



UNIVERSITAT DE  
BARCELONA

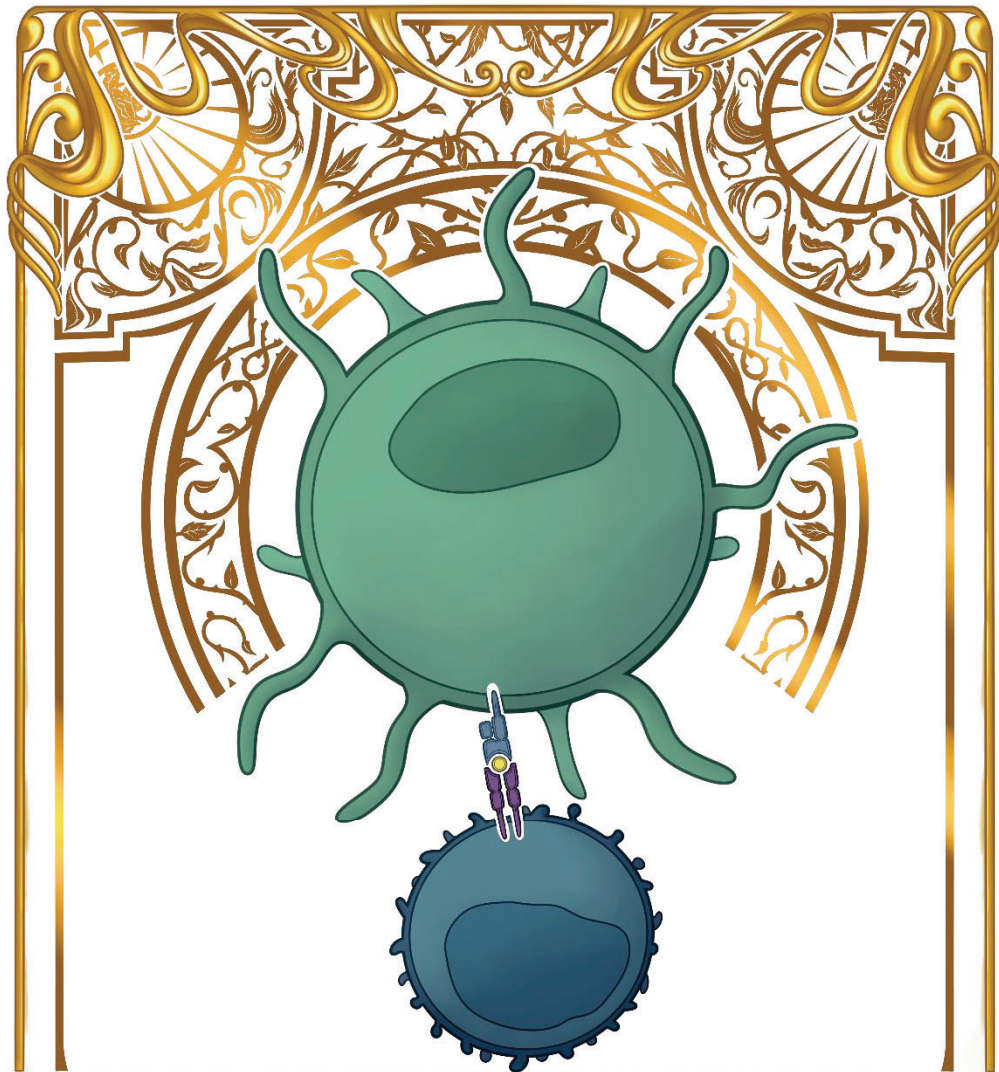
## A Cancer Prevention Dendritic Cell Vaccine: Frameshift-Derived Neoantigens as Immunopreventive Targets in Lynch Syndrome

Cristina Bayó Llorens

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



# A CANCER PREVENTION DENDRITIC CELL VACCINE:

*Frameshift-derived Neoantigens  
As Immunopreventive Targets in  
Lynch Syndrome*

CRISTINA BAYÓ LLORENS  
BARCELONA, 2025  
DOCTORAL THESIS







Programa de Doctorat en Biomedicina

Línia d'investigació en Immunologia

2020 – 2025



**A Cancer Prevention Dendritic Cell Vaccine: Frameshift-Derived Neoantigens as Immunopreventive Targets in Lynch Syndrome**

\*

**Vacunes De Cèl·lules Dendrítiques per la Prevenció del Càncer: Neoantígens *Frameshift-Derived* com a Nous Objectius Immunoterapèutics en Síndrome de Lynch**

Tesis doctoral presentada per:

Cristina Bayó Llorens

Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic de Barcelona

Departament d'immunologia, secció Immunoteràpia

La doctoranda

Cristina  
Bayó Llorens

El co-director

Dr. Francesc  
Balaguer

El co-director

Dr. Daniel  
Benítez Ribas

El Tutor

Dr. Manel  
Juan Otero



*Als meus pares.*

*“There's a grain of truth in every fairy tale”*

*— Andrzej Sapkowski, The Last Wish*



## **AGRAÏMENTS**

A tothom que, d'alguna manera o d'altre, m'hagi acompanyat durant aquest viatge, em vull adreçar a vosaltres per agrair-vos-ho com és necessari. Probablement em deixi coses, ja sabeu com soc, però estic segura que sabreu perdonar-me i sabeu qui sou els que heu d'estar aquí.

Vull començar agraint a tota la comunitat amb Síndrome de Lynch la participació activa en el nostre projecte, la motivació i el recolzament donat, i destacar les seves ganes de seguir avançant en ciència. Es un plaer investigar per a millorar la vida de comunitats com aquesta.

