Nonrelapse mortality (%)
Good	t(8;21) with WBC ≤20 Inv(16)/t(16;16) Mutated *CEBPA* (bi-allelic) Mutated *NPM1* (No *FLT3*–ITD mutation) Early first complete remission (after first cycle of chemotherapy) and MRD negative	35-40	15-20	0	10-15
Intermediate	t(8;21) with WBC >20 Cytogenetically normal (or loss of X and Y chromosomes), WBC count ≤100 and early first complete remission	50-55	20-25	≤2	<20–25
Poor	Otherwise good or intermediate, but not in complete remission after first cycle of chemotherapy Normal cytogenetics and WBC >100 Abnormal cytogenetics	70-80	30-40	≤3-4	<30
Very poor	Monosomal karyotype Abn3q26 Enhanced Evi-1 expression	>90	40-50	≤5	<40

Table 1.1: European LeukemiaNet (ELN) recommendations for allogeneic stem cell transplantation in patients with AML in first complete remission. ELN 2012 patient-specific risk assessment of AML relapse and non-relapse mortality following allo-SCT compared with chemotherapy consolidation. Recommendation of allo-SCT if the individual patient's disease-free survival benefit is at least 10%.
*Chemotherapy consolidation includes option of high-dose chemotherapy and autologous stem cell rescue.
HCT-CI: haematopoietic cell transplantation comorbidity index, *CEBPA*: CCAAT/enhancer-binding protein α, *NPM1*: nucleophosmin, *FLT3-ITD*: fms-like tyrosine kinase receptor-3 internal tandem duplication. Monosomal karyotype defined by presence of either 2 or more autosomal monosomies, or one monosomy plus one or more structural aberrations. Evi-1: Ecotropic viral integration site 1, WBC: white blood cell count at diagnosis, MRD: measurable residual disease.
Adapted from Cornelissen et al. [35].

provides additional prognostic information that complements pre-treatment genetic risk stratification. The presence of low amounts of MRD has been consistently associated with increased relapse and reduced OS in AML [37]. Two approaches may be used for MRD detection: (1) multiparameter flow cytometry, and (2) molecular techniques, including real-time quantitative PCR (RT-qPCR) and next generation sequencing (NGS). MRD using flow cytometry commonly involves the identification of a leukaemia-associated immunophenotype for the individual patient that differs from normal haematopoietic cells [38]. RT-qPCR assays are available for MRD detection of specific genetic lesions found in sub-groups of patients with AML, including *NPM1* mutations and *CBFB-MYH11*, *RUNX1-RUNX1T1*, and *BCR-ABL1* fusion genes. As a molecular marker can be detected in the majority of cases, NGS offers the possibility of tracking additional molecular markers in

the future. However, validation of markers is still required, as mutations in genes associated with pre-leukaemic clones (e.g. *DNMT3A*, *TET2*, and *ASXL1*) are frequently detected at remission but are poorly predictive of relapse [39]. A key goal is to be able to use MRD data to identify early recurrence of leukaemia and guide post-remission therapy. This could offer the opportunity to intervene pre-emptively to prevent morphological relapse, such as by administering further pre-transplant chemotherapy, increasing the intensity of transplant conditioning or introducing post-transplant therapy [36]. However, the optimal means of using MRD to guide treatment decisions has yet to be fully defined.

When treating relapsed AML with curative intent, allo-SCT is generally the most effective therapy. However, compared to transplants in first remission, outcomes are inferior due to an increase in both relapse (40–45%) and non-relapse mortality (25–35%) [35]. Breems *et al.* developed a prognostic index for AML patients at first relapse based on a study of 667 patients to predict clinical outcomes [40]. Three risk groups were defined using four clinical parameters: duration of remission, age at relapse, cytogenetics at diagnosis and whether patients had previously received an allo-SCT. Only 37% of patients in the cohort entered a second complete remission and 18% received an allo-SCT after relapse. However, in all risk groups, patients who received allo-SCT had superior outcomes.

1.6 Evidence for a graft-versus-leukaemia effect

A range of donor immune cells are likely to contribute to the GvL response, including T cells, NK cells, and B cells. However, current clinical evidence suggests that T cells exert the most potent and clinically relevant anti-leukaemia effect. Donor T cells recognise leukaemia through interactions between their T cell receptor (TCR) and major histocompatibility (MHC) molecules on the surface of AML cells. MHC, which are known as HLA in humans, present peptides to T cells. HLA class I molecules are expressed on all nucleated cells and present peptides to $CD8^+$ T cells, whereas class II molecules are primarily expressed on specialised antigen

presenting cells and present peptides to $CD4^+$ T cells. High levels of HLA class II expression is also seen in most cases of AML [41–43].

When allogeneic haematopoietic stem cell transplantation is planned, a donor is sought with HLA alleles matching the recipient. This reduces immune responses from patient-derived cells, which could result in graft rejection, and from donor-derived cells, which result in GvHD. However, even after a HLA-matched transplant, an immune response against leukaemia occurs. The importance of this GvL response in AML disease control is evidenced by several clinical observations:

1. An inverse correlation between GvHD and disease relapse.
2. Increased relapse in patients treated with T cell depleted grafts.
3. The successful use of DLIs post-transplantation to achieve disease control.
4. Cure of patients treated with reduced intensity conditioning (RIC) regimens that permit stable engraftment of donor haematopoietic cells, but are unlikely to directly kill all leukaemia cells [44].
5. Immune evasion mechanisms displayed by AML at post-transplant relapse indicate the presence of immune selection.

1.6.1 Graft-versus-Host Disease and relapse

Many retrospective analyses of patients who received allo-SCT for haematological malignancies have found a significant association between incidence of GvHD and reduced relapse rates [45–48]. Following transplantation for AML, Baron *et al.* found that grade I acute GvHD is associated with lower incidence of relapse (hazard ratio 0.7) and there was a trend toward an OS benefit. However, severe acute GvHD (grade III and IV) was associated with a worse OS due to non-relapse mortality. Similarly, the extent of chronic GvHD correlated with lower relapse rates but increased non-relapse mortality [47]. Recipients of identical twin (syngeneic) stem cell transplants have a higher incidence of relapse but a lower treatment-related mortality due to a reduction in GvHD [45].

A retrospective study of 48,111 allo-SCTs performed in adult patients between 1998 and 2007 for a range of haematological malignancies highlighted that GvHD

is strongly associated with a GvL effect in chronic myeloid leukaemia (CML) and acute lymphoblastic leukaemia (ALL), with GvL effect inferred from post-transplant relapse rates. In contrast, there was a weaker correlation between the GvL effect and GvHD in AML and myelodysplastic syndrome (MDS) [44, 49]. This suggests that some mechanisms of GvL may be distinct from GvHD and provides hope that these processes may be uncoupled therapeutically (Figure 1.2).

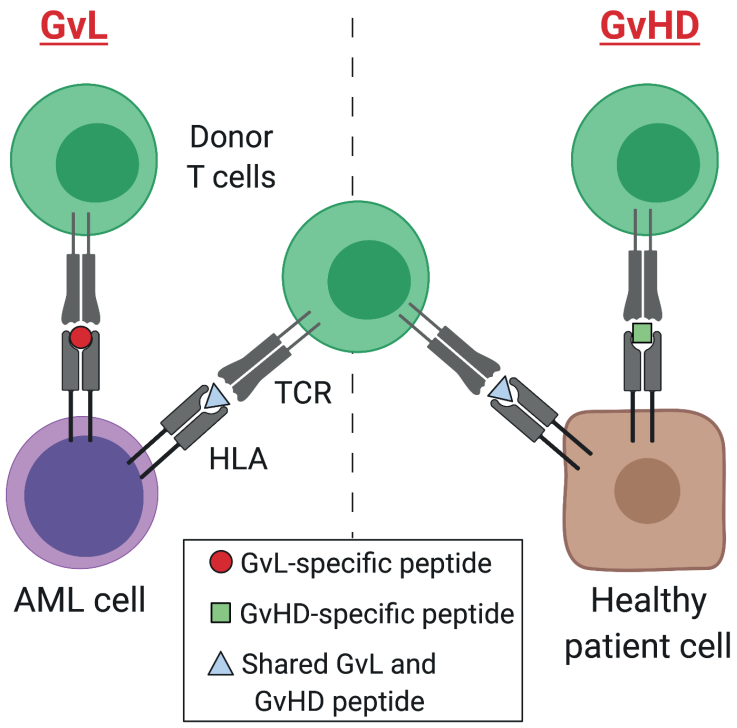


Figure 1.2: GvL and GvHD T cell responses. GvL and GvHD T cell responses. Separation of GvL from GvHD T cell responses according to tissue expression of immunogenic HLA-binding peptides. Peptides may result from germline differences between donor and recipient (minor histocompatibility antigens), somatic mutations (neoantigens), or overexpression of non-mutant peptides that are not expressed by healthy tissues (leukaemia-associated antigens).
Created with BioRender.com. Adapted from Sweeney and Vyas [1].

Studies comparing outcomes of AML patients treated with sibling or matched unrelated donor transplants have yielded conflicting results, with some reporting inferior survival with unrelated donors and others detecting no significant survival difference [50–52]. Although a more potent GvL effect may be expected with HLA-matched unrelated vs related donors, studies have failed to demonstrate a reduction

in relapse following unrelated donor transplants [51, 53]. It is possible that high resolution HLA typing and advances in supportive care have improved the outcomes of matched unrelated transplants in recent years by reducing transplant mortality [54]. However, 7/8 mismatched unrelated donor transplants (1 mismatched HLA allele) are associated with increased early mortality, but beyond 6 months survival rates are similar between donor types [55].

The most common source of donor stem cells is peripheral blood, which accounts for 81% allo-SCTs for AML in Europe [5]. Donors are treated with recombinant human granulocyte colony-stimulating factor (G-CSF) and sometimes with the CXCR4 antagonist plerixafor prior to harvesting by apheresis [56]. A bone marrow harvest is performed less commonly, by aspirating marrow from the pelvis, usually while the donor is under general anaesthetic. Some studies indicate a survival advantage using peripheral blood stem cell (PBSC) compared with bone marrow for treatment of malignant disease [57, 58], whereas others have shown no impact on OS [59, 60]. PBSC transplants have a higher T cell content and have been consistently associated with more rapid neutrophil and platelet engraftment. Lower relapse rates, suggestive of a GvL effect, was observed in a large randomised trial of PBSC vs bone marrow donation for haematological malignancy [58]. However, other studies have shown no effect. Consistent with the higher T cell content of PBSC grafts, a higher incidence of chronic GvHD has been consistently reported across studies and some studies also report an increase in acute GvHD [57–60]. Therefore, peripheral blood is more commonly used as the source of stem cells in the treatment of haematological malignancy due to the potential for a more potent GvL effect and the avoidance of general anaesthetic risks for the donor. In contrast, bone marrow harvest is generally preferred for the treatment of non-malignant conditions due to the reduced incidence of GvHD and the absence of benefit conferred by the higher number of alloreactive T cells in a PBSC graft [5].

1.6.2 T cell depletion and relapse

The commonest causes of death after allo-SCT for leukaemia are relapse followed by GvHD [61]. The role of T cells in GvHD was established in 1968 by van Dicke and colleagues, who used a murine transplantation model to show that mice transplanted with lymphocyte-depleted spleen fractions survived without GvHD, whereas mice receiving lymphocyte-replete fractions all died from GvHD [62]. When introduced in human transplants, it was hoped that performing T cell depletion would reduce morbidity and mortality from GvHD and also eliminate the requirement for post-transplant treatment of patients with cyclosporin/methotrexate immunosuppression. Whilst a low incidence of GvHD was confirmed with T cell depleted grafts, there was an increase in graft failure and disease relapse [63].

An early strategy for *ex vivo* T cell depletion of grafts was incubation with Campath-1H (alemtuzumab), the first humanised monoclonal antibody, together with complement from donor serum (Table 1.2) [64, 65]. Although this reduced the incidence of GvHD in patients transplanted for CML, the incidence of relapse approximately doubled [66]. Similarly, early experience in AML transplants found an increase in relapse with T cell depletion [45, 67]. Marmont *et al.* studied 1,154 AML patients treated with allo-SCT and found a 2.75-fold increased risk of relapse following T cell depletion. An increased incidence of graft failure was observed in both matched related and unrelated donor transplants, suggesting that donor T cells might be required to counterbalance the effect of recipient T cells rejecting the graft [68]. These findings suggested that *ex vivo* pan-T cell depletion strategies are not optimal even for unrelated donor transplantation [69]. An alternative method of T cell depletion uses $CD34^+$ selection of G-CSF-mobilised PBSCs, which results in 4–5 $\log_{10}$ reduction in T cells [14]. Due to significant reduction in GvHD, this strategy has enabled the transplantation of older patients. Despite concerns about increased relapse due to a reduction in donor T cells available for a GvL effect, this was not seen in a study of AML patients [70].

Other methods of negative selection have been used to balance the beneficial effects of donor T cells (a GvL effect and improved engraftment) against GvHD.

Depletion Strategy	Method
Ex vivo	
Positive selection	CD34 selection
Negative selection:	
- Pan-T cell	CD3 depletion
	Monoclonal anti-CD52 (Alemtuzumab)
- T cell subset	CD8 depletion
	CD3 / CD19 depletion
	αβ T cell / CD19 depletion
In vivo	
- Pre-transplant conditioning	Monoclonal anti-CD52 (Alemtuzumab)
	Polyclonal anti-thymocyte globulin (ATG)
	- Atgam® (horse)
	- Thymoglobulin (rabbit)
- Post-transplant	Cyclophosphamide

Table 1.2: Methods of T cell depletion. Depletion of T cells may be achieved by manipulating the stem cell graft *ex vivo* or by administering treatment to the transplant recipient *in vivo*.
Adapted from Sweeney and Vyas [1].

Treatment of 41 patients with $CD8^+$ depleted grafts led to engraftment of all recipients. However, GvHD risk was not reduced, with 61% grade II- IV reported, implying that other immune cell populations are also likely to be important for GvHD such as $CD4^+$ T cells or NK cells [71]. Combined elimination of T cells and B cells by targeting $CD3^+/CD19^+$ has been used to eliminate both T cells that are responsible for GvHD and B cells that can lead to post-transplant lymphoproliferative disorders. $CD3^+/CD19^+$ depletion was used in a study of 61 recipients of haploidentical transplants who lacked a HLA-matched donor. In this cohort of high-risk patients, 62% had AML and the incidence of grades II– IV acute GvHD and chronic GvHD were 46 and 18%, respectively and incidence of relapse was 31% [72].

The majority (95%) of circulating T cells have TCRs that comprise dimers of α and β glycoproteins and are implicated in the adaptive immune response that mediates GvHD and GvL. In contrast, $\gamma\delta$ T cells are part of the innate immune system and pre-clinical models have indicated they do not cause GvHD but promote

alloengraftment [73]. Given the protective role for $\gamma\delta$ T cells, there has been interest in specific $\alpha\beta$ T cell depletion. Reduction in acute leukaemia relapse was observed among recipients of $\alpha\beta$-depleted haploidentical donor grafts compared with pan-CD3 depletion [74]. The level of circulating $\gamma\delta$ T cells post-transplant correlates significantly with relapse-free survival, suggesting a GvL effect [75]. Furthermore, several subsequent studies have demonstrated anti-tumour activity of $\gamma\delta$ T cells against leukaemia and other cancers, highlighting their protective role [69, 76, 77].

An alternative approach is to T cell deplete *in vivo* by treating the recipient with either alemtuzumab [78] or anti-thymocyte globulin (ATG) [79] as part of pre-transplant conditioning, or by treating with post-transplant cyclophosphamide [80]. Alemtuzumab-containing regimens are consistently associated with a reduction in acute and chronic GvHD [81]. Malladi *et al.* retrospectively studied 88 AML patients treated with HLA-identical sibling transplants, with or without alemtuzumab conditioning. A significant reduction in chronic GvHD was seen and a trend toward increased relapse (35% with alemtuzumab and 19% in the untreated group) [82]. Furthermore, due to more pronounced immunosuppression compared with ATG, alemtuzumab is more commonly associated with reactivation of cytomegalovirus (CMV) and Epstein-Barr virus (EBV) [81, 83].

High-dose cyclophosphamide administered early post-transplantation depletes alloreactive T cells derived from host and donor. This approach has grown in popularity, particularly in the haploidentical allo-SCT setting, due to its efficacy at reducing GvHD and graft failure [84]. Cyclophosphamide selectively targets highly proliferative alloreactive T cells early post-transplant, while sparing the relatively quiescent non-alloreactive T cell and haematopoietic stem cell compartments [29]. Ciurea *et al.* reported a non-randomised comparison of 65 haploidentical transplants: T cell replete bone marrow transplants treated with post-transplant cyclophosphamide vs T cell depleted PBSC transplants without post-transplant immunosuppression. A significant reduction in chronic GvHD was observed in the T cell replete group (8 vs 18%) with an associated reduction in non-relapse mortality (16 vs 42%) and increase in OS at 1-year post-transplant (66 vs 30%).

Improved reconstitution of T cell subsets and a concomitant reduction in infectious complications was also observed in the T cell replete group [85]. Following its success in haploidentical allo-SCT, post-transplant cyclophosphamide has been demonstrated to provide effective GvHD prophylaxis and a good safety profile in the HLA-matched related and unrelated donor settings [86].

T cell depletion has been essential in enabling allo-SCTs to be delivered outside of the HLA-matched sibling setting, where the risk of GvHD would be otherwise high, such as mismatched sibling, unrelated donor and haploidentical settings. A balanced approach to T cell depletion is needed, tailored to the GvHD risk associated with the type of donor and the post-transplant immunosuppressive therapy used.

1.6.3 Donor lymphocyte infusions reinstate a GvL effect

Given the crucial role of a donor immune response in preventing relapse, in 1989 prophylactic infusions of donor lymphocytes were trialled immediately after transplant in AML patients in an attempt to establish a GvL response [87]. However, acute GvHD and mortality was increased. Subsequent work on CML in the 1990s established that using the original transplant donor and separating DLI from the transplant by at least 2 months produced complete haematological, cytogenetic and molecular remissions, indicating a robust GvL effect, often without accompanying GvHD [88, 89]. Of relapses, 67% achieved molecular remission following DLI and this translated into 95% 3-year survival. However, transplantation for CML in the current era has been marginalised by targeted tyrosine kinase inhibitor treatment, which is highly efficacious and spares patients the toxicity of transplantation.

Patients with AML or MDS have a lower response rate to DLI of 20–40% [90, 91]. A large retrospective analysis by EBMT of 399 post-transplant AML relapses, of whom 43% were treated with DLIs, showed a survival advantage at 2 years with DLI treatment (21% with DLIs vs 9% without) [92]. Although a GvL effect is evident from DLIs in AML, efficacy has been limited to patients with favorable cytogenetics or a low disease burden at relapse. AML cells are highly proliferative,

which may cause them to rapidly overwhelm the ability of donor T cells to exert immunological control in the presence of abundant disease.

Given that DLIs are most likely to exert a therapeutic effect with lower AML disease burden, there has been significant interest in incorporating prophylactic DLI into transplant regimens. For patients with high-risk AML, the FLAMSA regimen incorporated prophylactic DLI and was given in a non-randomised study at day +120 following stem cell infusion in patients who were not receiving immunosuppression and were free from GvHD [93]. Long-term follow-up revealed that at 7 years post-transplant, there was a significant survival advantage of DLI, with OS of 67% compared with 31% in a non-randomised control group [94]. A phase II randomised trial of prophylactic DLI post-transplant in high risk myeloid malignancy is currently underway (PRO-DLI, NCT02856464) [95].

1.6.4 Cure of patients treated with reduced intensity conditioning

Given that the median age of AML diagnosis is 65–70, delivery of transplants with intensive myeloablative conditioning would be associated with unacceptable toxicity and high transplant-related mortality in the majority of patients [8]. The development of RIC protocols has enabled allo-SCT to be delivered to older patients. This is particularly important because higher risk cytogenetic and molecular genetic features are commoner in older people and so this cohort would have a limited prognosis if treated with chemotherapy alone.

Myeloablative conditioning regimens use alkylating agents, with or without TBI, at doses predicted to prevent autologous haematopoietic recovery. In RIC regimens, doses of alkylating agents or TBI are reduced by at least 30% [96]. RIC must be sufficiently intense to prevent graft rejection by host haematopoietic cells, but preserve a GvL effect, which is especially important given that leukaemia cells are unlikely to be entirely eliminated by the conditioning treatment. A large retrospective analysis of 1,070 AML and MDS patients who received a RIC transplant in first complete remission demonstrated that this treatment can be safely delivered to older adults. The 2-year OS in AML patients aged 65 years or

older was 36%, which was not significantly different to survival of younger adults in a multivariate analysis [97]. Overall, compared with myeloablative transplants, most data show increased relapse following RIC conditioning, but this is likely to be offset by a reduction in non-relapse mortality, particularly in older adults [98].

1.7 Immune evasion mechanisms at post-transplant relapse

Immune evasion mechanisms at post-transplant relapse provide further evidence for the selection pressure exerted by a GvL effect. T cells play a major role in shaping the immunogenicity of cancers, which can lead to tumour cells altering their phenotype such that they are no longer controlled by the immune system, in a process known as immunoediting [99]. Loss of AML immunogenicity under donor immune selective pressure has been shown to lead to immune escape and relapse. There are multiple possible mechanisms that may be used by AML cells to evade the GvL response (Figure 1.3).

1.7.1 Reduction in HLA presentation

Following haploidentical transplantation for AML or MDS, disease from 5 of 17 patients who relapsed had genomic loss of the mismatched HLA haplotype due to acquired uniparental disomy of chromosome 6p [100]. This suggests that in these patients, major HLA mismatches were the critical targets of the GvL response and loss of mismatched HLA as a single large-scale genetic event was sufficient to evade immune detection. Genomic HLA loss has also been described following HLA-matched allo-SCT for AML [101, 102]. In a study of relapsed AML in 12 recipients of matched unrelated donor transplants, 2 patients were found to have focal deletions spanning HLA class I alleles. Given that HLA alleles were identical in patient and donor, it is likely that the pivotal GvL targets were peptide(s) presented by HLA. In this setting, loss of the HLA locus prevented these key peptides being presented to donor-derived T cells [102].

Recent studies have shown that patients with relapsed AML post-allo-SCT have dysregulation of pathways that influence immune function, including downregulation

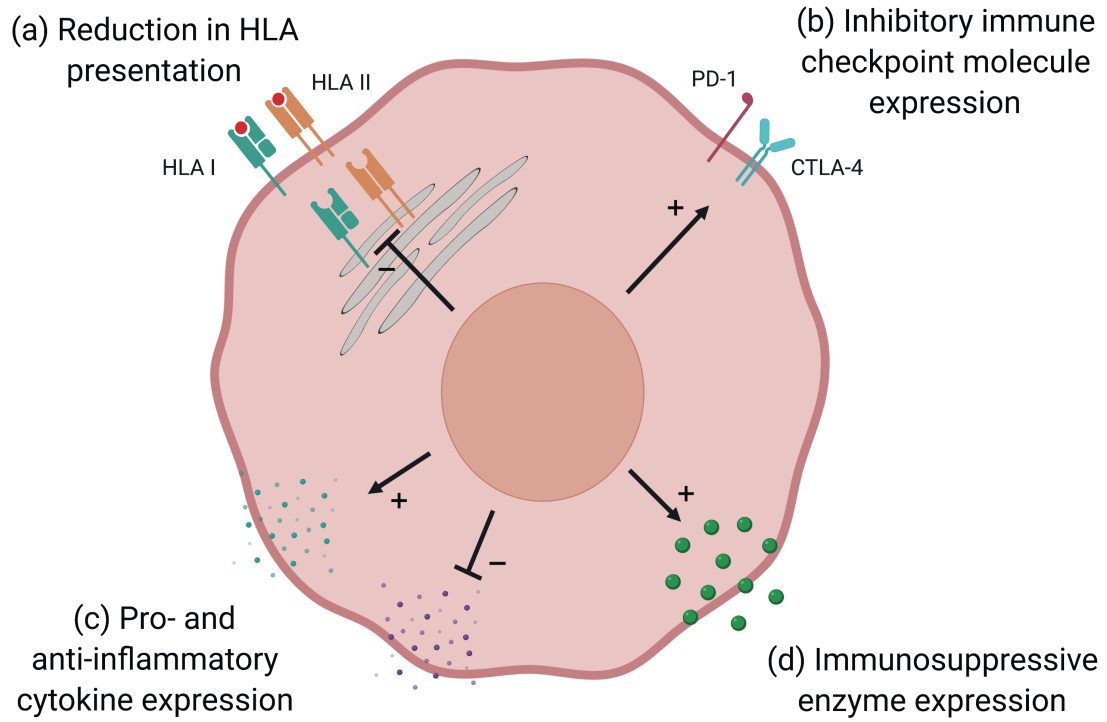


Figure 1.3: Possible mechanisms of post-transplant immune evasion. Loss of AML immunogenicity under immune selective pressure following allo-SCT leads to immune escape and relapse. There are multiple possible mechanisms: **(a)** Reduction in HLA presentation prevents donor-derived T cells identifying the AML cell. This is commonest in HLA-mismatched transplants and can result from genetic loss of part, or all, of the HLA locus. In other patients, downregulation of HLA expression via defects in transcriptional regulators may play a role. **(b)** Upregulation of immune checkpoint inhibitory molecules has been shown to suppress T cell responses at relapse in a subset of patients. **(c)** and **(d)** Expression of anti-inflammatory enzymes and cytokines, and suppression of pro-inflammatory cytokines, have an immunosuppressive effect in AML but their role has yet to be established in the allo-SCT setting.
PD-1: Programmed cell death-1, CTLA-4: Cytotoxic T-lymphocyte-associated protein 4
Created with BioRender.com. Adapted from Sweeney and Vyas [1].

of the expression of HLA class II genes and associated transcriptional regulators [42, 43]. Transcriptional downregulation of genes involved in antigen processing and HLA presentation, such as *IFI30*, *HLA-DMA*, *HLA-DMB*, and *CD74*, were observed at relapse in patients with reduced HLA class II expression. In contrast with genetic loss of HLA, reduction in HLA class II expression did not correlate with the number of donor-recipient HLA mismatches and also occurred after HLA-matched transplants. Reduction in HLA class II expression following a HLA-matched

transplant may favour immune evasion by substantially narrowing the repertoire of antigens presented to donor-derived T cells. Given that immune evasion mechanisms have involved genomic loss or transcriptional downregulation of both HLA class I and II molecules, roles for $CD4^+$ as well as $CD8^+$ T cells in GvL responses are likely.

1.7.2 Expression of immune checkpoint molecules

The balance of co-stimulatory and co-inhibitory signals conveyed to T cells influences whether the cell becomes activated when its TCR binds its cognate HLA-peptide complex. Upregulation of co-inhibitory signalling has been described in many cancers, including AML, and can result in a dysfunctional "exhausted" T cell state [103, 104]. In AML patients, $CD8^+$ cytotoxic T-cells expressing co-inhibitory receptors are functionally impaired and predict AML relapse [105]. Toffalori *et al.* demonstrated an increase in co-inhibitory ligand expression (PD-L1 and B7-H3) at post-transplant relapse compared with paired diagnostic samples, with a concomitant increase in the corresponding receptors on T cells [43]. Increased expression of co-inhibitory ligands and their receptors was mutually exclusive with downregulation of HLA class II, indicating distinct mechanisms of post-transplant immune evasion. Together, these modalities accounted for over two-thirds of relapses.

1.7.3 Immunosuppressive enzymes

Expression of metabolic enzymes by AML, such as arginase II [106] and indoleamine 2,3-dioxygenase 1 (IDO1) [107], has been shown to provide an immunosuppressive environment. AML blasts express and secrete arginase II, resulting in significantly elevated plasma levels and enhanced arginine metabolism compared with healthy controls. This is associated with impaired T cell proliferation and an immunosuppressive M2-like monocyte phenotype [106]. IDO1 is an interferon (IFN)-γ-inducible enzyme that metabolises tryptophan, leading to kynurenine production. Kynurenine inhibits effector T cell function and promotes the differentiation of immunosuppressive regulatory T cells (Tregs). Correspondingly, IDO1 expression by paediatric AML cells was associated with significantly worse 8-year event-free survival (16.4%) than non-expressing AMLs (48.0%) [107]. Although these

mechanisms have not been demonstrated in the allo-SCT setting, they could play a role in suppressing alloimmune GvL responses [108].

1.7.4 Alteration of pro- and anti-inflammatory cytokines

A potential method of immune evasion following allo-SCT is the modulation of pro- and anti-inflammatory cytokines. Suppression of pro- and elevation in anti-inflammatory cytokines are predicted to dampen effective GvL responses, though this has not been proven. AML cells express anti-inflammatory cytokines interleukin-4 (IL-4) and IL-10, both of which are known to reduce HLA class II expression [108, 109]. This mechanism has been shown to promote immune evasion in a mouse model of chronic lymphocytic leukaemia (CLL) [110].

Conversely, pro-inflammatory cytokines could promote GvL responses and therefore suppression of these cytokines might be used by leukaemia cells to evade alloreactivity. IL-15 is a pro-inflammatory cytokine expressed by multiple cell types, including dendritic cells and myeloid progenitors [111]. Reduced plasma levels early post-transplant is significantly associated with relapse, indicating reduced immune control of disease [112]. Furthermore, treatment of AML harbouring the internal tandem duplication in the gene encoding Fms-related tyrosine kinase 3 (FLT3) with the tyrosine kinase inhibitor, sorafenib, increases IL-15 expression in human AML. In a mouse model of AML with FLT3 internal tandem duplication (FLT3-ITD), sorafenib treatment promotes a GvL effect through IL-15 production [113]. A recent phase 1 trial evaluated the IL-15 superagonist complex, ALT-803, in patients with haematological malignancies who relapsed after allo-SCT. Responses were observed in 19% of evaluable patients with expansion and activation $CD8^+$ T cells and NK cells, without an increase in Tregs [114]. Further study is required to demonstrate whether therapeutic modulation of cytokines can enhance clinically relevant GvL effects and reduce relapse.

1.8 Candidate GvL antigens

Genetic differences between donor and recipient are crucial for a GvL effect, as is evident from the increased relapse rate seen in syngeneic twin transplants [45]. When allo-SCT is planned, a donor is sought with HLA-A, -B, -C, -DRB1, and -DQB1 alleles matching the recipient to reduce the risk of GvHD. Often HLA-DPB1 is not taken into account when selecting donors, as OS is unaffected by mismatching HLA-DPB1 alleles. However, mismatches have been associated with an increased risk of GvHD and reduction in relapse, consistent with a GvL effect, suggesting that mismatched HLA-DP may represent important antigens in otherwise HLA-matched allo-SCTs [115, 116].

Even in the setting of a fully HLA-matched transplant, immune reactivity against malignant cells provides a major contribution to disease control. Donor T cells recognise host antigens through interactions between their TCRs and peptide bound to a HLA molecule (Figure 1.4). A particular HLA allele, together with the peptide being presented, form a complex that is recognised by TCRs. T cells play a key role in GvL, even in the HLA-matched setting, which provides evidence that HLA-presented peptides are capable of stimulating an anti-leukaemic response [10, 117].

HLA-presented peptides responsible for eliciting a GvL response are likely to result from mismatched coding germline variants between patient and donor that are expressed in leukaemia cells [16]. These polymorphic peptides are known as minor histocompatibility antigens (miHAs). Most result from single nucleotide variants (SNVs) but insertions/deletions (indels) can also produce miHAs. In addition to germline variants, somatically acquired genetic variants may result in antigenic peptides, which are known as neoantigens. These are formed from SNVs, indels or chromosomal translocations acquired during leukaemia development. In order for miHAs or neoantigens to result from germline or somatic genetic variants, they must result in an amino acid change (non-synonymous variants), be expressed in AML cells and presented by HLA molecules. Another group of candidate GvL

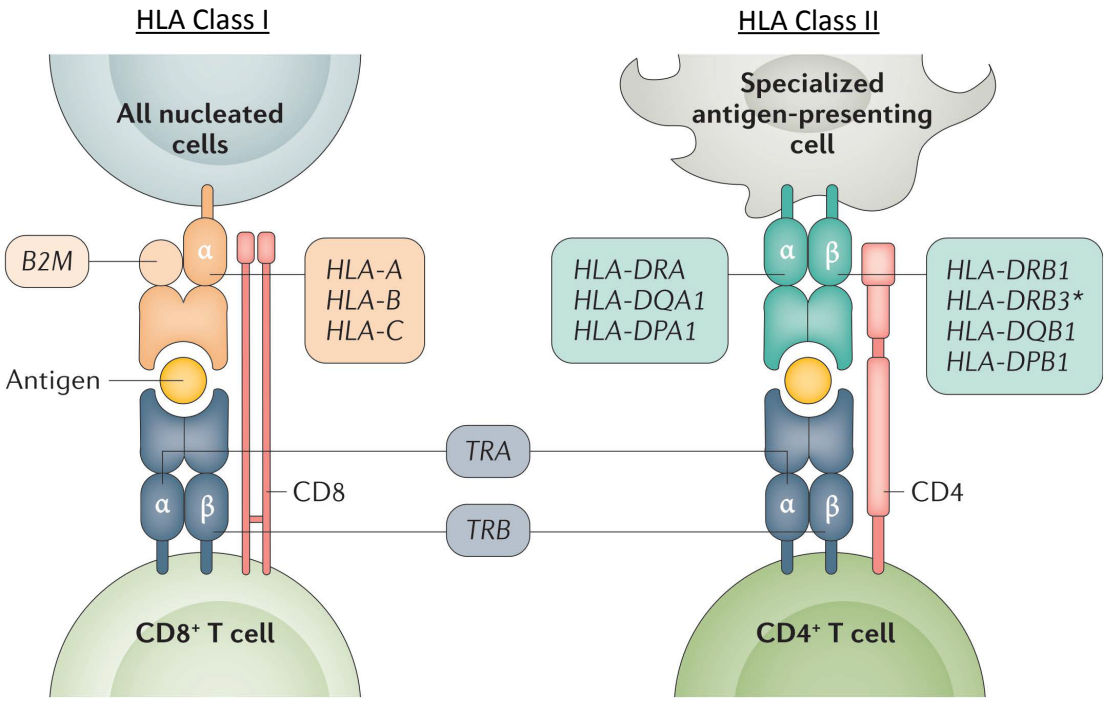


Figure 1.4: HLA presentation of antigen to T cells. All nucleated cells express HLA class I molecules on their surface, which are dimers comprising a polymorphic α chain and an invariant β2M chain. HLA class I molecules present peptide antigen to TCRs, in association with the CD8 co-receptor (Left). Professional antigen-presenting cells (i.e. dendritic cells, macrophages and B cells) express HLA class II molecules on their surface, which comprise α and β chains. HLA class II molecules present peptide antigen to TCRs, in association with the CD4 co-receptor (Right).
*Some individuals have several *HLA-DRB* genes, including *HLA-DRB4* and *HLA-DRB5*.
β2M: β2-microglobulin, TRA: T cell receptor α chain, TRB: T cell receptor β chain, TCR: T cell receptor
Adapted from Kornum et al. [118].

antigens are derived from proteins that are overexpressed in leukaemia cells, but are not expressed in healthy cells (leukaemia-associated antigens) [90].

Peptide presentation by HLA class I and II involves distinct intracellular pathways. HLA class I molecules present peptides derived from intracellular antigens, whereas HLA class II molecules are conventionally thought to present peptides from extracellular sources (Figure 1.5). In the HLA class I pathway, endogenous antigens are degraded into peptides by the proteasome. The peptides are then transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) protein. Immature HLA class I molecules are stabilised by the peptide loading complex, which comprises Tapasin, Erp57 and

Calreticulin (Crt). HLA class I complexes that are loaded with peptide antigen are transported to the cell surface through the Golgi apparatus. In the peptide HLA class II pathway, exogenous antigens are endocytosed. Immature class II complexes that are stabilised by the invariant chain (Ii) are released into MHC class II compartments (MIIC), where the Ii is cleaved. This leaves a short peptide sequence, class II-associated invariant chain peptide (CLIP), that is subsequently exchanged with the final binding peptide. Mature HLA class II molecules are then transported to the cell surface [119, 120].

In addition to presenting exogenous antigens, peptides derived from endogenous proteins can enter the MIIC through the autophagy pathway. Autophagy delivers cytoplasmic constituents for lysosomal degradation and resultant peptides can be presented by HLA class II. Autophagosomes form when an isolation membrane forms from membranes of ER, Golgi apparatus or the plasma membrane. Autophagosomes fuse with lysosomes, which deliver peptides into MIIC, where they can be loaded onto HLA class II molecules and presented on the cell surface [121].

Although evidence suggests that the HLA class II pathway is likely to be important in establishing a GvL response against AML [42, 43], it is not known whether presented peptides are derived from intracellular or extracellular proteins. If endogenous proteins can be loaded onto class II molecules, then pharmacological potentiation of autophagy could enhance tumour recognition by donor-derived T cells and strengthen a therapeutic GvL response [122].

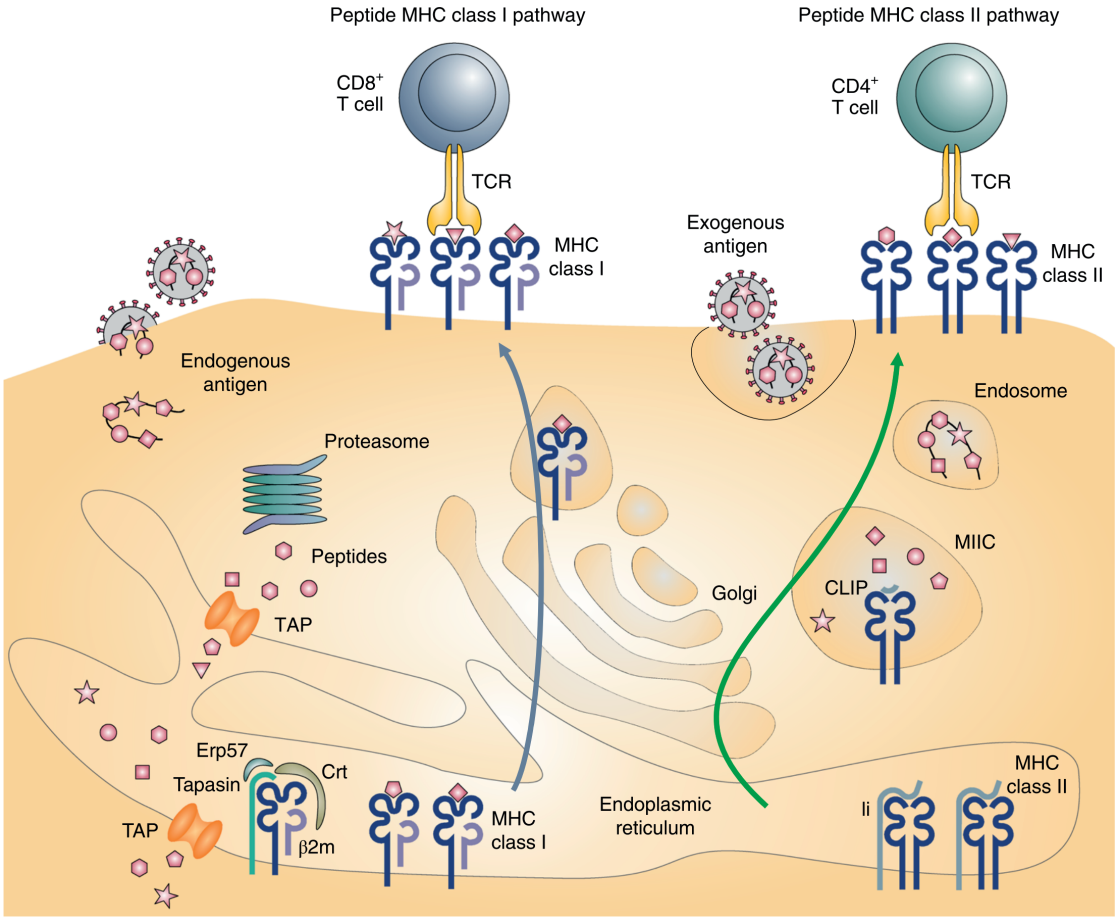


Figure 1.5: MHC class I and II antigen presentation pathways. MHC class I, known as HLA class I in humans, presents peptides derived from intracellular antigens, whereas MHC class II molecules present extracellularly-derived peptides. In the MHC class I pathway, endogenous antigens are degraded into peptides by the proteasome in the cytosol. The peptides are then transported into the ER by the TAP protein. Immature MHC class I molecules are stabilised by the peptide loading complex: Tapasin, Erp57 and calreticulin (Crt). MHC class I complexes that are loaded with peptide antigen are transported to the cell surface through the Golgi apparatus. In the peptide HLA class II pathway, exogenous antigens are endocytosed. Immature class II complexes that are stabilised by the invariant chain (Ii) are released into MIIC, where the Ii is cleaved, leaving a short peptide (CLIP) that is then exchanged with the final binding peptide. Mature MHC class II molecules are transported to the cell surface via the Golgi apparatus.
ER: Endoplasmic reticulum, TAP: Transporter associated with antigen processing, MIIC: MHC class II compartment, CLIP: Class II-associated invariant chain peptide
Adapted from Purcell and Ternette [120].

1.8.1 Minor histocompatibility antigens

Compared to neoantigens, there are a large number of mismatched germline variants between patient and donor, which suggests that miHAs might be more important in the GvL response due to the number of potential antigens presented to the donor immune system. From whole exome sequencing of nine patient-donor pairs, an average of 6,445 non-synonymous SNVs were found to be mismatched in an allo-SCT, which provides a large reservoir of potential miHAs [123]. Genomewide SNP array data from donors and their stem cell transplant recipients has demonstrated that the average mismatching for coding SNVs is 9.4% for HLA-identical sibling donors, increasing to 17.3% for HLA-identical unrelated donors [124]. Not all non-synonymous mismatched variants will produce immunogenic peptides. They must be expressed, processed, bind HLA and be presented at the cell surface. Approximately 1 in 200 non-synonymous SNPs is predicted to be expressed and capable of generating peptides that can be presented by HLA molecules [125]. After presentation, in order to generate an immune response, the HLA-peptide complex must bind a TCR and T cell activation is also influenced by the balance of co-stimulatory and co-inhibitory signals.

It is not known whether the number of miHAs triggering a GvL response in a given transplant is very large or whether a small number of antigens play a dominant role. miHAs that have a broad expression across tissues may result in GvHD, whereas haematopoietic-specific miHAs might result in a more targeted GvL response.

In early studies, miHA-specific cytotoxic T cells were identified from post-transplant samples by demonstrating selective cytotoxicity against host but not donor cells [126]. In this way, HA-1 to -5 antigens were described and were found to be recognised by T cells in a HLA-restricted manner [127]. To identify the HLA-A*02:01-restricted HA-2 epitope, HLA-bound peptides were eluted by acid treatment and fractionated by reversed-phase high-performance liquid chromatography (HPLC). Individual peptide fractions were tested for their ability to sensitise HLA-A*02:01 cells to lysis by an antigen-specific cytotoxic T cell clone. Tandem mass

spectrometry of the sensitising peptide fraction was then used to identify the amino acid sequence of the HA-2 antigen, which derived from a class I myosin protein [128].

Once it was appreciated that miHAs occurred as a result of mismatched germline genetic variants, more efficient methods were used to identify miHAs using genetic linkage analysis and genomewide association studies (GWAS) [129]. In most studies, miHA phenotypes were defined by assessing reactivity of miHA-specific T cell clones to EBV-transformed B cells derived from individual patients [130]. Phenotypes were then correlated with SNPs and microsatellite markers throughout the genome, to identify the genetic basis of miHAs [129]. This approach was successfully employed to identify the miHAs AAC-1 and ACC-2 resulting from SNPs in the *BCL2A1* gene [131], and LRH-1, from a single base deletion in *P2X5* leading to a frameshift [132].

Clinical data correlating individual miHA mismatches with relapse following allo-SCT are limited, as most studies assessing outcomes have focused on GvHD risk rather than anti-tumour effects [129]. However, a large cohort study of 849 patients treated with allo-SCT examined 10 previously described autosomal miHAs and correlated patient-donor mismatches with clinical outcomes [133]. They found that mismatched miHAs were associated with greater relapse-free survival only in patients with GvHD, but other studies failed to demonstrate this association, perhaps due to small sample sizes [129]. A study of three HA-1- and/or HA-2-positive patients with CML or myeloma demonstrated that expansion of miHA-specific T cells after DLI treatment from their miHA-negative donor correlates with clinical disease responses, suggestive of a GvL effect [134].

Gender mismatching between donor and recipient is associated with reduced survival, which supports a role for immune responses against antigens encoded on the Y chromosome. Stern *et al.* found that male recipients of haematopoietic stem cell grafts from a matched woman were more likely to experience severe acute GvHD than those receiving grafts from a man. In contrast, female recipients of male grafts had increased risk of graft rejection and worse survival than gender-matched recipients, likely as a result of immune reactivity in the host-versus-graft direction against Y-encoded proteins [135]. Evidence for a GvL effect mediated by reactivity

against Y-encoded antigens was provided by a study of 3,238 HLA-matched sibling allo-SCT recipients, which found that male recipients of female grafts had a lower risk of relapse than any other sex combination. Female/female and male/male transplants had statistically significant hazard ratios (HR) for relapse of 1.26 and 1.38, respectively, and female recipients of male grafts had a non-significant HR of 1.21 [136]. Based on *in vitro* T cell activation assays using samples from sex mismatched allo-SCTs, Y-encoded miHAs have been identified from several genes, including *DFFRY*, *JARID1D*, *DDX3Y*, *PCDH11Y*, and *UTY* [137, 138].

Antibodies against Y chromosome miHAs in female-to-male gender-mismatched transplants 3 months post-transplant have been shown to predict chronic GvHD [139]. This highlights the possibility that B cell responses may be important in GvHD/GvL responses and also demonstrates that antibodies against miHAs could represent biomarkers for these processes.

1.8.2 Neoantigens

In solid tumours with a high mutational burden, such as melanoma and lung cancer, tumour-specific somatic mutations have been shown to result in large numbers of predicted neoantigens [99]. In patients with non-small cell lung cancer, higher numbers of non-synonymous coding mutations were associated with improved response to immunotherapy with the programmed cell death-1 (PD-1) inhibitor, pembrolizumab [140]. The median number of non-synonymous somatic coding mutations in this lung cancer cohort was 200, most of which are predicted to be non-pathogenic "passenger" mutations. In contrast, AML has amongst the lowest mutation burden of any adult cancer, with an average of 13 mutations in genes, of which on average 5 are in genes that are recurrently mutated in AML patients [141]. Therefore, in AML the likelihood of mutations leading to pathologically or therapeutically relevant neoantigens is relatively low.

One proposed method of cancer immune evasion is the loss of mutations associated with neoantigens that trigger a T cell response. A study of paired non-small cell lung cancer samples taken before immune checkpoint inhibitor therapy and

at post-treatment relapse demonstrated that resistant clones had lost neoantigens that were present pre-treatment [142]. A study of AML patients similarly predicted neoantigens in diagnostic and relapse samples. However, they failed to find selection against putative neoantigens post-transplant, suggesting that neoantigens are not being targeted by donor-derived T cells [42]. This is likely to be because there is a smaller repertoire of somatic mutations available in AML to generate neoantigens.

Nevertheless, neopeptides resulting from common AML mutations have been shown to elicit T cell responses *in vitro*. Greiner and colleagues leveraged the SYFPEITHI database of known HLA-binding peptides to generate mutant NPM1 peptides and tested their ability to activate T cells from patients and healthy controls using Cr^{51} release and ELISpot assays for IFNγ and granzyme B [143]. They identified two immunogenic peptides that activated $CD8^+$ T cells from 33 to 44% of AML patients tested. It has similarly been demonstrated that a FLT3 internal tandem duplication (ITD) neopeptide can also induce HLA-restricted autologous $CD8^+$ T cell responses *ex vivo* [144].

1.8.3 Leukaemia-associated antigens

Leukaemia-associated antigens (LAAs) are expressed in leukaemia but not in most healthy tissues. Candidate LAAs, such as the cancer-testis family of antigens, may be expressed in immune sanctuary sites such as testes, or in a developmental time-restricted manner (e.g. placenta). Other antigens are not leukaemia-specific and are expressed at low levels in normal tissue (e.g. adrenal glands, ovaries and endometrium) but may be overexpressed in leukaemia. These antigens include PRAME (preferentially expressed antigen of melanoma), WT1 (Wilms' tumour), hTERT (human telomerase reverse transcriptase), RHAMM (receptor for hyaluronan-mediated motility), G250/CA9 (an LAA derived from carbonic anhydrase), survivin and proteinase-3 [145, 146]. Outside the transplant setting, LAA expression has been found to correlate with improved OS in AML. In a microarray analysis of 116 AML samples, high levels of mRNA for at least one of three LAAs, G250/CA9, PRAME, or RHAMM was associated with significantly

improved OS. Specific T cell responses were demonstrated against LAAs using Cr^{51} and ELISpot assays in 47–70% cases, suggesting that the presence of these antigens could potentially enable effective immune surveillance and disease control.

Unlike miHAs and neoantigens, LAAs are proteins that could be expressed by the transplant donor and therefore central T cell tolerance may prevent these proteins from generating a GvL response.

1.8.4 NK cells and the GvL effect

In addition to the established role of T cells in the GvL response, correlative evidence suggests that natural killer (NK) cells may also play a role as effector cells. The relative contributions from T and NK cells in a given transplant may be influenced by multiple factors, including mode of T cell depletion, conditioning regimen and post-transplant immunosuppression [147]. NK cells are lymphocytes that form part of the innate immune system and are the earliest lymphocytes that recover after haematopoietic stem cell transplantation. Therefore, they are available to exert protective anti-microbial and GvL effects before T cell reconstitution. Indeed, high levels of donor NK cell chimerism early post-transplant (day +14) have been associated with a lower rate of relapse [148].

NK cell effector function is regulated by the balance of signalling from activating and inhibitory surface receptors [149]. NK cells were initially functionally characterised by their ability to kill tumour cell lines that lack expression of MHC class I molecules, which led Ljunggren and Karre to propose the "missing self" hypothesis, whereby NK cells kill targets due to the lack of MHC class I presentation of self-peptides [150]. This recognition occurs from the interaction between inhibitory killer cell immunoglobulin-like receptors (KIRs) on NK cells and MHC class I on target cells. In addition to inhibitory KIR receptors, NK cells express a variety of additional inhibitory (e.g. CD94/NKG2A) and activating receptors (e.g. activating KIRs, CD94/NKG2C). When an NK cell encounters a target cell, the balance of signals from these receptors determines whether the NK cell exerts a cytotoxic effect.

Downregulation of class I molecules is an important mechanism for evading T cell immune surveillance that has been described in several tumours [151, 152]. However, this mechanism can lead to NK cell recognition and activation of cytotoxicity. A counter-mechanism employed by some tumours is upregulation of the non-canonical HLA-E that bind CD94/NKG2 receptors on NK cells and prevent tumour killing [153].

NK cell alloreactivity was demonstrated by Ruggeri *et al.*, who studied 57 transplanted AML patients and showed that KIR ligand incompatibility in the graft-vs-host direction predicted a reduction in relapse and improved survival [154]. KIR genes are polymorphic and different genotypes have been linked to GvL effects in AML. There are two kinds of haplotype, A and B, which are distinguished by the composition of genes for activating and inhibitory receptors. Type B haplotypes contain more activating genes and correspondingly have been linked to stronger cytotoxic reactions against both virus-infected and malignant cells [44]. Significant associations between donors with a type B KIR haplotype and reduced AML relapse post-transplant have been shown, which suggests that a clinically relevant GvL effect might be mediated by NK cells [155].

1.9 Identifying GvL antigens and T cells

1.9.1 Identifying GvL antigens

Although allo-SCT has been administered as a treatment for AML for over 50 years, our understanding of the antigenic basis of the therapeutic GvL effect is very limited. The current study identifies miHAs that are important in GvL responses by applying exome-based antigen prediction methods that have previously been employed for the identification of neoantigens. Studies to identify neoantigens have used either whole exome or RNA sequencing data from tumour cells to identify non-synonymous mutations [140, 156–159]. Sequencing of a germline sample from the same patient is used as a control to enable somatically acquired mutations to be distinguished from germline variants. RNA sequencing is used to filter expressed variants, which have the potential to be presented by HLA. Artificial

neural networks trained on HLA binding data have been used to predict which variants could bind HLA with high affinity [160, 161].

By applying these principles to the allo-SCT setting, miHAs can be predicted by identifying germline variants that are mismatched between a patient and their transplant donor. Candidate GvL antigens are those that are present in the patient but not the donor, are expressed by AML cells and are predicted to bind HLA.

HLA-binding predictions have limited accuracy and even if a peptide does bind HLA, this does not indicate that it can stimulate a T cell response. Therefore, to conclusively identify GvL antigens, functional validation is imperative. Putative HLA class I-binding neoantigens have been tested for their ability to elicit functional T cell responses using ELISpot assays, intracellular cytokine staining and fluorescent- and barcode-labelled tetramers [162]. However, only 4% of tested peptides elicited a T cell response [163]. This low rate highlights that additional factors beyond HLA binding affinity play an important role in determining a peptide's immunogenicity, which likely include protein degradation, peptide processing and immune tolerance. If mutant and wildtype peptide sequences are similar, immune tolerance to the wildtype sequence may also result in tolerance to the mutant sequence, reducing its immunogenicity. This factor is incorporated into a refinement to HLA class I binding predictions, by adding weighting for differences in binding strength between mutant and wildtype peptide sequences [164]. Despite the relatively low validation of T cell responses for individual peptides, van Buuren and colleagues were able to detect neoantigen-specific T cell responses in 6 of 8 melanoma patients studied [165].

As discussed in Section 1.7.1, peptides presented by class II are likely to play a key role in eliciting a GvL response against AML [42, 43]. However, *in silico* HLA-binding predictions are less robust for class II than class I, which poses a challenge to using this approach to identifying important neoantigens or miHAs. Computational methods trained on *in vitro* binding data are limited by insufficient training data and algorithmic constraints. Improvements in training datasets and machine learning frameworks have improved the accuracy of HLA-binding predictions, but these remain imperfect [166, 167].

Peptides bound by HLA on the surface of tumour cells can be directly assayed by first immunoprecipitating the HLA molecules and eluting bound peptides using acid denaturation. The peptides are then separated into fractions by high performance liquid chromatography (HPLC) and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify the amino acid sequence [120]. By integrating this data with genetic variants identified by next-generation sequencing, HLA-presented neoantigens have previously been identified [168]. In the current study, mismatched DNA variants between patients and their respective allo-SCT donors was used to identify candidate miHAs that could result in a GvL T cell response. Functional peptide-specific T cell responses can be elicited by a greater proportion of the candidate neoantigens identified using this immunopeptidomic approach than using *in silico* HLA-binding predictions, when assessed by *in vitro* and *in vivo* mouse studies [168, 169].

Although immunopeptidomics provides an unbiased analysis of the peptide cargo of HLA molecules, there are two major limitations. First, the number of peptides identified is restricted by the amount of cellular material available for analysis and with small numbers of cells, only the most abundant and readily ionisable peptides can be identified. Another limitation is that due to the physico-chemical properties of peptides, some have restricted capacity to ionise and therefore may be undetectable by LC-MS/MS [120].

1.9.2 Identifying antigen-specific T cells

The identification of T cells responding to specific antigens has several scientific and clinical applications. By analysing the phenotype of antigen-specific T cells, insights can be gained into the characteristics of adaptive immune responses. Additionally, it allows antigen-specific T cells to be isolated. T cells specific for tumour antigens can be expanded *ex vivo* and then reinfused to augment anti-tumour immune responses, which is an approach that has been used for the treatment of different cancers [170–174]. Furthermore, the amino acid sequence of antigen-specific TCRs can be identified from single cell RNA sequencing. By engineering T cells to express

these TCR sequences, adoptive T cell therapies can be produced to treat cancers that express the antigen [175, 176].

Antigen-specific T cells can be identified by flow cytometry using fluorescent HLA-peptide multimers (Figure 1.6). TCRs have a relatively low affinity for their cognate HLA-peptide complex, but by combining multiple HLA-peptides in a single complex, the affinity is greatly enhanced. The most commonly used multimer is a tetramer and these can be used to track antigen-specific T cell responses directed against tumours. By combining tetramer staining with other markers, the phenotype of these cells can be characterised by flow cytometry [177, 178].

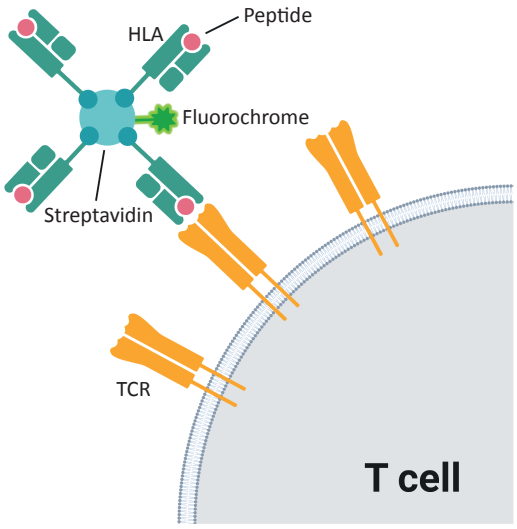


Figure 1.6: Fluorescent HLA-peptide tetramer. HLA molecules are biotinylated and bound to streptavidin, forming a tetravalent structure. They are conjugated with a fluorochrome, which allows bound T cells to be identified and isolated by flow cytometry. *Created with BioRender.com.*

Another way of identifying antigen-specific T cells is by stimulating a mixed population of T cells with the antigen and assessing which cells express activation markers by flow cytometry. IFNγ is a cytokine that is produced in response to a wide range of stimuli and is commonly used as an activation marker [179]. T cells can be sorted according to either multimer binding or activation in response to antigen. As well as enabling the development of engineered T cells, the identification of antigen-specific TCR sequences provides a means of tracking T cell responses directed against cancer. Targeted sequencing of TCRs from bulk samples taken at

different timepoints during a patient's treatment is performed. Then, the frequency of the antigen-specific TCR within the bulk repertoire is assessed. This method has been used to demonstrate that the frequency of tumour antigen-specific T cell clones increases with disease burden, reflecting increased antigenic stimulation [180].

1.10 Improving transplant outcomes in AML

To maximise the anti-leukaemic effect of allo-SCT, multiple factors must be tailored to the individual patient at different stages of treatment (Figure 1.7). Positive MRD prior to transplant has consistently been associated with inferior outcomes across multiple studies involving both myeloablative and RIC protocols [181] and using sibling, unrelated, and haploidentical donors [182, 183]. Therefore, the key aim of pre-transplant treatment is to achieve as deep a remission as possible. As discussed above, there is substantial risk of morbidity and mortality from the toxicity of transplant conditioning and from the immunological effects of GvHD. Therefore, when selecting the conditioning regimen and stem cell donor, these risks are balanced against the reduction in relapse that is achieved from intensive conditioning and a GvL effect.

Identification of GvL antigens is essential to understand *in vivo* alloimmune anti-leukaemic responses. Elucidating key miHAs may enable improved donor selection such that a genetic mismatch between patient and donor can produce a miHA that is predicted to elicit a T cell response. Whether a mismatch is also likely to result in GvHD may be guided by tissue-specific gene expression data, thus enabling enhancement of GvL in the absence GvHD.

Despite an effective GvL response, AML may relapse post-transplant due to the rapid proliferation overwhelming the protective GvL response. MRD assessment post-transplant may predict impending morphological relapse and although not yet established practice, this may enable treatment to be initiated when the disease burden is relatively low and easier to control. This may entail either:

1. Administering post-transplant anti-AML therapy that slows the pace of AML cell proliferation, to permit an effective donor GvL effect to take hold.

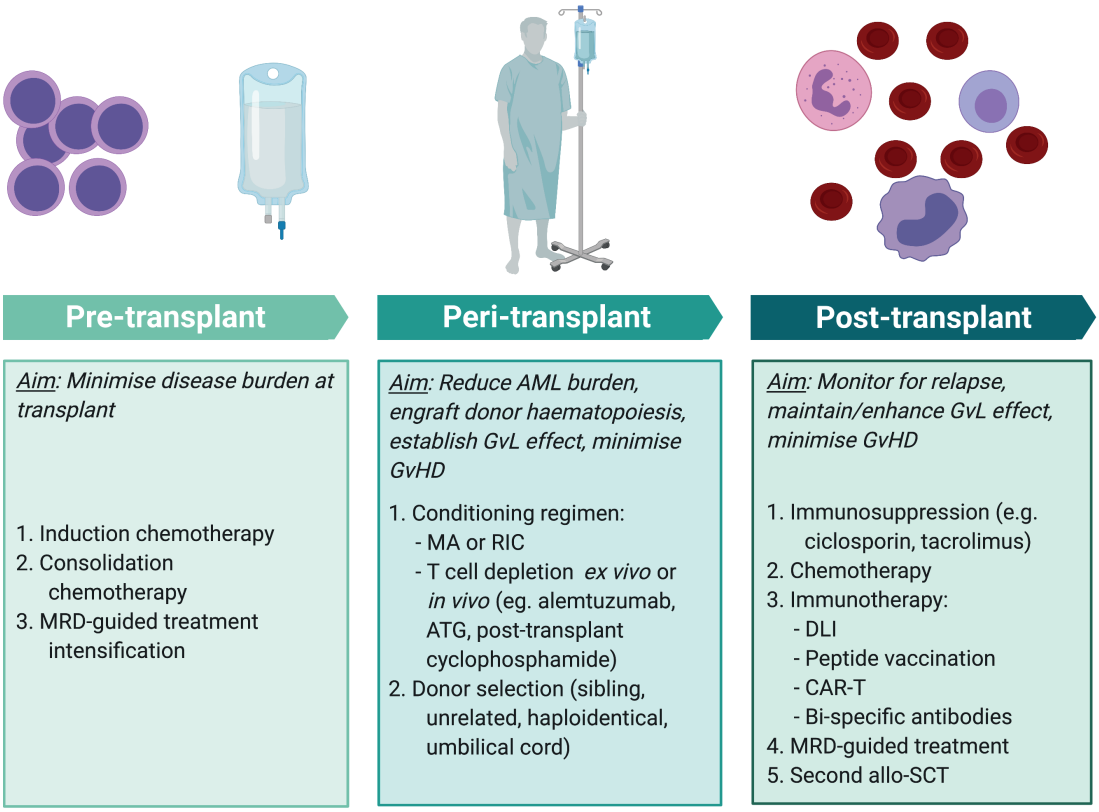


Figure 1.7: Methods of improving transplant outcomes in AML. Transplant outcomes for AML patients can be improved by optimising pre-, peri- and post-transplant factors for individual patients. MRD-guided treatment and some immunotherapy modalities (peptide vaccination, CAR-T, and bi-specific antibody therapy) are not currently standard practice.
MRD: measurable residual disease, MA: myeloablative conditioning, RIC: reduced intensity conditioning, ATG: anti-thymocyte globulin, DLI: donor lymphocyte infusion, CAR-T: chimeric antigen receptor T cell therapy.
Created with BioRender.com.

2. Using immunotherapy to enhance the alloreactive effect. This may be targeted to known antigens or employ non-specific immune activation, such as by DLI or immune checkpoint blockade.

Traditionally, post-transplant AML relapses have been treated with chemotherapy, but response rates to treatment are modest and remissions are usually short-lived. Amongst patients who achieve a complete remission with chemotherapy, consolidation treatment with donor cellular therapy, comprising either DLI or a second allo-SCT in selected patients, significantly improves 2-year OS (55% in patients who received cellular therapy vs 20% in those who did not) [184]. However,

many patients who relapse following allo-SCT are unable to tolerate further intensive chemotherapy. Treatment with the hypomethylating agent, azacitidine, is associated with relatively low toxicity and has demonstrated efficacy in this setting. Craddock *et al.* studied treatment of 181 patients with relapsed AML and MDS post-transplant and found that 25% responded to azacitidine but concurrent administration of DLI did not improve response rates [185]. It has yet to be established whether, in the absence of relapse, post-transplant maintenance chemotherapy has a role in improving outcomes. Treatment with CC-486, an oral formulation of the DNA methyltransferase inhibitor azacitidine, was well tolerated and associated with a 1-year survival of up to 86% in a single-arm, non-randomised study of AML and MDS patients in post-transplant remission [186]. A phase III randomised trial, AMADEUS, is currently evaluating whether maintenance therapy improves relapse-free survival (EudraCT 2018-001012-30) [187]. Azacitidine increases the expression of epigenetically silenced antigens and has been shown to result in upregulation of the cancer-testis antigen, MAGE [188]. A related hypomethylating agent, decitibine, has been shown to induce an activated cyototoxic phenotype in $CD8^+$ T cells, which have enhanced killing capacity [189]. Therefore, in addition to a direct effect on AML cell growth, azacitidine may also enhance the T cell-mediated GvL response.

Our deepening understanding of the molecular landscape of AML in recent years has led to the development of targeted therapies. In relapsed FLT3-ITD AML post-transplant, treatment with the tyrosine kinase inhibitor sorafenib has demonstrated efficacy, both alone and when combined with azacitidine or DLI [190, 191]. The second-generation tyrosine kinase inhibitors, quizartinib and gilteritinib, exhibit higher specificity for FLT3 and may be more efficacious [192]. A randomised phase III trial of single agent quizartinib in relapsed and refractory AML demonstrated higher CR rates (48 vs 27%) and prolonged OS (hazard ratio 0.76, 95% CI 0.58–0.98) compared with salvage chemotherapy. However, although patients who relapsed after allo-SCT were included in the trial, this sub-group was not reported separately [193]. An ongoing randomised, placebo-controlled phase II trial is currently assessing

gilteritinib maintenance therapy, starting at 30–90 days following allo-SCT and continuing for 24 months (NCT02997202) [194].

Mutations in isocitrate dehydrogenase (IDH) enzymes are found in 5–15% of patients with AML and inhibitors to both IDH1 (ivosidenib) and IDH2 (enasidenib) mutant enzymes have been approved by the FDA for treatment of relapsed and refractory AML patients. In this group of patients who are challenging to treat, they have achieved complete remission rates of 30.4% (ivosidenib [195]) and 26.1% (enasidenib [196]). Both drugs are currently being assessed as maintenance therapy following allo-SCT in patients with AML harbouring the relevant mutations (NCT03564821, NCT03515512) [197, 198].

Non-targeted augmentation of post-transplant immune responses using DLI is an established therapy that improves outcomes in some AML patients who relapse post-transplant [92]. Another method of non-specific immune activation is using antibodies against immune checkpoint inhibitors, such as PD-1 or CTLA-4 (cytotoxic T-lymphocyte-associated protein 4). Given their success in treating a range of cancers [199–201], they have also been used in recent studies to treat patients with haematological malignancies following allo-SCT. Although these might enhance the GvL response and thereby improve disease control, non-specific immune activation leads to substantial morbidity and mortality from GvHD [202]. Concern has also been raised about the possibility of increased risk of veno-occlusive disease (VOD) of the liver with immune checkpoint inhibitor therapy. A retrospective study of lymphoma patients treated with PD-1 blockade prior to RIC allo-SCT found an increased rate of severe VOD compared with a retrospective series of RIC transplants (8 vs 2.1%) [203]. Therefore, it is currently unclear whether immune checkpoint inhibitors could form part of the repertoire of post-transplant therapies for AML.

1.11 GvL antigens are potential targets for immunotherapy

As well as being of fundamental importance in understanding the immune control of AML following allo-SCT, the identification of GvL antigens has substantial

clinical significance. Patients and donors are currently HLA-matched to reduce the likelihood of GvHD. However, matching is not performed to select a donor whose graft is more likely to elicit a therapeutic GvL effect. If miHAs important in eliciting a GvL response were known, a donor could be selected who is mismatched with the patient for the relevant DNA variant and therefore is predicted to result in an alloreactive T cell response.

GvL antigens also represent attractive targets for immunotherapy. Whenever an allo-SCT is delivered, many different donor-derived T cell clones are introduced into the patient. By selecting a target antigen that is expressed by AML cells but is either absent or expressed at a low level in normal tissues, GvHD morbidity may be minimised. As GvL antigens in AML are likely to be miHAs resulting from germline DNA variation, they may also be targetable in other cancers that express the appropriate gene.

One method of producing a targeted GvL response is by vaccination with peptide antigen after transplantation. This leads to activation and proliferation of antigen-specific T cell clones. These T cells then target leukaemia cells that also present the peptide, leading to immune control of disease. WT1 is a leukaemia-associated antigen that is highly expressed in many cancers, including most cases of AML [204]. A polyvalent vaccine has been used in early phase clinical trials of AML in complete remission outside the transplant setting, with functional T cell responses demonstrated *in vitro* by IFNγ ELISpot and an increase in WT1 tetramer-positive T cells [205, 206]. In contrast to WT1, GvL antigens are likely to be germline-derived miHAs. To target these with peptide vaccination, treatment would have to be in the context of allo-SCT with a donor lacking the miHA. This is because the source of responsive T cells must be derived from an individual lacking the miHA. Otherwise, negative selection in the thymus is expected to have eliminated any T cells expressing TCRs that recognise the antigen.

An alternative method of targeting antigens expressed by HLA on cancer cells is using adoptive T cell therapy. T cells can be engineered to express antigen-specific TCRs and then administered to the patient. Chapuis *et al.* targeted

WT1 post-transplant using a TCR gene therapy approach. A high affinity WT1 antigen-specific TCR was inserted into donor $CD8^+$ T cells, which were infused prophylactically following allo-SCT into 12 patients. No relapses were observed after median follow-up of 44 months [178]. Pre-clinical studies have used lentiviral transfer of TCRs and demonstrated killing of AML by targeting neoantigens derived from *NPM1* mutations and the *CBFB-MYH11* gene fusion that results from the inv(16) chromosomal translocation. These genetic abnormalities are found in approximately 30% and 7% of adult AML cases, respectively. Killing of AML cells was demonstrated *in vitro* [175, 176] and *in vivo* in a mouse xenograft model of AML [176].

Alternative modes of immunotherapy that direct T cells against specific antigens could also be employed to enhance the GvL effect, including bi-specific antibodies [207] and chimeric antigen receptor T cell (CAR-T) therapy [208]. Unlike TCR-based therapies, these therapies can target surface molecules that are not limited to HLA-peptide complexes and therefore they are not restricted by the HLA alleles of the patient.

1.12 Aims

The overall aims of the study were:

(1) To establish a workflow to predict alloreactive antigens following allo-SCT that are responsible for therapeutic GvL T cell responses in AML patients.

(2) To functionally screen predicted alloreactive peptides for their ability to activate donor-derived T cells following allo-SCT.

(3) To characterise the antigens responsible for the GvL response to AML in patients, including the nature of antigens, the phenotype of alloreactive T cells and the HLA restriction of responses.

(4) As a proof of concept to demonstrate that these alloreactive antigens serve as therapeutic targets, by developing a T cell therapy using T cells that express antigen-specific TCRs.

Chapter 2

Materials and Methods

Contents

2.1 Collection and processing of human blood and bone marrow . . . 44
2.2 Flow cytometry staining and sorting of leukaemia and T cells . . . 44
2.3 Whole exome sequencing . . . 45
2.4 Bulk RNA sequencing . . . 47
2.5 HLA binding prediction . . . 48
2.6 Immunopeptidomic analysis . . . 49
2.6.1 Antibody purification and resin preparation . . . 49
2.6.2 HLA immunoprecipitation and elution of peptides . . . 50
2.6.3 Liquid chromatography-tandem mass spectrometry . . . 50
2.6.4 Data analysis . . . 51
2.7 *Ex vivo* and cultured IFNγ ELISpot peptide screening 52
2.8 Intracellular cytokine staining after peptide stimulation 53
2.9 HLA allele restriction and epitope mapping . . . 54
2.10 Isolation of antigen-specific T cells and sequencing their T cell receptors (TCRs) . . . 55
2.11 Processing and analysis of single cell TCR-Seq and RNA-Seq data . . . 56
2.12 Bulk TCR sequencing . . . 57
2.13 PADI4 genotyping . . . 58
2.14 Assessing PADI4 expression . . . 58
2.14.1 RT qPCR for *PADI4* gene expression . . . 58
2.14.2 Western blotting for PADI4 . . . 59
2.15 Virus production and transduction . . . 59
2.15.1 TCR cloning . . . 59
2.15.2 Retrovirus production . . . 60
2.15.3 Retroviral transduction of T cells . . . 60
2.15.4 Lentivirus production . . . 61
2.15.5 Lentiviral transduction of T cells . . . 61
2.16 *In vitro* T cell activation and killing assays . . . 62
2.16.1 T cell activation assays . . . 62
2.16.2 Generating monocyte-derived dendritic cells . . . 63
2.16.3 Killing assays . . . 63
2.17 *In vivo* killing of AML . . . 64

2.1 Collection and processing of human blood and bone marrow

Bone marrow (BM) and peripheral blood (PB) samples were obtained from patients with AML with informed consent (UK protocol MREC 06/Q1606/110). Samples were processed within 48 hours of collection and Ficoll gradient separation used to isolate mononuclear cells (MNCs). Fresh PB/BM was diluted in phosphate-buffered saline (PBS; Thermo Fisher Scientific) in a 1:1 ratio. This diluted solution was layered onto Ficoll-Paque (Sigma-Aldrich) in a 2:1 ratio, centrifuged at 400 *g* for 30 minutes at room temperature with no brake and the interphase of MNCs collected. Samples were cryopreserved in fetal bovine serum (FBS, Sigma-Aldrich) with 10% dimethyl sulfoxide (DMSO) and stored in liquid nitrogen in gaseous phase. Prior to use, vials were thawed at 37°C in a water bath and washed twice with media prior to culture or were directly used in an experiment.

For retroviral and lentiviral transduction of human T cells, fresh PB samples from healthy donors were obtained with informed consent (Weatherall Institute of Molecular Medicine phlebotomy consent). Ficoll gradient separation was performed, as indicated above, before use. Samples were not frozen.

2.2 Flow cytometry staining and sorting of leukaemia and T cells

Antibodies used for all flow cytometry analysis and sorting experiments are listed in Table B.1.

Diagnostic AML and pre-transplant remission BM samples were thawed as outlined in Section 2.1. Live cells were counted using trypan blue exclusion and stained with purified mouse anti-human antibodies against: CD3 (PE Cy5), CD45 (APC eFluor780) and CD235ab (Pacific Blue). Antibody incubations were carried out in flow cytometry staining buffer: Iscove's Modified Dulbecco's Medium (IMDM, Thermo Fischer Scientific) + 10% FBS + 20 μg/ml DNAse for 20 minutes at 4°C

in the dark. Up to 5x10^6 cells were stained in 100 μl volume and scaled up, as appropriate, based on cell number. Stained cells were filtered through a 35 μm strainer. Following antibody incubations, cells were washed with 1 ml flow cytometry staining buffer, centrifuged at 300 *g* for 5 minutes and resuspended in flow cytometry staining buffer containing the live/dead stain Hoechst 33258 (final concetration 1 μg/ml; Invitrogen). Unstained, single-stained and fluorescence-minus-one (FMO) controls were used to determine background staining and compensation in each channel. Single-stained controls used the Anti-Mouse Ig, k/Negative Control Compensation Particle Set (Becton Dickinson, Oxford UK) and a Hoechst-only stained sample.

Flow cytometry cell sorting was performed using the BD Fusion instrument with an 85 μm nozzle and analysis with FACS Diva v8.1 and FlowJo v10.7.1 software. Doublets and dead cells were excluded. AML cells from diagnostic BM samples were sorted as Hoechst$^-$CD235ab$^-$CD45$^+$CD3$^-$. Where possible, pre-transplant remission samples were used to obtain T cells but where unavailable, diagnostic samples were used to sort T cells: Hoechst$^-$CD235ab$^-$CD45$^+$CD3$^+$. The leukaemia and T cell sorting purity were assessed by sorting cells in the appropriate gate and re-analysing sorted cells.

For whole exome sequencing, cells were sorted into 1.5 ml centrifuge tubes and placed on ice. After sorting, samples were centrifuged at 350 *g*, the supernatant was discarded and cell pellets were stored at -20°C prior to DNA extraction. For bulk RNA sequencing of leukaemia cells, 5000 cells were sorted directly into 75 μl RLT buffer (Qiagen) with 0.75 μl β-mercaptoethanol 14.3M (Sigma-Aldrich). Samples were snap frozen using dry ice and stored at -80°C until further processing.

2.3 Whole exome sequencing

DNA was extracted from cell pellets of FACS-purified populations using the Qiagen QIAamp DNA Micro kit and quantification was performed using High Sensitivity DNA Qubit (Thermo Fisher Scientific). Quality was assessed by 1% agarose gel electrophoresis (100V, 400mA, 40 minutes) of 20ng DNA. DNA from

allo-SCT donors was obtained from the National Health Service (NHS) Transplant Immunology and Immunogenetics Laboratory (Oxford, UK). For each sample, a minimum of 100ng of DNA was used to produce DNA libraries by a commercial vendor (Novogene, https://en.novogene.com). DNA was fragmented by Covaris shearing, Illumina adapters added and whole exome capture performed using biotinylated RNA baits (Agilent SureSelect Human All Exon v6). Next-generation sequencing was performed on the Illumina HiSeq platform with 150 base pair (bp) paired-end reads and 100X coverage for leukaemia samples and 50X for patient T cells and donor DNA.

Sequence data were aligned using the BWA-MEM algorithm (BWA v0.7.17) with the Homo sapiens genome hg38 as reference. For pre-processing of sequencing data, Picard was used for trimming and removal of PCR duplicates and Genome Analysis Toolkit (GATK v4.0.5.1) to re-calibrate the base quality scores.

Somatic variants were identified by analysing BAM files for variants using 3 variant callers: Mutect2 (GATK) [227], Strelka2 [210] and Vardict (2018.10.18) [211]. Sequences from AML and pre-transplant T cell DNA (germline control) were compared to define somatic variants present in tumour but not germline DNA, and were annotated using ANNOVAR (2018.04.16) [212]. Variants were filtered to only include variants that: (1) are nonsynonymous, (2) are exonic, (3) have VAF (variant allele frequency) $\geq 5\%$, and (4) are associated with $< 1\%$ VAF in germline (or $< 15\%$ if the variant is in a gene that is recurrently mutated in AML). Non-coding RNAs and untranslated regions (UTRs) were excluded. Variants were retained only if called by at least 2 of the 3 variant callers. After filtering, variants were manually inspected using the Integrative Genomics Viewer (IGV) tool (http://software.broadinstitute.org/software/igv/).

Mismatched germline variants between each patient and their respective stem cell transplant donor were identified using a separate analysis pipeline. Germline variants were called independently from Patient and Donor DNA samples using the variant caller, Platypus (v0.8.1.2) [213], and were annotated with ANNOVAR (2018.04.16). Default filters from Platypus were applied on a per-variant basis to

account for factors such as allele bias, strand bias and mapping quality. Variants were filtered to only include variants that: (1) are nonsynonymous, (2) are exonic, and (3) have ≥ 6 reads covering the nucleotide position in exome sequencing from both patient and donor. A threshold of 6 reads was chosen because at this read depth, the probability of failing to identify a heterozygous variant is <0.05. Non-coding RNAs and untranslated regions (UTRs) were excluded.

Germline variants were compared between patients and their respective donors to identify amino acid sequences present in the patient but not the donor (Figure 3.8)

2.4 Bulk RNA sequencing

For bulk RNA sequencing of leukaemia cells, 5000 cells were sorted directly into 75 μl RLT buffer (Qiagen) with 0.75 μl β-mercaptoethanol 14.3M (Sigma-Aldrich) and snap frozen with dry ice. cDNA libraries were prepared using SMARTer Ultra Low RNA Kit v4 (Clontech, Takara). Illumina adapters and paired-end indices were added by tagmentation using the Nextera XT kit (Illumina). The quality of cDNA libraries was assessed using a High Sensitivity DNA Bioanalyzer chip (Agilent Technologies) on a 2100 Bioanalyzer instrument (Agilent Technologies), both after the SMARTer protocol and after indexing. Libraries were quantified using High Sensitivity DNA Qubit (Thermo Fisher Scientific) prior to pooling. Next-generation sequencing was performed with 75 bp paired-end reads using the Illumina NextSeq 500 platform (high-output kit, 150 cycles), aiming for an average of 25 million reads/sample. The sequencing libraries were loaded onto the flow cell at a concentration of 1.8 pM.

Quality control of raw FASTQ files was performed using FASTQC. Data from each NextSeq lane were then merged and FASTQC repeated. FASTQ files were trimmed using TrimGalore (v0.3.1) to remove adapter sequences and low quality reads, using a Phred score threshold of 20. BAM files were generated, mapped to the *Homo sapiens* reference transcriptome (hg38) using the STAR alignment algorithm (v2.4.2a) and filtered for uniquely mapped reads. featureCounts was then used to count the number of reads that uniquely mapped to individual genes.

For the patient-donor mismatched variants that were identified from exome sequencing, expression was assessed within the mapped RNA-Seq reads, using in-house scripts. First, SAMtools mpileup was used to generate a pileup of read bases using alignment to the hg38 reference genome. Then, variant bases were counted at the positions of interest. Expression of the mismatched base >1 count per million mapped reads (CPM) were taken forward for prediction of HLA binding affinity.

2.5 HLA binding prediction

NetMHCpan 4.0 was used to predict HLA class I binding affinity of mismatched peptides to patient-specific HLA alleles. All possible 8-11 amino acid peptide windows across mismatched variants were assessed. Similarly, NetMHCIIpan 3.0 was used to predict HLA class II binding affinity. All possible 15 amino acid peptide windows across mismatched variants were assessed. For all predictions, only those present at > 1 CPM from RNA-Seq were retained. Peptides were excluded if the sequence was present elsewhere in the Homo sapiens reference proteome (Uniprot).

To identify peptides for functional screening, high affinity predicted HLA binders were taken forward (percentile rank affinity $< 0.5\%$ for HLA class I and $IC_{50} <$ 150nM for class II predictions). The difference in units for HLA class I and II binding is due to an updated version of the class I binding algorithm (NetMHCpan 4.0), which found that percentile ranks are superior to fixed affinity values across different HLA molecules [160]. Where multiple peptides crossing the same mismatched variant were predicted to bind patient specific HLA allele(s) (different length peptide or different sliding window across the position), peptides were combined to produce a long amino acid sequence. For overlapping peptide sequences ≤ 20 amino acids in length, the entire sequence was used for IFNγ ELISpot screening. If > 20 amino acids, the long peptide was split into 2 peptides, each 20 amino acids in length, and these were used for screening.

2.6 Immunopeptidomic analysis

The methods used for immunoprecipitation and mass spectrometry analysis are based on Partridge *et al.* [214]. All steps up to the data analysis were performed by Dr Annalisa Nicastri and Dr Nicola Ternette.

2.6.1 Antibody purification and resin preparation

Antibodies were purified by harvesting supernatant from hybridomas and clearing at 2,500 × g for 25 minutes at 4°C. Antibodies were then then filtered through a 0.2 μm SteriCup Filter (Millipore) and adjusted to pH 8.0 with 1 M Tris. Antibody was captured by addition of 2–3 ml protein A resin (Protein A Sepharose, PAS, Expedeon) suspended in PBS for 30 minutes at room temperature on an orbital shaker at 100 rpm. Beads were collected by gravity flow through a chromatography column and then washed with 20–30 ml PBS. Captured antibody was eluted by addition of 5 ml of 100 mM glycine (pH 3.0) to the beads and then 1 ml fractions were collected into tubes containing 50 μl of 1 M Tris (pH 9.5). Buffer was exchanged for PBS by concentration through a 5 kDa molecular weight cutoff centrifugal filter. Antibody purity was assessed by observing the presence of light and heavy chains via sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE).

For antibody cross-linking with beads, 500 μg purified antibody per 250 μl PAS was incubated at room temperature for 30 minutes. Antibodies used were: W6/32 (anti-class I), L243 (anti-HLA-DR) and IVA12 (anti-class II). PAS beads were first collected by gravity flow through a chromatography column, washed and equilibrated with 2.5 ml borate buffer (0.05 M boric acid, 0.05 M KCl, 4 mM NaOH). Antibody was then cross-linked with PAS beads with 2.5 ml of 40 mM dimethyl pimelimidate in borate buffer (pH 8.6) for 30 minutes at room temperature. The cross-linking reaction was terminated by addition of 2.5 ml ice cold 0.2 M Tris (pH 8.0). Unlinked antibody was removed by flowing through 2.5 ml citrate buffer (pH 3.0), followed by 2–3 washes with 2.5 ml PBS to neutralise pH.