Primerament, donar les gràcies als meus directors de tesis, Daniel i Francesc, que m'han donat suport i guia des del primer moment. Gràcies per dipositar-me la confiança de dur a terme aquest projecte, per recolzar les meves idees (i criticar les que feien falta) i permetre'm dur-les a terme, i sobretot per ser un suport en la ciència quan més ho he necessitat. M'heu fet sentir segura de mi mateixa, i no podria haver caigut en millors mans. Al Dani, sobretot, gràcies per atendre'm en tot moment i per aguantar-me quan estava treballant amb tu quan cobria la posició de QA durant tota la pandèmia, malgrat tot així es va fer molt més amè. Gràcies també al meu tutor de tesis i cap de servei, el Manel Juan, per donar suport al projecte i a les persones treballant-hi sempre que ho hem necessitat.

A tot l'equip de Gastroenterologia, per permetre que el circuit per obtenir les mostres necessàries per aquest projecte fos el més fluït possible, i per donar-me suport i acollir-me a la seva planta quan no tenia altre lloc de treball. Sobretot també a Joaquin, per ajudar-me en tot moment en tasques que a mi se m'escapen completament, i per l'entusiasme que li has posat al projecte des que vas afegir-t'hi.

I want to also acknowledge and appreciate the support given by the group from Jolanda de Vries in Netherlands and from the group of Beatriz Carreño in Pennsylvania. I am deeply grateful for your knowledge and insights, for kindly sharing samples back and forth, and for the collaborations that have surged over these years, that have allowed us to advance in ways we could not have. I hope to keep up this relationship in the future.

A Giancarlo, gracias por aceptar meterte en este proyecto cuando aún no era nada y no tenía ni idea de bioinformática. Por tu carácter y tu amabilidad, por tu paciencia y por todas las horas mirando secuencias juntos e intentando descifrar como funcionaban programas que ninguno habíamos visto nunca. Por todo ello, esto ha podido tirar adelante.

A tota la gent que m'ha ensenyat, guiat, ajudat o ha compartit moments bons i dolents amb mi al laboratori: Començant per el "DC team" de l'inici, que em vau conèixer al màster i em vau ensenyar mil coses quan anava perdudíssima: A Luchy, Libertad, Carla i JR, per acollir-me i fer de mi una més, i Carol, per això, i per tots els moments compartits, i per deixar-me casa teva (i conèixer al Lolo). I a tots els que s'han anat afegint després i han permès que les nostres immunoteràpies segueixin avançant.

A la Gina, gràcies, des de l'entrevista del principi de tot, passant per mil coses i fins que has tornat de Holanda i t'has tornat a implicar en el projecte amb molta il·lusió, has estat un suport importantíssim. I a tu, Raquel, per agafar el relleu de guiar-me quan no hi era la Gina i per donar-me la oportunitat (i confiar en mi) per poder cobrir la teva posició durant la baixa, cosa que em va permetre començar el doctorat. Sense tu, no seria on soc ara.

A Marta por hacerme compañía durante el inicio, por las salidas con los resis a las que me acogisteis, y por la de horas que estuvimos entre papeles y organoides.

A tu, Ari, per acompanyar-me durant tota aquesta odissea, per compartir mil moments, plorar i riure juntes, ajudar-nos en tot i passar per mil trasllats. De veritat, valoro molt que estiguessis allà, i espero que continuem la amistat molt molt temps.

I això em porta al primer trasllat, al CELLEX. Gràcies a tots els de la 4B, els Mitos, els Vàsculs i els Predis, per l'ambient que vam crear i per tota la ajuda que ens vam donar encara que treballéssim en grups diferents. I tot això entre dinars al Goiko, que arreglaven completament una setmana de merda.

Agraeixo també aquí haver conegut a la Berta, i ja continuar juntes fins al CEK. Berta, gràcies per tot, per la amistat i tots els moments que hem compartit, tant al lab com fora. Els cafès de cotilleo de tant en tant son indispensables per mi ara.

A tota la gent del CEK on he acabat aterrant: Lorena, Sergio, Sergi, Alexis, Berta, Joan, Blanca, Alejandro, Mario, Mercedes, Rober, Andrea, Elisabet, Erik, Valeria, Mireia, Nela i Maria (A tu, sobretot, agrair-te especialment haver trobat algú friki per compartir les meves frikades). Gràcies per haver fet del grup una família, a vegades disfuncional, que dona suport en tot moment, encara que cadascú treballi en projectes diferents. I sobretot per compartir la vostra amistat tant dins com fora del lab.

I per últim, agrair a tota la gent que m'ha acompanyat fora de la feina, que m'ha aguantat mentre estava en mals moments i ha celebrat amb mi els bons.

Primer de tot a la meva família, als meus pares, que m'han permès seguir el camí que he volgut. M'heu donat suport incondicional en tot moment i heu estat sempre al meu costat; suposo que ho dic menys del que hauria, però moltes gràcies per ser com sou. I per aguantar venir a la tesis i posar atenció al que us explico, encara que molts cops no m'entengueu.

A mi gente de MQ, a todos y especialmente a Sara y Víctor, por los Amovinos, los Viajes, las salidas, las sesiones de hablar de nuestra vida. No sabéis lo importante que ha sido (y es) para mí. Dese micro y hasta siempre, love u.

A la gente del máster y en especial a Luís, sé que soy un desastre manteniendo el contacto, eres una inspiración y una persona maravillosa, y espero que vuelvan las quedadas de juegos de mesa.

Xavi, no necessito dir-te res que no sàpigues perquè ens entenem sense paraules. Gràcies per estar allà sempre i treure'm de casa (o envair-la).

I acabant, i probablement el més important, Manel (el meu), no puc imaginar-me aquest camí sense tu al meu costat, has hagut d'aguantar mil canvis de humor, moments en els que estava derrotada i has compartit amb mi totes les alegries. T'estimo, ara i sempre.



# INDEX

|           |                                                                                               |    |
|-----------|-----------------------------------------------------------------------------------------------|----|
| 1         | Introduction.....                                                                             | 33 |
| 1.1       | Lynch Syndrome .....                                                                          | 33 |
| 1.1.1     | Epidemiology and diagnosis.....                                                               | 33 |
| 1.1.2     | Surveillance .....                                                                            | 35 |
| 1.1.3     | Molecular Mechanisms.....                                                                     | 36 |
| 1.1.3.1   | DNA damage and repair systems.....                                                            | 36 |
| 1.1.3.2   | Base mismatches and the MMR system.....                                                       | 38 |
| 1.1.3.3   | MMR deficiency and microsatellite instability (MSI) .....                                     | 39 |
| 1.1.4     | Characteristics of tumours associated to LS .....                                             | 43 |
| 1.1.5     | Challenges, treatment and immunotherapy.....                                                  | 44 |
| 1.1.5.1   | Challenges in LS management and psychological impact.... <b>¡Error! Marcador no definido.</b> |    |
| 1.1.5.2   | Current prevention and prophylaxis for LS-related cancers .....                               | 44 |
| 1.1.5.3   | LS MSI tumour treatment strategies.....                                                       | 47 |
| 1.1.5.4   | Introduction to immunotherapy.....                                                            | 48 |
| 1.1.5.5   | Immunotherapy strategies for LS (ICI, cancer vaccines) .....                                  | 49 |
| 1.1.5.5.1 | Immune checkpoint inhibitors (ICIs).....                                                      | 49 |
| 1.1.5.5.2 | Cancer vaccines.....                                                                          | 49 |
| 1.2       | Immunology: The adaptive cellular response .....                                              | 52 |
| 1.2.1     | Dendritic cells.....                                                                          | 53 |
| 1.2.1.1   | Dendritic cell Subtypes.....                                                                  | 54 |
| 1.2.1.1.1 | Conventional Dendritic Cells.....                                                             | 55 |
| 1.2.1.1.2 | Plasmacytoid Dendritic cells.....                                                             | 56 |
| 1.2.1.1.3 | Monocyte-derived Dendritic cells.....                                                         | 57 |
| 1.2.1.2   | Maturation and antigen processing .....                                                       | 58 |
| 1.2.1.2.1 | The MHC system.....                                                                           | 58 |
| 1.2.1.2.2 | Antigen processing.....                                                                       | 61 |
| 1.2.1.2.3 | Cross-presentation .....                                                                      | 64 |
| 1.2.1.3   | Antigen presentation and Lymphocyte activation .....                                          | 65 |

|           |                                                                      |    |
|-----------|----------------------------------------------------------------------|----|
| 1.2.1.3.1 | DC migration to LNs.....                                             | 65 |
| 1.2.2     | T-lymphocytes .....                                                  | 66 |
| 1.2.2.1   | T-cell ontogeny.....                                                 | 67 |
| 1.2.2.2   | T-cell activation: Antigen presentation and the TCR.....             | 67 |
| 1.2.2.2.1 | The TCR complex.....                                                 | 69 |
| 1.2.2.3   | CD4+ T-helper cells differentiation.....                             | 70 |
| 1.2.2.3.1 | Th1 cells.....                                                       | 71 |
| 1.2.2.3.2 | Th2 cells.....                                                       | 71 |
| 1.2.2.3.3 | Th9 cells.....                                                       | 71 |
| 1.2.2.3.4 | Th17 cells.....                                                      | 72 |
| 1.2.2.3.5 | T-reg cells .....                                                    | 72 |
| 1.2.2.4   | CD4+ and CD8+ T-cell interaction.....                                | 73 |
| 1.2.2.5   | T-cell Memory and effector differentiation.....                      | 74 |
| 1.2.2.5.1 | CD8+ subsets.....                                                    | 74 |
| 1.2.2.5.2 | CD4+ subsets.....                                                    | 75 |
| 1.2.3     | Immunology in Lynch Syndrome.....                                    | 76 |
| 1.2.3.1   | Immune surveillance in LS and antitumour immunity.....               | 76 |
| 1.2.3.2   | Immune escape .....                                                  | 78 |
| 1.3       | Cancer immunoprevention in LS .....                                  | 79 |
| 1.3.1     | Immunology of cancer prevention vaccines .....                       | 80 |
| 1.3.2     | Vaccine vectors: Cellular vaccines .....                             | 82 |
| 1.3.3     | Antigen choice.....                                                  | 84 |
| 1.3.3.1   | Tumour-Associated Antigens.....                                      | 84 |
| 1.3.3.2   | Cancer-Testis Antigens.....                                          | 84 |
| 1.3.3.3   | Neoantigens.....                                                     | 84 |
| 1.3.4     | Identification of candidate neoantigens .....                        | 85 |
| 1.3.5     | State of the art in neoantigen-based cancer prevention vaccines..... | 87 |
| 2         | Hypothesis and objectives.....                                       | 90 |
| 2.1       | Hypothesis .....                                                     | 90 |
| 2.2       | Objectives .....                                                     | 91 |