2.6.2 HLA immunoprecipitation and elution of peptides

Cells were washed with PBS, pelleted and then 0.5 ml IGEPAL buffer (0.5% IGEPAL 630, 50 mM Tris pH 8.0, 150 mM NaCl and 1 tablet of cOmplete Protease Inhibitor Cocktail EDTA-free (Roche) per 10 ml buffer) was added to each sample. Cells were lysed by mixing for 45 minutes at 4°C. Lysates were cleared by sequential centrifugation steps at 2,000 x g for 10 minutes, then 20,000 x g for 15 minutes at 4°C. Peptide–HLA complexes were captured by incubation of lysate with 500 μg of the relevant antibody coupled to PAS overnight at 4°C on a rotator. The lysate was flowed through a pre-washed chromatography column and the beads were then consecutively washed with 2.5 ml wash buffer 1 (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA), 2.5 ml wash buffer 2 (50 mM Tris pH 8.0, 150 mM NaCl), 2.5 ml wash buffer 3 (50 mM Tris pH 8.0, 450 mM NaCl), and 2.5 ml wash buffer 4 (50 mM Tris pH 8.0). Peptide–HLA complexes were eluted by addition of 1.25 ml of 10% acetic acid and then dried in a vacuum drier.

To purify HLA-bound peptides, the eluted peptides and subunits of the HLA complex were loaded on Sep-Pak tC18 (Waters, Milford, MA) cartridges that were pre-washed with 80% acetonitrile (ACN) in 0.1% trifluoroacetic acid (TFA) and the with 0.1% TFA alone. After loading, the cartridges were washed with 0.1% TFA. The peptides were separated from the much more hydrophobic HLA heavy chains on C18 cartridges by elution with 30% ACN in 0.1% TFA. They were further purified using Silica C18 column tips (Harvard Apparatus, Holliston, MA) and eluted again with 30% ACN in 0.1% TFA. The peptides were concentrated and the volume reduced to 15 μl using vacuum centrifugation. For liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, 5 μl was used, which is highly enriched for peptides eluted from HLA [215].

2.6.3 Liquid chromatography-tandem mass spectrometry

Purified peptides were resuspended in 20 μl of 0.1% TFA / 1% ACN and then 9 μl was injected into an Ultimate 3000 high performance liquid chromatography (HPLC) system coupled online to a Fusion Lumos mass spectrometer (ThermoFisher

Scientific). Peptides were separated with a 75 μm × 50 cm PepMap RSLC C18 EasySpray column using a linear gradient of buffer B (0.1% TFA in ACN) in buffer A (0.1% TFA in water) at a flow rate of 250 nl/min for 30 minutes. For peptides eluted from HLA class I, a gradient of 3 to 25% of buffer B in buffer A was used, whereas 3 to 30% was used for HLA class II peptides.

Peptides were introduced to the Fusion Lumos using an EasySpray source. Precursors were selected in top-speed mode within a 2 second cycle time (accumulation time of 120 ms) and an isolation width of 1.2 amu (atomic mass units) for fragmentation. For HLA class I peptides, higher energy collision dissociation (HCD) with a collision energy setting of 28 was performed on the peptides with a charge state of 2–5, while a higher collision energy of 32 was applied to singly charged precursor ions that were selected with lower priority. For HLA class II peptides, HCD with a collision energy setting of 25 was performed on the peptides with a charge state of 2-5. MS resolution was set at 120,000 and MS^2 resolution was set at 30,000. All fragmented precursor ions were actively excluded from repeated selection for 30 seconds.

2.6.4 Data analysis

Data were imported into PEAKS 8 software (Bioinformatic Solutions) as .raw files. A database (20,243 entries) containing all annotated human Swiss-Prot entries (current at 10/08/2017) including RNA-Seq expression data for each sample was used for interpretation of MS/MS spectra. Precursor and fragment error tolerances were set at 5 ppm and 0.03 Da, respectively. A false discovery rate (FDR) of 1% (unless otherwise indicated) was defined using parallel decoy database searches. No fixed or variable modifications were set. Peptide spectrum matches were exported as .csv files and peptides < 8 amino acids in length were excluded from all analyses. Where two possible peptides differing only by a single leucine to isoleucine residue were identified, both matches were retained in the dataset. Sequence logos of identified peptides were produced using the default settings in WebLogo 3, which gives a graphical representation of conserved amino acid residues in aligned peptide sequences [216].

2.7 *Ex vivo* and cultured IFNγ ELISpot peptide screening

Peptides were synthesised with crude purity (> 70% purity) (GenScript Netherlands), reconstituted in DMSO (40 mg/ml) and stored at -80°C until use. Peptides were combined in pools with up to 35 peptides per pool for IFNγ ELISpot screening.

Post-transplant peripheral blood mononuclear cell (PBMC) samples were thawed, washed with media and rested overnight. For *ex vivo* IFNγ ELISpots, cells were directly used for the assay. For cultured ELISpots, after overnight rest, cells were counted and seeded at a density of 1.5-2 million cells/ml with peptide (each at 2 μg/ml) and interleukin-7 (IL-7) 25 ng/ml (R and D Systems). Cells were cultured in RAB-10 media (RPMI-1640 with 10% heat-inactivated human AB serum (Sigma-Aldrich), 1% penicilllin/streptomycin, 1% sodium pyruvate (Sigma-Aldrich), 1% HEPES buffer (Sigma-Aldrich), 1% MEM non-essential amino acids (Gibco), 1% GlutaMAX (2mM final concentration; Gibco), 0.1% β-mercaptoethanol (50 μM final concentration; Gibco) and 25U/ml Benzonase (Millipore)). The maximum number of peptides in each culture was 100, which maintained a DMSO concentration below 0.5%. Interleukin-2 (IL-2; Proleukin) was added after commencing culture with peptide on days 3, 7 and 10 . Further RAB-10 media was added as required, to maintain the culture density. After 12 days culture with peptide, cells were washed 4 times with PBS and rested in RAB-10 overnight in the absence of cytokines or peptides.

Prior to setting up ELISpot experiments, 96-well plates with polyvinylidene difluoride (PVDF) membranes were activated with 70% ethanol, washed 3 times with PBS and then coated with 10 μg/ml anti-IFNγ antibody (1-D1K; 1:100 dilution; Mabtech) in PBS. Plates were incubated at 4°C for 1-7 days. On the day of setting up ELISpots, plates were washed 3 times with PBS and then blocked with RAB-10 media for 3 hours.

For *ex vivo* IFNγ ELISpot testing, 200,000 cells/well were used and for cultured IFNγ ELISpots, 50,000 cells/well. Peptide pools were tested with each peptide

at a concentration of 2 μg/ml. When deconvoluting positive pools, individual component peptides were also tested at 2 μg/ml. Positive controls were treatment with phytohaemagglutinin (PHA; 10 μg/ml) and a CMV/EBV/Flu peptide pool (Mabtech CEF Pool). For a negative control, DMSO treatment was used at the same concentration as the peptides being tested. Conditions were each tested with 2 or 3 technical replicates.

After 16-20 hours incubation, ELISpot plates were developed. Media was removed and wells were washed with PBS/0.05% Tween 6 times. Biotinylated anti-IFNγ antibody (7-B6-1; Mabtech) was added to the wells at a concentration of 1 μg/ml (1:1000 dilution) in PBS/0.5% bovine serum albumin (BSA). The solution was sterile filtered with a 22 μm filter prior to addition to the wells and incubated for 4 hours at room temperature. Then, plates were washed 6 times with PBS/0.05% Tween. Avidin-peroxidase conjugate (Vector Laboratories) was added to the wells and incubated for 1 hour. Wells were washed 3 times with PBS/0.05% Tween and then 3 times with PBS, before addition of substrate.

AEC substrate solution was prepared before addition to plate wells: 43.7ml deionised water; 3.5ml 0.1M acetic acid; 280 μl 3M sodium acetate; 3-amino-9-ethylcarbazole tablet (AEC; Sigma-Aldrich) dissolved in 2.5ml dimethylformamide; 25 μl of 30% hydrogen peroxide. The solution was filtered with a 0.45 μm filter and added to wells for 3-4 minutes, before washing off thoroughly with water. Plates were left to air dry before analysis using an AID machine and AID ELISpot software (Autoimmun Diagnostika, Strasburg, Germany). Positive responses were defined as either greater than 50 spot forming units (SFU) per million cells, or 3 x DMSO negative control, whichever value was greater.

2.8 Intracellular cytokine staining after peptide stimulation

Post-transplant PBMCs were thawed, cultured with peptide and rested, as indicated in Section 2.7. To block Golgi transport and stain with CD107a, cells were then incubated with CD107a (BV650) antibody, together with monensin

(Golgistop; 2 μM; BD Biosciences) and brefeldin A (GolgiPlug; 10 μg/ml; BD Biosciences). At the same time, cells were stimulated with 10 μg/ml peptide and placed in a 37°C incubator (5% CO_2) for 6 hours. Following this period of stimulation, cells were washed with PBS and resuspended in PBS with Fc block (TruStain, Biolegend) for 10 minutes at room temperature. Then, Fixable Live/Dead Aqua stain (Thermo Fisher Scientific) was added with incubation for 15 minutes at room temperature. Cells were washed with PBS and stained with surface antibodies, CD25 (APC) and CD137 (PE), in FACS buffer. After a 15 minute incubation at room temperature, cells were washed with FACS buffer. Fixation and permeabilisation (Fix/Perm; BD Biosciences) was performed at 4°C for 20 minutes, followed by 2 washes with Perm/Wash solution, as per the manufacturer's instructions. Cells were then intracellularly stained in Perm/Wash buffer with CD3 (BV711), CD4 (APC Cy7), CD8 (Alexa Fluor700), IFNγ (PE Cy7), TNFα (FITC) and IL-2 (PerCP eFluor710) at 4°C in the dark. Cells were washed twice in Perm/Wash buffer and resuspended in PBS for flow cytometry. Analysis was performed using a BD Fortessa X20 instrument.

2.9 HLA allele restriction and epitope mapping

To assess allele restriction of peptide responses, post-transplant PBMCs were first thawed, cultured with peptide for 12 days and rested, as outlined in Section 2.7. Cells were then incubated with purified anti-HLA antibodies (10 μg/ml) for 2 hours at 37°C. The antibodies used were: W6/32 (pan-class I), Tü39 (pan-class II), BB7.2 (HLA-A2), L243 (HLA-DR), Tü169 (HLA-DQ) and B7/21 (HLA-DP).

IFNγ ELISpot plates were prepared and 50,000 cells added per well, without washing away anti-HLA antibodies. Peptide solution was added at 2 µg/ml final concentration. Plates were incubated and developed, as outlined in Section 2.7.

To map the optimal peptide epitope, peptides of different length (9-18 amino acids) crossing the mismatched amino acid in different positions were tested by IFNγ ELISpot.

2.10 Isolation of antigen-specific T cells and sequencing their T cell receptors (TCRs)

Post-transplant PBMCs were thawed, cultured with 10 μg/ml peptide for 12 days and rested, as indicated in Section 2.7. After resting for 24 hours, cells were stimulated with 10 μg/ml peptide and incubated for 6 hours at 37°C. Cells were then washed with cold buffer (PBS, 0.5% BSA, 2mM EDTA at 4°C) and IFNγ catch used to label IFNγ-secreting cells, as per the manufacturer's instructions (IFNγ Secretion Assay; Miltenyi Biotec), in order to identify cells being activated by peptide. In brief, IFNγ catch reagent was added and cells were incubated on ice for 5 minutes. Warm RAB-10 medium was then added to the cells, which were incubated at 37°C for 45 minutes. Cell tubes were rotated every 5 minutes to resuspend settled cells during the incubation. Cold buffer was then added to the cells and samples were centrifuged at 300 g for 10 minutes and supernatant aspirated. Cells were resuspended in PBS with Fc block (TruStain, Biolegend) at 1:20 dilution. IFNγ-PE detection antibody was added to the cell suspension, together with flow cytometry anti-human antibodies CD3 (PE Cy7), CD4 (APC) and CD8 (FITC). Cells were incubated in the dark at 4°C for 30 minutes, washed with cold flow cytometry staining buffer and then resuspended in flow cytometry staining buffer containing Hoechst for live/dead separation. Cell populations were sorted using an 85 μm nozzle with a BD Fusion instrument.

For each population, up to 10,000 cells were loaded into a 10X Chromium chip. Sample processing for single cell gene expression (scRNA-Seq) and T cell receptor V(D)J clonotypes (scTCR-Seq) was performed using the Chromium Single Cell 5' Library and Gel Bead Kit (10X Genomics), following the manufacturer's user guide. After the Gel Bead-in-Emulsion reverse transcription (GEM-RT) reaction and clean-up, 14 cycles of PCR amplification of cDNA were performed to obtain sufficient material for both RNA-Seq library construction and TCR V(D)J targeted enrichment followed by V(D)J library construction. TCR V(D)J enrichment was

performed using the Human T Cell Chromium Single Cell V(D)J Enrichment Kit (10X Genomics). Libraries for RNA-Seq and V(D)J were prepared, profiled using the Bioanalyzer High Sensitivity DNA kit (Agilent Technologies) and quantified with High Sensitivity DNA Qubit (Thermo Fisher Scientific). Libraries were sequenced using NovaSeq 6000 (Illumina) by a commercial vendor (Novogene, Cambridge, UK), aiming for 100,000 reads/cell for the transcriptome libraries and 5,000 reads/cell for the V(D)J-enriched TCR libraries.

2.11 Processing and analysis of single cell TCR-Seq and RNA-Seq data

scTCR-Seq data were processed with Cell Ranger software for each sample using the command 'cellranger vdj' with the *Homo sapiens* reference genome, GRCh38. This demultiplexes and aligns sequencing reads, corrects and counts the unique molecular identifiers (UMIs) for each gene in each cell and produces an output file with TCRα and TCRβ chain CDR3 nucleotide sequences, organised by clonotype. Each clonotype comprises all single cells with the same TCRα and TCRβ chain.

Expression data were processed using the command 'cellranger count' on the human reference set of 30,727 genes. Cell Ranger performed default filtering for quality control, and produced for each sample a barcodes.tsv, genes.tsv, and matrix.mtx file containing counts of transcripts for each gene in each single cell, which were parsed using a script in R to create a count matrix for each sample. Using the Seurat package, low quality cells and multiple cells within droplets were excluded by filtering to include only cells with between 500 and 7000 genes and also to exclude cells with $> 10\%$ transcripts originating from mitochondrial genes. After quality filtering, RNA expression data was normalised by total expression, multiplied by a scaling factor of 10,000 and the results were log-transformed. Seurat was then used to analyse the data, visualise with UMAP (Uniform Manifold Approximation and Projection) and perform differential gene expression analysis.

Annotation of cell cluster identity was performed using published literature. This was supported by independent analyses that map the transcriptome of each

single cell to the best correlated immune reference signature, which were obtained from bulk RNA-Seq of sorted human immune cells. Analyses were performed using: (1) SingleR, with reference data from Monaco *et al.* [217, 218], and (2) Cluster Identity Predictor (CIPR) with reference data from the Database of Immune Cell Expression (DICE) [219, 220].

For gene set enrichment analysis (GSEA), reference gene sets were downloaded in GMT format from MSigDB (http://software.broadinstitute.org/gsea/downloads.jsp). Analysis was performed using the Fgsea package with default parameters [221] on lists of genes pre-ranked according to differential expression between groups of cells, as stated in the text. Pathways with an adjusted p-value, FDR (false discovery rate) <0.05 are highlighted as statistically significant after accounting for multiple hypothesis testing.

2.12 Bulk TCR sequencing

Ex vivo post-transplant PBMCs were thawed, washed and stained with flow cytometry anti-human antibodies CD3 (PE Cy7), CD4 (APC) and CD8 (FITC), as indicated above. Hoechst was used as a live/dead stain. Live $CD4^+$ and $CD8^+$ cells were sorted directly into RLT buffer (Qiagen) using a BD Fusion instrument (BD Biosciences). A total of $2.8x10^5$ $CD4^+$ and $1.1x10^5$ $CD8^+$ cells were sorted. UMI-labelled 5' RACE TCRα and TCRβ complementary DNA (cDNA) libraries were prepared using a Human TCR Profiling Kit (MiLaboratory LLC). All extracted RNA was used for cDNA synthesis, and all synthesised cDNA was used for PCR amplification. Libraries were prepared in parallel using the same number of PCR cycles and sequenced using 150 bp paired-end reads on a NovaSeq 6000 machine (Illumina) by a commercial vendor (Novogene, Cambridge, UK).

Data quality control was assessed by FASTQC and MULTIQC. UMIs were demultiplexed and UMI read groups were assembled by consensus sequence using MIGEC. Read groups were retained if they were supported by at least 5 reads with the same UMI. V, D and J segments were mapped, CDR3 sequences extracted and clonotypes assembled. This approach yielded the following number of assembled

UMI read groups (each representing a single UMI-labelled TCR cDNA molecule): $CD4^+$ TCRα – 16,963; $CD4^+$ TCRβ – 28,928; $CD8^+$ TCRα – 8,741; $CD8^+$ TCRβ – 11,379. TCR repertoire diversity analysis was performed using VDJtools software and annotation with known antigen specificity of TCRs was completed using VDJdb [284].

2.13 PADI4 genotyping

PADI4 genotyping was performed on primary patient AML samples, AML cell lines and normal samples by either Sanger sequencing or real-time quantitative PCR (RT qPCR). For each technique, genomic DNA was isolated using a QIAmp DNA Micro kit (Qiagen).

Sanger sequencing of purified PCR product was performed. First, a 369 base pair amplicon was generated using custom primers (Forward: CTCCAAGTGCAGCG-TAGTCT / Reverse: TGGAACACCACAGCCTCG), Q5 PCR mastermix (New England Biolabs, NEB) and 30 PCR cycles. Then, a sample of the PCR product was run on a 2% agarose gel to confirm amplification of a single DNA fragment and the remaining PCR product was cleaned using a PCR Purification kit (Qiagen), before Sanger sequencing.

Genotyping by RT qPCR used a PADI4 TaqMan SNP Genotyping Assay (Applied Biosystems, Thermo Fisher Scientific) and was performed according to the manufacturer's instructions.

2.14 Assessing PADI4 expression

2.14.1 RT qPCR for *PADI4* gene expression

RNA was extracted from AML samples and PBMCs using Qiashredder (Qiagen) and an RNeasy Micro kit (Qiagen). RNA was then quantified by NanoDrop and the quality assessed using a Bioanalyzer RNA 6000 pico assay (Agilent Technologies) on a 2100 Bioanalyzer instrument (Agilent Technologies).

RT qPCR was performed using a PADI4 TaqMan gene expression assay (Hs01057483-m1, Applied Biosystems, Thermo Fisher Scientific) with TaqMan Gene expression

Master mix (Applied Biosystems, Thermo Fisher Scientific) on a 7500 HT Fast Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific), following the manufacturer's instructions. Housekeeping control Taqman assays were GAPDH (Hs03929097-g1) and β-actin (Hs99999903-m1). PADI4 expression was normalised to GAPDH.

2.14.2 Western blotting for PADI4

Protein was extracted from AML cells using radioimmunoprecipitation assay (RIPA) buffer (Thermo Fisher Scientific) containing Halt protease inhibitor (Thermo Fisher Scientific) and quantified using bicinchoninic acid (BCA) assay. For each sample, equal quantities of protein extract was diluted in Laemmli sodium dodecyl sulphate (LDS) with 1 M 1,4-dithiothreitol (DTT) and added to lanes of precast 4-12% NuPAGE Bis-Tris gels (Invitrogen) in NuPAGE MOPS running buffer (Invitrogen). Following polyacrylamide gel electrophoresis, protein was transferred to a polyvinylidene difluoride (PVDF) membrane.

The membranes were blocked with 5% milk in Tris-buffered saline containing 0.1% Tween-20 (TBS-T) for 1 hour at 4°C. They were then incubated overnight at 4°C in TBS-T with primary anti-human antibody: mouse anti-PADI4 (1:1,000; Abcam ab128086) or mouse anti-GAPDH (1:2,000; Santa Cruz Biotechnology sc-47724). The membranes were then washed 4-6 times with TBS-T. Secondary goat anti-mouse antibody (1:10,000; Affinity Biosciences) conjugated to horseradish peroxidase was incubated in TBS-T for 1 hour at room temperature, followed by 4-6 washes with TBS-T. ECL Prime detection kit (Merck) was used for development and blots were imaged using a Bio-Rad ChemiDoc instrument.

2.15 Virus production and transduction

2.15.1 TCR cloning

TCRα and TCRβ chain nucleotide sequences from the 2 largest PADI4 antigen-specific T cell clones identified from scRNA-Seq were synthesised by GeneArt (Life Technologies), with human codon optimisation. Paired TCRα and TCRβ variable

segments were inserted into a retroviral (RV) vector, pMP71. Sub-cloning was performed using restriction endonuclease digestion and DNA ligation (T4 ligase, New England Biolabs, NEB). NotI / SacII were used for TCRα, and NcoI / PmlI for TCRβ (all restriction enzymes from New England Biolabs).

TCR variable and murine constant regions were transferred from the pMP71 RV vector into a lentiviral (LV) vector, pHRsin, using NEBuilder HiFi DNA Assembly (New England Biolabs). Sequences of all vectors were confirmed by Sanger sequencing.

2.15.2 Retrovirus production

pMP71-TCR vectors (α variable–murine α constant–P2A-β variable-murineβ constant-T2A-truncated murine CD19) were co-transfected with pCL-Ampho retrovirus packaging vector using FuGENE®(Promega) into the Phoenix Ampho packaging cell line at up to 80% confluence. After a 48-hour incubation at 37°C, viral supernatant was harvested and centrifuged for 5 minutes to remove cellular debris.

2.15.3 Retroviral transduction of T cells

Prior to RV transduction, fresh PBMCs from healthy donors were activated with Dynabeads T-activator CD3/CD28 (25 μl beads/ml) (Thermo Fisher Scientific) for 48 hours at a density of 1.5-2 million cells/ml in media (RPMI-1640 with 10% heat-inactivated FBS, 1% penicilllin/streptomycin, 1% L-glutamine and 30 U/ml IL-2 (Roche)).

To prepare plates for transduction, RetroNectin (Takara) was diluted with PBS to a concentration of 30 μg/ml and 750 μl was added to each well of a non-treated 24-well tissue culture plate and incubated at 4°C overnight. Immediately before transduction, RetroNectin was removed, wells were blocked with PBS/2% BSA for 30 minutes at room temperature and washed twice with PBS. To each RetroNectin-coated well, 1 million activated T cells were added in 50 μl of media. To this, 500 μl of RV supernatant was added. Plates were centrifuged at 2000 rpm for 90 minutes at 32°C. Viral supernatant was then removed and replaced with 2 ml of fresh media (10 U/ml IL-2) per well. Transduced cells were then incubated at 37°C.

Transduction efficiency was assessed by flow cytometry at 72 hours, using anti-murine CD19 (APC Cy7) and anti-murine TCRβ (BV711). Anti-human flow cytometry antibodies used in the assay were: CD3 (PE Cy7), CD4 (APC) and CD8 (FITC). Live/dead staining was with Hoechst.

2.15.4 Lentivirus production

The pHRsin-TCR vectors (SFFV–β variable–murine β constant–P2A-α variable-murine α constant) were co-transfected with pMD-G (VSV-G) envelope and pQ8.91 packaging vectors using TurboFectin (Origene) into the HEK-293T packaging cell line at 90% confluence. After a 48-hour incubation at 37°C, viral supernatant was harvested and centrifuged to remove cellular debris.

2.15.5 Lentiviral transduction of T cells

Prior to LV transduction, fresh PBMCs from healthy donors were activated with Dynabeads T-activator CD3/CD28 (25 μl beads/ml) (Thermo Fisher Scientific) for 48 hours at a density of 1.5-2 million cells/ml in T cell media: RPMI-1640 with 5% heat-inactivated human AB serum (Sigma-Aldrich), 1% penicilllin/streptomycin (Sigma-Aldrich), 1% sodium pyruvate (Sigma-Aldrich), 1% HEPES buffer (Sigma-Aldrich), 1% MEM non-essential amino acids (Gibco), 1% GlutaMAX (2mM final concentration; Gibco), 0.1% β-mercaptoethanol (50 μM final concentration; Gibco), 1000 IU/ml IL-2 (Proleukin, Clinigen) and 10 ng/ml IL-15 (PeproTech).

To prepare plates for transduction, RetroNectin was diluted with PBS to a concentration of 30 μg/ml, added to wells of non-treated tissue culture plates and incubated at 4°C overnight. Immediately before transduction, RetroNectin was removed, wells were blocked with PBS/2% BSA for 30 minutes at room temperature and washed once with PBS. LV supernatant was added to the RetroNectin-coated wells and the plate centrifuged for 90 minutes at 2000 g at 4°C. Supernatant was removed from wells and pre-stimulated T cells were added in T cell media to each well at $5x10^5$ cells/ml. The plate was then briefly centrifuged at 2000 rpm for 1 minute. Transduced cells were incubated at 37°C.

Transduction efficiency was assessed by flow cytometry at 72 hours, using anti-murine TCRβ (BV711). Anti-human flow cytometry antibodies used in the assay were: CD3 (PE Cy7), CD4 (APC) and CD8 (FITC). Live/dead staining was with Hoechst. Also at 72 hours, transduced cell cultures were separated into $CD4^+$ and $CD8^+$ cell fractions using microbeads and MS columns (Miltenyi), according to the manufacturer's instructions. $CD4^+$ and $CD8^+$ cells were separately cultured for up to 2 weeks before functional assays.

2.16 *In vitro* T cell activation and killing assays

To prepare transduced T cells for activation and killing assays, $CD4^+$ and $CD8^+$ T cells were washed 3 times with RPMI-1640 and rested overnight in T cell media at 37°C. $CD4^+$ and $CD8^+$ T cells were then combined in a 1:1 ratio. Target cells were thawed and washed. For peptide-pulsed assays, target cells were incubated with 10 μM peptide at 37°C at cell density 2.5x10^6 cells/ml for 2 hours. Cells were then washed twice with RPMI-1640.

2.16.1 T cell activation assays

For T cell activation assays, co-culture of T cells with AML target cells was performed at a 1:2 cell ratio in T cell media containing 5% human serum and cytokines: SCF (25 ng/ml), FLT3L (10 ng/ml), TPO (10 ng/ml), IL-3 (10 ng/ml), G-CSF (20 ng/ml), GM-CSF (5 ng/ml). After the 8-hour incubation, cells were prepared and analysed by flow cytometry as outlined in Section 2.8. Surface antibody staining was with CD137 (PE), CD134 (BB700) and murine TCRβ (APC) antibodies. Intracellular antibody staining was with CD3 (BV711), CD4 (APC Cy7), CD8 (Alexa Fluor700), IFNγ (PE Cy7) and TNFα (FITC) antibodies.

For T cell activation assays using extracellular PADI4 antigen, AML supernatant was produced by culturing a primary AML sample at high density (4x10^6 cells/ml) for 72 hours, then centrifuging at 150 g for 3 minutes and retaining the supernatant. To produce AML cell lysate, an AML sample was snap frozen with dry ice, then

thawed at 37°C. It was then mixed thoroughly by vortexing and centrifuged at 150 *g* for 3 minutes to remove cell debris.

2.16.2 Generating monocyte-derived dendritic cells

Fresh healthy donor PBMCs with the HLA-DPA1*01:03-DPB1*04:01 haplotype were obtained by Ficoll gradient separation, as outlined in Section 2.1. PBMCs were plated in flasks at a density of $1x10^6/cm^2$ in pre-warmed RPMI-1640 with 5% heat-inactivated human AB serum, 1% penicilllin/streptomycin and 1% HEPES for 2 hours at 37°C. Non-adherent cells were then removed by washing 5 times with warm media. Media supplemented with 300 U/ml IL-4 (Peprotech) and 100 U/ml GM-CSF (Peprotech) was added to cover the adherent cells.

After 3 days, media was removed and replaced with fresh pre-warmed media, supplemented with 300 U/ml IL-4 and 100 U/ml GM-CSF. The media that had been removed was centrifuged. Non-adherent cells in the pellet were resuspended in media with IL-4 and GM-CSF and added back to the flask. After a further 3 days, monocyte-derived dendritic cells (MDDCs) were harvested using a cell scraper.

2.16.3 Killing assays

For T cell-mediated killing assays, target cells were thawed, washed and labelled with Cell Trace Violet (Thermo Fisher Scientific), according to the manufacturer's protocol. For peptide-pulsed assays, target cells were incubated with 10 μM peptide at 37°C at cell density $2.5x10^6$ cells/ml for 2 hours. Cells were then washed twice with PBS. Co-culture of T cells with target cells was performed at different ratios, up to 10:1, for 24 hours. Cells were then stained for flow cytometry using Live/Dead Fixable Far Red (Thermo Fisher Scientific) and antibodies: CD3 (PE Cy7), CD4 (APC), CD8 (FITC) and murine TCRβ (BV711). Cells were quantified using Flow Count beads (Molecular Probes) and % target cell killing calculated by the following formula:

% Specific lysis = [(Target cell lysis with T cell clone – Spontaneous lysis)/(Maximum lysis – Spontaneous lysis)]*100

For all assays, analysis was performed using a BD Fortessa X20 instrument.

2.17 *In vivo* killing of AML

Biobanked PBMCs from blood or bone marrow taken from AML patients at diagnosis were thawed and AML cells were stained with purified mouse anti-human antibodies against CD3 (PE Cy5), CD45 (APC eFluor780) and CD235ab (Pacific Blue). AML cells were purified by FACS to exclude dead cells and T cells, as outlined in Section 2.2. T cells were removed to reduce the risk of GvHD in the mice. Between 6.5x10^5 and 2x10^6 of each sample were injected intra-tibially into each immunodeficient NSGW (NOD/SCID/$Il2r\gamma^{-/-}/Kit^{W41}$) mouse aged 10-20 weeks to generate patient-derived xenografts.

Engraftment was assessed in peripheral blood every 4 weeks until human AML cells were detectable. Freshly expanded T cells (3x10^6 per mouse) were then injected into the tail vein and AML burden was monitored by flow cytometry, using antibodies against murine CD45 (BV605), murine TCRβ and human antibodies against CD45 (APC eFluor780), CD33 (PE), CD34 (APC), CD19 (FITC), CD3 (PE Cy7). Hoechst was used as a live/dead stain. At the termination of experiments blood, bone marrow and spleen were harvested and AML engraftment assessed.

Chapter 3

Prediction of graft-versus-leukaemia antigens

Contents

3.1 Multi-omic approach to antigen prediction 65
3.1.1 Experimental overview 65
3.1.2 Patient cohort 68
3.1.3 Cell sorting for whole exome sequencing and RNA-Seq . 69
3.2 Identifying mismatched exonic variants 72
3.2.1 Germline and somatic variant calling 72
3.2.2 Germline mismatched variants 79
3.2.3 Somatic variants 79
3.2.4 Population frequency of mismatched variants 83
3.3 RNA expression of variants 84
3.3.1 Gene expression in AML 84
3.3.2 Gene expression in normal tissues 85
3.4 HLA binding prediction 86
3.5 Immunopeptidomic analysis of AML 89
3.5.1 Approach to identify HLA-bound peptides using tandem mass spectrometry 89
3.5.2 Detection of HLA-bound peptides 90
3.5.3 Comparison of peptide and RNA expression 93
3.5.4 Comparing HLA-binding prediction to mass spectrometry detection 95
3.5.5 Candidate alloreactive peptides identified by mass spectrometry 96
3.6 Conclusions: Antigen prediction 98

3.1 Multi-omic approach to antigen prediction

3.1.1 Experimental overview

An unbiased, exome-guided approach was taken to identify putative antigens responsible for the GvL response in 15 patients. To identify candidate alloreactive peptides in transplant recipients, samples were analysed using 3 techniques (Figure 3.1):

1. Whole exome sequencing (WES) - If a variant is present in the patient but not the donor, it could lead to an amino acid sequence that is presented by HLA

and recognised as foreign by donor-derived T cells, leading to an immune response. WES was used to identify germline DNA variants in each patient's pre-transplant T cells (patient germline) and in their allo-SCT donor. For each patient-donor pair, variants were compared to identify mismatched germline variants. In addition, leukaemia-specific variants were identified by performing somatic variant calling on AML, using pre-transplant T cells as a germline control. Variants present in the AML sample, but not germline, are likely to have been acquired during leukaemia development and therefore will not be present in normal healthy cells.

2. Bulk RNA sequencing - In order to be presented by HLA on the surface of AML cells, the gene containing a mismatched variant must be expressed. Therefore, RNA-Seq was performed on AML blasts.

3. Mass spectrometry of HLA-bound peptides (immunopeptidomics) - To detect HLA-presented peptides directly, AML cells were lysed and HLA class I and II molecules immunoprecipitated. HLA-bound peptides were then eluted and analysed by mass spectrometry. The amino acid sequence of peptides detected by mass spectrometry was determined using a reference proteome. Patient-donor mismatched peptides were distinguished with reference to the mismatched variants identified by exome sequencing.

Data from each of these 3 techniques was integrated to predict alloreactive HLA-binding peptides *in silico*.

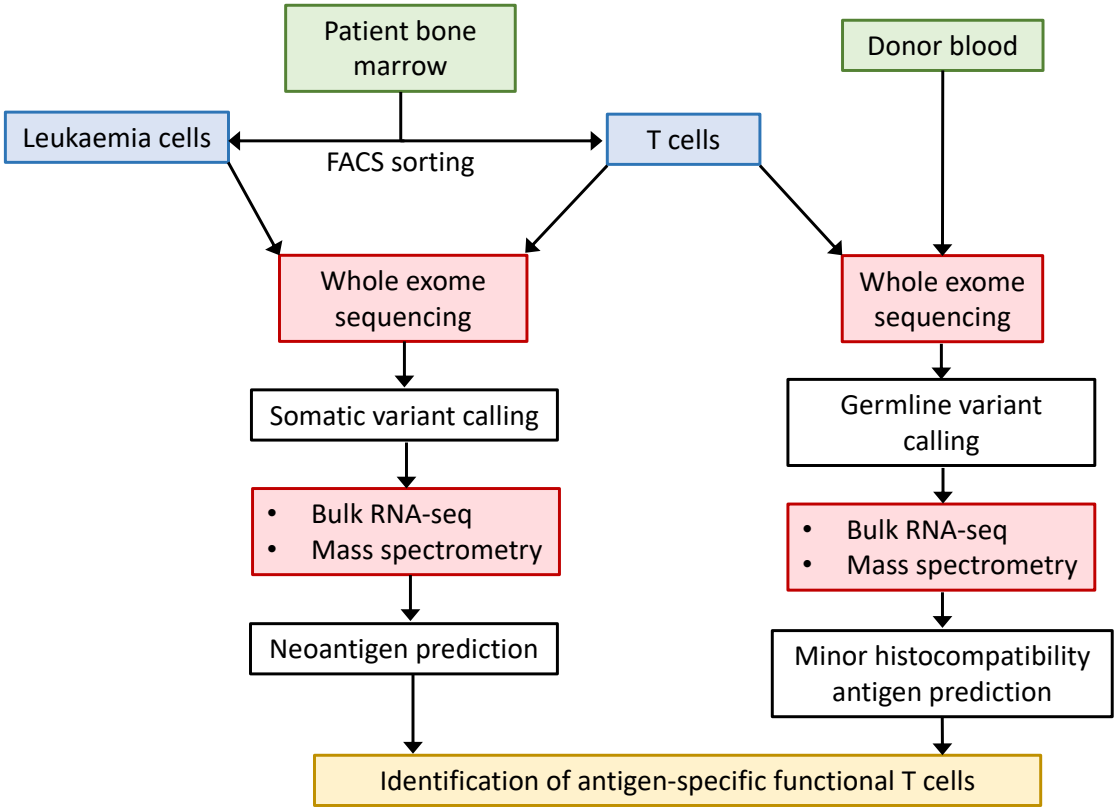


Figure 3.1: Overview of exome-guided antigen prediction. Whole exome sequencing was used to identify: (1) leukaemia-specific (somatic) variants, and (2) germline differences between patients and their respective allo-SCT donors. By also accounting for expression in AML (RNA-Seq) and presentation by HLA (immunopeptidomics), neoantigens and minor histocompatibility antigens were predicted. These were then functionally validated (Chapters 4 and 5).

This "reverse immunology" approach of starting from the sequence of the gene product and then predicting and functionally validating immunogenic peptides has been successfully employed to identify neoantigens [165, 168, 169]. A large number of exome variants are identified. Expression analysis (RNA-Seq) and *in silico* HLA-binding prediction reduces the number of candidate neoantigens. Some studies also incorporate mass spectrometry of HLA-bound peptides to further filter potential antigens [168] (Figure 3.2). To confirm that the antigens are capable of eliciting immune responses, functional testing is required.

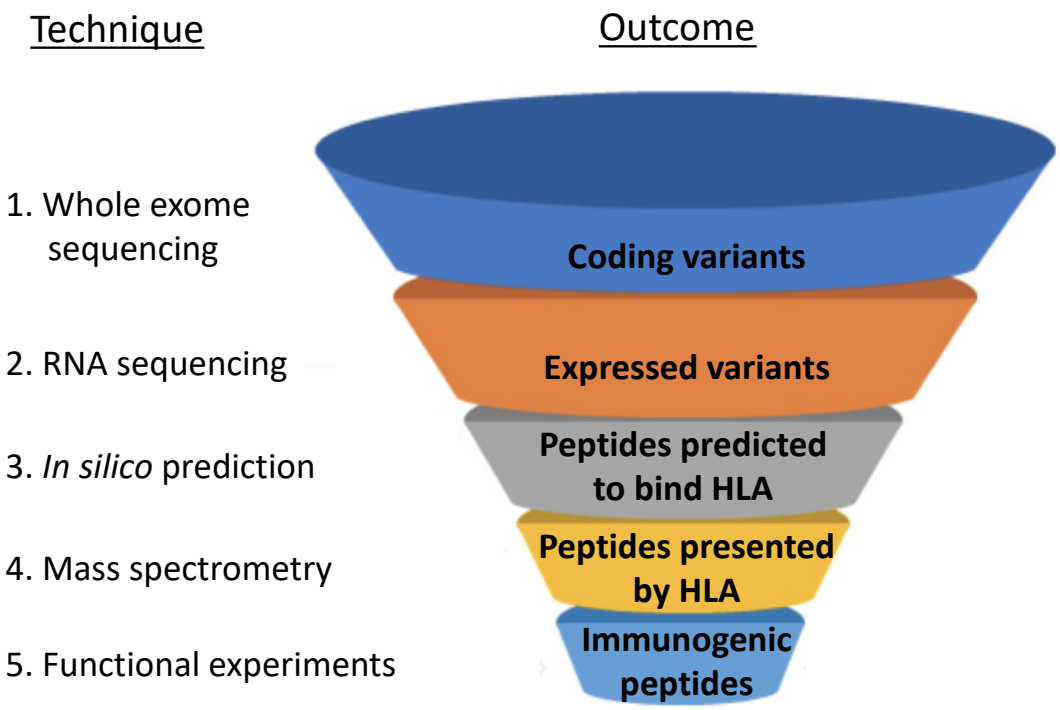


Figure 3.2: Reverse immunology approach to antigen prediction. The experimental techniques (left of figure) are performed and sequentially analysed to predict antigens. The number of candidate variants/antigens reduces on descending the funnel.

3.1.2 Patient cohort

To perform a thorough characterisation of GvL responses to AML in individual patients, it was essential to study primary patient-derived samples. In total, 15 AML patients were selected, who all had serial biobanked bone marrow and blood samples taken before and after allo-SCT. The cohort had a mean age at diagnosis of 45.6 years (SD 11.3 years) (Table A.1). Although younger than average for AML patients, who have a mean age of 68 years at diagnosis, the age distribution is comparable to other AML cohorts undergoing allo-SCT [222, 223]. Due to the toxicity associated with treatment, only younger patients with fewer co-morbid medical conditions are eligible for allo-SCT. 14/15 patients had *de novo* AML and 1/15 had AML secondary to myelodysplastic syndrome (MDS). Baseline cytogenetic and molecular genetic features of disease were diverse, with all patients classified as having either favourable or intermediate risk disease [36].

All 15 patients were treated with intensive chemotherapy and of these, 4 patients

did not respond to first-line treatment. A further 5 patients relapsed prior to transplant, of whom 2 had a morphological relapse and 3 had an increase in measurable residual disease (MRD) detected by real-time quantitative PCR (RT qPCR) (Table A.2).

The type of allo-SCT donor was mixed:

- 7 HLA-matched sibling donors.

- 6 unrelated donors (12/12 HLA match in 4 patients; 11/12 HLA match with HLA-DP mismatch in 2 patients).

- 1 mismatched sibling donor (11/12 HLA match with mismatch at HLA-A).

- 1 haploidentical sibling donor. A haploidentical sibling shares one HLA haplotype with the patient and therefore at least half of HLA alleles will be matched.

The majority of patients were in remission at the time of allo-SCT. However, 1 patient (OX905) had 16% blasts on examination of the bone marrow aspirate prior to transplantation, indicating persistent disease. All surviving patients have been followed up for at least 3.5 years post-transplant and 4/15 patients relapsed. Of these, 3 patients died from relapsed disease. Although there was no allo-SCT treatment-related mortality, GvHD of varying severity was reported in 12/15 patients, which required immunosuppressive therapy (Table A.3). However, all surviving patients have subsequently stopped immunosuppression.

3.1.3 Cell sorting for whole exome sequencing and RNA-Seq

To obtain DNA for WES, fluorescence-activated cell sorting (FACS) was used to purify live leukaemia cells ($Hoechst^-/CD45^+/CD235^-/CD3^-$) from diagnostic bone marrow samples (Figure 3.3 (a)). To obtain germline control DNA, T cells were sorted ($Hoechst^-$ $CD45^+$ $CD235^-$ $CD3^+$) from pre-transplant remission samples, where available (Figure 3.3 (b)), or from the diagnostic sample. A minimum of $1x10^5$ cells per sample were sorted, which yields approximately 300 ng DNA (Figure 3.3 (c)). A minimum of 100 ng DNA is required for WES library preparation for next-generation sequencing. Donor DNA extracted from unfractionated peripheral blood was obtained from the Transplant Immunology Laboratory in Oxford. T

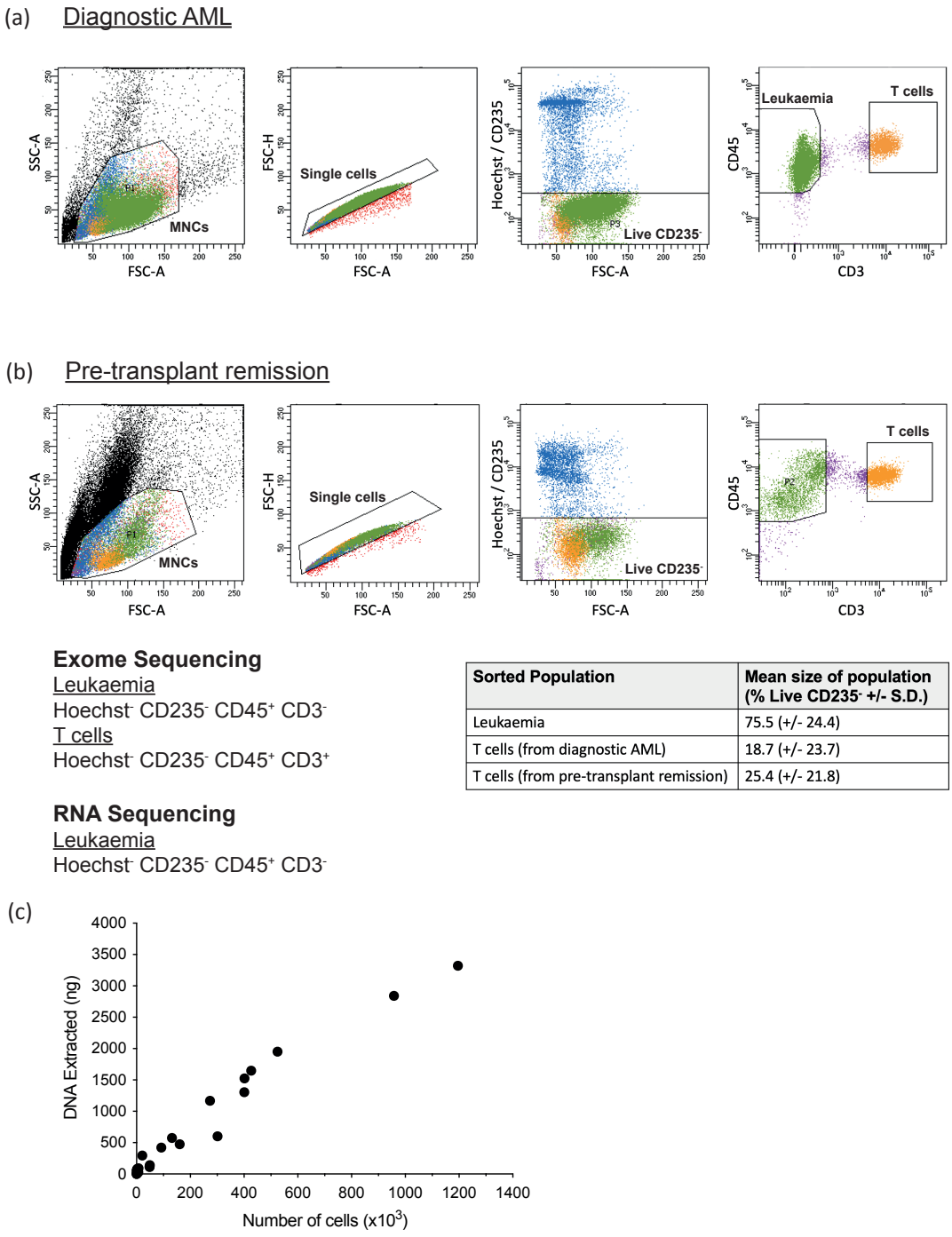
cells are suitable for WES germline control because T lymphopoiesis branches from myeloid development at an early stage of haematopoietic differentiation and therefore most are not part of the malignant clone [224]. However, a caveat is that somatic mutations commonly associated with clonal haematopoiesis, such as *DNMT3A*, have been reported in T cells of AML patients [224–226].

For bulk RNA-Seq, 5000 AML cells (Hoechst$^-$ CD45$^+$ CD235$^-$ CD3$^-$) were sorted.

In order to call leukaemia-specific somatic variants, it is essential that the purity of germline control DNA is high. If there is contamination with DNA from malignant cells, then variant reads from AML within the germline control could result in failure to call somatic variants from the AML exome. In all cases, the purity of the sorted T cell population was over 93% and the purity of AML cells was over 97% (Table 3.1).

Patient	% AML Purity	T Cell Source	% T Cell Purity
OX747	100	Pre-allograft	100
OX905	NA	Diagnostic	95
OX607	100	Pre-allograft	96
OX628	NA	Pre-allograft	99
OX866	NA	Diagnostic	98
OX813	NA	Pre-allograft	100
OX802	97	Pre-allograft	98
OX1149	NA	Pre-allograft	99
OX690	98	Pre-allograft	97
OX1072	100	Diagnostic	99
OX993	100	Diagnostic	98
OX1098	100	Pre-allograft	98
OX289	100	Diagnostic	100
OX885	NA	Pre-allograft	100
OX452	100	Diagnostic	93

Table 3.1: Purity of sorted AML and T cells. Purity of FACS-sorted populations was assessed by sorting cells and then reanalysing to determine the % of cells that fall within the sorting gate. AML cells were used for WES and bulk RNA-Seq. T cells were used for WES. The 'T Cell Source' column indicates whether a diagnostic or pre-allograft remission sample was used to obtain T cells.
NA: Not assessed.



Sorted Population	Mean size of population (% Live CD235- +/- S.D.)
Leukaemia	75.5 (+/- 24.4)
T cells (from diagnostic AML)	18.7 (+/- 23.7)
T cells (from pre-transplant remission)	25.4 (+/- 21.8)

Figure 3.3: Flow cytometry sorting of AML and T cells for exome sequencing and RNA-Seq. (a) Leukaemia cells were sorted from diagnostic samples (shown for Patient OX1072). **(b)** Where samples were available, T cells were sorted from post-transplant remission samples (shown for Patient OX690). Alternatively, they were sorted from diagnostic samples. The table shows the mean size of sorted populations across all samples, as the % of Live CD235$^-$ cells (+/- standard deviation). **(c)** Amount of DNA extracted from different numbers of cells. A minimum of 100 ng was required for exome sequencing.
MNCs: Mononuclear cells

3.2 Identifying mismatched exonic variants

3.2.1 Germline and somatic variant calling

Next-generation sequencing was performed with 100X coverage for leukaemia samples and 50X for patient T cells and donor DNA. Germline and somatic variant calling were performed using distinct bioinformatic pipelines (Figure 3.4). Germline variant calling uses the assumption of a diploid genome, which is important for informing the genotype likelihoods and variant quality calculations. This can result in variant calls as heterozygous or homozygous. In contrast, when calling somatic variants in cancer, a range of allele frequency is possible due to variation in clone size, genomic copy number and tumour purity [227]. Therefore, somatic variant callers do not use a fixed ploidy assumption.

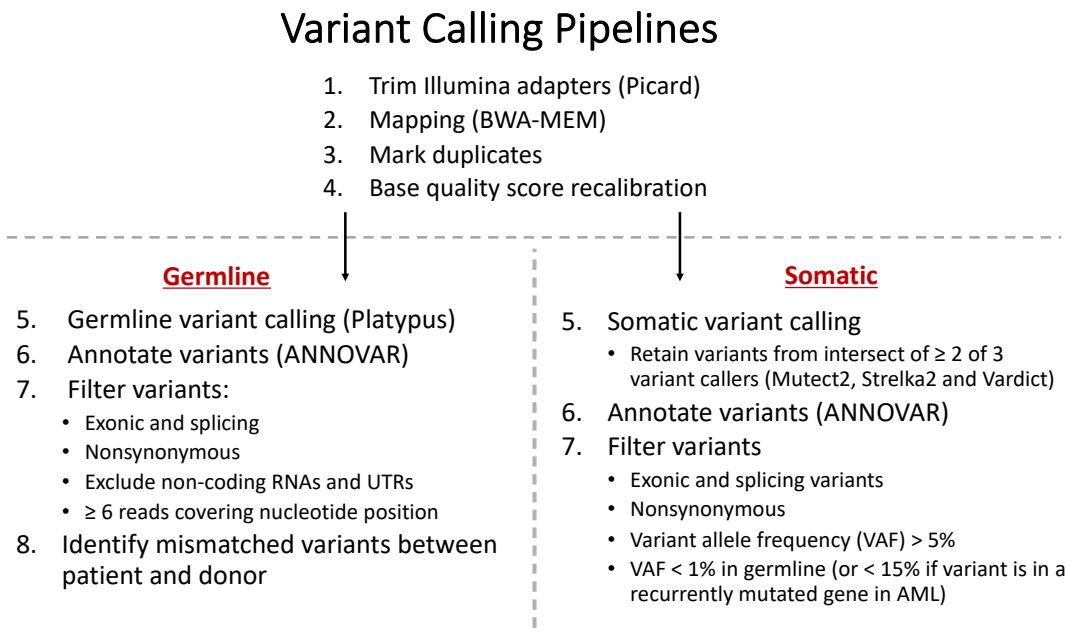


Figure 3.4: Overview of germline and somatic variant calling pipelines. Initial data pre-processing steps were performed in the same way for germline and somatic pipelines. WES reads were trimmed, mapped and processed as outlined in Section 2.3. After base quality score recalibration, the pipelines branch. Platypus was used for germline variant calling. Somatic variant calling was performed using Mutect2, Strelka2 and Vardict, with the intersect of at least 2 of the 3 callers used to retain variants.
BWA-MEM: Burrows-Wheeler aligner maximal exact match. UTR: Untranslated region.

Germline variant calling using Platypus was performed independently for each patient and their respective stem cell transplant donor, in comparison to the hg38 reference genome [213]. Variants were filtered to retain only those that: (1) are exonic or involve splice sites, (2) result in an amino acid change (i.e. are nonsynonymous), and (3) have ≥ 6 reads covering the variant position from both patient and donor (Figure 3.5). A threshold of 6 reads was chosen because at this read depth, the chance of failing to identify a true heterozygous variant is < 0.05.

A smaller number of variants were called in the donor samples for patients OX628 and OX690 (Figure 3.5 (c) - (g)). This is due to poor overall exome coverage in these samples, leading to fewer variants meeting the minimum threshold of 6 reads for calling. For some donor samples, including OX628 and OX690, the genomic DNA used to prepare sequencing libraries was degraded when assessed by agarose gel electrophoresis (Figure 3.6 (b)). In contrast, the genomic DNA for all patient germline was intact (Figure 3.6 (a)). For donor samples OX628 and OX690, there was a smaller average insert size within the DNA fragments of the library (Figure 3.6 (c) - (d)). The proportion of OX628 donor reads that were appropriately aligned was significantly lower than for other samples (76.9% compared with >95% in all other samples). The average base coverage was lower for both OX628 and OX690, which is likely to account for the lower number of variants called in these samples (Figure 3.6 (e) - (f)).

(a)

	Patient mean	Donor mean
Total variants called – min. 1 read coverage	533,552	466,998
Total variants called – min. 6 reads coverage	97,210	87,474
PASS filter	85,637	73,756
Exonic/splicing variants	20,171	17,830
Nonsynonymous variants	10,119	8,892
Remove if coverage < 6 reads in corresponding Patient/Donor sample	9,970	8,884

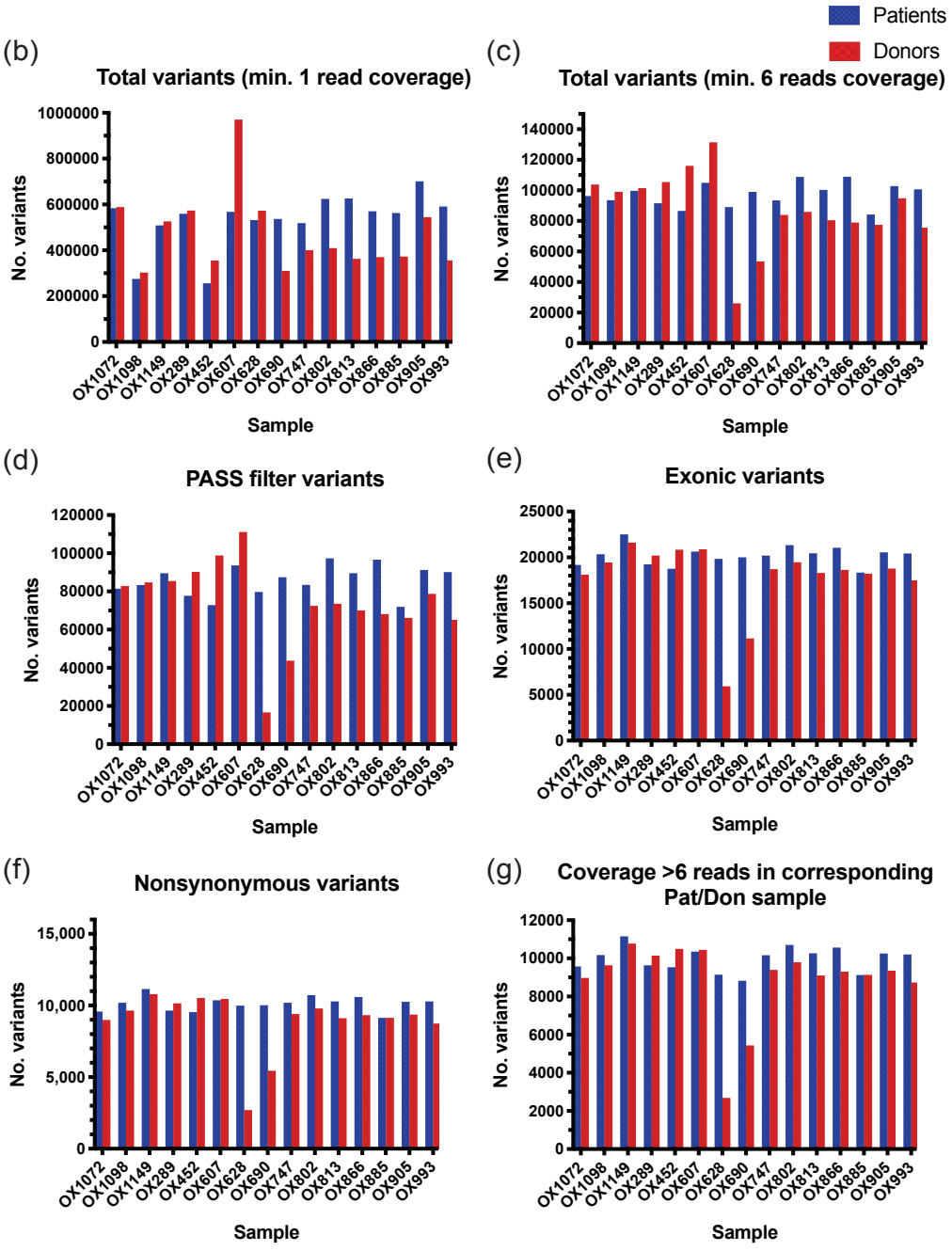


Figure 3.5: Filtering of germline variants. Germline variants were called independently for each patient and donor. **(a)** Each row in the Table shows the number of variants retained at each step. PASS filter applies default filters from the Platypus variant caller on a per-variant basis [213] (See Section 2.3). For **(b)** - **(g)**, the number of variants retained at each step is shown for each patient (blue bars) and donor (red bars).

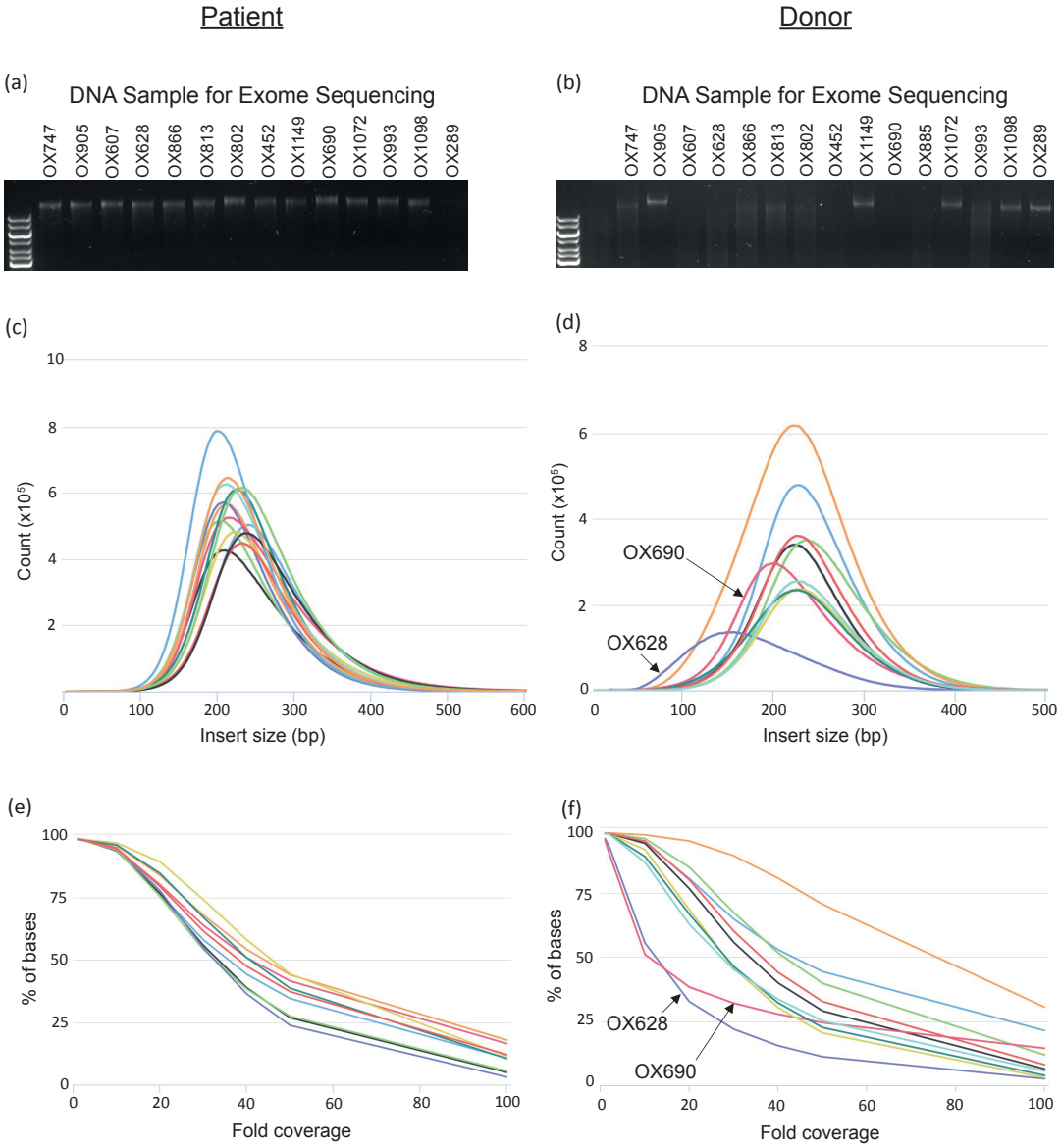


Figure 3.6: Germline DNA libraries for exome sequencing. **(a)** and **(b)** Agarose gel electrophoresis with 20 ng genomic DNA from each sample. For patient samples, there is a high molecular weight band, indicating good quality genomic DNA. In contrast, many donor samples have a smear or are difficult to visualise on the gel, indicating degradation of genomic DNA. Note there is no OX885 patient sample, as exome sequencing data for this sample was obtained from an earlier project. **(c)** and **(d)** Distribution of insert sizes for DNA library fragments (base pairs, bp), showing the number of fragments (y-axis) and insert size (x-axis). OX628 and OX690 donor samples, which have shorter average insert sizes, are highlighted. **(e)** and **(f)** The % of all target bases with at least the indicated fold coverage. OX628 and OX690 donor samples have lower average base coverage.

The majority of variants called from the patient and donor germline surpassed the minimum required coverage of 6 reads (Figure 3.7 (a)). As expected for a diploid genome, there was a bimodal pattern of variant allele frequency (VAF), with peaks at 0.5 and 1.0 (Figure 3.7 (b)).

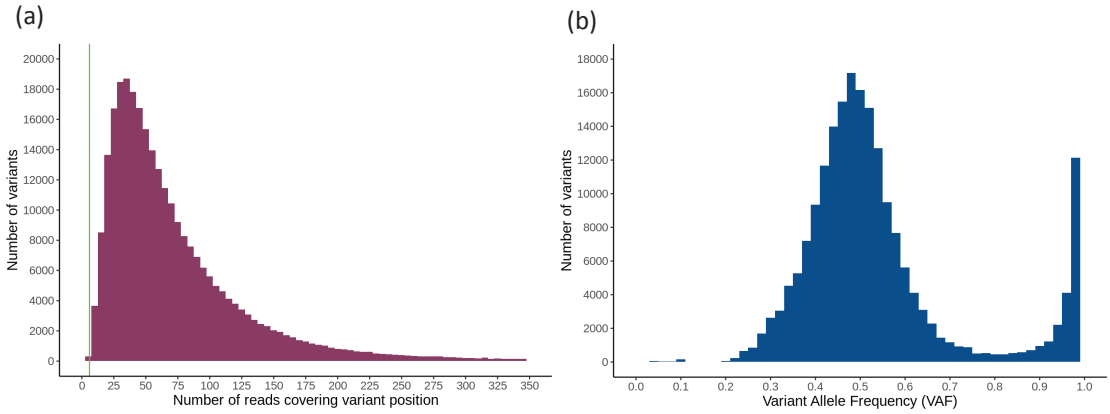


Figure 3.7: Exome sequencing coverage and VAF of germline variants. (a) For all germline variants detected across the 15 patients and donors studied, the number of reads covering each variant position is shown. The green vertical line indicates the threshold of ≥ 6 read coverage for variant calling. **(b)** VAF of germline variants across all patients and donors.
VAF: Variant allele frequency

Leukaemia-specific variants were identified by somatic variant calling from AML WES, using pre-transplant T cell WES as a germline control. Three variant callers were used (Mutect2, Strelka2 and Vardict, [210, 211, 227]) and variants were retained only if identified by ≥ 2 variant callers. Variants were filtered to only include variants that: (1) are nonsynonymous, (2) are exonic or involve splice sites, (3) have VAF $\geq 5\%$, and (4) are associated with $< 1\%$ VAF in germline (or $< 15\%$ if the variant is in a gene that is recurrently mutated in AML).

Germline mismatches can only produce GvL peptide antigens if the amino acid sequence is present in the patient but not the donor, as this results in a sequence that could be recognised as foreign by donor T cells. This can occur if: (1) The donor is homozygous for the reference allele and the patient has the variant allele (potential alloreactive response against the variant allele), or (2) the donor is homozygous for the variant allele and the patient has the reference allele (potential alloreactive response against the reference allele) (Figures 3.8 and 3.9).

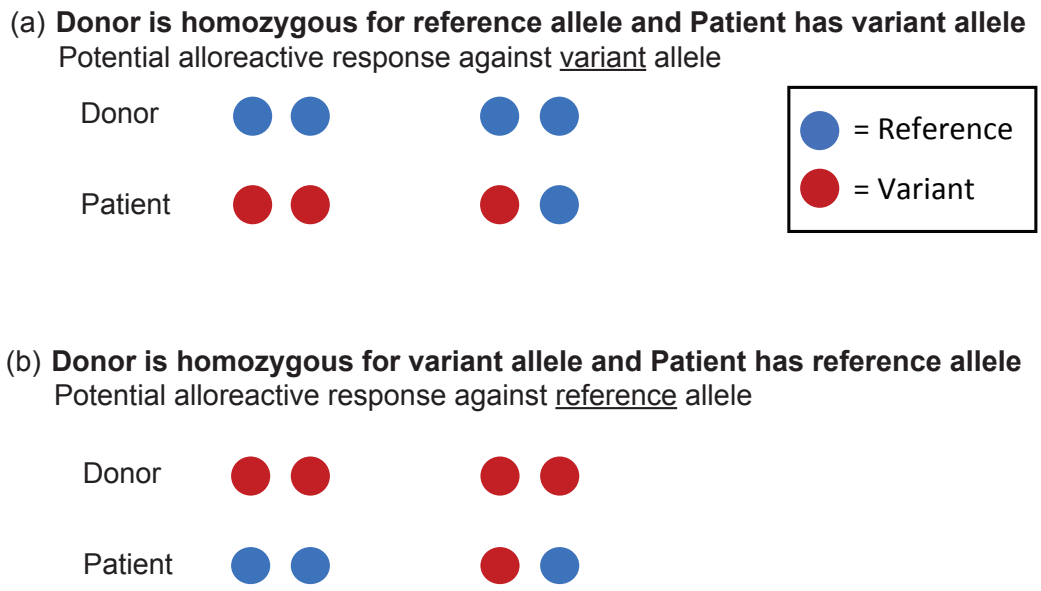


Figure 3.8: Identifying mismatched germline variants between patient and donor. Alloreactive T cell responses are possible where a nonsynonymous nucleotide sequence is present in the patient but not the donor. Responses may be directed against peptide antigens resulting from **(a)** the variant allele (red), or **(b)** the reference allele (blue).

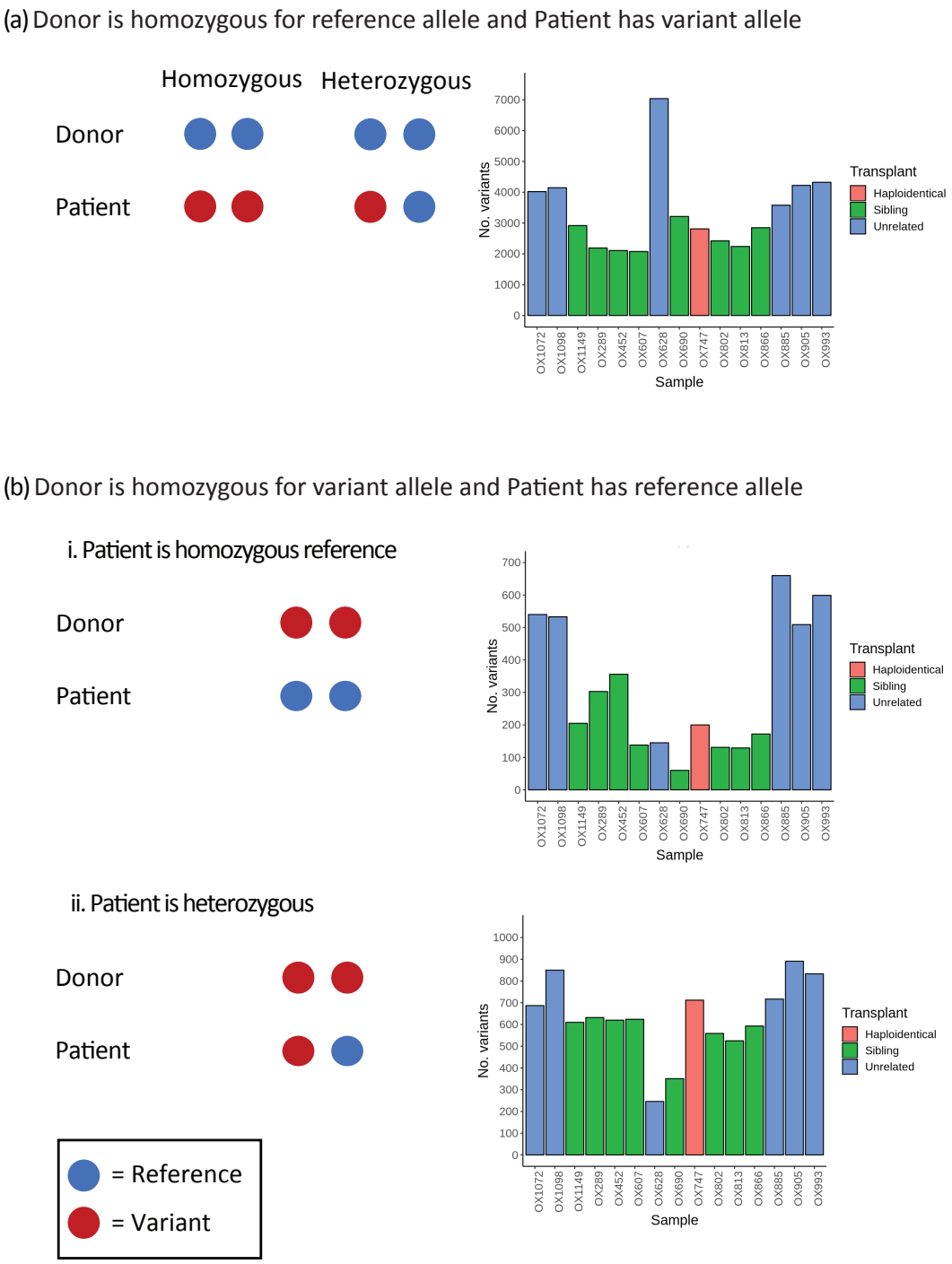


Figure 3.9: Number of patient-donor mismatched germline variants. For each patient-donor pair, the number of mismatched variants is shown where an alloreactive response could be directed against peptide antigens resulting from **(a)** the variant allele (red), or **(b)** the reference allele (blue). The type of stem cell transplant donor used for each patient is indicated by the colour of the bar: haploidentical (red), sibling (green) and unrelated (blue).

The majority of mismatched germline variants were single nucleotide variants (SNVs). A smaller number of variants (5.84%) were a result of substitutions of > 1 nucleotide, without changing the length of the amino acid sequence. Insertions and deletions comprised 1.60% and 2.36% of variants, respectively (Figure 3.10 (a)). The most common coding consequence was missense (single amino acid substitution) (Figure 3.10 (b)).

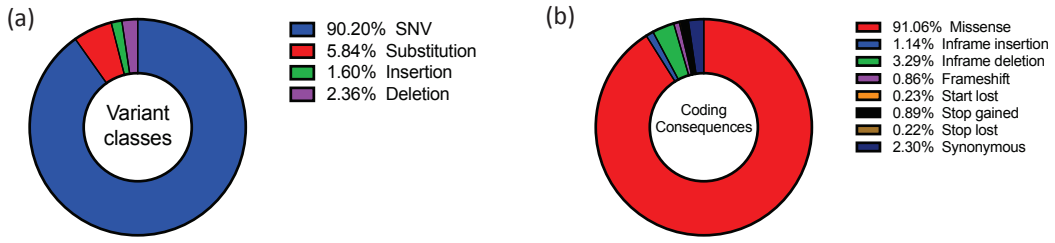


Figure 3.10: Annotation of nucleotide variants. (a) Classes of nucleotide variant detected by exome sequencing. **(b)** Consequences of variants on the amino acid sequence. Data obtained using Ensembl Variant Effect Predictor (VEP).
SNV: Single nucleotide variant.

3.2.2 Germline mismatched variants

The number of mismatched germline variants in patient-donor pairs was significantly different with sibling and unrelated donor transplants. As expected, approximately twice as many variants were found in the context of an unrelated donor compared with a sibling donor (average for an unrelated donor of 5745 vs 3253 with a sibling donor), which is in keeping with findings from previous studies [123, 228] (Figure 3.11).

3.2.3 Somatic variants

An average of 52.1 leukaemia-specific somatic variants were identified per patient. This accounts for only 1.3% of the total genetic variation between a patient's AML and their donor, with the remaining 98.7% accounted for by germline variation. For each patient, an average of 3.9 somatic mutations were in genes that are recurrently mutated in AML patients, many of which are putative driver mutations (Figure 3.12 and Table 3.2). Recurrently mutated genes were defined from previous large studies

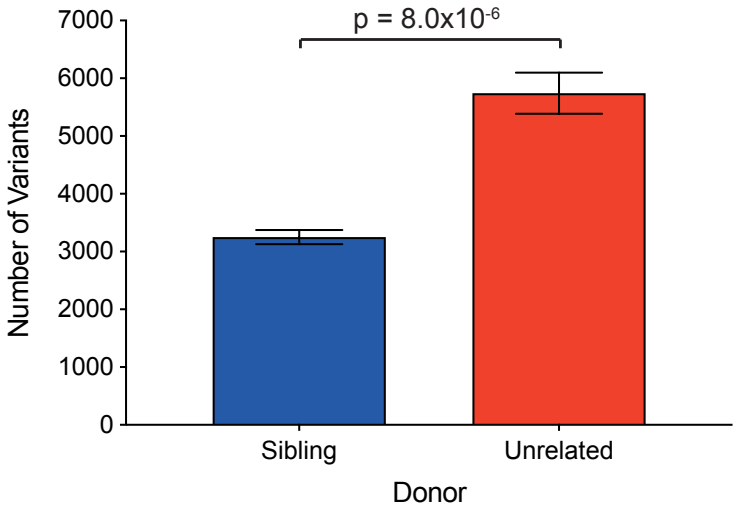


Figure 3.11: Mismatched germline variants between patient and donor. Mean number of mismatched germline variants between patient and donor according to donor type +/- SEM.
SEM: Standard error of the mean

of AML patients [4, 141, 229–231]. The other mutations are likely to be principally passenger mutations, which do not impact on the function of the leukaemia cell, but have been acquired during the development of leukaemia.

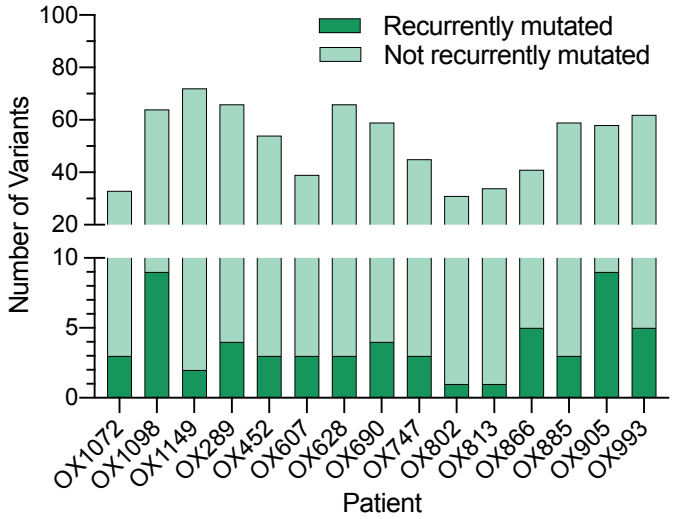


Figure 3.12: Number of somatic variants in AML. Total number of somatic variants (total bar) and those in recurrently mutated genes in AML (dark green bar) [4, 141, 229–231].

Mutations in 31 recurrently mutated genes were detected across the 15 patients

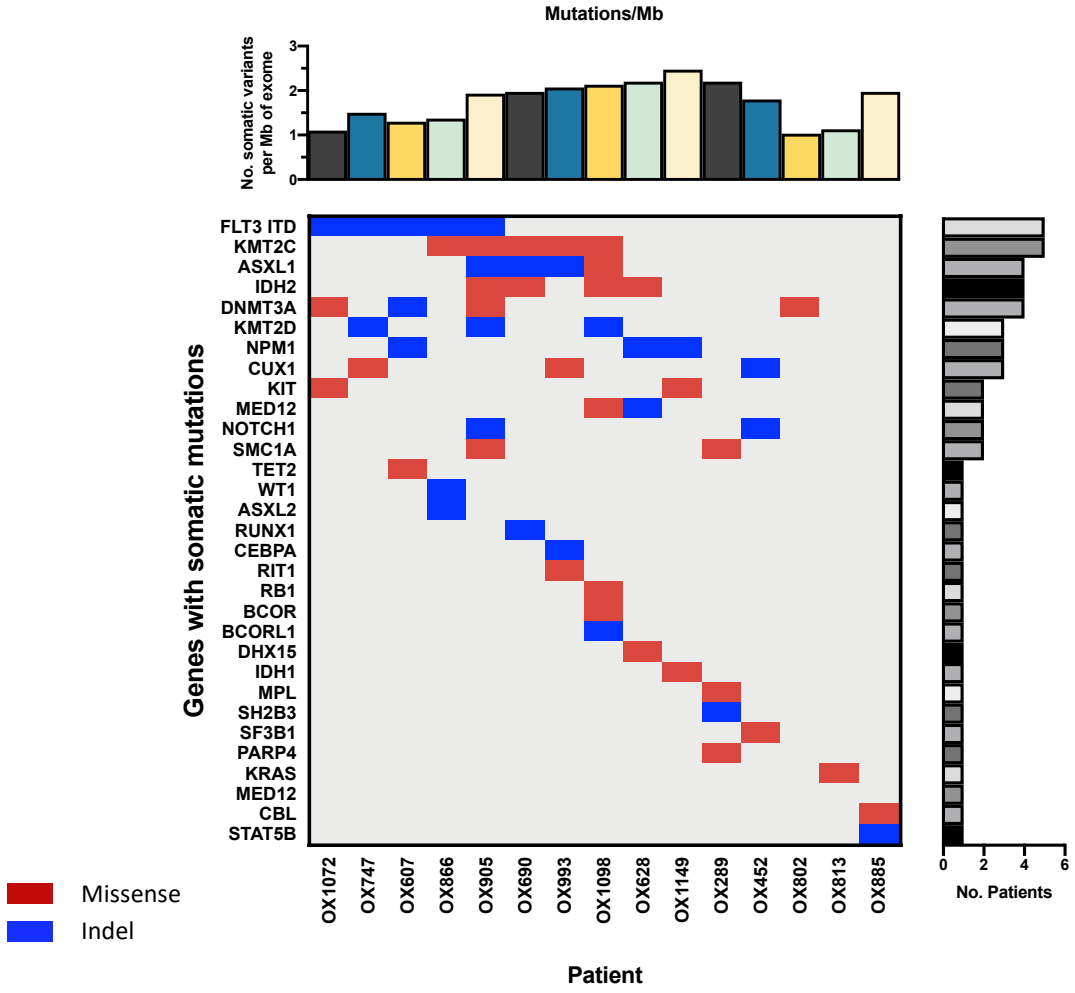
Patient	Donor Type	No. somatic variants in recurrently mutated AML genes	Somatic variants in recurrently mutated AML genes (VAF%)	Mismatched germline variants per patient/donor pair
OX289	Sib	4	**MPL** – 36%; **SH2B3** – 97%; **PARP4** – 19%; **SMC1A** – 19%	3127
OX452	Sib	3	**SF3B1** – 6%; **CUX1** – 7%; **NOTCH1** – 6%	3083
OX607	Sib	3	**DNMT3A** – 49%; **TET2** – 97%; **FLT3 ITD** – 31%	2838
OX628	MUD	3	**DHX15** – 5%; **IDH2** – 50%; **MED12** – 12%	7428
OX690	Sib	4	**KMT2C** – 30%; **IDH2** – 36%; **ASXL1** – 41%; **RUNX1** – 51%	3624
OX747	Haplo	3	**CUX1** – 6%; **KMT2D** – 6%; **FLT3 ITD** – 38%	3723
OX802	Sib	1	**DNMT3A** – 5%	3114
OX813	Sib	1	**KRAS** – 11%	2894
OX866	Sib	5	**ASXL2** – 5%; **KMT2C** – 11%; 2 x **WT1** – 14%, 16%; **FLT3 ITD** – 29%	3613
OX885	MUD	3	**CBL** – 4%; 2 x **STAT5B** – 4%, 7%	4956
OX905	MUD	9	**DNMT3A** – 47%; 2 x **KMT2C** – 18%, 10%; **NOTCH1** – 4%; **KMT2D** – 5%; **IDH2** – 48%; **ASXL1** – 5%; **SMC1A** – 37%; **FLT3 ITD** – 25%	5622
OX993	MUD	5	**RIT1** – 33%; **CUX1** – 10%; **KMT2C** – 6%; **CEBPA** – 10%; **ASXL1** – 35%	5686
OX1072	MUD	3	**DNMT3A** – 48%; **KIT** – 45%; 2 x **FLT3 ITD –** 15%, 10%	5248
OX1098	MUD	9	2 x **DNMT3A** – 23%, 8%; **KMT2C** – 7%; **KMT2D** – 4%; **RB1** – 8%; **IDH2** – 25%; **BCOR** – 17%; **MED12** – 12%; **BCORL1** – 14%	5529
OX1149	Sib	2	**IDH1** – 43%; **KIT** – 68%	3733
	Mean	**3.9**	**Mean**	**4281**
			Mean (Sib)	**3253**
			Mean (MUD)	**5745**

Table 3.2: Somatic variants detected by exome sequencing. Nonsynonymous somatic variants in genes that are recurrently mutated in AML. The VAF (%) of each mutation is shown.
Sib: Sibling, MUD: Matched unrelated donor, Haplo: Haploidentical, VAF: Variant allele frequency.

studied, with a combination of missense mutations and insertions/deletions. Increasing numbers of mutations per megabase (Mb) of exonic DNA correlates with response to immune checkpoint inhibitor therapy, which is likely to be due to the formation of leukaemia-specific neoantigens. The presence of > 10 mutations/Mb is associated with a higher likelihood of response to therapy [99, 232]. AML has a relatively low mutational burden and therefore neoantigens are less likely to play an important role. Indeed, in all AML patients studied, there were < 3 mutations/Mb exonic DNA (Figure 3.13 (a)).

The prevalence of somatic mutations in this study was compared with 3 larger published cohorts of 113, 200 and 1540 AML patients [4, 141, 223]. The spectrum of mutations across the cohorts was similar and for most genes, there was not a significant difference in the frequency of patients with mutations. However, our cohort was significantly enriched for the presence of mutations in *ASXL1* when compared to each of the studies (Figure 3.13 (b)).