|           |                                                                           |     |
|-----------|---------------------------------------------------------------------------|-----|
| 3         | Materials and Methods.....                                                | 92  |
| 3.1       | Samples and sequences sources.....                                        | 92  |
| 3.1.1     | Bioinformatic sequences.....                                              | 92  |
| 3.1.1.1   | Data collection for FSDN predictions.....                                 | 92  |
| 3.1.1.2   | Colorectal adenomas (CrAD) and ECs FFPE samples source.....               | 92  |
| 3.1.1.3   | Data source for CRC and CrAD DNA sequences.....                           | 93  |
| 3.1.1.4   | Data source for sporadic MMRd genetic variants .....                      | 93  |
| 3.1.1.5   | Data source for MSS and MSI cell lines .....                              | 93  |
| 3.1.2     | Biopsies and Peripheral blood.....                                        | 94  |
| 3.1.3     | Cell lines .....                                                          | 94  |
| 3.2       | Bioinformatic analysis.....                                               | 95  |
| 3.2.1     | Neoantigen prediction .....                                               | 95  |
| 3.2.1.1   | Data filtering and neoantigens prioritization.....                        | 96  |
| 3.2.1.2   | NMD .....                                                                 | 96  |
| 3.2.1.3   | Pathogen and self-similarity.....                                         | 96  |
| 3.2.2     | Library preparations and sequencing samples from LS cancer patients ..... | 97  |
| 3.2.3     | DNA sequencing analysis and variant calling .....                         | 99  |
| 3.2.4     | pVACseq analysis.....                                                     | 100 |
| 3.2.5     | RNA sequencing analysis .....                                             | 100 |
| 3.2.6     | Statistic tests.....                                                      | 100 |
| 3.3       | In-vitro validation of FSDN.....                                          | 101 |
| 3.3.1     | FSDN synthesis .....                                                      | 101 |
| 3.3.2     | Primary cultures.....                                                     | 101 |
| 3.3.2.1   | PBMCs, PBLs and moDCs.....                                                | 101 |
| 3.3.2.2   | Tumour-infiltrating lymphocytes (TILs) from Colorectal Biopsies .....     | 102 |
| 3.3.2.2.1 | Colorectal Biopsies processing .....                                      | 102 |
| 3.3.2.2.2 | Polyclonal expansion of TILs and TRLs .....                               | 102 |
| 3.3.3     | Cell lines .....                                                          | 103 |

|           |                                                                                                |     |
|-----------|------------------------------------------------------------------------------------------------|-----|
| 3.3.3.1   | <i>Generation and culture of monoallelic K562-based Cell Lines expressing vaccine peptides</i> | 103 |
| 3.3.3.1.1 | <i>TMG construction process</i>                                                                | 104 |
| 3.3.3.1.2 | <i>Bacterial transformation</i>                                                                | 105 |
| 3.3.3.1.3 | <i>Lentiviral production</i>                                                                   | 106 |
| 3.3.3.1.4 | <i>Lentiviral titration</i>                                                                    | 106 |
| 3.3.4     | <i>Cell phenotyping and characterization assays</i>                                            | 107 |
| 3.3.4.1   | <i>Staining and detection of surface cellular markers by flow cytometry</i>                    | 107 |
| 3.3.4.2   | <i>Single cell multiplex cytokine profiling</i>                                                | 107 |
| 3.3.4.3   | <i>TCR<math>\gamma</math><math>\beta</math> clonogram analysis</i>                             | 108 |
| 3.3.4.3.1 | <i>Cluster and expansions analysis of TCR <math>\gamma</math><math>\beta</math> clonogram</i>  | 109 |
| 3.3.5     | <i>Co-cultures and cell reactivity assays</i>                                                  | 109 |
| 3.3.5.1   | <i>TILs and moDCs co-cultures for bulk IFN-<math>\gamma</math> ELISPOTs</i>                    | 109 |
| 3.3.5.2   | <i>Specific T-cell expansion with moDCs</i>                                                    | 110 |
| 3.3.5.3   | <i>Specific T-cell expansion with autologous PBMCs</i>                                         | 111 |
| 3.3.6     | <i>Statistical analysis</i>                                                                    | 112 |
| 3.4       | <b>CLINICAL TRIAL</b>                                                                          | 113 |
| 3.4.1     | <i>Drug substance</i>                                                                          | 113 |
| 3.4.1.1   | <i>Manufacturing process</i>                                                                   | 113 |
| 3.4.2     | <i>Design</i>                                                                                  | 115 |
| 3.4.2.1   | <i>Study objectives and endpoints</i>                                                          | 115 |
| 3.4.3     | <i>Patients</i>                                                                                | 117 |
| 3.4.3.1   | <i>Inclusion, exclusion and discontinuation</i>                                                | 117 |
| 3.4.4     | <i>Treatment and clinical schedule</i>                                                         | 117 |
| 3.4.5     | <i>Immune monitoring</i>                                                                       | 118 |
| 3.4.6     | <i>Clinical efficacy and safety evaluation</i>                                                 | 118 |
| 3.4.6.1   | <i>Safety</i>                                                                                  | 118 |
| 3.4.6.2   | <i>Clinical efficacy</i>                                                                       | 119 |
|           |                                                                                                | 121 |
| 4         | <b>CHAPTER I: NEOANTIGENS PREDICTION, VALIDATION AND PRIORITIZATION</b>                        | 123 |

|         |                                                                                                        |            |
|---------|--------------------------------------------------------------------------------------------------------|------------|
| 4.1     | Prediction and characterization of shared neoantigens in LS-related tumours.....                       | 123        |
| 4.2     | Occurrence of FSDN variants in LS pre-cancers and cancers.....                                         | 127        |
| 4.2.1   | <i>Analysis of variants from targeted cMS NGS sequencing from LS patients CrAD and EC samples.....</i> | <i>127</i> |
| 4.2.1.1 | <i>NGS sequencing data analysis.....</i>                                                               | <i>127</i> |
| 4.2.1.2 | <i>FSDN HLA-binding prediction for analysed variants.....</i>                                          | <i>129</i> |
| 4.2.2   | <i>Analysis of WGS data from LS CrADs and CRCs sequences .....</i>                                     | <i>130</i> |
| 4.3     | Occurrence of FSDN variants in non-LS MSI-h cancers.....                                               | 134        |
| 4.3.1   | <i>pFSDN associated variants in non-LS MMRd human tumours.....</i>                                     | <i>134</i> |
| 4.3.2   | <i>pFSDN associated variants in MSI and MSS cancer cell lines .....</i>                                | <i>136</i> |
| 4.3.2.1 | <i>Analysis of FSDN mutations occurrence on CL DNA sequences .....</i>                                 | <i>137</i> |
| 4.3.2.2 | <i>Analysis of FSDN mutations occurrence on CL mRNA sequences.....</i>                                 | <i>139</i> |
| 4.4     | <i>In-vitro bulk FSDN-loaded moDCs recognition assays .....</i>                                        | <i>141</i> |
| 4.5     | Prioritization of optimal FSDN .....                                                                   | 144        |
| 4.6     | Peptide pooling for the vaccine product .....                                                          | 147        |
| 5       | CHAPTER II: IDENTIFICATION AND CHARACTERIZATION OF DELAY FSDN-SPECIFIC T-CELLS .....                   | 153        |
| 5.1     | Patients and samples characteristics .....                                                             | 153        |
| 5.1.1   | <i>Demographics and HLA allele distribution analysis.....</i>                                          | <i>153</i> |
| 5.1.2   | <i>Lynch syndrome patients T-cell populations characterization .....</i>                               | <i>156</i> |
| 5.1.2.1 | <i>Surface phenotype markers analysis .....</i>                                                        | <i>158</i> |
| 5.1.2.2 | <i>TCR V<math>\beta</math> repertoire analysis.....</i>                                                | <i>162</i> |
| 5.1.2.3 | <i>Single cell secretome analysis.....</i>                                                             | <i>167</i> |
| 5.1.3   | <i>FSDNs do not impair Dendritic Cell maturation capacity.....</i>                                     | <i>172</i> |
| 5.2     | DELAY-FSDN-specific T-cells are present on normal mucosa and pre-cancers of LS carriers .....          | 174        |
| 5.2.1   | <i>Detection of mucosa-infiltrating neoantigen specific cells by dextramers .....</i>                  | <i>174</i> |
| 5.2.2   | <i>Detection of neoantigen-specific T-cells by IFN-<math>\gamma</math> production.....</i>             | <i>177</i> |
| 5.3     | Creating a DELAY-FSDN-expressing target cell line .....                                                | 182        |