(a)



(b)

Gene	Current study (n=15)	Quek (Blood Adv 2016) (n=113)		TCGA (NEJM 2013) (n=200)		Papaemmanuil (NEJM 2016) (n=1540)	
	No. patients (%)	No. patients (%)	Fisher's exact test (two-tailed)	No. patients (%)	Fisher's exact test (two tails)	No. patients (%)	Fisher's exact test (two tails)
ASXL1	4 (26.7%)	8 (7.1%)	0.0349	5 (2.5%)	0.0164	71 (4.6%)	0.0045
DNMT3A	4 (26.7%)	19 (16.8%)	0.4712	51 (25.5%)	1	362 (23.5%)	0.7621
FLT3 ITD	5 (33.3%)	43 (38.1%)	0.7847	39 (19.5%)	0.1973	345 (22.4%)	0.3494
FLT3 PM	0 (0%)	7 (6.2%)	1	17 (8.5%)	0.6145	203 (13.2%)	0.2422
IDH1	1 (6.7%)	13 (11.5%)	1	39 (19.5%)	1	106 (6.9%)	1
IDH2	4 (26.7%)	14 (12.4%)	0.2257	20 (10%)	0.0703	151 (9.8%)	0.0541
KIT	2 (13.3%)	10 (8.8%)	0.6324	8 (4%)	0.1477	66 (4.3%)	0.1373
NPM1	3 (20%)	17 (15%)	0.704	54 (27%)	0.764	440 (28.6%)	0.5761
NRAS	0 (0%)	12 (10.6%)	0.3577	15 (7.5%)	0.6066	270 (17.5%)	0.0887
PTPN11	0 (0%)	5 (4.4%)	1	9 (4.5%)	1	120 (7.8%)	0.6224
RUNX1	1 (6.7%)	17 (15%)	0.6931	19 (9.5%)	1	137 (8.9%)	1
SRSF2	0 (0%)	11 (9.7%)	0.3597	1 (0.5%)	1	93 (6%)	1
TET2	1 (6.7%)	21 (18.6%)	0.4654	17 (18.5%)	1	156 (10.1%)	1
TP53	0 (0%)	6 (5.3%)	1	16 (8%)	0.6097	101 (6.6%)	0.6176
WT1	1 (6.7%)	8 (7.1%)	1	12 (6%)	1	77 (5%)	0.5395

Figure 3.13: Somatic variants in recurrently mutated genes. (a) Mutations in genes that are recurrently mutated in AML (rows) in each of the patients studied (columns). Missense mutations are red and insertions/deletions (indels) are blue. The total number of patients with a mutation in the gene is indicated on the right side. At the top of the figure, the total number of somatic mutations per megabase (Mb) of exonic DNA for each patient is shown. **(b)** Comparison of the study cohort with 3 published cohorts of AML: Quek et al [223], the Cancer Genome Atlas (TCGA) [141] and Papaemmanuil et al [4]. Fisher's exact test was used to compare the frequency of mutations in the indicated genes between our cohort and each of the other studies. The value for *ASXL1* (highlighted in green) was significantly different to the other studies ($p < 0.05$).

3.2.4 Population frequency of mismatched variants

GvL variants that are common in the population are the most attractive candidates for translation into clinical practice. If used as an immunotherapy target, a greater number of patients could be treated by targeting the resultant antigen. If used to select a donor that is predicted to elicit a productive GvL response, common antigens could also have greater clinical utility. However, at very high allele frequencies, it would become more challenging to produce a mismatch between patient and donor because only a small proportion of the donor population would lack the variant.

The median population allele frequency of all germline variants detected was 18.55% (Figure 3.14) [233]. When comparing across all 15 patients studied, 55.25% of variants were present in only 1 patient-donor pair, whereas the remaining variants were present in up to 10 patient-donor pairs (Figure 3.15 (a)). As expected, when present in a greater number of the patients studied, the population allele frequency from the gnomAD database was also higher (Figure 3.15 (b)).

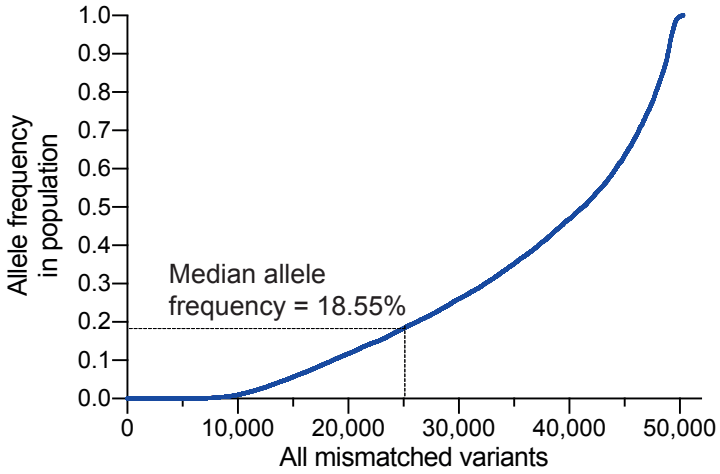


Figure 3.14: Allele frequency of germline variants. Variants are arranged on the x-axis in order of increasing allele frequency in the population (gnomAD database). The median allele frequency is 18.55%.

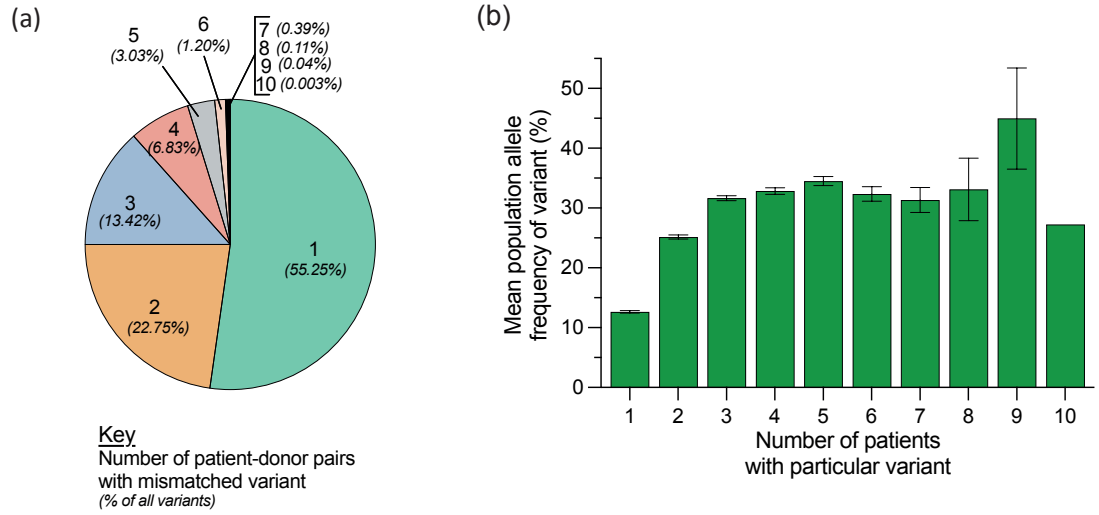


Figure 3.15: Variants detected across multiple patients. (a) 55.25% of variants were mismatched in just 1 patient-donor pair. Of 15 patients studied, the maximum number of patient-donor pairs sharing a given variant is 10. **(b)** The mean population frequency of variants (y-axis) according to the number of patient-donor pairs with the variant (x-axis) +/- SEM
SEM: Standard error of the mean

3.3 RNA expression of variants

3.3.1 Gene expression in AML

The variant DNA nucleotide must be transcribed in order for the amino acid sequence to be subsequently translated, processed, bound by HLA and presented to donor-derived T cells. Therefore, bulk RNA sequencing was performed on FACS-sorted AML blasts to assess expression of patient-donor mismatched variants. The presence of the mismatched nucleotide was assessed within RNA-Seq reads. Expression was detected for 28.3% of germline variants and 20.8% of somatic variants. The lower number for somatic variants is likely to be due to the presence of variants in small AML sub-clones, in which case the sensitivity to detect expression is limited by the number of RNA-Seq reads covering the variant position (Figure 3.16).

The majority of genetic differences between patient and donor (98.7%) are of germline origin and therefore it is very likely that minor histocompatibility antigens

are more important than neoantigens in driving therapeutic GvL responses.

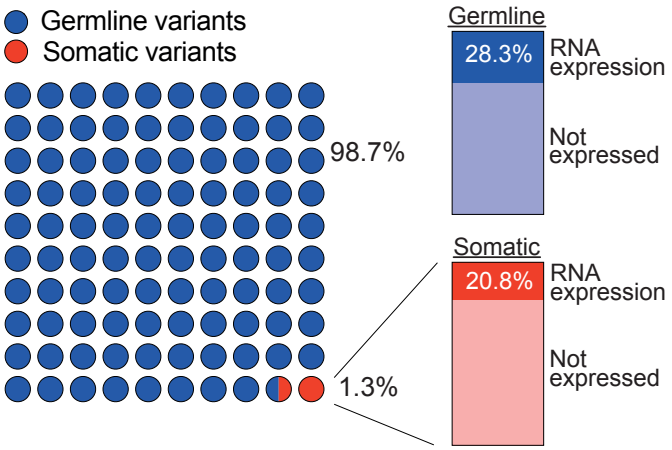


Figure 3.16: RNA expression of variants. Of all mismatched variants between patients and their respective donors, 98.7% are germline and only 1.3% are somatic in origin. RNA expression of the variant nucleotide(s) was detected for 28.3% of germline variants and 20.8% of somatic variants.

3.3.2 Gene expression in normal tissues

If a mismatched variant is expressed in normal tissues, it could result in T cell responses directed against healthy tissues and manifest as GvHD. Therefore, expression of each gene containing a mismatched variant was assessed in 37 normal tissues, using publicly available RNA-Seq data from the Human Protein Atlas v18.1 [234]. For all of the patient-donor mismatched variants across the 15 patients, expression of the gene containing each variant was first assessed in AML. Of all 34,210 variants, gene expression in AML was found for 11,503 variants (33.6%). Of all variants expressed in AML, 4.5% had either absent or very low expression in normal tissues and therefore could potentially be targeted therapeutically with a lower risk of triggering GvHD (Figure 3.17).

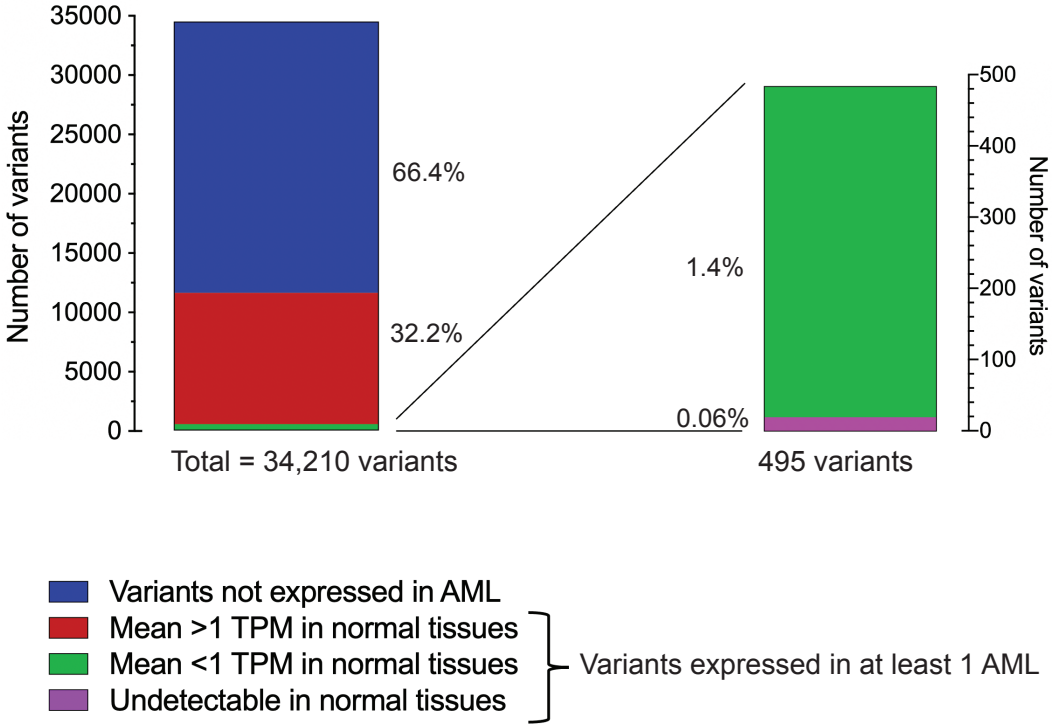


Figure 3.17: Normal tissue expression of genes containing patient-donor mismatched variants. The bar on the left represents all patient-donor mismatched variants across the 15 patients. The bar on the right is the lower portion of the left bar, which has been enlarged. Gene expression in AML is represented in red, green and purple according to the mean expression in normal tissues, using RNA-Seq data from the Human Protein Atlas v18.1. The average RNA-Seq expression level was calculated using all available samples in the database to calculate a mean TPM value. In red, mean expression $>$ 1 TPM in normal tissues; in green, expression $<$ 1 TPM in normal tissues; in purple, undetectable expression in normal tissues.
TPM: Transcripts per million

3.4 HLA binding prediction

In order to elicit alloreactive T cell responses, patient-donor peptides must be presented by HLA molecules on the cell surface as minor histocompatibility antigens. HLA-binding affinity prediction was performed *in silico* using NetMHCpan 4.0 [160] and NetMHCIIpan 3.0 [161], which use artificial neural networks trained on existing HLA-binding affinity data. This method integrates: (1) the patient-donor mismatched variants from exome sequencing, (2) RNA-Seq expression, and (3) the

patient's HLA alleles. HLA typing was performed by the National Health Service (NHS) Transplant Immunology and Immunogenetics Laboratory (Oxford, UK). The accuracy of machine learning predictions is dependent on a large training dataset and therefore HLA-binding predictions are better for common HLA alleles, which have a larger amount of available affinity data.

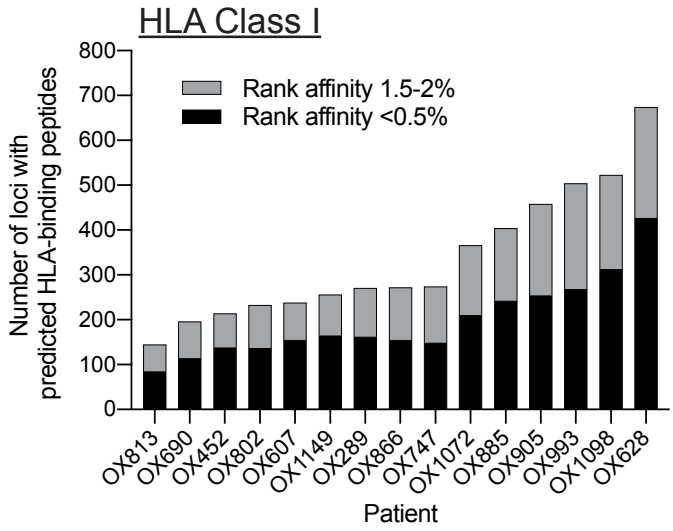
Alloreactive T cell responses can be directed against amino acid sequences resulting from variant alleles (Figure 3.8 (a)) or from the reference allele (Figure 3.8 (b)), dependent on which is uniquely present in the patient but not their stem cell transplant donor. Accordingly, HLA-binding peptides were predicted for either the variant or reference sequence.

For HLA class I binding predictions, all possible 8-11 amino acid peptide windows across mismatched variants were assessed and assigned a rank affinity value. This value is derived by comparing the predicted binding to a large panel of random natural peptides and those in the top 0.5% binding affinity are designated as high affinity binders, whereas those between 0.5% and 2% are low affinity binders. For HLA class II binding prediction, all possible 15-amino acid peptide windows across mismatched variants were assessed and given an IC_{50} (concentration of peptide required to displace 50% of a known binding peptide). Different units were used for for HLA class I and II binding due to an updated version of the class I binding algorithm (NetMHCpan 4.0), which found that percentile ranks are superior to fixed affinity values when comparing across different HLA molecules [160].

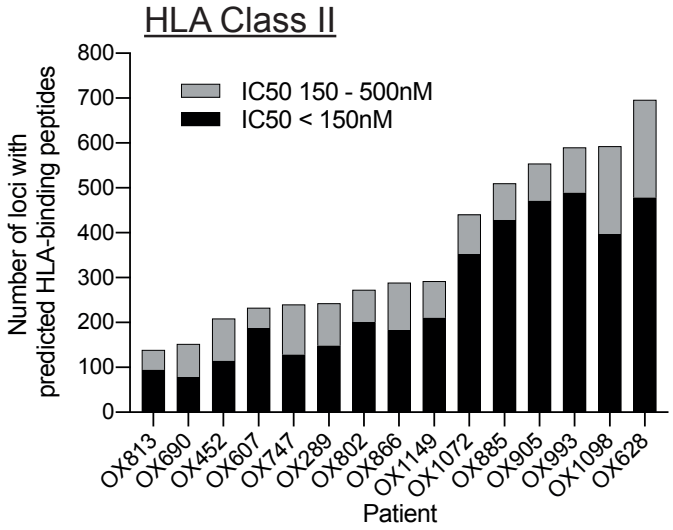
In order to be bound by HLA, the gene must be expressed. Therefore for all predictions, only those variants present at >1 count per million (CPM) reads from RNA-Seq were retained. Peptides were excluded if the sequence was present elsewhere in the *Homo sapiens* reference proteome (Uniprot), as this would be present in the donor's proteome. In this scenario, the donor is likely to have immune tolerance to the peptide sequence. For each patient, encompassing both high and low affinity predicted binding, a total of 145 - 674 peptides were predicted to bind HLA class I (Figure 3.18 (a)) and 139 - 696 peptides were predicted to bind HLA

class II (Figure 3.18 (b)) . There was substantial overlap between peptides predicted to bind HLA class I and II, with 43.7% predicted to bind both (Figure 3.18 (c)).

(a)



(b)



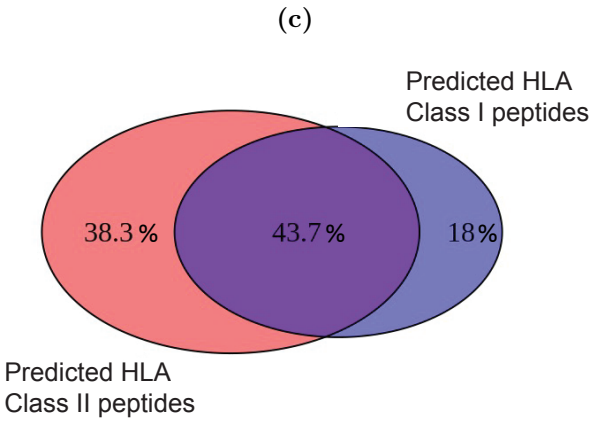


Figure 3.18: HLA binding prediction. (a) and **(b)** Total number of patient-donor mismatched peptides predicted to bind patient-specific HLA class I and II molecules, respectively [160, 161]. Where multiple peptide windows across the mismatched amino acid were predicted to bind, only the peptide with highest predicted binding affinity was counted. Black bars represent high affinity binders (Class I, rank affinity <0.5%; Class II, IC_{50} <150nM) and grey bars represent low affinity binders (Class I, rank affinity 1.5 - 2%; Class II, IC_{50} 150 - 500nM). **(c)** Overlap between predicted HLA class I and II binding peptides. 43.7% of variants had peptides that were predicted to bind to both class I and II.

3.5 Immunopeptidomic analysis of AML

Preparation of samples and analysis by mass spectrometry was performed by Dr Annalisa Nicastri, Dr Nicola Ternette and Dr Silvia Salatino. I analysed the data and identified the resultant candidate GvL antigens.

3.5.1 Approach to identify HLA-bound peptides using tandem mass spectrometry

The accuracy of HLA-binding predictions, particularly for class II and for uncommon class I alleles, is very limited. A more direct approach to identify candidate alloreactive peptides is by analysing peptides that are bound by HLA on AML cells. AML cells were lysed and HLA molecules immunoprecipitated sequentially. Firstly, an anti-HLA class I antibody (W6/32) was used to purify class I. HLA-DR was then immunoprecipitated using an anti-HLA-DR antibody (L243). Finally, HLA-DP and -DQ were immunoprecipitated from the lysate using a pan-class II antibody (IVA12). As HLA-DR was removed in the prior step, only HLA-DP

and -DQ remain. Peptides were dissociated from the immunoprecipitated HLA molecules upon acetic acid denaturation. Peptides were then separated into different fractions using high performance liquid chromatography (HPLC) and analysed by tandem mass spectrometry (MS/MS) [120].

Peptides were identified from the mass spectra by comparison with a reference proteome, using PEAKS DB software [235]. This entails a combination of *de novo* sequencing to derive the peptide sequences directly from the MS/MS spectra and a database search. Individualised reference proteome databases were generated for each patient, incorporating the variants identified from exome sequencing. Any peptides identified by MS/MS that incorporate a patient-donor mismatched variant could then be identified as candidate alloreactive peptides (Figure 3.19).

Unlike *in silico* prediction of HLA binding (Section 3.4), peptides identified by elution from HLA on patients' AML cells have demonstrated that the protein of origin can be processed by the cell and produce peptides that are presented by HLA.

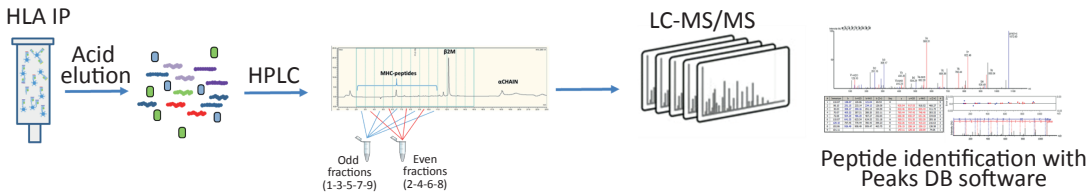


Figure 3.19: Tandem mass spectrometry of peptides eluted from HLA molecules on AML. AML cells were lysed and HLA molecules purified by immunoprecipitation from the lysate. HLA-associated peptides were dissociated from HLA molecules by acid elution and then separated by HPLC. The peptide fraction was analysed by LC-MS/MS.
IP: Immunoprecipitation, HPLC: High performance liquid chromatography, LC-MS/MS: Liquid chromatography-tandem mass spectrometry

3.5.2 Detection of HLA-bound peptides

The mean number of unique peptides from each patient identified from HLA class I, HLA-DR and -DP/-DQ fractions was 870, 1026 and 337, respectively (Figure 3.20 (a)). The mean total number of unique peptides per patient, accounting for peptides bound to all HLA molecules, was 2233 (Figure 3.20 (b)).

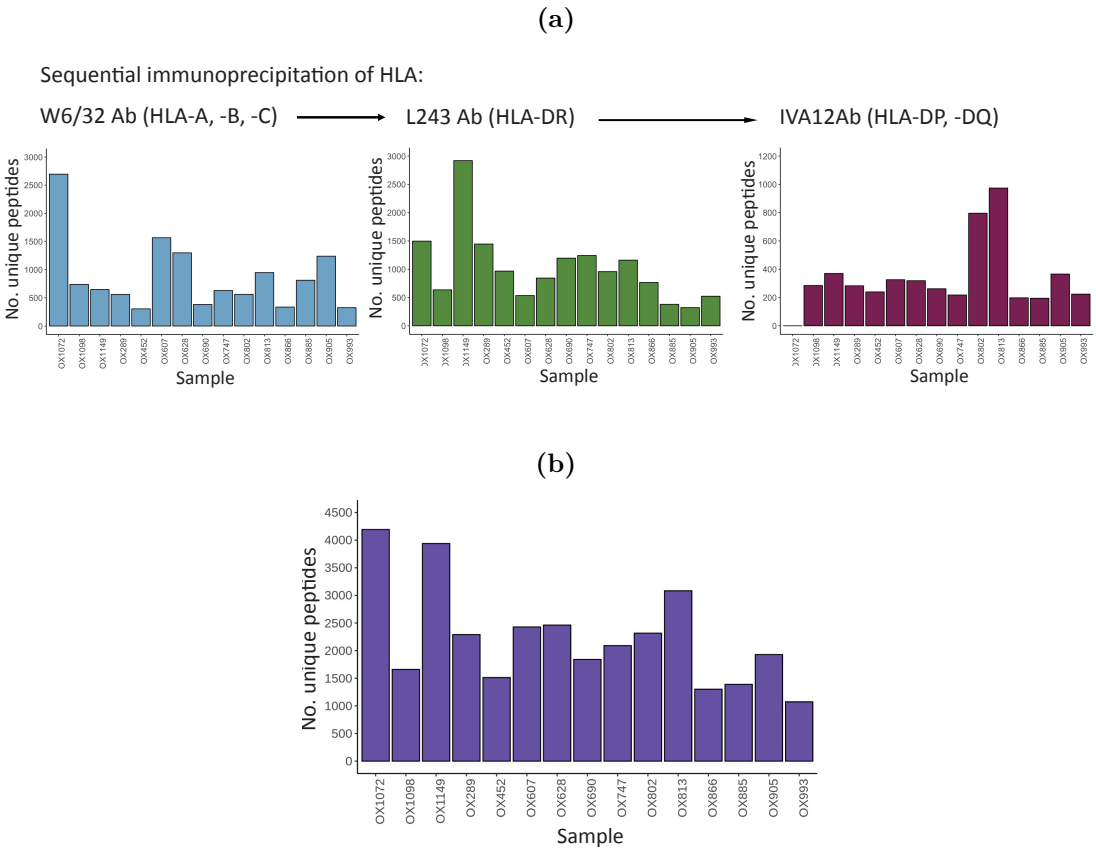


Figure 3.20: HLA molecules were immunoprecipitated, peptides eluted and then analysed by mass spectrometry. (a) HLA molecules were first immunoprecipitated using an anti-HLA class I antibody (W6/32). HLA-DR was then immunoprecipitated using an anti-HLA-DR antibody (L243). Finally, HLA-DP and -DQ were immunoprecipitated using a pan-class II antibody (IVA12). Peptides were dissociated from HLA by acid elution. Graphs show the total number of unique peptides detected per patient from each immunoprecipitation. **(b)** Total number of unique peptides detected per patient across all immunoprecipitations.

As expected, the average peptide length was shorter for class I than class II and the distribution of peptide lengths was narrower. The class I binding groove is closed at both ends by conserved tyrosine residues, which restricts the peptide length to 8-11 amino acids [236]. The most common peptide length was 9 amino acids (Figure 3.21 (a)). In contrast, the class II binding groove is open at both ends and therefore peptides with a wider length distribution are able to bind, up to 25 amino acids in length (Figure 3.21 (b) and (c)).

Peptides bound to HLA class I have conserved amino acid residues at anchor positions, which are particularly important for peptide binding. Anchor residues

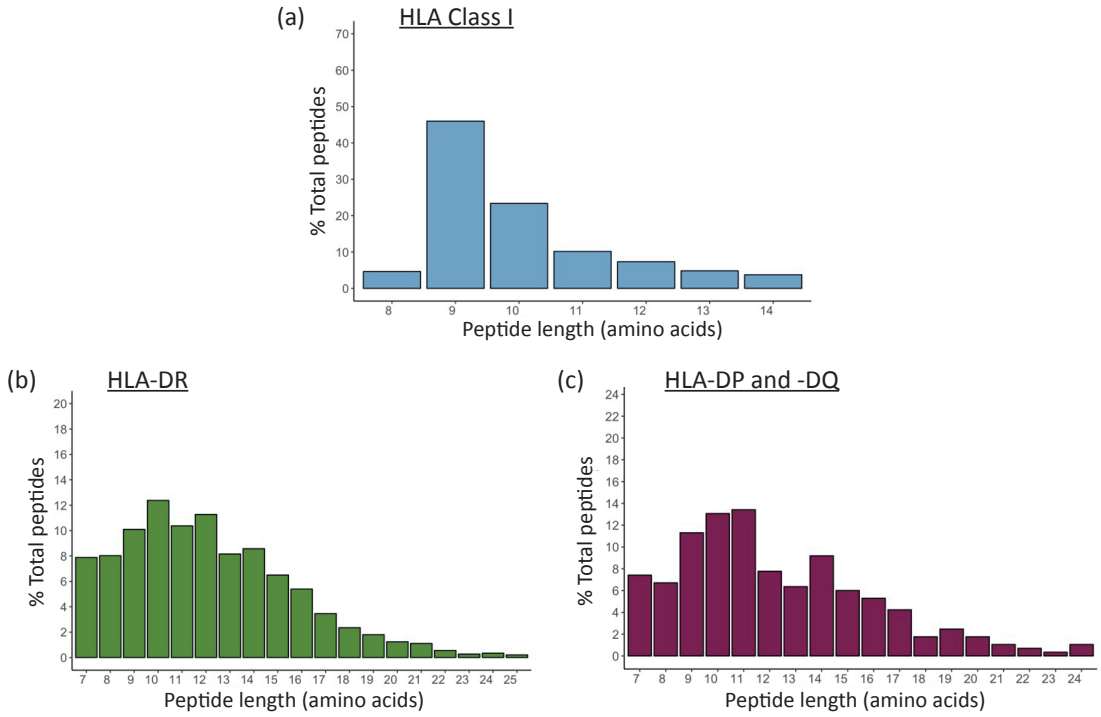


Figure 3.21: Length of peptides eluted from HLA molecules. Proportion of peptides (%) detected of different amino acid lengths bound to **(a)** HLA class I, **(b)** HLA-DR, and **(c)** HLA-DP and -DQ.

are at the second and last (C-terminal) positions [237]. The amino acid preference at these positions varies according to the restricting HLA allele. The presence of conserved residues at positions 2 and 9 was confirmed for the 9-amino acid peptides eluted from HLA class I from each patient. Patients OX1072 and OX905 are homozygous for HLA-A*02:01. Prior studies have found that most HLA-A*02-binding peptides are 9 amino acids in length and have the anchor residues leucine (L) at position 2, and leucine (L) or valine (V) at position 9 [238]. Accordingly, these residues were found at anchor positions for these patients (Figure 3.22 (a) and (b)). Patients OX607 and OX885, who have alternative HLA-A alleles (OX607: A*24:02 / A*26:01 and OX813: A*68:01 / A*11:01) showed different amino acid preference at anchor positions (Figure 3.22 (c) and (d)).

Although peptides bound to HLA class II are usually longer than those bound to class I, residues within the binding pocket of the antigen binding cleft of class II molecules typically interact with a 9-amino acid binding core within the peptide

[239]. As this core can be in different positions along the peptide, it is more difficult to identify conserved anchor residues.

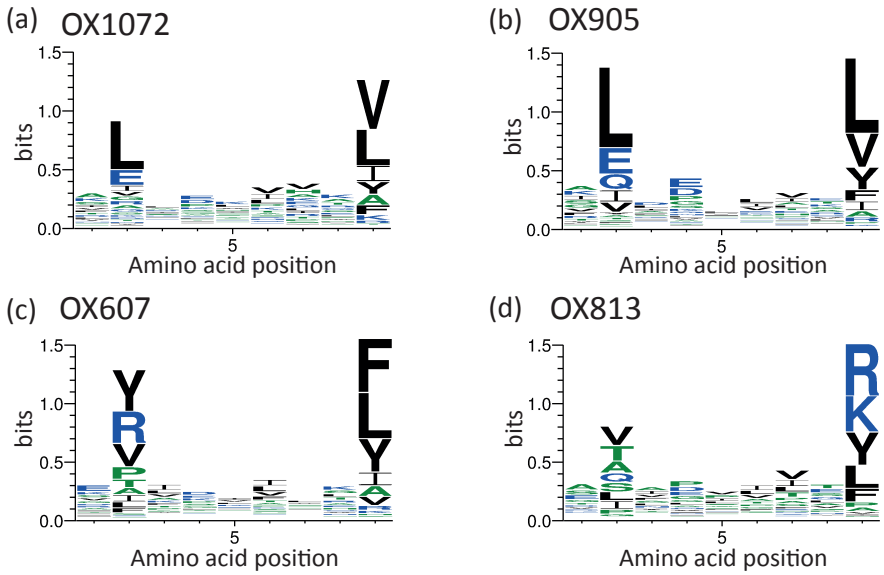


Figure 3.22: Amino acid representation by position along peptide sequences. **(a)** - **(d)** A graphical representation of the amino acid sequences of 9-amino acid peptides eluted from HLA class I is shown for 4 patients (OX1072, OX905, OX607 and OX813). Patients OX1072 and OX905 are homozygous for HLA-A*02:01, whereas OX607 and OX813 have alternative HLA-A alleles. Each sequence logo consists of stacks of amino acid symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position. *Plots were produced using WebLogo 3 [216].*

3.5.3 Comparison of peptide and RNA expression

The mean number of unique peptides detected per sample was 2233, which arise from 978 genes (Figure 3.23 (a)). In many cases, multiple peptides from the same gene were detected by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Across all patients, the most commonly represented genes were metabolic enzymes (aldolase, enolase 1, GAPDH and pyruvate kinase), cytoskeletal proteins (keratin, β-tubulin, α-actin 1 and vimentin) and albumin.

There was a strong correlation between RNA expression and the detection of a HLA-bound peptide. For 81% of peptides, RNA expression for the corresponding gene was $>$ 1 RPKM (reads per kilobase of transcript, per million mapped reads) (Figure 3.23 (b)). For peptides eluted from each of the immunoprecipitates (class

I, HLA-DR and HLA-DP/-DQ), detection of peptides by mass spectrometry was associated with significantly higher mean number of RNA transcripts from the corresponding gene (Class I, p = 4.96×10^{-10}; HLA-DR, p = 1.42×10^{-7}; IVA12, p = 2.53×10^{-7}).

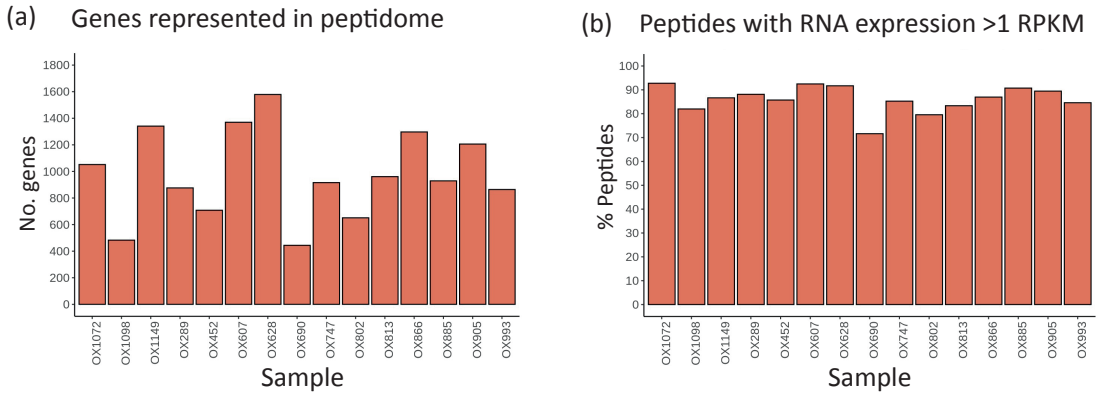


Figure 3.23: Genes represented in the immunopeptidome. For each patient sample, **(a)** the total number of genes with peptides detected by mass spectrometry, and **(b)** % peptides with RNA expression (> 1 RPKM) of the corresponding gene.
RPKM: reads per kilobase of transcript, per million mapped reads

For each patient-donor pair, the mean number of genes containing mismatched variants was 2864. From RNA-Seq of AML cells, RNA expression (> 1 RPKM) was detectable for 1598 (55.8%) of these genes. However, a peptide was only detected by LC-MS/MS for 215 (7.5%) genes. Given that only a small proportion of genes are represented in the peptidome, this suggests that the technique has limited capacity to detect mismatched peptides. The cause of this limited sensitivity is likely to be multifactorial. First, due to the limited availability of bone marrow samples from the patients studied, only between 1.5×10^6 and 5.6×10^7 cells per sample were used. Typically, immunopeptidomic studies use over 1×10^8 input cells. Granados et al used 5×10^8 EBV-transformed B-lymphoblastoid cells and detected 4468 unique peptides, which is almost double the number detected in the current study (2233 peptides) [125]. Another factor reducing the detection of peptides is that due to their physicochemical properties, some peptides are not readily detected by LC-MS/MS [240].

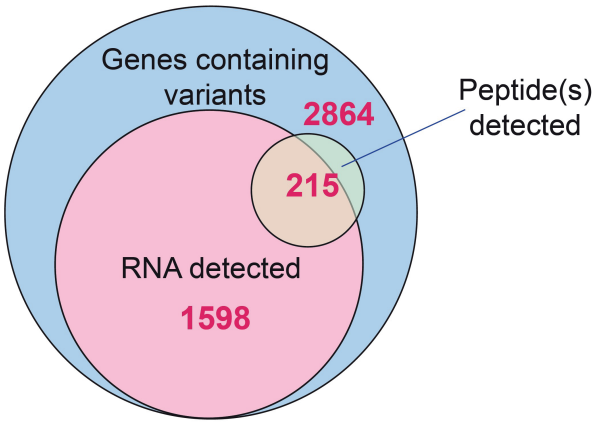


Figure 3.24: Immunopeptidomic analysis has limited sensitivity to detect patient-donor mismatched peptides. For each patient-donor pair, the mean number of genes with mismatched variants was 2864. RNA expression (> 1 RPKM) was detectable for 1598 (55.8%) of these genes. However, a peptide was only detected by LC-MS/MS for 215 (7.5%) genes. This suggests that LC-MS/MS has limited sensitivity to detect patient-donor mismatched peptides.
RPKM: reads per kilobase of transcript, per million mapped reads, LC-MS/MS: Liquid Chromatography-Tandem Mass Spectrometry

3.5.4 Comparing HLA-binding prediction to mass spectrometry detection

Patient-donor mismatched HLA-binding peptides were identified using 2 methods: *in silico* prediction and LC-MS/MS. The methods were compared by taking all peptides identified by mass spectrometry and predicting their binding to patient-specific HLA alleles. For HLA class I peptides, the proportion of correct predictions varied across patients (Figure 3.25 (a)). Overall, 67.22% peptides were predicted to bind HLA class I, of which 58.62% were predicted high affinity 'Strong' binders (Figure 3.25 (b)).

For HLA class II peptides, predictions were made from peptides eluted from HLA-DR, -DP and -DQ (NetMHCIIpan 3.2 [161]). These predictions performed less well and only 17.98% peptides were predicted to bind HLA class II, of which 12.77% were predicted high affinity binders (Figure 3.26). Class II prediction is challenging due to the variable length of peptides and their binding promiscuity across different HLA molecules [166]. Additionally, the molecules are highly polymorphic and in the case of HLA-DP and -DQ, both chains of the heterodimeric protein complexes can vary, resulting in up to 4 different combinations in a single individual. The accuracy of predictions is dependent on the availability of binding affinity training

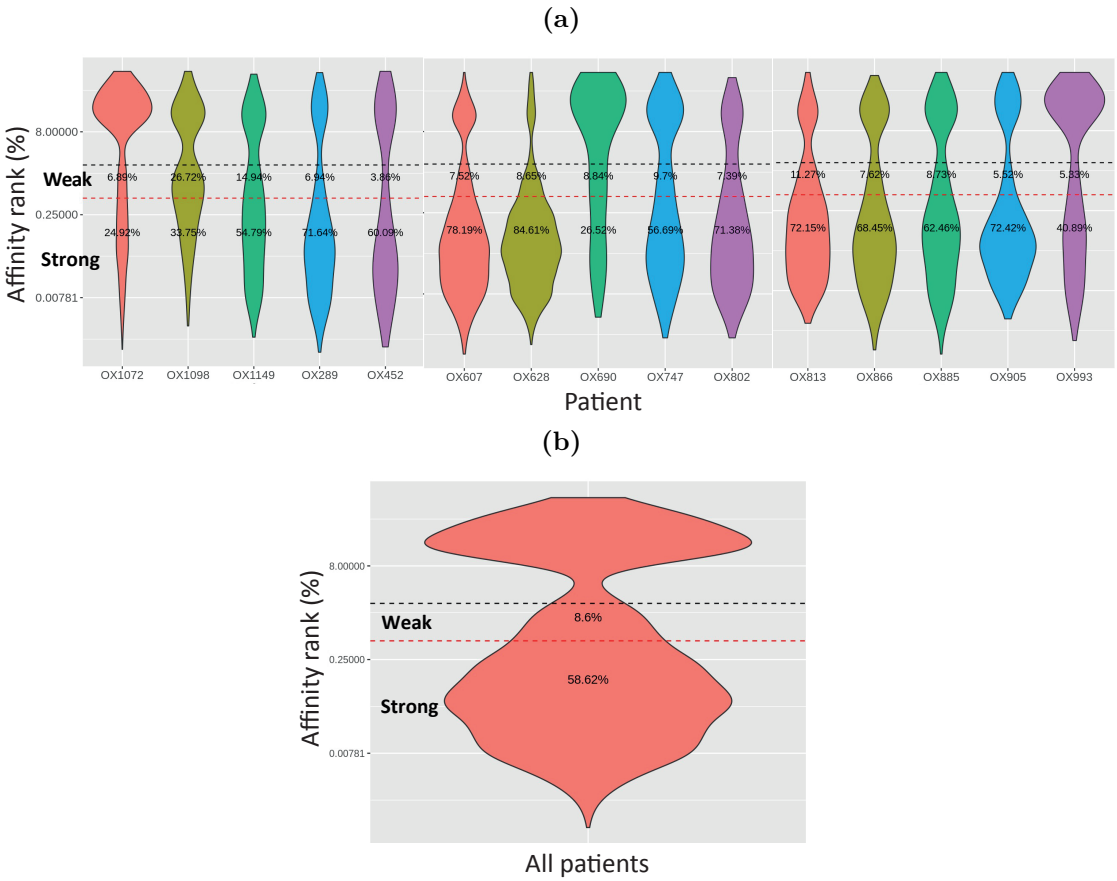


Figure 3.25: *In silico* binding prediction of HLA class I peptides detected by LC-MS/MS. (a) Violin plots representing HLA class I peptides from each patient sample detected by LC-MS/MS. The *in silico* predicted binding affinity (Affinity rank) to patient-specific HLA molecules is shown on the y-axis. The portion of each violin plot below the red dashed line represents peptides predicted to bind with high affinity ($< 0.5\%$ Affinity rank, 'Strong'). The portion of each plot between red and black dashed lines represents peptides predicted to bind with low affinity (0.5 - 2%, 'Weak') **(b)** Summary of peptides detected by LC-MS/MS across all patient samples. Overall, 67.22% peptides were predicted to bind HLA class I, of which 58.62% were predicted high affinity 'Strong' binders (below red dashed line) and 8.6% were low affinity 'Weak' binders (between red and black dashed lines).

set data that is used by artificial neural networks. For less common HLA alleles, data are sparse and therefore predictions are less accurate.

3.5.5 Candidate alloreactive peptides identified by mass spectrometry

Peptides identified by mass spectrometry that incorporate a patient-donor mismatched variant could be recognised as foreign by donor-derived T cells and therefore are candidate alloreactive peptides. Given homology in amino acid sequences

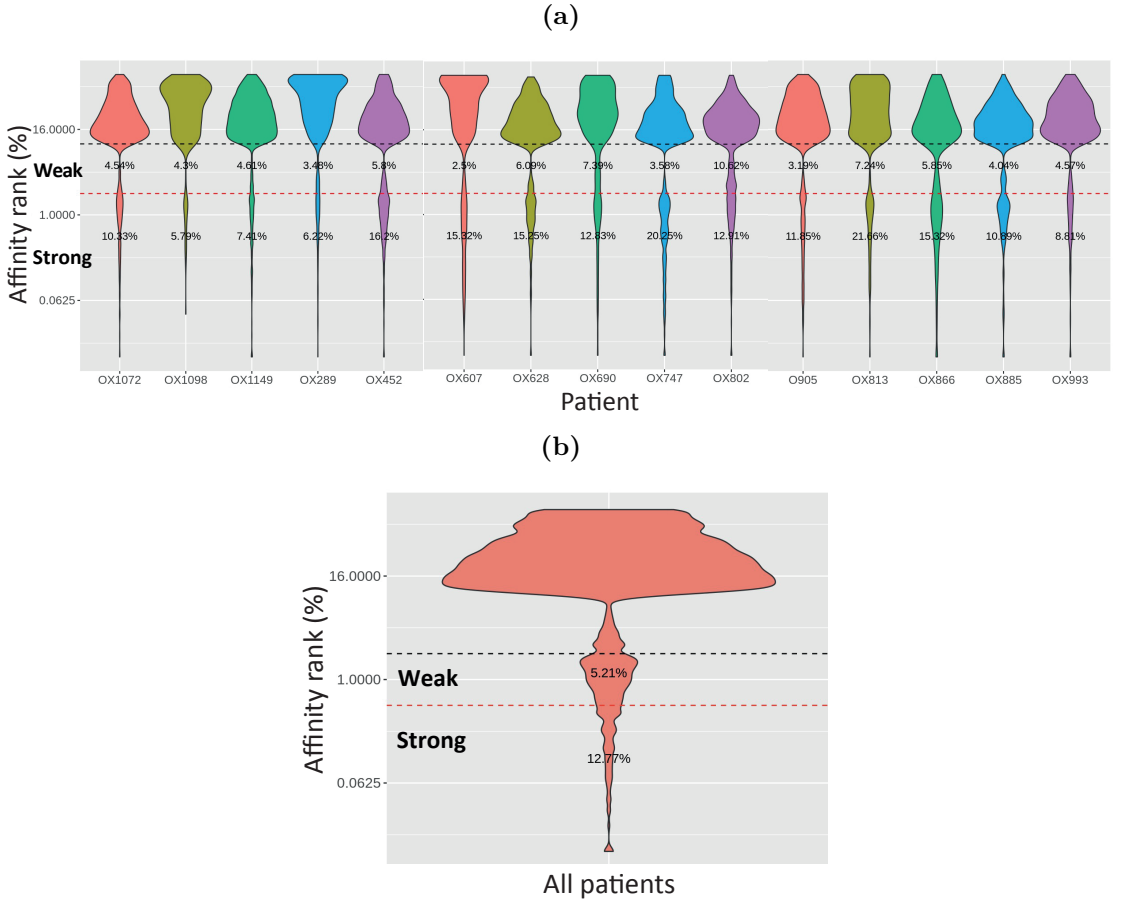


Figure 3.26: *In silico* binding prediction of HLA class II peptides detected by LC-MS/MS. (a) Violin plots for HLA class II peptides from each patient sample detected by LC-MS/MS. The *in silico* predicted binding affinity (Affinity rank) to patient-specific HLA molecules is shown on the y-axis. The portion of each violin plot below the red dashed line represents peptides predicted to bind with high affinity (< 2% Affinity rank, 'Strong'). The portion of each plot between red and black dashed lines represents peptides predicted to bind with low affinity (2 - 10%, 'Weak') **(b)** Summary of peptides detected by LC-MS/MS across all patient samples. Overall, only 17.98% peptides were predicted to bind HLA class II, of which 12.77% were predicted high affinity 'Strong' binders (below red dashed line) and 5.21% were low affinity 'Weak' binders (between red and black dashed lines).

between different genes, if the amino acid sequence is present elsewhere in the proteome, it was excluded from further analysis, because the donor is likely to express the sequence and have immunological tolerance to the peptide. Across all 15 patients studied, 31 candidate alloreactive peptides were identified. Of these, 14 were eluted only from HLA class I, 14 were eluted only from HLA class II and 3 peptides were eluted from both class I and II (Figure 3.27).

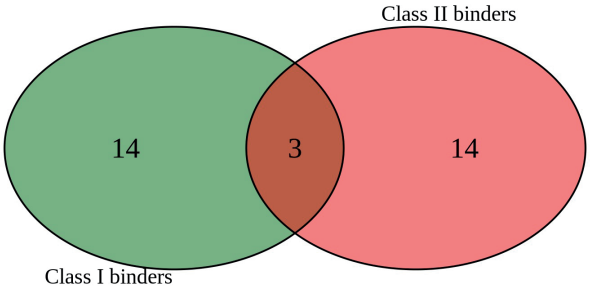


Figure 3.27: A total of 31 patient-donor mismatched peptides were detected by LC-MS/MS. Across the 15 patients studied, a total of 31 patient-donor mismatched HLA-binding peptides were detected. Of these, 3 were eluted from both HLA class I and II molecules.

3.6 Conclusions: Antigen prediction

An unbiased reverse immunology approach was used to predict alloreactive antigens in 15 patients treated with allogeneic stem and immune cell transplantation. This high-throughput method of antigen discovery starts by using whole exome sequencing data to identify differences in coding DNA sequences between patients and their respective donors. An average of 5745 differences were found with an unrelated donor transplant and 3253 with a sibling donor transplant. The majority of exonic differences between patient AML and donor germline (98.7%) were due to inherited differences between individuals rather than leukaemia-specific mutations. Overall, 28.2% of patient-donor mismatched variants were detectable in RNA-Seq reads from AML cells.

For peptides resulting from mismatched, expressed peptides, binding to patient-specific HLA alleles was predicted *in silico* using a machine learning approach.

For each patient, 145 - 674 peptides were predicted to bind HLA class I, and 139 - 696 were predicted to bind class II molecules, with significant overlap between those predicted to bind class I and II. In conjunction with *in silico* prediction, an immunopeptidomic approach was used to directly assay HLA-bound peptides on the AML cells from the patients being studied. However, this yielded an average of only 2.1 mismatched peptides per patient-donor pair.

HLA-binding predictions have limited accuracy and even if the prediction is correct, the peptide may not be able to activate donor-derived T cells from the patient following allo-SCT. Therefore, functional screening of peptides was essential to identify alloreactive antigens, as described in Chapter 4.

Chapter 4

Identification and characterisation of a GvL antigen

Contents

4.1 Approach to screening predicted GvL antigens 102
4.1.1 Screening using IFNγ ELISpot 102
4.1.2 Patient selection 102
4.1.3 Compiling a peptide library for screening 104
4.2 IFNγ ELISpot screening 105
4.2.1 *Ex vivo* screening 105
4.2.2 *In vitro* culture with peptide increases sensitivity of IFNγ ELISpot 108
4.3 Characterisation of T cell responses to the PADI4 antigen 111
4.3.1 Epitope mapping and functional titration 111
4.3.2 Antigen-specific T cells are $CD4^+$ 114
4.3.3 T cell responses are HLA-DP restricted 117
4.4 Clinical utility of the PADI4 antigen 120
4.5 PADI4 expression in human tissues 120
4.5.1 The function of PADI4 120
4.5.2 Expression in normal tissue 121
4.5.3 Expression in cancer 125
4.6 Phenotyping antigen-specific T cells using single cell RNA sequencing 128
4.6.1 Isolating antigen-specific T cells 128
4.6.2 Antigen-specific T cells are $CD4^+$ T helper 1 cells with a cytotoxic phenotype 131
4.6.3 Pathway analysis 136
4.7 Identifying antigen-specific TCR sequences 139
4.8 Antigen-specific T cells are not $CD8^+$ 142
4.9 *Ex vivo* post-transplant T cell receptor repertoires . . 145
4.10 Conclusions: PADI4 antigen identification and characterisation 147

4.1 Approach to screening predicted GvL antigens

4.1.1 Screening using IFNγ ELISpot

In individual recipients of an allogeneic stem cell transplant, it is not known whether a small number of antigens drive GvL responses or whether responses are the cumulative result of many antigen-specific responses. For each patient studied, between 177 and 871 variants resulted in peptides that were predicted to bind patient-specific HLA molecules. However, HLA-binding predictions have limited accuracy [166, 167]. Furthermore, HLA binding affinity does not equate to immunogenicity because in order to produce a T cell response, the peptide must be processed and loaded onto HLA, the HLA-peptide complex presented on the cell surface and a T cell with the cognate TCR must be present and capable of becoming activated.

IFNγ ELISpot (enzyme-linked immune absorbent spot) was used to functionally screen for antigen-specific T cells in post-transplant blood samples. This highly sensitive assay allows a large number of peptides to be screened in pools, using a limited amount of post-transplant blood (Figure 4.1). As a pilot in a single patient, all predicted HLA-binding peptides were synthesised and screened for T cell responses by IFNγ ELISpot.

4.1.2 Patient selection

Patient OX866 was chosen for detailed functional characterisation of alloreactive T cell responses due to clinical evidence of a GvL effect. The patient's AML did not respond to first-line chemotherapy treatment with daunorubicin and cytarabine. He was therefore given 2 cycles of more intensive treatment (fludarabine / cytarabine / idarubicin / granulocyte colony-stimulating factor, G-CSF) and still had residual disease (7% AML blasts in the bone marrow) after the first cycle. Although his disease was in remission at the time of allo-SCT, he relapsed 3 months later (Table A.2). Given the initial refractoriness to chemotherapy and early post-transplant relapse, his prognosis was very poor. He was treated with further chemotherapy

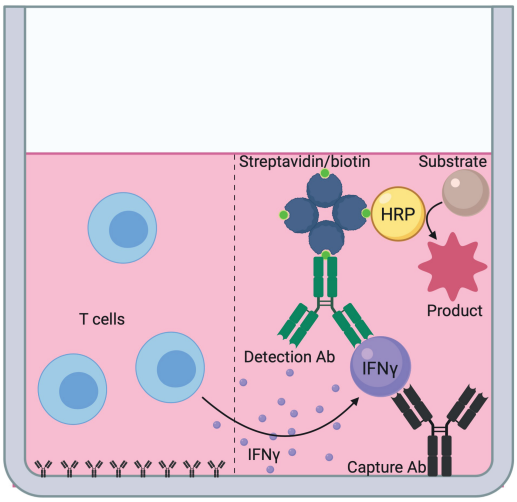


Figure 4.1: IFNγ ELISpot. T cells are incubated with a stimulus (e.g. peptide). If a HLA-presented peptide is recognised by its cognate TCR on a T cell, it can lead to T cell activation and release of IFNγ, which is bound by anti-IFNγ capture antibody coating the wells. After a 16-20 hour incubation, cells are washed away and an anti-IFNγ detection antibody is added. HRP is conjugated to the detection antibody and substrate is added. The production of an insoluble product is visible as spots on the bottom of wells, which are detected and counted using an ELISpot reader. Each spot represents IFNγ released from a single T cell and therefore by counting spots, the strength of responses can be assessed.
Ab: Antibody, HRP: Horseradish Peroxidase, IFNγ: Interferon-γ
Created with BioRender.com.

and, importantly, 3 donor lymphocyte infusions (DLIs). This treatment produced disease remission and he remains in remission more than 7 years since his transplant. Given that his post-transplant relapse occurred following several courses of intensive chemotherapy, including transplant conditioning treatment, he is very unlikely to have been curable with chemotherapy alone. Interestingly, despite evidence of an immune response from DLI directed against AML, he did not experience GvHD. This suggests that donor-derived T cells are targeting an antigen that is not present in normal healthy tissues. GvL antigens could represent targets for cancer immunotherapy. Therefore, by identifying antigens from a patient without GvHD, the chance of off-target effects may be minimised.

Other factors considered when selecting a patient to study included: (1) availability of post-transplant blood samples, (2) not receiving systemic immunosuppression

at the time of blood sampling, which could reduce the number of T cells available for assays, and (3) the presence of common HLA alleles, such as HLA-A*02:01 (Table B.3). This is because HLA-binding predictions are more robust for common alleles, so positive results from screening peptide libraries may be more likely. The clinical impact of results would also be greater, as a larger proportion of patients would have the restricting allele. Furthermore, fluorescent-labelled HLA-peptide tetramers can be more readily produced for common HLA alleles, which would allow tracking of antigen-specific T cells in post-transplant samples by flow cytometry.

4.1.3 Compiling a peptide library for screening

In many cases, different predicted HLA-binding peptides overlap the same DNA variant. This is due to different lengths of peptides and also their position at different sliding windows across the variant position. In these cases, a long peptide was produced that encompassed all overlapping peptides (Figure 4.2 (a)). Screening peptides $>$ 20 amino acids in length may result in less efficient T cell activation. Therefore, whenever the overlapping peptide was greater than 20 amino acids in length, it was split into 2 peptides, each of 20 amino acids in length (Figure 4.2 (b)).

Alloreactive responses can be directed against amino acid sequences resulting from variant alleles (Figure 3.8 (a)) or from the reference allele (Figure 3.8 (b)). Accordingly, HLA-binding peptides were predicted for either the variant or reference sequence, depending on which was present in the patient but not the donor.

For Patient OX866, 335 peptides were synthesised for screening:

- 234 genetic variants produced peptide sequence(s) that were predicted to bind at least 1 patient-specific HLA allele.

- 84 of the 234 peptides are $>$ 20 amino acids in length and therefore were split into 2 peptides.

- 14 additional peptides resulting from alternative transcripts.

- 3 mismatched peptides identified from mass spectrometry analysis.

Of all 335 peptides, only 1 resulted from a somatic variant. The remaining peptides are a result of germline differences between patient and donor.

(a)

Chromosome	Position	Gene	HLA Allele	Peptide (variant aa)
9	271638	DOCK8	HLA-B*27:05	NRYSSVEIRK
			HLA-B*27:05	NRYSSVEIR
			HLA-DRB1*08:01	ALKINRYSSVEIRKQ
			HLA-DRB1*08:01	FALKINRYSSVEIRK
			HLA-DRB1*08:01	RAFALKINRYSSVEI
			HLA-DRB1*08:01	ERRAFALKINRYSSV
			HLA-DRB1*08:01	RRAFALKINRYSSVE
			HLA-DRB1*08:01	AFALKINRYSSVEIR
			HLA-DRB1*08:01	SSVEIRKQFTLPPNL
		Overlapping sequence	ERRAFALKINRYSSVEIRKQFTLPPNL	

(b)

	Amino acid residues from long peptide (Total peptide length = n)	Sequence (variant aa)
Split peptide 1	1 to 20	ERRAFALKINRYSSVEIRKQ
Split peptide 2	(n-19) to n	KINRYSSVEIRKQFTLPPNL

Figure 4.2: Producing overlapping peptides for screening. (a) For one example locus in the *DOCK8* gene, all of the predicted HLA-binding peptides across the patient-donor mismatched variant are listed (variant amino acid in red). The composite overlapping sequence is highlighted in yellow. **(b)** As the overlapping sequence is > 20 amino acids in length, it was split into 2 peptides, each 20 amino acids in length.
aa: Amino acid

4.2 IFNγ ELISpot screening

4.2.1 *Ex vivo* screening

In an initial screen, PBMCs were used *ex vivo*, without cell culture and expansion. Therefore, T cell responses to peptides tested using this approach represent the frequency of T cells in the original blood sample. Peptides were screened in pools of up to 20 peptides using PBMCs from a sample taken 5 years post-transplant (Figure 4.3). Positive controls were: (1) phytohaemagglutinin (PHA), which non-specifically activates T cells, and (2) a pool of 23 peptides derived from CMV, EBV and influenza (CEF; Mabtech) that elicits HLA class I responses in almost 90% of Caucasions.

No responses were detected to *ex vivo* testing. The absence of responses may indicate that there are genuinely no antigen-specific T cells in the sample.

Alternatively, responses may not have been detected because the frequency of antigen-specific T cells is below the limit of detection of the assay. IFNγ ELISpot quantifies T cells capable of immediate IFNγ secretion upon antigen stimulation, which are likely to be mainly effector memory T cells. *In vitro* culture with peptide may expand antigen-specific central memory T cells, which can then be detected by ELISpot. This increases the sensitivity of the assay, particularly if the patient's T cells have not recently been exposed to antigen [241].

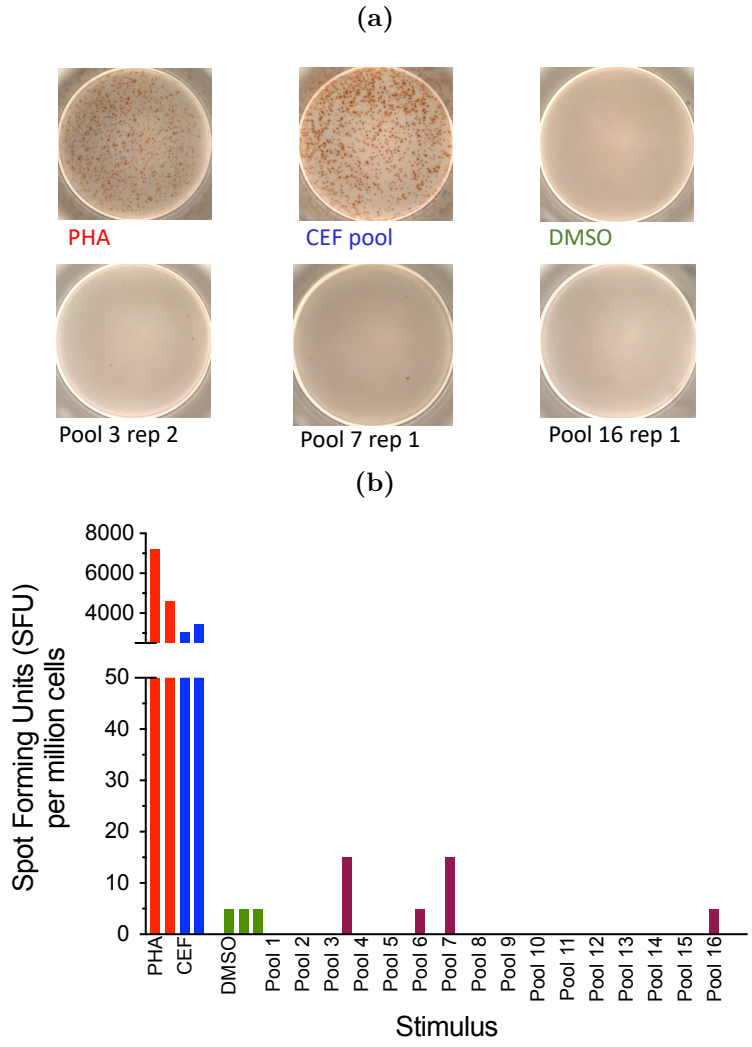


Figure 4.3: *Ex vivo* screening of peptide pools. Pooled screening by *ex vivo* IFNγ ELISpot, using blood taken 5 years post-transplant. **(a)** ELISpot wells following treatment with PHA (positive control), CMV/EBV/Influenza (CEF) peptides (positive control), DMSO (negative control) and Peptide pools 3, 7 and 16. **(b)** Summary of ELISpot results. Each peptide pool and positive control was tested in duplicate. DMSO negative controls were tested with 4 replicates.
PHA: Phytohaemagglutinin, CEF: Cytomegalovirus / Epstein-Barr virus / Influenza, DMSO: Dimethyl sulfoxide

4.2.2 *In vitro* culture with peptide increases sensitivity of IFNγ ELISpot

Using the same post-transplant sample as for *ex vivo* IFNγ ELISpot, cells were cultured for 12 days with peptide pools, IL-2 and IL-7 (Figure 4.4). Cells were grown in separate cultures with up to 80 peptides (2 μg/ml each peptide). The total number of peptides in each culture was limited by the concentration of DMSO, which was used to dissolve peptides, and was maintained below 0.5% to minimise its negative effect on T cell proliferation and activation [242]. After 12 days of culture, cells were washed to remove peptide and cytokines and rested overnight.

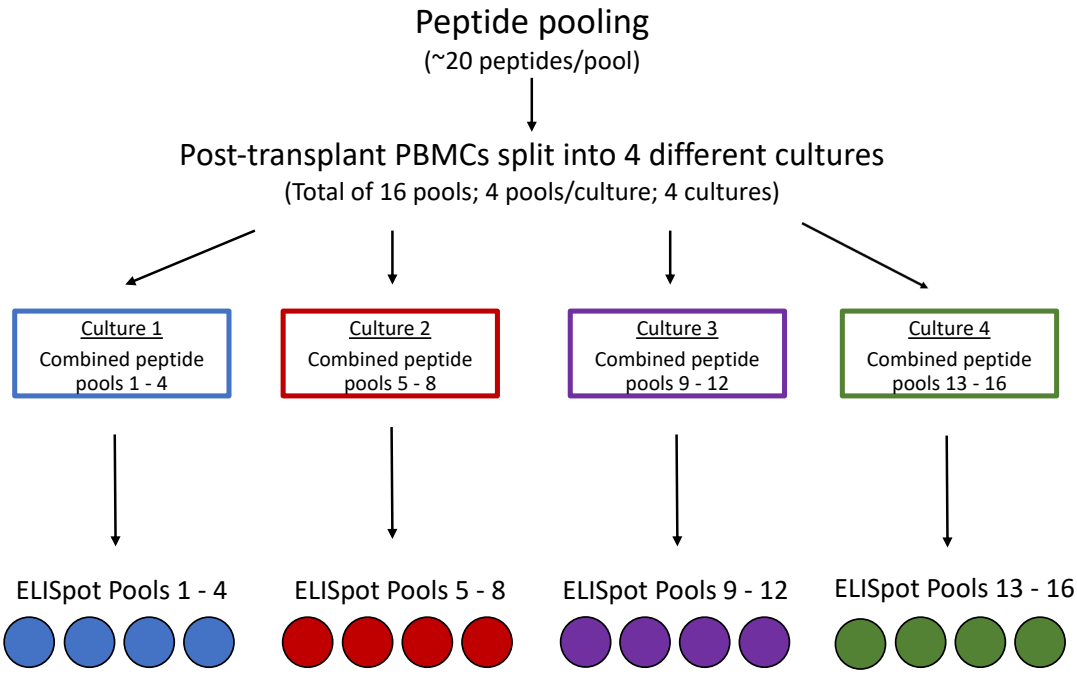
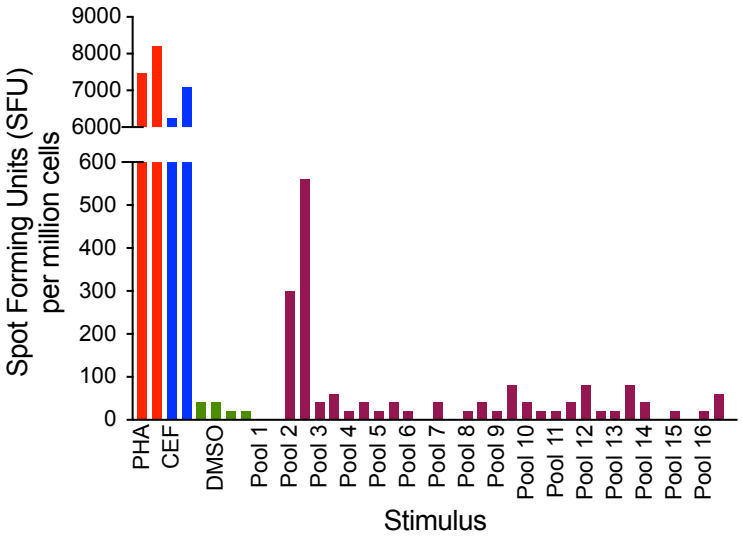


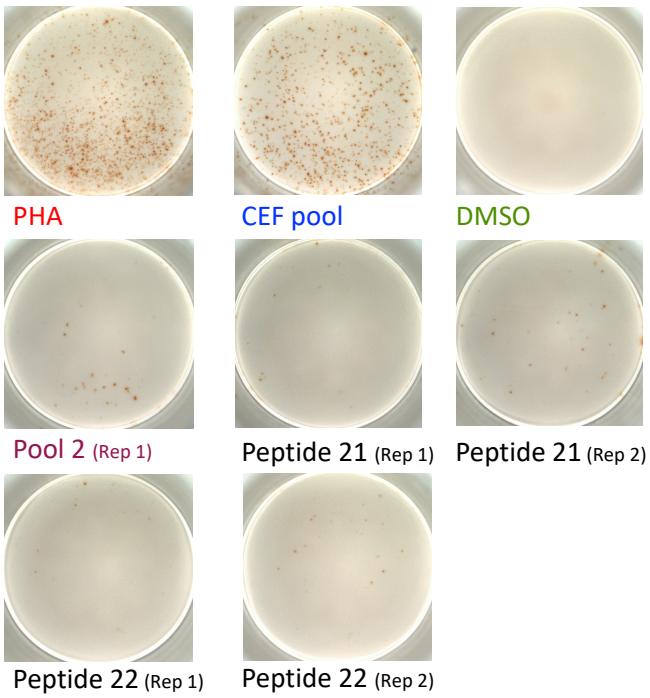
Figure 4.4: Approach to screening predicted GvL peptides for T cell responses. Post-transplant PBMCs were cultured for 12 days with up to 80 peptides, each 2 μg/ml, in the presence of IL-2 and IL-7. After culture, cells were washed and rested overnight. The component peptide pools of each culture were then tested by IFNγ ELISpot.

IFNγ ELISpot of pools was then performed, using the same peptides as were present in their culture. This demonstrated responses to Pool 2 across technical replicates (Figure 4.5 (a)). By testing the individual component peptides, responses were found to be driven by peptides 21 and 22, which overlie the same genetic variant in the *PADI4* (peptidyl arginine deiminase type IV) gene (Figure 4.5 (b) and (c)).

(a)


Spot Forming Units (SFU)
per million cells
9000
8000
7000
6000
600
500
400
300
200
100
0
PHA
CEF
DMSO
Pool 1
Pool 2
Pool 3
Pool 4
Pool 5
Pool 6
Pool 7
Pool 8
Pool 9
Pool 10
Pool 11
Pool 12
Pool 13
Pool 14
Pool 15
Pool 16
Stimulus

(b)


PHA
CEF pool
DMSO
Pool 2 (Rep 1)
Peptide 21 (Rep 1)
Peptide 21 (Rep 2)
Peptide 22 (Rep 1)
Peptide 22 (Rep 2)

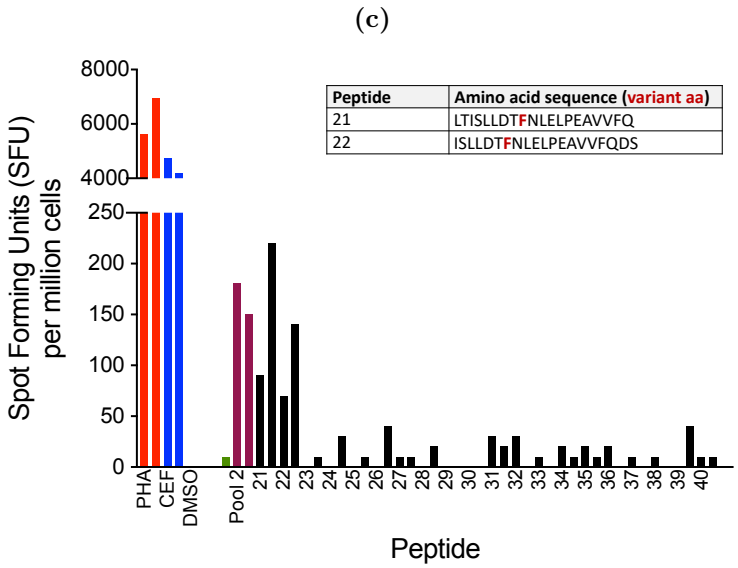


Figure 4.5: Cultured IFNγ ELISpot screening of peptide pools, followed by deconvolution of positive Pool 2. **(a)** Cultured IFNγ ELISpot screening of 16 peptide pools demonstrated responses to Pool 2 (n=2). **(b)** Component peptides from Pool 2 were individually tested. ELISpot wells are shown following treatment with PHA (positive control), CMV/EBV/Influenza (CEF) peptides (positive control), DMSO (negative control), Pool 2 and peptides 21 and 22. **(c)** Testing individual peptides demonstrated responses to peptides 21 and 22 (n=2). The amino acid sequences of these overlapping peptides is shown inset, with the variant amino acid highlighted in red.

T cells recognising the antigen were still detectable 6 years post-transplant (Figure 4.6).

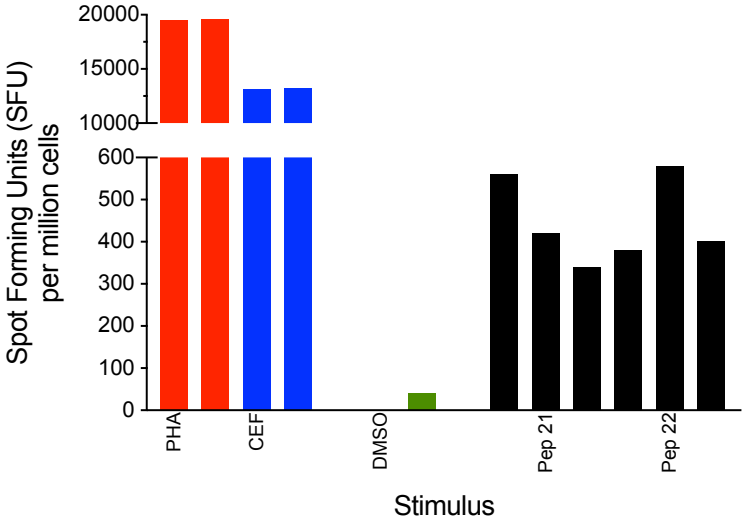


Figure 4.6: OX866 6 years post-transplant. Responses to PADI4 peptides 21 and 22 by cultured IFNγ ELISpot were still detectable using blood taken from the patient 6 years post-transplant (n=3).

4.3 Characterisation of T cell responses to the PADI4 antigen

4.3.1 Epitope mapping and functional titration

The reference PADI4 amino acid sequence will be expressed in the donor and therefore they are likely to have undergone immunological tolerance and negative selection of T cells responsive to the epitope. Indeed, when the reference peptide sequence was tested, which differs from the variant sequence by just 1 amino acid substitution, there were no detectable responses by IFNγ ELISpot (Figure 4.7).

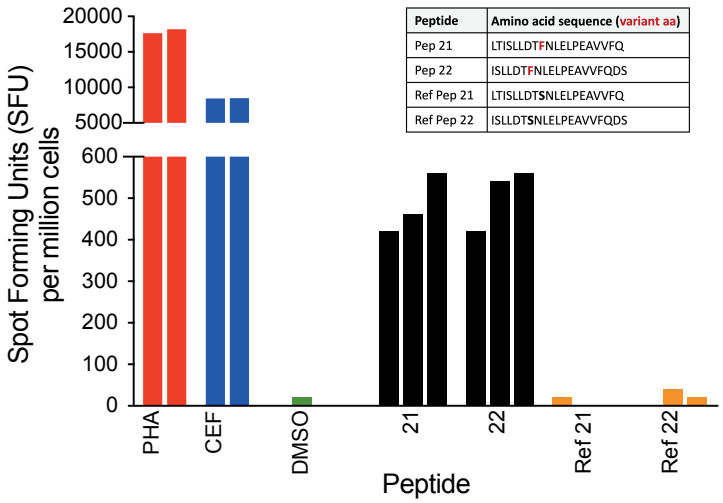


Figure 4.7: Absence of T cell responses to the PADI4 reference peptide. IFNγ ELISpot using post-transplant PBMCs from Patient OX866 to test the reference (donor) amino acid sequence showed an absence of response (yellow bars), whereas patient-donor mismatched peptides 21 and 22 elicited T cell responses (black bars) (n=3).
Ref: Reference

HLA class I molecules bind peptides 8-11 amino acids in length, whereas class II molecules bind a wider range of peptide lengths, usually between 13 and 25 amino acids [236]. Peptides 21 and 22 are each 20 amino acids in length. It was not known which part of the peptide is required for binding to HLA and eliciting T cell responses upon engagement of its cognate TCR. Given that the 2 amino acids at the N-terminus of peptide 21, and the 2 amino acids at the

C-terminus of peptide 22 are not present in the other overlapping peptide, these are likely to be dispensable for binding.

In order to map the minimal optimal peptide for binding and eliciting T cell responses, peptides of different lengths were functionally tested in different windows across the variant position using IFNγ ELISpot (Table 4.1). Peptides of length 15 and 16 amino acids elicited the strongest responses from cultured T cells (Figure 4.8). This length of peptides is strongly suggestive of a HLA class II-restricted response. The peptide position along the overlapping peptide suggests that peptides towards the C-terminal end are particularly important.

Overlap peptide	**ISLLDTFNLELPEAVVFQ (variant aa)**
20 amino acids	
Peptide 21	LTISLLDTFNLELPEAVVFQ
Peptide 22	ISLLDTFNLELPEAVVFQDS
18 amino acids	
18aa-1	ISLLDTFNLELPEAVVFQ
17 amino acids	
17aa-1	ISLLDTFNLELPEAVVF
17aa-2	SLLDTFNLELPEAVVFQ
16 amino acids	
16aa-1	ISLLDTFNLELPEAVV
16aa-2	SLLDTFNLELPEAVVF
16aa-3	LLDTFNLELPEAVVFQ
15 amino acids	
15aa-1	ISLLDTFNLELPEAV
15aa-2	SLLDTFNLELPEAVV
15aa-3	LLDTFNLELPEAVVF
15aa-4	LDTFNLELPEAVVFQ

14 amino acids	
14aa-1	ISLLDTFNLELPEA
14aa-2	SLLDTFNLELPEAV
14aa-3	LLDTFNLELPEAVV
14aa-4	LDTFNLELPEAVVF
14aa-5	DTFNLELPEAVVFQ
13 amino acids	
13aa-1	ISLLDTFNLELPE
13aa-2	SLLDTFNLELPEA
13aa-3	LLDTFNLELPEAV
13aa-4	LDTFNLELPEAVV
13aa-5	DTFNLELPEAVVF
13aa-6	TFNLELPEAVVFQ
12 amino acids	
12aa-1	ISLLDTFNLELP
12aa-2	SLLDTFNLELPE
12aa-3	LLDTFNLELPEA
12aa-4	LDTFNLELPEAV
12aa-5	DTFNLELPEAVV
12aa-6	TFNLELPEAVVF
12aa-7	FNLELPEAVVFQ

Table 4.1: Peptides used for epitope mapping. Peptides of varying lengths at different sliding windows across the variant amino acid (in red) were tested in Figure 4.8.

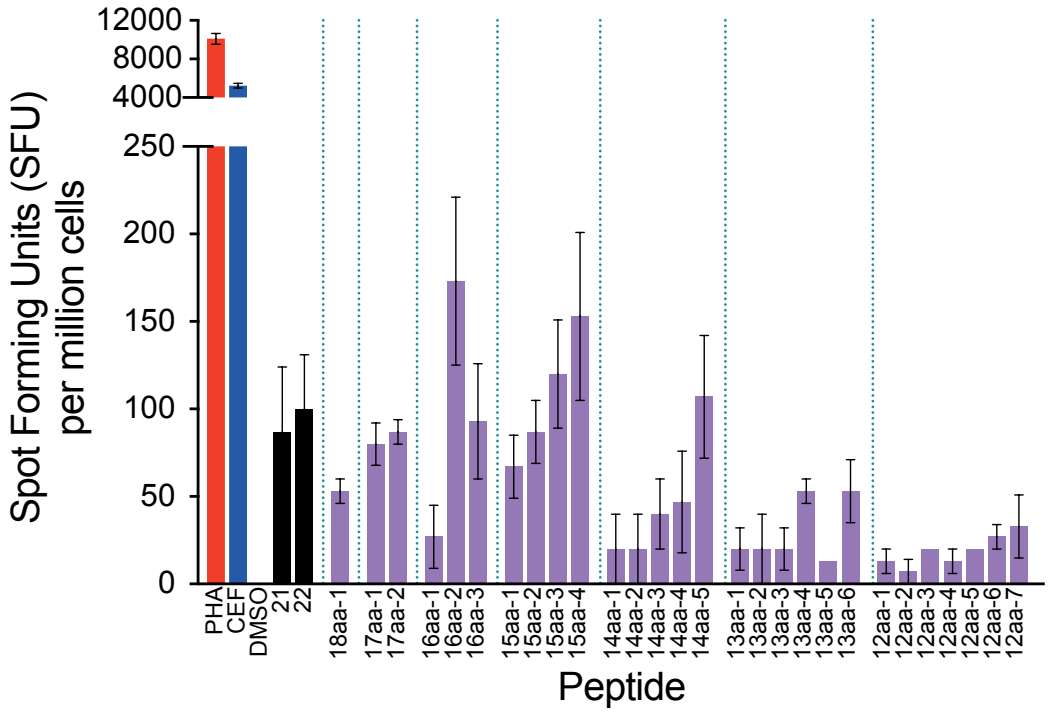


Figure 4.8: PADI4 antigen epitope mapping. Peptides of varying lengths at different sliding windows across the variant amino acid (Table 4.1) were tested by IFNγ ELISpot. Data shown are mean +/- SEM of 3 replicates.
SEM: Standard error of the mean

Functional titration of the 2 optimally mapped peptides showed a dose-dependent response, with saturation of the response between 1 and 10 μg/ml peptide. A concentration of 2 μg/ml was used in original screening, indicating that this is sufficient to stimulate near-maximal responses. As expected, there were no responses with the reference peptide sequences (Ref Peptides 21 + 22) (Figure 4.9) .

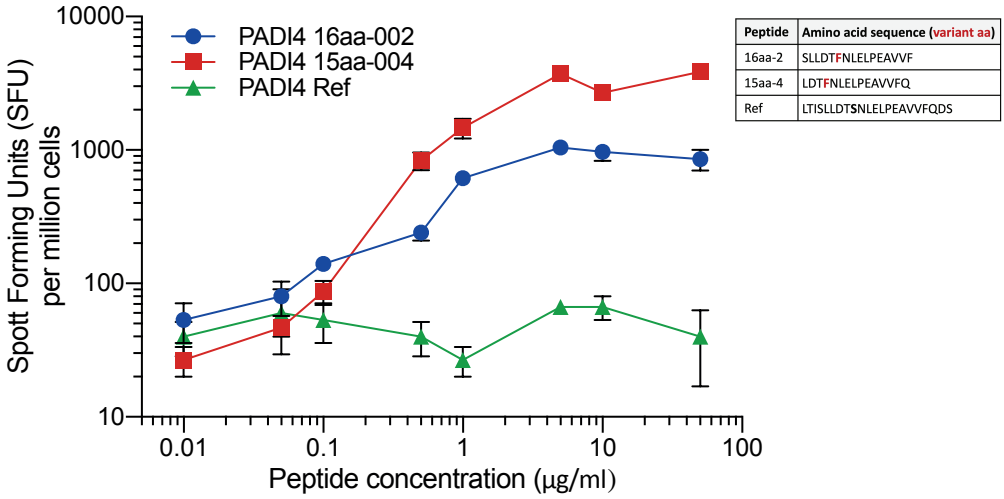


Figure 4.9: Functional titration of peptide responses. Functional responses to peptide were assayed by IFNγ ELISpot using a range of peptide concentrations. The 2 optimally mapped peptides, 16aa-2 and 15aa-004 (Figure 4.8) were used. Reference peptide sequences for peptides 21 and 22 were pooled as a negative control. Data shown are mean +/- SEM of 3 replicates.
SEM: Standard error of the mean

4.3.2 Antigen-specific T cells are CD4$^+$

To assess the phenotype of T cells that are responding to PADI4 peptide, flow cytometry with intracellular cytokine staining was performed following stimulation with peptide. Prior to the assay, PBMCs were cultured with peptide to enrich for antigen-specific T cells. CD4$^+$ cells, but not CD8$^+$ cells, were responding to peptide, as demonstrated by an increase in expression of intracellular cytokines (IFNγ and TNFα) and CD107a, a marker of T cell degranulation (Figure 4.10). This indicates a HLA class II-restricted response, which is consistent with the finding that longer peptides, of length 15-16 amino acids, elicit the strongest responses by IFNγ ELISpot. The surface markers of T cell activation, CD137 and CD25, were not increased above the background level in untreated control cells. High levels of CD25 (IL2RA, IL-2 receptor-α) were found across all conditions and likely represents upregulation in response to IL-2 during *in vitro* culture, which has previously been described [243].

The cytokine profile of T cells responding to PADI4 peptides is distinct from the

profile seen with the other stimuli tested. Almost half of PADI4 peptide-responsive $CD4^+$ cells produce only IFNγ without TNFα and IL-2 (47.81%). Most of the remaining responsive cells (39.84%) express both IFNγ and TNFα. Very little IL-2 is expressed (Figure 4.11). In contrast, the majority of T cells responding to PHA expressed single cytokines. $CD4^+$ cells expressed either IFNγ or TNFα, and $CD8^+$ cells expressed IFNγ. CEF stimulated $CD8^+$ cells that dually express IFNγ and TNFα. Expression of multiple cytokines, as seen in 40.06% of $CD4^+$ cells responding to PADI4 peptide, has been associated with polyfunctional T cells, which result in productive immune responses to pathogens and vaccines [244, 245]. Although this could indicate improved function of PADI4 antigen-specific T cells, it is not known whether T cells expressing multiple cytokines have improved ability to kill malignant cells, such as seen in a GvL response following allo-SCT.

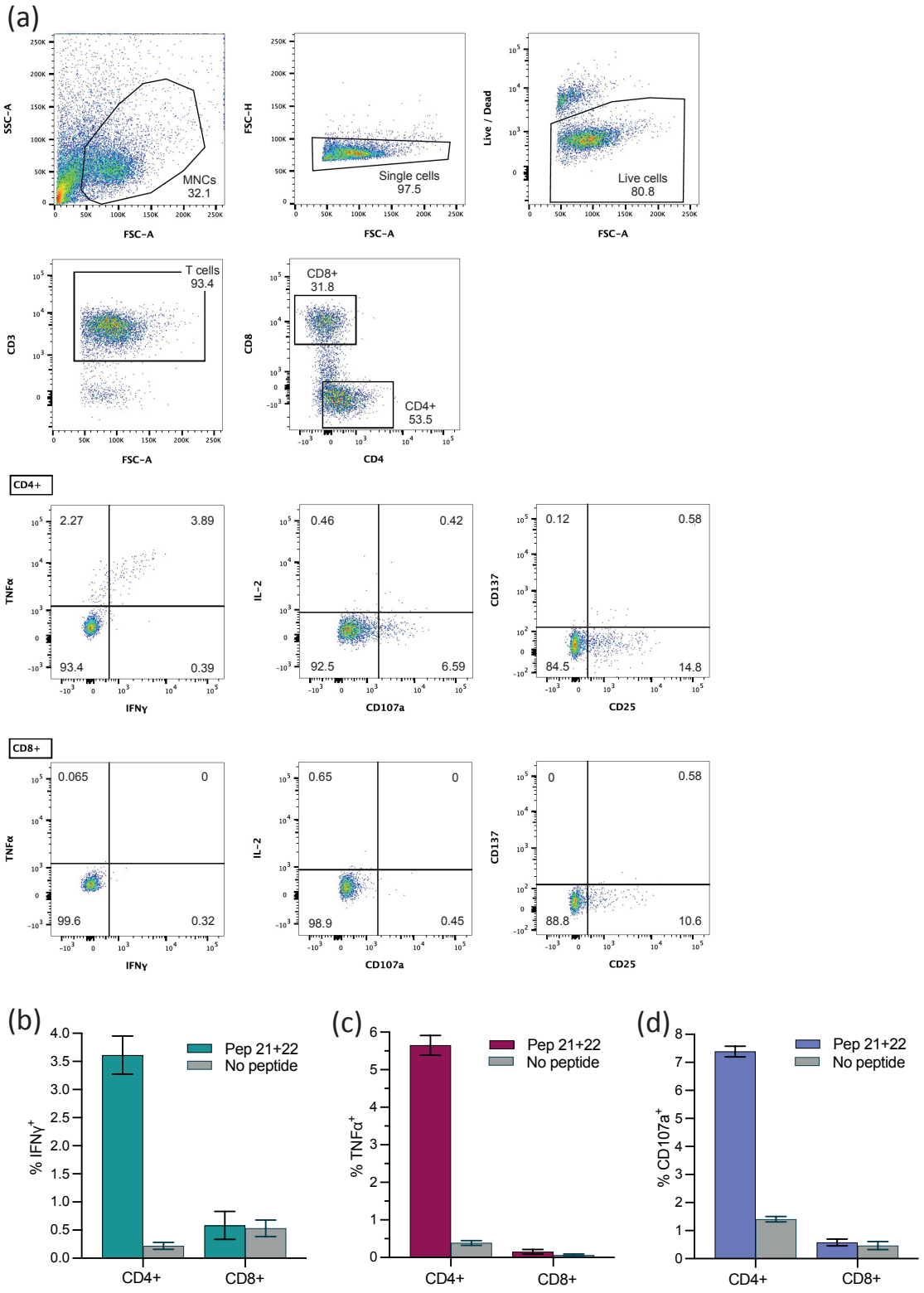


Figure 4.10: Immunophenotyping of PADI4 T cell responses. **(a)** Immunophenotyping of cultured post-transplant PBMCs using intracellular cytokine staining. Live $CD3^+CD4^+$ and $CD3^+CD8^+$ cells were gated and activation markers were assessed for each cell type. Numbers on the plots are % of the parent population. Expression of **(b)** IFNγ, **(c)** TNFα and **(d)** CD107a in $CD4^+$ and $CD8^+$ T cells in response to peptide. For each graph, the coloured bars indicate % T cells expressing the activation marker in response to peptide. Grey bars indicate cells expressing the marker in the absence of peptide. Data shown are mean +/- SEM of 3 replicates.

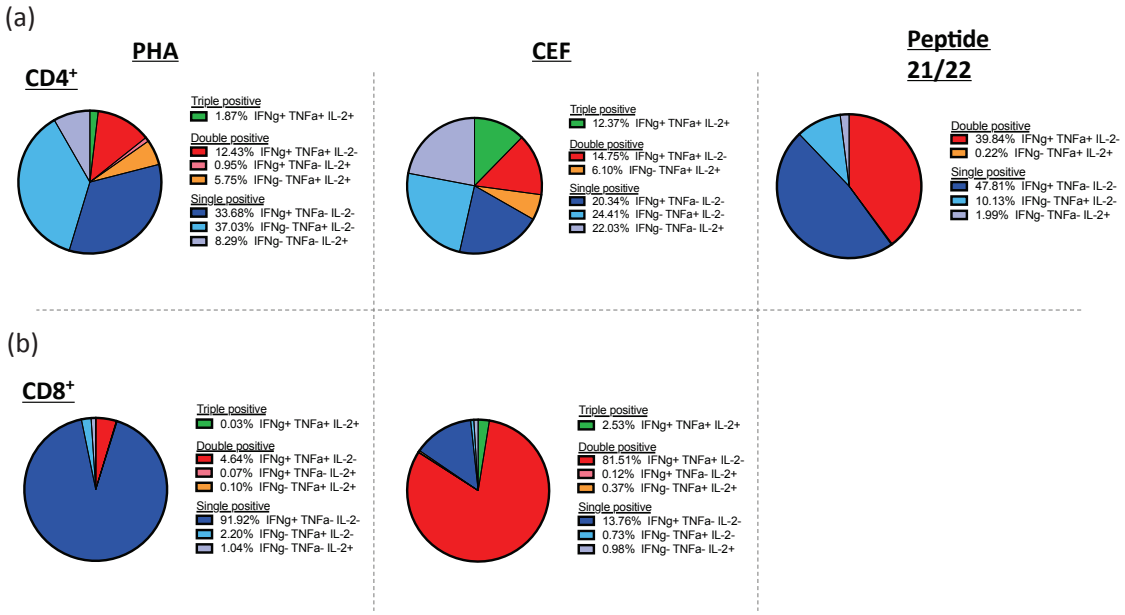


Figure 4.11: Cytokine responses. Pie charts showing the proportion of **(a)** $CD4^+$ T cells (top row), and **(b)** $CD8^+$ T cells (bottom row) expressing cytokines in response to stimuli: PHA, CEF peptides and peptides 21/22.
PHA: Phytohaemagglutinin, CEF: Cytomegalovirus / Epstein-Barr virus / Influenza

4.3.3 T cell responses are HLA-DP restricted

The HLA molecule presenting the PADI4 peptide was assessed using IFNγ ELISpot with HLA-blocking antibodies. T cell responses were abrogated using a pan-HLA class II-blocking antibody (Tü39). This is consistent with the immunophenotyping data, which indicated $CD4^+$ cells were responding. Blockade of single HLA molecules (HLA-DR, -DQ or -DP) revealed a HLA-DP restricted response. Overlapping peptides 21 and 22 each showed an identical pattern of response with HLA blockade (Figure 4.12).

In parallel with HLA-blocking antibody assays, HLA restriction was assayed using an antigen-presenting cell line that expresses single HLA alleles. The B lymphoblastoid cell line, .221, lacks HLA class I expression and when transfected with single HLA alleles, allows assessment of HLA restriction of peptide responses. .221 cells were selected that expressed single HLA class I alleles found in Patient OX866. Post-transplant PBMCs from Patient OX866 were cultured with peptide for 12 days, as performed for the other IFNγ ELISpot assays. .221 cells were then

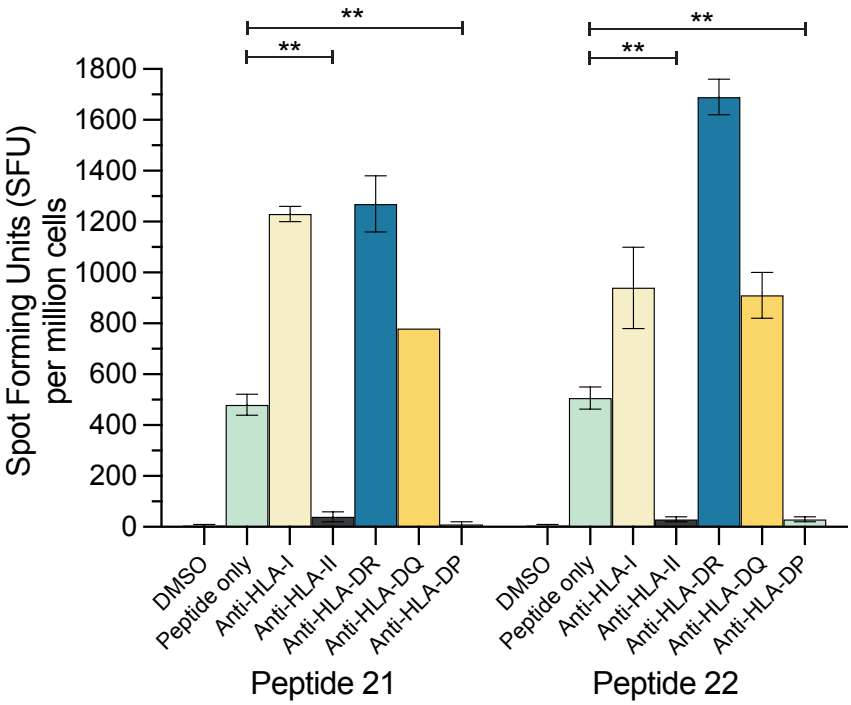


Figure 4.12: Identifying HLA restriction using HLA-blocking antibodies. T cell responses to peptides 21 and 22 were separately assessed by IFNγ ELISpot using anti-HLA antibodies to block responses mediated by particular HLA molecules. HLA molecules were blocked by antibodies: Pan-class I (W6/32), Pan-class II (Tü39), HLA-DR (L243), HLA-DQ (Tü169) and HLA-DP (B7/21). Data shown are mean +/- SEM of 2 replicates. Statistical significance was tested with two-tailed Student's t-test. **p < 0.01. SEM: Standard error of the mean

pulsed with peptide and incubated with the T cell-enriched culture. T cell activation was assessed by IFNγ ELISpot (Figure 4.13 (a)). As expected, in the absence of a transfected HLA molecule, there was no activation. Nor was activation observed without peptide pulsing. When .221 cells expressing HLA-A*02:01 were pulsed with peptide, a small increase in activation was seen with peptide 21 but not peptide 22. Despite this result, it is unlikely to represent a true positive response. Overlapping peptides 21 and 22 are likely to share the same epitope and therefore a response would also be expected with peptide 22. Furthermore, the data from HLA-blocking experiments, immunophenotyping and epitope mapping all indicate HLA class II-mediated activation of $CD4^+$ T cells in response to the PADI4 peptides and failed to show any class I-mediated $CD8^+$ responses (Figures 4.12, 4.10 and 4.8).

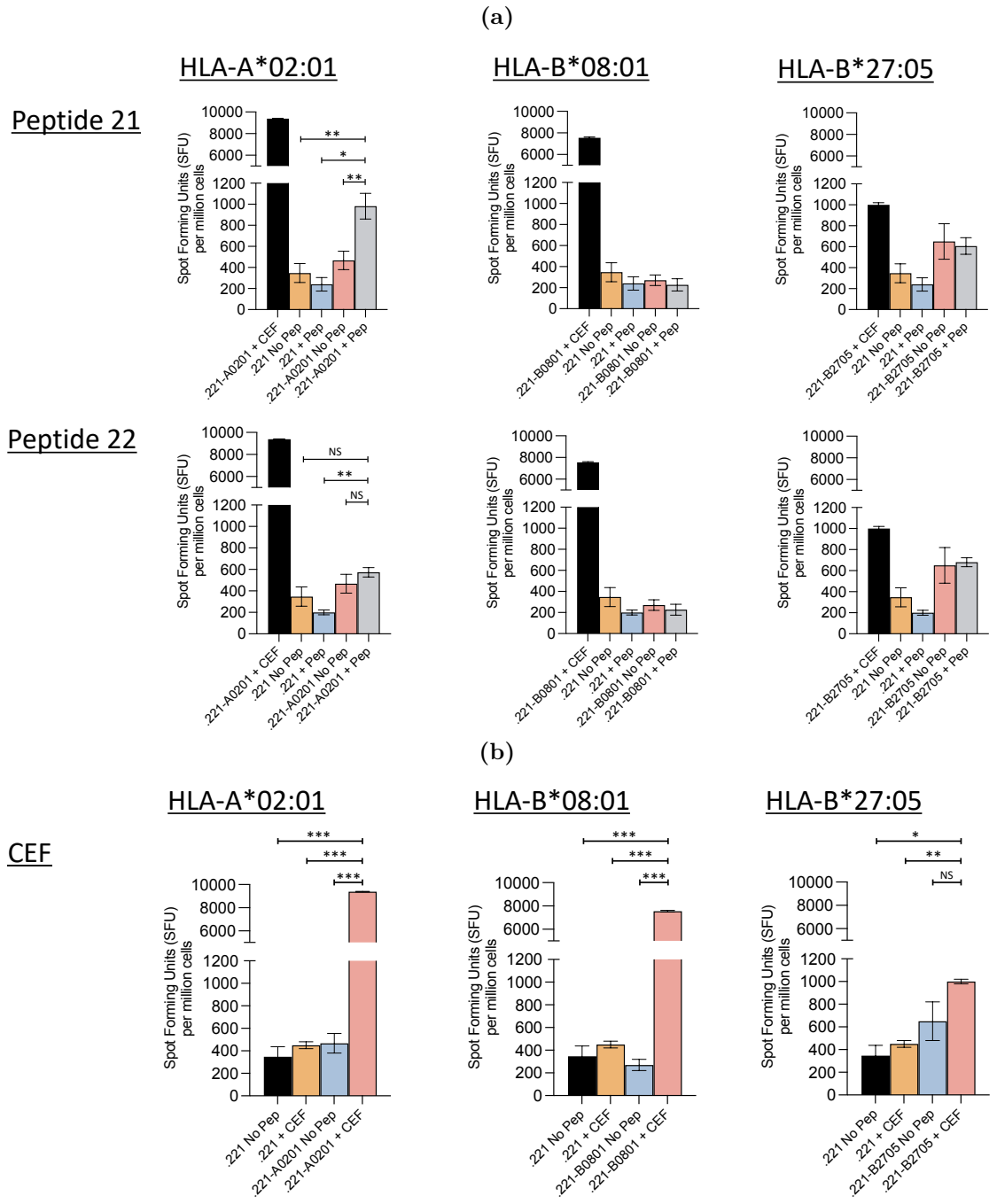


Figure 4.13: Investigating HLA class I-restricted responses using single HLA-transfected .221 cells. (a) .221 cells expressing single HLA molecules present in Patient OX866 (HLA-A*02:01, HLA-B*08:01, HLA-B*27:05) were pulsed with either peptide 21 (row 1) or 22 (row 2). PBMCs were initially cultured with peptide for 12 days to enrich for antigen-specific T cells, before incubation with the .221 cells. Responses were detected using IFNγ ELISpot. For each peptide and .221 transfectant, responses are shown to (1) CEF peptides (positive control, black bars), (2) .221 without HLA expression or peptide pulse (orange bars), (3) .221 without HLA expression but with peptide pulse (blue bars), (4) .221 with HLA expression but no peptide pulse (red bars), and (5) .221 with HLA expression and peptide pulse (grey bars). **(b)** As a positive control, responses to CEF peptides were tested in a similar manner. CEF peptides are known to produce HLA class I-restricted responses from T cells derived from almost 90% of Caucasians. Data shown are mean +/- SEM of 3 replicates. Statistical significance was tested with two-tailed Student's t-test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, NS: not significant.

4.4 Clinical utility of the PADI4 antigen

If the HLA-peptide complex were used as a therapeutic target, the proportion of patients that could be treated by targeting the HLA-presented PADI4 antigen is dependent on the proportion of patients that have the combination of the *PADI4* single nucleotide polymorphism (SNP) and the restricting HLA allele. The SNP that leads to the PADI4 antigen has a population allele frequency of 4.27%. This figure is derived from the gnomAD database, which contains mixed ethnic populations and 54.5% of entries are European [233]. Patient OX866 is homozygous at the HLA-DP locus for the haplotype DPA1*01:03-DPB1*04:01, which has a frequency of 41.7% in an English population [246]. Accounting for both heterozygosity and homozygosity of the *PADI4* SNP and HLA-DP haplotype, 8.4% of people have at least 1 PADI4 variant allele and 66% have at least 1 of the restricting HLA-DP alleles. Therefore, 5.5% of people have the combination of PADI4 SNP and HLA-DP haplotype and so are estimated to be treatable.

4.5 PADI4 expression in human tissues

4.5.1 The function of PADI4

PADI4 is a member of a gene family that encodes enzymes responsible for post-translational modification of proteins by converting arginine to citrulline. Citrullination, or deimination, counteracts arginine methylation by protein arginine methyltransferases (PRMTs). PADI4 can citrullinate arginine residues on histone tails of H3 and H4 and this epigenetic mark has been implicated in both activating and repressive effects on gene expression [247, 248]. Arginine methylation by PRMT4 (coactivator-associated arginine methyltransferase 1) at position H3R17 activates gene expression. By citrullinating this arginine residue, PADI4 counteracts this effect and thereby has a repressive effect on gene expression [249].

PADI4 has also been found to activate gene expression by distinct mechanisms. Citrullination of histone H4R3 inhibits binding of the repressor protein, heterochro-

matin protein 1 α (HP1α) [250]. A further activatory mechanism was found in a study using human breast cancer cell lines. Direct citrullination of the transcription factor, Elk-1, facilitates subsequent phosphorylation of Elk-1 by ERK. Elk-1 then activates transcription of its well-defined target, the oncogene c-Fos [251]. Therefore, the impact of PADI4 on gene expression is context-specific.

PADI4 is highly expressed in haematopoietic tissues, which suggests an important role in the regulation of haematopoiesis. Here, it has been found to regulate the proliferation of bone marrow multipotent progenitor cells by controlling c-myc expression [252].

PADI4 has been implicated in rheumatoid arthritis, an autoimmune disease characterised by inflammation of synovial tissue in joints and subsequent joint destruction. Antibodies recognising citrullinated peptides are highly specific for rheumatoid arthritis and often predate the onset of disease, suggesting a role in its pathogenesis [253]. Multiple genetic linkage studies performed on East Asian populations have shown that *PADI4* is a susceptibility locus for rheumatoid arthritis [254–256]. People with a haplotype associated with increased risk are also more likely to have detectable autoantibodies to citrullinated peptides [257]. However, more recent studies using populations of European descent have shown mixed results and a large study of 5591 rheumatoid arthritis patients in the UK showed no association with *PADI4* genotype [253, 258, 259].

4.5.2 Expression in normal tissue

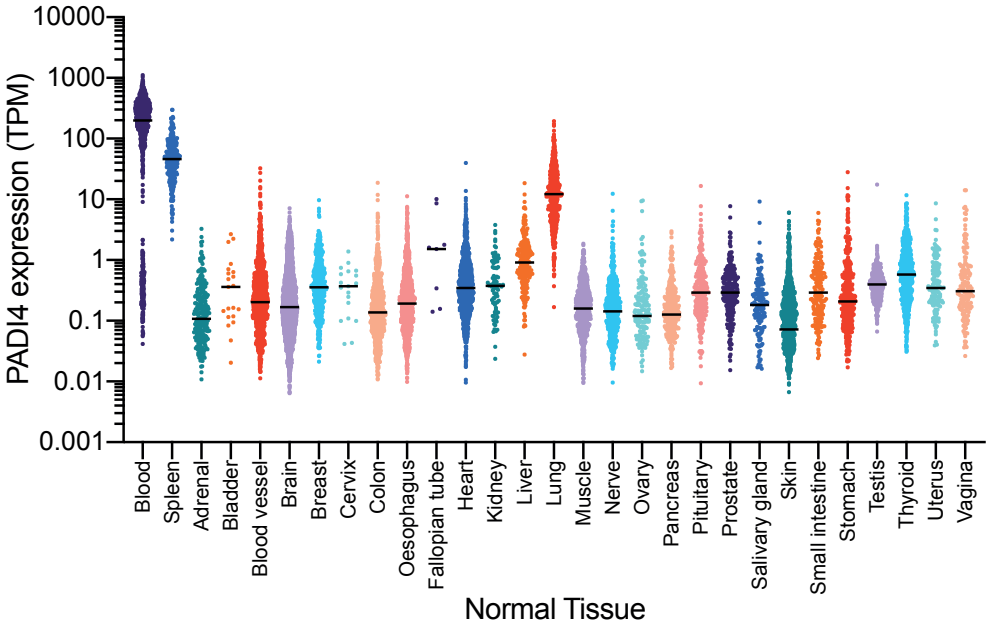
Expression of *PADI4* is principally in the haematopoietic tissues of blood, bone marrow and spleen (Figure 4.14), but was also found in the lung in the GTEx (Genotype-Tissue Expression [260]) dataset (Figure 4.14 (a)). PADI4 is found in alveolar type II cells in the lung (Human Protein Atlas [234]), which are responsible for synthesis and secretion of surfactant proteins and also have a regenerative role following lung injury [261].

Given that Patient OX866 had good clinical evidence of a GvL effect without experiencing GvHD either after allo-SCT or DLIs (Section 4.1.2), it may be that

a GvL response was generated against an antigen that is not expressed in non-haematopoietic tissues. Following allo-SCT, haematopoietic cells are donor-derived. Therefore, expression of the PADI4 antigen will not be found in the engrafted haematopoietic cells and therefore T cell clones targeting the antigen will not recognise the healthy post-transplant blood cells.

PADI4 is expressed throughout the haematopoietic system, with a predilection for more mature myeloid cells such as neutrophils, eosinophils, basophils and monocytes (Figure 4.15).

(a)



(b)

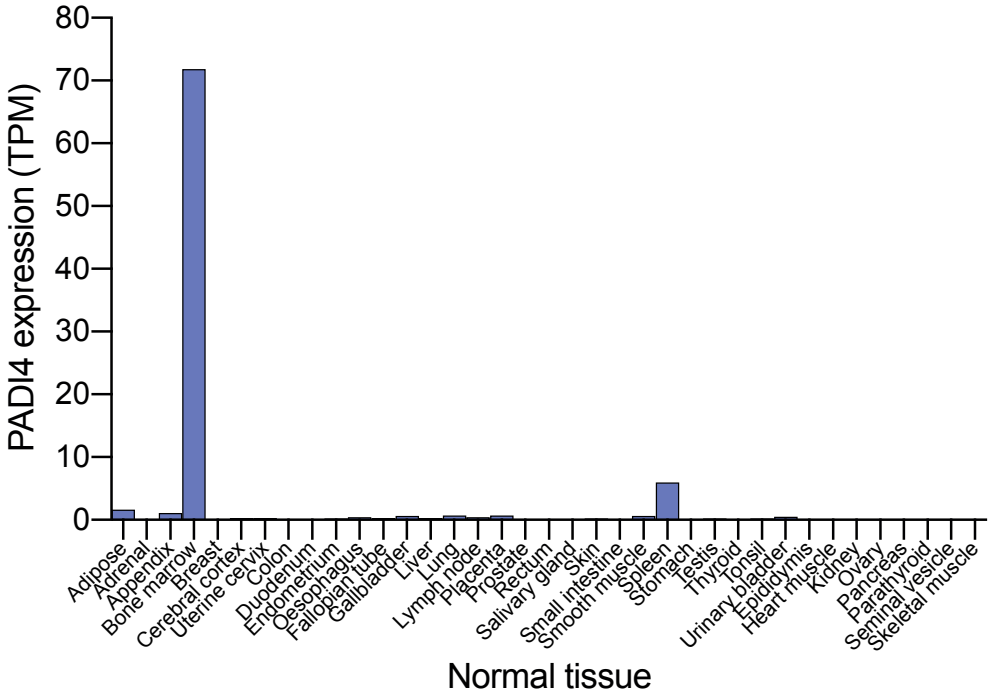
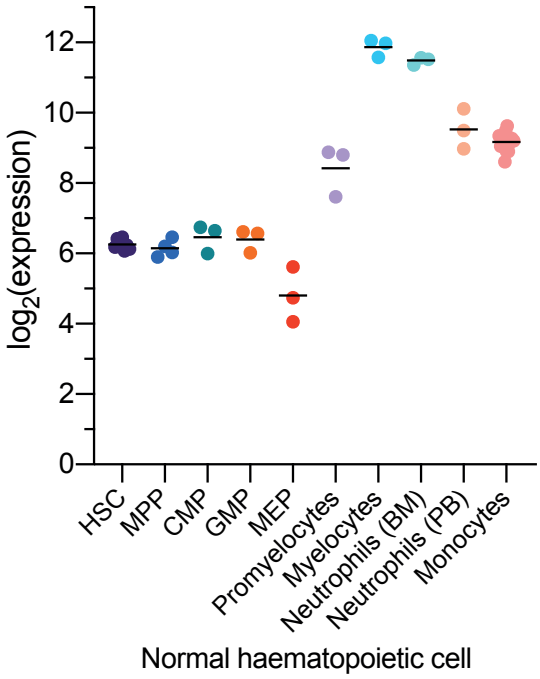


Figure 4.14: Expression of PADI4 in normal tissues. (a) Expression of PADI4 in 29 normal human tissues from RNA-Seq data (GTEx, Genotype-Tissue Expression [260]). Each dot shows expression from a single sample (total of 17,382 samples). **(b)** Expression of PADI4 across 37 human tissues from RNA-Seq data (Human Protein Atlas [234]). TPM: Transcripts per million.

(a)



(b)

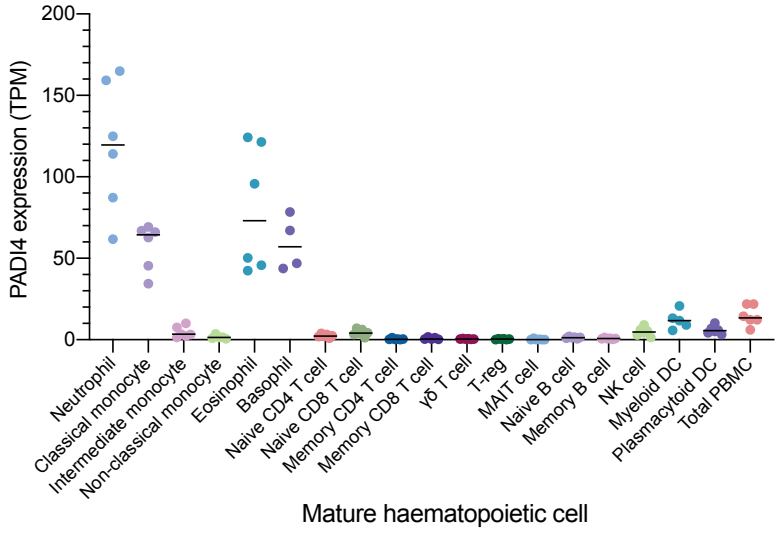


Figure 4.15: PADI4 expression in normal haematopoietic cells. **(a)** and **(b)** PADI4 expression (RNA-Seq) in different immature and mature haematopoietic cell types. Each dot represents RNA-Seq expression data from a single sample. All data were obtained from Hemaexplorer, accessed through Bloodspot [262].
BM: Bone marrow, DC: Dendritic cell, HSC: Haematopoietic stem cell, MPP: Multipotent progenitor, CMP: Common myeloid progenitor, GMP: Granulocyte-monocyte progenitor, MAIT cells: Mucosal-associated invariant T cells, MEP: Megakaryocyte-erythrocyte progenitor, NK cell: Natural killer cell, PB: Peripheral blood, PBMC: Peripheral blood mononuclear cells, TPM: Transcripts per million.

4.5.3 Expression in cancer

PADI4 was highly expressed in patients from the current study and 2 large studies of AML, Beat AML [263] and TCGA [141] (Figure 4.16). Certain karyotypic abnormalities are found recurrently in AML and have prognostic significance. PADI4 is expressed at a similar level across AMLs with common karyotypic abnormalities (Figure 4.17 (a)). Expression is also high in other myeloid malignancies, MDS and chronic myeloid leukaemia (CML) [264] (Figure 4.17 (b)).

In comparison with AML, expression at a lower level is seen in the B cell malignancies (B-cell non-Hodgkin lymphoma and chronic lymphocytic leukaemia) and a further reduction in expression is found in solid tumours (TCGA data, Figure 4.16).

Upregulation of PADI4 has been found in solid tumours, including oesophageal and lung carcinoma, where expression is associated with cell proliferation, invasion and metastasis [265, 266]. A study has found that inhibition of PADI4 using potent small molecule inhibitors, F- and Cl-amidine, results in cytotoxicity towards multiple cancer cell lines, including AML, breast and colon cancer lines. No cytotoxic effect was observed in noncancerous cell lines, suggesting that PADI4 may have the potential to be therapeutically targeted [267].

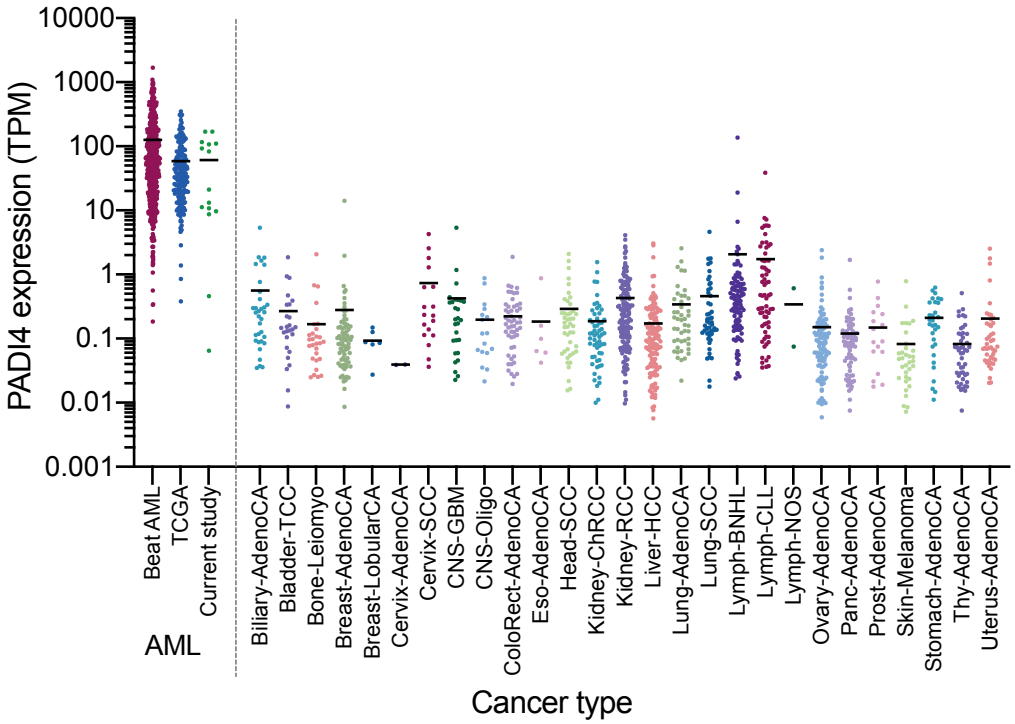


Figure 4.16: Expression of PADI4 in AML and other cancers. Expression of PADI4 in AML samples from 3 different studies: (1) Beat AML (494 samples [263]), (2) The Cancer Genome Atlas (TCGA, 179 samples [141]) and (3) the current study (15 samples). Expression data is shown from an additional 27 tumour types (TCGA, 1,393 samples) [268]. Each dot represents RNA-Seq expression data from a single sample.

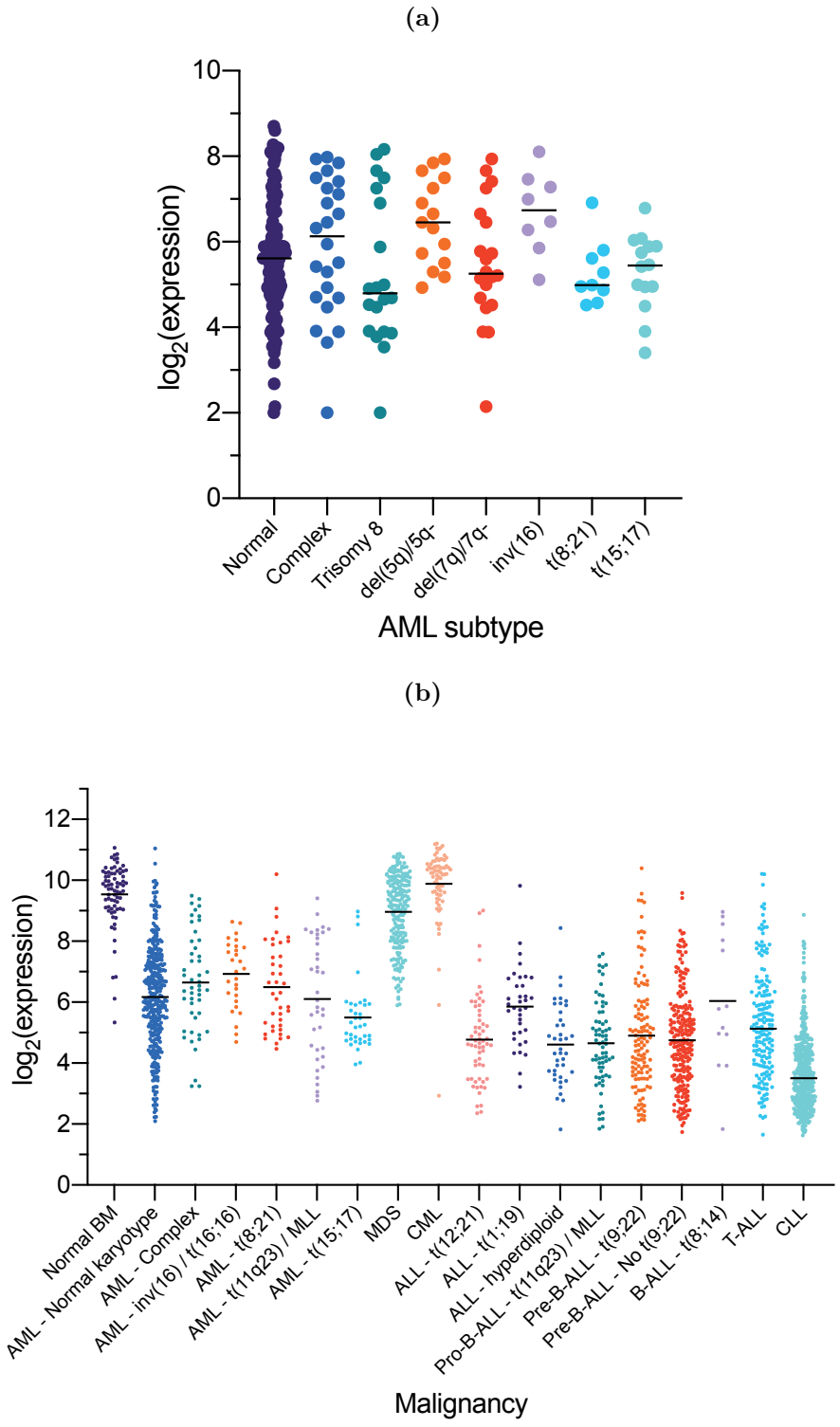


Figure 4.17: PADI4 expression in haematological malignancies. **(a)** Expression of PADI4 in different cytogenetic subtypes of AML (TCGA [141]), and **(b)** in other haematological malignancies (MILE, Microarray Innovations in LEukemia [264]). Each dot represents RNA-Seq expression data from a single sample.
BM: Bone marrow, AML: Acute myeloid leukaemia, MDS: Myelodysplastic syndrome, CML: Chronic myeloid leukaemia, ALL: Acute lymphoblastic leukaemia, CLL: Chronic lymphocytic leukaemia

4.6 Phenotyping antigen-specific T cells using single cell RNA sequencing

4.6.1 Isolating antigen-specific T cells

The phenotype of GvL T cells and the mechanisms they use to kill AML cells are not well understood. Having identified an antigen expressed by AML that is recognised by donor-derived T cells, these T cells were further characterised using single cell RNA sequencing (scRNA-Seq). Compared with a bulk approach, a single cell method has the advantage of allowing distinct phenotypes of heterogeneous cell populations to be distinguished. Crucially, a single cell approach also allows the TCRα and β pairing to be determined within individual T cells. The paired TCR sequences from antigen-specific T cells can be used to artificially engineer T cells that recognise and kill cells presenting their target antigen, such as AML cells.

Antigen-specific T cells were identified and isolated prior to analysis using IFNγ capture, which allows FACS-sorting of cells that are specifically activated by the peptide antigen. After peptide culture to enrich for antigen-specific T cells, the cultured cells were stimulated with peptide for 6 hours and then cells were sorted according to IFNγ expression for scRNA-Seq (Figure 4.18). IFNγ capture was used (Figure 4.18 (a)), which employs a bispecific antibody against CD45 and IFNγ to 'catch' IFNγ molecules as they are released by the cell. Subsequently, a fluorochrome-labelled anti-IFNγ antibody is added to enable the detection of cells that have released IFNγ. This approach was essential because conventional IFNγ staining using fixation, permeabilisation and intracellular staining with fluorescent-labelled antibodies would preclude subsequent RNA sequencing.

Four populations of T cells were sorted for scRNA-Seq: CD4$^+$IFNγ^+ (9,529 cells), CD4$^+$IFNγ^- (10,000 cells), CD8$^+$IFNγ^+ (3,711 cells) and CD8$^+$IFNγ^- (10,000 cells) (Figure 4.19). The purity of the activated (IFNγ^+) sorted populations were assessed by flow cytometry and were 96% and 99% for CD4$^+$ and CD8$^+$ cells, respectively.

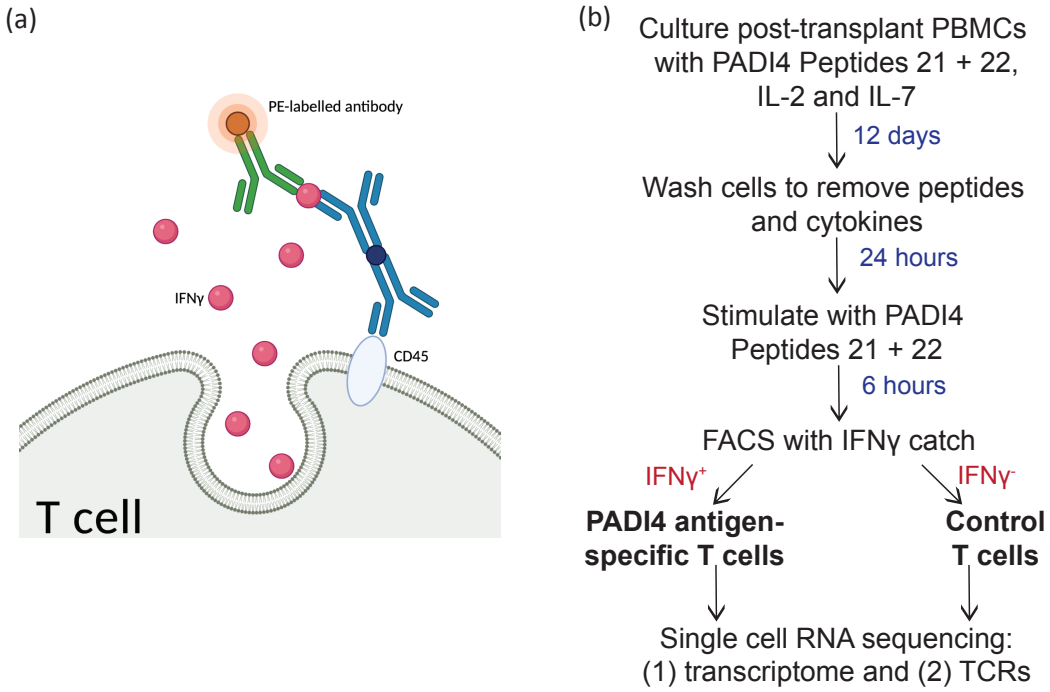


Figure 4.18: Isolation of peptide-responsive T cells for scRNA-Seq. (a) IFNγ capture uses a bispecific antibody against CD45 and IFNγ to 'catch' IFNγ molecules as they are released by the cell. Subsequently, a fluorochrome-labelled anti-IFNγ antibody is added to enable the detection of cells that have released IFNγ. **(b)** Method of isolating antigen-specific T cells for scRNA-Seq. First, post-transplant PBMCs were cultured with peptide and then washed. Cells were then stimulated with peptide and FACS-sorted according to IFNγ expression into putative antigen-specific (IFNγ^+) and negative control (IFNγ^-) populations for scRNA-Seq using the 10X Chromium platform.
PBMCs: Peripheral blood mononuclear cells, FACS: Fluorescence-activated cell sorting, scRNA-Seq: Single cell RNA-sequencing

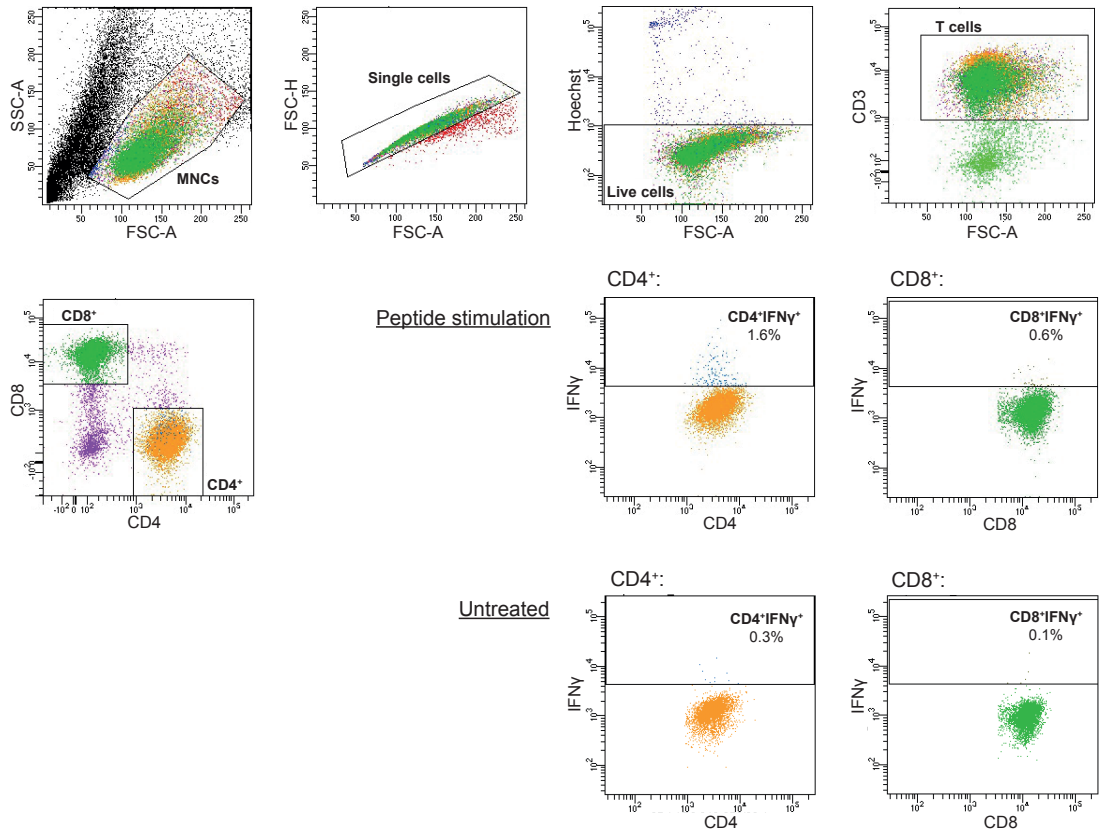


Figure 4.19: Sorting antigen-specific T cells for scRNA-Seq. Sorting strategy for purifying T cells releasing IFNγ in response to PADI4 peptides 21 + 22. IFNγ^+ and IFNγ^- populations were both sorted from samples that had been stimulated with peptide. Untreated negative control plots are also shown. The sizes of populations shown are the % of all CD4$^+$ or CD8$^+$ cells.

4.6.2 Antigen-specific T cells are CD4$^+$ T helper 1 cells with a cytotoxic phenotype

Across all FACS-sorted populations, a total of 2.09x10^9 mapped RNA-Seq reads were obtained from 12,162 T cells. Unbiased clustering of CD4$^+$ Activated (IFNγ^+) and Control (IFNγ^-) cell populations according to their transcriptomes resulted in their separation from each other when visualised by UMAP (Figure 4.20 (a)). In total, the cells were annotated in 10 clusters, with clusters 3, 5 and 6 comprising the putative antigen-specific activated cells (Figure 4.20 (b)). The cell identity within UMAP clusters was annotated by comparing the gene expression from each single cell to reference datasets derived from pure cell types, using: (1) SingleR, with immune cell reference data from Monaco *et al.* [217, 218], and (2) Cluster Identity Predictor (CIPR), with reference data from the Database of Immune Cell Expression (DICE) [219, 220]. The most common annotation(s) for cells within each cluster were used to label the cluster. The Activated clusters, which are putatively antigen-specific, all have an activated T helper 1 (Th1) phenotype, which is associated with the promotion of cell-mediated effector responses. In the Control population, multiple cell subsets were distinguished, including Th1, Th2, regulatory T cells (Treg), Th17 and naïve T cells (Figure 4.20 (c)).

In total, there were 987 differentially expressed genes between Activated and Control clusters, of which 629 (63.7%) had higher expression in the Activated population (Figure 4.21). The most significantly differentially expressed genes included genes that encode secreted cytokines and chemokines. *CCL3* and *CCL4* are members of the C-C motif ligand family of chemokines. These proteins interact with each other, forming a heterodimer, and are involved in trafficking and homing of neutrophils, monocytes and natural killer cells [269]. *CSF2* was also selectively upregulated in Activated cells and encodes GM-CSF (granulocyte-monocyte colony-stimulating factor), which stimulates haematopoietic progenitor and precursor cells to proliferate and differentiate into granulocytes and monocytes [270].

Known markers of T cell activation showed significantly increased expression in the Activated population, including the co-stimulatory immune checkpoint receptor

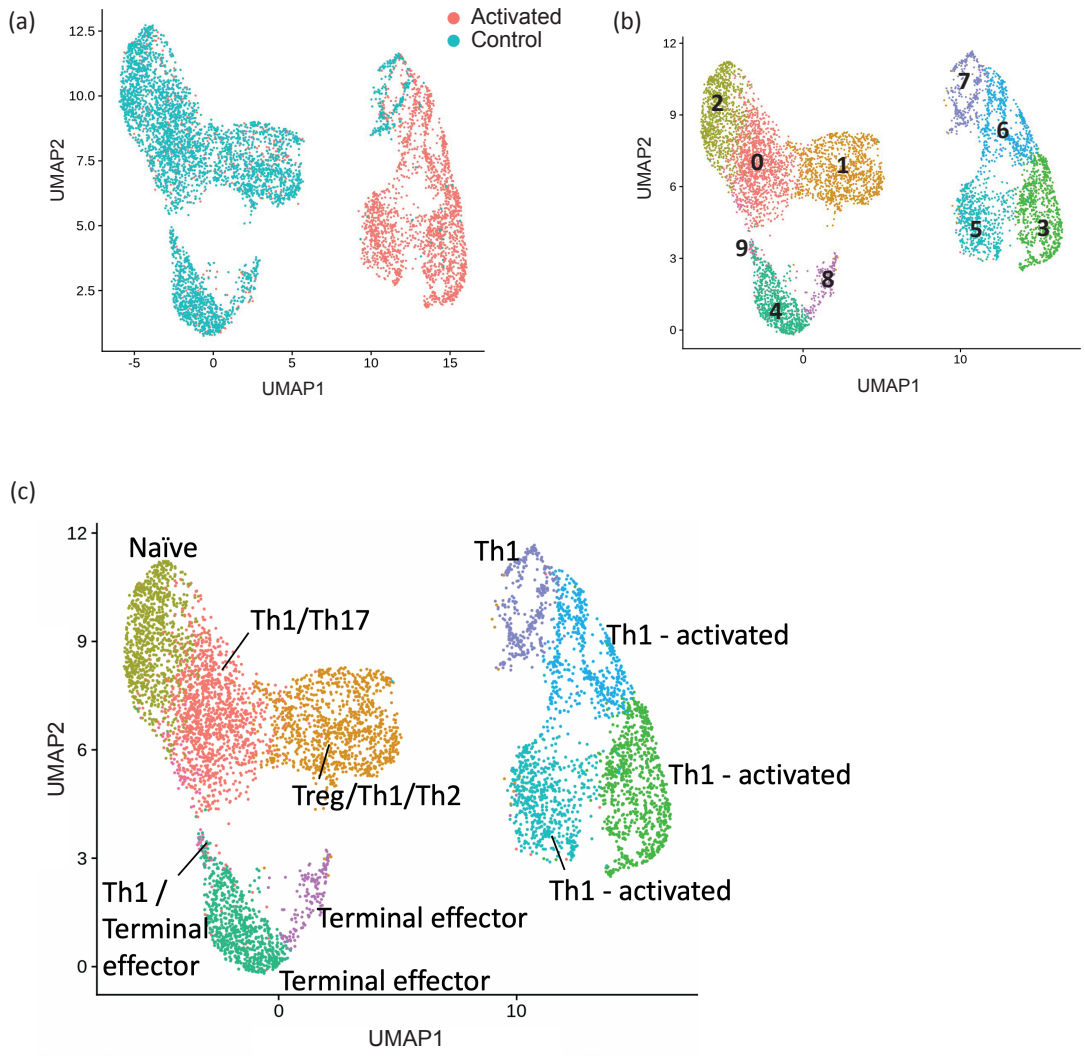


Figure 4.20: CD4$^+$ T cells cluster according to activation by PADI4 peptide. **(a)** UMAP plot for CD4$^+$ T cells with unbiased clustering, labelled according to whether they are in the Activated or Control group (red = Activated; blue = Control). **(b)** and **(c)** 10 cell clusters were identified. The transcriptomes of each single cell was compared to a reference dataset to assign its likely identity, using SingleR with the Monaco reference set [217, 218], and CIPR with the DICE reference set [219, 220].
Th1: T helper 1 cells, Th2: T helper 2 cells, Th17: T helper 17 cells, Treg: Regulatory T cells, CIPR: Cluster Identity Predictor, DICE: Database of Immune Cell Expression.

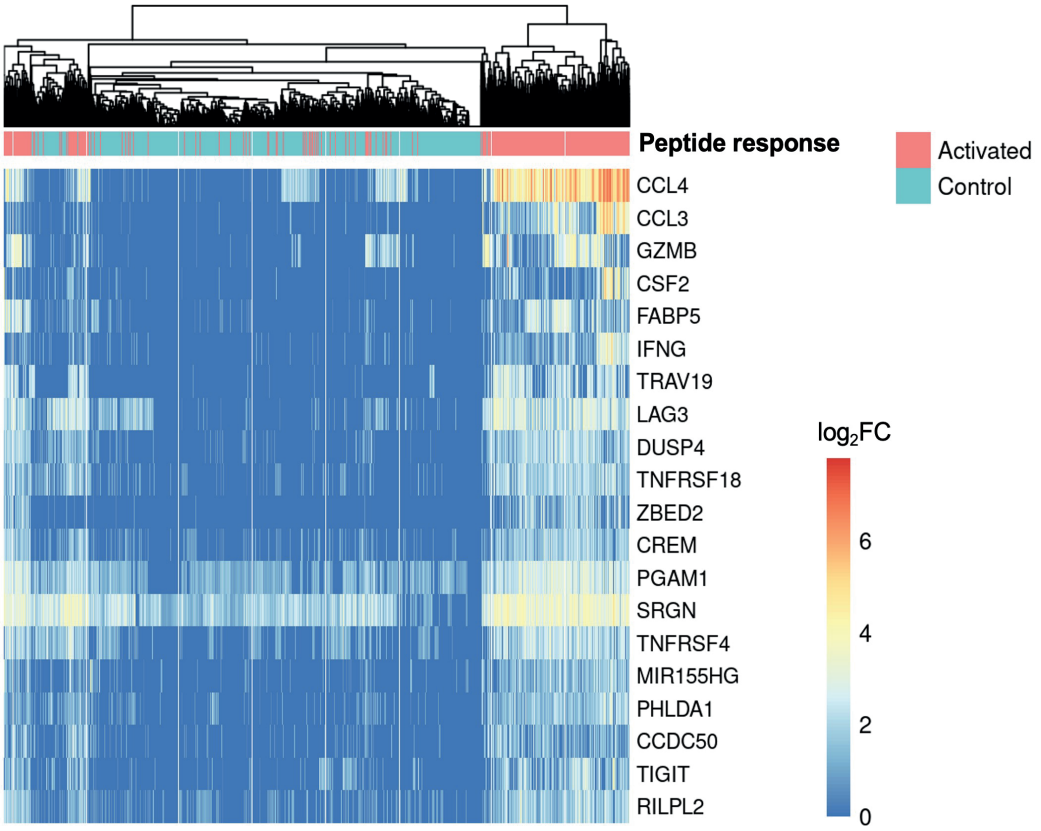


Figure 4.21: Differentially expressed genes between peptide-responsive and control T cells. Heatmap showing expression ($\log_2$FC) of the top 20 differentially expressed genes between T cells that are activated in response to peptide and those that are not activated (negative control). Cells are represented by columns and genes by rows. The identity of each cell (Activated or Control) is indicated by the coloured bar above the heatmap.
$\log_2$FC: $\log_2$(Fold Change)

TNFRSF4 (OX40), *IFNG*, *IL2RA* (CD25) and *TNFRSF9* (CD137) (Figure 4.22). Markers of chronic T cell activation were also upregulated, including *LAG3*, *TIGIT*, *PDCD1*, *HAVCR2* and *CTLA4*), which are also found in T cells that exhibit a hypofunctional 'exhausted' state (Figure 4.23) [271].

The Activated population expressed genes associated with a Th1 phenotype, including the key transcription factor *TBX21* (T-bet), as well as *IFNG*, *TNF* (TNFα) and *LTA* (TNFβ) (Figure 4.24 (a)).

Markers of cytotoxicity also showed elevated expression in Activated cells, such as *GZMB*. This encodes a serine protease that is contained within granules of cytotoxic

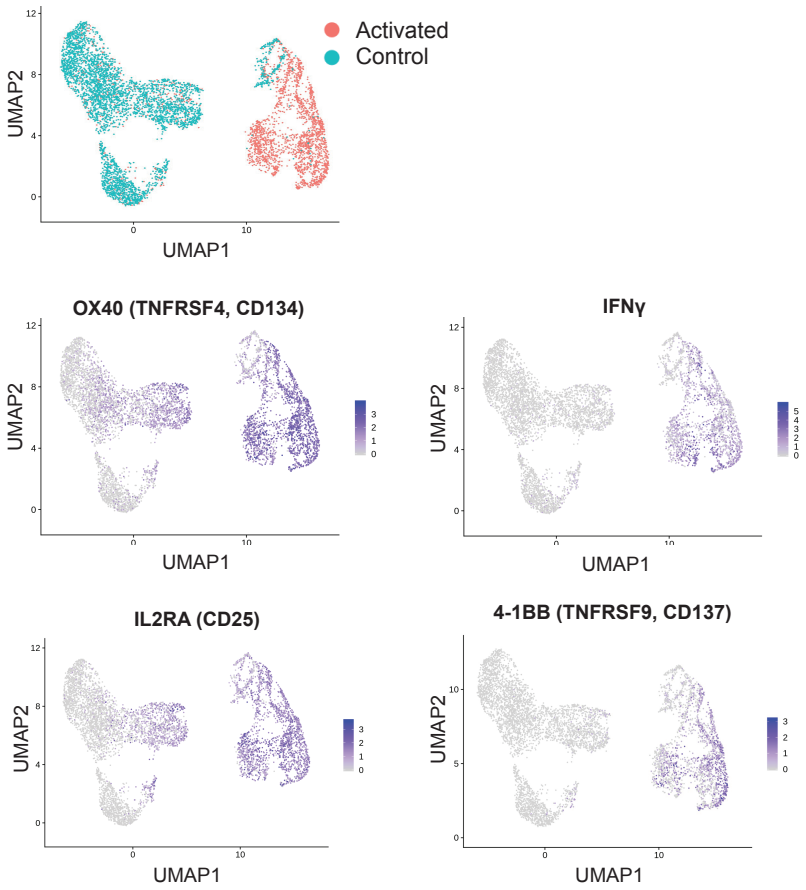


Figure 4.22: Markers of T cell activation. The top plot shows a UMAP projection for CD4$^+$ T cells with unbiased clustering, labelled according to whether they are in the Activated or Control group (red = Activated; blue = Control). The lower 4 plots show relative intensity of expression of selected genes associated with T cell activation, superimposed on the UMAP projection.

T cells. When secreted together with the pore-forming protein, perforin, it can enter target cells and mediate apoptosis [272]. *LAMP1* (CD107a) is an indicator of T cell degranulation but it displayed uniform RNA expression across both Activated and Control populations, which contrasts with the increased surface expression that was detected by FACS following peptide stimulation (Figure 4.10 (d)). This is because upregulation is mediated not by increased RNA expression, but by fusion of lysosomal membranes containing CD107a with the plasma membrane. A further possible mechanism by which the Activated T cells could result in cytotoxicity is through their increased expression of *FASLG*, which induces apoptosis upon engagement with its receptor by forming the death-inducing signaling complex

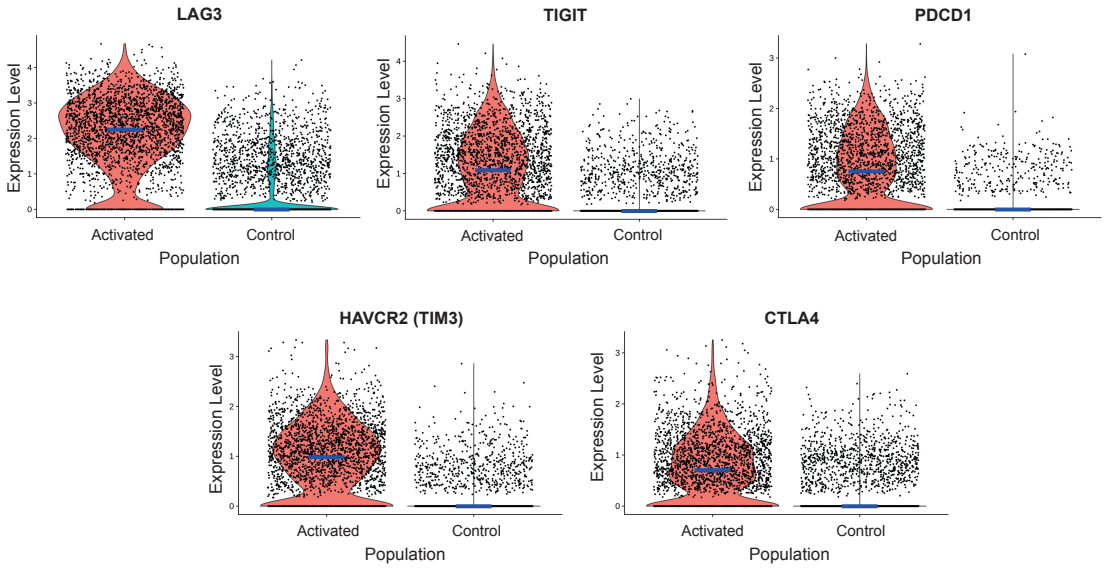


Figure 4.23: Expression of markers of chronic activation / exhaustion. For each of the 5 markers of chronic activation / exhaustion, the violin plots show the relative expression in Activated and Control cells. Each dot represents a single cell and the horizontal blue bars show the mean expression level.

(DISC) (Figure 4.24 (b)) [273].

UMAP clusters 4 and 8, which contain cells from Control populations, express *TBX21* and *GZMB*, suggesting that these are also Th1 cells, but do not recognise the PADI4 antigen (Figures 4.20 (b) and 4.24).

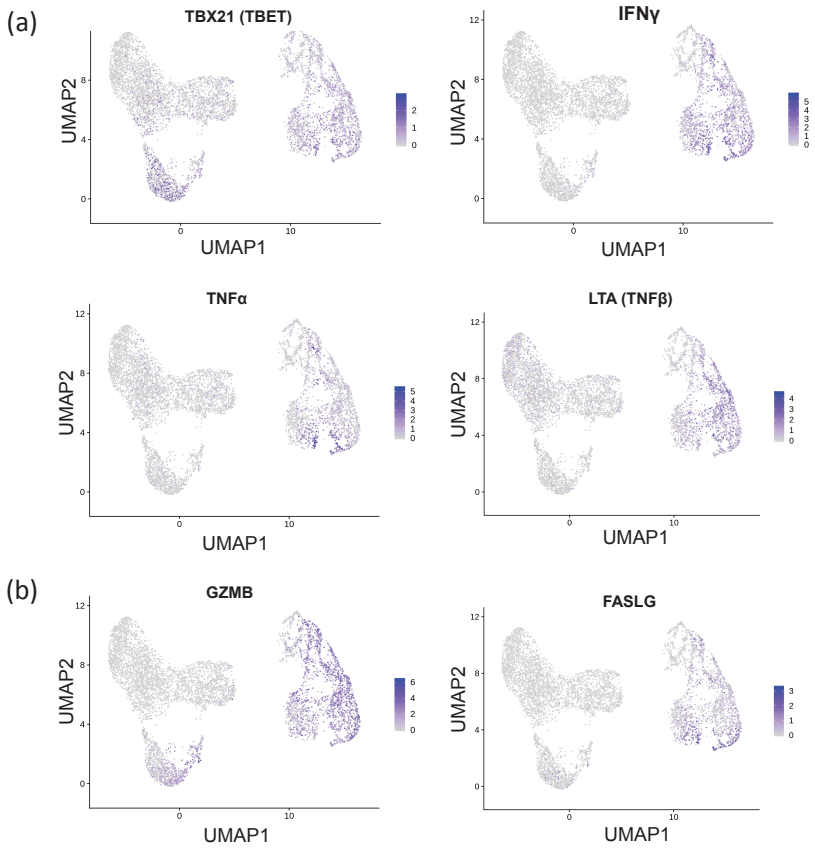


Figure 4.24: Activated CD4$^+$ are Th1 cells with a cytotoxic phenotype. The reference UMAP plot showing Activated and Control cells is at the top of Figure 4.22. The plots show relative intensity of expression of the selected genes associated with **(a)** a Th1 phenotype, and **(b)** cytotoxicity, superimposed on the UMAP projection.
Th1: T helper 1

4.6.3 Pathway analysis

To gain further insight into the biological relevance of transcriptional differences between T cells activated by peptide and negative control cells, analysis of the pathways linking the differentially expressed genes was performed. Gene set enrichment analysis (GSEA) is a computational method that takes a ranked gene expression list and a previously defined reference set of genes and determines whether the members of the reference set are randomly distributed throughout the ranked gene expression list, or whether they are primarily found at the top or bottom [274]. This allows a normalised enrichment score to be calculated, which indicates the degree to which a gene set is over-represented in a ranked list of genes.

Using the 987 differentially expressed between Activated and Control cells, GSEA was performed using Hallmark gene reference sets, which represent well-defined biological states and processes. In the Activated cells, there was significant enrichment for TNFα signaling via NFκB and allograft rejection, after accounting for multiple hypothesis testing (Figure 4.25). Unexpectedly, in the Control cells, there was enrichment for genes in the IFNα response pathway, which is part of the innate immune system, mediating responses to environmental exposures such as viral infections [275].

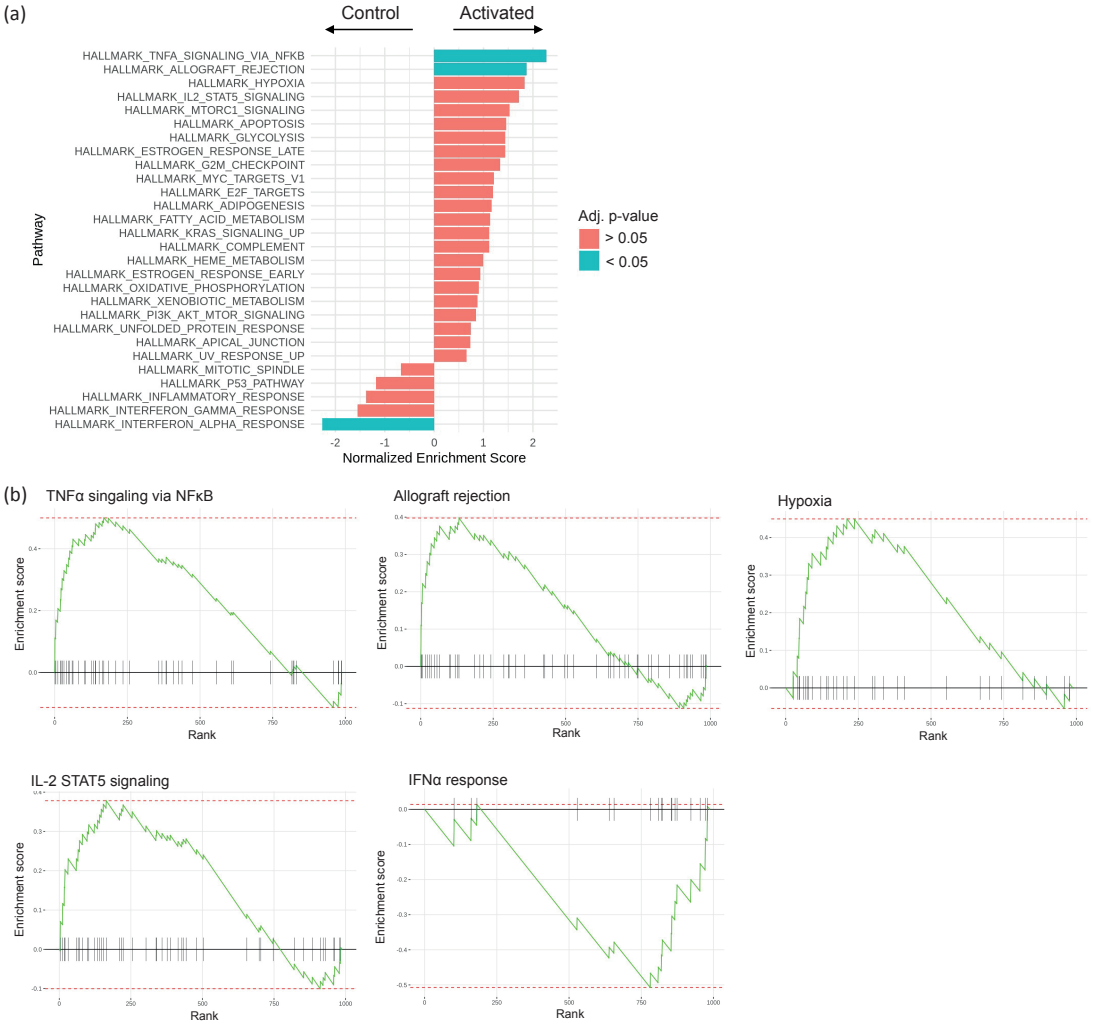


Figure 4.25: GSEA of differentially expressed genes between Activated and Control cells. (a) GSEA was performed using Hallmark reference gene sets. The plot summarises the normalised enrichment scores for all pathways with individual test p-values < 0.05. After adjusting for multiple hypothesis testing, FDR adjusted p-values < 0.05 are shown in blue and values > 0.05 are in red. Positive bars are pathways that are enriched in Activated cells, whereas negative bars are pathways that are enriched in Control cells. **(b)** GSEA plots showing enrichment of TNFα signaling via NFκB, allograft rejection, hypoxia and IL-2 STAT5 signaling in Activated cells. The bottom plot shows enrichment of IFNα response in Control cells.
GSEA: Gene set enrichment analysis

4.7 Identifying antigen-specific TCR sequences

TCRs are heterodimers comprising α and β chains. By assessing the sequences of paired TCRα and TCRβ within single cells, T cells from a single clone can be distinguished. In the population of T cells identified by flow cytometry that were activated by peptide ($CD4^+IFN\gamma^+$) there were 2 expanded clones, accounting for 31.5% and 29.1% of the total repertoire (Figure 4.26). In contrast, the largest clone in the population of T cells that were not activated by peptide ($CD4^+IFN\gamma^-$) was much smaller and comprised only 6.7% of the repertoire. Expanded Clone 1 was present in the Control repertoire at low frequency (0.18%) but Clone 2 was not detected. This indicates that the expanded clones have been effectively enriched using the IFNγ capture assay.

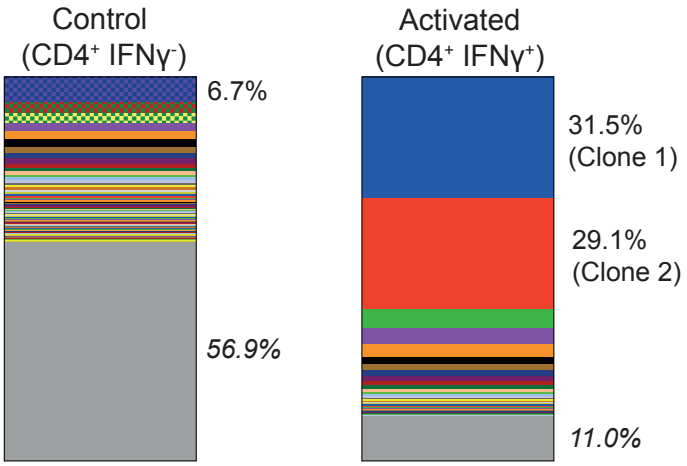
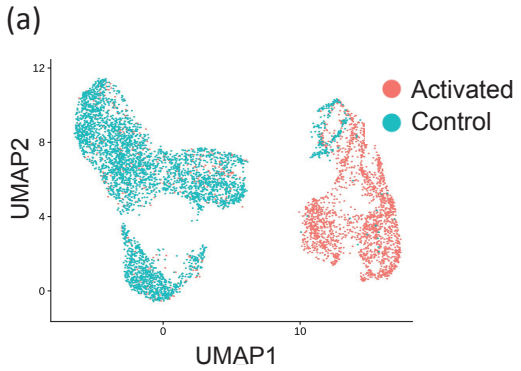


Figure 4.26: $CD4^+$ clones in Activated and Control repertoires. Each large bar represents all T cell clones identified from the Control ($CD4^+IFN\gamma^-$, left bar) and Activated ($CD4^+IFN\gamma^+$, right bar) TCR repertoires. The grey portions of each bar represent the sum of all clones which are individually $< 0.1\%$ of the total repertoire. In the repertoire of TCRs from cells activated by PADI4 peptide (right bar), Clones 1 and 2 represent 31.5% and 29.1% of the total repertoire, respectively.
TCR: T cell receptor

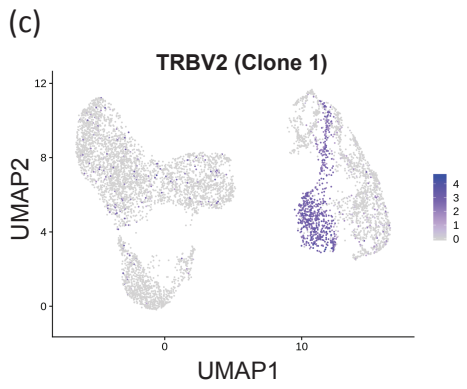
Clone 2 expressed 2 different TCRα chains and 1 TCRβ chain and therefore 2 possible pairings of TCR chains are possible (Figure 4.27 (b)). However, the 4th

largest clone in the activated repertoire (4.0% of total repertoire) expressed just 1 of these TCRα chains, together with the TCRβ chain. Therefore, it is likely that the combination found in the 4th largest clone is the pairing that recognises the PADI4 antigen, as it has been enriched during *in vitro* culture with peptide. Unbiased clustering of cells from CD4$^+$ Activated and Control populations according to their transcriptomes resulted in segregation between Clones 1 and 2, indicating that they have distinct transcriptional profiles (Figure 4.27 (c) and (d)).



(b)

Clone		Sequence			
		V	D	J	C
1	TRA	TRAV19		TRAJ32	TRAC
	TRB	TRBV2		TRBJ2-7	TRBC2
2	TRA	TRAV19		TRAJ15	TRAC
	TRA	TRAV12-2		TRAJ29	TRAC
	TRB	TRBV5-1	TRBD1	TRBJ1-2	TRBC1



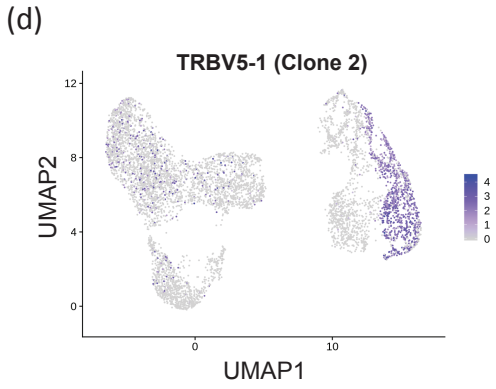


Figure 4.27: Expanded Clones 1 and 2 have distinct TCRs and transcriptional profiles. (a) UMAP plot for CD4$^+$ cells with unbiased clustering, labelled according to whether they are in the Activated or Control group (red = Activated; blue = Control). **(b)** Gene segments (V, D, J and C) for the TCRα and TCRβ from Clones 1 and 2. **(c)** and **(d)** UMAP plots highlighting the expression of TRBV gene segments found in Clones 1 and 2. Although the clusters lie close to each other, they are transcriptionally distinct. TRA: T cell receptor α, TRB: T cell receptor β.

There were 352 differentially expressed genes between Clones 1 and 2. The individual component TCR gene segments involved in producing the distinct TCRs

for each clone were the most significantly differentially expressed genes. TRBV2 and TRBC2 were expressed in Clone 1, whereas TRAV12-2 and TRBV5-1 were expressed in Clone 2. Interestingly, several genes that have been implicated in cytolytic activity were specifically expressed by cells from Clone 1, but not Clone 2, including *GNLY*, *CTSW* and *LAYN*. Granulysin (*GNLY*) is present in T cell cytolytic granules and produces pores in the membrane of target cells, which enables death-inducing enzymes to enter the cell [276]. Cathepsin W (*CTSW*) and layilin (*LAYN*) are expressed in cytotoxic T cells and are involved in the regulation of cytotoxic activity [277, 278]. It is unclear whether these differences between the 2 putatively antigen-specific T cell clones translates into functional differences in their ability to recognise and kill tumour cells *in vivo*.

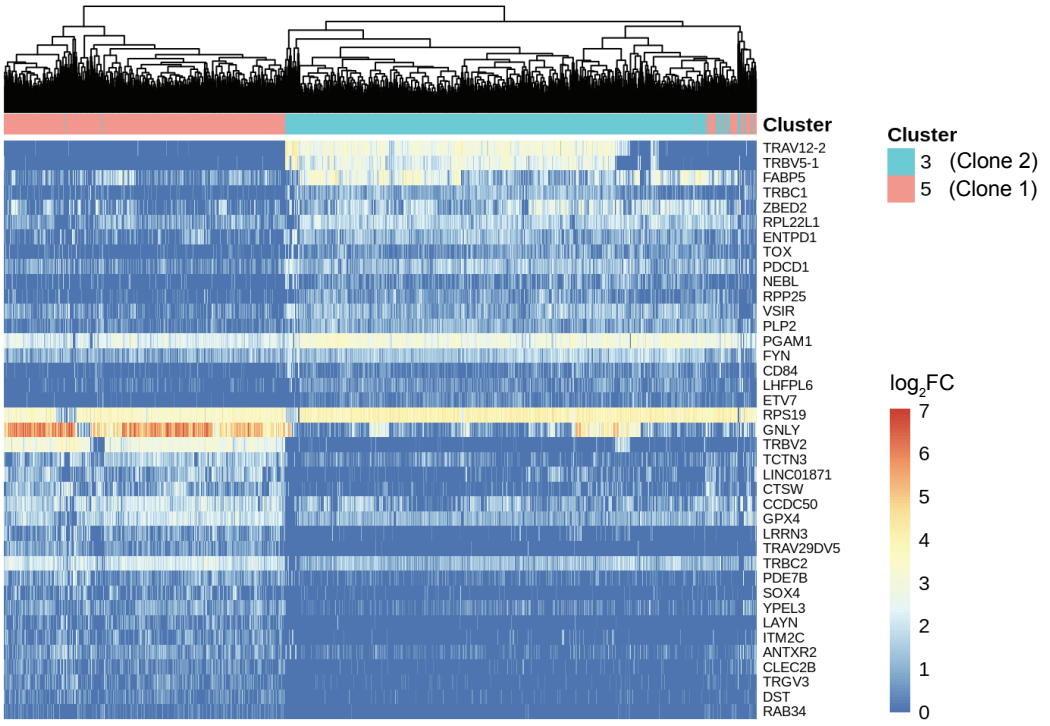


Figure 4.28: Differentially expressed genes between expanded Clones 1 and 2. Heatmap showing expression ($\log_2$FC) of differentially expressed genes between Clones 1 and 2. A total of 40 genes are shown, comprising the top 20 differentially expressed genes that are upregulated in Clone 2 (top half of heatmap) and top 20 differentially expressed genes that are upregulated in Clone 1 (bottom half of heatmap). Cells are represented by columns and genes by rows. The identity of each cell is indicated by the coloured bar above the heatmap: Clone 1 (Cluster 5) in red, and Clone 2 (Cluster 3) in blue.
$\log_2$FC: $\log_2$(Fold Change)

4.8 Antigen-specific T cells are not CD8$^+$

In addition to sorting populations of CD4$^+$ cells according to IFNγ capture following peptide stimulation, CD8$^+$ cells were sorted. When populations were clustered together and visualised using UMAP, both of the sorted populations, CD8$^+$IFNγ^+ and CD8$^+$IFNγ^-, overlap each other suggesting they are not transcriptionally distinct. In contrast, the sorted CD4$^+$IFNγ^+ and CD4$^+$IFNγ^- populations segregate on the UMAP plot (Figure 4.29). This supports the prior data indicating that T cell responses to the PADI4 antigen are mediated by CD4$^+$ cells and not

$CD8^+$ cells. Therefore, the FACS-sorted $CD8^+IFN\gamma^+$ cells represent a background level of $IFN\gamma^+$ cells rather than antigen-specificity.

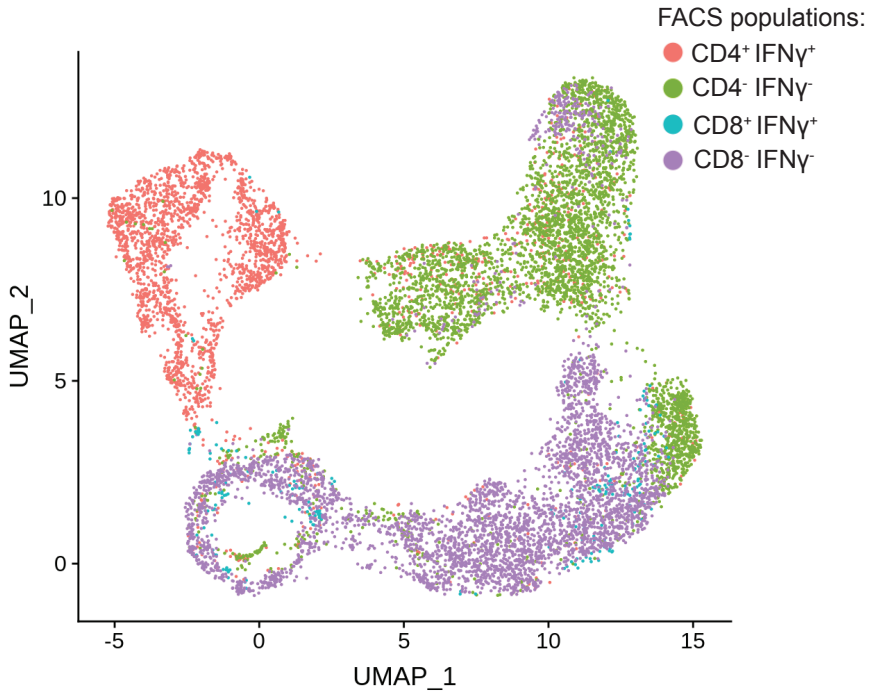


Figure 4.29: Clustering of CD4 and CD8 T cells. Visualisation of all FACS-sorted populations ($CD4^+IFN\gamma^+$, $CD4^+IFN\gamma^-$, $CD8^+IFN\gamma^+$ and $CD8^+IFN\gamma^-$) with UMAP.

Further evidence that $CD8^+$ cells are not responding the PADI4 peptide was provided by the TCR repertoires, where there was no significant T cell clonal expansion. The largest clone in the $CD8^+IFN\gamma^+$ FACS-sorted population was 6.6%, whereas the largest clone in the $CD8^+IFN\gamma^-$ was 10.9% (Figure 4.30). Unlike for $CD4^+$ cells, there was substantial overlap in the clones detected in the sorted $CD8^+IFN\gamma^+$ and $CD8^+IFN\gamma^-$ repertoires, suggesting that clones are not being specifically enriched by the $IFN\gamma$ capture technique (Figure 4.31).

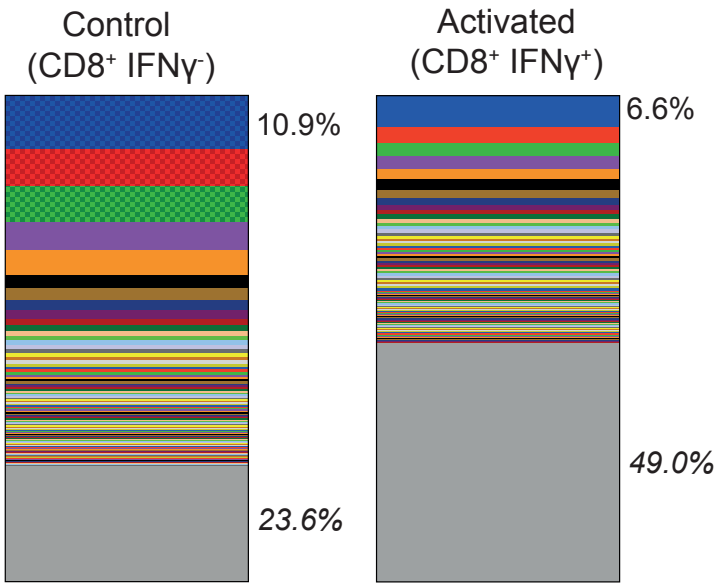


Figure 4.30: CD8 clonotypes in IFNγ^+ and IFNγ^- repertoires. Each large bar represents all T cell clones identified from the Control (CD8$^+$IFNγ^-, left bar) and Activated (CD8$^+$IFNγ^+, right bar) TCR repertoires. The grey portions of each bar represent the sum of all clones that individually comprise $< 0.1\%$ of the total repertoire.

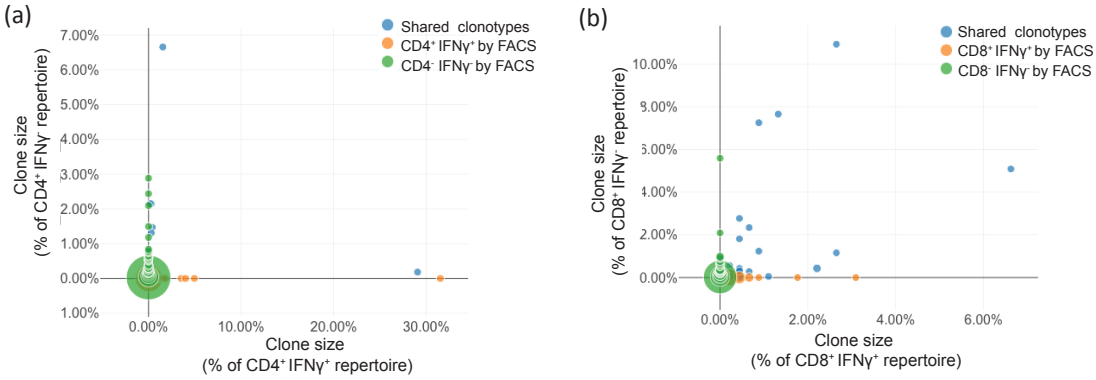


Figure 4.31: Overlap between Activated and Control repertoires. Overlap in paired TCRα and TCRβ sequences between IFNγ^+ and IFNγ^- repertoires for **(a)** CD4, and **(b)** CD8 populations. Orange dots represent clones present only in the IFNγ^+ repertoires; green dots represent clones present only in the IFNγ^- repertoires; blue dots represent shared clones between IFNγ^+ and IFNγ^- repertoires.