|         |                                                                        |            |
|---------|------------------------------------------------------------------------|------------|
| 6       | CHAPTER III: A PHASE Ib FIRST-IN-HUMAN CLINICAL TRIAL .....            | 187        |
| 6.1     | Validation results.....                                                | 187        |
| 6.1.1   | <i>Product generation and characterization .....</i>                   | <i>187</i> |
| 6.1.1.1 | <i>Product specification compliance.....</i>                           | <i>188</i> |
| 6.1.1.2 | <i>Product phenotype and potency.....</i>                              | <i>189</i> |
| 6.1.2   | <i>Immunomonitoring – basal vaccine immunogenicity assessment.....</i> | <i>192</i> |
| 7       | Discussion and Conclusions .....                                       | 195        |
| 7.1     | Discussion .....                                                       | 195        |
| 7.1.1   | <i>Chapter I .....</i>                                                 | <i>195</i> |
| 7.1.2   | <i>Chapter II.....</i>                                                 | <i>198</i> |
| 7.1.3   | <i>Chapter III .....</i>                                               | <i>205</i> |
| 7.2     | Conclusions.....                                                       | 209        |
| 8       | Supplementary material.....                                            | 238        |

## **RESUM I PARAULES CLAU**

El Síndrome de Lynch (LS) és la causa més comuna de càncer colorectal hereditari i està associat a variants patogèniques germinals en els gens de reparació d'errors de replicació de l'ADN (MMR, Mismatch Repair). La deficiència en aquest sistema de reparació provoca una acumulació de mutacions principalment en seqüències repetitives, generant inestabilitat de microsatèl·lits (MSI) i la presència de neoantígens específics derivats de mutacions de canvi de marc de lectura (*frameshift*). Malgrat els avenços en les estratègies de vigilància i prevenció, els portadors de LS continuen tenint un risc elevat de desenvolupar múltiples tumors al llarg de la seva vida.

Aquesta tesi investiga el potencial dels neoantígens generats per mutacions *frameshift* en seqüències mononucleotídiques repetitives (MNR) com a dianes immunoterapèutiques per a la prevenció del càncer en portadors de LS. S'han analitzat dades genòmiques de mutacions freqüents en tumors i lesions pre-canceroses MSI amb l'objectiu d'identificar variants compartides entre pacients i prioritzar aquelles amb més potencial immunogènic. A través d'eines bioinformàtiques, es van seleccionar seqüències amb una alta probabilitat de ser presentades per diverses molècules del Complex Major d'Histocompatibilitat (MHC) i, per tant, reconegudes pel sistema immunitari de la majoria de la població.

Un cop seleccionades aquestes seqüències, es va validar la presència de les seves mutacions associades en diferents tipus de tumors, descobrint que un percentatge significatiu d'aquests neoantígens és compartit entre diversos càncers MSI, tant associats com no associats a LS. Això suggereix que existeix un perfil comú de neoantígens en tumors amb deficiència en MMR, la qual cosa podria permetre l'ampliació de l'abast de la nostra vacuna.

Els experiments *in vitro* van confirmar la immunogenicitat dels neoantígens seleccionats, demostrant que poden ser reconeguts per limfòcits T de pacients amb LS, tant si tenen antecedents de càncer com si no. Això dona suport a la hipòtesi que aquests individus desenvolupen respostes immunitàries i eliminen lesions MSI fins i tot abans que siguin clínicament detectables, generant una memòria immunològica que podria ser ampliada mitjançant la vacunació.

Sobre aquesta base, s'ha desenvolupat una vacuna preventiva basada en cèl·lules dendrítiques, composta per una selecció dels neoantígens més prometedors. L'ús d'aquestes cèl·lules en un estadi preventiu manté tot el seu potencial en la inducció de respostes immunes robustes i fisiològiques. Aquesta formulació s'avaluarà en un assaig clínic de fase Ib, l'objectiu principal del qual és determinar la seguretat de la vacuna i la seva capacitat per induir una resposta immune protectora en portadors de LS. Els resultats preliminars obtinguts en les fases de validació de l'assaig mostren que l'estratègia és factible i que existeix un reconeixement específic d'alguns neoantígens per part de cèl·lules T en individus abans de la vacunació, fet que suggereix un potencial preventiu rellevant.