4.9 *Ex vivo* post-transplant T cell receptor repertoires

For the IFNγ capture assay, post-transplant cells were cultured for 12 days with peptide and cytokines. Therefore, the proportion of antigen-specific T cells detected after flow cytometry and scRNA-Seq in this assay will not be representative of the repertoire in the patient. Bulk TCR sequencing was performed *ex vivo* on T cells from the patient's blood 6 years post-transplant in order to describe the overall repertoire and also to enable the size of antigen-specific clones to be ascertained. $CD4^+$ and $CD8^+$ cells were separately sorted and TCRα and TCRβ repertoire sequencing was performed, using the 5' RACE (rapid amplification of cDNA ends) technique for first strand cDNA synthesis. Unique molecular identifiers were used to allow accurate transcript-level quantitation from sequencing data that is unbiased by PCR amplification [279, 280].

Overall, there was less diversity in the $CD8^+$ compared with $CD4^+$ repertoires (Figure 4.32 (b) and (d)). There were fewer different clones detected and individual clones comprised a larger proportion of the overall repertoires, in keeping with other studies [281, 282]. PADI4 Clone 2 was detected in the TCRα and TCRβ repertoires at frequencies of 0.00827% and 0.00482%, respectively (Figure 4.32 (a) and (c)). Only one of the TCRα chains of Clone 2 was detectable in the repertoire, which is the same chain that is present in the 4th largest clone identified from scRNA-Seq, and is likely to be the chain that is important in producing the antigen-specific TCR (See Section 4.7).

Understanding the relationship between a TCR sequence and binding specificity is an outstanding question. A limited number of specificities can be assigned using currently available curated databases of TCR sequences with known antigen specificities [283]. Using VDJdb, sequences were annotated with their antigen for between 1.08% and 13.95% of the repertoires [284] (Figure 4.32). These were principally viral antigens in $CD8^+$ cells, which reflects the composition of TCRs in the reference database.

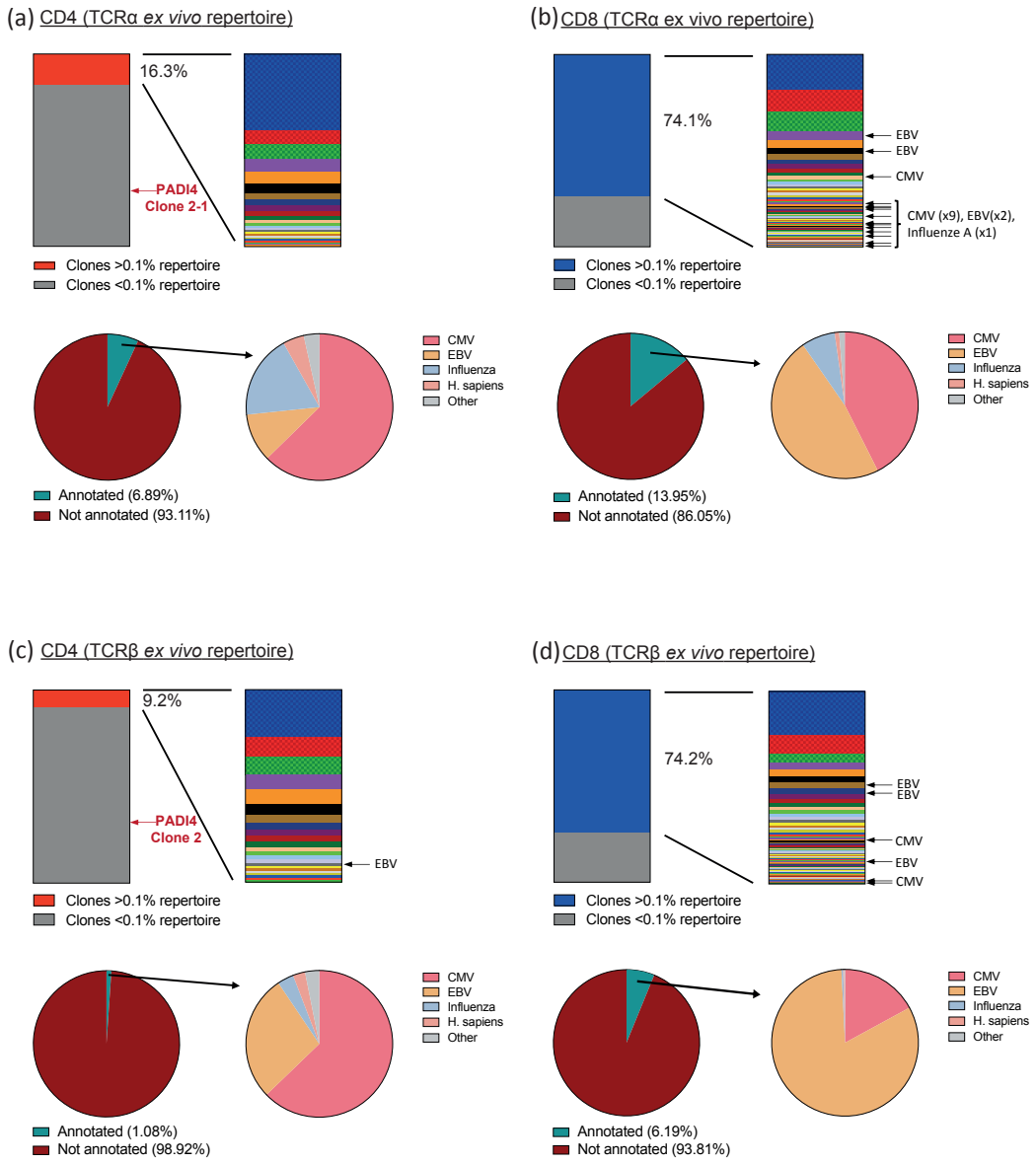


Figure 4.32: *Ex vivo* post-transplant T cell repertoires. TCRα (**(a)** and **(b)**) and TCRβ (**(c)** and **(d)**) *ex vivo* repertoires from CD4$^+$ and CD8$^+$ populations from blood taken 6 years post-transplant from Patient OX866. For each sub-figure, the left bar represents the total repertoire. The grey portion represents the sum of T cell clones that individually comprise $< 0.1\%$ of the total repertoire. The coloured portion (red for CD4 / blue for CD8) represents all clones that individually comprise $> 0.1\%$ of the repertoire. The right bar in each sub-figure shows the size of these clones. PADI4 Clone 2 was detected in the CD4$^+$ repertoires, as indicated. Annotation of the target antigen for some clones was possible using VDJdb, as indicated with arrows [284]. The pie charts indicate the proportion of each repertoire that was annotated by VDJdb (left pie charts) and the source of antigen (right pie charts). Those labelled as H. sapiens were previously found to recognise neoantigens or tumour-associated antigens.
CMV: Cyclomegalovirus, EBV: Epstein-Barr virus

4.10 Conclusions: PADI4 antigen identification and characterisation

Functional screening of all *in silico* predicted peptide antigens from Patient OX866 revealed T cell responses to a peptide derived from the PADI4 protein. Characterisation of donor-derived T cell responses from post-allo-SCT blood samples demonstrated that responses are HLA-DP-restricted and result in activation of $CD4^+$ T cells with a cytotoxic Th1 phenotype. PADI4 is widely expressed in AML samples. In normal tissues, it is selectively expressed in haematopoietic cells, which may explain why Patient OX866 had clinical evidence of a GvL response without GvHD.

In this chapter, one antigen was characterised from a single patient. The antigen identification and characterisation workflow was validated by studying a further 9 AML patients treated with allo-SCT in Chapter 5, which gives insight into the antigens and associated alloreactive T cell responses that may be responsible for GvL.

Putative antigen-specific TCR sequences were identified using single cell RNA-Seq of T cells that are activated in response to the PADI4 peptide. Chapter 6 describes how these TCRs were used to develop a PADI4 antigen-specific T cell therapy to target AML cells.

Chapter 5

Applying the antigen-identification workflow to characterise further GvL responses

Contents

5.1 Characterising GvL responses in a cohort of transplant recipients 149
5.1.1 Aims 149
5.1.2 Clinical evidence for a GvL response 151
5.1.3 Functional screening of predicted GvL antigens 153
5.2 Patient OX1149 154
5.2.1 Clinical features 154
5.2.2 Screening for alloreactive peptides 155
5.2.3 Characterising T cell responses to peptide 190 160
5.2.4 Characterising T cell responses to peptide 290 162
5.3 Patient OX289 163
5.3.1 Clinical features 163
5.3.2 Screening for alloreactive peptides 164
5.3.3 Immunophenotyping of antigen-specific T cell responses 176
5.3.4 HLA restriction of peptide responses 178
5.4 Summary of alloreactive peptides 180
5.5 Predicting antigen binding to common HLA alleles . . 183
5.6 Expression of antigens in normal tissues 185
5.7 Identifying antigen-specific TCR sequences 188
5.8 Functional screening of peptides identified by mass spectrometry 191
5.8.1 *Ex vivo* testing 191
5.8.2 Testing following *in vitro* culture 191
5.9 Conclusions: Identifying and characterising alloreactive antigens 193

5.1 Characterising GvL responses in a cohort of transplant recipients

5.1.1 Aims

For each of the 15 patients studied, there were a large number of candidate HLA-binding peptides that were predicted *in silico* (Section 3.4). However, the

machine learning approach to predicting HLA-binding peptides (NetMHCpan 4.0 and NetMHCIIpan 3.0) is imperfect, particularly for less common HLA alleles, due to the limited amount of data used to train the artificial neural networks. Furthermore, although a particular amino acid sequence is present within a protein, this does not necessarily mean that the peptide will be processed and presented by HLA on the cell surface. Even if a peptide is presented by HLA, the presenting cell must encounter a T cell expressing its cognate TCR in order to generate a T cell response.

For these reasons, functional screening of peptides was vital and the approach was successful in identifying one antigen in Patient OX866 (Chapter 4). IFNγ screening was similarly used to identify further antigens in an additional 9 patients. In each case, post-transplant blood from the patient in whom the HLA-binding predictions were made was used to look for peptide-responsive donor-derived T cells. Post-transplant blood was unable to be obtained for the remaining 5 patients, and as this was essential, screening could not be performed for these patients. Of these 5 patients, 3 died from relapsed disease within 16 months of allo-SCT. The remaining 2 patients could not be contacted to request a blood donation (Table A.2). The aims of peptide screening were:

(1) To demonstrate that the antigen identification workflow is robust enough to identify alloreactive antigens in multiple allo-SCT recipients.

(2) To provide insight into the antigenic landscape of GvL in AML patients, including the nature of antigens, the phenotype of alloreactive T cells and the HLA restriction of responses.

(3) To identify antigens that could be of clinical importance as targets for immunotherapy. Identifying antigens that elicit GvL responses could also be used to improve donor selection for individual allo-SCT recipients, by choosing a donor that is mismatched for the relevant germline DNA variant and therefore is predicted to result in an alloreactive T cell response.

5.1.2 Clinical evidence for a GvL response

Evidence for a GvL response, which serves to reduce the risk of relapse, is provided by multiple clinical observations (Section 1.6). However, for individual patients receiving allo-SCT, it is not possible to directly measure this therapeutic effect. Through analysis of clinical features of a patient's disease and treatment, the possibility of a GvL effect can be predicted. Antigens identified from patients with clinical evidence of a GvL effect are attractive candidates for clinical translation, as they could have resulted in a productive therapeutic response *in vivo*.

In individual patients, clinical features suggestive of a possible GvL response are:

(1) Survival from poor-risk disease, defined either based on genetic features of AML at baseline [36] or response to chemotherapy. If the patient does not respond adequately to chemotherapy (refractory disease) or relapses, their prognosis is worse. In these cases, given that AML was present despite chemotherapy treatment, their disease was less likely to have been curable with chemotherapy alone and a GvL response might have been important in controlling disease.

(2) Residual detectable AML prior to allo-SCT, but cured following transplantation. Disease may be detectable using histology, cytology, or higher sensitivity MRD methods (RT qPCR or flow cytometry).

(3) Disease response to DLI treatment following post-allo-SCT relapse.

(4) The presence of GvHD, which correlates with reduced relapse [45–48]. Unlike a GvL response, GvHD can be directly assessed and there are likely to be shared mechanisms and antigens for both responses.

Clinical evidence for a GvL response was evaluated for all 10 patients undergoing functional screening of predicted antigenic peptides (Table 5.1). All patients have survived at least 3.5 years post-transplant and are presumed to be cured. Patient OX866 had the strongest evidence for a GvL response and was studied in Chapter 4.

Patient	(1) AML prognostic risk		(2) Disease detected prior to allo-SCT	(3) Response of post-allo-SCT relapsed disease to DLI	(4) GvHD	Features suggestive of a GvL response
	ELN 2017 genetic risk	Response to chemotherapy				
OX866	Intermediate	Refractory to first cycle of chemotherapy. Residual disease after salvage chemotherapy	Nil	Excellent response	Nil	1. High risk disease: • Refractory to first-line chemotherapy • Residual disease after salvage chemotherapy • Relapse post-allo-SCT 2. Response to DLI
OX1149	Intermediate	Relapse after 4 cycles of chemotherapy, detectable by *NPM1* MRD	MRD positive	NA	Very mild, oral	1. High risk disease: • Relapse pre-allo-SCT 2. MRD positive pre-allo-SCT 3. GvHD
OX885	Intermediate	Relapse after 3 cycles of chemotherapy	Nil	NA	Mild, skin	1. High risk disease: • Relapse pre-allo-SCT 2. GvHD
OX993	Intermediate	Relapse after 4 cycles of chemotherapy	Nil	NA	Mild, skin	1. High risk disease: • Relapse pre-allo-SCT 2. GvHD
OX1098	Intermediate	Refractory to first cycle of chemotherapy	Nil	NA	Nil	1. High risk disease: • Refractory to first-line chemotherapy
OX813	Favourable	Relapse after 4 cycles of chemotherapy, detectable by *CBFB-MYH1* MRD	MRD positive	NA	Mild, skin	1. High risk disease: • Relapse pre-allo-SCT 2. MRD positive pre-allo-SCT 3. GvHD
OX289	Intermediate	Good response	Nil	NA	Chronic: mouth, gut and liver	GvHD
OX802	Intermediate	Good response	Nil	NA	Chronic: skin, mouth, gut, liver and ocular. Autoimmune neutropenia	GvHD
OX747	Intermediate	Good response	Nil	NA	Skin	GvHD
OX628	Favourable	Good response	Nil	NA	Mild, skin	GvHD

Table 5.1: Clinical evidence for a GvL response. Clinical features that could indicate a GvL response are indicated in the columns numbered (1) - (4). Adverse prognostic features indicate a lower likelihood of cure with chemotherapy alone and given that all patients are long-term survivors, a GvL response might have been important in controlling disease. The ELN 2017 genetic risk was determined according to cytogenetic and molecular genetic features of disease at diagnosis [36] (Table A.1). MRD assessment was by RT qPCR. DLI treatment is only shown when given for post-allo-SCT relapse and not for reduced donor CD3 chimerism. A summary of features that could indicate a GvL response are shown in the final column, highlighted in yellow.
MRD: Measurable residual disease, Allo-SCT: Allogeneic haematopoietic stem cell transplantation, DLI: Donor lymphocyte infusion, RT qPCR: Real-time quantitative polymerase chain reaction

5.1.3 Functional screening of predicted GvL antigens

For each patient, libraries of *in silico* predicted HLA-binding peptides were synthesised (Section 3.4). Where multiple peptides overlapped the same genetic variant, long peptides overlapping all of these predicted HLA-binding peptides were designed, as outlined in Section 4.1.3.

IFNγ ELISpots were performed after post-transplant PBMCs were cultured with peptide pools, which increases the sensitivity of the assay (Section 4.2.2). First, peptides were combined in pools of up to 35 peptides. Post-transplant PBMCs from each patient were cultured with pools of peptides that had been predicted to bind their own HLA molecules, together with IL-2 and IL-7. Cells were grown in separate cultures, which each combined up to 3 pools, in total representing up to 100 peptides (2 μg/ml each peptide). After 12 days of culture, cells were washed to remove peptide and cytokines and rested overnight. IFNγ ELISpot of pools was then performed, using the same peptides as were present in their culture (Figure 5.1). Positive responses were defined for averaged values across replicates as either greater than 50 spot forming units (SFU) per million cells, or 3 x DMSO negative control, whichever value was greater.

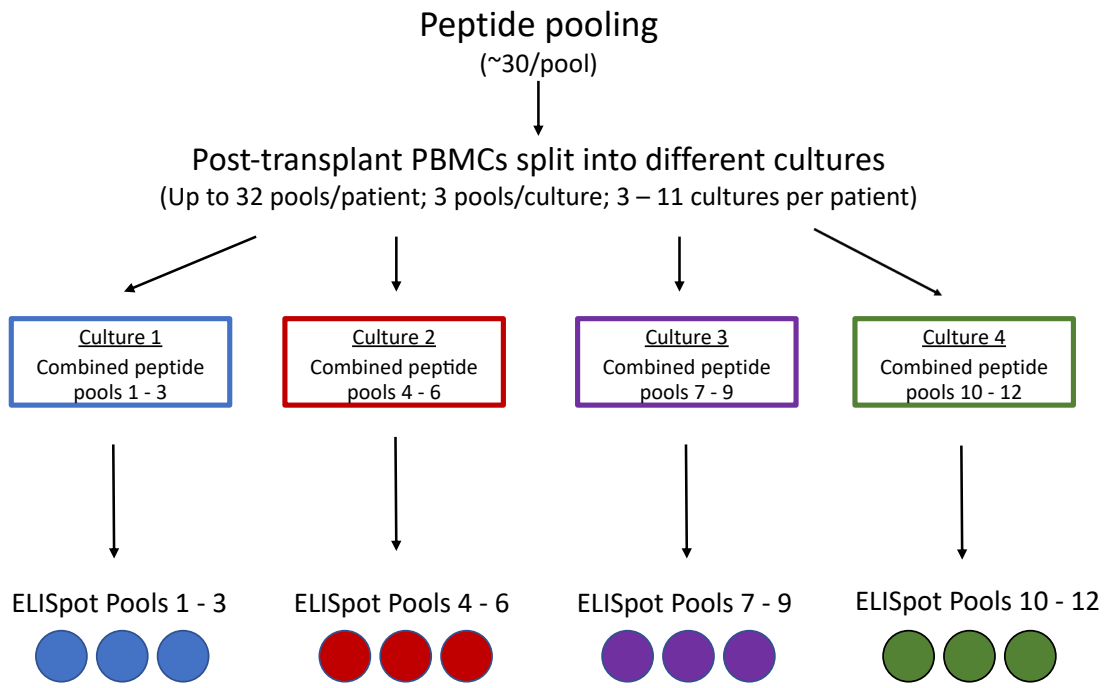


Figure 5.1: Approach to screening predicted GvL peptides for T cell responses. Post-transplant PBMCs were cultured for 12 days with up to 100 peptides, each at 2 μg/ml, in the presence of IL-2 and IL-7. After culture, cells were washed and rested overnight. The component peptide pools of each culture were then tested by IFNγ ELISpot.
PBMCs: Peripheral blood mononuclear cells, IL-2: Interleukin-2, IL-7: Interleukin-7

5.2 Patient OX1149

5.2.1 Clinical features

Patient OX1149 was diagnosed with AML aged 41. His disease was associated with an *NPM1* mutation and *FLT3* was wild type. Trisomy 4 was identified by karyotyping, which classifies his disease as intermediate risk according to European LeukemiaNet (ELN) criteria (Table A.1) [36]. He responded well to first-line chemotherapy. Monitoring of *NPM1* mutant transcripts from bone marrow biopsies was performed using RT qPCR, which is a highly sensitive method allowing the detection of low levels of AML. After completing chemotherapy, increasing levels of *NPM1* mutant transcripts were detected, indicating impending clinical relapse. Although an allo-SCT was not initially planned, due to his molecular relapse, he

was treated with salvage chemotherapy and a sibling allo-SCT (Table A.2).

The patient experienced limited oral GvHD, which was successfully treated with steroid mouthwash in addition to routine systemic immunosuppression with ciclosporin. All immunosuppression was stopped 6 months post-transplant and the patient remains alive and relapse-free after 3.5 years (Table A.3).

5.2.2 Screening for alloreactive peptides

Cultured IFNγ ELISpot screening was performed for 399 predicted HLA-binding peptides in 15 pools. High background signal was evident in the DMSO negative control from culture 1 and a similar level of signal was seen with CEF (CMV/EBV/Flu) peptides and all peptide pools tested using cells from this culture (Figure 5.2, upper panel). Therefore, responses to these pools were interpreted as negative. However, responses above background were detected for pools 7 and 10 across 3 technical replicates (Figure 5.2, lower panel).

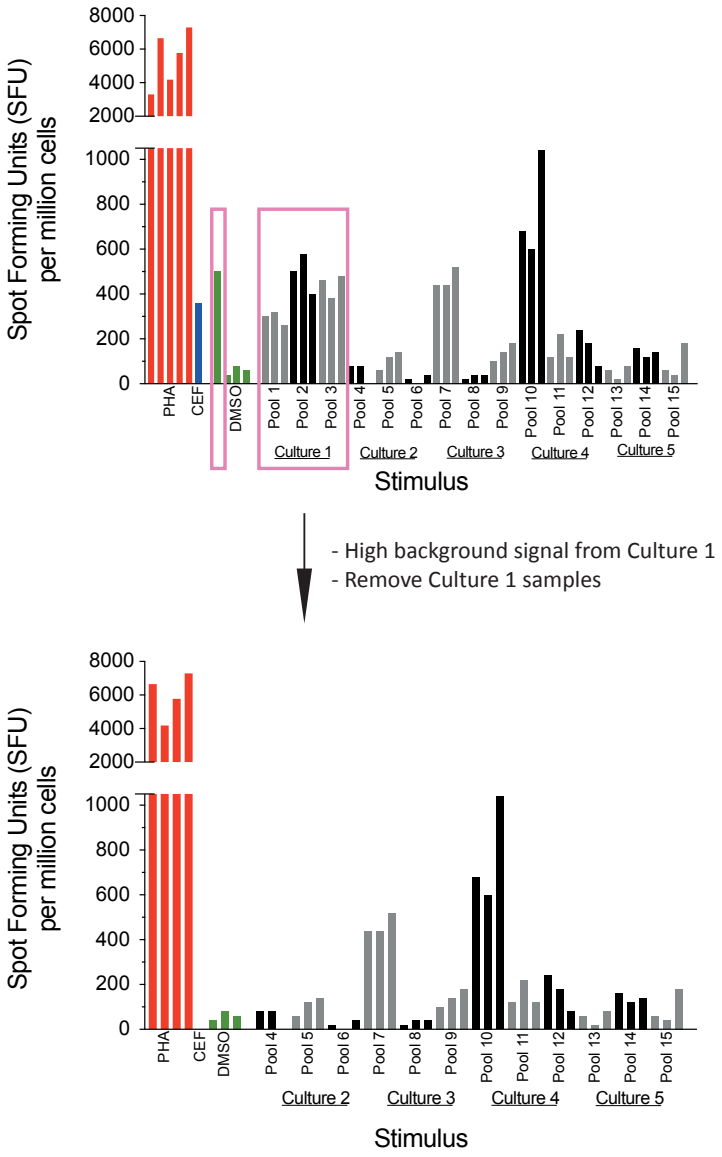
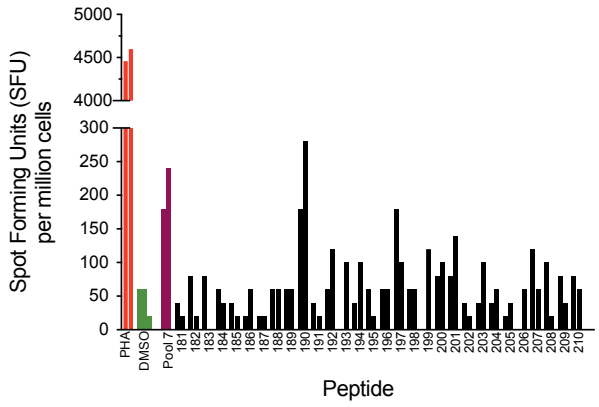


Figure 5.2: OX1149 pooled peptide screening by IFNγ ELISpot. Pooled screening of peptides by IFNγ ELISpot. For each of the 5 cultures, a PHA (positive control) and DMSO (negative control) was included. CEF peptides were included for cultures 1 and 2. Each peptide pool was tested in triplicate. High background signal from Culture 1 was seen in the negative control (upper graph), with no pools from this culture above the background level. Therefore, Culture 1 samples were removed (lower graph). PHA:Phytohaemagglutinin, CEF: Cytomegalovirus / Epstein-Barr virus / Influenza, DMSO: Dimethyl sulfoxide

Testing the individual component peptides of positive pools allows the identification of peptides driving T cell responses. IFNγ ELISpot testing of peptides from pool 7 revealed responses to peptide 190 across 2 replicates (Figure 5.3 (a)). This peptide results from a heterozygous SNP in the *RNH1* gene, which is present in the patient but not his allo-SCT donor. This peptide was predicted to bind 2 of the patient's HLA-DQ molecules, each with the same DQA1 chain (Figure 5.3 (b)). Interestingly, peptide 191 overlaps peptide 190 and contains the same mismatched amino acid. However, as it did not elicit T cell responses, this suggests that the N-terminal amino acids in peptide 190 that are not present in peptide 191 are likely to be important for HLA binding (underlined in Figure 5.3 (c)).

(a) OX1149 Pool 7



Peptide 190
Heterozygous SNP in *RNH1*
Allele frequency 11.21%

Peptide	Amino acid sequence (variant aa)
OX1149-190	AASCEPLASVLRAKLTSRSS
Reference	AASCEPLASVLRAKPTSRSS

(b)

HLA Class II Allele	IC50 (nM)
HLA-DQA1*05:05/DQB1*06:03	77.23
HLA-DQA1*05:05/DQB1*03:01	121.67

(c)

Peptide	Amino acid sequence (variant aa)
OX1149-190	AASCEPLASVLRAKLTSRSS
OX1149-191	VLRAKLTSRSSRLATTTSMR

Figure 5.3: OX1149 Pool 7 IFNγ ELISpot screening. **(a)** Screening found responses to peptide 190 across 2 technical replicates. The population allele frequency is an average of all samples in the gnomAD database [233]. Amino acid sequences for the peptide and reference are shown, with the variant amino acid highlighted in red. **(b)** Patient-specific HLA molecules that were predicted to bind peptide 190. Predicted IC_{50} < 150 nM indicates high affinity binding. **(c)** Amino acid sequences of overlapping peptides 190 and 191, with the variant amino acid highlighted in red. Only peptide 190 elicited T cell responses. The amino acid sequence uniquely present in peptide 190 is underlined.

Alloreactive T cell responses are expected to be directed against amino acid sequences present in patient but not donor. To confirm that T cells are specifically recognising the patient-donor mismatched peptide, the reference control peptide

(present in the donor) was also tested and as expected, there were no responses (Figure 5.4). Immune tolerance in the donor will have eliminated any T cells targeting this reference sequence.

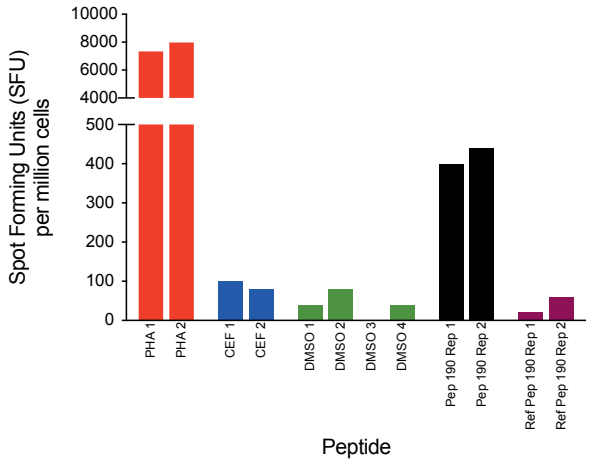
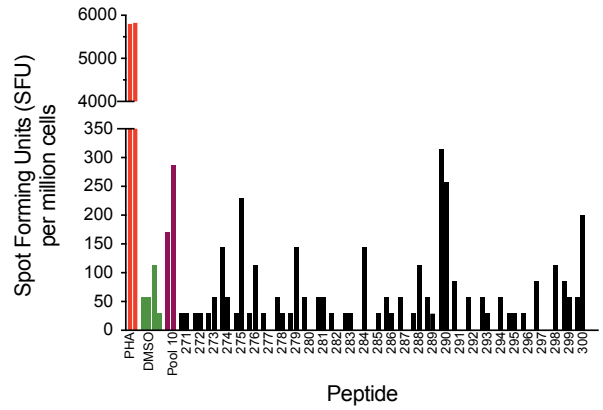


Figure 5.4: Peptide 190 reference control peptide. IFNγ ELISpot testing of peptide 190 (black bars) and donor reference peptide 190 (maroon bars), with 2 replicates. PHA: Phytohaemagglutinin, CEF: Cytomegalovirus / Epstein-Barr virus / Influenza, DMSO: Dimethyl sulfoxide

By testing peptides from pool 10, responses were found to peptide 290 across 2 replicates (Figure 5.5 (a)). This peptide results from a heterozygous SNP in the *SLC26A6* gene, which is present in the patient but not his allo-SCT donor. This peptide was predicted to bind 2 of the patient's HLA-DR molecules (Figure 5.5 (b)). In addition to peptide 290, 3 further peptides overlie the same variant amino acid and are the result of alternative transcripts (peptide 288, 289 and 291). However, these did not elicit T cell responses, suggesting that the amino acid sequence uniquely present in peptide 290 is likely to be important for HLA binding (underlined in Figure 5.5 (c)).

Testing the reference control (WT) sequence for peptide 290, which is present in the donor, unexpectedly elicited T cell responses (Figure 5.6). The responses were even were greater in magnitude to those seen with peptide 290. In the original screen of Pool 10 peptides, 4 peptides were tested that cover this genomic locus (peptides 288-291), from different transcripts and only peptide 290 elicited responses (Figure 5.5 (c)). However, the transcript (ENST00000414944) for peptide 290 does

(a) OX1149 Pool 10



Peptide 290
Heterozygous SNP in *SLC26A6*
Allele frequency 9.84%

Peptide	Amino acid sequence (variant aa)
OX1149-290	GAGPDPLRLHGHLPVRTSCP
Reference	GAGPDPLRLRGHLPVRTSCP

(b)

HLA Class II Allele	IC50 (nM)
HLA-DRB1*11:02	121.45
HLA-DRB1*13:01	121.45

(c)

Peptide	Amino acid sequence (variant aa)
OX1149-288	HISVGGAGPDPLRLHGHLPV
OX1149-289	WPLPGGAGPDPLRLHGHLPV
OX1149-290	GAGPDPLRLHGHLPVRTSCP
OX1149-291	GRPTGGAGPDPLRLHGHLPV

Figure 5.5: OX1149 Pool 10 IFNγ ELISpot screening. **(a)** Screening found responses to peptide 290 across 2 technical replicates. The population allele frequency is an average of all samples in the gnomAD database [233]. Amino acid sequences for the peptide and reference are shown, with the variant amino acid highlighted in red. **(b)** Patient-specific HLA molecules that were predicted to bind peptide 290. Predicted IC_{50} < 150 nM indicates high affinity binding. **(c)** Amino acid sequences of peptides resulting from alternative transcripts, with the variant amino acid highlighted in red. Only peptide 290 elicted T cell responses. The amino acid sequence uniquely present in peptide 290 is underlined.

not result in a protein due to nonsense-mediated decay (Ensembl). Therefore, the donor will not have undergone immune tolerance to the reference peptide, which could explain why a T cell response was detected.

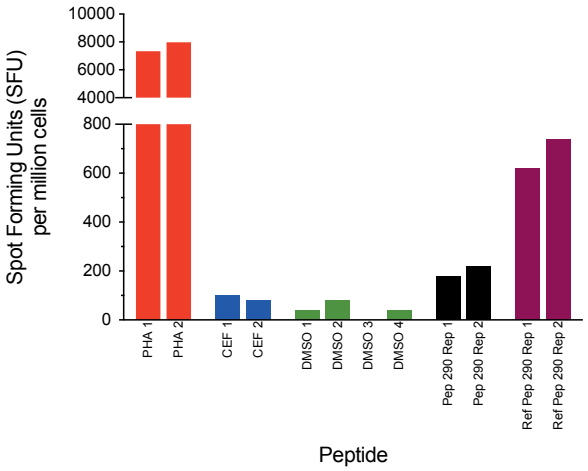


Figure 5.6: Peptide 290 reference control peptide elicits responses. IFNγ ELISpot testing of peptide 190 (black bars) and donor reference peptide 290 (maroon bars).

5.2.3 Characterising T cell responses to peptide 190

T cells responding to peptide 190 were phenotyped with flow cytometry using surface markers of activation and intracellular cytokine staining (ICS). In comparison to control cells that were not treated with peptide, a high frequency $CD4^+$ T cell response was detected after 6 hours of peptide 190 stimulation, with elevated expression of IFNγ, TNFα, IL-2 and CD107a on responding cells (Figure 5.7). Although surface CD25 (IL-2 receptor α chain) expression was detected on a high proportion (68.34%) of $CD4^+$ T cells, a high background level of $CD25^+$ cells was detected in unstimulated cells, which is likely to be due to upregulation in response to IL-2 in the culture media [285]. $CD8^+$ T cells did not upregulate expression of activation markers in response to peptide stimulation.

As expected, the reference (WT, wild type) peptide sequence did not elicit detectable responses by flow cytometry (Figure 5.7 (b)), which is in accordance with the IFNγ ELISpot results (Figure 5.4).

IFNγ ELISpot with blockade of HLA showed abrogation of responses only in wells containing either a pan-HLA class II blocking antibody (Tü39) or HLA-DR blocking antibody (L243), indicating a HLA-DR restricted response (Figure 5.8). The finding

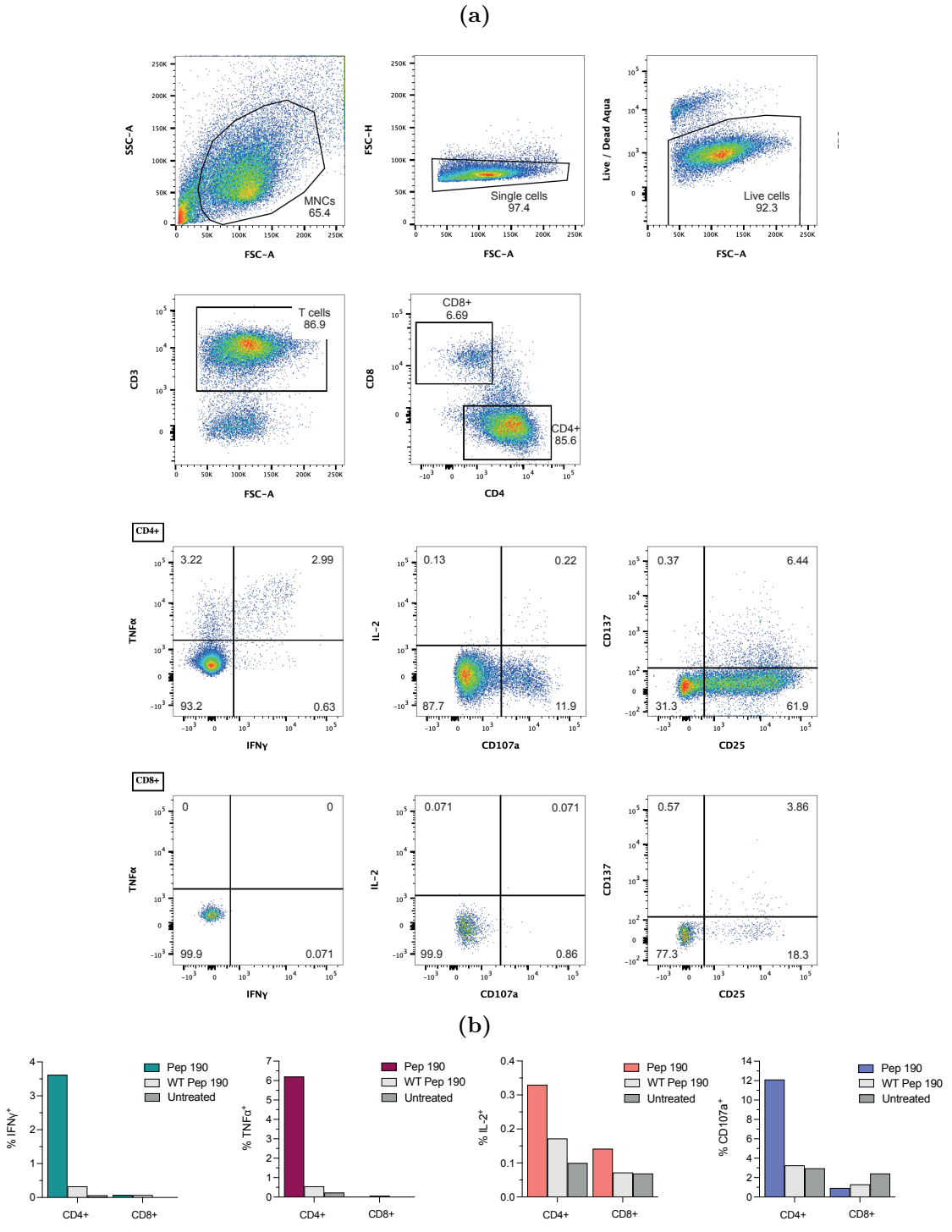


Figure 5.7: Peptide 190 stimulates CD4$^+$ T cells. (a) Flow cytometry plots showing T cells stimulated with peptide 190. The cells used in the assay were cultured for 12 days with peptide 190, IL-2 and IL-7. They were then washed and rested for 24 hours before stimulation. Live CD3$^+$CD4$^+$ and CD3$^+$CD8$^+$ cells were gated and activation markers were assessed for each cell type. Numbers on the plots are % of the parent population. **(b)** Summary of flow cytometry data, showing the % of CD4$^+$ and CD8$^+$ T cells expressing the indicated marker in response to peptide 190 (coloured bars), reference (WT) peptide (white bars) or no treatment (grey bars).
WT: Wild type

of a HLA class II restricted response is in keeping with immunophenotyping results that showed that only $CD4^+$ T cells were responding to peptide 190.

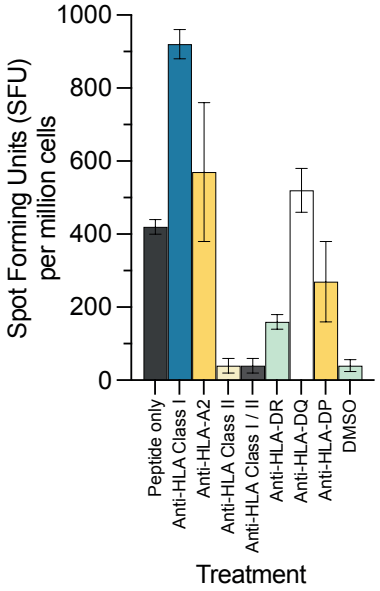


Figure 5.8: T cell responses to peptide 190 are HLA-DR restricted. IFNγ ELISpot with blockade of HLA molecules. HLA blocking antibodies used were: W6/32 (anti-HLA class I), BB7.2 (anti-HLA-A2), Tü39 (anti-HLA class I / II), L243 (anti-HLA-DR), Tü169 (anti-HLA-DQ) and B7/21 (anti-HLA-DP). DMSO is the negative control without peptide or blocking antibody. Error bars show mean +/- SEM of 2 replicates. SEM: Standard error of the mean

5.2.4 Characterising T cell responses to peptide 290

Flow cytometry demonstrated that responses to peptide 290 are driven by $CD4^+$ T cells (Figure 5.9). In comparison with control cells that were not treated with peptide, a $CD4^+$ T cell response was detected after 6 hours of peptide 290 stimulation, with elevated expression of IFNγ and TNFα. Expression levels of IL-2 and CD107a were similar to the untreated cells. $CD8^+$ T cells did not upregulate expression of activation markers in response to peptide stimulation. Consistent with the ELISpot results, reference control peptide sequences also elicited responses.

IFNγ ELISpot with blockade of HLA showed abrogation of responses in the presence of either a pan-HLA class II blocking antibody (Tü39) or a HLA-DR blocking antibody (L243), indicating a HLA-DR restricted response (Figure 5.10). The finding

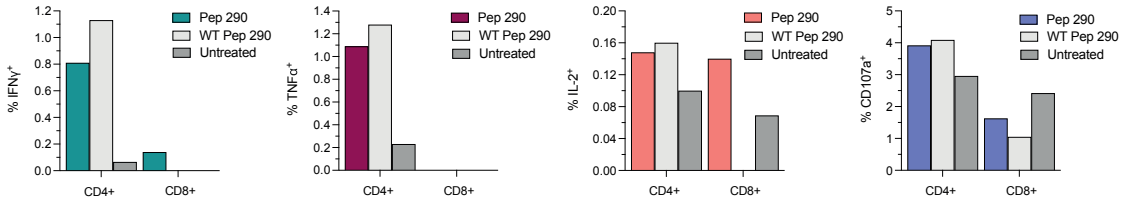


Figure 5.9: Peptide 290 stimulates CD4$^+$ T cells. Summary of flow cytometry data, showing the proportion of CD4$^+$ and CD8$^+$ T cells (%) expressing the indicated marker in response to peptide 290 (coloured bars), reference (WT) peptide (white bars) or no treatment (grey bars).

of a HLA class II restricted response is in keeping with the immunophenotyping results, which showed that only CD4$^+$ T cells were responding to peptide 290.

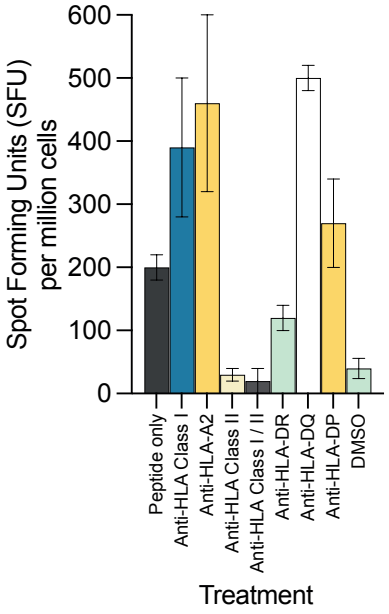


Figure 5.10: T cell responses to peptide 290 are HLA-DR restricted. IFNγ ELISpot with blockade of HLA molecules. HLA blocking antibodies used: W6/32 (anti-HLA class I), BB7.2 (anti-HLA-A2), Tü39 (anti-HLA class I / II), L243 (anti-HLA-DR), Tü169 (anti-HLA-DQ) and B7/21 (anti-HLA-DP). DMSO is the negative control without peptide or blocking antibody. Values shown are mean +/- SEM.
SEM: Standard error of the mean

5.3 Patient OX289

5.3.1 Clinical features

Patient OX289 was diagnosed with AML secondary to myelodysplastic syndrome (MDS) aged 52. Her disease was associated with a normal karyotype and an absence

of *NPM1* and *FLT3* mutations, which classified her disease genetic risk as intermediate (Table A.1) [36]. After first-line chemotherapy (daunorubicin / cytarabine), her normal blood cells failed to regenerate, rendering her dependent on blood transfusions. Bone marrow biopsy showed an absence of normal haematopoietic cells (aplasia), which meant that further planned cycles of chemotherapy could not be given due to the risk of excess toxicity to normal bone marrow. She had a HLA-matched sibling donor allo-SCT while in remission from AML (Table A.2).

The patient experienced significant chronic GvHD affecting the mouth, gastrointestinal tract and liver and was treated with systemic immunosuppression for 17 months following allo-SCT. Betamethasone steroid mouthwash was continued for 29 months (Table A.3). She remains disease-free 8 years following her transplant and therefore is very likely to have been cured.

5.3.2 Screening for alloreactive peptides

Cultured IFNγ ELISpot screening was performed for 298 peptides in 12 pools. A large number of pools elicited T cell responses across 3 replicate ELISpot wells: pools 1, 2, 3, 6, 7, 8, 10 and 11, with particularly strong responses to pools 1 and 10 (Figure 5.11). The DMSO negative control responses were low. Although CEF (CMV/EBV/Flu) peptides (Mabtech) elicit responses from almost 90% caucasians, responses were absent for Patient OX289. Testing the component peptides from positive pools revealed 10 peptides that produce T cell responses.

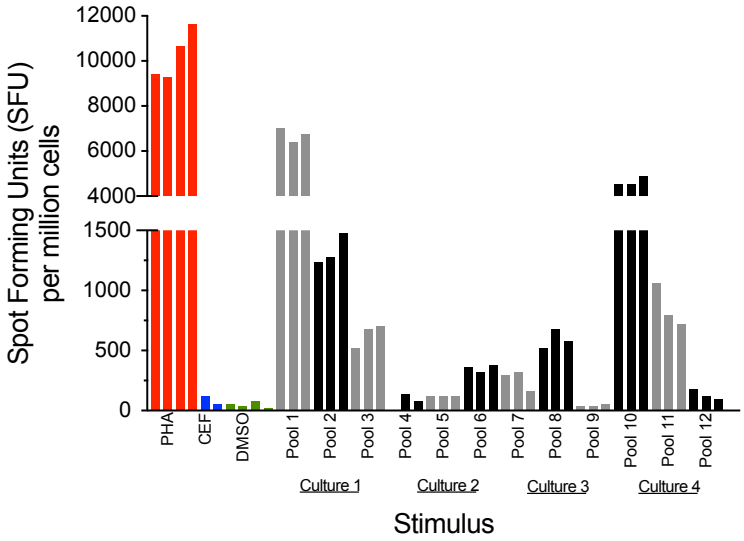
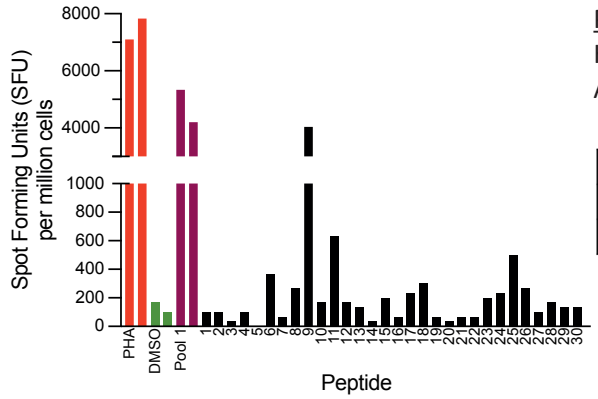


Figure 5.11: OX289 pooled peptide screening by IFNγ ELISpot. Pooled screening of peptides by IFNγ ELISpot, using 3 technical replicates per pool. For each of the 5 cultures, a PHA (positive control) and DMSO (negative control) was included. A CEF (CMV/EBV/Flu) peptide pool was tested for cultures 1 and 2.

Testing the individual peptides from pool 1 revealed a strong response to peptide 9, of similar magnitude to the pooled testing (Figure 5.12 (a)). This peptide results from a heterozygous SNP in the *NAGPA* gene, which is present in the patient but not her allo-SCT donor. This peptide was predicted to bind 1 of the patient's HLA-A molecules and 3 HLA-DR molecules (Figure 5.12 (b)).

(a) OX289 Pool 1



Peptide 9
Heterozygous SNP in *NAGPA*
Allele frequency 30.07%

Peptide	Amino acid sequence (variant aa)
OX289-009	ALAFLLLISIAANLSLLLSR
Reference	ALAFLLLISTAANLSLLLSR

(b)

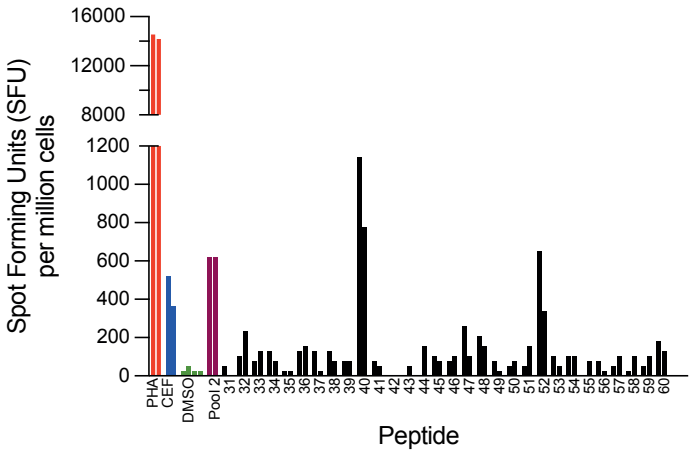
HLA Class I Allele	Rank affinity (%)
HLA-A*02:05	0.3155
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	96.59
HLA-DRB3*02:02	81.17
HLA-DRB4*01:01	73.71

Figure 5.12: OX289 Pool 1 IFNγ ELISpot screening. (a) A response was detected to peptide 9. Due to limited cell numbers, no replicates were performed. Amino acid sequences for the peptide and reference are shown, with the variant amino acid highlighted in red. The population allele frequency is an average of all samples in the gnomAD database [233]. **(b)** Patient-specific HLA molecules that were predicted to bind peptide 9. For class I, rank affinity $< 0.5\%$ indicates high affinity binding. For class II, predicted $IC_{50} < 150$ nM indicates high affinity binding.

From testing the individual peptides of pool 2, peptides 40 and 52 each elicited responses across replicates (Figure 5.13 (a)). Peptide 40 results from a heterozygous SNP in the *RNF123* gene, which is present in the patient but not her allo-SCT donor (Figure 5.13 (b)). This peptide was predicted to bind 1 of the patient's HLA-B molecules and 2 HLA-DR molecules (Figure 5.13 (c)).

Peptide 52 is the result of a heterozygous SNP in *LLGL2*, which has a high allele frequency (50.19%) (Figure 5.13 (d)). This peptide was predicted to bind 2 of the patient's HLA class I molecules and 3 HLA-DR molecules, (Figure 5.13 (e)). Peptide 51 overlaps peptide 52 and contains the same mismatched amino acid. However, as it did not elicit T cell responses, this suggests that the C-terminal amino acids in peptide 52 that are not present in peptide 51 are likely to be important for HLA binding (underlined in Figure 5.13 (f)).

(a) OX289 Pool 2



(b) Peptide 40
Heterozygous SNP in *RNF123*
Allele frequency 5.07%

Peptide	Amino acid sequence (variant aa)
OX289-040	EVQDCLKQLMMSLLQLYRFS
Reference	EVQDCLKQLMMSLLRLYRFS

(c)

HLA Class I Allele	Rank affinity (%)
HLA-B*58:01	0.2234
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	48.45
HLA-DRB4*01:01	46.82

(d) Peptide 52
Heterozygous SNP in *LLGL2*
Allele frequency 50.19%

Peptide	Amino acid sequence (variant aa)
OX289-052	GYSPSLHILAIGTRSGAIKL
Reference	GYSPSLRILAIGTRSGAIKL

(e)

HLA Class I Allele	Rank affinity (%)
HLA-A*23:01	0.1518
HLA-C*04:01	0.3086
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*03:01	125.89
HLA-DRB1*07:01	15.8
HLA-DRB3*02:02	54.36

(f)

Peptide	Amino acid sequence (variant aa)
OX289-051	PHQPSALGYSPSLHILAIGT
OX289-052	GYSPSLHILAIGTRSGAIKL

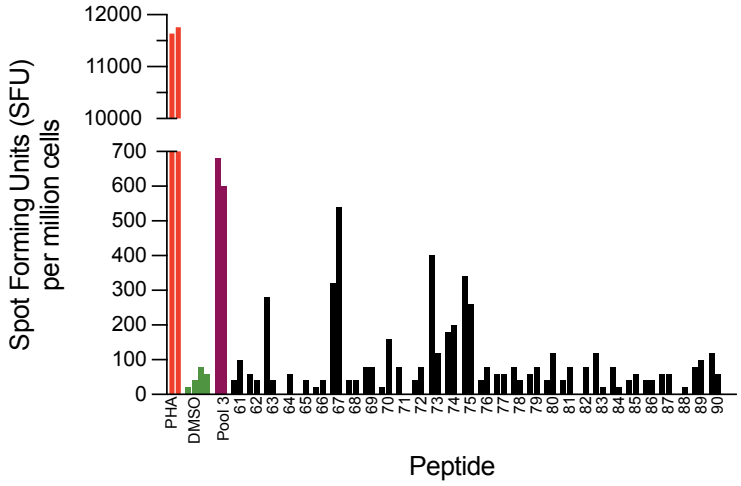
Figure 5.13: OX289 Pool 2 IFNγ ELISpot screening. **(a)** Responses were detected to peptides 40 and 52 (2 replicates each). **(b)** Amino acid sequences for peptide 40 and the reference are shown, with the variant amino acid highlighted in red. **(c)** Patient-specific HLA molecules that were predicted to bind peptide 40. **(d)** Amino acid sequences for peptide 52 and the reference are shown, with the variant amino acid highlighted in red. **(e)** Patient-specific HLA molecules that were predicted to bind peptide 52. **(f)** Amino acid sequences of overlapping peptides 51 and 52, with the variant amino acid highlighted in red. Only peptide 52 elicited T cell responses and its unique amino acid sequence is underlined.
Population allele frequencies are from gnomAD [233]. For class I, predicted rank affinity $< 0.5\%$ indicates high affinity binding. For class II, predicted $IC_{50} < 150$ nM indicates high affinity binding.

Testing the individual peptides from pool 3 revealed responses to 3 peptides: peptide 67, overlapping peptides 73 / 74, and peptide 75 (Figure 5.14 (a)). Peptide 67 results from a heterozygous SNP in the *ENDOD1* gene, which is present in the patient but not her allo-SCT donor (Figure 5.14 (b)). This peptide was predicted to bind 1 of the patient's HLA-A molecules and 1 HLA-DR molecule (Figure 5.14 (c)).

Overlapping peptides 73 and 74 result from a heterozygous SNP in *HENMT1* and were predicted to bind to multiple patient-specific HLA molecules (Figure 5.14 (d) and (e)).

Peptide 75 results from a heterozygous SNP in *ACADS* and was predicted to bind to 1 of the patient's HLA-C molecules and 2 HLA-DR molecules (Figure 5.14 (f) and (g)). Peptide 75 overlaps peptide 76 and contains the same mismatched amino acid. However, as it did not elicit T cell responses, this suggests that the N-terminal amino acids in peptide 75 that are not present in peptide 76 are likely to be important for HLA binding (underlined in Figure 5.14 (h)).

(a) OX289 Pool 3



(b) Peptide 67
Heterozygous SNP in *ENDOD1*
Allele frequency 25.36%

Peptide	Amino acid sequence (variant aa)
OX289-067	VATFPVYTMVAIPIVCKD
Reference	VATFPVYTMGAIPIVCKD

(c)

HLA Class I Allele	Rank affinity (%)
HLA-A*02:05	0.1551
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	113.75

(d) Peptides 73/74
Heterozygous SNP in *HENMT1*
Allele frequency 1.93%

Peptide	Amino acid sequence (variant aa)
OX289-073	PPLYRQRYQFIKNLVDQHEP
OX289-074	PLYRQRYQFIKNLVDQHEPK
Reference	PPLYRQRYQFVKNLVDQHEPK

(e)

HLA Class I Allele	Rank affinity (%)
HLA-A*23:01	0.3011
HLA-C*07:01	0.0195
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	134.41
HLA-DRB3*02:02	95.48
HLA-DRB4*01:01	81.95

(f) Peptide 75
Heterozygous SNP in *ACADS*
Allele frequency 26.58%

Peptide	Amino acid sequence (variant aa)
OX289-075	VSRPELLRESISAFLVPMPT
Reference	VSRPELLREGISAFLVPMPT

(g)

HLA Class I Allele	Rank affinity (%)
HLA-C*07:01	0.0195
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	42.00
HLA-DRB3*02:02	120.56

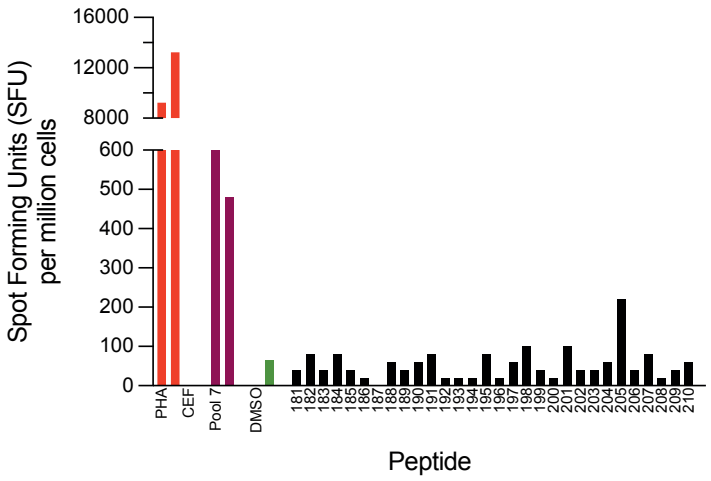
(h)

Peptide	Amino acid sequence (variant aa)
OX289-075	VSRPELLRESISAFLVPMPT
OX289-076	ELLRESISAFLVPMPTPGLT

Figure 5.14: OX289 Pool 3 IFNγ ELISpot screening. (a) Responses were detected to peptide 67, overlapping peptides 73/74 and peptide 75 (2 replicates each). **(b)** Amino acid sequences for peptide 67 and the reference are shown, with the variant amino acid highlighted in red. **(c)** Patient-specific HLA molecules that were predicted to bind peptide 67. **(d)** Amino acid sequences for peptides 73 /74 and the reference are shown, with the variant amino acid highlighted in red. **(e)** Patient-specific HLA molecules that were predicted to bind peptides 73/74. **(f)** Amino acid sequences for peptide 75 and the reference are shown, with the variant amino acid highlighted in red. **(g)** Patient-specific HLA molecules that are predicted to bind peptide 75. **(h)** Amino acid sequences of overlapping peptides 75 and 76, with the variant amino acid highlighted in red. Only peptide 75 elicited T cell responses and its unique amino acid sequence is underlined. Population allele frequencies are from gnomAD [233]. For class I, predicted rank affinity $< 0.5\%$ indicates high affinity binding. For class II, predicted $IC_{50} < 150$ nM indicates high affinity binding.

Testing the component peptides of pool 7 demonstrated responses to peptide 205. Interestingly, this results from the same SNP in *ACADS* that produced peptide 75, but is due to an alternate transcript (Figure 5.15 (a) and (b)). The C-terminal amino acids of the peptides overlap (underlined in Figure 5.15 (d)), whereas the N-terminal sequence differs, suggesting that the C-terminal amino acids are likely to be important for peptide binding.

(a) OX289 Pool 7



(b) Peptide 205
Heterozygous SNP in *ACADS*
Allele frequency 26.58%

Peptide	Amino acid sequence (variant aa)
OX289-205	TDRALQNKSISAFLVPMPTP
Reference	TDRALQNKGISAFLVPMPTP

(c)

HLA Class II Allele	IC50 (nM)
HLA-DRB1*07:01	74.17

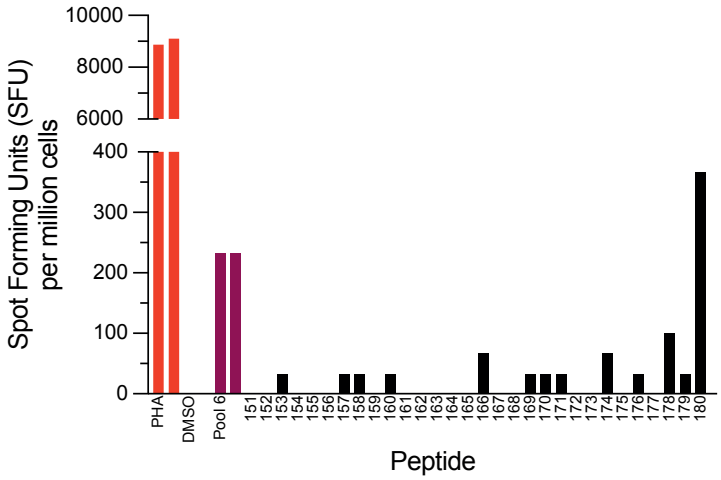
(d)

Peptide	Amino acid sequence (variant aa)
OX289-205	TDRALQNKSISAFLVPMPTP
OX289-075	VSRPELLRESISAFLVPMPT

Figure 5.15: OX289 Pool 7 IFNγ ELISpot screening. **(a)** Low frequency responses were detected to peptide 205. Due to limited cell numbers, no replicates were performed. **(b)** Amino acid sequences for the peptide and reference are shown, with the variant amino acid highlighted in red. The population allele frequency is an average of all samples in the gnomAD database [233]. **(c)** Patient-specific HLA-DR molecule that was predicted to bind peptide 205. Predicted $IC_{50} < 150$ nM indicates high affinity binding. **(d)** Amino acid sequences of overlapping peptides 205 and 75, which result from alternate transcripts and each produce T cell responses. Overlapping amino acids are underlined.

Testing the component peptides of pool 6 showed that responses were driven by peptide 180, which results from a heterozygous SNP in *TIMM23B* and is predicted to bind 1 of the patient's HLA-DR molecules (Figure 5.16).

(a) OX289 Pool 6



(b) Peptide 180

Homozygous SNP in *TIMM23B*

Allele frequency 9.11%

Peptide	Amino acid sequence (variant aa)
OX289-180	NPLSPYLNVDPRYLVQDT
Reference	NPLCPYLNVDPRYLVQDT

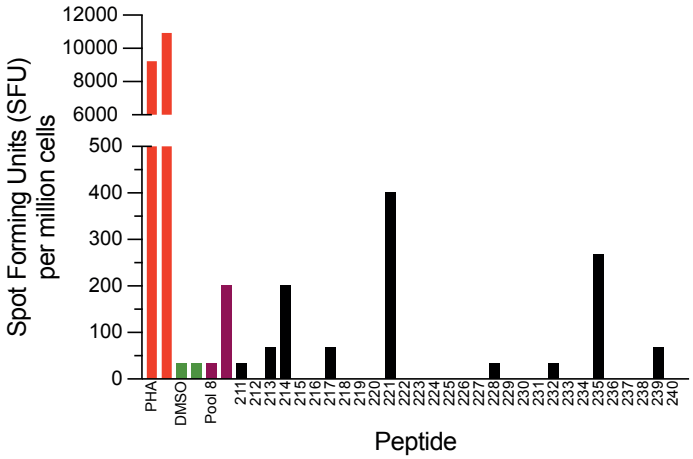
(c)

HLA Class II Allele	IC50 (nM)
HLA-DRB1*03:01	25.93

Figure 5.16: OX289 Pool 6 IFNγ ELISpot screening. **(a)** Responses were detected to peptide 180. Due to limited cell numbers, no replicates were performed. **(b)** Amino acid sequences for the peptide and reference are shown, with the variant amino acid highlighted in red. The population allele frequency is an average of all samples in the gnomAD database [233]. **(c)** Patient-specific HLA-DR molecule that was predicted to bind peptide 180. Predicted $IC_{50} < 150$ nM indicates high affinity binding.

IFNγ ELISpot screening of the component peptides of pools 8, 10 and 11 identified 3 further alloreactive responses to peptides 221, 284 and OX1098 peptide 614 (Figures 5.17, 5.18 and 5.19). For each of these responses, the allo-SCT donor was homozygous for the SNP and the patient was heterozygous. Each peptide driving T cell responses derived from the reference amino acid sequence, which is uniquely present in the patient. The allele frequency of these SNPs were high (20.48%, 64.78% and 72.36%). For each, there were overlapping peptides that did not elicit IFNγ ELISpot responses from donor-derived T cells, which helps to map the amino acids within long peptides that are likely to be important for HLA binding (underlined in **Figures** 5.17 (d), 5.18 (d) and 5.19 (d)).

(a) OX289 Pool 8



(b) Peptide 221
Heterozygous SNP in *LRR1*
Allele frequency 20.48%

Peptide	Amino acid sequence (reference aa)
OX289-221	EPPVDICLSKAISSSLKGFL
Variant	EPPVDICLSKANSSSLKGFL

(c)

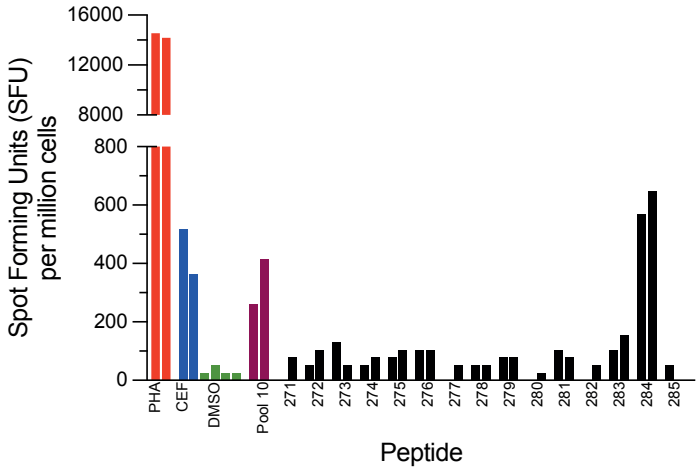
HLA Class I Allele	Rank affinity (%)
HLA-B*58:01	0.317
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	27.69
HLA-DRB3*02:02	144.43

(d)

Peptide	Amino acid sequence (reference aa)
OX289-221	EPPVDICLSKAISSSLKGFL
OX289-222	CLSKAISSSLKGFLSAMRLA

Figure 5.17: OX289 Pool 8 IFNγ ELISpot screening. **(a)** Responses were detected to peptide 221. Due to limited cell numbers, no replicates were performed. **(b)** The amino acid sequence for the peptide is shown. The allo-SCT donor is homozygous for the variant sequence. The population allele frequency is an average of all samples in the gnomAD database [233]. **(c)** Patient-specific HLA molecules that were predicted to bind peptide 221. For class I, rank affinity $< 0.5\%$ indicates high affinity binding. For class II, predicted $IC_{50} < 150$ nM indicates high affinity binding. **(d)** Amino acid sequences of overlapping peptides 221 and 222, with the variant amino acid highlighted in red. Only peptide 221 elicited T cell responses and its unique amino acid sequence is underlined.

(a) OX289 Pool 10



(b) Peptide 284
Heterozygous SNP in *FTSJ3*
Allele frequency 64.78%

Peptide	Amino acid sequence (reference aa)
OX289-284	GETGMFSL**S**TIRGHQYATY
Variant	GETGMFSL**C**TIRGHQYATY

(c)

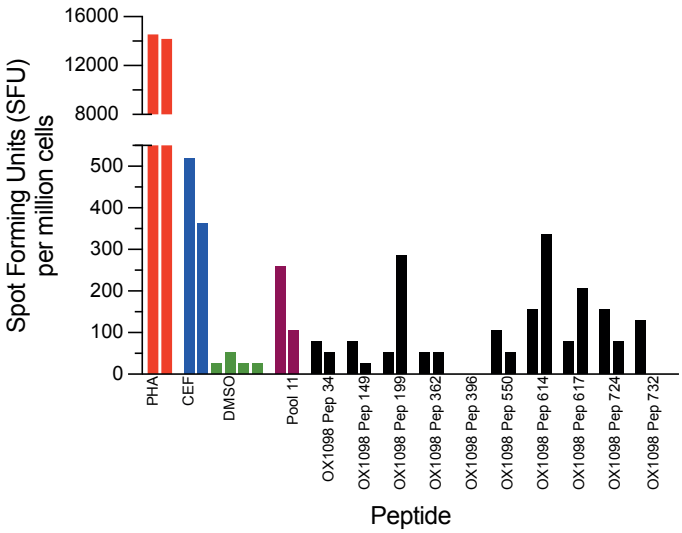
HLA Class II Allele	IC50 (nM)
HLA-DRB1*07:01	46.36

(d)

Peptide	Amino acid sequence (reference aa)
OX289-284	GETGMFSL**S**TIRGHQYATY
OX289-271	GETGMFSL**S**TIRGHQLLEEV

Figure 5.18: OX289 Pool 10 IFNγ ELISpot screening. **(a)** Responses were detected to peptide 284 (2 replicates). **(b)** The amino acid sequence for the peptide is shown. The allo-SCT donor is homozygous for the variant sequence. The population allele frequency is an average of all samples in the gnomAD database [233]. **(c)** Patient-specific HLA molecules that were predicted to bind peptide 284. For class II, predicted $IC_{50} <$ 150 nM indicates high affinity binding. **(d)** Amino acid sequences of overlapping peptides 284 and 271, with the variant amino acid highlighted in red. Only peptide 284 elicited T cell responses and its unique amino acid sequence is underlined.

(a) OX289 Pool 11



(b) OX1098 Peptide 614

Heterozygous SNP in *LGALS8*

Allele frequency 72.36%

Peptide	Amino acid sequence (reference aa)
OX1098-614	MNYVSKRLPFAARLNTPMGP
Variant	MNYVSKSLPFAARLNTPMGP

(c)

HLA Class I Allele	Rank affinity (%)
HLA-A*23:01	0.3691
HLA-B*58:01	0.3921
HLA-C*07:01	0.0355
HLA Class II Allele	**IC50 (nM)**
HLA-DRB1*07:01	78.58

(d)

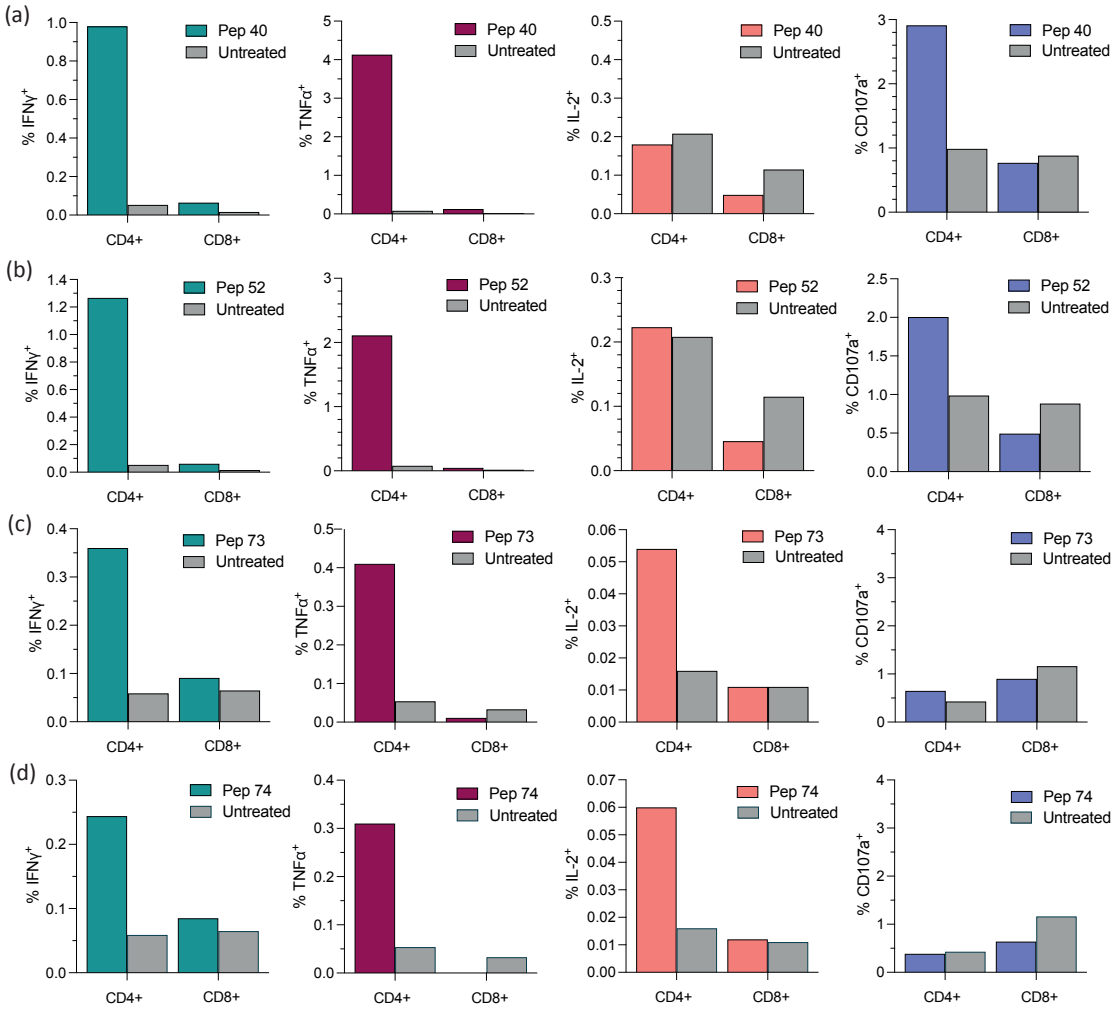
Peptide	Amino acid sequence (reference aa)
OX1098-614	MNYVSKRLPFAARLNTPMGP
OX289-241	PPMNYVSKRLPFAARLNTPM
OX289-285	SGTPQLRLPFAARLNTPMGP (alt transcript)

Figure 5.19: OX289 Pool 11 IFNγ ELISpot screening. **(a)** Responses were detected to peptide OX1098-614 (2 replicates). **(b)** The amino acid sequence for the peptide is shown. The allo-SCT donor is homozygous for the variant sequence. The population allele frequency is an average of all samples in the gnomAD database [233]. **(c)** Patient-specific HLA molecules that were predicted to bind peptide OX1098-614. For class I, rank affinity $< 0.5\%$ indicates high affinity binding. For class II, predicted $IC_{50} <$ 150 nM indicates high affinity binding. **(d)** Amino acid sequences of overlapping peptides OX1098-614, OX289-241 and OX289-285, with the variant amino acid highlighted in red. Only peptide OX1098-614 elicited T cell responses and its unique amino acid sequence is underlined.

5.3.3 Immunophenotyping of antigen-specific T cell responses

Of the 10 peptides from Patient OX289 that elicited alloreactive T cell responses, 7 peptides stimulated donor-derived $CD4^+$ T cells (Figure 5.20) and 2 stimulated $CD8^+$ T cells (Figure 5.21). The final peptide (OX1098 peptide 614) was not evaluable because it did not elicit detectable responses by flow cytometry.

The cytokines IFNγ and TNFα were the most useful markers for identifying activated T cells. For all peptides, more T cells expressed these cytokines compared with untreated cells. IL-2 was only expressed in response to some peptides (overlapping peptides 73/74 and 75/205, as well as peptides 180, 9 and 67) and in each case the % cells expressing IL-2 was lower than those expressing IFNγ, TNFα and CD107a.



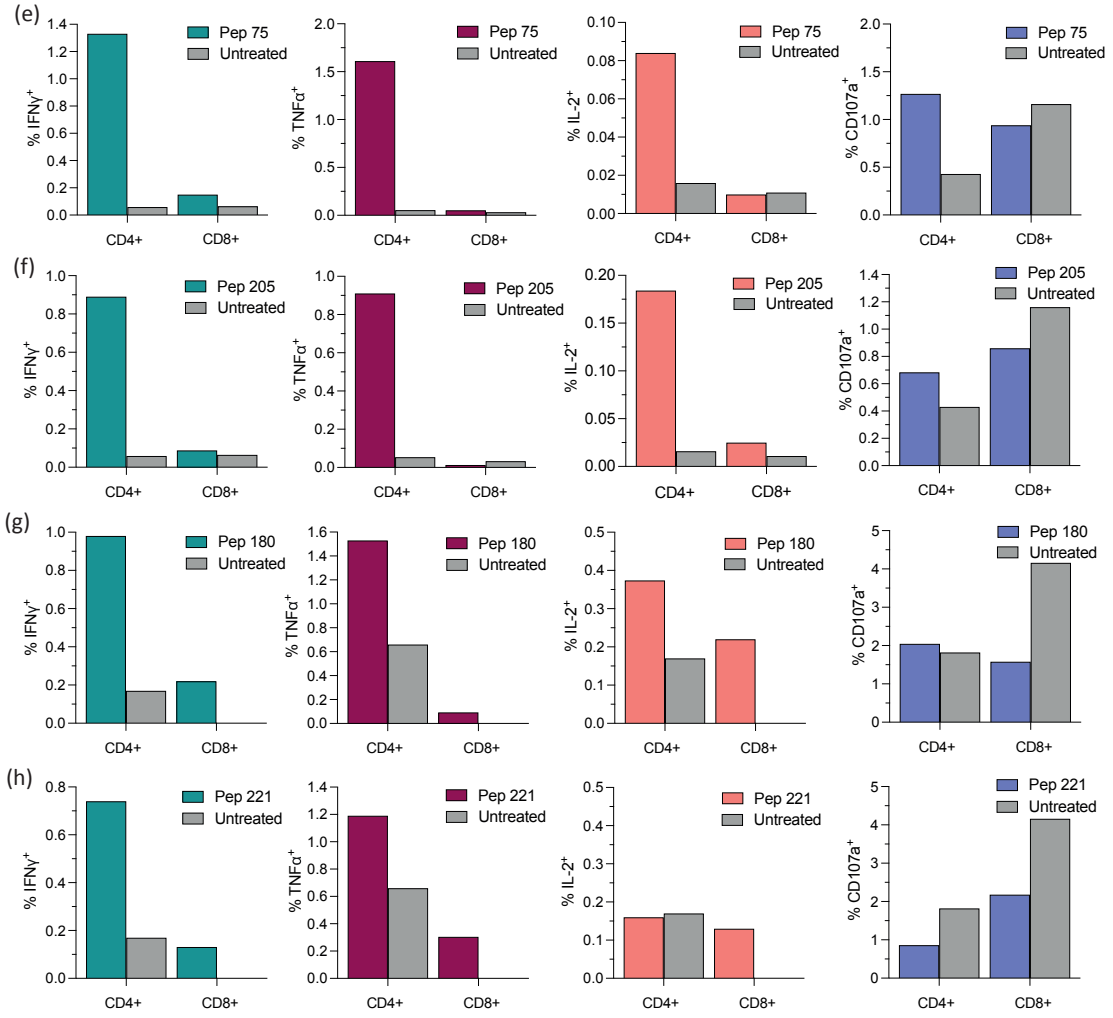


Figure 5.20: Immunophenotyping of CD4$^+$ T cell responses. (a) - (h) Immunophenotyping using intracellular cytokine staining. Live CD4$^+$ and CD8$^+$ T cells were gated and activation markers were assessed. Expression of IFNγ, TNFα, IL-2 and CD107a in CD4$^+$ and CD8$^+$ T cells are shown. For each graph, the coloured bars indicate % of CD4$^+$ or CD8$^+$ T cells expressing the activation marker in response to peptide and grey bars indicate cells expressing the marker in the absence of peptide.

Peptides 9 (pool 1) and 284 (pool 10) both elicited very high frequency T cell responses when tested by IFNγ ELISpot (Figure 5.11) and these strong responses were reproduced when assessed by immunophenotyping (Figure 5.21 (a) and (c)). The finding of a CD8$^+$ T cell response is in keeping with these large responses, as CD8$^+$ T cell clones are often larger than CD4$^+$ clones [281, 282].

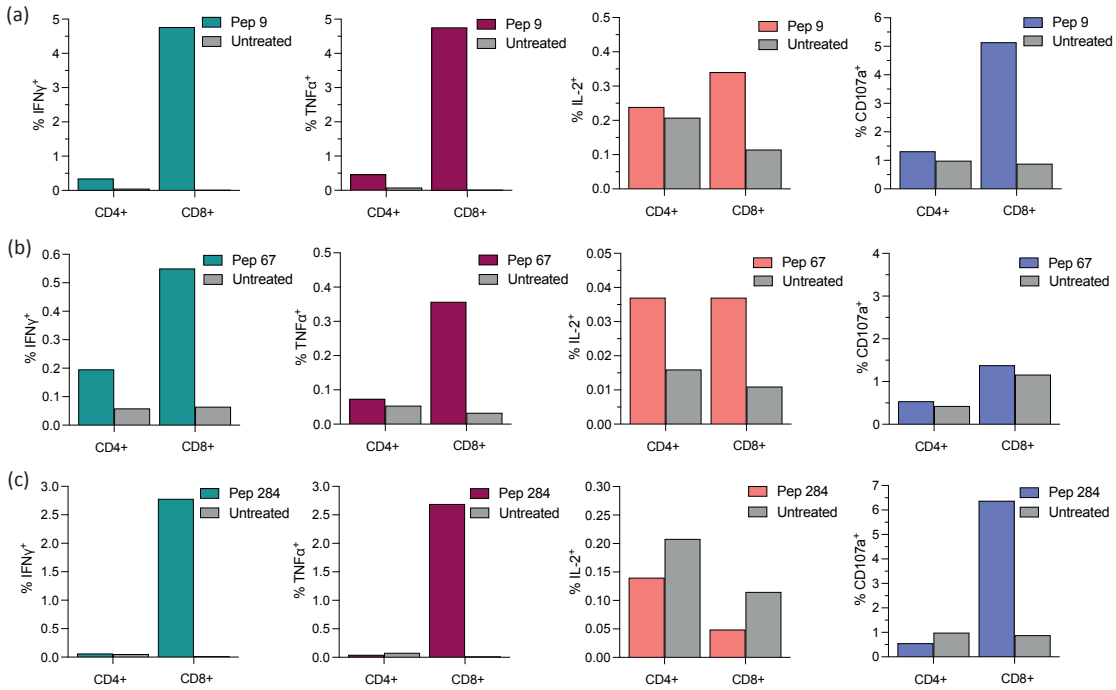


Figure 5.21: Immunophenotyping of CD8$^+$ T cell responses. **(a) - (c)** Immunophenotyping using intracellular cytokine staining. Live CD4$^+$ and CD8$^+$ T cells were gated and activation markers were assessed. Expression of IFNγ, TNFα, IL-2 and CD107a in CD4$^+$ and CD8$^+$ T cells are shown. For each graph, the coloured bars indicate % of CD4$^+$ or CD8$^+$ T cells expressing the activation marker in response to peptide and grey bars indicate cells expressing the marker in the absence of peptide.

5.3.4 HLA restriction of peptide responses

The HLA allele restriction of T cell responses was assessed for 7 peptide antigens from Patient OX289, using IFNγ ELISpot with HLA blocking antibodies. For each of the peptides that elicited CD4$^+$ T cell responses, HLA blockade suggested restriction either by HLA-DR (peptides 52 and 180) or HLA-DP (peptides 40, 221 and OX1098-614) (Figure 5.22). However, blockade with the pan-class II antibody (Tü39)) did not effectively abrogate responses. This could be due to impaired blockade of the patient's specific HLA molecules. In some cases, responses were unexpectedly much stronger in the presence of HLA blocking antibodies and the reason for this is unclear.

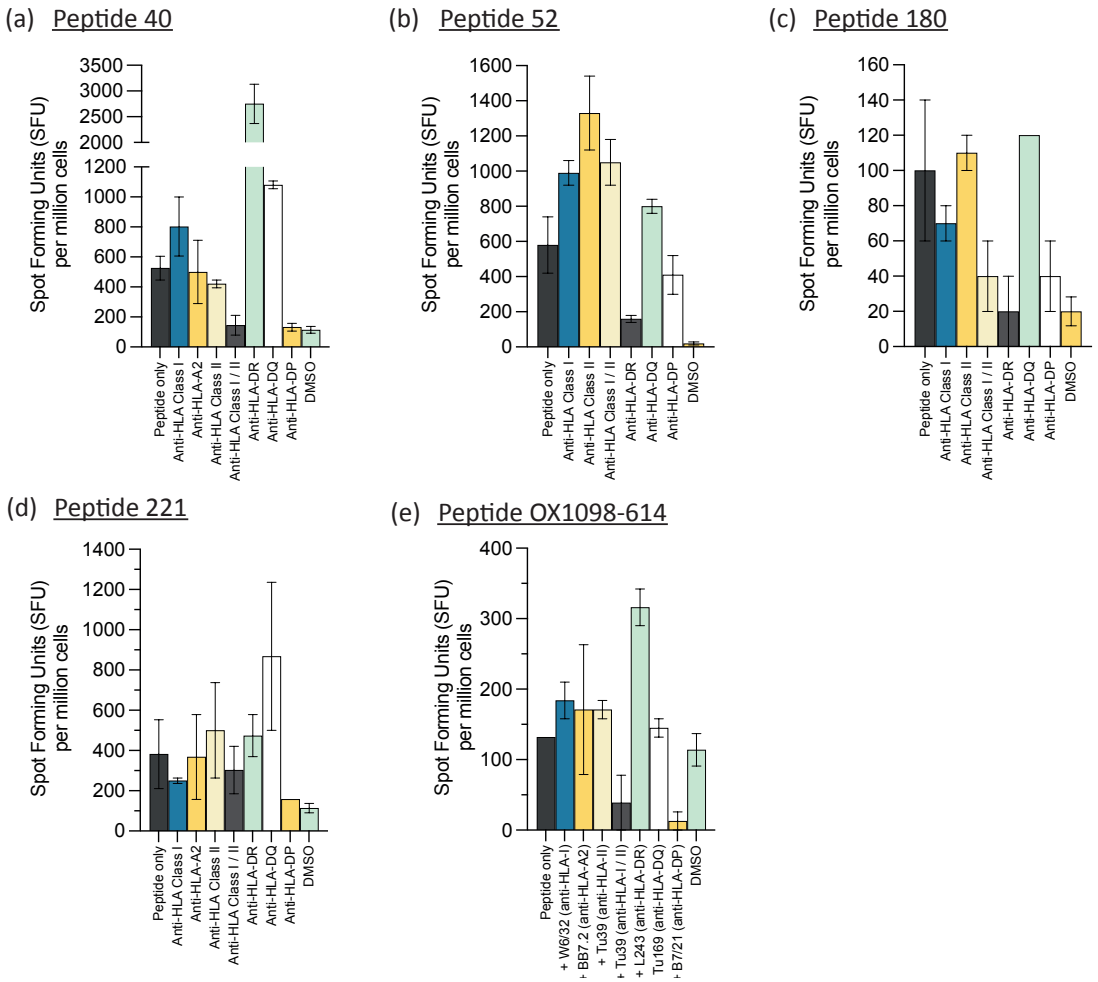


Figure 5.22: Assessing HLA restriction of CD4$^+$ T cell responses. (a) - (e) T cell responses to the indicated peptides were assessed by IFNγ ELISpot using anti-HLA antibodies to block responses mediated by particular HLA molecules. HLA blocking antibodies used: W6/32 (anti-HLA class I), BB7.2 (anti-HLA-A2), Tü39 (anti-HLA class I / II), L243 (anti-HLA-DR), Tü169 (anti-HLA-DQ) and B7/21 (anti-HLA-DP). DMSO is the negative control, without peptide or blocking antibody. Values are mean +/- SEM of 2 replicates.
SEM: standard error of the mean

Two CD8$^+$ T cell responses were assessed by HLA blockade. Responses to peptide 9 were reduced by class I blockade, but not completely blocked (Figure 5.23 (a)). No antibodies blocked responses to peptide 284 (Figure 5.23 (b)). As a positive control, responses to CMV / EBV / Influenza pooled peptides, which are known to elicit HLA class I-restricted responses, were tested. Responses were reduced by class I blockade but not completely eliminated (Figure 5.23 (c)).

In all cases, the likely HLA restriction was in keeping with the phenotype of

responding T cells. HLA class I responses were from $CD8^+$ T cells, whereas class II responses were mediated by $CD4^+$ T cells.

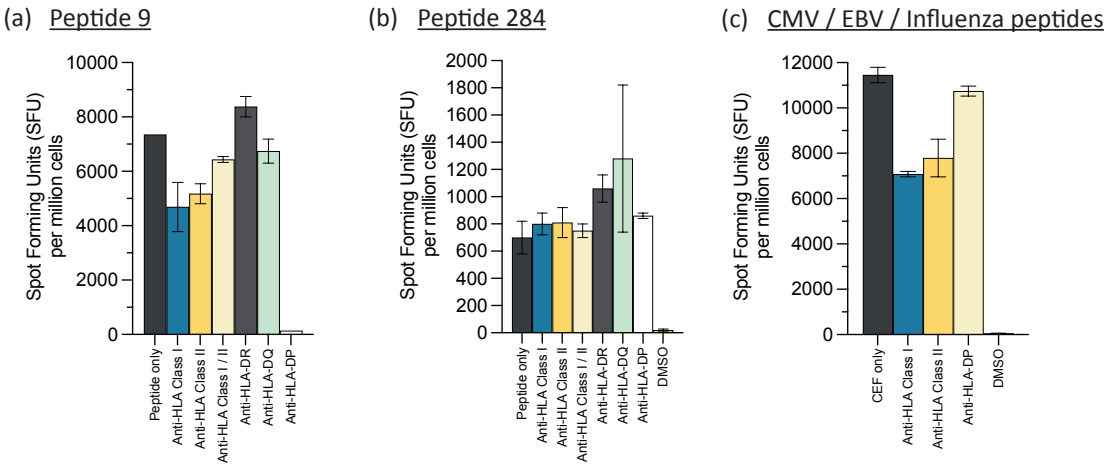


Figure 5.23: Assessing HLA restriction of $CD8^+$ T cell responses. **(a) - (c)** T cell responses to the indicated peptides were assessed by IFNγ ELISpot using anti-HLA antibodies to block responses mediated by particular HLA molecules. HLA blocking antibodies used: W6/32 (anti-HLA class I), Tü39 (anti-HLA class I / II), L243 (anti-HLA-DR), Tü169 (anti-HLA-DQ) and B7/21 (anti-HLA-DP). DMSO is the negative control, without peptide or blocking antibody. Values are mean +/- SEM of 2 replicates.
SEM: standard error of the mean

5.4 Summary of alloreactive peptides

For 10 of the 15 patients studied, all predicted HLA-binding peptides were functionally screened (5340 peptides) using IFNγ ELISpot and the nature of T cell responses were characterised. The data for 3 of these patients has been shown: OX866 (Chapter 4) OX1149 (Section 5.2) and OX289 (Section 5.3). In each case, post-transplant blood from the patient in whom the HLA-binding predictions were made was used to look for peptide-responsive donor-derived T cells.

Patient	Peptide(s)	Gene	Allele frequency (%)	CD4 / CD8 responses	HLA restriction	Freq. of response (% IFNγ$^+$ cells amongst CD4$^+$ or CD8$^+$ by FACS)
OX866	OX866-21/22	PADI4	4.27	CD4	HLA-DP	1.1
OX1149	OX1149-190	RNH1	11.21	CD4	HLA-DR	3.62
	OX1149-290	SLC26A6	9.84	CD4	HLA-DR	0.81
OX813	OX813-101/102	RHOT2	51.51	Not known	Not known	Not detected
OX802	OX802-321/322	COA6	52.52	CD4	Not tested	5.47/3.23
OX289	OX289-9	NAGPA	30.07	CD8	Class I	4.77
	OX289-040	RNF123	5.07	CD4	HLA-DP	0.98
	OX289-052	LLGL2	50.19	CD4	HLA-DR	1.27
	OX289-067	ENDOD1	25.36	CD8	Not tested	0.55
	OX289-073/074	HENMT1	1.93	CD4	Not tested	0.36/0.24
	OX289-075/205	ACADS	26.58	CD4	Not tested	1.33/0.89
	OX289-180	TIMM23B	9.11	CD4	HLA-DR	0.98
	OX289-221	LRR1	20.48	CD4	HLA-DP	0.74
	OX289-284	FTSJ3	64.78	CD8	Class I	2.78
	OX1098-614	LGALS8	72.36	Likely CD4	HLA-DP	Not detected
OX747	OX747-157/158	TMBIM4	43.73	CD4	Not tested	0.325/0.45
	OX747-185	DOCK8	24.54	Likely CD4	HLA-DR	Not detected
	OX747-190/191	METTL22	22.2	Likely CD4	HLA-DR	Not detected
	OX628-305	DCAF13	23.64	Not known	Not known	Not detected
OX993	OX993-413/414	DENND6B	31.06	CD4	HLA-DR	0.4
OX628	OX628-80/81	LITAF	0.00122	CD4	HLA-DR	2.36/1.12
	OX628-564	PGLS	0.08	Not known	Not tested	Not detected
OX885	OX885-559/560	GSTZ1	30.57	CD4	Not tested	0.83
	OX628-308/309	WASHC4	94.67	CD4	Not tested	1.96/1.06

Table 5.2: Summary of antigens and their T cell responses. Summary of all antigens that were functionally validated by IFNγ ELISpot. For each row, the patient and corresponding peptide are shown. Where 2 peptides are indicated, they overlap the same variant amino acid and both elicited T cell responses. The allele frequency is from all samples in gnomAD [233]. The nature of responding T cells (CD4$^+$ or CD8$^+$) was assessed by flow cytometry. Where 'Likely CD4' is stated, responses were not detected by flow cytometry, but a CD4$^+$ response is suggested by blockade of responses using HLA class II antibodies. 'Not known' indicates that HLA blockade did not indicate the HLA restriction and no response was detectable by flow cytometry. Not all responses were assessed with HLA blockade.

Of the 10 patients who were screened, 9 had detectable responses to at least 1 peptide (Table 5.2). A mean of 2.4 functionally validated alloreactive peptides were detected per patient. Many result from very common SNPs, with 12 of 24 peptides present at allele frequency $> 25\%$. For many of those at high allele frequency, the donor is homozygous for the variant allele and the T cell response is directed against the reference allele, which is present in the patient but not the donor (Table B.2).

Interestingly, 86% of responses that were evaluable by flow cytometry were mediated by CD4$^+$ T cells. This result is in keeping with other studies that have shown the importance of HLA class II presentation in AML patients treated with allo-

SCT [42, 43]. The peptides eliciting $CD8^+$ responses were all from Patient OX289 and 2 of the responses were strong (peptides 9 and 284, Figure 5.21 (a) and (c)). This is in keeping with larger average clone sizes found in $CD8^+$ T cell repertoires compared with $CD4^+$ [281, 282]. A total of 10 peptide responses were found for Patient OX289, which is more than in any other patient studied. She experienced more GvHD than any of the other patients and therefore it is possible that some of these responses are directed against normal healthy tissues rather than AML.

The HLA allele restriction was tested for 16 of the 24 functionally validated peptides and a result was interpretable for 13 of these responses (Table 5.2). HLA class II responses were classifiable according to the particular molecule (e.g. HLA-DR and -DP), whereas HLA class I responses could not be sub-classified because reliable blocking antibodies are not available for HLA-A, -B and -C individually. *In silico* prediction correctly identified the restricting HLA allele in 11 of 13 cases (84.6%), including 9 high affinity and 2 low affinity binding peptides. In some cases, restriction could not be determined because there was no reduction in peptide response with any HLA-blocking antibody tested, which could be due to poor antibody binding to the particular HLA molecules found in the patient. In all cases, the HLA restriction results were consistent with the flow cytometric characterisation: $CD4^+$ T cell responses were HLA class II restricted and $CD8^+$ responses were class I restricted. HLA types of the patients studied are shown in Tables B.3 (a) and (b).

For 6 of the 24 peptides identified by IFNγ ELISpot, responses were not detected by flow cytometry, which is likely to be due to limited sensitivity of the assay. Another factor is that in some cases, different peptide cultures were used for ELISpot and flow cytometry assays. It is likely that sometimes the peptide-responsive clone does not adequately expand to enable detection, particularly where low frequency T cell clones are present in the sample.

If genes associated with alloreactive antigens share common features such as location in subcellular compartments or involvement in cellular processes, this could suggest common mechanisms of antigen presentation. This possibility was

investigated using pathway analysis with GSEA and Gene Ontology [221]. However, features linking the genes were not evident.

5.5 Predicting antigen binding to common HLA alleles

Experiments using HLA neutralising antibodies demonstrated the HLA restriction of responses. Precise identification of the restricting HLA allele was possible for the PADI4 antigen described in Chapter 4, due to homozygosity for the HLA-DP haplotype in the patient studied (Sections 4.3.3 and 4.4). However, in most cases HLA-blocking antibody experiments do not enable the identification of the precise restricting allele because patients have multiple alleles of the same HLA sub-type. For example, in each patient, there are up to 4 different HLA-DRB alleles: 2 DRB1 and up to 2 further alleles from DRB3/4/5.

Peptides can bind promiscuously to multiple HLA molecules, particularly for class II. The ability to bind to multiple HLA molecules would expand the potential clinical utility for targeting antigens, because a greater number of patients would have HLA that could present the peptide. In order to estimate binding to other HLA molecules, for each of the 24 peptides identified, HLA-binding predictions were performed for common alleles. Predictions were performed for 10 common HLA class I alleles, covering 91.7% of the world's population with at least one allele (Figure 5.24) [286, 287]. All 8-11 amino acid peptide windows across each variant amino acid were used in the predictions and if any peptide was predicted to bind a particular HLA molecule, the antigen was counted as a binder [160]. All 24 peptides were predicted to bind at least 3 of these HLA molecules. Peptides were predicted to bind a mean of 2.8 HLA molecules with high affinity (rank affinity $< 0.5\%$) and 2.8 with low affinity (rank affinity 0.5-2.0%).

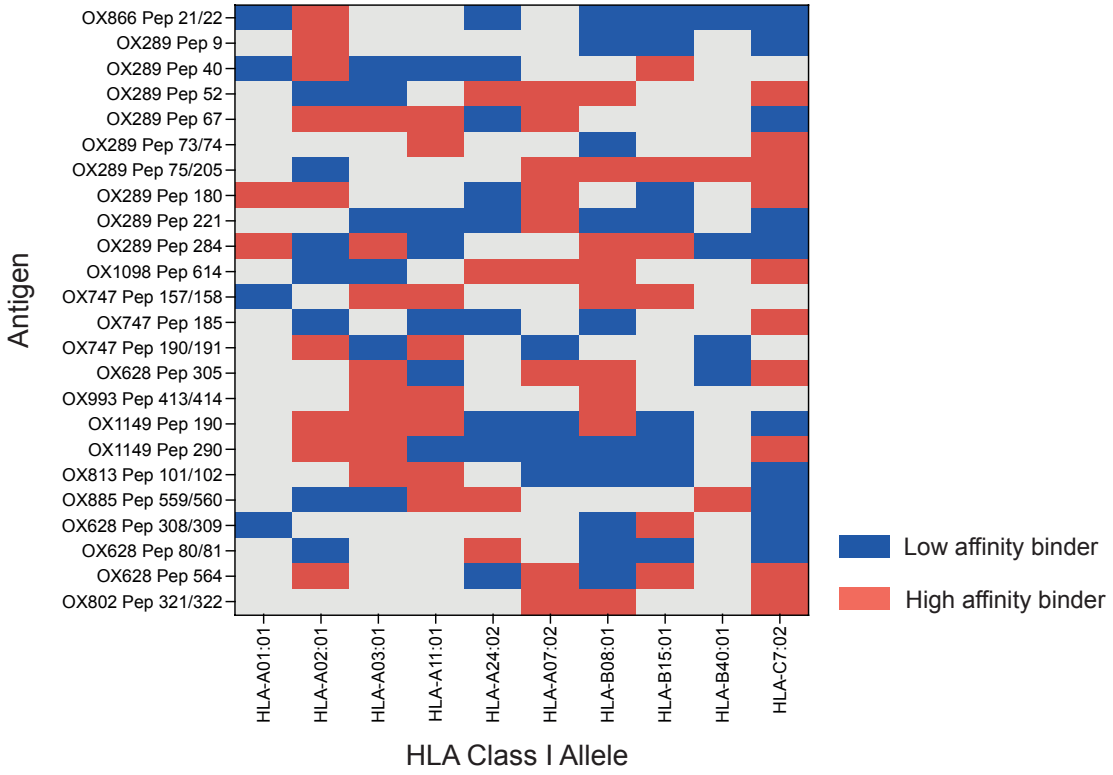


Figure 5.24: Predicted binding of GvL peptide antigens to common HLA class I molecules. For each of the 24 peptide antigens (rows), binding was predicted to 10 common HLA class I molecules (columns) [286, 287]. Predictions were performed using all 8-11 amino acid windows across the variant amino acid, using NetMHCpan 4.0 [160]. High affinity binding was defined as predicted rank affinity $< 0.5\%$ and low affinity binding was defined as rank affinity 0.5 - 2.0%.

Predictions were also performed for 27 common HLA class II alleles, with an estimated coverage of over 98% of the world's population with at least one allele (Figure 5.25) [288]. All 15-amino acid peptide windows across the variant amino acid were used in the predictions and if any peptide was predicted to bind a particular HLA molecule, the antigen was counted as a binder [161]. Compared with class I, there were fewer predicted binders per HLA molecule tested, which is likely to be at least in part because a smaller number of peptide windows were tested for class II. Whereas class I predictions involved up to 38 peptide windows per antigen (all 8-11 amino acid windows), class II predictions used up to 15 peptide windows per antigen.

Two peptide antigens (OX289 overlapping peptides 73/74 and OX628 peptide 305) were predicted to bind many of the HLA class II molecules: 22 and 20,

respectively. Two peptide antigens were not predicted to bind any of the common alleles: OX289 peptide 221, OX1149 peptide 190. Peptides were predicted to bind a mean of 1.6 HLA molecules with high affinity (rank affinity $< 0.5\%$) and 4.0 with low affinity (rank affinity 0.5-2.0%).

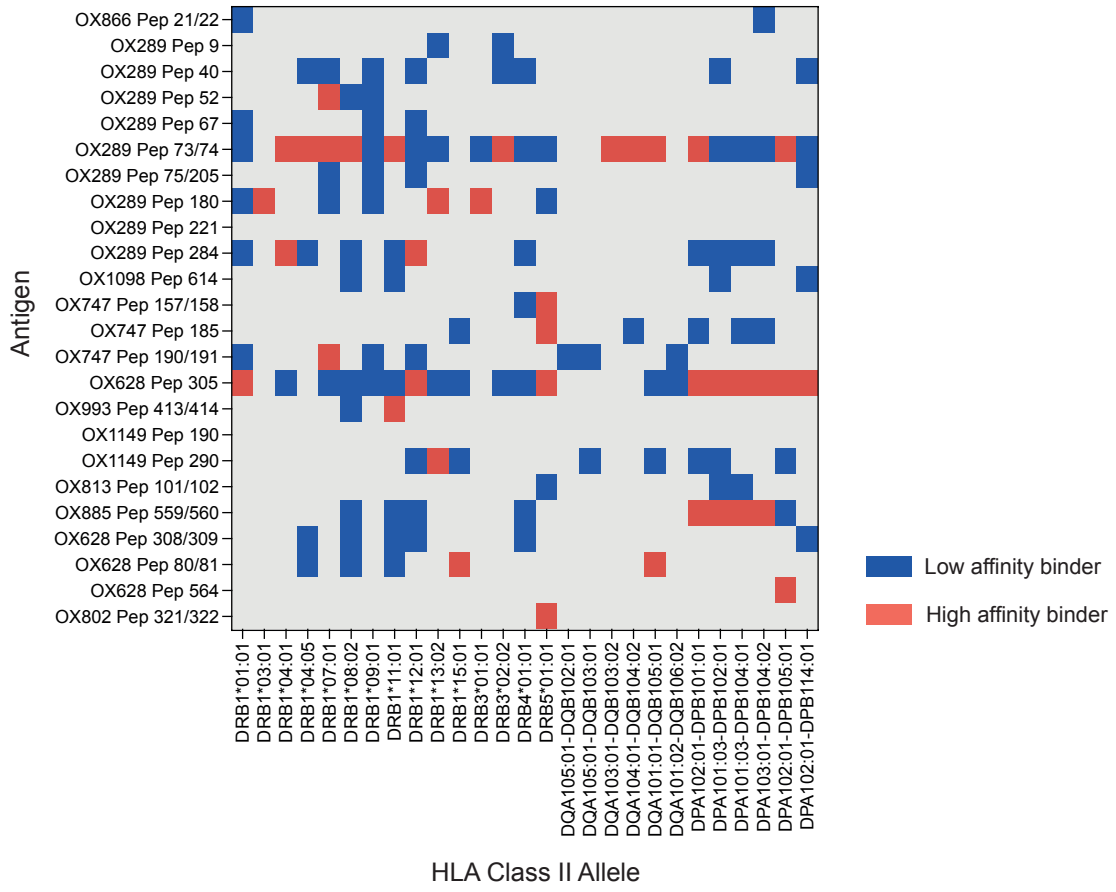


Figure 5.25: Predicted binding of GvL peptide antigens to common HLA class II molecules. For each of the 24 peptide antigens (rows), binding was predicted to 27 common HLA class II molecules (columns) [288]. Predictions were performed using all 15-amino acid windows across the variant amino acid, using NetMHCIIpan 4.0 [161]. High affinity binding was defined as predicted rank affinity $< 0.5\%$ and low affinity binding was defined as rank affinity 0.5 - 2.0%.

5.6 Expression of antigens in normal tissues

For use as a therapeutic target, antigens must be expressed by HLA on the surface of AML cells. If antigens are also expressed in normal healthy tissues, this could present a barrier to their clinical use due to the risk of toxicity. In

order to assess this risk, the expression of each of the 24 antigens was assessed across 37 healthy tissues (Figure 5.26).

The consideration of GvHD differs according to whether targeted immunotherapy is to be delivered following allo-SCT, when haematopoietic and lymphoid cells are derived from the transplant donor. In this setting, if the patient has the SNV that produces the antigen, but the donor lacks this SNV, then the donor-derived haematopoietic and lymphoid tissues would not be targeted, even if the gene is expressed in these tissues. Therefore, haematopoietic and lymphoid tissues (bone marrow, spleen, lymph node and tonsil) were considered separately. Testis was also considered separately because the testes are in an immune privileged site and therefore targeting cancer-testis antigens does not generally cause toxicity to the testes. [289].

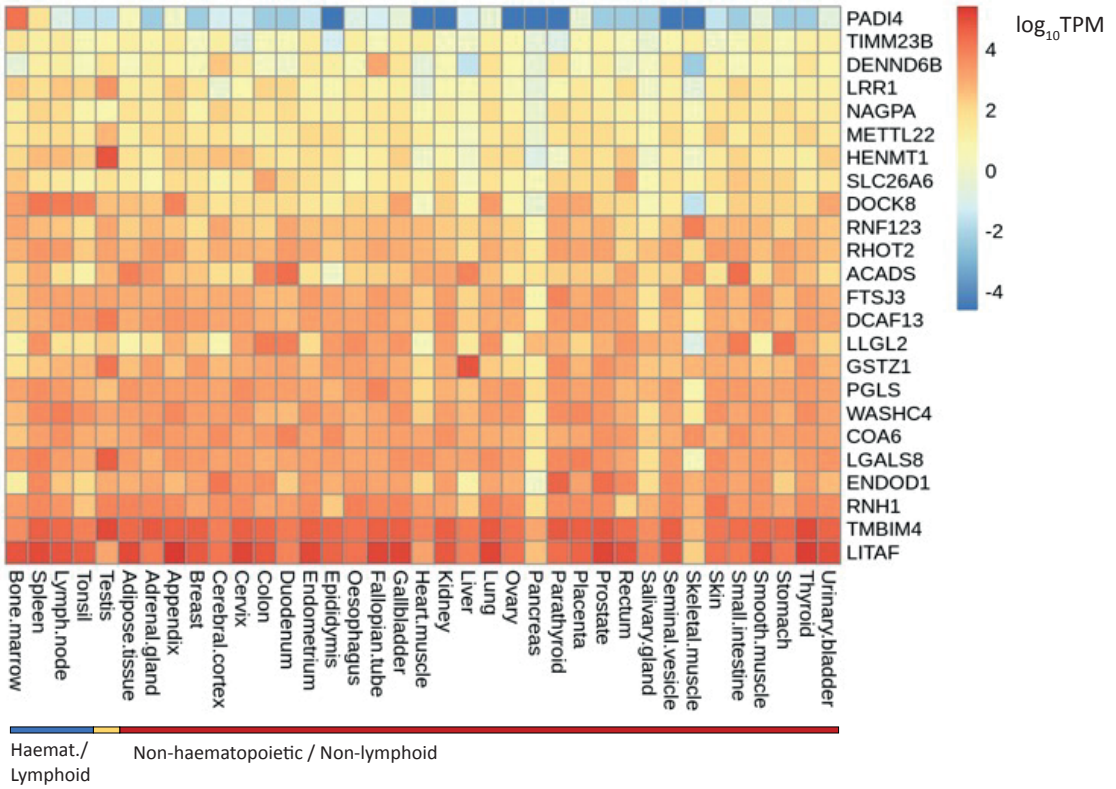


Figure 5.26: Expression of GvL antigens in normal tissues. Heatmap depicting expression of the genes from which the antigens are derived (rows) in 37 normal healthy tissues (columns), using RNA-Seq data from The Human Protein Atlas [290]. At the bottom of the heatmap, tissue groupings are indicated: haematopoietic / lymphoid tissues (blue bar), testis (yellow bar) and other tissues (red bar).
$\log_{10}$TPM: $\log_{10}$ Transcripts per million

PADI4 has selective haematopoietic / lymphoid tissue expression in the bone marrow and spleen, whereas all other genes have expression in other tissues. Although GvHD can affect almost any tissue, the gastrointestinal tract, liver and skin are commonly affected. Therefore, expression of the antigens was assessed specifically in these tissues. In addition to *PADI4*, some other genes were also expressed at only a low level in tissues that are commonly affected by GvHD, including *DENND6B*, *TIMM23B*, *METTL22*, *NAGPA* and *HENMT1* (Figure 5.27).

Despite a gene being expressed in a certain tissue, targeting the corresponding antigen may not lead to GvHD, especially if the antigen is presented by HLA class II, because expression is largely restricted to antigen-presenting cells (dendritic cells, B cells and macrophages). However, expression may be increased in normal tissues

in the context of inflammation through the action of IFNγ on the MHC class II transactivator (*CIITA*), which regulates HLA class II expression [291–293].

Upregulation of HLA class II expression has also been reported in multiple cancers, including bladder, breast, melanoma, ovarian, glioma, lung, colorectal, prostate and Hodgkin's lymphoma [294, 295] As the GvL antigens identified are germline-encoded and not AML-specific, this raises the exciting possibility of targeting these antigens in the treatment of other tumours.

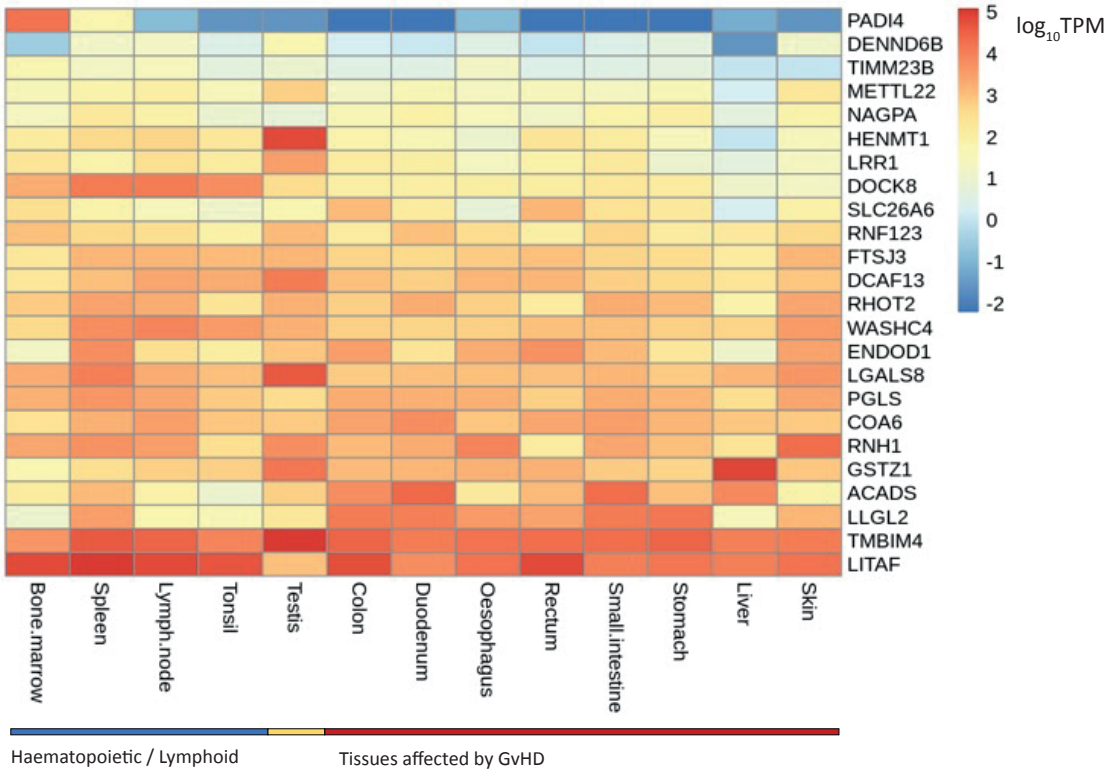


Figure 5.27: Expression of GvL antigens in tissues affected by acute GvHD. Heatmap depicting expression of the genes from which the antigens are derived (rows) in 37 normal healthy tissues that are commonly affected by GvHD (columns), using RNA-Seq data from The Human Protein Atlas [290]. At the bottom of the heatmap, tissue groupings are indicated: haematopoietic / lymphoid tissues (blue bar), testis (yellow bar) and other tissues (red bar).
$\log_{10}$TPM: $\log_{10}$ Transcripts per million, GvHD: Graft-versus-host Disease

5.7 Identifying antigen-specific TCR sequences

Using the same method as for the PADI4 antigen (Section 4.6.1), T cells responding to the additional antigens were isolated using IFNγ capture and analysed

by scRNA-Seq. In each case, post-allo-SCT blood samples from the patient with the associated antigen were used, which enabled the detection of responses from donor-derived T cells. Of the 17 antigens with detectable IFNγ responses by flow cytometry (Table 5.2), 16 were detectable using the IFNγ capture assay and therefore the T cells could be FACS-sorted and analysed by scRNA-Seq. For each antigen, only $CD4^+$ or $CD8^+$ cells were analysed, depending which were peptide-responsive.

For all 16 antigens, the TCR repertoires from T cells activated by peptide ($IFN\gamma^+$) contained at least one clone that was larger than the largest of the control repertoire ($IFN\gamma^-$), suggesting that successful enrichment of antigen-specific T cells had been achieved. The TCR repertoires from all $CD4^+$ T cell responses from Patients OX1149 and OX289 are summarised in Figures 5.29 and 5.28. In some cases, a single expanded clone dominates the repertoire, such as OX289 peptide 40 (largest clone 85.2% of the repertoire), whereas in other cases multiple expanded clones were detected, such as OX1149 peptide 190. From these analyses, the nucleotide and amino acid sequences of putative antigen-specific TCRα and β pairs were identified.

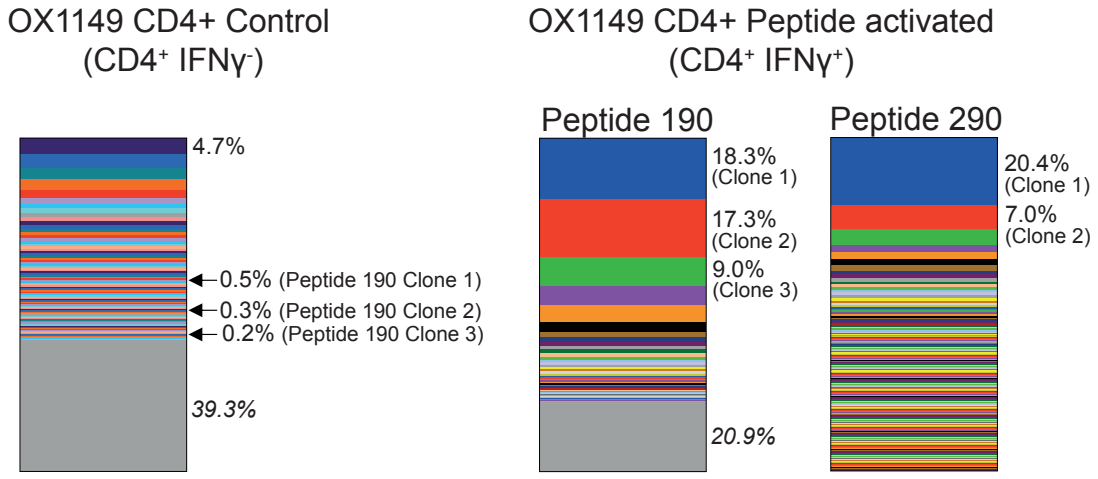


Figure 5.28: OX1149 $CD4^+$ TCR repertoires following peptide stimulation. Each large bar represents all T cell clones identified from the control ($CD4^+IFN\gamma^-$, left bar) and activated ($CD4^+IFN\gamma^+$, right bars) TCR repertoires. The grey portions of each bar represent the sum of all clones that are individually $< 0.2\%$ of the total repertoire. The control T cells were FACS-sorted from the peptide 190 culture. In the peptide 190 and 290 activated repertoires, the labelled clones are those that are larger than the largest clone in the control repertoire. Their size is indicated as % of the total repertoire. In the control repertoire, peptide 190 clones 1, 2 and 3 were present, but at much lower frequency than in the activated repertoire.

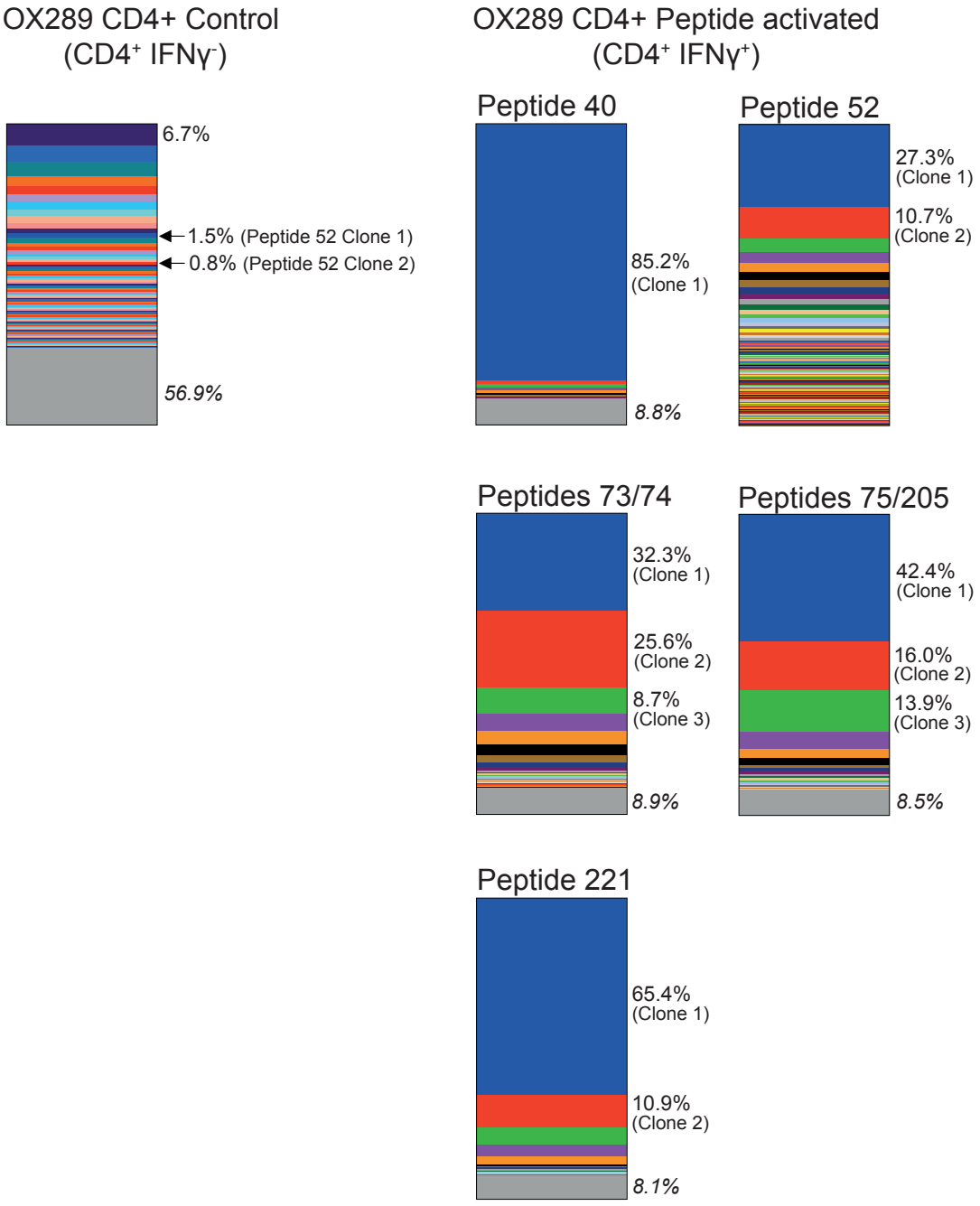


Figure 5.29: OX289 CD4$^+$ TCR repertoires following peptide stimulation. Each large bar represents all T cell clones identified from the control (CD4$^+$IFNγ^-, left bar) and activated (CD4$^+$IFNγ^+, right bars) TCR repertoires. The grey portions of each bar represent the sum of all clones that are individually $< 0.2\%$ of the total repertoire. The control T cells were FACS-sorted from the peptide 52 culture. In the activated repertoires for peptides 40, 52, 73/74, 75/205 and 221, the labelled clones are those that are larger than the largest clone in the control repertoire. Their size is indicated as % of the total repertoire. In the control repertoire, peptide 52 clones 1 and 2 were present, but at much lower frequency than in the activated repertoire.

TCR repertoires from the remaining T cell responses are being analysed. The putative antigen-specific TCR sequences from expanded clones could be inserted into T cells from healthy donors and used to therapeutically target their corresponding antigen on AML and possibly other cancer cells (See Chapter 6).

Single cell gene expression data is also currently being used to assess the phenotype of post-allo-SCT T cells responding to each peptide. This analysis will determine whether alloreactive T cells exhibit a consistent phenotype when responding to different peptides or whether distinct phenotypes are observed.

5.8 Functional screening of peptides identified by mass spectrometry

5.8.1 *Ex vivo* testing

In addition to screening large libraries of *in silico* predicted HLA-binding peptides, patient-donor mismatched peptides identified by LC-MS/MS (Section 3.5.5) were also functionally tested by IFNγ ELISpot. Given that these peptides were directly eluted from immunoprecipitated HLA molecules on AML cells, we hypothesised that there may be a greater rate of successful functional validation. Post-transplant blood samples were available from 8 patients, which enabled testing of 25 of the 31 patient-donor mismatched peptides identified by LC-MS/MS.

Analysing samples taken directly from patients without culture provides a better estimation of the number of antigen-specific T cells circulating in the patient. Therefore, patient-donor mismatched peptides were initially tested by *ex vivo* IFNγ ELISpot. Across all 8 patient samples tested, there were no detectable responses (Figure 5.30).

5.8.2 Testing following *in vitro* culture

In vitro culture with peptide that is recognised by a TCR can induce T cell proliferation, increase the number of peptide-responsive T cells and thereby increase the sensitivity of the assay. After culturing post-transplant PBMCs with peptide and then performing IFNγ ELISpot, 1 of the 25 peptides (OX866 peptide MS6) elicited responses across replicates (Figure 5.31). However, this was not reproducible

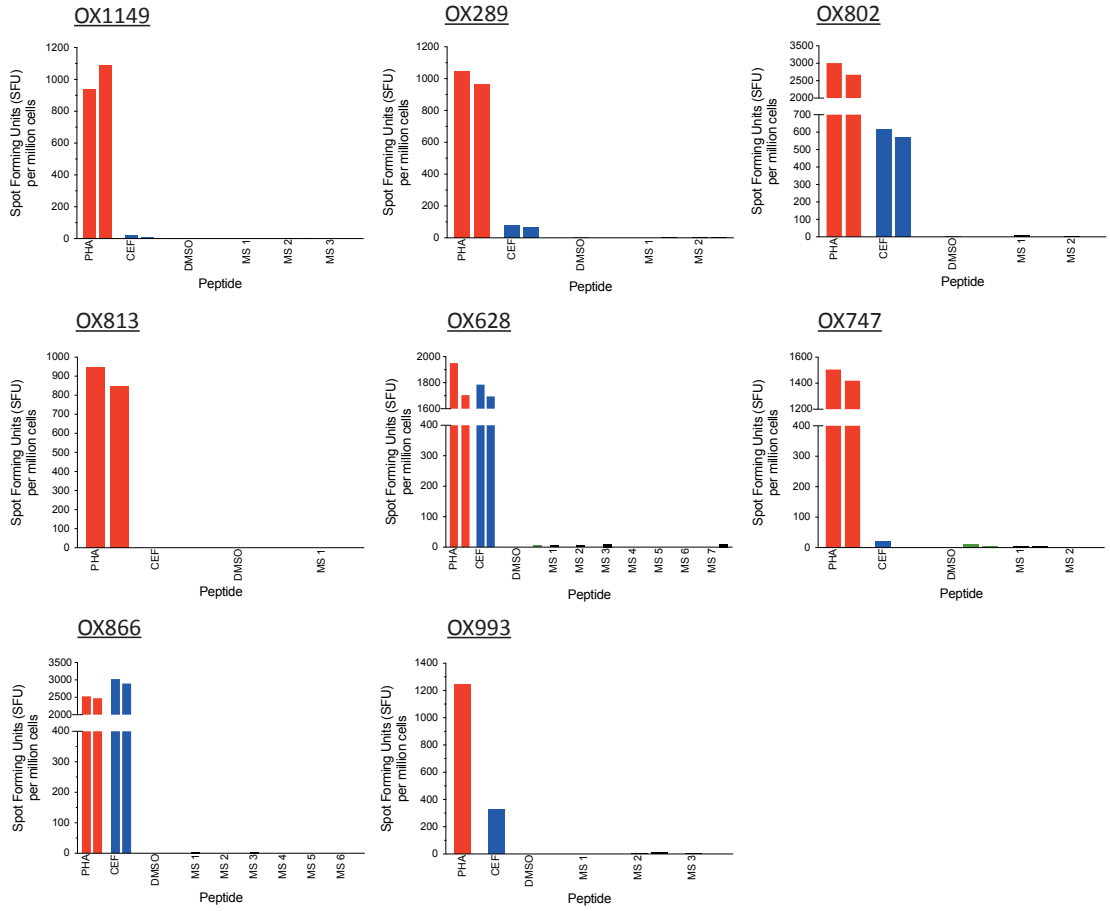


Figure 5.30: *Ex vivo* IFNγ ELISpot testing of patient-donor mismatched peptides identified by LC-MS/MS. Each plot shows data from a single patient. Peptides are labelled MS 1 - 7 and are different for each patient. Peptides were each tested with 2 replicates (black bars). Controls - PHA (positive control, red bars), CEF peptides (positive control, blue bars) and DMSO (negative control, green bars).

on repeat testing. OX866 peptide MS6 was predicted *in silico* to bind one of the patient's HLA-DR alleles (HLA-DRB3*01:01) and therefore had previously been screened in OX866 Pool 8 without eliciting responses (Figure 4.5 (a)).

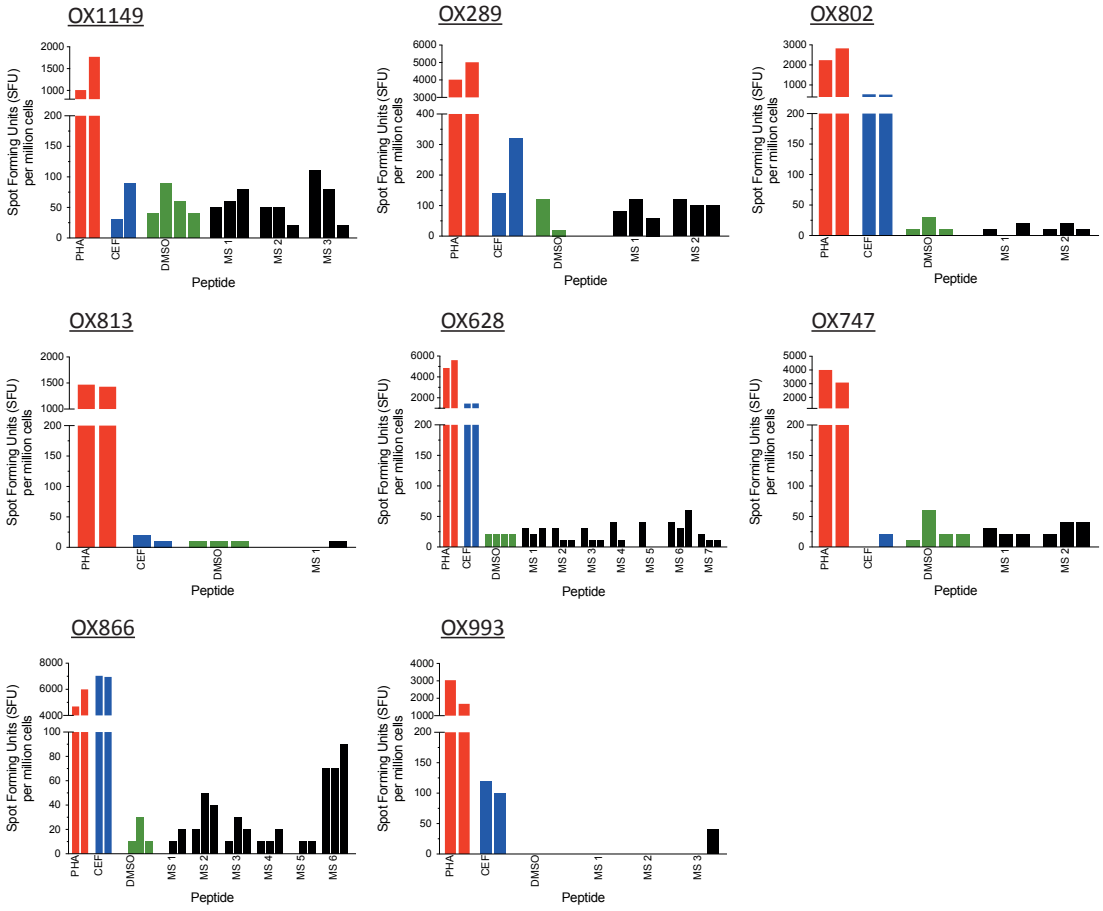


Figure 5.31: IFNγ ELISpot testing of peptides identified by mass spectrometry, following *in vitro* culture. Post-transplant PBMCs were cultured with peptide, IL-2 and IL-7 for 12 days, then washed and rested for 24 hours. IFNγ ELISpot testing of the peptide(s) was then performed with 3 replicates (black bars). Each plot shows data from a single patient. Peptides are labelled MS 1 - 7 and are different for each patient. Controls - PHA (positive control, red bars), CEF peptides (positive control, blue bars) and DMSO (negative control, green bars).

5.9 Conclusions: Identifying and characterising alloreactive antigens

A total of 5340 peptides that were predicted to bind patient-specific HLA molecules from 10 AML patients were screened using cultured IFNγ ELISpot assays. Donor-derived T cell responses from post-allo-SCT blood samples were detected for 24 peptide antigens. 86% of responses were mediated by $CD4^+$ T cells, which is consistent with previous studies that have demonstrated the importance of

antigens presented by HLA class II in AML patients following allo-SCT [42, 43]. Where evaluable, the HLA restriction always corresponded to the T cell responses: HLA-DP and DR-restricted antigens activated $CD4^+$ T cells, whereas HLA class I-restricted antigens activated $CD8^+$ T cells.

For use as therapeutic targets, antigens must be expressed by the cancer cells. However, expression in normal tissues could present a barrier to clinical use due to the risk of toxicity. *PADI4* exhibits the most haematopoietic tissue-restricted expression, but several other genes that result in antigens are expressed at only a low level in tissues that are commonly affected by GvHD (*DENND6B*, *TIMM23B*, *METTL22*, *NAGPA* and *HENMT1*). Due to the restricted tissue distribution of HLA class II expression, most of the identified antigens may not be presented to T cells despite gene expression. Therefore, these antigens may still be targetable without significant toxicity.

Applying the workflow to patients with clinical evidence of GvL in the absence of GvHD could enrich for the discovery of antigens with greater therapeutic potential. Patients who have had AML remission following DLI would be ideal candidates to study.

Chapter 6

Targeting the PADI4 antigen using adoptive T cell therapy

Contents

6.1 Targeting the PADI4 antigen with T cell therapy . . . 195
6.2 TCR transfer using viral vectors 196
6.2.1 Retroviral TCR transfer 196
6.2.2 Lentiviral TCR transfer 198
6.3 Identifying target cells for functional assays 204
6.4 T cell activation by AML 210
6.5 *In vitro* killing of AML by targeting the PADI4 antigen 213
6.6 Assessing the mechanism of PADI4 antigen presentation 215
6.7 *In vivo* killing of AML by targeting the PADI4 antigen 219
6.8 Conclusions: Targeting the PADI4 antigen using adoptive T cell therapy 221

6.1 Targeting the PADI4 antigen with T cell therapy

Adoptive T cell therapy with T cells expressing transgenic TCRs has shown safety and efficacy in the treatment of a range of cancers, including AML, myeloma, melanoma and multiple other solid tumours [178, 296, 297]. In most cases, this involves genetically redirecting the specificity of autologous T cells to target a neoantigen or tumour-associated antigen. However, GvL antigens identified in Chapters 4 and 5 could also be targeted using adoptive T cell therapy and given that the antigens are germline-derived, they may be targetable in multiple cancers that express the appropriate gene.

To be recognised by the TCR, the patient must have the correct DNA variant and restricting HLA allele, which can substantially reduce the proportion of patients who can be treated, depending on the SNP and HLA allele frequencies in the population. Patients with the combination of DNA variant and the restricting HLA

allele could be treated with adoptive T cell therapy, using T cells that express an antigen-specific TCR. For the PADI4 antigen, the SNP allele frequency is 4.27% and the restricting HLA-DP allele has a frequency of 41.7%. By applying the Hardy-Weinberg principle, 5.5% of patients are estimated to have the combination and therefore be treatable (Section 4.4).

As a proof of concept, this antigen was chosen to develop a targeted T cell-based immunotherapy. The patient with this antigen had clinical evidence of a GvL response without GvHD (Section 4.1.2), suggesting that the antigen may serve as an effective target. Unlike many of the other GvL antigens identified, *PADI4* is principally expressed in the haematopoietic system and therefore toxicity to healthy tissues is predicted to reduced (Section 4.5.2).

The most efficient method to introduce TCRs into primary human T cells is using viral vectors and this method has been used in the majority of clinical trials [298]. In this study, 2 types of vector were assessed for TCR transfer: (1) a gamma-retrovirus, and (2) a lentivirus.

6.2 TCR transfer using viral vectors

6.2.1 Retroviral TCR transfer

Using a post-transplant blood sample from Patient OX866, T cell clones were identified as being activated in response to the PADI4 antigen by IFNγ capture and single cell RNA-Seq. Of these clones, 2 were markedly expanded in the activated repertoire and are putatively antigen-specific (Section 4.7). The nucleotide sequences for the TCRα and β variable regions of the 5 largest clones were inserted into a retroviral vector (Figure 6.1). These comprised the V, D and J gene segments identified from single cell RNA sequencing (Figure B.1). Clone 2 expressed 2 different TCRα chains and 1 TCRβ chain and therefore 2 possible pairings of TCR chains are possible (Section 4.7, Figure 4.27 (b)). Both combinations were cloned for testing of PADI4 antigen recognition.

The synthetic TCRs in the retroviral vector incorporate murine constant regions, which prevent mis-pairing with endogenously expressed TCR chains in the

transduced cells, which could lead to multiple expressed TCR dimers in each cell. Using murine constant regions also enables expression of the inserted TCR to be assessed using an anti-murine TCR antibody by flow cytometry, which distinguishes it from endogenous human TCR molecules. The variable and constant regions were human codon optimised, which preserves the amino acid sequence but uses nucleotide codons that are optimised for human expression. This may improve translation elongation rate and accuracy of translation [299]. Murine constant regions were equipped with an additional disulfide bond to enhance TCRα/β chain pairing [300]. A truncated murine CD19 molecule was also included as a selection marker, separated from TCRs by a T2A self-cleavage site.

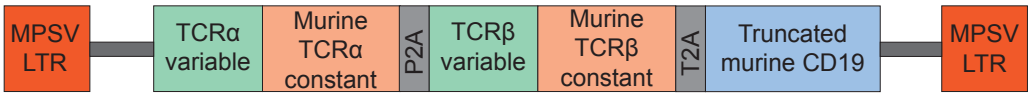


Figure 6.1: pMP71-TCR retroviral vector. Vector map showing the transgene region that integrates into the host cell genome. This region is flanked by LTRs derived from MPSV, which drives transcription [301]. Self-cleavage sites (P2A and T2A) separate the TCRα chain, TCRβ chain and truncated murine CD19.
MPSV: Myeloproliferative sarcoma virus, LTR: Long terminal repeat, TCR: T cell receptor

Retrovirus was produced and used to transduce T cells from a healthy human donor. The efficiency of TCR transgene expression was assessed by flow cytometry after 72 hours. Due to variations in cell surface expression levels of human TCRs, murine CD19 expression was used to assess efficiency of transduction, as this may differ from efficiency of TCR expression. Low expression of TCRs may impair the function of engineered T cells [300]. Murine CD19 and TCRβ expression were similar, indicating that TCR expression was detected in most or all transduced cells, with 24.9% cells expressing the inserted TCR (Figure 6.2).

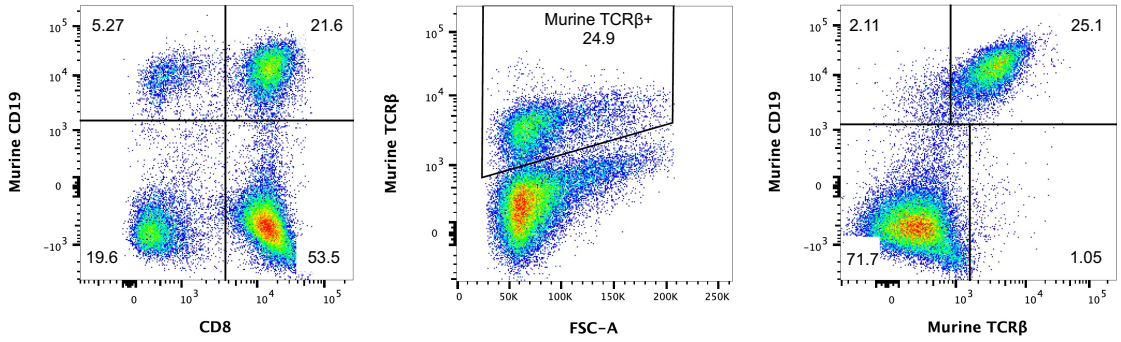


Figure 6.2: Retroviral transduction of T cells. Flow cytometry plots showing assessment of transduction efficiency (murine CD19) and expression of the inserted TCR (murine TCRβ) in human T cells (Live CD3$^+$). Murine CD19 expression was found in 26.87% (left panel) and murine TCRβ in 24.9% (centre panel). Most cells that expressed murine CD19 also expressed murine TCRβ (right panel). Unless otherwise stated, the markers indicated on axes are human.

6.2.2 Lentiviral TCR transfer

Due to the modest and variable transduction efficiencies observed in experiments using retrovirus, TCR transfer using lentivirus was attempted. Both retrovirus and lentivirus have RNA genomes that are incorporated into host cell DNA by reverse transcription and result in stable transgene expression. However, compared with retrovirus, lentivirus has several advantages:

1. Lentivirus can infect non-dividing cells, which broadens the cell types that can be transduced and may lead to higher transduction efficiency.

2. Typically lentiviral titres and transduction efficiencies are higher than with retrovirus.

3. Lentiviruses may be safer for use in humans. The use of retrovirus in early gene therapy trials was associated with a high risk of T cell acute lymphoblastic leukaemia (T-ALL) as a result of virus insertion site mutagenesis [302, 303]. In contrast, there are no reported cases of leukaemia associated with newer generation lentiviral vectors. Additional safety features include replication incompetence, so virus can not be reproduced in the target cell, and a deletion in the 3' long terminal repeat (LTR) rendering the virus self-inactivating after integration.

PADI4 TCRα and β variable sequences were sub-cloned into the pHRsin-TCR backbone from the retroviral vector (Figure 6.3). This vector lacks the murine

CD19 tag, reducing its size, which may result in higher lentiviral titres.



Figure 6.3: pHRsin-TCR lentiviral vector. Vector map showing the transgene region that integrates into the host cell genome, which is flanked by LTRs. SFFV drives transcription of the TCRβ and TCRα chains as a single transcript, separated by a P2A self-cleavage site.
LTR: Long terminal repeat, SFFV: Spleen focus-forming virus, TCR: T cell receptor

Highly efficient expression of the inserted TCRs was achieved, with up to 78.4% in CD4$^+$ and 67.8% in CD8$^+$ T cells (Figure 6.4). Efficiency was consistently higher in CD4$^+$ cells than CD8$^+$ cells across transductions. There was some difference in TCR expression between the different clones, with a consistent pattern across transduction experiments. Negative control Clone 1G4, which is a HLA-A2-restricted TCR that recognises an NY-ESO-1 epitope, always had the lowest expression (Figure 6.5). Transduction 3 was associated with lower TCR expression across all clones and the effect was most marked for CD8$^+$ cells.

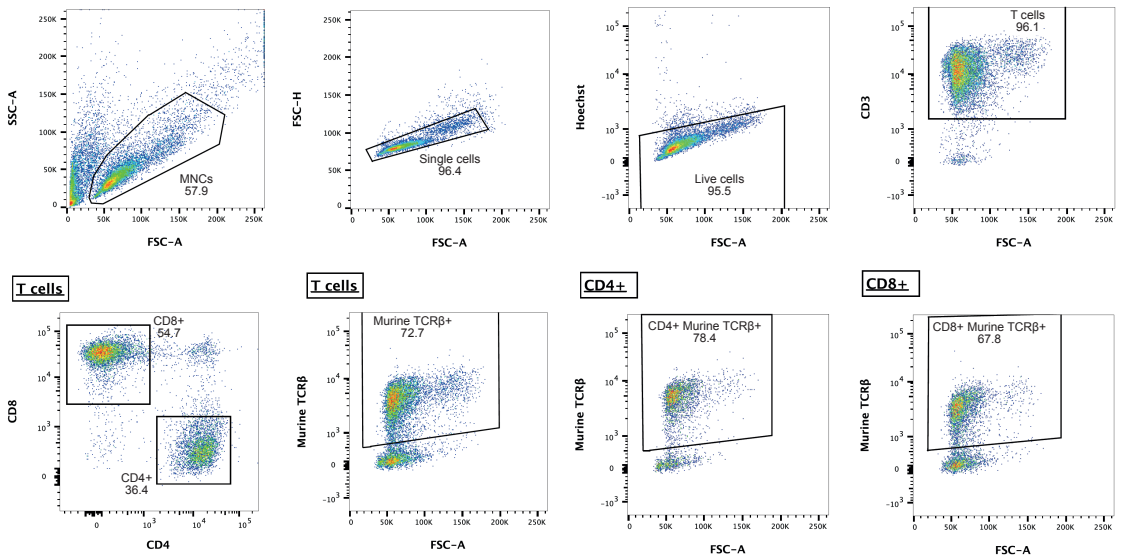


Figure 6.4: Lentiviral transduction of T cells. Representative flow cytometry plots showing expression of the inserted TCR (murine TCRβ) in human T cells. Overall, 72.7% T cells express the transgene, with 78.4% and 67.8% of CD4$^+$ and CD8$^+$ cells, respectively. Unless otherwise stated, the markers indicated on axes are human.

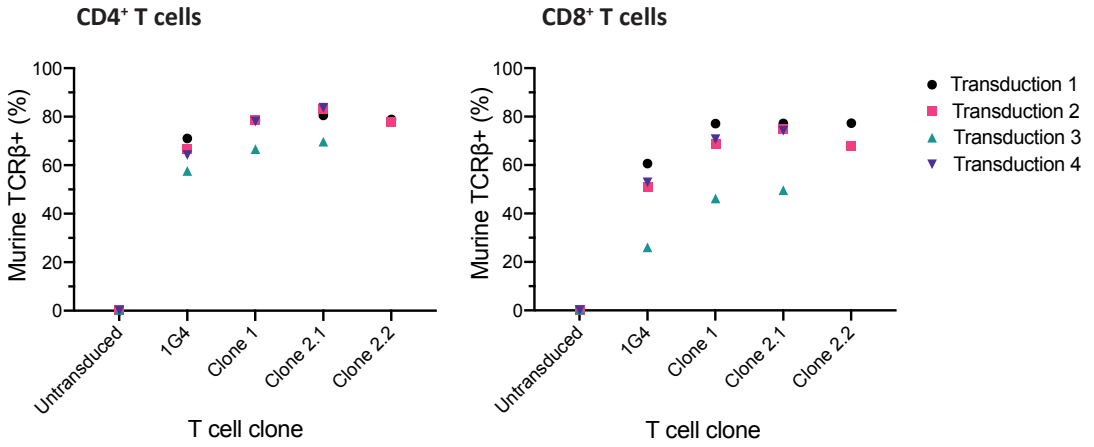
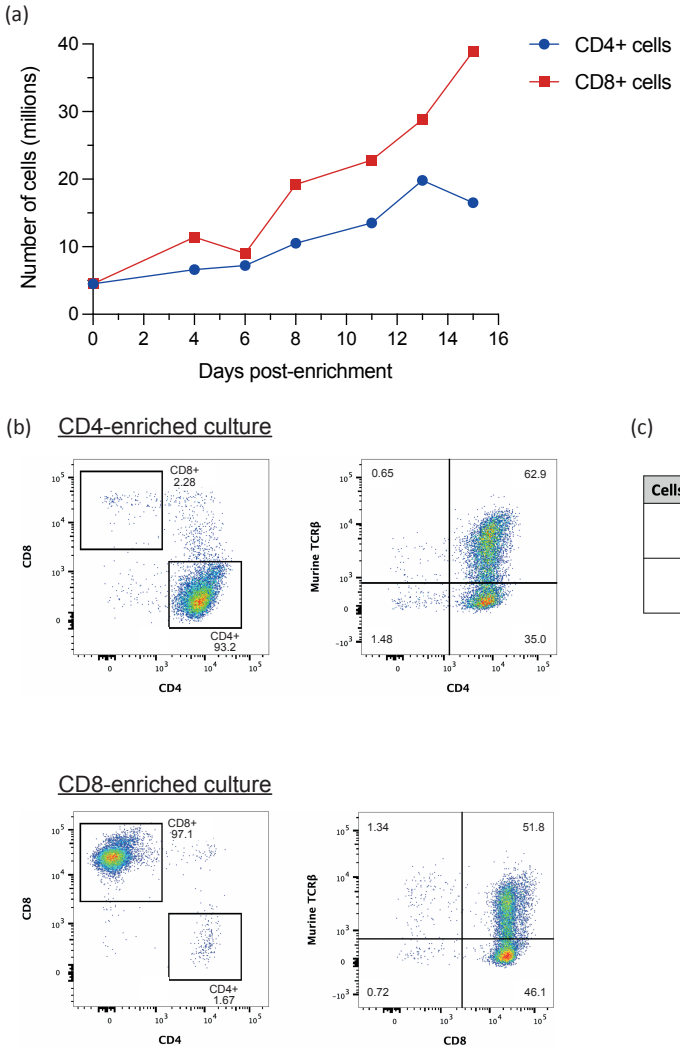


Figure 6.5: Serial transduction efficiencies. Surface expression of the inserted TCR was assessed by flow cytometry using anti-murine TCRβ. Untransduced, 1G4 control and PADI4 Clones 1, 2.1 and 2.2 are shown. Clone 2.2 was not included for transductions 3 and 4 due to the lack of response to PADI4 antigen.

At 72 hours post-transduction, $CD4^+$ and $CD8^+$ cells were separated using magnetic beads (Miltenyi) and expanded in culture for up to 2.5 weeks. They were cultured separately to enable the ratio of $CD4^+$ and $CD8^+$ cells in downstream assays to be controlled. CD8 cells expanded more than CD4 cells and in a representative experiment, CD4 expanded 3.5-fold and CD8 expanded 8.3 fold during the 15 days after the cells were separated (Figure 6.6 (a)).

Following culture, the proportion of cells expressing CD4 or CD8 was maintained, but there was a reduction in murine TCRβ expression (Figure 6.6 (b) and (c)).



(c)

		Day post-enrichment	
Cells in culture	**FACS marker**	**0**	**15**
CD4	CD4 (%)	99.4	97.9
	Murine TCRβ (%)	74.6	62.9
CD8	CD8 (%)	94.2	97.9
	Murine TCRβ (%)	57.5	51.8

Figure 6.6: CD4$^+$ and CD8$^+$ enriched T cell cultures. Representative data from 1G4 T cells cultured following transduction and CD4 / CD8 cell enrichment. **(a)** Number of CD4$^+$ and CD8$^+$ cells in culture over time. Day 0 is taken from when CD4$^+$ and CD8$^+$ cells were magnetically separated at 3 days post-transduction. **(b)** Flow cytometry plots showing expression of CD4, CD8 and murine TCRβ at the end of the culture period. Numbers on the plots are % of of live CD3$^+$ T cells. Unless otherwise stated, the markers indicated on axes are human. **(c)** Table showing % expression of the indicated surface marker at days 0 and 15.

Prior to use in functional assays, recognition of PADI4 antigen by the TCRs was first assessed by IFNγ ELISpot after combining $CD4^+$ and $CD8^+$ cells for each clone in a 1:1 ratio. As a source of T cells for transduction, a healthy donor was selected who was heterozygous for the restricting HLA allele (HLA-DPA1*01:03-DPB1*04:01). The PADI4 peptide, but not WT PADI4 peptide, elicited strong responses from PADI4 Clone 2.1 (Figure 6.7). The WT PADI4 peptide differs from PADI4 peptide only by a single amino acid substitution that results from the DNA variant. Of the two combinations of TCRα/β pairings in Clone 2, Clone 2.1 was expected to be correct given that the 4th largest clone in the activated repertoire from single cell RNA-Seq (4.0% of total repertoire) expressed only this TCRα chain, together with the TCRβ chain (Section 4.7). It is unclear why responses were not observed with Clone 1. Although high surface expression levels of this TCR were previously demonstrated, it is possible that improper folding of the TCR could lead to failure of antigen recognition.

It is likely that T cells in the culture present the peptide to other T cells expressing the inserted TCR, leading to activation and IFNγ release. Although resting T cells do not express HLA class II, expression is upregulated following activation and HLA-DP expression was detectable in 49.8% of the T cells in culture, which likely results from the presence of IL-2 and IL-15 (Figure 6.8).

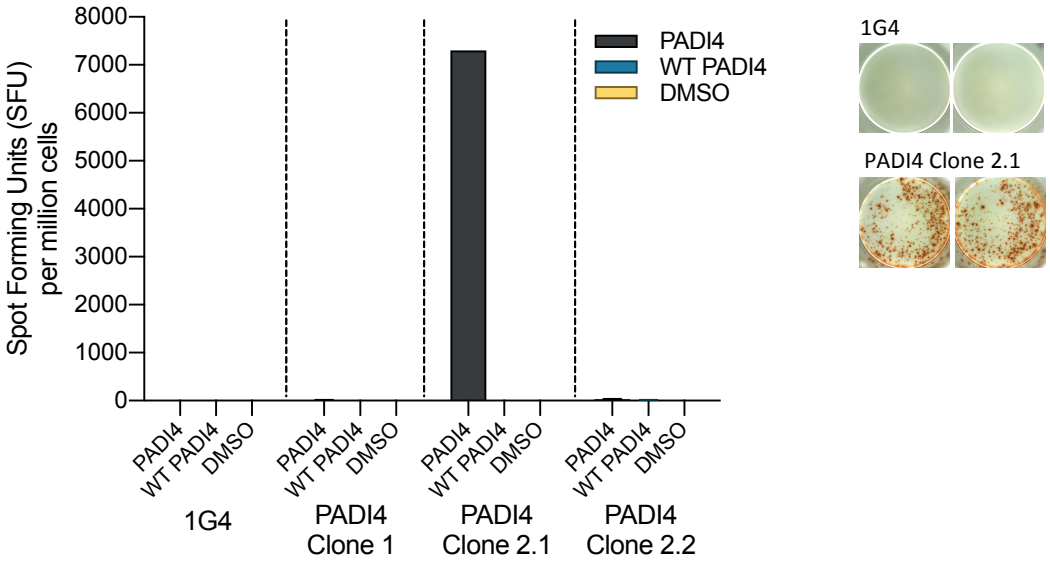


Figure 6.7: Clone 2.1 exhibits PADI4 antigen-specific responses. IFNγ ELISpot assessment of responses to PADI4 peptide by transduced T cell clones. Each clone was treated with either PADI4 peptide (10 μg/ml), WT PADI4 peptide (10 μg/ml) or DMSO for 18 hours, before ELISpot plates were developed. Values are mean +/- SEM of 2 replicates. Error bars are not visible due to similar results across replicates for PADI4 Clone 2.1 and very low responses for the other conditions. On the right of the panel, ELISpot wells from 1G4 control and PADI4 Clone 2.1 wells are shown, each treated with PADI4 peptide.
DMSO: Dimethyl sulfoxide, SEM: Standard error of the mean

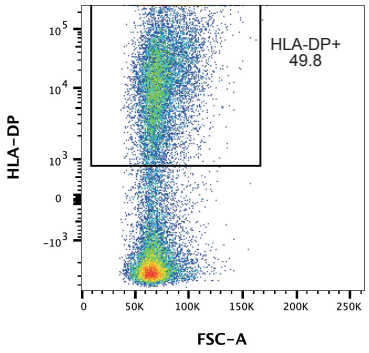


Figure 6.8: HLA-DP is expressed on cultured T cells. Flow cytometry plot showing HLA-DP surface expression on T cells at the end of culture. Gating was first performed for live CD3$^+$ cells and shows that 49.8% express HLA-DP.

6.3 Identifying target cells for functional assays

For a cell to be recognised by the PADI4 antigen-specific T cells, it must have the following characteristics:

1. *PADI4* SNP that result in the peptide antigen.
2. Appropriate HLA-DP type (HLA-DPA1*01:03-DPB1*04:01).
3. Expression of *PADI4*.
4. Cell surface expression of HLA-DP.

AML cell lines and primary human AML samples with available biobanked cells were tested for these characteristics, to identify suitable samples for assays assessing AML recognition and killing by PADI4 T cell clones.

AML cell lines and primary human samples were genotyped to look for the *PADI4* SNP (dbSNP ID = rs1748020) using either Sanger sequencing or a TaqMan SNP Genotyping Assay (Applied Biosystems, Thermo Fisher) (Figure 6.9). Of the AML cell lines tested (THP-1, MOLM-13, OCI-AML3, ME-1, K562, HL-60, Kasumi, HEL, CMK, GDM-1, KG-1, ML-2, MONOMAC-1, MV4;11, NOMO-1, P31/FUJ, SHI-1, U937), the SNP was detected in 0 of 18. The SNP was detected in 41 of 499 (8.2%) primary AML samples tested. This is similar to the expected proportion of the population with the SNP (8.4%), based on the gnomAD database [233] (Section 4.4).

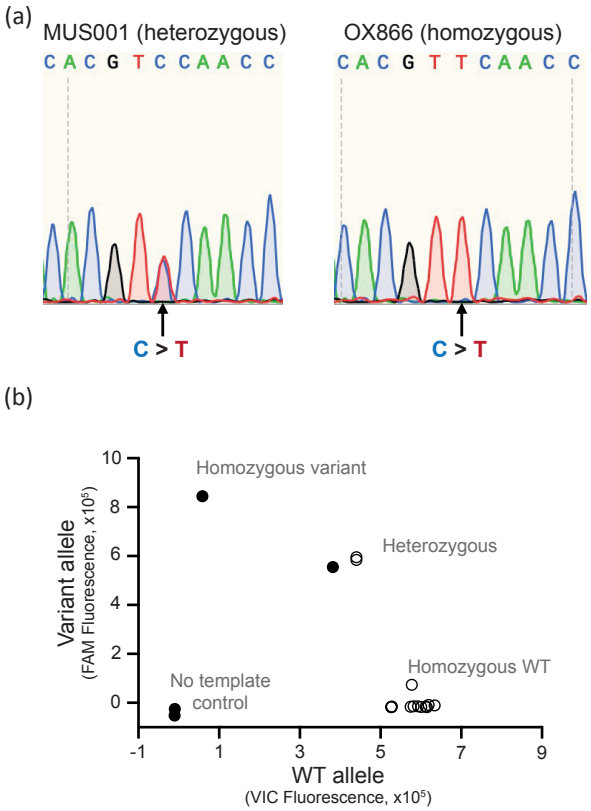


Figure 6.9: PADI4 genotyping. (a) Representative Sanger sequencing traces showing samples that are heterozygous (MUS001) and homozygous (OX866) for the *PADI4* SNP (C > T transition) at position chr1:17342114 (GRCh38). **(b)** Genotyping by RT qPCR using a TaqMan SNP Genotyping Assay, which incorporates probes complementary to the WT and variant alleles, each associated with a different fluorescent dye (VIC or FAM). The ratio of fluorescence intensity for each dye indicates the genotype.
RT qPCR: Real-time quantitative polymerase chain reaction, WT: Wild type

In order for the PADI4 peptide antigen to be presented by HLA, the PADI4 protein must be expressed. Protein expression by human AML samples was assessed using Western blot (Figure 6.10). Samples were tested that had the *PADI4* SNP and HLA-DPB1*04:01 allele. Samples were also tested that lacked these characteristics, to identify suitable negative control samples for downstream assays, which would enable us to demonstrate that recognition and killing of target cells is dependent on the combination of both the *PADI4* SNP and HLA-DPB1*04:01 allele. In the Western blots shown, expression was detected for samples OX1433, GCH001, NOT114, OX903 and positive control sample OX1093. There was also likely weak expression in SAL005, AYR001 and STB390.

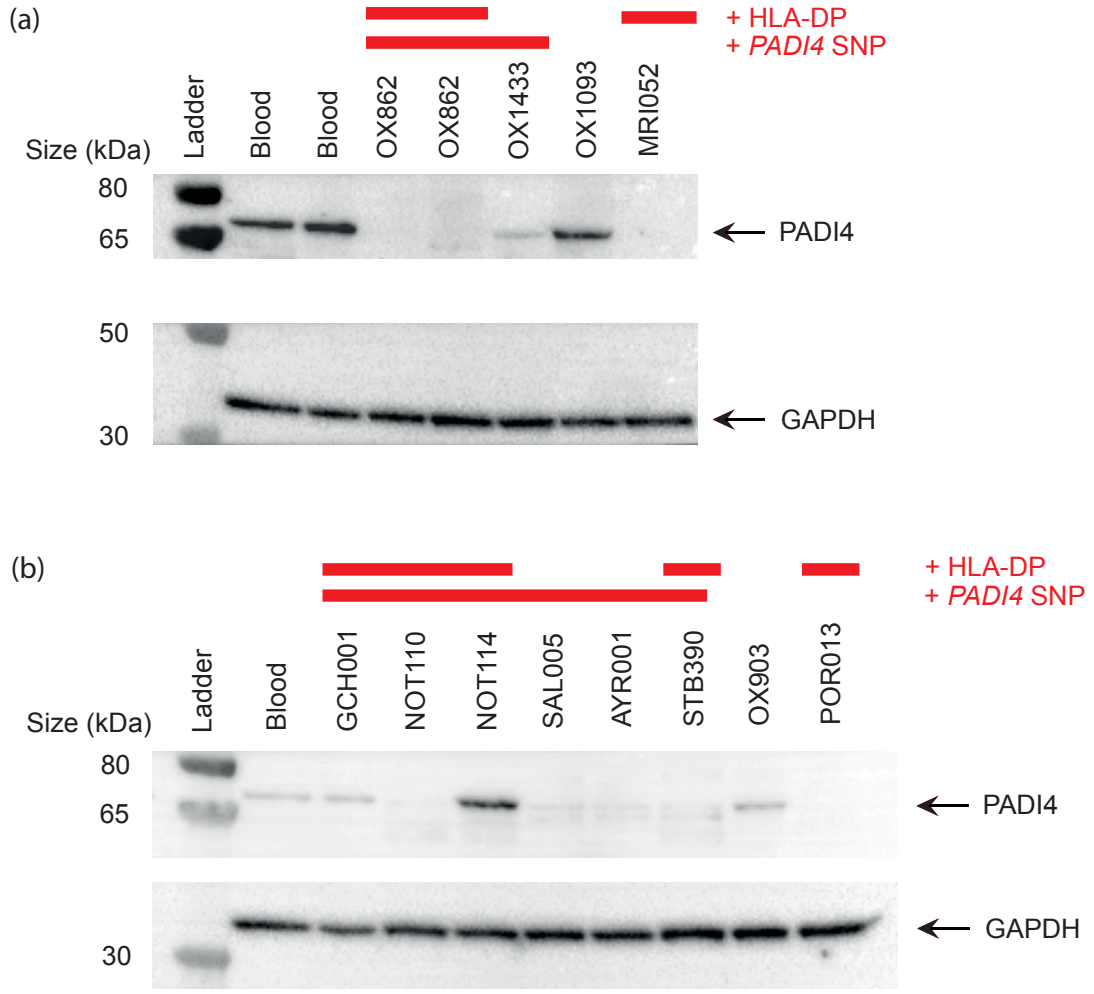


Figure 6.10: PADI4 protein expression. (a) and **(b)** Western blots depicting PADI4 expression (74 kDa) using protein extracts from primary AML samples. Positive controls: Blood (leukocytes) and OX1093 (AML sample with high PADI4 expression from RNA-Seq). Red bars above the blots indicate AML samples with the HLA-DPB1*04:01 allele and *PADI4* SNP. GAPDH was included as a loading control (36 kDa).

For some samples, expression of PADI4 was also assessed by real-time quantitative PCR (RT qPCR). RNA expression was detected for samples OX1093 and OX1433, with low expression by MRI052 (Figure 6.11). Correlation between RNA and protein expression was imperfect. Although OX1433 was had higher expression by RNA, OX1093 had greater expression by Western blot.

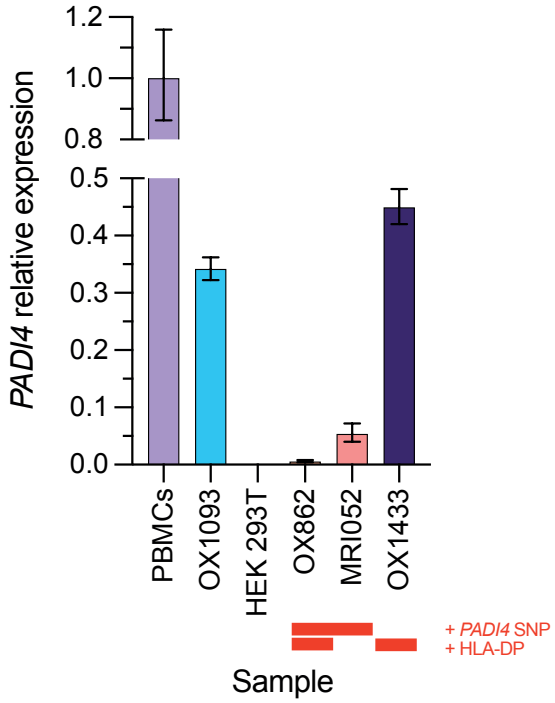


Figure 6.11: PADI4 RNA expression. RT qPCR assessment of *PADI4* expression. Positive controls: RNA from PBMCs and OX1093 (AML sample with high *PADI4* expression from RNA-Seq). Negative control: HEK 293T cell RNA. *PADI4* expression values were normalised to the housekeeping gene, *GAPDH*. Expression values shown are mean +/- SD of 2 technical replicates. Red bars below the plot indicate AML samples with the *PADI4* SNP and HLA-DPB1*04:01 allele.
RT qPCR: Real-time quantitative PCR, PBMCs: Peripheral blood mononuclear cells, SD: Standard deviation

Prior studies have shown that the majority of AML samples express high levels of HLA-DP [304]. Given that the PADI4 antigen is HLA-DP restricted, surface expression of HLA-DP was assessed in candidate target AML samples using flow cytometry, as this is required for recognition of target cells by TCRs (Figure 6.12).

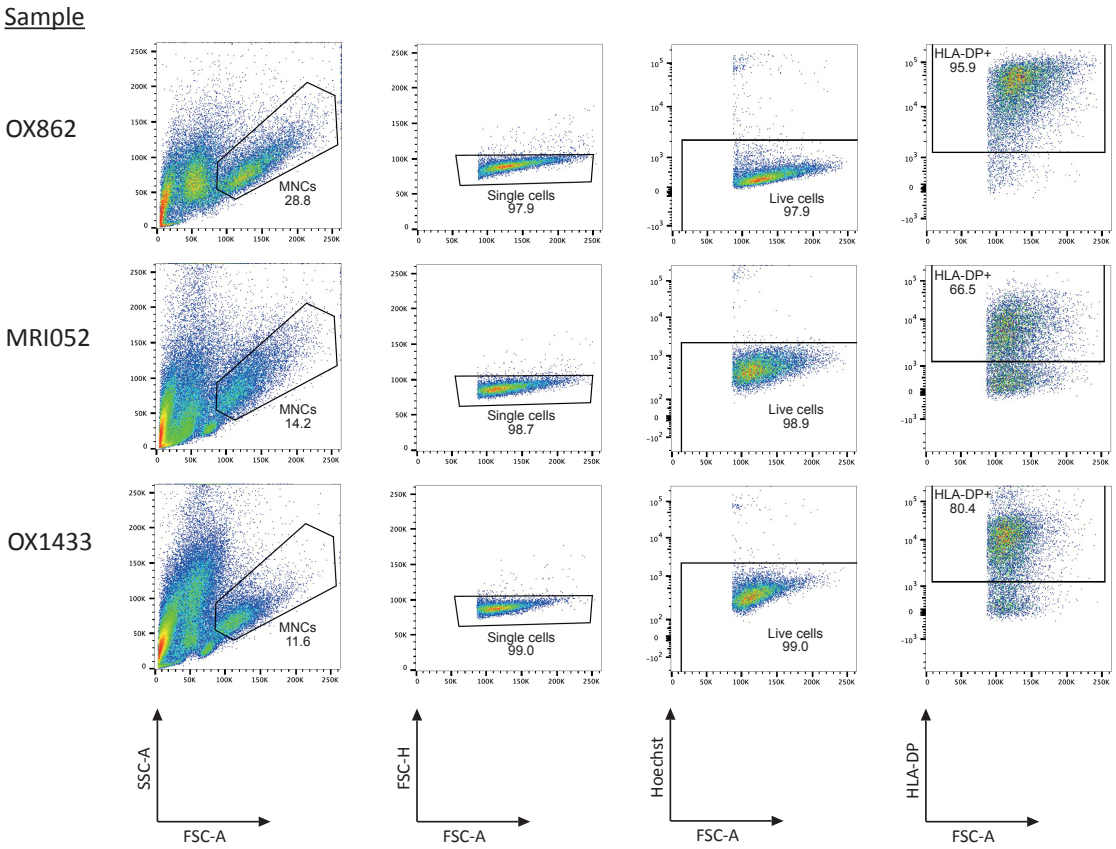


Figure 6.12: HLA-DP is expressed on AML cells. Flow cytometry was used to evaluate HLA-DP expression on candidate AML samples for functional assays. The sample name is shown on the left side. Gating of mononuclear cells (MNCs), single cells, live cells and HLA-DP$^+$ cells is shown.
MNCs: Mononuclear cells

Of the 41 AML samples with the *PADI4* variant, 40 were heterozygous and 1 was homozygous. Where DNA was available, HLA-DP typing was performed for samples with the *PADI4* variant. 56.4% had the restricting HLA-DP allele (HLA-DPB1*04:01), of which 22.7% were homozygous and 77.3% were heterozygous (Figure 6.13).