En conjunt, aquesta tesi demostra que els neoantígens derivats de mutacions *frameshift* representen una estratègia prometedora per al desenvolupament d'immunoteràpies i estratègies d'immunoprevenció en LS i altres tumors MSI, fent un pas endavant cap a la millora de la qualitat de vida d'aquests individus.

**Paraules clau:** Síndrome de Lynch, Vacuna, Cèl·lula Dendrítica, Càncer, Neoantígens

## ***SUMMARY AND KEYWORDS***

Lynch Syndrome (LS) is the most common cause of hereditary colorectal cancer and is associated with pathogenic germline variants in DNA mismatch repair (MMR) genes. The deficiency in this repair system leads to an accumulation of mutations, particularly in repetitive sequences, resulting in microsatellite instability (MSI) and the presence of specific neoantigens derived from frameshift mutations. Despite advances in surveillance and prevention strategies, LS carriers continue to face a high lifetime risk of developing multiple tumours.

This thesis investigates the potential of neoantigens generated by frameshift mutations in mononucleotide repeat (MNR) sequences as immunotherapeutic targets for cancer prevention in LS carriers. Genomic data from frequent mutations in MSI tumours and precancerous lesions were analysed with the goal of identifying variants shared among patients and of prioritizing those with the highest immunogenic potential. Using bioinformatics tools, frameshift sequences were selected based on their high likelihood of being presented by various Human Leukocyte Antigen (HLA) alleles, and thus recognized by the immune system of the majority of the population.

Once these sequences were selected, the presence of their associated mutations was validated in different tumour types, revealing that a significant proportion of these neoantigens are shared among various MSI cancers, both LS-related and unrelated. This suggests the existence of a common neoantigen profile in MMR-deficient tumours, potentially allowing for broader application of our vaccine.

In vitro experiments confirmed the immunogenicity of the selected neoantigens, showing they can be recognized by T lymphocytes from LS patients, regardless of cancer history. This supports the hypothesis that these individuals develop immune responses and eliminate MSI lesions even before clinical detection, generating immunological memory that could be enhanced through vaccination.

On this basis, a preventive dendritic cell (DC)-based vaccine was developed, composed of a selection of the most promising neoantigens. Using these cells at a preventive stage retains their full potential to induce robust and physiological immune responses. This formulation will be evaluated in a phase Ib clinical trial, whose main goal is to determine

the vaccine's safety and its ability to induce a protective immune response in LS carriers. Preliminary results from the validation phases of the trial show that the strategy is feasible and that there is specific recognition of certain neoantigens by T cells in individuals prior to vaccination, suggesting relevant preventive potential.

Overall, this thesis demonstrates that neoantigens derived from frameshift mutations represent a promising strategy for the development of immunotherapies and immunoprevention approaches in LS and other MSI tumours, marking a step forward in improving the quality of life for these individuals.