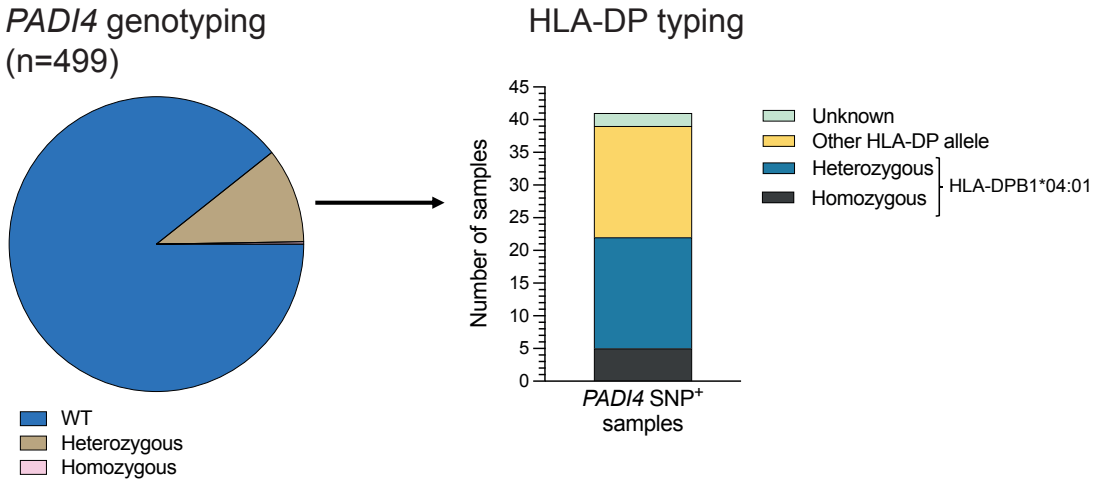


Figure 6.13: Summary of PADI4 genotyping and HLA-DP typing of AML samples. In total, 499 AML samples were genotyped and the *PADI4* SNP that produces the antigen was found in 41 samples (left of figure). The restricting HLA-DP allele (DPB1*04:01) was detected in 22 of these samples. 'Unknown' indicates samples with insufficient DNA available for typing.

Protein expression of PADI4 and HLA-DP was assessed in samples with the *PADI4* variant and HLA-DPB1*04:01 allele (Figure 6.14). PADI4 protein expression was detectable in 10 AML samples by Western blot and HLA-DP was expressed in 16 samples. A total of 8 samples expressed both PADI4 and HLA-DP.

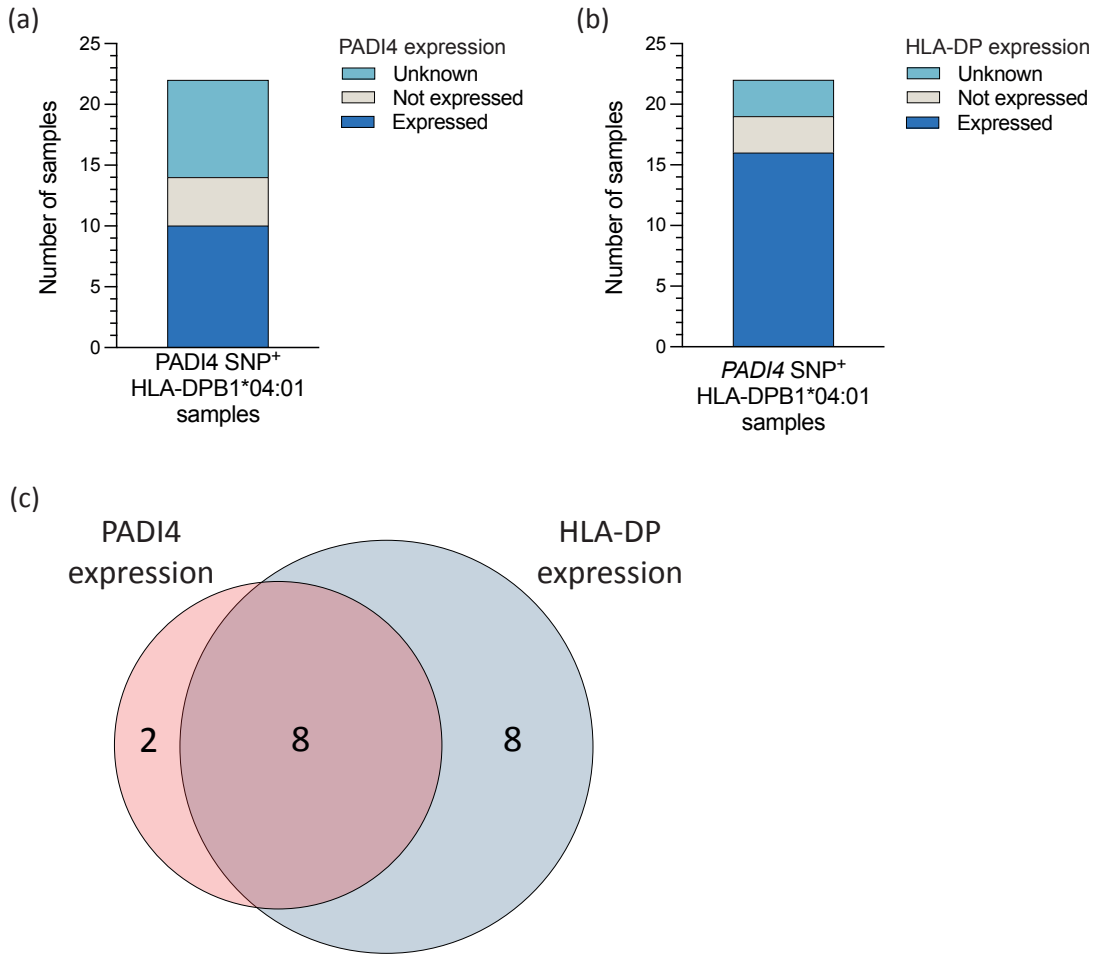


Figure 6.14: Summary of PADI4 and HLA-DP expression by AML samples. Protein expression of **(a)** PADI4, and **(b)** HLA-DP in samples with the *PADI4* SNP and HLA-DPB1*04:01 allele. Expression of PADI4 was assessed by Western blot and HLA-DP by flow cytometry. 'Unknown' indicates samples that were insufficient for analysis. **(c)** Venn diagram showing overlap of samples with expression of PADI4 and HLA-DP.

6.4 T cell activation by AML

The ability of AML cells to activate PADI4 T cell clones was assessed using an *in vitro* co-culture assay. First, T cells from a healthy donor were transduced and expanded for 2.5 weeks. They were then cultured with a primary AML sample that had the *PADI4* SNP and HLA-DPB1*04:01 allele (SNP+/HLA-DP+, sample OX862), and T cell activation was assessed by flow cytometry. To confirm that both the SNP and HLA-DP allele are required, control AML samples were included

that lacked one of these: SNP+/HLA-DP- (sample MRI052) and SNP-/HLA-DP+ (sample OX1433). Surface expression of HLA-DP was confirmed for all 3 samples used in the assay (Figure 6.12). Each AML was either untreated or had been pulsed with PADI4 peptide. Peptide pulsing involves addition of exogenous peptide antigen and will either bind surface HLA molecules directly or be endocytosed and presented. Untreated samples test for endogenous processing of the PADI4 protein and presentation of the peptide antigen by HLA.

Responses were only seen following peptide pulsing and were seen in both AML samples with the HLA-DPB1*04:01 allele (samples OX862 and OX1433), but not when this allele was absent (sample MRI052) (Figure 6.16). In keeping with the IFNγ ELISpot results, only PADI4 Clone 2.1 elicited responses. Both $CD4^+$ and $CD8^+$ cells were activated, as demonstrated by elevated expression of intracellular cytokines (IFNγ and TNFα) and CD107a in cells expressing murine TCRβ. Compared with peptide pulsing of T cells, pulsing of AML cells led to greater T cell activation, suggesting that antigen presentation by AML cells is more efficient. In the original patient used to identify the PADI4 antigen and cognate TCR, only $CD4^+$ cells were antigen-specific. However, activation of both $CD4^+$ and $CD8^+$ transduced T cells that express the TCR indicates that interaction of the HLA-peptide complex with the TCR is not dependent on the CD4 co-receptor.

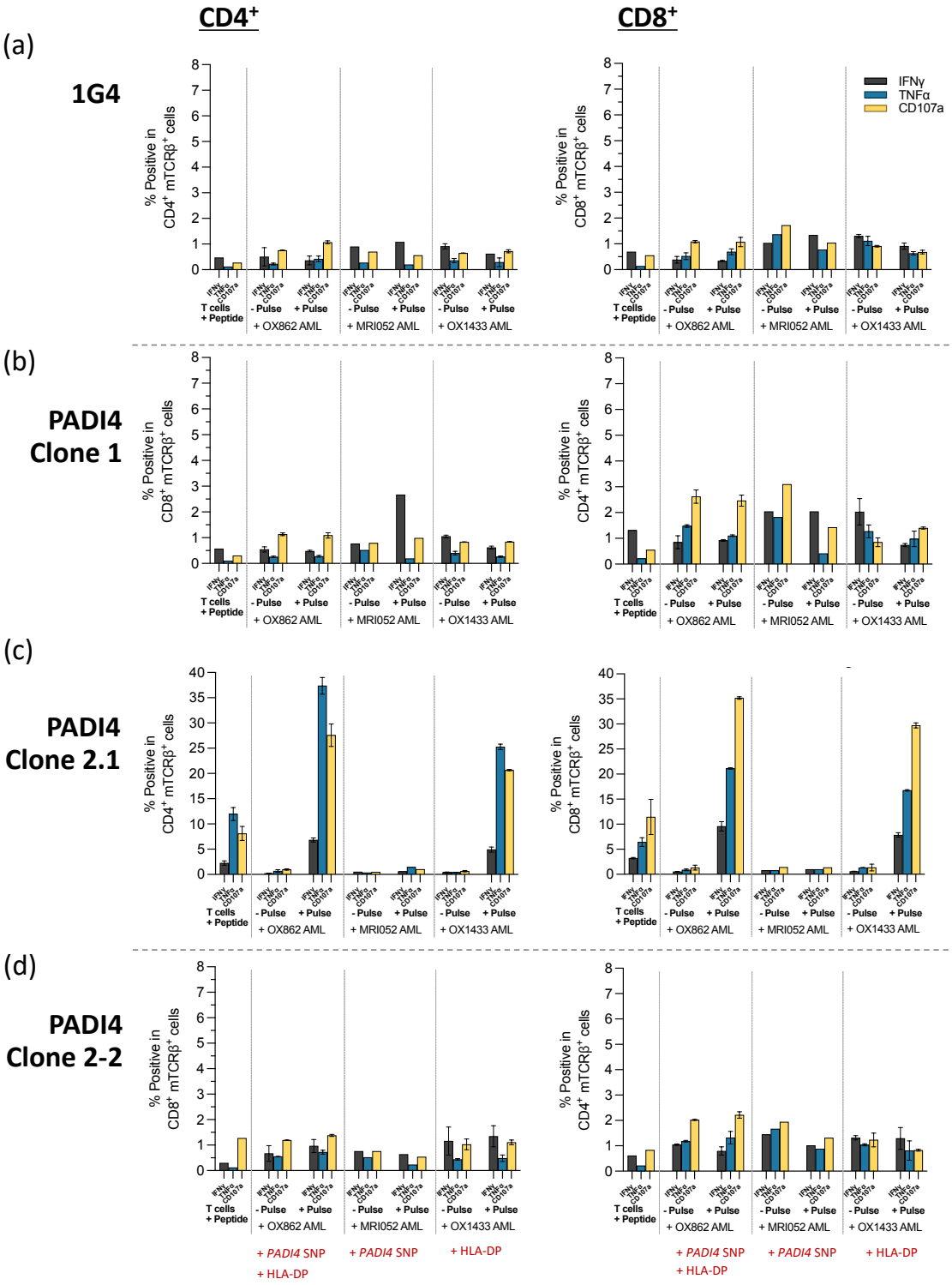

CD4+
CD8+
(a)
1G4
(b)
PADI4
Clone 1
(c)
PADI4
Clone 2.1
(d)
PADI4
Clone 2-2
% Positive in CD4+ mTCRβ+ cells
% Positive in CD8+ mTCRβ+ cells
IFNγ
TNFα
CD107a
T cells + Peptide
- Pulse
+ Pulse
+ OX862 AML
+ MRI052 AML
+ OX1433 AML
+ PADI4 SNP
+ HLA-DP

Figure 6.16: Transduced T cells are activated by AML that has been pulsed with PADI4 peptide. Summary of flow cytometry data with intracellular cytokine staining assessing T cell clones after 8-hour co-culture with AML. Left plots show $CD4^+$ T cells and right plots show $CD8^+$ T cells. Different T cell clones are shown in each row: **(a)** 1G4, **(b)** PADI4 Clone 1, **(c)** PADI4 Clone 2.1 and **(d)** PADI4 Clone 2.2. In each graph, T cells are shown that have been incubated with 3 different target AML samples (+/- pulsing with PADI4 peptide). In red below the plots, AML samples with the *PADI4* SNP and HLA-DPB1*04:01 allele are indicated. On the left of each graph, T cells are shown that have been pulsed with PADI4 peptide. Each bar shows the proportion of $CD4^+$mTCRβ^+ or $CD8^+$mTCRβ^+ cells that express IFNγ (grey bars), TNFα (blue bars) of CD107a (yellow bars). Values are mean +/- SEM of 2 replicates.
mTCRβ: Murine T cell receptor β

6.5 *In vitro* killing of AML by targeting the PADI4 antigen

A flow cytometry assay was next used to assess whether activation of PADI4 T cell clones in response to AML was associated with T cell-mediated killing. Target AML cells (sample OX862), which express HLA-DP but not *PADI4*, were first labelled with a cytoplasmic dye (cell trace violet) to enable detection by flow cytometry. They were then pulsed with PADI4 peptide. In 'spontaneous lysis' reactions, $5x10^4$ AML cells were cultured in a 200 μl volume on their own. When co-cultured with T cells, $5x10^4$ - $5x10^5$ T cells were added per reaction (ratios of T:AML cells of 1:1 to 10:1) in a final volume of 200 μl. Following a 24-hour co-culture with effector cells (PADI4 T cell clones or 1G4 control), AML cells were identified using flow cytometry and % specific lysis of target cells was calculated using the formula:

% Specific lysis = [(Target cell lysis with T cell clone – Spontaneous lysis) / (Maximum lysis – Spontaneous lysis)]*100

AML cell killing was observed with PADI4 Clone 2.1 and was proportional to the Effector:Target ratio (Figure 6.17). PADI4 Clone 1 did not result in killing, which is consistent with earlier data demonstrating lack of activation in response to the PADI4 peptide. Unexpectedly, the % specific lysis calculation for most reactions gave a negative value, as a result of the spontaneous death of AML cells being greater than

seen in the presence of T cells. Primary AML cells are challenging to culture *ex vivo* and there is significant variability in behaviour between samples [305]. The reason for the excess death in the spontaneous reactions may be that AML cells do not survive at the lower density seen in the absence of T cells. Therefore, in subsequent killing assays, the overall cell density was maintained at a constant, higher level by the addition of Clone 1G4 negative control T cells to the PADI4 T cell co-cultures.

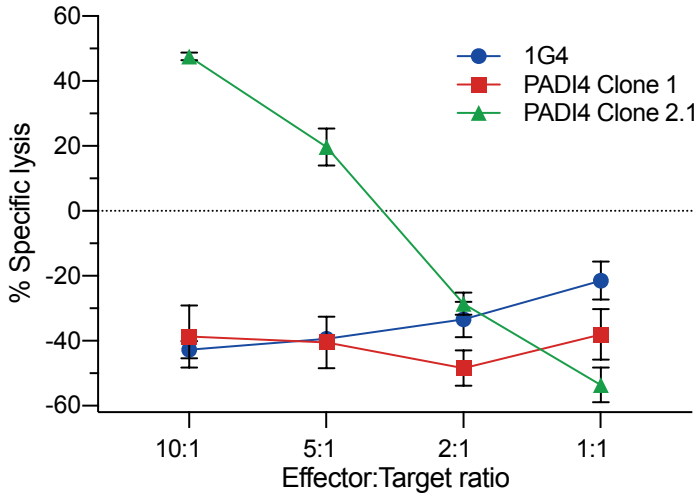


Figure 6.17: PADI4 Clone 2.1 kills AML cells that have been pulsed with PADI4 peptide. Specific lysis of AML sample OX862 (Target) by T cell clones (Effector) at decreasing Effector:Target ratios. T cell clones are: 1G4 control (blue), PADI4 Clone 1 (red) and PADI4 Clone 2.1 (green). Values are mean +/- SEM of 3 replicates.

To assess whether PADI4 Clone 2.1 can kill AML cells that endogenously process and present PADI4 antigen, a target sample was selected (GCH001) that had the *PADI4* SNP, restricting HLA-DP allele and protein expression of PADI4 and HLA-DP. Peptide pulsing of this AML target sample again resulted in cell killing by PADI4 Clone 2.1, but not 1G4 or PADI4 Clone 1. However, in the absence of peptide pulsing, there was no killing of AML cells, suggesting that endogenous expression of PADI4 protein is not leading to presentation of PADI4 antigen by HLA (Figure 6.18).

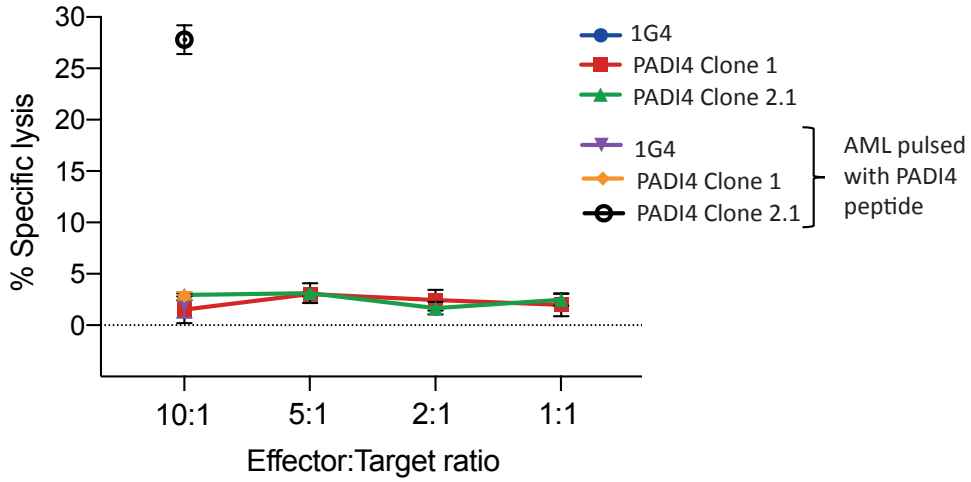


Figure 6.18: PADI4 Clone 2.1 kills AML cells only when pulsed with PADI4 peptide. Specific lysis of AML sample GCH001 (Target) by T cell clones (Effector) at decreasing Effector:Target ratios. To test endogenous presentation of antigen, AML without PADI4 peptide pulsing was tested at all Effector:Target ratios. Pulsed AML was only tested at an Effector:Target ratio of 10:1. The total number of T cells present in each reaction was maintained at a constant level across different Effector:Target ratios by the addition of 1G4 control cells. Values are mean +/- SEM of 3 replicates.

6.6 Assessing the mechanism of PADI4 antigen presentation

The absence of AML killing without peptide pulsing may be because to enter the HLA class II presentation pathway, extracellular PADI4 must be taken up by a professional antigen presenting cell by endocytosis. This possibility was tested by first generating monocyte-derived dendritic cells (MDDCs) from a healthy donor who is homozygous for HLA-DPB1*04:01. MDDCs express high levels of HLA-DP. Two sources of extracellular material containing PADI4 protein with the variant amino acid were tested: (1) culture media (supernatant, S/N) from an AML sample that has the PADI4 variant and expresses PADI4, and (2) cell lysate from the same AML sample, generated by freeze-thaw treatment. MDDCs were incubated with these for 18 hours, before addition of T cell clones (Figure 6.19). T cell activation was assessed by FACS after co-culture for 8 hours.

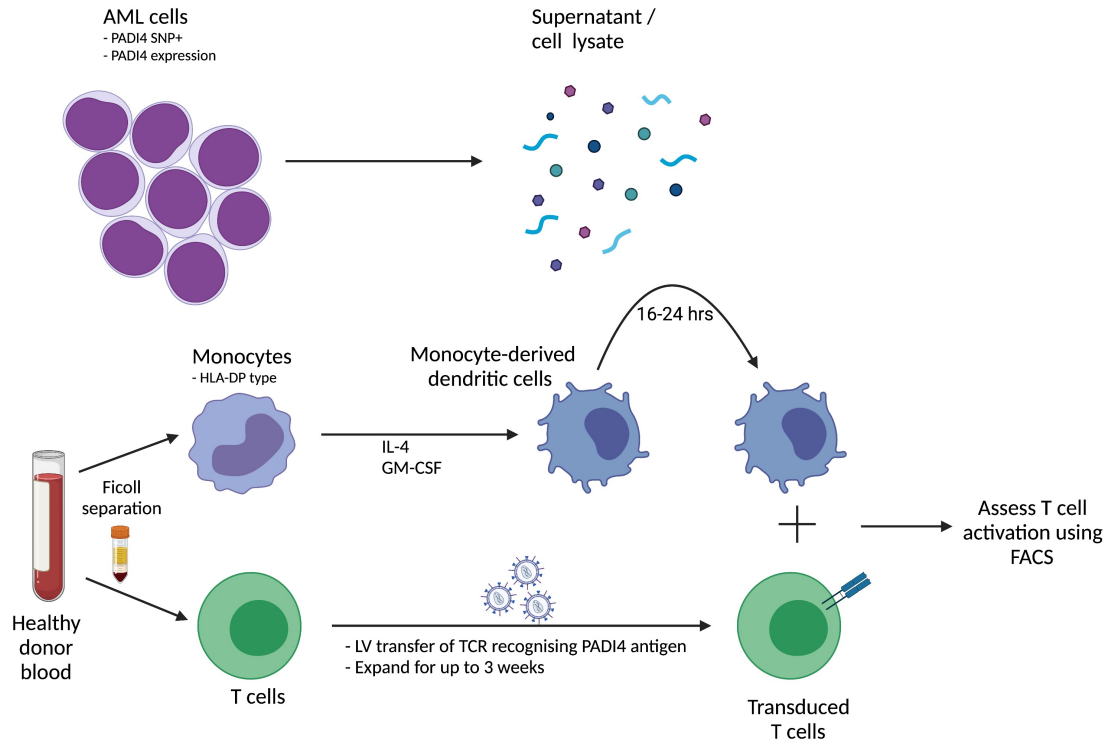


Figure 6.19: Testing extracellular antigen presentation by monocyte-derived dendritic cells Extracellular material containing PADI4 protein was prepared using two methods: (1) culture media (supernatant) from an AML sample that has the PADI4 variant and expresses PADI4, and (2) cell lysate from the same AML sample, generated by freeze-thaw treatment. Monocyte-derived dendritic cells (MDDCs) were produced from healthy donor PBMCs (homozygous for HLA-DPB1*04:01) and incubated with either AML culture supernatant or lysate. PADI4 or 1G4 control T cell clones were then incubated with the MDDCs for 8 hours and T cell activation was assessed by FACS. *Created with BioRender.com*

There was evidence of $CD4^+$, but not $CD8^+$, T cell activation when PADI4 Clone 2.1 was incubated with MDDCs treated with either S/N or lysate (Figure 6.20). There was an increase in cells expressing TNFα from 0.50% with untreated MDDCs to 3.30% / 2.61% with S/N or lysate treatment, respectively. This increase was not seen with 1G4 control T cells.

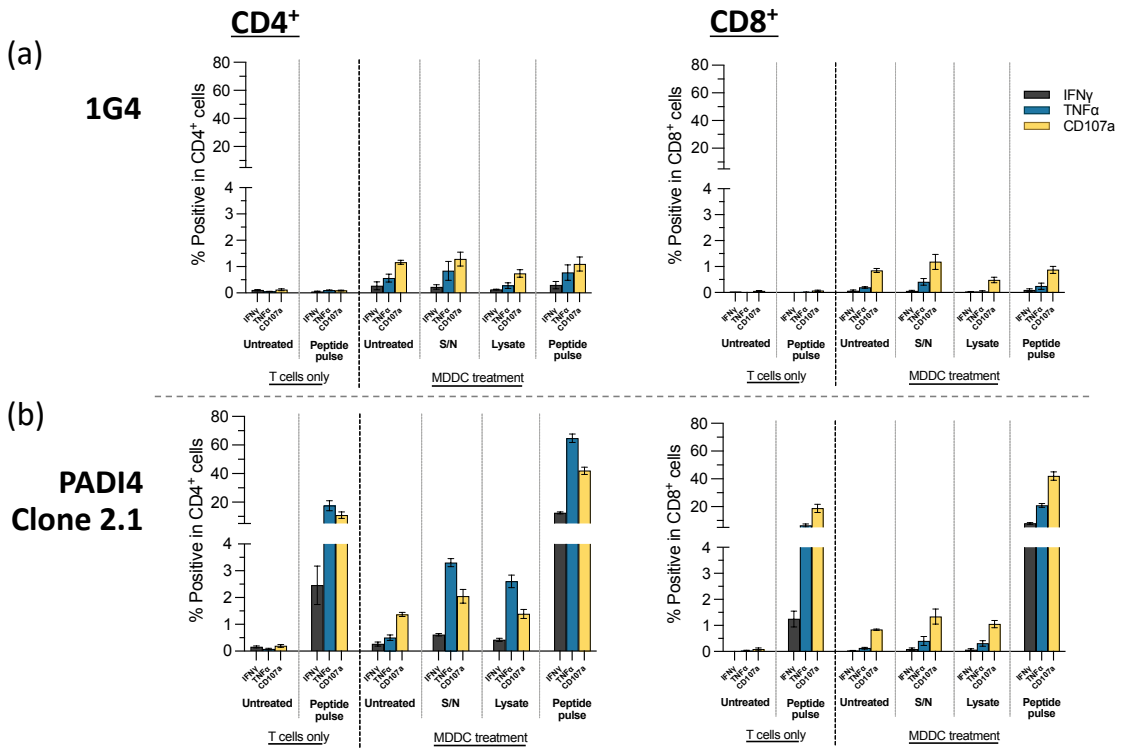


Figure 6.20: MDDCs present exogenous PADI4 antigen. Summary of flow cytometry data assessing T cell activation after 8-hour co-culture with MDDCs. Left plots show $CD4^+$ T cells and right plots show $CD8^+$ T cells. T cell clones are shown in each row: **(a)** 1G4 and **(b)** PADI4 Clone 2.1. On the left of each graph, T cells are shown that have been pulsed with PADI4 peptide. In each graph, MDDCs have been treated in different ways and then incubated with T cells. MDDC treatments were: (1) Untreated, (2) AML supernatant, (3) AML lysate, and (4) PADI4 peptide. Each bar shows the proportion of $CD4^+$mTCRβ^+ or $CD8^+$mTCRβ^+ cells that express IFNγ (grey bars), TNFα (blue bars) of CD107a (yellow bars). Values are mean +/- SEM of 3 replicates. S/N: Supernatant, MDDC: Monocyte-derived dendritic cell

An interesting possibility is that AML cells could act as professional antigen presenting cells (APCs). Although this phenomenon has not been shown *in vivo*, AML can be differentiated *in vitro* into cells with the immunophenotype of APCs and enhanced antigen presentation capability [306]. This could have clinical significance in patients with AML. After being killed by chemotherapy, AML cells release their contents and PADI4 within this milieu may be endocytosed by residual AML cells, which then present antigen. Therefore, presentation of extracellular antigen by AML was tested with a similar approach as used for MDDCs (Figure 6.21).

$CD4^+$, but not $CD8^+$, T cells from Clone 2.1 were activated. Responses were seen with T cell incubation with untreated AML, suggesting that endogenously presented

peptide is able to activate T cells, with no increment seen with the prior incubation of AML with extracellular antigen. There was no activation of 1G4 control cells.

When either MDDCs or AML were used as APCs, the activation of T cells by extracellular antigen was much lower in magnitude to that seen following peptide pulsing of either the T cells directly, or pulsing AML. This is likely to be because peptide pulsing causes a much greater amount of peptide antigen to bind HLA-DP on the cell surface. Peptide pulsing also led to activation of both $CD4^+$ and $CD8^+$ cells. Therefore, although CD4 is not essential for activation in the presence of abundant antigen, it may be important for lower magnitude responses.

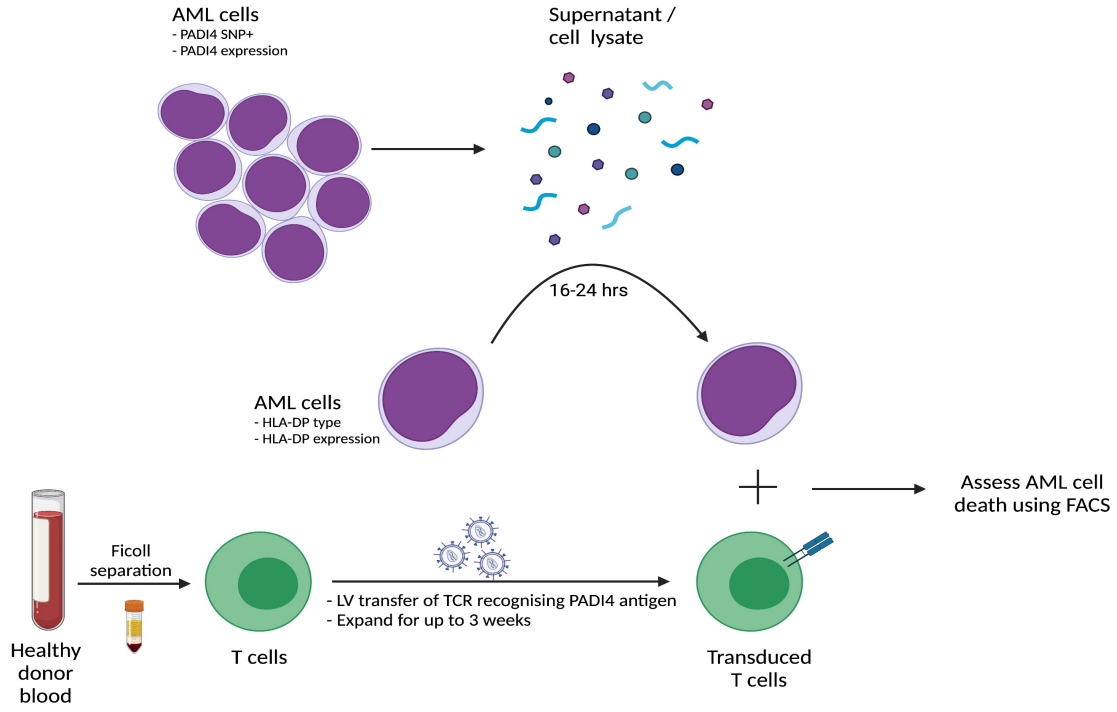


Figure 6.21: Testing antigen presentation by AML cells. Extracellular material containing PADI4 protein was prepared using two methods: (1) culture media (supernatant) from an AML sample that has the PADI4 variant and expresses PADI4, and (2) cell lysate from the same AML sample, generated by freeze-thaw treatment. AML cells (heterozygous for HLA-DPB1*04:01 and HLA-DP surface expression confirmed) were incubated with either AML culture supernatant or lysate. PADI4 or 1G4 control T cell clones were then incubated with the AML for 8 hours and T cell activation was assessed by FACS.
Created with BioRender.com

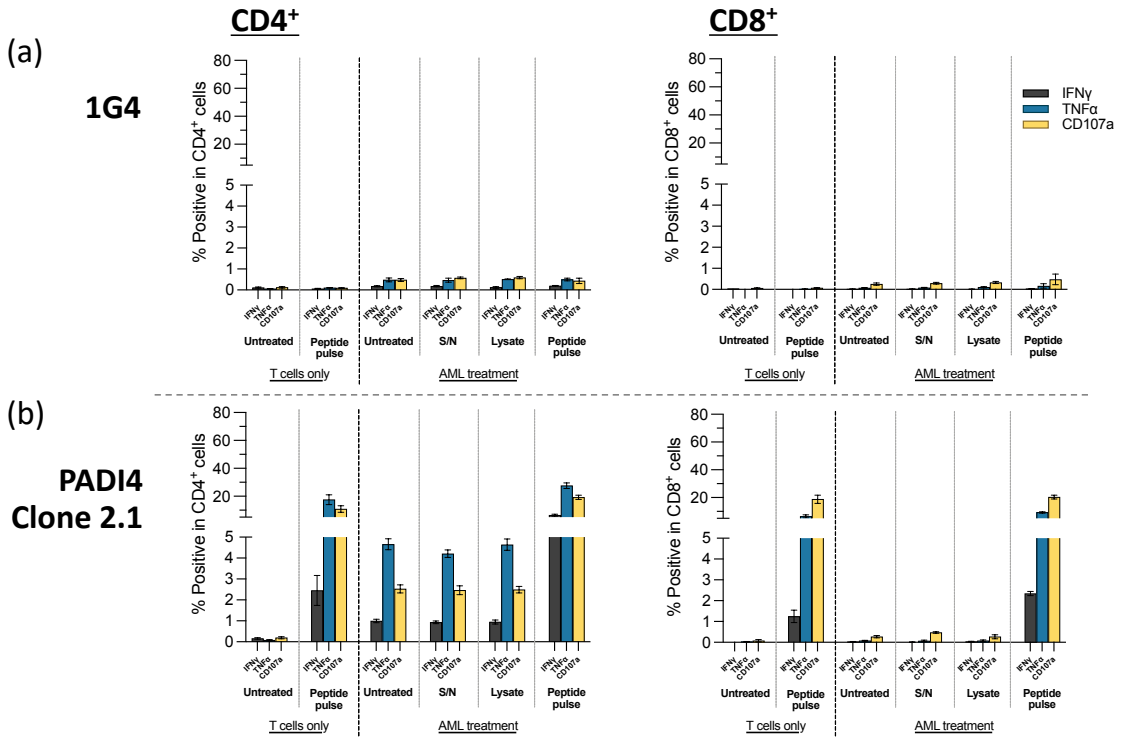


Figure 6.22: AML activates T cells. Summary of flow cytometry data assessing T cell activation after 8-hour co-culture with AML. Left plots show $CD4^+$ T cells and right plots show $CD8^+$ T cells. T cell clones are shown in each row: **(a)** 1G4 and **(b)** PADI4 Clone 2.1. On the left of each graph, T cells are shown that have been pulsed with PADI4 peptide. In each graph, AML has been treated in different ways and then incubated with T cells. AML treatments were: (1) Untreated, (2) AML supernatant, (3) AML lysate, and (4) PADI4 peptide. Each bar shows the proportion of $CD4^+$mTCRβ^+ or $CD8^+$mTCRβ^+ cells that express IFNγ (grey bars), TNFα (blue bars) of CD107a (yellow bars). Values are mean +/- SEM of 3 replicates.
S/N: Supernatant

6.7 *In vivo* killing of AML by targeting the PADI4 antigen

In ongoing experiments, the anti-leukaemic efficacy of PADI4 Clone 2.1 is being evaluated *in vivo*. Primary human AML samples with the *PADI4* SNP, HLA-DP type and protein expression of PADI4 and HLA-DP have been injected intra-tibially into immunodeficient NSGW (NOD/SCID/$Il2r\gamma^{-/-}/Kit^{W41}$) mice to generate patient-derived xenografts. These mice have functional defects in T and B cell receptors and lack the IL-2 gamma chain. They also have a mutation in *Kit*, which permits

engraftment of human haematopoietic cells without preceding irradiation. Following AML xenotransplantation, rates of engraftment are dependent on mouse strain, with up to two-thirds of human AML samples engrafting in previous studies [307]. Following AML engraftment, either PADI4 Clone 2.1 or 1G4 control T cells are injected intravenously. Both $CD4^+$ and $CD8^+$ T cells are injected in a 1:1 ratio. The outcomes being assessed are: (1) mouse survival, and (2) % AML engraftment in blood each week after T cell injection.

Four different AML samples have been injected into mice and two have engrafted. Each sample exhibits different kinetics of engraftment. AML is detected in blood at different times following intra-tibial injection and the rate of progression in the mice differs. As many T cell immunotherapies are most effective in the setting of low levels of disease, we aim to model a similarly low disease burden in the mice [175, 178].

Mice injected with a primary AML sample that engrafted rapidly, within 6 weeks (Sample OX862), were injected intravenously with either PADI4 Clone 2.1 or control T cells (Figure 6.23). However, the rapid engraftment was associated with a very high level of AML disease burden (98%) in each mouse and the mice were sacrificed before the experiment could be completed. In ongoing experiments, T cell clones will be administered to mice with lower AML burden.

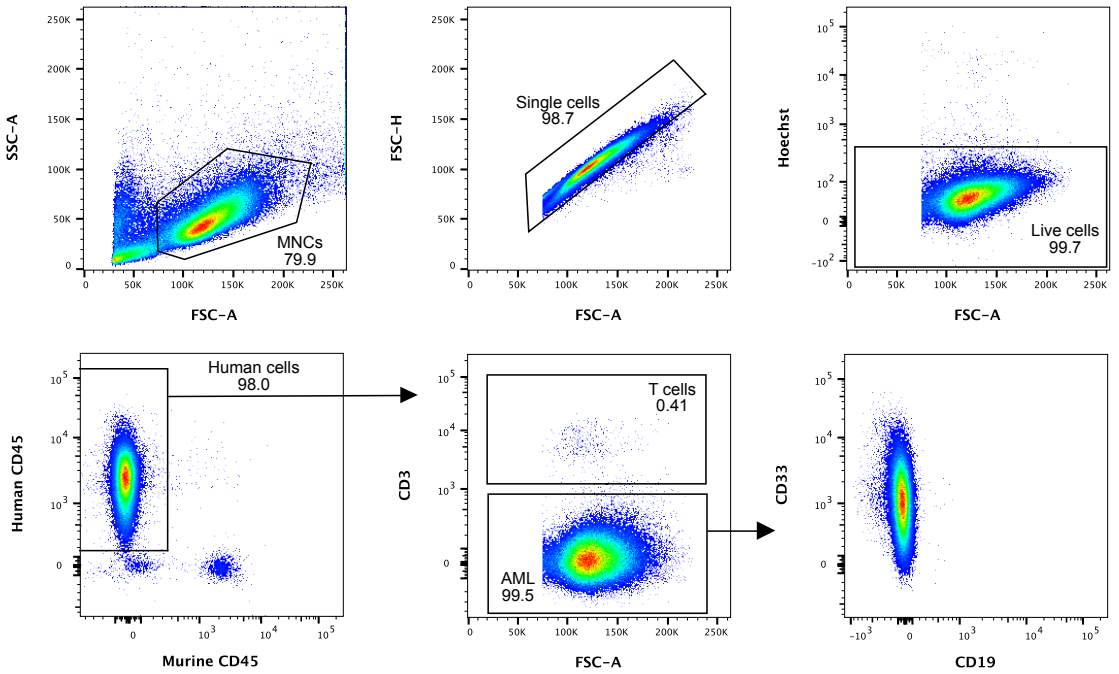


Figure 6.23: Preliminary testing *in vivo* killing of AML by PADI4 T cell Clone 2.1. Intra-tibial injection of AML was followed 6 weeks later by intravenous injection of PADI4 Clone 2.1 T cells. The flow cytometry plots show peripheral blood assessed from a mouse 3 days later, at the time of culling. Human $CD45^+$ cells account for 98% of live cells. Of these, 0.41% were $CD3^+$ T cells. Unless otherwise stated, the markers indicated on axes are human.

CD34-enriched bone marrow cells from healthy donors have also been injected into mice, which will enable the assessment of PADI4 Clone 2.1 targeting normal human haematopoietic cells, which could indicate a source of toxicity if used therapeutically in patients.

6.8 Conclusions: Targeting the PADI4 antigen using adoptive T cell therapy

A T cell therapy targeting the PADI4 antigen was developed using putative antigen-specific TCRs identified in Section 4.7. Lentiviral transfer of TCRs into healthy donor human T cells resulted in highly efficient transduction and surface expression of the inserted TCRs.

Antigen specificity was confirmed for PADI4 Clone 2.1. Recognition and killing of AML cells with the restricting HLA-DP allele was demonstrated after pulsing the AML cells with exogenous PADI4 peptide. This indicates that the peptide binds to

HLA-DP on the surface of the AML cells and the HLA-peptide complex is then able to bind the TCRs. However, to understand how the the antigen can be processed and presented to T cells in patients, we assessed whether endogenous and/or exogenous sources of PADI4 protein can enter the HLA class II presentation pathway. Results varied between different AML samples. Endogenous presentation by AML cells was shown, as was presentation of exogenous antigen by monocyte-derived dentritic cells.

In ongoing experiments, killing by PADI4 Clone 2.1 is being assess using further AML samples *in vitro*, and also *in vivo*, using immunodeficient mice engrafted with human AML in a patient-derived xenograft model.

Chapter 7

Discussion

Contents

7.1 Clinical importance of identifying of GvL antigens . . 223
7.2 Prediction of GvL antigens 224
7.3 Identification and characterisation of the PADI4 antigen 226
7.3.1 Functional screening of predicted HLA-binding peptides 226
7.3.2 Phenotype of PADI4 antigen-specific T cells 229
7.4 The antigenic landscape following allogeneic stem cell transplantation 230
7.4.1 Tissue distribution of antigen expression 230
7.4.2 HLA restriction of antigens 231
7.4.3 Tracking antigen-specific T cell responses post-transplant 232
7.4.4 Clinical impact of alloreactive antigens following allogeneic stem cell transplantation 232
7.5 Targeting GvL antigens using adoptive T cell therapy 233
7.5.1 Targeting the PADI4 antigen 233
7.5.2 Clinical implications of TCR therapy 235

7.1 Clinical importance of identifying of GvL antigens

Allogeneic haematopoietic stem cell transplantation produces a therapeutic effect in AML patients by two mechanisms. First, by re-establishing the haematopoietic system with donor cells, it enables very high doses of chemoradiotherapy to be delivered that would otherwise be unacceptably toxic to normal haematopoiesis. Second, donor cells produce a specific alloimmune response against recipient AML cells, which directly targets the disease [1]. T cells have an established role in eliminating malignant cells and can reduce relapse rates through the GvL effect. For a GvL effect to be established, donor-derived T cells must recognise AML cells as foreign and mount a killing response. The current approach to donor selection aims to reduce immune responses directed against healthy tissues, which manifests

as GvHD [29]. However, no selection occurs to increase the likelihood of a GvL effect, with the exception of avoidance of identical twin donors [45].

In this study, minor histocompatibility antigens that could represent GvL antigens have been identified and the associated T cell responses characterised. The identification of alloreactive antigens that produce responses from donor-derived T cells against AML has two potential clinical applications: (1) improvement in donor selection for individual transplant recipients, to introduce patient-donor DNA mismatches that result in GvL antigens, and (2) GvL antigens could be targeted by immunotherapy, such as by the administration of antigen-specific T cells (adoptive T cell therapy).

7.2 Prediction of GvL antigens

As reported in Chapter 3, an unbiased reverse immunology approach was used to predict and identify immunogenic peptides following allo-SCT. This high-throughput method of antigen discovery starts from differences in coding DNA sequences between patients and their respective donors. The majority of exonic differences between patient AML and donor germline (98.7%) are due to inherited differences between the individuals rather than cancer-specific mutations acquired during leukaemia development. Expression of patient-donor mismatched variants in RNA-Seq reads were detected for 28.3% of germline and 20.8% of somatic variants. The lower proportion of somatic variant detection is likely to be due the sensitivity of RNA-Seq being limited by sequencing read depth. Due to the subclonal structure of AML, certain somatic variants are only be present in a small proportion of cells, whereas germline variants are present in all cells. Therefore, the RNA of somatic variants present in small subclones may not have been detectable.

For peptides resulting from patient-donor mismatched variants that are expressed in AML, *in silico* prediction of HLA-binding affinity was used to produce libraries of candidate immunogenic peptides for functional testing. Predictions for HLA class I peptide ligands have a high level of accuracy, especially following the incorporation of mass spectrometry data into more recent training datasets for

machine learning algorithms [308]. However, this is not the case for HLA class II ligands, which remain difficult to predict. In the current study, the accuracy of *in silico* prediction of HLA-binding affinity was tested using the peptides detected by tandem mass spectrometry (LC-MS/MS). For each of the 15 patients, all unique peptide sequences detected by LC-MS/MS were inserted into NetMHCpan and NetMHCIIpan prediction algorithms with patient-specific HLA alleles. Prediction was correct for 67% of HLA class I peptides but only 18% for class II peptides.

Compared with HLA class I, modelling of class II ligand binding presents multiple challenges: (1) The HLA-binding affinity training datasets for the machine learning algorithms are smaller. (2) Class II peptides are long and current methods rely on predicting the position of the binding core within the peptide sequence, which is important for interactions with the peptide-binding groove [161]. (3) Class II molecules have greater degeneracy and less specific motifs. (4) HLA-DP and -DQ are hetermodimers and both subunits are polymorphic, which increases the diversity of alleles [309]. Another factor that could influence efficient prediction of binding by class II ligands is the mode of action of the complex machinery of antigen presentation through this pathway. Conformational flexibility may be important in loading peptides onto class II molecules, which is difficult to predict from the amino acid sequences. In this study, a total of 24 antigens were identified (Table 5.2), but the total number of true alloreactive antigens that could elicit a GvL effect in these patients is unknown because the limited accuracy of HLA-binding prediction reduces the sensitivity of antigen detection.

Due to the limitations of *in silico* prediction, in parallel an immunopeptidomic approach was used to directly identify HLA-bound peptides (Section 3.5). Previous studies have found that functional peptide-specific T cell responses are elicited by a greater proportion of the antigens identified using this method compared with *in silico* prediction [168, 169]. HLA-peptide complexes were immunoprecipitated and the peptide cargo assayed by LC-MS/MS. However, only a small number of peptides were detected. To saturate the ability of this technique to detect unique peptides, $5x10^8$ cells are typically required as input. However, due to limited

availability of samples from patients, only between $5x10^6$ and $5.6x10^7$ cells could be analysed. Therefore, only the most abundant peptides are likely to have been identified. Furthermore, even when input cell number is not limiting, due to the physico-chemical properties of peptides, some are not readily ionised and therefore may be undetectable by LC-MS/MS [120].

The approach employed to predict antigens in this study identifies neoantigens and minor histocompatibility antigens, but other classes of cancer antigen are not detected. Leukaemia-associated antigens (e.g. PRAME, and WT1), which are expressed in AML but not in most healthy tissues, do not result from genetic differences between patient and donor and were not investigated. Dysregulation of RNA splicing is common in human cancers and can lead to neoepitopes by the formation of novel exon-exon junctions through intron retention, exon skipping, mutually exclusive exons and alternative 3' and 5' splice site changes [310–312]. Splice variant proteins can result from mutations that create splice sites or from mutations of the spliceosome components themselves, which lead to dysregulated splicing. Mutations in spliceosome components, such as *SF3B1* or *SRSF2*, are found in up to 20% of AML. Furthermore, post-translational protein modifications such as phosphorylation and glycosylation could also produce neoepitopes [313].

T cells have the most established role in producing a GvL effect following allo-SCT and these responses were the focus of the current study. However, other immune cells including NK cells are also likely to be important [90].

7.3 Identification and characterisation of the PADI4 antigen

7.3.1 Functional screening of predicted HLA-binding peptides

Functional screening is necessary because HLA-binding predictions have limited accuracy and even if the prediction is correct, the peptide may not be capable of activating T cells. To elicit a T cell response, the peptide must be processed within the cell, loaded onto HLA, the HLA-peptide complex presented on the cell surface and a T cell with the cognate TCR must be present and capable of becoming

activated. The importance of factors beyond HLA binding was emphasised by the immunopeptidomic results from LC-MS/MS, which analysed peptides eluted directly from HLA molecules on AML cells. A total of 25 patient-donor mismatched peptides were screened by IFNγ ELISpot, but none reproducibly elicited T cell responses (Section 5.8). This finding highlights that additional factors beyond HLA binding play an important role in immunogenicity. Therefore, it was essential to functionally screen predicted HLA-binding peptides for their ability to elicit T cell responses.

During normal T cell development, self-reactive T cells are negatively selected in the thymus and undergo apoptosis to produce central tolerance. Following allo-SCT, T cells are present that have never been exposed to recipient antigens and therefore have not been eliminated by donor thymic selection. Therefore, these T cells can elicit immune responses against recipient tissues, which may manifest as GvL or GvHD. This study sought to identify these T cell clones that have not been eliminated by immune tolerance in the allo-SCT donor. For IFNγ ELISpot screening, post-transplant blood samples were used from the patient in whom the peptide antigens had been predicted.

Previous studies investigating cancer-specific neoantigens found that approximately 4% of predicted HLA-binding peptides can be validated, either by the detection of antigen-specific T cells using HLA-peptide multimers or by demonstration of T cell activation in response to peptide [163]. In the current study, the vast majority of screened peptides did not result from cancer-specific DNA variants, but inherited differences between patient and donor. In this novel setting, the expected proportion that would be functionally validated was unknown.

Of 5,340 peptides screened, 24 were functionally validated (0.45%). Although blood samples were cultured with peptides to enrich for antigen-specific T cells, some low frequency alloreactive clones may still be undetectable. The blood used for screening was taken between 1 and 6 years following allo-SCT. It is possible that clinically significant alloreactive T cell responses were present early post-allo-SCT, but that the clones driving these responses diminished over time as AML cells were eradicated, thereby removing the antigenic stimulus that was maintaining them.

Immune tolerance mechanisms following allo-SCT may also reduce the activity of potential alloreactive T cells and therefore reduce the number of antigens detected by IFNγ ELISpot screening. Successful immune reconsititution following allo-SCT requires the donor immune cells to develop tolerance to alloantigens that reduces morbidity from GvHD, while retaining the ability to recognise and respond to pathogens. However, tolerance mechanisms are also likely to dampen therapeutic GvL responses directed against residual AML cells. Three mechanisms have been implicated in promoting immune tolerance after allo-SCT:

(1) T cell negative selection in the thymus removes T cells with high avidity TCRs to HLA-peptide complex on antigen-presenting cells in the thymus. However, in AML patients, this central tolerance mechanism may be impaired by thymic damage inflicted by chemotherapy and radiation delivered in the transplant conditioning regimen [314].

(2) Active immune suppression by regulatory T cells (Tregs), which are generated in the thymus and peripheral lymphoid tissues. The level of Tregs inversely correlates with the severity of GvHD and adoptive transfer of *ex vivo* expanded Tregs into patients has successfully been used to treat GvHD [315, 316].

(3) Induction of T cell anergy, which is a reversible state of inactivation in which T cells are unable to respond to antigen despite encountering HLA-peptide complex due to inadequate co-stimulatory signals [317].

The post-transplant blood used for screening was taken up to 6 years following allo-SCT, including in Patient OX866 (Section 4.2). The detection of antigen-specific T cell clones at this stage raises the interesting possibility of ongoing immune surveillance of AML. It is possible that alloreactive T cells that are producing the strongest and most clinically relevant anti-cancer responses are those that are present at later timepoints post-allo-SCT. In the first months following allo-SCT, T cell repertoires have low diversity and are dominated by pathogen-specific T cells. In particular, CMV and EBV are often associated with large T cell clones [318]. Therefore, screening for alloreactive peptides at early timepoints may not yield positive results as a productive GvL response has yet to be established.

The peptide screening method by ELISpot is dependent on IFNγ secretion by antigen-specific T cells. This cytokine is released predominantly by activated lymphocytes, including CD4$^+$ Th1 cells, cytotoxic T cells, $\gamma\delta$ T cells and NK cells, in response to diverse stimuli [319]. Nevertheless, this method is biased in favour of detecting responses from T cell clones releasing IFNγ and certain responses could be missed, such as those from Th2 cells. Other screening methods that could be used include TNFα ELISpot or HLA-peptide multimers [162]. However, the latter method would be much more expensive and labour-intensive and therefore unsuitable for high-throughput screening of thousands of peptides.

7.3.2 Phenotype of PADI4 antigen-specific T cells

HLA class I-restricted CD8$^+$ T cell responses are the focus of much research in the cancer immunology field. However, HLA class II is expressed at a high level in AML cells and studies have observed transcriptional downregulation of HLA class II expression at post-allo-SCT relapse in up to half of cases, including following HLA-matched transplants [42, 43]. This common mechanism of immune evasion emphasises the importance of peptides presented by HLA class II for GvL responses against AML. Correspondingly, the majority of antigens identified in this study (86%), including the PADI4 antigen, were class II-restricted and elicited responses from donor-derived CD4$^+$ T cells.

The phenotype of PADI4 antigen-specific T cells from Patient OX866 was assessed using single cell RNA-Seq and demonstrated activated CD4$^+$ T helper cells with a cytotoxic phenotype (Section 4.6.2). CD4$^+$ T cells with cytotoxic activity have been described in different immune responses, including to viruses and cancer, and are characterised by their ability to secrete granzyme B and perforin [272]. They have recently been described as playing a key role in autologous responses against human bladder cancer by targeting HLA class II-restricted epitopes [295]. A marked increase in CD4$^+$ cytotoxic T cells has also been described in very elderly individuals, resulting from marked expansion of a small number of clones, but the functional significance of this finding is unclear [320]. However, a caveat to the

interpretation in the current study is that the phenotype of the T cells in the patient may have been altered by 12 days of *in vitro* culture with peptide, IL-2 and IL-7 prior to analysis and therefore might not accurately reflect the phenotype of PADI4 antigen-specific T cells responding to the antigen *in vivo*.

7.4 The antigenic landscape following allogeneic stem cell transplantation

7.4.1 Tissue distribution of antigen expression

Screening of predicted HLA-binding patients led to the identification of 24 functionally validated alloreactive antigens in 9 patients. The functions of the proteins from which the peptides are derived are unlikely to be vital for their roles as antigens. However, a high level of protein expression and ability to enter the HLA class I or II presentation pathways are important, as this increases the density of HLA-presented peptide on the cell surface.

Expression in normal tissues could be a barrier to the clinical utility of antigens due to the risk of toxicity. However, peptides were not excluded from functional screening on this basis. Unlike HLA class I, which is ubiquitously expressed on nucleated cells, class II is expressed only expressed on certain cell types. Therefore, despite expression of genes that results in antigens, class II-restricted antigens are not presented in the majority of cells and therefore may not result in toxicity if targeted therapeutically. Interestingly, HLA class II is upregulated in inflammatory states, including cancer, which may enable targeting of the identified germline-derived antigens in multiple cancers [292, 294]. However, a pro-inflammatory environment can also cause increased HLA class II expression in healthy cells, which could lead to GvHD.

The patient with most GvHD (Patient OX289) also had the greatest number of antigens identified, suggesting that some of these antigens could be responsible for immune responses against normal tissues. Of the 10 antigens identified in the patient, 3 elicited $CD8^+$ T cell responses, which were not detected in any other patient.

Due to widespread HLA class I expression across healthy tissues, class I-restricted antigens that drive $CD8^+$ T cell responses may be more likely to result in GvHD.

PADI4 has selective expression in haematopoietic tissues and the patient with this antigen (Patient OX866) had evidence of GvL without GvHD (Table 5.1). In this allo-SCT setting, the PADI4 antigen is not found in the engrafted donor-derived haematopoietic cells and therefore T cell clones targeting the antigen recognise patient-derived AML cells, but not the healthy post-transplant blood cells.

7.4.2 HLA restriction of antigens

For 14 of 24 antigens studied, the HLA restriction of peptide responses was charactered using HLA-blocking antibodies. Patient OX866 is homozygous for the HLA-DPA1*01:03-DPB1*04:01 haplotype and therefore accurate allele restriction was determined following blockade of responses with anti-HLA-DP antibody (Section 4.3.3). However, in most cases there are multiple alleles and therefore HLA antibody blockade does not precisely identify the restricting allele. For the HLA class I-restricted responses, effective individual HLA-A, -B and -C blocking antibodies are not available and therefore full restriction could not be determined.

Classic approaches to determine HLA restriction involve antigen presentation assays with panels of cell lines that are matched or mismatched at HLA loci. The pattern of T cell activation can help identify the restricting allele(s) [321]. An alternative approach is to use APCs that have been transfected with single HLA alleles. T cell responses observed after incubation with these APCs that have been pulsed with peptide antigen indicates the HLA restriction. This technique was attempted for the PADI4 antigen, using .221 B lymphoblastoid cells that were transfected with single HLA class I alleles from Patient OX866 (Figure 4.13). However, the majority of T cell responses detected in this study were HLA class II-restricted and to assess restriction with this approach, a panel of APCs would need to be used, each cell transfected with both α and β chains of the HLA class II heterodimer.

7.4.3 Tracking antigen-specific T cell responses post-transplant

The identification of alloreactive antigens and their restricting HLA allele enables antigen-specific T cell responses to be tracked at serial timepoints post-allo-SCT. Flow cytometry using fluorescent-labelled HLA-peptide tetramers could be used to specifically label antigen-specific T cells. Additional cell surface markers could be used to identify T cell subsets and functional states, which would enable assessment of the kinetics and phenotype of GvL and GvHD T cell responses. One could also investigate mechanisms of immune evasion at post-allo-SCT relapse by assessing whether relapse coincides with a reduction in frequency of antigen-specific T cells or if they exhibit a hypofunctional, exhausted state.

A further method that could be used to track the size of antigen-specific T cell clones is to perform bulk TCR repertoire sequencing *ex vivo*, as performed for Patient OX866 (Section 4.9), at serial timepoints after allo-SCT. Antigen-specific TCR sequences may be identified using IFNγ capture and single cell RNA sequencing (Section 4.7). By looking for antigen-specific TCR sequences within bulk *ex vivo* repertoires, the size of clones could be tracked over time post-allo-SCT.

7.4.4 Clinical impact of alloreactive antigens following allogeneic stem cell transplantation

Patient-donor mismatched exome variants that are at high allele frequency in the population are the most attractive candidates for clinical translation, as more patients could be treated with immunotherapy that targets the resultant antigens. If used to select allo-SCT donors that are predicted to elicit a GvL response, common antigens could also have greater clinical utility. However, at very high allele frequencies, it would become more challenging to produce a mismatch between patients and donors because only a small proportion of the donor population would lack the variant.

Applying the described workflow to patients with clinical evidence of GvL but without GvHD may enable important GvL antigens to be selectively identified. Patients who have had a clinical response to DLI would be ideal candidates to study and may enrich for antigens of greater therapeutic potential.

The clinical outcomes of patients could serve as a means to infer the importance of particular alloreactive antigens. Clinical evidence of a GvL effect in the presence of a given antigen and its restricting HLA allele would highlight it as a clinically useful antigen. However, unlike a GvHD effect, which can be directly assessed by disease affecting normal tissues, a GvL response can not be directly measured, but is inferred from reduced relapse rates. In individual patients, a GvL effect may be more likely in the following scenarios: (1) survival from poor risk disease, (2) achieving a cure despite residual detectable AML prior to allo-SCT, (3) response to DLI treatment, and (4) in the presence of GvHD, as there are likely to be shared mechanisms with GvL [1].

By studying the clinical outcomes of a large cohort of transplant recipients, the importance of individual antigens could be assessed by genotyping patients and their respective donors for variants that result in minor histocompatibility antigens. Correlation of a patient-donor mismatched variant in conjunction with its restricting HLA allele with a GvL effect in a multivariate analysis would highlight it as a clinically important antigen. Measurable clinical outcomes include overall survival, relapse and GvHD. Co-variates relating to demographics, disease status and treatment would also need to be accounted for in the multivariate analysis.

7.5 Targeting GvL antigens using adoptive T cell therapy

7.5.1 Targeting the PADI4 antigen

GvL antigens represent potential therapeutic targets. In the patient used to identify a particular antigen, the presence of GvL and GvHD could indicate the likely incidence of these phenomena if it were targeted in other patients. As a proof of concept to demonstrate that the identified alloreactive antigens could serve as therapeutic targets, a T cell therapy was developed targeting the PADI4 antigen.

Both retroviral and lentiviral vectors have been used to insert TCRs into T cells in clinical trials and were compared in this study [298]. Lentiviral transfer was consistently associated with higher transduction efficiency and also has the advantage of lower rates of insertional mutagenesis, which improves safety for clinical use.

To demonstrate *in vitro* targeting the PADI4 antigen presented by AML cells, assays were performed to assess activation of T cells in response to AML cells and also to assess the ability of T cells to kill AML cells. When AML cells with the appropriate HLA-DP type were pulsed exogenously with PADI4 peptide, T cell activation and killing of AML cells was consistently observed (Sections 6.4 and 6.5). This finding indicates that the peptide binds to HLA-DP on the surface of the AML cells and the HLA-peptide complex is then able to bind antigen-specific TCRs.

However, to test its utility as a therapeutic target, we wanted to understand how the PADI4 antigen could be presented to T cells in patients. Conventionally, peptides enter the HLA class II presentation pathway by endocytosis of extracellular protein. However, endogenous protein can also enter the class II presentation pathway by autophagy [121]. This process delivers cytoplasmic constituents to lysosomes, where proteins are degraded and resultant peptides can be loaded onto HLA class II. Therefore, the capacity of antigen-specific PADI4 Clone 2.1 to be activated or kill AML cells presenting either endogenous or exogenous antigen was tested *in vitro*. Endogenous presentation was assessed by co-culturing T cells with untreated AML cells. To test exogenous presentation, prior to co-culture with T cells, the AML cells were incubated with either culture supernatant or cell lysate from AML cells with expression of PADI4 protein with the antigenic variant. Monocyte-derived dendritic cells were used as a positive control for HLA class II presentation of extracellularly-derived antigen.

AML cells are heterogeneous and by testing different AML samples, distinct mechanisms of antigen presentation were identified. For some samples, no detectable T cell responses to endogenously presented antigen were found (Figure 6.18), whereas for one sample, responses were observed (Figure 6.22). Dendritic cells showed evidence of presenting extracellular antigen. Given that there was activation of PADI4 Clone 2.1 with untreated AML cells but no further increase after pre-treatment of AML cells with extracellular protein, it is unclear whether AML cells can similarly present extracellularly-derived antigen. This pathway could be clinically relevant because administration of adoptive T cell therapy or allo-SCT is

preceded by conditioning chemotherapy, which kills AML cells. The cellular debris from these cells could then be endocytosed and enter the HLA class II presentation pathway, leading to recognition by antigen-specific T cells.

Increasing the expression of antigen by AML cells could enhance the efficacy of immunotherapy. HLA class II presentation of antigens that are derived from endogenous proteins may be enhanced pharmacologically by inducing autophagy. Many drugs that are already licensed for other clinical indications have been found to potentiate autophagy, including sirolimus, metformin, lithium carbonate, carbamazepine and sodium valproate [122].

Although in Patient OX866, PADI4 antigen-specific T cells were $CD4^+$ with a cytotoxic phenotype, this subset of T cells is found at a low frequency in blood. In developing the T cell therapy, rather than restricting TCR transfer to $CD4^+$ cells, TCRs were inserted into all T cells from healthy donors. $CD8^+$ cytotoxic T cells are much more frequent than $CD4^+$ cytotoxic T cells and may therefore result in an engineered T cell pool that is more capable of killing target cells. Activation of both $CD4^+$ and $CD8^+$ T cells was observed following peptide pulsing, indicating that the CD4 co-receptor is not required for interactions between TCR and HLA-peptide complex. However, the lack of $CD8^+$ T cell responses in the absence of pulsing suggests that CD4 may be important for lower magnitude responses (Figure 6.22).

7.5.2 Clinical implications of TCR therapy

GvL antigens are potential immunotherapy targets in AML and as they are germline in origin rather than leukaemia-specific, they could also be targeted in other cancers that express the antigen. Adoptive T cell therapy has several advantages over other forms of cancer immunotherapy, such as peptide vaccination or immune checkpoint inhibition, which depend on the *in vivo* generation of sufficient numbers of anti-cancer T cells to mediate tumour regression [298]. During *in vitro* preparation for adoptive T cell therapy, cells are released from inhibitory signals that are present in the patient and large numbers of tumour-reactive T cells can be expanded. It is

also possible for T cells to undergo *in vitro* selection for TCRs with high avidity for their target antigen, as well as for optimal effector functions.

In addition to TCR recognition of HLA-peptide complex, T cell activation requires co-stimulatory signalling, for example from CD28 on the T cell surface interacting with CD80 or CD86 on APCs. Co-inhibition counteracts co-stimulation and the balance of these signals determines the activity of the effector T cell. After chronic antigen stimulation, adoptively transferred T cells can become hypofunctional and "exhausted" due to increased co-inhibitory signalling [322]. Therapeutic inhibition of co-inhibitory checkpoint receptors (e.g. PD-1, CTLA-4) has been used synergistically with CAR-T cell therapy in clinical studies to increase the activity of T cell therapy [323].

If there is redundancy in the function of the target protein, therapy could lead to downregulation of protein expression as a means of immune escape, thereby reducing efficacy [322]. Strategies to overcome this problem include targeting proteins that are essential for the survival of the cancer cells, or combinatorial targeting of multiple antigens.

The affinity of interactions between a TCR and HLA-peptide complex is often relatively low and therefore to strengthen therapeutic responses, affinity may be enhanced by altering amino acids in the antigen-binding sites of the TCR. However, caution must be taken due to the risk unexpected off-target effects, which can be difficult to predict [298]. The first two patients treated with a particular TCR directed against the tumour-associated antigen, MAGE-A3, experienced fatal cardiogenic shock as a result of off-target reactivity against a HLA-presented peptide derived from titin in cardiac muscle [324]. To mitigate against toxicity, viral constructs that deliver TCRs can include a "suicide gene" that is targetable in the event of severe toxicity, allowing the engineered T cells to be selectively removed. A construct that incorporates epitopes from the cell surface marker CD20 has been produced. In the event of toxicity, such constructs can be targeted using the monoclonal antibody, rituximab, which is already in widespread clinical use for the treatment of B cell malignancies and autoimmune disorders [325].

As with most adoptive T cell therapy, identified GvL antigens could be clinically targeted using engineered T cells that have been individually prepared using autologous T cells. Challenges associated with this approach include difficulty obtaining sufficient starting material, especially following chemotherapy or allo-SCT, and long delays caused by individualised manufacture. The availability of off-the-shelf allogeneic T cell products is an alternative method that overcomes these problems. However, allogeneic T cell products may be associated with dual rejection: infused cells may result in GvHD against the host, or the host immune system may reject the infused cells [326].

Appendices

Chapter A

Appendices

Patient	Age at Diagnosis	Gender	AML sub-type	ELN 2017 genetic risk	Cytogenetics	Molecular Genetics	
						NPM1/FLT3	Mutations in recurrently mutated genes (Whole exome sequencing)
OX00289	52	F	Secondary (MDS)	Intermediate	46, XX	WT	MPL, PARP4, SH2B3, SMC1A
OX00452	50	F	De novo	Intermediate	MLL rearrangement	WT	CUX1, NOTCH1, SF3B1
OX00607	50	F	De novo	Favourable	Normal	NPM1 mut / FLT3 ITD (allele ratio ITD/WT <0.5)	DNMT3A, FLT3 ITD, TET2
OX00628	48	M	De novo	Favourable	Normal	NPM1 mut	DHX15, IDH2, MED12
OX00690	29	F	De novo	Intermediate	47, XX, +21 (66/100 cells)	WT	ASXL1, IDH2, KMT2C, RUNX1
OX00747	48	F	De novo	Intermediate	46 XX, add(3)(q21), add(5)(q3?4)	NPM1 WT / FLT3 ITD (allele ratio ITD/WT <0.5)	CUX1, FLT3 ITD, KMT2D
OX00802	48	F	De novo	Intermediate	Normal	WT	DNMT3A
OX00813	16	M	De novo	Favourable	Inv(16)	WT	KRAS
OX00866	51	M	De novo	Intermediate	46, XY, t(5;11;9) 3-way translocation	NPM1 WT / FLT3 ITD (allele ratio ITD/WT <0.5)	ASXL2, FLT3 ITD, KMT2C, WT1
OX00885	42	F	De novo	Intermediate	Normal	NPM1 WT / FLT3 ITD (allele ratio ITD/WT <0.5)	CBL, STAT5B
OX00905	52	F	De novo	Intermediate	t(2;21)	NPM1 WT / FLT3 ITD (allele ratio ITD/WT <0.5)	ASXL1, DNMT3A, FLT3 ITD, IDH2, KMT2C, KMT2D, NOTCH1, SMC1A
OX00993	59	M	De novo	Intermediate	Normal	WT	ASXL1, CEBPA, CUX1, KMT2C, RIT1
OX01072	38	F	De novo	Intermediate	47, XX, +8	NPM1 WT / FLT3 ITD (allele ratio ITD/WT <0.5)	Not performed
OX01098	60	F	De novo	Intermediate	Karyotype failed	WT	DNMT3A, FLT3 ITD, KIT
OX01149	41	M	De novo	Intermediate	47, XY, +4	NPM1 mut / FLT3 WT	IDH1, KIT

Table A.1: Baseline characteristics of AML patients. AML sub-type indicates whether the disease was secondary to either MDS or chemotherapy/radiotherapy, or whether there was no identifiable cause (*de novo*). The ELN 2017 genetic risk was determined according to cytogenetic and molecular genetic features of disease at diagnosis [36]. NPM1 and FLT3 results are from PCR fragment analysis by electrophoresis, performed in an NHS laboratory. The FLT3 ITD / WT allele ratio is used to assign ELN 2017 risk categories. Whole exome sequencing was performed following exome capture with RNA baits (Agilent SureSelect Human All Exon v6). Leukaemia-specific variants were determined by comparing leukaemia variants to those identified in a matched patient germline control sample (pre-transplant T cell DNA).
M: Male, F: Female, MDS: Myelodysplastic syndrome, WT: Wild Type, FLT3 ITD: FLT3 internal tandem duplication, NHS: National Health Service

Patient	Pre-transplant			Transplant				Post-transplant			
	First-line chemotherapy	Refractory disease / relapse	Salvage chemotherapy	Disease status at transplant	Donor type	Transplant type	Conditioning	Relapse	Chemotherapy	DLI	Clinical status at final follow-up (time post-transplant)
OX289	DA (x1) – No further chemotherapy due to poor blood count recovery	No	-	CR1	Matched sibling	RIC	Flu/Mel	No	-	-	Alive (8 years)
OX452	DA (x2); HiDAC (x1)	No	-	CR1	Matched sibling	RIC	Flu/Mel	6 months	-	-	Died from relapse (10 months)
OX607	DA (x1); DA + Lestaurtinib/Placebo (x1); HiDAC + Lestaurtinib/Placebo (x1)	No	-	CR1	Matched sibling	RIC	Flu/Mel. Post-transplant Cy	No	-	-	Alive (9 years)
OX628	DA (x1); DA + Everolimus/Placebo (x1); HiDAC + Everolimus/Placebo (x1)	Molecular relapse (NPM1 MRD)	FLAG-Ida (x1)	Morphological CR2	12/12 Matched unrelated	RIC	Flu/Mel/Campth	No	-	Yes (x2)	Alive (8 years)
OX690	DA (x1); DA + Everolimus/Placebo (x1); HiDAC + Everolimus/Placebo (x1)	No	-	CR1	Matched sibling	MA	Bu/Cy. Post-transplant Cy	No	-	-	Alive (8 years)
OX747	DA (x2)	No	-	CR1	Sibling haploidentical*	RIC	Cy/Flu/TBI	No	-	-	Alive (8 years)
OX802	DA (x2); HiDAC (x1)	No	-	CR1	Matched sibling	RIC	Flu/Mel. Post-transplant MTX	No	-	-	Alive (7 years)
OX813	DA (x1); DA + Gemtuzumab (x1); HiDAC (x2)	Molecular relapse (CBFB-MYH11 MRD)	FLAG-Ida (x1)	Morphological remission. MRD positive	11/12 sibling (HLA-A mismatch - likely crossover)	MA	Thiotepa/Bu/Flu/ATG	No	-	-	Alive (7 years)
OX866	DA (x1)	Refractory	FLAG-Ida (x2) – residual disease (7% blasts) after cycle 1	CR	Matched sibling	RIC	Flu/Mel	3 months	FLAG-Ida (x1); FLA (x1)	Yes (x3)	Alive (7 years)
OX885	DA (x2); HiDAC (x1)	Morphological relapse	FLAG-Ida (x1)	CR2	11/12 matched unrelated (DP mismatch)	RIC	Flu/Mel/Campath	No	-	Yes (x2)	Alive (6 years)
OX905	DA (x1) – No further chemotherapy due to poor blood count recovery	Refractory	-	16% blasts	12/12 matched unrelated	RIC	Flu/Amsacrine/ Ara-C/Bu/ATG	6 months	FLAG-Ida (x1)	Yes (x1)	Died from relapse (10 months)
OX993	DA (x2); HiDAC (x1)	Morphological relapse	FLAG-Ida (x1)	CR2	11/12 matched unrelated (DP mismatch)	RIC	Flu/Amsacrine/ Ara-C/Bu/ATG	No	-	-	Alive (4 years)
OX1072	DA (x1)	Refractory	FLAG-Ida (x1), FLA (x1)	CR	12/12 matched unrelated	MA	Thiotepa/Bu/Flu/ATG	8 months	HiDAC + Amsacrine (x1); Azacitidine (x3)	Yes (x3)	Died from relapse (16 months)
OX1098	DA (x1)	Refractory	FLAG-Ida (x1)	CR	12/12 matched unrelated	RIC	Flu/Amsacrine/ Ara-C/Bu/ATG	No	-	-	Alive (4 years)
OX1149	DA + Gemtuzumab (x1); DA (x1); HiDAC (x2)	Molecular relapse (NPM1 MRD)	FLAG-Ida (x1)	Morphological remission. MRD positive	Matched sibling	RIC	Bu/Cy. Post-transplant MTX	No	-	-	Alive (3.5 years)

Table A.2: Treatment summary for patient cohort. Treatment administered to each patient prior to transplant, during allo-SCT and afterwards. Pre-transplant salvage treatment was only given where the patient was refractory to first-line treatment or relapsed. The degree of donor match refers to HLA alleles. Post-transplant chemotherapy treatment was given to patients who relapsed. DLIs were administered for relapsed disease or for reduced donor T cell chimerism. *Patient OX747 was initially given an unrelated donor transplant, but this failed to engraft. Further conditioning chemotherapy was administered, followed by a sibling haploidentical graft.
DA: Daunorubicin / Cytarabine, HiDAC: High-dose Cytarabine; FLAG-ida: Fludarabine / Cytarabine / Granulocyte colony-stimulating factor (G-CSF) / Idarubicin, MRD: Measurable Residual Disease, CR1/CR2: First / Second Complete Remission, MA: Myeloablative, RIC: Reduced Intensity Conditioning, Flu: Fludarabine; Mel: Melphalan, Cy: Cyclophosphamide, TBI: Total Body Irradiation, MTX: Methotrexate, Bu: Busulfan, ATG: Anti-thymocyte Globulin, DLI: Donor Lymphocyte Infusion.

Patient	Graft-versus-Host Disease (GvHD)	Post-transplant immunosuppressive therapy	Donor CD3 chimerism	DLI + indication
OX289	Chronic GvHD of mouth, gastrointestinal tract and liver	Systemic immunosuppression stopped after 17 months. Betamethasone mouthwash continued for up to 29 months.	Month 3 - 84.9%; Month 4 - 77.3%; Month 5 - 98.1%	Nil
OX452	Nil	Not known	Month 3 - 96%; Month 4 - 98%; Month 5 - 98%	Nil
OX607	Nil	Ciclosporin stopped after 5 months	Month 3 - 98%; Month 6 - 98%; Month 8 - 96.5%	Nil
OX628	Mild skin GvHD	Ciclosporin stopped after 4 months. Topical steroids continued for a further 7 months.	Month 11 - 46.1% - DLI treatment; Month 12 - 67.9% - DLI treatment; Month 13 - 62.7%; Month 14 - 81.1% - DLI treatment; Month 16 - 99.3%; Month 20 - 100%	Yes (x3). For reduced donor CD3 chimerism
OX690	GvHD, mainly affecting liver	Oral prednisolone stopped 6 months post-transplant. Ciclosporin weaned and stopped after 15 months.	Month 5 - 86.9%; Month 12 - 92.5%; Month 14 - 97.4%; Month 16 - 99.8%; Month 19 - 100%	Nil
OX747	Acute skin GvHD	Prednisolone stopped 8 months post-transplant. Mycophenolate mofetil stopped after 4 months and ciclosporin after 11 months.	100% throughout	Nil
OX802	Skin, mouth, gut and liver chronic GvHD. Autoimmune neutropenia and ocular GvHD over 18 months post-transplant.	Ciclosporin, budesonide (5 day course), topical steroids and steroid mouthwash. Course of prednisolone for liver GvHD 6 months post-transplant. Ciclosporin stopped after 6 months, but later reintroduced for autoimmune neutropenia.	100% throughout	Nil
OX813	Mild skin GvHD	Topical steroids during first few months post-transplant	Month 1 - >99%; Month 9 - >99%. Then, 100% thereafter	Nil
OX866	No GvHD, including after DLIs	Stopped ciclosporin 3 months post-transplant	Month 1 - 97.9%; Month 3 - 80% (relapse); Month 5 - 97.8% - DLI treatment; Month 9 - 100% - DLI treatment; Month 11 - 100% - DLI treatment	Yes (x3). Treatment of relapse.
OX885	Mild skin GvHD	Topical steroids. Ciclosporin stopped after 4 months.	Month 8 - 13.5% - DLI treatment; Month 13 - 44%; Month 15 - 62% - DLI treatment; Month 20 - 92%	Yes (x2). For reduced donor CD3 chimerism
OX905	Acute gut GvHD	Gut GvHD treated with prednisolone and budesonide. Still on weaning doses of prednisolone and ciclosporin when relapsed 6 months post-transplant.	Month 2 - 99.9%; Month 6 - 1.2% (relapse); Month 8 - 98.7% – DLI treatment	Yes (x1). Treatment of relapse.
OX993	Mild acute skin GvHD	Mycophenolate mofetil stopped 2 months post-transplant. Ciclosporin after 6 months.	Month 3 - 90%; Month 5 - 90%; Month 11 - 90%	Nil
OX1072	Mild gut GvHD	Budesonide for gut GvHD. Still taking ciclosporin when relapsed 8 months post-transplant.	Month 3 - 100%; Month 8 - 60%; (relapse); Month 10 - 100% - DLI treatment (x2); Month 13 - 40%	Yes (x3). Treatment of relapse.
OX1098	No GvHD	Ciclosporin stopped after 5 months	Month 14 - 84%; Month 23 - 83%; Month 44 - 87%	Nil
OX1149	Limited oral GvHD	Steroid mouthwash. Stopped ciclosporin after 6 months.	Month 2 - 99%); Month 3 - 97%; Month 6 - 100%; Month 25 - 100%	Nil

Table A.3: GvHD, immunosuppression and treatment with donor lymphocyte infusions. Organs affected by GvHD in individual patients are detailed. Where the clinical information was available, the GvHD severity and whether it was acute or chronic is indicated. Post-transplant immunosuppression was administered to prevent and to treat GvHD. Donor CD3 chimerism indicates the % of T cells in the sample that originate from the allo-SCT donor. DLIs were administered for either relapsed disease or for reduced donor T cell chimerism. GvHD: Graft-versus-Host Disease, DLI: Donor Lymphocyte Infusion.

Chapter B

Appendices

Target	Clone	Fluorochrome	Company	Dilution	Assay
CD3	HIT3A	PE Cy5	Biolegend	1:80	Sorting for WES and RNA-Seq
CD45	HI30	APC eFluor780	Invitrogen	1:40	Sorting for WES and RNA-Seq / Mouse engraftment
CD235ab	HIR2	Pacific blue	Biolegend	1:40	Sorting for WES and RNA-Seq
CD3	UCHT1	BV711	Biolegend	1:100	*In vitro* T cell activation
CD4	OKT4	APC CY7	Biolegend	1:50	*In vitro* T cell activation
CD8	RPA-T8	Alexo Fluor700	BD	1:200	*In vitro* T cell activation
TNFα	Mab11	FITC	BD	1:500	*In vitro* T cell activation
IL-2	MQ1-17H12	PerCP eFluor 710	eBioscience	1:50	*In vitro* T cell activation
IFNγ	B27	PE Cy7	BD	1:50	*In vitro* T cell activation
CD25	BC96	APC	Biolegend	1:80	*In vitro* T cell activation
CD134	ACT35	BB700	BD	1:50	*In vitro* T cell activation
Mouse TCRβ	H57-597	APC	BD	1:40	*In vitro* T cell activation
CD137	4B4-1	PE	Biolegend	1:80	*In vitro* T cell activation
CD107a	H4A3	BV650	Biolegend	1:100	*In vitro* T cell activation
Live/Dead Aqua	NA	NA	Thermo Fisher	1:1000	*In vitro* T cell activation
CD3	HIT3A	PE Cy7	Biolegend	1:50	Transduction assessment / IFNγ catch for single cell RNA-Seq / *In vitro* T cell killing
CD4	OKT4	APC	Biolegend	1:100	Transduction assessment / IFNγ catch for single cell RNA-Seq / *In vitro* T cell killing
CD8	SK1	FITC	Biolegend	1:50	Transduction assessment / IFNγ catch for single cell RNA-Seq / *In vitro* T cell killing
Mouse TCRβ	H57-597	BV711	BD	1:40	Transduction assessment / *In vitro* T cell killing
Mouse CD19	1D3	APC Cy7	Biolegend	1:80	Transduction assessment
Live/Dead near IR	NA	NA	Thermo Fisher	1:1000	*In vitro* T cell killing
Cell trace violet	NA	NA	Thermo Fisher	1:2500	*In vitro* T cell killing
Mouse CD45	30-F11	BV605	Biolegend	1:40	Mouse engraftment
CD33	WM-53	PE	Invitrogen	1:100	Mouse engraftment
CD34	581	APC	Biolegend	1:50	Mouse engraftment
CD19	HIB19	FITC	Invitrogen	1:40	Mouse engraftment
Hoechst 33258 (10ml/ml)	NA	NA	Invitrogen	1:10000	Sorting for WES and RNA-Seq / IFNγ catch for single cell RNA-Seq / Mouse engraftment

Table B.1: Antibodies and other stains used in flow cytometry assays.

Antigen number	Patient	Time of blood sample post-allo-SCT (years)[1]	Peptide(s)	Alloreactive peptide amino acid sequence (mismatch underlined)	Donor amino acid sequence	Gene	Chromosome	Position[2]	dbSNP ID[3]	T cell response against reference or variant sequence[4]	Allele frequency (%)[5]
1	OX866	6	OX866-021	LTISLLDTFNLELPEAVVFQ	LTISLLDTSNLELPEAVVFQ	PADI4	1	17342114	rs1748020	Variant	4.27
			OX866-022	ISLLDTFNLELPEAVVFQDS	ISLLDTSNLELPEAVVFQDS						
2	OX1149	1	OX1149-190	AASCEPLASVLRAKLTSRSS	AASCEPLASVLRAKPTSRSS	RNH1	11	499120	rs17585	Variant	11.21
3	OX1149	1	OX1149-290	GAGPDPLRLHGHLPVRTSCP	GAGPDPLRLRGHLPVRTSCP	SLC26A6	3	48632014	rs13324142	Variant	9.84
4	OX813	5.5	OX813-101	TQRSVLLCKVVGARGVGKSA	TQRSVLLCKVVGACGVGKSA	RHOT2	16	672331	rs3177338	Reference	51.51
			OX813-102	KVVGARGVGKSAFLQAFLGR	KVVGACGVGKSAFLQAFLGR						
5	OX802	5	OX802-321	GQKSPRFRRVSCFLRLGRST	GQKSPRFRRVTCFLRLGRST	COA6	1	234373513	rs10910420	Reference	52.52
			OX802-322	RFRRVSCFLRLGRSTLLELE	RFRRVTCFLRLGRSTLLELE						
6	OX289	6	OX289-009	ALAFLLLISIAANLSLLLSR	ALAFLLLISTAANLSLLLSR	NAGPA	16	5025632	rs7188856	Variant	30.07
7	OX289	6	OX289-040	EVQDCLKQLMMSLLQLYRFS	EVQDCLKQLMMSLLRLYRFS	RNF123	3	49700521	rs35620248	Variant	5.07
8	OX289	6	OX289-052	GYSPSLHILAIGTRSGAIKL	GYSPSLRILAIGTRSGAIKL	LLGL2	17	75556104	rs1671036	Variant	50.19
9	OX289	6	OX289-067	VATFPVYTMVAIPIVCKD	VATFPVYTMGAIPIVCKD	ENDOD1	11	95129413	rs3740861	Variant	25.36
10	OX289	6	OX289-073	PPLYRQRYQFIKNLVDQHEP	PPLYRQRYQFVKNLVDQHEP	HENMT1	1	108657486	rs36100901	Variant	1.93
			OX289-074	PLYRQRYQFIKNLVDQHEPK	PLYRQRYQFVKNLVDQHEPK						
11	OX289	6	OX289-075	VSRPELLRESISAFLVPMPT	VSRPELLREGISAFLVPMPT	ACADS	12	120738280	rs1799958	Variant	26.58
			OX289-205	TDRALQNKSISAFLVPMPTP	TDRALQNKGISAFLVPMPTP						
12	OX289	6	OX289-180	NPLSPYLNVDPRYLVQDT	NPLCPYLNVDPRYLVQDT	TIMM23B	10	49945053	rs148307270	Variant	9.11
13	OX289	6	OX289-221	EPPVDICLSKAISSSLKGFL	EPPVDICLSKANSSSLKGFL	LRR1	14	49607404	rs17121605	Variant	20.48
14	OX289	6	OX289-284	GETGMFSLSTIRGHQYATY	GETGMFSLCTIRGHQYATY	FTSJ3	17	63823837	rs2727288	Reference	64.78
15	OX289	6	OX1098-614	MNYVSKRLPFAARLNTPMGP	MNYVSKSLPFAARLNTPMGP	LGALS8	1	236543562	rs2243525	Reference	72.36
16	OX747	6	OX747-157	LGSLGLIFALTLNRHKYPLN	LGSLGLIFALILNRHKYPLN	TMBIM4	12	66152320	rs8793	Variant	43.73
			OX747-158	LGLIFALTLNRHKYPLNLYL	LGLIFALILNRHKYPLNLYL						
17	OX747	6	OX747-185	APISLSSFFNVSTLEREVTD	APISLSSFFSVSTLEREVTD	DOCK8	9	334337	rs10970979	Reference	24.54
18	OX747	6	OX747-190	LELGAGTGLASIIAATMART	LELGAGTGLTSIIAATMART	METTL22	16	8635267	rs2302607	Reference	22.2
			OX747-191	AGTGLASIIAATMARTVYCT	AGTGLTSIIAATMARTVYCT						
19	OX747	6	OX628-305	VPREYVRALNATKLERVFAK	VPREYIRALNATKLERVFAK	DCAF13	8	103420317	rs3134253	Variant	23.64
20	OX993	2	OX993-413	LHRDKALLKRLLKGMQKKRP	LHRDKALLKRLLKGVQKKRP	DENND6B	22	50313721	rs68178377	Variant	31.06
			OX993-414	KALLKRLLKGMQKKRPSDVQ	KALLKRLLKGVQKKRPSDVQ						
21	OX628	6	OX628-80	ITVQTVYVQHLITFLDRPIQ	ITVQTVYVQHPITFLDRPIQ	LITAF	16	11553659	rs761746674	Variant	0.00122
			OX628-81	QTVYVQHLITFLDRPIQMCC	QTVYVQHPITFLDRPIQMCC						
22	OX628	6	OX628-564	PGLISMFSSSQELGAALAQL	PGLISVFSSSQELGAALAQL	PGLS	19	17511703	rs368232818	Variant	0.08
23	OX885	4	OX885-559	WRVRIALALKGIDYETVPIN	WRVRIALALKGIDYKTVPIN	GSTZ1	14	77326864	rs7975	Reference	30.57
			OX885-560	VRIALALKGIDYETVPINLI	VRIALALKGIDYKTVPINLI						
24	OX885	4	OX628-308	DRAEKFNRGIRKLGITPEGQ	DRAEKFNRGIRKLGVTPEGQ	WASHC4	12	105152394	rs1663564	Variant	94.67
			OX628-309	EKFNRGIRKLGITPEGQSYL	EKFNRGIRKLGVTPEGQSYL						

Table B.2: Summary of functionally validated alloreactive antigens. [1]Length of time after allo-SCT that the blood sample was taken for peptide screening and antigen characterisation. [2]Genomic position taken from the hg38 human genome reference build (NCBI). [3]Unique ID for nucleotide variant taken from NCBI Database of Single Nucleotide Polymorphisms, dbSNP (https://www.ncbi.nlm.nih.gov/snp/) [4]Indicates whether the antigen results from the reference or variant allele, as explained in Section 3.2.1. [5]The population allele frequency is an average of all samples in the gnomAD database [233].

(a) HLA class I

Patient	A	A	B	B	C	C
OX813	*68:01:01	*11:01:01	*55:01:01	*15:01:01	*03:04:01	*03:03:01
OX993	*24:02	*02:01:01	*40:01	*18:01:01	*12:03	*03:04
OX1098	*02:01:01	-	*57:01:01	*40:01	*06:02	*03:04:01
OX607	*24:02/09N/11N/14/15/20-144	*26:01/08/11N/12/14-50	*27:05/13/28/37/38/45-67	*55:01-03/05/10/11/15-43	*01:02/06/08/11/15-18/20-40	*03:03/13/18/20N/22Q-88
OX289	*02:05	*23:01/05/06/08/10/11-19	*44:03/13/26/30/32/36-61	*58:01/04/11/13/15	*04:01/05/08/10/12/16-33	*07:01/06/16/18/20/21/24-57
OX690	*24:02/09N/11N/15/17-144	*11:01/05/07/09/12/13-64	*13:02/03/14-16/19/30-38	*51:01/03/11N/12/14/17-96	*03:03/13/18/20N/22Q-88	*06:02/03/07/09/10/12-45
OX885	*02:05:01	*24:02	*18:01:01	*49:01	*07:01	-
OX628	*24:02	*03:01:01	*44:02:01	*08:01	*07:01	*05:01
OX452	*02:01:01	*03:01:01	*35:01/07/15/29/40-42/52-94	*44:02/19/22-24/27/33-88	*04:01/28/30/41	*05:01/03
OX905	*02:01:01	-	*40:01	*15:01:01	*04:01/82	*03:04:01
OX866	*26:01/08/11N/12/14-77	*02:01/06/09/21/25/26-374	*27:05/13/28/37/38/45-87	*08:01/11/13-15/18/19-94	*07:01/06/16/18/20/21-263	*01:02/06/08/11/15-73
OX1072	*02:01:01	-	*18:01:01	*44:02	*05:01:01	*07:01
OX802	*32:01-03/05/06/09-46	*11:01/05/07/09-13/21N-64	*35:01/05/17/23/30/32-202	*08:01/07/11/14/15/18-93	*04:01/05/08/10/12/16-127	*07:02/10/13/15/23/32N-262
OX747	*01:01:01	-	*51:01:01	*40:01	*03:04:01	*06:02:01
OX1149	*34:02:01	*24:02	*55:01:01	*58:02	*06:02	*01:02

(b) HLA class II

Patient	DRB1	DRB1	DRB3	DRB3	DRB4	DRB5	DQA1	DQA1	DQB1	DQB1	DPB1	DPB1
OX813	*14:54	*15:01	*02:02/28/29N			*01:01:01	*01:02/06	*01:01/04/05/07	*06:02:01	*05:03:01	*05:01	*04:01:01
OX993	*01:01	*11:01	*02:02/28/29N				*01:01/02/04-09/11/12	*05:05/08/09/11	*03:01:01	*05:01:01	*04:01:01	-
OX1098	*07:01:01	*15:01			*01:03:01:02N	*01:01:01	*01:02/06/08/09/11	*02:01	*06:02:01	*03:03:02	*04:01:01	-
OX607	*01:01:01	*13:02:01	*03:01-03				*01:02/06	-	*06:04/27/36/38/39	*05:01	*04:02	*02:01
OX289	*03:01	*07:01:01	23		*01:01-05/07				*02:02	*02:01/02/04	*04:01:01	*02:01:02
OX690	*07:01:01	-			*01:01-04/07		*02:01	-	*02:02	*02:01/02/05	*17:01	*13:01
OX885	*11:04:01	*04:05:01	*02:02/28/29N		*01:03		*03:01-03	*05:05/08/09/11	*03:01:01	*03:02:01	*04:01:01	*04:02:01
OX628	*03:01	*15:01	*01:01/04			*01:01:01	*01:02/06	*05:01	*02:01:01	*06:02:01	*04:01:01	-
OX452	*01:03	*04:01:01			*01:01-05/07		*01:01/04/05/07	*03:01-03	*03:02:01	*05:01:01	*10:01	*04:01:01
OX905	*12:01/06/10/17	*13:02:01	*03:01//02	*02:02/28/29N//05			*01:02/06/08/09	*05:05/08/09/11	*03:01:01	*06:04:01	*03:01:01	-
OX866	*03:01	*08:01	*01:01/03-06/09-13/15				*04:01-04	*05:01	*02:01	*04:02	*04:01/*120:01N/*121:01/*128:01/*134:01	-
OX1072	*15:01	*11:01	*02:02/28/29N			*01:01:01	*01:02/06/08/09/11	*05:05/08/09/11	*03:01:01	*06:02:01	*04:01:01	-
OX802	*03:01	*11:01	29N	-			*05:01	*05:05/08-11	*02:01	*03:01	*04:01/*121:01	-
OX747	*04:04:01	*13:02:01	*03:01		*01:03		*01:02/06/08/09	*03:01-03	*03:02:01	*06:04:01	*10:01	*03:01:01
OX1149	*13:01	*11:02:01	*02:02/28/29N	*01:01/04			*01:03/10	*05:05/08/09/11	*03:01:01	*06:03:01	*02:01:02	-

Table B.3: HLA types of the patients studied. **(a)** HLA class I and **(b)** class II types of the patients studied.

Clone 1

TCRα

GCGGCCGCGCCACCATGCTTACAGCTTCTCTGCTGAGAGCCGTGATCGCCAGCATCTGTGTGGTGTCTAGCATGGCCCAG
AAAGTGACACAGGCCCAGACCGAGATCAGCGTGGTGGAAAAAGAAGATGTGACCCTGGACTGCGTGTACGAGACACGG
GACACCACCTACTACCTGTTCTGGTACAAGCAGCCTCCTAGCGGCGAGCTGGTGTTCCTGATCAGACGGAACAGCTTCGA
CGAGCAGAACGAGATCTCCGGCCGGTACAGCTGGAACTTCCAGAAGTCCACCAGCAGCTTCAACTTCACCATCACCGCCA
GCCAGGTGGTGGATAGCGCCGTGTATTTTTGCGCCCTGAGCGAGAGGCCTTACGGCGGAGCTACAAACAAGCTGATCTTC
GGCACCGGCACACTGCTGGCTGTTCAACCTAACATCCAGAACCCCGACCCCGCGG

TCRβ

CCATGGATACCTGGCTCGTGTGCTGGGCCATCTTCAGCCTGCTGAAAGCCGGACTGACCGAGCCTGAAGTGACCCAGACA
CCTAGCCACCAAGTGACACAGATGGGCCAAGAAGTGATCCTGCGCTGCGTGCCCATCAGCAACCACCTGTACTTCTACTG
GTACAGACAGATCCTGGGCCAGAAAGTGGAATTCCTGGTGTCCTTCTACAACAACGAGATCAGCGAGAAGTCCGAGATCT
TCGACGACCAGTTCAGCGTGGAAAGACCCGACGGCAGCAACTTCACCCTGAAGATCAGAAGCACCAAGCTCGAGGACAG
CGCCATGTACTTTTGCGCCAGCAGCGAGAGAAGAACCCAGCCTGCCTACGAGCAGTACTTCGGCCCTGGCACAAGACTGA
CCGTGACAGAGGACCTGCGGAACGTGACCCCCCCCAAGGTGTCCCTGTTCGAGCCCAGCAAGGCCGAGATCGCCAACAA
GCAGAAAGCCACACTGGTCTGTCTGGCTAGGGGCTTCTTCCCCGACCACGTG

Clone 2

TCRα (1)

GCGGCCGCGCCACCATGAAGTCTCTGAGAGTGCTGCTGGTCATCCTGTGGCTGCAGCTGTCTTGGGTCTGGTCCCAGCAG
AAAGAGGTGGAACAGAACAGCGGCCCTCTGTCTGTTCCTGAAGGCGCTATCGCCAGCCTGAACTGCACCTACAGCGATA
GAGGCAGCCAGAGCTTCTTCTGGTACAGACAGTACAGCGGCAAGAGCCCCGAGCTGATCATGAGCATCTACAGCAACGG
CGACAAAGAGGACGGCCGGTTTACAGCCCAGCTGAACAAGGCCAGCCAGTACGTGTCCCTGCTGATCAGAGATAGCCAG
CCTAGCGACAGCGCCACCTATCTGTGTGCCGTTAGAGGCCAGGCTGGCACAGCCCTGATCTTTGGCAAGGGCACAACACT
GAGCGTGTCCAGCAACATCCAGAACCCCGACCCCGCGG

TCRα (2)

GCGGCCGCGCCACCATGCTTACAGCTTCTCTGCTGAGAGCCGTGATCGCCAGCATCTGTGTGGTGTCTAGCATGGCCCAG
AAAGTGACACAGGCCCAGACCGAGATCAGCGTGGTGGAAAAAGAAGATGTGACCCTGGACTGCGTGTACGAGACACGG
GACACCACCTACTACCTGTTCTGGTACAAGCAGCCTCCTAGCGGCGAGCTGGTGTTCCTGATCAGACGGAACAGCTTCGA
CGAGCAGAACGAGATCTCCGGCCGGTACAGCTGGAACTTCCAGAAGTCCACCAGCAGCTTCAACTTCACCATCACCGCCA
GCCAGGTGGTGGATAGCGCCGTGTACTTTTGTGCCCTGCCTGCCGGAAATACCCCTCTGGTGTTTGGCAAGGGCACCAGA
CTGTCTGTGATCGCCAACATCCAGAACCCCGACCCCGCGG

TCRβ

CCATGGGCAGCAGACTGCTGTGTTGGGTGCTGCTGTGTCTGCTTGGAGCCGGACCTGTGAAAGCTGGCGTGACCCAGAC
ACCTAGATACCTGATCAAGACCAGAGGCCAGCAAGTGACCCTGAGCTGCTCTCCTATCAGCGGCCACAGAAGCGTGTCCT
GGTATCAGCAGACACCTGGACAGGGCCTGCAGTTCCTGTTCGAGTACTTCAGCGAGACACAGCGGAACAAGGGCAACTT
CCCCGGCAGATTTTCCGGCAGACAGTTCAGCAACAGCCGCAGCGAGATGAACGTGTCCACACTGGAACTGGGCGACAGC
GCCCTGTATCTGTGTGCCTCTTCTCTGACCGGCGGCAACTACGGCTACACATTTGGCAGCGGCACCAGACTGACAGTGGT
CGAGGACCTGCGGAACGTGACCCCCCCCAAGGTGTCCCTGTTCGAGCCCAGCAAGGCCGAGATCGCCAACAAGCAGAAA
GCCACACTGGTCTGTCTGGCTAGGGGCTTCTTCCCCGACCACGTG

Clone 3

TCRα

GCGGCCGCGCCACCATGTTGTTGCTGCTGATTCCTGTGCTGGGCATGATCTTCGCCCTGAGGGATGCTAGAGCCCAGTCC
GTGTCTCAGCACAACCACCACGTGATCCTGTCTGAGGCCGCCTCTCTGGAACTGGGCTGCAATTACAGCTACGGCGGCAC
CGTGAACCTGTTTTGGTACGTGCAGTACCCCGGCCAGCATCTCCAGCTGCTGCTGAAGTACTTTAGCGGCGACCCTCTGGT
CAAGGGCATCAAGGGATTCGAGGCCGAGTTCATCAAGAGCAAGTTCAGCTTCAACCTGCGGAAGCCCAGCGTGCAGTGG
AGCGATACAGCCGAGTACTTTTGTGCCGTGGGCACCAATGCCGGCAAGAGCACATTTGGCGACGGCACCACACTGACCG
TGAAGCCTAACATCCAGAACCCCGACCCCGCGG

TCRβ

CCATGGGCCCTCAGCTGCTGGGATATGTGGTGCTGTGTCTGCTTGGAGCCGGACCTCTGGAAGCCCAAGTGACACAGAAC
CCCAGATACCTGATCACCGTGACCGGCAAGAAACTGACCGTGACCTGCAGCCAGAACATGAACCACGAGTACATGAGCT
GGTACAGACAGGACCCTGGCCTGGGCCTGAGACAGATCTACTACAGCATGAACGTGGAAGTGACCGACAAGGGCGACG
TGCCCGAGGGCTACAAGGTGTCCAGAAAAGAGAAGCGGAACTTCCCACTGATCCTGGAAAGCCCATCTCCTAACCAGACC
AGCCTGTACTTCTGCGCCAGCAGCAGAAAAGGACTGGCTGGCGGAGGACCTACCGGCGAGCTGTTTTTTGGCGAGGGCA
GCAGACTGACAGTGCTCGAGGACCTGCGGAACGTGACCCCCCCCAAGGTGTCCCTGTTCGAGCCCAGCAAGGCCGAGAT
CGCCAACAAGCAGAAAGCCACACTGGTCTGTCTGGCTAGGGGCTTCTTCCCCGACCACGTG

Clone 4

TCRα(1) and TCRβ from Clone 2

Clone 5

TCRα

GCGGCCGCGCCACCATGGATAAGATTCTGGGCGCCAGCTTCCTGGTGCTGTGGCTGCAACTTTGTTG
GGTGTCCGGCCAGCAGAAAGAGAAGTCCGACCAGCAGCAAGTGAAACAGAGCCCTCAGAGCCTGA
TCGTGCAGAAAGGCGGCATCAGCATCATCAACTGCGCCTACGAGAATACCGCCTTCGACTACTTCCC
CTGGTATCAGCAGTTCCCCGGCAAGGGACCTGCTCTGCTGATCGCCATTAGACCCGACGTGTCCGAG
AAGAAAGAGGGCAGATTCACCATCAGCTTCAACAAGAGCGCCAAGCAGTTCAGCCTGCACATCATG
GATAGCCAGCCTGGCGACAGCGCCACCTATTTTTGTGCTGCTGTTACAGGCGGCGGAGCCGATGGC
CTGACATTTGGAAAGGGCACCCACCTGATCATCCAGCCTTACATCCAGAACCCCGACCCCGCGG

TCRβ

CCATGGACACCAGAGTGCTGTGCTGCGCCGTGATCTGTCTGCTTGGAGCCGGACTGTCTAATGCCG
GCGTGATGCAGAACCCCAGACACCTCGTTCGGAGAAGAGGCCAAGAGGCCAGACTGAGATGCAGC
CCTATGAAGGGCCACAGCCATGTGTACTGGTACAGACAGCTGCCCGAAGAGGGCCTGAAGTTCATG
GTGTACCTGCAGAAAGAGAACATCATCGACGAGAGCGGCATGCCCAAAGAGCGGTTCTCTGCCGAG
TTTCCCAAAGAGGGCCCCAGCATCCTGAGAATCCAGCAGGTTGTGCGGGGAGATAGCGCCGCCTAC
TTTTGTGCTAGCAGCCCTCAGGGCTACAACGAGCAGTTTTTCGGCCCTGGCACCAGACTGACAGTGC
TCGAGGACCTGCGGAACGTGACCCCCCCCAAGGTGTCCCTGTTCGAGCCCAGCAAGGCCGAGATCG
CCAACAAGCAGAAAGCCACACTGGTCTGTCTGGCTAGGGGCTTCTTCCCCGACCACGTG

Figure B.1: Nucleotide sequences of T cell receptors for PADI4 antigens
Human codon optimised nucleotide sequences of candidate TCRs recognising the PADI4 antigen, which were cloned into the pMP71 retroviral vector. In black, sequences of the 5' and 3' restriction sites are shown (NotI / SacII for TCRα; NcoI / PmlI for TCRβ). At the 3' end, part of the murine constant region is also included. Gene segments are in different colours: V (red), D (purple); J (green).
TCR: T cell receptor

References

[1] Sweeney, C. and Vyas, P. (2019). The graft-versus-leukemia effect in AML. *Frontiers in Oncology* 9: 1217.

[2] De Kouchkovsky, I. and Abdul-Hay, M. (2016). Acute myeloid leukemia: A comprehensive review and 2016 update. *Blood Cancer Journal* 6(7): e441.

[3] Dohner, H., Weisdorf, D.J. and Bloomfield, C.D. (2015). Acute myeloid leukemia. *New England Journal of Medicine* 373(12): 1136–52.

[4] Papaemmanuil, E., Gerstung, M., Bullinger, L., et al. (2016). Genomic classification and pr ognosis in acute myeloid leukemia. *New England Journal of Medicine* 374(23): 2209–21.

[5] Passweg, J.R., Baldomero, H., Bader, P., et al. (2016). Hematopoietic stem cell transplantation in Europe 2014: More than 40,000 transplants annually. *Bone Marrow Transplantation* 51(6): 786–92.

[6] Bower, H., Andersson, T., Bjorkholm, M., et al. (2016). Continued improvement in survival of acute myeloid leukemia patients: An application of the loss in expectation of life. *Blood Cancer Journal* 6: e390.

[7] Burnett, A., Wetzler, M., and Löwenberg, B. (2011). Therapeutic advances in acute myeloid leukemia. *Journal of Clinical Oncology* 29(5): 487–94.

[8] Goyal, G., Gundabolu, S., Vallabhajosyula, S., et al. Reduced-intensity conditioning allogeneic hematopoietic cell transplantation for older adults with acute myeloid leukemia. *Cancer Control* 18(4): 237–45.

[9] Vyas, P., Appelbaum, F.R., and Craddock, C. (2015). Allogeneic hematopoietic cell transplantation for acute myeloid leukemia. *Biology of Blood and Marrow Transplantation* 21(1): 8–15.

[10] Zilberberg, J., Feinman, R., and Korngold, R. (2015). Strategies for the identification of T cell-recognized tumor antigens in hematological malignancies for improved graft-versus-tumor responses after allogeneic blood and marrow transplantation. *Biology of Blood and Marrow Transplantation* 21(6): 1000–7.

[11] Ferrara, J.L.M., Levine, J., Reddy, P., et al. (2009). Graft-versus-host disease. *Lancet* 373: 1550–61.

[12] Rowe, J.M. (2008). Graft-versus-disease effect following allogeneic transplantation for acute leukaemia. *Best Practice and Research: Clinical Haematology* 21(3): 485–502.

[13] Barker, J.N., and Wagner, J.E. (2003). Umbilical-cord blood transplantation for the treatment of cancer. *Nature Reviews Cancer* 3(7): 526–32.

[14] Hobbs, G., and Perales, M-A. (2015). Effects of T-cell depletion on allogeneic hematopoietic stem cell transplantation outcomes in AML patients. *Journal of Clinical Medicine* 4(3): 488–503.

[15] Kolb, H.J. (2008). Graft-versus-leukemia effects of transplantation and donor lymphocytes. *Blood* 112(12): 4371–83.

[16] Griffioen, M., Van Bergen, C.A.M., and Falkenburg, J.H.F. (2016). Autosomal minor histocompatibility antigens: How genetic variants create diversity in immune targets. *Frontiers in Immunology* 7: 100.

[17] Center for International Blood and Marrow Transplant. (2019). U.S. transplant data by disease report, AML - Acute myelogenous leukemia, patient survival 2011 - 2015. *U.S. Transplant Data by Disease Report*.

[18] Barnes, D.W.H., Corp, M., Loutit, J., et al. (1956). Treatment of murine leukaemia with x rays and homologous bone marrow: preliminary communication. *British Medical Journal* 2(4993): 626–7.

[19] Thomas, E.D., Lochte, H., Lu, W., et al. (1957). Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. *New England Journal of Medicine* 257(11): 491–6.

[20] Henig, I., and Zuckerman, T. (2014). Hematopoietic stem cell transplantation—50 years of evolution and future perspectives. *Rambam Maimonides Medical Journal* 5(4): e0028.

[21] Dausset, P.J. (1958). Iso-leuco-anticorps. *Acta Haematologica* 166(20): 156–66.

[22] Thorsby, E. (2009). A short history of HLA. *Tissue Antigens* 74(2): 101–16.

[23] Mathe, G., Amiel, J., Schwarzenberg, A., et al. (1965). Successful allogenic bone marrow transplantation in man: chimerism, induced specific tolerance and possible anti-leukemic effects. *Blood* 25(2): 179–96.

[24] Matsukura, M., Mery, A., Amiel, J., et al. Investigation on a test of histocompatibility for allogeneic grafts. *Transplantation* 1(1): 61–4.

[25] Thomas, E.D., Buckner, D., Banaji, M., et al. (1977). One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation. *Blood* 49(4): 511–33.

[26] Thomas, E.D., Buckner, C., Clift, R., et al. (1979). Marrow transplantation for nonlymphoblastic leukemia in first remission. *New England Journal of Medicine* 301(11): 597–9.

[27] Weiden, P.L., Sullivan, K., Flournoy, N., et al. (1981). Antileukemic effect of chronic graft-versus-host disease. *New England Journal of Medicine* 304(25): 1529–33.

[28] Hansen, J.A., Clift, R., Thomas, E.D., et al. (1980). Transplantation of marrow from an unrelated donor to a patient with acute leukemia. *New England Journal of Medicine* 303(10): 565–7.

[29] Singh, A.K., and McGuirk, J.P. (2016). Allogeneic stem cell transplantation: A historical and scientific overview. *Cancer Research* 76(22): 6445–51.

[30] World Marrow Donor Association (WMDA). Total number of donors and cord blood units. URL: https:statistics.wmda.info (accessed on 06/25/2019).

[31] Gragert, L., Eapen, M., Williams, E., et al. (2014). HLA match likelihoods for hematopoietic stem-cell grafts in the U.S. Registry. *New England Journal of Medicine* 371(4): 339–48.

[32] Yang, B., Yu, R., Cai, L., et al. (2019). Haploidentical versus matched donor stem cell transplantation for patients with hematological malignancies: a systemic review and meta-analysis. *Bone Marrow Transplantation* 54(1): 99–122.

[33] Duarte, R.F., Labopin, M., Bader, P., et al. (2019). Indications for haematopoietic stem cell transplantation for haematological diseases, solid tumours and immune disorders: current practice in Europe. *Bone Marrow Transplantation* 54: 1525–52.

[34] Sorror, M.L., Maris, M., Storb, R., et al. (2005). Hematopoietic cell transplantation (HCT)– specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood* 106(8): 2912–20.

[35] Cornelissen, J.J., Gratwohl, A., Schlenk, R., et al. (2012). The European LeukemiaNet AML Working Party consensus statement on allogenic HSCT for patients with AML in remission: an integrated-risk adapted approach. *Nature Reviews Clinical Oncology* 9(10): 579–590.

[36] Dohner, H., Estey, E., Grimwade, D., et al. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 129(4): 424–48.

[37] Short, N.J., and Ravandi, F. (2019). How close are we to incorporating measurable residual disease into clinical practice for acute myeloid leukemia? *Haematologica* 104(8): 1532–41.

[38] Ravandi, F., Walter, R.B., and Freeman, S.D. (2018). Evaluating measurable residual disease in acute myeloid leukemia. *Blood Advances* 2(11): 1356-66.

[39] Jongen-Lavrencic, M., Grob, T., Hanekamp, D., et al. (2018). Molecular minimal residual disease in acute myeloid leukemia. *New England Journal Medicine* 378(13): 1189–99.

[40] Breems, D.E., Van Putten, W., Huijens, P., et al. (2005). Prognostic Index for adult patients with acute myeloid leukemia in first relapse. *Journal of Clinical Oncology* 23(9): 1969–78.

[41] Seliger, B., Kloor, M., and Ferrone, S. (2017). HLA class II antigen-processing pathway in tumors: molecular defects and clinical relevance. *OncoImmunology* 6(2): 1–10.

[42] Christopher, M.J., Petti, A., Rettig, M., et al. (2018). Immune escape of relapsed AML cells after allogeneic transplantation. *New England Journal of Medicine* 379(24): 2330-41.

[43] Toffalori, C., Zito, L., Gambacorta, V., et al. (2019). Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. *Nature Medicine* 25(4): 603–11.

[44] Warren, E.H., and Deeg, H.J. (2013). Dissecting graft-versus-leukemia from graft-versus-host disease using novel strategies. *Tissue Antigens* 81(4): 183–93.

[45] Horowitz, M., Gale, R., Snodel, P., et al. (1990). Graft-versus-leukemia reactions after bone marrow transplantation. *Proceedings of the American Mathematical Society* 75(3): 555–62.

[46] Valcárcel, D., Martino, R., Caballero, D., et al. (2008). Sustained remissions of high-risk acute myeloid leukemia and myelodysplastic syndrome after reduced-intensity conditioning allogeneic hematopoietic transplantation: Chronic graft-versus-host disease is the strongest factor improving survival. *Journal of Clinical Oncology* 26(4): 577–84.

[47] Baron, F., Labopin, M., Niederwieser, D., et al. (2012). Impact of graft-versus-host disease after reduced-intensity conditioning allogeneic stem cell transplantation for acute myeloid leukemia: A report from the Acute Leukemia Working Party of the European group for blood and marrow transplantation. *Leukemia* 26(12): 2462–68.

[48] Kato, M., Kurata, M., Kanda, J., et al. (2019). Impact of graft-versus-host disease on relapse and survival after allogeneic stem cell transplantation for pediatric leukemia. *Bone Marrow Transplantation* 54(1): 68–75.

[49] Stern, M., de Wreede, L., Brand, R., et al. (2014). Sensitivity of hematological malignancies to graft-versus-host effects: an EBMT megafile analysis. *Leukemia* 28: 2235-40

[50] Woolfrey, A., Lee, S., Gooley, T., et al. (2010). HLA-allele matched unrelated donors compared to HLA-matched sibling donors: Role of cell source and disease risk category. *Biology of Blood and Marrow Transplantation* 16(10): 1382–87.

[51] Ringden, O., Pavletic, S., Anasetti, C., et al. (2009). The graft-versus-leukemia effect using matched unrelated donors is not superior to HLA-identical siblings for hematopoietic stem cell transplantation. *Blood* 113(13): 3110-18.

[52] Schlenk, R.F., Dohner, K., Mack, S., et al. (2010). Prospective evaluation of allogeneic hematopoietic stem-cell transplantation from matched related and matched unrelated donors in younger adults with high-risk acute myeloid leukemia: German-Austrian trial AMLHD98A. *Journal of Clinical Oncology* 28(30): 4642–48.

[53] Brissot, E., Labopin, M., Stelljes, M., et al. (2017). Comparison of matched sibling donors versus unrelated donors in allogeneic stem cell transplantation for primary refractory acute myeloid leukemia: A study on behalf of the Acute Leukemia Working Party of the EBMT. *Journal of Hematology and Oncology* 10(1): 1–10.

[54] Lee, S.J., Klein, J., Haagenson, M., et al. (2007). High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood* 110(13): 4576–84.

[55] Saber, W., Opie, S., Rizzo, J., et al. (2012). Outcomes after matched unrelated donor versus identical sibling hematopoietic cell transplantation in adults with acute myelogenous leukemia. *Blood* 119(17): 3908–16.

[56] Bensinger, W.I. (2012). Allogeneic transplantation: peripheral blood vs. bone marrow. *Current Opinion in Oncology* 24(2): 191–96.

[57] Bensinger, W.I., Martin, P., Storer, B., et al. (2001). Transplantation of bone marrow as compared with peripheral-blood cells from HLA-identical relatives in patients with hematologic cancers. *New England Journal of Medicine* 344(3): 175–81.

[58] Couban, S., Simpson, D., Barnett, M., et al. (2002). A randomized multicenter comparison of bone marrow and peripheral blood in recipients of matched sibling allogeneic transplants for myeloid malignancies. *Blood* 100(5): 1525–31.

[59] Blaise, D., Kuentz, C., Fortanier, C., et al. (2000). Randomized trial of bone marrow versus lenograstim-primed blood cell allogeneic transplantation in patients with early-stage leukemia: A report from the Société Française de Greffe de Moelle. *Journal of Clinical Oncology* 18(3): 537.

[60] Heldal, D., Brinch, Y., Floisand, Y., et al. (2000). Long-term (> 5 years) follow-up of a randomized study of allogeneic stem cell transplantation with stem cells from blood or bone marrow. *Bone Marrow Transplantation* 25: 1129–36.

[61] Gratwohl, A., Brand, R., Frassoni, F., et al. (2005). Cause of death after allogeneic haematopoietic stem cell transplantation (HSCT) in early leukaemias: An EBMT analysis of lethal infectious complications and changes over calendar time. *Bone Marrow Transplantation* 36(9): 757–69.

[62] van Dicke, K.A., Hooft, J.I., and van Bekkum, D.W. (1968). The selective elimination of immunologically competent cells from bone marrow and lymphatic cell mixtures. II. Mouse spleen cell fractionation on a discontinuous albumin gradient. *Transplantation* 6(4): 562–70.

[63] Ho, V.T., and Soiffer, R.J. (2001). The history and future of T-cell depletion as graft-versus-host disease prophylaxis for allogeneic hematopoietic stem cell transplantation. *Blood* 98(12): 3192–3204.

[64] Waldmann, H., Hale, G., Cividalli, G., et al. (1984). Elimination of graft-versus-host disease by in-vitro depletion of alloreactive lymphocytes with a monoclonal rat anti-human lymphocyte antibody (Campath-1). *The Lancet* 324(8401): 483–86.

[65] Heit, W., Bunjes, D., Wiesneth, M., et al. (1986). Ex vivo T-cell depletion with the monoclonal antibody Campath-1 plus human complement effectively prevents acute graft-versus-host disease in allogeneic bone marrow transplantation. *British Journal of Haematology* 64(3): 479–86.

[66] Apperley, J.F., Jones, L., Hale, G., et al. (1986). Bone marrow transplantation for patients with chronic myeloid leukaemia: T-cell depletion with Campath-1 reduces the incidence of graft-versus-host disease but may increase the risk of leukaemic relapse. *Bone Marrow Transplantation* 1(1): 53–66.

[67] Marmont, A.M., Horowitz, M., Gale, R., et al. (1991). T-cell depletion of HLA-identical transplants in leukemia. *Blood* 78(8): 2120–30.

[68] Kernan, N.A., Bordignon, C., Heller, G., et al. (1989). Graft failure after T-cell-depleted human leukocyte antigen identical marrow transplants for leukemia: I. Analysis of risk factors and results of secondary transplants. *Blood* 74(6): 2227–36.

[69] Saad, A., and Lamb, L.S. (2017). Ex vivo T-cell depletion in allogeneic hematopoietic stem cell transplant: Past, present and future. *Bone Marrow Transplantation* 52(9): 1241–48.

[70] Pasquini, M.C., Devine, S., Mendizabal, A., et al. (2012). Comparative outcomes of donor graft CD34+ selection and immune suppressive therapy as graft-versus-host disease prophylaxis for patients with acute myeloid leukemia in complete remission undergoing HLA-matched sibling allogeneic hematopoietic cell transplantation. *Journal of Clinical Oncology* 30(26): 3194–3201.

[71] Ho, V. T., Kim, H., Li, S., et al. (2004). Partial CD8+ T-cell depletion of allogeneic peripheral blood stem cell transplantation is insufficient to prevent graft-versus-host disease. *Bone Marrow Transplantation* 34(11): 987–94.

[72] Federmann, B., Bornhauser, M., Meisner, C., et al. (2012). CD3/CD19 depletion and reduced intensity conditioning: A phase II study. *Haematologica* 97(10): 1523–31.

[73] Drobyski, W.R., Majewski, D., and Hanson, G. (1999). Graft-facilitating doses of ex vivo activated gd T cells do not cause lethal murine graft-vs.-host disease. *Biology of Blood and Marrow Transplantation* 5(4): 222–30.

[74] Mehta, J., Singhal, S., Gee, A., et al. (2004). Bone marrow transplantation from partially HLA-mismatched family donors for acute leukemia: Single-center experience of 201 patients. *Bone Marrow Transplantation* 33(4): 389–96.

[75] Lamb, L.S., Gee, A., Hazlett, L., et al. (1999). Influence of T cell depletion method on circulating *gd* T cell reconstitution and potential role in the graft-versus-leukemia effect. *Cytotherapy* 1(1): 7–19.

[76] Meeh, P.F., King, M., O'Brien, R.L., et al. (2006). Characterization of the T cell response to acute leukemia. *Cancer Immunology, Immunotherapy* 55(9): 1072–80.

[77] Beck, B.H., Kim, H., Kim, H., et al. (2010). Adoptively transferred ex vivo expanded $\gamma\ \delta$-T cells mediate in vivo antitumor activity in preclinical mouse models of breast cancer. *Breast Cancer Research and Treatment* 122(1): 135–44.

[78] Chakraverty, R., Orti, G., Roughton, M., et al. (2010). Impact of in vivo alemtuzumab dose before reduced intensity conditioning and HLA-identical sibling stem cell transplantation: Pharmacokinetics, GVHD, and immune reconstitution. *Blood* 116(16): 3080–88.

[79] Bacigalupo, A., Lamparelli, P., Bruzzi, P., et al. (2001). Antithymocyte globulin for graft-versus-host disease prophylaxis in transplants from unrelated donors: 2 randomized studies from Gruppo Italiano Trapianti Midollo Osseo (GITMO). *Blood* 98(1): 2942–47.

[80] Luznik, L., O'Donnell, P., Symons, H., et al. (2008). HLA-haploidentical bone marrow transplantation for hematologic malignancies using nonmyeloablative conditioning and high-dose, posttransplantation cyclophosphamide. *Biology of Blood and Marrow Transplantation* 14(6): 641–50.

[81] Ali, R., Ramdial, S., Algaze, S., et al. (2017). The role of anti-thymocyte globulin or alemtuzumab-based serotherapy in the prophylaxis and management of graft-versus-host disease. *Biomedicines* 5(4): 67.

[82] Malladi, R.K., Peniket, A, Littlewood, T., et al. (2009). Alemtuzumab markedly reduces chronic GVHD without affecting overall survival in reduced-intensity conditioning sibling allo-SCT for adults with AML. *Bone Marrow Transplantation* 43(9): 709–15.

[83] Schmidt-Hieber, M., Schwarck, S., Stroux, A., et al. (2010). Immune reconstitution and cytomegalovirus infection after allogeneic stem cell transplantation: The important impact of in vivo T cell depletion. *International Journal of Hematology* 91(5): 877–85.

[84] Fuchs, E.J. (2012). Haploidentical transplantation for hematologic malignancies: where do we stand? *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* 230–236.

[85] Ciurea, S.O., Mulanovich, V., Saliba, R., et al. (2012). Improved early outcomes using a T cell replete graft compared with T cell depleted haploidentical hematopoietic stem cell transplantation. *Biology of Blood and Marrow Transplantation* 18(12): 1835–44.

[86] Namdaroglu, S., Kaya, A., Batgi, H., et al. (2019). Impacts of post-transplantation cyclophosphamide treatment after allogeneic hematopoietic stem cell transplantation in acute myeloid leukemia. *Scientific Reports* 9(1): 1–6.

[87] Sullivan, K.M., et al. (1989). Graft-versus-host disease as adoptive immunotherapy in patients with advanced hematologic neoplasms. *New England Journal of Medicine* 320(13): 828–34.

[88] Kolb, H.J., Storb, R., Buckner, C., et al. (1990). Donor leukocyte transfusions for treatment of recurrent chronic myelogenous leukemia in marrow transplant patients. *Blood* 76(12): 2462–5.

[89] Dazzi, F., Szydlo, R.M., Cross, N.C., et al. (2000). Durability of responses following donor lymphocyte infusions for patients who relapse after allogeneic stem cell transplantation for chronic myeloid leukemia. *Blood* 96(8): 2712–6.

[90] Dickinson, A.M., Norden, J., Li, S., et al. (2017). Graft-versus-leukemia effect following hematopoietic stem cell transplantation for leukemia. *Frontiers in Immunology* 8: 496.

[91] Orti, G., Barba, P., Fox, L., et al. (2017). Donor lymphocyte infusions in AML and MDS: Enhancing the graft-versus-leukemia effect. *Experimental Hematology* 48: 1–11.

[92] Schmid, C., Labopin, M., Nagler, A., et al. (2007). Donor lymphocyte infusion in the treatment of first hematological relapse after allogeneic stem-cell transplantation in adults with acute myeloid leukemia: A retrospective risk factors analysis and comparison with other strategies by the EBMT Acute Leukemia Working Party. *Journal of Clinical Oncology* 25(31): 4938–45.

[93] Schmid, C., Schleuning, M., Ledderose, G., et al. (2005). Sequential regimen of chemotherapy, reduced-intensity conditioning for allogeneic stem-cell transplantation, and prophylactic donor lymphocyte transfusion in high-risk acute myeloid leukemia and myelodysplastic syndrome. *Journal of Clinical Oncology* 23(24): 5675–87.

[94] Jedlickova, Z., Schmid, C., Koenecke, C., et al. (2016). Long-term results of adjuvant donor lymphocyte transfusion in AML after allogeneic stem cell transplantation. *Bone Marrow Transplantation* 51(5): 663–667.

[95] Prophylactic and pre-emptive DLI for myeloid malignancies (PRO-DLI). URL: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/prophylatic-and-pre-emptive-dli-for-myeloid-malignancies-pro-dli/ (accessed on 06/28/2019).

[96] Bacigalupo, A., Ballen, K., Rizzo, D., et al. (2009). Defining the intensity of conditioning regimens: working definitions. *Biology of Blood and Marrow Transplantation* 15(12): 1628–33.

[97] McClune, B.L., Weisdorf, D., Pedersen, T., et al. (2010). Effect of age on outcome of reduced-intensity hematopoietic cell transplantation for older patients with acute myeloid leukemia in first complete remission or with myelodysplastic syndrome. *Journal of Clinical Oncology* 28(11): 1878–87.

[98] Weisdorf, D.J. (2017). Reduced-intensity versus myeloablative allogeneic transplantation. *Hematology/Oncology and Stem Cell Therapy* 10(4): 321–26.

[99] Schumacher, T.N., and Schreiber, R.D. (2015). Realising the promise: neoantigens in cancer immunotherapy. *Science* 348: 69–74.

[100] Vago, L., Perna, S., Zanussi, M., et al (2009). Loss of mismatched HLA in leukemia after stem-cell transplantation. *New England Journal of Medicine* 361(5): 478–88.

[101] Toffalori, C . (2012). To the editor: Genomic loss of patient-specific HLA in acute myeloid leukemia relapse after well-matched. *Blood* 119(20): 4813–15.

[102] Jan, M., Leventhal, M., Morgan, E., et al. (2019). Recurrent genetic HLA loss in AML relapsed after matched unrelated allogeneic hematopoietic cell transplantation. *Blood Advances* 3(14): 2199–2204.

[103] Ahmadzadeh, M., Johnson, L., Heemskerk, B., et al. (2009). Tumor antigen specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired Tumor antigen –specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired. *Blood* 114(8): 1537–44.

[104] Puram, S.V., Tirosh, I., Parikh, A., et al (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171(7): 1611–24.

[105] Norde, W.J., Maas, F., Hobo, W., et al (2011). PD-1/PD-L1 interactions contribute to functional T-cell impairment in patients who relapse with cancer after allogeneic stem cell transplantation. *Cancer Research* 71(15): 5111–22.

[106] Mussai, F., Santo, C., Abu-Dayyeh, I., et al. (2019). Acute myeloid leukemia creates an arginase-dependent immunosuppressive microenvironment. *Blood* 122(5): 749–59.

[107] Folgiero, V., Goffredo, B., Filippini, P., et al. (2013). Indoleamine 2,3-dioxygenase 1 (IDO1) activity in leukemia blasts correlates with poor outcome in childhood acute myeloid leukemia. *Oncotarget* 5(8): 2052–64.

[108] Zeiser, R., and Vago, L. (2019). Mechanisms of immune escape after allogeneic hematopoietic cell transplantation. *Blood* 133(12): 1290–97.

[109] Park, H.H., Kim, M., Lee, B., et al. (2006). Intracellular IL-4, IL-10, and IFN-g levels of leukemic cells and bone marrow T cells in acute leukemia. *Annals of Clinical and Laboratory Science* 36(1): 7–15.

[110] Dilillo, D.J., Weinberg, J., Yoshizaki, A., et al. (2013). Chronic lymphocytic leukemia and regulatory B cells share IL-10-competence and immunosuppressive function. *Leukemia* 27(1): 170–82.

[111] Colpitts, S.L., Stonier, S., Stoklasek, T., et al. (2013). Transcriptional regulation of IL-15 expression during hematopoiesis. *Journal of Immunology* 191: 3017–24.

[112] Thiant, S., Magro, L., Trauet, J., et al. (2010). Plasma levels of IL-7 and IL-15 in the first month after myeloablative BMT are predictive biomarkers of both acute GVHD and relapse. *Bone Marrow Transplantation* 45(10): 1546–52.

[113] Mathew, N.R., Baumgartner, F., Braun, L., et al. (2018). Sorafenib promotes graft-versus-leukemia activity in mice and humans through IL-15 production in FLT3-ITD-mutant leukemia cells. *Nature Medicine* 24(3): 282–91.

[114] Romee, R., Cooley, S., Berrien-Elliott, M., et al. (2018). First-in-human phase 1 clinical study of the IL-15 superagonist complex ALT-803 to treat relapse after transplantation. *Blood* 131(23): 2515–28.

[115] Rutten, C.E., Luxemburg-Heijs, S., Halkes, C. et al. (2013). Patient HLA-DP-specific CD4 + T cells from HLA-DPB1-mismatched donor lymphocyte infusion can induce graft-versus-leukemia reactivity in the presence or absence of graft-versus-host disease. *Biology of Blood* and *Marrow Transplantation* 19: 40–48.

[116] Petersdorf, E.W., Malkki, M., O'hUigin, C., et al. (2019). High HLA-DP Expression and Graft-versus-Host Disease. *New England Journal of Medicine* 373(7): 599–609.

[117] Beyar-Katz, O., and Saar Gill, S. (2018). Novel approaches to acute myeloid leukemia immunotherapy. *Clinical Cancer Research* 24(22): 5502-15

[118] Kornum, B.R., Knudsen, S., Ollilla, H.M., et al. (2017). Narcolepsy. *Nature Reviews Disease Primers* 3: 16100.

[119] Kobayashi, K.S., and Van Den Elsen, P.J. NLRC5: A key regulator of MHC class I-dependent immune responses. *Nature Reviews Immunology* 12(12): 813–20.

[120] Purcell, A.W., Ramarathinam, S.H., and Ternette, N. (2019). Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nature Protocols* 14(6): 1687–1707.

[121] Münz, C. (2011). Antigen processing by macroautophagy for MHC presentation. *Frontiers in Immunology* 2(42): 1–7.

[122] Thellung, S., Corsaro, A., Nizzari, M., et al. (2019). Autophagy activator drugs: A new opportunity in neuroprotection from misfolded protein toxicity. *International Journal of Molecular Sciences* 20(4): 901.

[123] Jameson-Lee, M., Koparde, V., Griffith, P., et al. (2014). In silico derivation of HLA-specific alloreactivity potential from whole exome sequencing of stem-cell transplant donors and recipients: Understanding the quantitative immunobiology of allogeneic transplantation. *Frontiers in Immunology* 5(529): 1–11.

[124] Martin, P.J., Levine, D., Storer, B., et al. (2017). Genome-wide minor histocompatibility matching as related to the risk of graft-versus-host disease. *Blood* 129(6): 791–98.

[125] Granados, D.P., Sriranganadane, D., Daouda, T., et al. (2014). Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. *Nature Communications* 5: 3600.

[126] van Els, C.A.C.M., D'Amaro, J., Pool, K., et al. (1992). Immunogenetics of human minor histocompatibility antigens: their polymorphism and immunodominance. *Immunogenetics* 35(3): 161–65.

[127] van der Harst, D., Goulmy, E., Falkenburg, J., et al. (1994). Recognition of minor histocompatibility antigens on lymphocytic and myeloid leukemic cells by cytotoxic T-cell clones. *Blood* 83(4): 1060–66.

[128] den Haan, J.M.M., Sherman, N., Blokland, E., et al. (1995). Identification of a graft versus host disease-associated human minor histocompatibility antigen. *Science* 268(5216): 1476–80.

[129] Oostvogels, R., Lokhorst, H.M., and Mutis, T. (2016). Minor histocompatibility Ags: Identification strategies, clinical results and translational perspectives. *Bone Marrow Transplantation* 51(2): 163–71.

[130] Van Bergen, C.A.M., Rutten, C., Van Der Meijden, E., et al. (2010). High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. *Cancer Research* 70(22): 9073–83.

[131] Akatsuka, Y., Nishida, T., Kondo, E., et al. (2003). Identification of a polymorphic gene, BCL2A1 , encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. *Journal of Experimental Medicine* 197(11): 1489–1500.

[132] De Rijke, B., Horssen-Zoetbrood, A., Beekman, J., et al. (2005). A frameshift polymorphism in P2X5 elicits an allogeneic cytotoxic T lymphocyte response associated with remission of chronic myeloid leukemia. *Journal of Clinical Investigation* 115(12): 3506–16.

[133] Spierings, E., Kim, Y., Hendriks, M., et al. (2013). Multicenter analyses demonstrate significant clinical effects of minor histocompatibility antigens on GvHD and GvL after HLA-matched related and unrelated. *Biology of Blood and Marrow Transplantation* 19(8): 1244–53.

[134] Marijt, W.A.E., Heemskerk, M., Kloosterboer, F., et al. (2003). Hematopoiesis-restricted minor histocompatibility antigens HA-1- or HA-2-specific T cells can induce complete remissions of relapsed leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 100(5): 2742–47.

[135] Stern, M., Passweg, J., Locasciulli, A., et al. (2006). Influence of donor/recipient sex matching on outcome of allogeneic hematopoietic stem cell transplantation for aplastic anemia. *Transplantation* 82(2): 218–26.

[136] Randolph, S.S.B., Gooley, T., Warren, E., et al. (2004). Female donors contribute to a selective graft-versus-leukemia e˙ect in male recipients of HLA-matched , related hematopoietic stem cell transplants. *Blood* 103(1): 347–53.

[137] Vogt, M.H.J., Paus, R., Voogt, P., et al. (2000). DFFRY codes for a new human male-specific minor transplantation antigen involved in bone marrow graft rejection. *Blood* 95(3): 1100–06.

[138] Ofran, Y., Kim, H., Brusic, V., et al. (2010). Diverse patterns of T-cell response against multiple newly identified human Y chromosome–encoded minor histocompatibility epitopes. *Clinical Cancer Research* 16(5): 1642–52.

[139] Nakasone, H., Sahaf, B., Schoenrock, K., et al. (2015). Allogeneic HY antibodies detected 3 months after female-to-male HCT predict chronic GVHD and nonrelapse mortality in humans. *Blood* 125(20): 3193–3201.

[140] Rizvi, N.A., Hellmann, M.D., Snyder, A., et al. (2016). Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* 348(6230): 124-128.

[141] The Cancer Genome Atlas Research Network. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *New England Journal of Medicine* 368(22): 2059–74.

[142] Anagnostou, V., Smith, K., Forde, P., et al. (2017). Evolution of neoantigen landscape during immune checkpoint blockade in non–small cell lung cancer. *Cancer Discovery* 7(3): 264-76.

[143] Greiner, J., Ono, Y., Hofmann, S., et al. (2012). Mutated regions of nucleophosmin 1 elicit both CD4+ and CD8+ T-cell responses in patients with acute myeloid leukemia. *Blood* 97(7): 1282–90.

[144] Graf, C., Heidel, F., Tenzer, S., et al. (2007). A neoepitope generated by an FLT3 internal tandem duplication (FLT3-ITD) is recognized by leukemia-reactive autologous CD8+ T cells. *Blood* 109(7): 2985–89.

[145] Anguille, S., Van Tendeloo, V.F., and Berneman, Z.N. (2012). Leukemia-associated antigens and their relevance to the immunotherapy of acute myeloid leukemia. *Leukemia* 26(10): 2186–96.

[146] Goswami, M., Hensel, N., Smith, B., et al. (2014). Expression of putative targets of immunotherapy in acute myeloid leukemia and healthy tissues. *Leukemia* 28(5): 1167–70.

[147] Zhao, X.-Y., Chang, Y.-J., and Huang, X.-J. (2011). Conflicting impact of alloreactive NK cells on transplantation outcomes after haploidentical transplantation: Do the reconstitution kinetics of natural killer cells create these differences? *Biology of Blood and Marrow Transplantation* 17(10): 1436–42.

[148] Baron, F., Petersdorf, E., Gooley, T., et al. (2009). What is the role for donor natural killer cells after nonmyeloablative conditioning? *Biology of Blood and Marrow Transplantation* 15: 580–88.

[149] Velardi, A. (2008). Role of KIRs and KIR ligands in hematopoietic transplantation. *Current Opinion in Immunology* 20: 581–87.

[150] Karre, K., Ljunggren, H., Piontek, G., et al. (1986). Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature* 319(6055): 675–78.

[151] Angell, T.E., Lechner, M., Jang, J., et al. (2014). MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in vitro. *Clinical Cancer Research* 20(23): 6034–44.

[152] del Campo, A., Kyte, J., Carretero, J., et al. (2014). Immune escape of cancer cells with beta2-microglobulin loss over the course of metastatic melanoma. *International Journal of Cancer* 134(1): 102–13.

[153] Seliger, B., Jasinski, S., Quandt, D., et al. (2016). HLA-E expression and its clinical relevance in human renal cell carcinoma. *Oncotarget* 7(41): 67360–72.

[154] Ruggeri, L., Capanni, M., Urbani, E., et al. (2002). Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science* 9(1): 2097–2100.

[155] Venstrom, J., Pittari, G., Gooley, T., et al. (2012). HLA-C-dependent prevention of leukemia relapse by donor activating KIR2DS1. *New England Journal of Medicine* 367(9): 805–16.

[156] McGranahan, N., Furness, A., Rosenthal, R., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351(6280): 1463–69.

[157] Schischlik, F., Jager, R., Rosebrock, F., et al. (2019). Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. *Blood* 134(2):199-210

[158] van Rooij, N., Van Buuren, M., Philips, D., et al. (2013). Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *Journal of Clinical Oncology* 31(32): 439–42.

[159] Rajasagi, M., Shukla, S., Fritsch, E., et al. (2014). Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 124(3): 453–63.

[160] Jurtz, V., Paul, S., Andreatta, M., et al. (2017). NetMHCpan-4.0: Improved peptide MHC class I interaction predictions integrating eluted ligand and peptide binding aÿnity data. *Journal of Immunology* 199(9): 3360–68.

[161] Jensen, K.K., Andreatta, M., Buus, S., et al. (2018). Improved methods for predicting peptide binding affnity to MHC class II molecules. *Immunology* 154(3): 394–406.

[162] Bentzen, A.K., Marquard, A., Lyngaa, R., et al. (2016). Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nature Biotechnology* 34(10): 1037–45.

[163] Bjerregaard, A.-M., Nielsen, M., Jurtz, V., et al. (2017). An analysis of natural T cell responses to predicted tumor neoepitopes. *Frontiers in Immunology* 8: 1566.

[164] Bjerregaard, A.-M., Nielsen, M., Hadrup, S., et al. (2017). MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunology, Immunotherapy* 66(9): 1123–30.

[165] Van Buuren, M.M., Calis, J.J.A., and Schumacher, T.N.M. (2014). High sensitivity of cancer exome-based CD8 T cell neo-antigen identification. *OncoImmunology* 3(5): e28836.

[166] Chen, B., Khodadoust, M., Olsson, N., et al. (2019). Predicting HLA class II antigen presentation through integrated deep learning. *Nature Biotechnology* 37(11): 1332–43.

[167] Abelin, J.G., Harjanto, D., Malloy, M., et al. (2019). Defining HLA-II ligand processing and binding rules with mass spectrometry enhances cancer epitope prediction. *Immunity* 51(4): 1-14.

[168] Yadav, M., Jhunjhunwala, S., Phung, Q., et al. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515(7528): 572–76.

[169] Gubin, M.M., Zhang, X., Schuster, H., et al. (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 515(7528): 577–81.

[170] Rosenberg, S.A., Jang, J., Sherry, R., et al. (2011). Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clinical Cancer Research* 17(13): 4550–57.

[171] Andersen, A., Donia, M., Ellebaek, E., et al. (2016). Long-lasting complete responses in patients with metastatic melanoma after adoptive cell therapy with tumor-infiltrating lymphocytes and an attenuated IL2 regimen. *Clinical Cancer Research* 22(15): 3734–45.

[172] Van Den Berg, J.H., Heemskerk, B., Van Rooij, N., et al. (2020). Tumor infiltrating lymphocytes (TIL) therapy in metastatic melanoma: Boosting of neoantigen-specific T cell reactivity and long-term follow-up. *Journal for ImmunoTherapy of Cancer* 8(2): 1–11.

[173] Stevanovic, S., Helman, S., Wunderlich, J., et al. (2019). A Phase II study of tumor-infiltrating lymphocyte therapy for human papillomavirus–associated epithelial cancers. *Clinical Cancer Research* 25(5): 1486–93.

[174] Zhou, X., Wu, J., Duan, C., et al. (2020). Retrospective analysis of adoptive TIL therapy plus anti-PD1 therapy in patients with chemotherapy-resistant metastatic osteosarcoma. *Journal of Immunology Research* 2020: 7890985.

[175] Biernacki, M.A., Foster, K., Woodward, K., et al. (2020). CBFB-MYH11 fusion neoantigen enables T cell recognition and killing of acute myeloid leukemia. *Journal of Clinical Investigation* 130(10): 5127–41.

[176] Van Der Lee, D.I., Reijmers, R., Honders, M., et al. (2019). Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia. *Journal of Clinical Investigation* 129(2): 774–85.

[177] Alspach, E., Lussier, D., Miceli, A., et al. (2019). MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* 574(7780): 696–701.

[178] Chapuis, A.G., Egan, D., Bar, M., et al. (2019). T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant. *Nature Medicine* 25: 1064–72.

[179] Keskin, D.B., Anandappa, A., Sun, J., et al. (2019). Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* 565(7738): 234–39.

[180] Oliveira, G., Stromhaug, K., Klaeger, S., et al. (2021). Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature* 596(7870): 119–25.

[181] Walter, R.B., Gyurkocza, B., Storer, B., et al. (2015). Comparison of minimal residual disease as outcome predictor for AML patients in first complete remission undergoing myeloablative or nonmyeloablative allogeneic hematopoietic cell transplantation. *Leukemia* 29(1): 137–44.

[182] Anthias, C., Dignan, F., Morilla, R., et al. (2014). Pre-transplant MRD predicts outcome following reduced-intensity and myeloablative allogeneic hemopoietic SCT in AML. *Bone Marrow Transplantation* 49(5): 679–83.

[183] Canaani, J., Labopin, M., Huang, X., et al. (2018). Minimal residual disease status predicts outcome of acute myeloid leukaemia patients undergoing T-cell replete haploidentical transplantation. An analysis from the Acute Leukaemia Working Party (ALWP) of the European Society for Blood and Marrow Transplantation. *British Journal of Haematology* 183: 411–20.

[184] Schmid, C., Labopin, M., Nagler, A., et al. (2012). Treatment, risk factors, and outcome of adults with relapsed AML after reduced intensity conditioning for allogeneic stem cell transplantation. *Blood* 119(6): 1599–1606.

[185] Craddock, C., Labopin, M., Robin, M., et al. (2016). Clinical activity of azacitidine in patients who relapse after allogeneic stem cell transplantation for acute myeloid leukemia. *Haematologica* 101(7): 879–83.

[186] de Lima, M., Oran, B., Champlin, E., et al. (2018). CC-486 maintenance after stem cell transplantation in patients with acute myeloid leukemia or myelodysplastic syndromes. *Biology of Blood and Marrow Transplantation* 24: 2017–24.

[187] AMADEUS: A double-blind, Phase III, randomised study to compare the efficacy and safety of oral Azacitidine (CC-486) versus placebo in subjects with acute myeloid leukaemia (AML) or myelodysplastic syndromes (MDS) as maintenance after allogeneic haematopoietic stem cell transplantation. URL: https//www.impactpartnership.org.uk/amadeus/ (accessed July 25, 2019).

[188] Goodyear, O., Agathanggelou, A., Novitzky-Basso, I., et al. (2009). Induction of a CD8 + T-cell response to the MAGE cancer testis antigen by combined treatment with azacitidine and sodium valproate in patients with acute myeloid leukemia and myelodysplasia. *Blood* 116(11): 1908–19.

[189] Loo Yau, H., Bell, E., Ettayebi, I., et al. (2021). Article DNA hypomethylating agents increase activation and cytolytic activity of CD8 + T cells Article DNA hypomethylating agents increase activation and cytolytic activity of CD8 + T cells. *Molecular Cell* 81(7): 1–15.

[190] De Freitas, T., Marktel, S., Piemontese, S., et al. (2015). High rate of hematological responses to sorafenib in FLT3-ITD acute myeloid leukemia relapsed after allogeneic hematopoietic stem cell transplantation. *European Journal of Hematology* 96(17): 629–36.

[191] Metzelder, S.K., Schroeder, T., Finck, A., et al. (2012). High activity of sorafenib in FLT3-ITD-positive acute myeloid leukemia synergizes with allo-immune effects to induce sustained responses. *Leukemia* 26(11): 2353–59.

[192] Rautenberg, C., Germing, U., Haas, R., et al. (2019). Relapse of acute myeloid leukemia after allogeneic stem cell transplantation: Prevention, detection, and treatment. *International Journal of Molecular Sciences* 20(1): 1–20.

[193] Cortes, J.E., Khaled, S., Martinelli, G., et al. Quizartinib versus salvage chemotherapy in relapsed or refractory FLT3-ITD acute myeloid leukaemia (QuANTUM-R): a multicentre, randomised, controlled, open-label, Phase 3 trial. *The Lancet Oncology* 20(7): 984–97.

[194] A trial of the FMS-like tyrosine kinase 3 (FLT3) inhibitor gilteritinib administered as maintenance therapy following allogeneic transplant for patients with FLT3/internal tandem duplication (ITD) acute myeloid leukemia (AML). URL: https://clinicaltrials.gov/ct2/show/NCT02997202 (accessed on 10/01/2019).

[195] DiNardo, C.D., Stein, E., de Botton, S., et al. (2018). Durable remissions with ivosidenib IDH-mutated relapsed or refractory AML. *New England Journal of Medicine* 378(25): 2386–98.

[196] Stein, E.M., DiNardo, C., Pollyea, D., et al. (2017). Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood* 130(6): 722–31.

[197] Ivosidenib as maintenance therapy in treating patients with IDH1-mutant acute myeloid leukemia, myelodysplastic syndrome, or chronic myelomonocytic leukemia following stem cell transplant. URL: https://www.cancer.gov/about-cancer/treatment/clinical-trials/search/v?id=NCI-2018-03528%7B%5C&%7Dr=1 (accessed on 10/01/2019).

[198] IDH2 inhibition using enasidenib as maintenance therapy for IDH2-mutant myeloid neoplasms following allogeneic stem cell transplantation. URL: https://clinicaltrials.gov/ct2/show/NCT03515512 (accessed on 10/01/2019).

[199] Motzer, R.J., Rini, B., McDermott, D., et al. (2015). Nivolumab for metastatic renal cell carcinoma: Results of a randomized Phase II trial. *Journal of Clinical Oncology* 33(13): 1430–37.

[200] Royal, R.E., Levy, C., Turner, K., et al. (2010). Phase 2 trial of single agent ipilimumab (Anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *Journal of Immunotherapy* 33(8): 828–33.

[201] Schadendorf, D., Hodi, F., Robert, C., et al. (2015). Pooled analysis of long-term survival data from Phase II and Phase III trials of ipilimumab in unresectable or metastatic melanoma. *Journal of Clinical Oncology* 33(17): 1889-94.

[202] Ijaz, A., Khan, A., Malik, S., et al. (2019). Significant risk of graft-versus-host disease with exposure to checkpoint inhibitors before and after allogeneic transplantation. *Biology of Blood and Marrow Transplantation* 25(1): 94–99.

[203] Merryman, R.W., Kim, H., Zinzani, P., et al. Safety and efficacy of allogeneic hematopoietic stem cell transplant after PD-1 blockade in relapsed/refractory lymphoma. *Blood* 129(10): 1380–88.

[204] Cilloni, D., and Saglio, G. (2004). WT1 as a universal marker for minimal residual disease detection and quantification in myeloid leukemias and in myelodysplastic syndrome. *Acta Haematologica* 112(1-2): 79–84.

[205] Brayer, J., Lancet, J., Powers, J., et al. (2015). WT1 vaccination in AML and MDS: a pilot trial with synthetic analog peptides. *American Journal of Haematology* 90(7): 602–7.

[206] Maslak, P.G., Dao, T., Bernal, Y., et al. (2018). Phase 2 trial of a multivalent WT1 peptide vaccine (galinpepimut-S) in acute myeloid leukemia. *Blood advances* 2(3): 224–34.

[207] Guy, D.G., and Uy, G.L. (2018). Bispecific antibodies for the treatment of acute myeloid leukemia. *Current Hematologic Malignancy Reports* 13(6): 417–25.

[208] Tasian, S.K. (2018). Acute myeloid leukemia chimeric antigen receptor T-cell immunotherapy: how far up the road have we traveled? *Therapeutic Advances in Hematology* 9(6): 135–48.

[209] Benjamin, D., Sato, T., Cibulskis, K., et al. (2019). Calling somatic SNVs and indels with mutect2. *bioRxiv* 861054.

[210] Kim, S., Scheffler, K., Halpern, A., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* 15(8): 591–94.

[211] Lai, Z., Markovets, A., Ahdesmaki, M., et al. (2016). VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Research* 44(11): 1–11.

[212] Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 38(16): 1–7.

[213] Rimmer, A., Phan, H., Mathieson, I., et al. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics* 46(8): 912–18.

[214] Partridge, T., Nicastri, A., Kliszczak, A., et al. (2018). Discrimination between human leukocyte antigen class I-bound and co-purified HIV-derived peptides in immunopeptidomics workflows. *Frontiers in Immunology* 9: 912.

[215] Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L., et al. (2015). Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Molecular and Cellular Proteomics* 14(3): 658–73.

[216] Crooks, G.E., Hon, G., Chandonia, J., et al. (2004). WebLogo: A sequence logo generator. *Genome Research* 14(6): 1188–90.

[217] Aran, D., Looney, A., Liu, L., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* 20(2): 163–72.

[218] Monaco, G., Lee, B., Xu, W., et al. (2019). RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Reports* 26(6): 1627–40.

[219] Ekiz, H.A., Conley, C., Stephens, W., et al. (2020). CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. *BMC Bioinformatics* 21(1): 191.

[220] Schmiedel, B.J., Singh, D., Madrigal, A., et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 175(6): 1701–15.

[221] Korotkevich, G., Sukhov, V., Budin, N., et al. (2021). Fast gene set enrichment analysis. *bioRxiv* 060012.

[222] Sharma, A., Bhatt, N., St Martin, A., et al. (2021). Clinical characteristics and outcomes of COVID-19 in haematopoietic stem-cell transplantation recipients: an observational cohort study. *The Lancet Haematology* 8(3): e185–93.

[223] Quek, L., Ferguson, P., Metzner, M., et al. (2016). Mutational analysis of disease relapse in patients allografted for acute myeloid leukemia. *Blood Advances* 1(3): 193–204.

[224] Schlenner, S.M., Madan, V., Busch, K., et al. (2010). Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus. *Immunity* 32(3): 426–36.

[225] Shlush, L.I., Mitchell, A., Heisler, L., et al. (2017). Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* 547(7661): 104–8.

[226] Young, A.L., Challen, G., Birmann, B., et al. (2016). Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature Communications* 7: 1–7.

[227] do Valle, I.F., Giampieri, E., Simonetti, G., et al. (2016). Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *BMC Bioinformatics* 17(Suppl. 12): 341.

[228] Lansford, J.L., Dharmasiri, U., Chai, S., et al. (2018). Computational modeling and confirmation of leukemia-associated minor histocompatibility antigens. *Blood Advances* 2(16): 2052–62.

[229] Jaiswal, S., Fontanillas, P., Flannick, J., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine* 371(26): 2488–98.

[230] Genovese, G., Kahler, A., Handsaker, R., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *New England Journal of Medicine* 371(26): 2477–87.

[231] Xie, M., Lu, C., Wang, J., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine* 20(12): 1472–78.

[232] Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *New England Journal of Medicine* 377(25): 2500–01.

[233] Karczewski, K.J., Francioli, L, Tiao, G., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581(7809): 434–43.

[234] Uhlén, M., Fagerberg, L., Hallstrom, B., et al. (2015). Tissue-based map of the human proteome. *Science* 347(6220): 1260419.

[235] Zhang, J., Xin, L., Shan, B., et al. (2012). PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular and Cellular Proteomics* 11(4): M111.010587.

[236] Wieczorek, M., Abualrous, E., Sticht, J., et al. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation. *Frontiers in Immunology* 8: 292.

[237] Guillaume, P., Picaud, S., Baumgaertner, P., et al. (2018). The C-terminal extension landscape of naturally presented HLA-I ligands. *Proceedings of the National Academy of Sciences of the United States of America* 115(20): 5083–88.

[238] Parker, K. C., Bednarek, M., Hull, L., et al. (1992). Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *Journal of Immunology* 149(11): 3580–87.

[239] Chatterjee, D., Priyadarshini, P., Das, D., et al. (2020). Deciphering the structural enigma of HLA class-II binding peptides for enhanced immunoinformatics-based prediction of vaccine epitopes. *Journal of Proteome Research* 19(11): 4655–69.

[240] Nielsen, M.L., Savitski, M., Kjeldsen, F., et al. (2004). Physicochemical properties determining the detection probability of tryptic peptides in fourier transform mass spectrometry. A correlation study. *Analytical Chemistry* 76(19): 5872–77.

[241] Calarota, S.A., Chiesa, A., Zelini, P., et al. (2013). Detection of Epstein-Barr virus-specific memory CD4+ T cells using a peptide-based cultured enzyme-linked immunospot assay. *Immunology* 139(4): 533–544.

[242] Holthaus, L., Lamp, D., Gavrisan, A., et al. (2018). CD4+ T cell activation, function, and metabolism are inhibited by low concentrations of DMSO. *Journal of Immunological Methods* 463: 54–60.

[243] Sereti, I., Gea-Banacloche, J., Kan, M.Y., et al. (2000). Interleukin 2 leads to dose-dependent expression of the alpha chain of the IL-2 receptor on CD25-negative T lymphocytes in the absence of exogenous antigenic stimulation. *Clinical Immunology* 97(3): 266–76.

[244] Han, Q., Bagheri, N. Bradshaw, E., et al. (2012). Polyfunctional responses by human T cells result from sequential release of cytokines. *Proceedings of the National Academy of Sciences of the United States of America* 109(5): 1607–12.

[245] Haining, W.N. (2012). Travels in time: Assessing the functional complexity of T cells. *Proceedings of the National Academy of Sciences of the United States of America* 109(5): 1359–60.

[246] Allele Frequency Net Database. URL: http://www.allelefrequencies.net (accessed on 13/10/2021)

[247] Kolodziej, S., Kuvardina, O., Oellerich, T., et al. (2014). PADI4 acts as a coactivator of Tal1 by counteracting repressive histone arginine methylation. *Nature Communications* 5: 3995.

[248] Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128: 693–705.

[249] Li, P., Wang, D., Yao, H., et al. (2010). Coordination of PAD4 and HDAC2 in the regulation of p53-target gene expression. *Oncogene* 29(21): 3153–62.

[250] Sharma, P., Azebi, S., England, P., et al. (2012). Citrullination of histone H3 interferes with HP1-mediated transcriptional repression. *PLoS Genetics* 8(9): 1–15.

[251] Zhang, X., Gamble, M., Stadler, S., et al. (2011). Genome-wide analysis reveals PADI4 cooperates with Elk-1 to activate C-Fos expression in breast cancer cells. *PLoS Genetics* 7.6: e1002112.

[252] Nakashima, K., Arai, S., Suzuki, A., et al. (2013). PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression. *Nature Communications* 4: 1836.

[253] Burr, M.L., Naseem, H., Hinks, A., et al. (2010). PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population. *Annals of the Rheumatic Diseases* 69(4): 666–70.

[254] Ikari, K., Kuwahara, M., Nakamura, T., et al. (2005). Association between PADI4 and rheumatoid arthritis: A replication study. *Arthritis and Rheumatism* 52(10): 3054–57.

[255] Kang, C.P., Lee, H., Ju, H., et al. (2006). A functional haplotype of the PADI4 gene associated with increased rheumatoid arthritis susceptibility in Koreans. *Arthritis and Rheumatism* 54(1): 90–96.

[256] Takata, Y., Inoue, H., Sato, A., et al. (2008). Replication of reported genetic associations of PADI4, FCRL3, SLC22A4 and RUNX1 genes with rheumatoid arthritis: Results of an independent Japanese population and evidence from meta-analysis of East Asian studies. *Journal of Human Genetics* 53(2): 163–73.

[257] Suzuki, A., Yamada, R., Change, X., et al. (2003). Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature Genetics* 34(4): 395–402.

[258] Martinez, A., Valdivia, A., Pascual-Salcedo, D., et al. (2005). PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population. *Rheumatology* 44(10): 1263–66.

[259] Plenge, R.M., Padyukov, L., Remmers, E.F., et al. (2005). Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: Association of susceptibility with PTPN22, CTLA4, and PADI4. *American Journal of Human Genetics* 77(6): 1044–60.

[260] Aguet, F., Brown, A., Castel, S., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550(7675): 204–13.

[261] Liao, D., and Li, H. (2020). Dissecting the niche for alveolar type II cells with alveolar organoids. *Frontiers in Cell and Developmental Biology* 8: 419.

[262] Bagger, F.O., Rapin, N., Theilgaard-Monch, K., et al. (2013). HemaExplorer: A database of mRNA expression profiles in normal and malignant haematopoiesis. *Nucleic Acids Research* 41(D1): 1034-39.

[263] Tyner, J.W., Tognon, C., Bottomly, D., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature* 562(7728): 526–31.

[264] Haferlach, T., Kohlmann, A., Wieczorek, L., et al. (2010). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: Report from the international microarray innovations in leukemia study group. *Journal of Clinical Oncology* 28(15): 2529-37.

[265] Liu, C., Tang, J., Li, C., et al. (2019). PADI4 stimulates esophageal squamous cell carcinoma tumor growth and up-regulates CA9 expression. *Molecular Carcinogenesis* 58(1): 66–75.

[266] Liu, M., Qu, X., Teng, X., et al. (2019). PADI4-mediated epithelial-mesenchymal transition in lung cancer cells. *Molecular Medicine Reports* 19(4): 3087–94.

[267] Slack, J.L., Causey, C.P., and Thompson, P.R. (2011). Protein arginine deiminase 4: A target for an epigenetic cancer therapy. *Cellular and Molecular Life Sciences* 68(4): 709–20.

[268] Weinstein, J.N., Collisson, E., Mills, G.B., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics* 45(10): 1113–20.

[269] Guan, E., Wang, J., and Norcross, M.A. (2001). Identification of human macrophage inflammatory proteins 1 alpha and 1 beta as a native secreted heterodimer. *Journal of Biological Chemistry* 276(15):12404-09.

[270] Francisco-Cruz, A., Aguilar-Santelises, M., Ramos-Espinosa, O., et al. (2014). Granulocyte-macrophage colony-stimulating factor: Not just another haematopoietic growth factor. *Medical Oncology* 31: 774.

[271] Thommen, D.S., and Schumacher, T.N. (2018). Perspective T cell dysfunction in cancer. *Cancer Cell* 33(4): 547–62.

[272] Takeuchi, A., and Sito, T. (2017). CD4 CTL, a cytotoxic subset of CD4+ T cells, their differentiation and function. *Frontiers in Immunology* 8: 194.

[273] Huang, D.C.S., Hahne, M., Schroeter, M., et al. (1999). Activation of Fas by FasL induces apoptosis by a mechanism that cannot be blocked by Bcl-2 or Bcl-xL. *Proceedings of the National Academy of Sciences of the United States of America* 96(26): 14871–76.

[274] Subramanian, A., Tamayo, P., Mootha, V., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102(43): 15545-50.

[275] Lee, A.J., and Ashkar, A.A. (2018). The dual nature of type I and type II interferons. *Frontiers in Immunology* 9: 2061.

[276] Krensky, A.M., and Clayberger, C. (2009). Biology and clinical relevance of granulysin. *Tissue Antigens* 73(3): 193–98.

[277] Patel, S., Homaei, A., and El-seedi, H.R. (2018). Cathepsins: Proteases that are vital for survival but can also be fatal. *Biomedicine and Pharmacotherapy* 105: 526–32.

[278] Mahuron, K.M., Moreau, J., Glasgow, J., et al. (2020). Layilin augments integrin activation to promote antitumor immunity. *Journal of Experimental Medicine* 217(9): e20192080.

[279] Brenna, E., Davydov, A., Ladell, K., et al. (2020). CD4+ T follicular helper cells in human tonsils and blood are clonally convergent but divergent from non-Tfh CD4 + Cells. *Cell Reports* 30(1): 137–52.

[280] Mamedov, I.Z., Britanova, O., Zvyagin, I., et al. (2013). Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Frontiers in Immunology* 4: 456.

[281] Li, H.M., Hiroi, T., Zhang, Y., et al. (2016). TCR repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. *Journal of Leukocyte Biology* 99(3): 505–13.

[282] Qi, Q., Liu, Y., Cheng, Y., et al. (2014). Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences of the United States of America* 111(36): 13139–44.

[283] Vujovic, M., Degn, K., Marin, F., et al. (2020). T cell receptor sequence clustering and antigen specificity. *Computational and Structural Biotechnology Journal* 18: 2166–73.

[284] Bagaev, D.V., Vroomans, R., Samir, J., et al. (2020). VDJdb in 2019: Database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Research* 48(D1): 1057–62.

[285] Shatrova, A.N., Mityushova, E., Vassilieva, I., et al. (2016). Time-dependent regulation of IL-2R-chain (CD25) expression by TCR signal strength and IL-2-induced STAT5 signaling in activated human blood T lymphocytes. *PLoS ONE* 11(12): 1–19.

[286] Nelde, A., Bilch, T., Heitmann, J., et al. (2021). SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nature Immunology* 22(1): 74–85.

[287] Bui, H.H., Sidney, J., Dinh, K., et al. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7: 153.

[288] Greenbaum, J., Sidney, J., Chung, J., et al. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* 63(6): 325–35.

[289] Jakobsen, M.K., and Gjerstor, M.F. (2020). CAR T-cell cancer therapy targeting surface cancer/testis antigens. *Frontiers in Immunology* 11: 1568.

[290] Uhlen, M., Zhang, C., Lee, S., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357(6352): eann2507.

[291] Steimle, V., Siegrist, C., Mottet, A., et al. (1994). Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* 265(5168): 106–09.

[292] Wosen, J.E., Muhopadhyay, D., MacAubas, C., et al. (2018). Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Frontiers in Immunology* 9: 2144.

[293] Butticè, G., Miller, J., Wang, L., et al. (2006). Interferon-gammainduces major histocompatibility class II transactivator (CIITA), which mediates collagen repression and major histocompatibility class II activation by human aortic smooth muscle cells. *Circulation Research* 98(4): 472–79.

[294] Axelrod, M.L., Cook, R., Johnson, D., et al. (2019). Biological consequences of MHC-II expression by tumor cells in cancer. *Clinical Cancer Research* 25(8): 2392–2402.

[295] Oh, D.Y., Kwek, S., Raju, S., et al. (2020). Intratumoral CD4+ T cells mediate anti-tumor cytotoxicity in human bladder cancer. *Cell* 181: 1–14.

[296] Rapoport, A.P., Stadtmauer, E., Binder-Scholl, G., et al (2015). NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. *Nature Medicine* 21(8): 914–21.

[297] Robbins, P.F., Kassim, S., Tran, T., et al. (2015). A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: Long-term follow-up and correlates with response. *Clinical Cancer Research* 21(5): 1019–27.

[298] Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 348(6230): 62–68.

[299] Plotkin, J.B. and Kudla, G. (2011). Synonymous but not the same. *National Review of Genetics* 12(1): 32–42.

[300] Thomas, S., Mohammed, F., Reijmers, R., et al. (2019). Framework engineering to produce dominant T cell receptors with enhanced antigen-specific function. *Nature Communications* 10(1): 1–15.

[301] Engels, B., Cam, H., Schuller, T., et al. (2003). Retroviral vectors for high-level transgene expression in T lymphocytes. *Human Gene Therapy* 14(12): 1155–68.

[302] Hacein-Bey-Abina, S., Hauer, J., Lim, A., et al. (2010). Efficacy of gene therapy for X-linked severe combined immunodeficiency. *New England Journal of Medicine* 363(4): 355–64.

[303] Braun, C.J., Boztug, K., Paruzynski, A., et al. (2014). Gene therapy for Wiskott-Aldrich syndrome - long-term efficacy and genotoxicity. *Science Translational Medicine* 6(254): 1–15.

[304] Herr, W., Eichinger, Y., Beshay, J., et al. (2017). HLA-DPB1 mismatch alleles represent powerful leukemia rejection antigens in CD4 T-cell immunotherapy after allogeneic stem-cell transplantation. *Leukemia* 31(2): 434–45.

[305] Pabst, C., Krosl, J., Fares, I., et al. (2014). Identification of small molecules that support human leukemia stem cell activity ex vivo. *Nature Methods* 11(4): 436–42.

[306] Boyer, M.W., Waller, E., Bray, R., et al. (2000). Cytokine upregulation of the antigen presenting function of acute myeloid leukemia cells. *Leukemia* 14(3): 412–18.

[307] Sanchez, P.V., Perry, R., Sarry, J., et al. (2009). A robust xenotransplantation model for acute myeloid leukemia. *Leukemia* 23(11): 2109–17.

[308] Reynisson, B., Alvarez, B., Paul, S., et al. (2021). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research* 48: W449–54.

[309] Gfeller, D., and Bassani-Sternberg, M. (2018). Predicting antigen presentation - What could we learn from a million peptides? *Frontiers in Immunology* 9: 1716.

[310] Kahles, A., Lehmann, K., Toussaint, N., et al. (2018). Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 34(2): 211–24.

[311] Smart, A.C., Margolis, C., Pimentel, H., et al. (2018). Intron retention is a source of neoepitopes in cancer. *Nature Biotechnology* 36(11): 1056–58.

[312] Jayasinghe, R.G., Cao, S., Gao, Q., et al. (2018). Systematic analysis of splice-site-creating mutations in cancer. *Cell Reports* 23(1): 270–81.

[313] Roerden, M., Nelde, A., and Walz, J.S. (2019). Neoantigens in hematological malignancies—ultimate targets for immunotherapy? *Frontiers in Immunology* 10: 3004.

[314] Chaudhry, M.S., Velardi, E., Malard, F., et al. (2017). Immune reconstitution after allogeneic hematopoietic stem cell transplantation: Time to T up the thymus. *Journal of Immunology* 198(1): 40–46.

[315] Ogonek, J., Juric, M., Ghimire, S., et al. (2016). Immune reconstitution after allogeneic hematopoietic stem cell transplantation. *Frontiers in Immunology* 7: 507.

[316] Elias, S., and Rudensky, A.Y. (2021). Therapeutic use of regulatory T cells for graft-versus-host disease. *British Journal of Haematology* 187(1): 25–38.

[317] Sykes, M. (2007). Immune tolerance: Mechanisms and application in clinical transplantation. *Journal of Internal Medicine* 262(3): 288–310.

[318] Zvyagin, I.V., Mamedov, I., Tatarinova, O., et al. (2017). Tracking T-cell immune reconstitution after TCR/CD19-depleted hematopoietic cells transplantation in children. *Leukemia* 31(5): 1145–53.

[319] Castro, F., Cardoso, A., Goncalves, R., et al. (2018). Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Frontiers in Immunology* 9: 847.

[320] Hashimoto, K., Kouno, T., Ikawa, T., et al. (2019). Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proceedings of the National Academy of Sciences of the United States of America* 116(48): 24242–51.

[321] McKinney, D.M., Southwood, S., Hinz, D., et al. (2013). A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. *Immunogenetics* 65(5): 357–70.

[322] Shah, N.N., and Fry, T.J. (2019). Mechanisms of resistance to CAR T cell therapy. *Nature Reviews Clinical Oncology* 16(6): 372–85.

[323] Song, W., and Zhang, M. (2020). Use of CAR-T cell therapy, PD-1 blockade, and their combination for the treatment of hematological malignancies. *Clinical Immunology* 214(1): 108382.

[324] Linette, G.P., Stadtmauer, E., Maus, M.V. et al. (2013). Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* 122(6): 863–71.

[325] Philip, B., Kokalaki, E., Mekkaoui, L., et al. (2014). A highly compact epitope-based marker/suicide gene for easier and safer T-cell therapy. *Blood* 124(8): 1277–87.

[326] Perez, C., Gruber, I., and Arber, C. (2020). Off-the-shelf allogeneic T cell therapies for cancer: Opportunities and challenges using naturally occurring "universal" donor T cells. *Frontiers in Immunology* 11: 583716.