**Keywords:** Lynch Syndrome, Vaccine, Dendritic Cell, Cancer, Neoantigens

## ***ABBREVIATIONS***

### **A**

**Ab:** Antibody

**ACN:** Acetonitrile

**ACT:** Adoptive cell therapy

**AEMPs:** Agencia Española de Medicamentos y Productos sanitarios

**AF:** Allele Frequency

**ANOVA:** Analysis of Variance

**APA:** American Psychology Association

**APC:** Antigen Presenting Cell

**ATCC:** American Type Culture Collection

**ATP:** Adenosine Triphosphate

### **B**

**BCR:** B cell receptors

**BER:** Base Excision Repair

**BLAST:** Basic Local Alignment Search Tool

**BSC:** Barcelona Supercomputing Center

**BWA:** Burrow-Wheeler Alignment

### **C**

**CAR-T:** Chimeric Antigen Receptor T-cells

**CD:** Cluster of Differentiation

**cDC:** Conventional Dendritic Cells

**CDR:** Complementarity-Determining Region

**CL:** Cell Line

**cMNR:** coding Mononucleotide Repeat

**cMS:** coding Microsatellite

**CP:** Colon Polyp

**CPV:** Cancer Prevention Vaccine

**CrAD:** Colorectal Adenoma

**CRC:** Colorectal Cancer

**CTA:** Caner Testis Antigen

**CTL:** Cytotoxic T lymphocytes

**CTLA4:** Cytotoxic T-Lymphocyte Antigen 4

## D

**DAMP:** Damage-Associated Molecular Patterns

**DC:** Dendritic Cell

**DELAY:** DEndritic ceLL-based vAccine for lYnch syndrome

**DMEM:** Dulbecco's Modified Eagle Medium

**DMSO:** Dimethyl sulfoxide

**DNA:** DeoxyriboNucleic Acid

**DSB:** Double-strand Break

## E

**EA:** Endometroid Adenocarcinoma

**EC:** Endometrial Cancer

**ELISPOT:** Enzyme-Linked ImmunoSpot Assay

**EOCRC:** Early Onset Colorectal Cancer

**EPCAM:** Epithelial Cell Adhesion Molecule

**ER:** Endoplasmatic Reticulum

## F

**FBS:** Fetal Bovine Serum

**FDA:** Food and Drug Administration

**FFPE:** Formalin-fixed paraffined-embedded

**FITC:** Fluorescein Isothiocyanate

**FSDN:** FrameShift Derived Neoantigen

**FSP:** FrameShift Peptide

## G

**GFP:** Green Fluorescence Protein

**GM-CSF:** Granulocyte Monocyte-Colony Stimulating Factor

**GMDP:** Granulocyte, monocyte, and DC progenitors

**GMP:** Good Manufacture Practices

**GRCh:** Genome Reference Consortium Human Build

## H

**HCB:** Hospital Clinic de Barcelona

**HGD:** High Grade Dysplasia

**HLA:** Human Leukocyte Antigen

**HNPCC:** Hereditary Non-Polyposis Colon Cancer

**HS:** Human Serum

**HSC:** Hematopoietic Stem Cells

**IC50:** Half maximal inhibitory concentration

**ICB:** Immune Checkpoint Blockade

**ICI:** Immune Checkpoint Inhibitors

**IEDB:** Immune Epitope Database

**IFN:** Interferon

**IHC:** ImmunoHistoChemistry

**IL:** Interleukin

**InDel:** Insertion and Deletion

**IQR:** InterQuartile Range

**IRF-7:** Interferon Regulatory Factor-7

**IS:** Immune System

**LGD:** Low Grade Dysplasia

**LN:** Lymphnode

**LS:** Lynch Syndrome

**mAb:** Monoclonal Antibody

**MDP:** Monocyte and DC progenitors

**MHC:** Major Histocompatibility Complex

***MLH1***: mutL Homolog 1

**MMR**: MisMatch Repair

**MMRd**: MisMatch Repair deficiency

**MNR**: Mononucleotide Repeat

**moDC**: Monocyte-derived Dendritic Cells

**MOI**: Multiplicity Of Infection

**MS**: MicroSatellite

***MSH2***: MutS, Homolog 2

***MSH6***: MutS Homolog 6

**MSI**: MicroSatellite Instability

**MSS**: Microsatellite Stability

## N

**NCBI**: National Center for Biotechnology Information

**NCCN**: National Comprehensive Cancer Network

**NCI**: National Cancer Institute

**NCT**: National Clinical Trial

**NER**: Nucleotide excision repair

**NGS**: Next Generation Sequencing

**NIH**: National Institute of Health

**NK**: Natural Killer cells

**NKT**: Natural Killer T cells

**NMD**: Nonsense Mediated RNA decay

# O

**ON:** Over Night

**ORF:** Open Reading Frame

# P

**PAMP:** Pathogen-Associated Molecular Patterns

**PBLs:** Peripheral Blood Lymphocytes

**PBMCs:** Peripheral Blood Mononuclear Cells

**PCC:** Pearson Correlation Coefficient

**PCNA:** Proliferation-Cell-Nuclear Antigen

**PD-1:** Programmed Death 1 (CD279)

**PD-L1:** Programmed Death-ligand 1 (CD274)

**PE:** Phycoerythrin

**PEI:** Polyethylenimine

**PGE2:** Prostaglandine E2

**PMA:** Phorbol-12-myristate-13-acetate

**PMS2:** PMS1 Homolog 2

**PRR:** Pattern Recognition Receptor

**PS:** Penicillin/Streptomycin

**PTC:** Premature Termination Codon

# R

**RFC:** Replication Factor C

**RNA:** Ribonucleic Acid

**RPA:** Replication Protein A

**RPMI:** Roswell Park Memorial Institute media

**RT:** Room Temperature

S

**SAE:** Serious Adverse Events

**SB:** Strong Binder

**SNV:** Single Nucleotide Variants

**SRA:** Sequence Read Archive

T

**TA:** Tubular Adenoma

**TAA:** Tumour-Associated Antigen

**TAP:** Transporter associated with Antigen Processing

**TCGA:** The Cancer Genome Atlas

**TCGA-BRCA:** The Cancer Genome Atlas - Breast Cancer

**TCGA-CESC:** The Cancer Genome Atlas - Cervix Uteri Cancer

**TCGA-COAD:** The Cancer Genome Atlas - Colorectal Adenocarcinoma

**TCGA-READ:** The Cancer Genome Atlas - Rectum neoplasms and adenomas

**TCGA-STAD:** The Cancer Genome Atlas - Stomach Cancer

**TCGA-UCEC:** The Cancer Genome Atlas - Corpus Uteri Cancer

**TCM:** T-Central Memory Cell

**TCR:** T cell receptors

**TEFF:** T-Effector Cell

**TEM:** T-Effector Memory Cell

**TEMRA:** Terminal effector memory T cells

**TGF- $\beta$ :** Transforming Growth Factor  $\beta$

**TILs:** Tumour infiltrating Lymphocytes

**TLR:** Toll-Like Receptor

**TMB:** Tumoural mutation burden

**TMC:** Tandeme Minigene Construct

**TME:** Tumour microenvironment

**TMG:** Tandem Minigene

**TNF:** Tumour Necrosis Factor

**TRL:** Tissue-Resident Lymphocytes

**TRM:** T-Resident Memory Cell

**TSCM:** T-Stem Cell Memory

**TTE:** T-Terminal Effector Cell

**TTM:** T-Transitional Memory Cell

**TVA:** Tubulovillous Adenoma

## V

**VAF:** Variant Allele Frequency

**VWB:** Very Weak Binder

## W

**WB:** Weak Binder

**WES:** Whole Exome Sequencing

**WGS:** Whole Genome Sequencing





## 1 INTRODUCTION

### 1.1 Lynch Syndrome

Lynch syndrome (LS) is the most common form of hereditary colorectal cancer (CRC) accounting for up to 1-3% of all cases<sup>1</sup> and up to 8% of all early onset CRC (EOCRC) cases<sup>2</sup>, with an estimated population-based prevalence of 1 in 279 individuals. It is an autosomal dominant disorder caused by a germline pathogenic mutation in one of the DNA mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*), by 3' deletions in the *EPCAM* gene which result in an epigenetic silencing of *MSH2*, or by a constitutional methylation of the *MLH1* promoter<sup>34</sup>. After a somatic inactivation of the previously wildtype MMR allele, the DNA repair machinery is impaired and leads to accumulation of errors during DNA replication, mainly insertions and deletions (InDels) in repetitive nucleotide sequences called microsatellites<sup>5</sup>. This predisposes LS carriers to the development of different tumours with MMR deficiency, starting at young ages, which are phenotypically characterized by microsatellite instability (MSI) and/or loss of MMR protein expression<sup>6</sup>. While previously known as Hereditary Non-Polyposis Colon Cancer (HNPCC), it should be noted that the terms LS and HNPCC are often used interchangeably. However, LS specifically refers to patients and family members with known germline variants<sup>7,8</sup>.

#### 1.1.1 Epidemiology and diagnosis

The prevalence of LS is commonly underrepresented due to the challenging diagnosis of LS carriers. This is due to the absence of pathognomonic clinical signs before developing a small lesion, adenocarcinomas or fully developed tumours, where the patient would be tested for MMR deficiency or MSI. In general, epidemiologic and cancer incidence data for LS has been obtained through retrospective studies. It is estimated that in the general population, 1 in 279 individuals is a carrier for an MMR gene pathogenic mutation (*MLH1*= 1 in 1946, *MSH2*= 1 in 2841, *MSH6*= 1 in 758, *PMS2*= 1 in 714)<sup>9</sup>.

In patients with newly diagnosed CRC, in a pooled-data analysis study of 10206 participants from 4 large cohorts between 1994 and 2010, 3.1% were MMR gene pathogenic variant carriers. Furthermore, among individuals with an MMR deficient

tumour, germline mutations in the MMR genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* were found in approximately 37%, 41%, 13%, and 9%, respectively<sup>6</sup>. In 2017, Dr. Pal Moller *et. al.* intended to establish a more robust estimate on cancer risk and the effects of surveillance in LS patients in a comprehensive prospective study<sup>10</sup>. On a cohort of 1942 mutation carriers without previous cancer and over a follow-up period ranging from 1 to 33 years, they concluded that the cumulative risk by age to first cancer varied amongst different MMR gene mutation carriers, with a 72% incidence from 25 to 70 years of age in *MLH1* and *MSH2* carriers and lower in *MSH6* (52%) and *PMS2* (18%) carriers. CRC was the most frequent first cancer observed (incidence from 25 to 70 years of 46%, 35%, 20% and 0% in *MLH1*, *MSH2*, *MSH6* and *PMS2*), followed by endometrial cancer (EC, incidence in females from 25 to 70 years of 34%, 51%, 49% and 24% in *MLH1*, *MSH2*, *MSH6* and *PMS2* carriers, respectively)<sup>10,11</sup>(Table 1). Of note, the estimated risks largely vary among studies due to their prospective/retrospective nature, ascertainment biases, population analysed, and statistical methods used.

*Table 1. 75-years cumulative cancer risk related with gene and sex in Lynch Syndrome individuals, adapted from<sup>11</sup>.*

| Cancer Type                                             | Incidence (population %) |        |             |        |             |        |             |        |
|---------------------------------------------------------|--------------------------|--------|-------------|--------|-------------|--------|-------------|--------|
|                                                         | <i>MLH1</i>              |        | <i>MSH2</i> |        | <i>MSH6</i> |        | <i>PMS2</i> |        |
|                                                         | Male                     | Female | Male        | Female | Male        | Female | Male        | Female |
| Brain Cancer                                            | 0.7                      | 1.6    | 7.7         | 2.9    | 1.8         | 1.2    | -           | -      |
| Stomach, duodenum, bile duct, gall bladder and pancreas | 21.8                     | 11     | 19.5        | 12.8   | 7.9         | 18.2   | 3.6         | 3.6    |
| Colorectal Cancer                                       | 57.1                     | 44.3   | 51.4        | 46.6   | 18.2        | 20.3   | 10.4        | 10.4   |
| Breast Cancer                                           | -                        | 12.3   | -           | 14.6   | -           | 13.7   | -           | 15.2   |
| Ovarian Cancer                                          | -                        | 11     | -           | 17.4   | -           | 10.8   | -           | 3      |
| Endometrial Cancer                                      | -                        | 37     | -           | 48.9   | -           | 41.1   | -           | 12.8   |
| Ureter/Kidney Cancer                                    | 4.9                      | 3.8    | 17.6        | 18.7   | 1.7         | 5.5    | 3.7         | 3.7    |
| Prostate Cancer                                         | 13.8                     | -      | 23.8        | -      | 8.6         | -      | 4.9         | -      |
| Urinary bladder Cancer                                  | 6.8                      | 5.4    | 12.8        | 7.9    | 8           | 1.2    | -           | -      |

Despite the fact that regular colonoscopy surveillance every 1-2 years from the ages of 20-25 to detect premalignant lesions or early cancers has shown to reduce CRC incidence and mortality in LS carriers<sup>12,13</sup>, CRC incidence remains an issue. Moreover, in patients that already had a CRC, up to 45% of patients will develop a second tumour in the colon, which sometimes requires a second surgery.

The presence of a germline pathogenic variant is required to establish a definitive diagnosis, which can occur only when a lesion is detected, or a family is studied due to a new diagnosis<sup>14,15</sup>.

Current diagnostic strategies include the study of families or individuals with familial aggregation and/or early onset of Lynch syndrome-associated tumours and universal MMR deficiency testing in all patients with CRC, which consists of tumour immunohistochemistry analysis (IHC) of MMR proteins and/or MSI testing<sup>2</sup>.

Given the high risk of metachronous CRC in the follow-up, total colectomy with ileorectal anastomosis (excision of the entire colon preserving the last 15-20 cm of rectum) is usually recommended in patients with CRC, especially at young ages<sup>15</sup>.

There is less evidence on the effectiveness of preventive strategies for other tumours in LS. Regarding endometrial and ovarian tumours, annual ultrasound and endometrial biopsies are recommended from the ages of 30-35 years until achieving gestational desire, followed by preventive resection of the uterus and ovaries (hysterectomy and double oophorectomy). Unfortunately, there are no proven preventive measures for other tumours<sup>16</sup>.

### 1.1.2 Surveillance

The National Comprehensive Cancer Network (NCCN) recommends a specific surveillance strategy for CRC prevention in individuals with LS. For *MLH1* and *MSH2* carriers, guidelines recommend colonoscopy every 1–2 years starting at ages 20–25. For carriers of *MSH6* and *PMS2*, the recommendation is to start the screening at ages 30-35 and to repeat it every 1-3 years<sup>17</sup>. The detection and resection of colorectal precursor lesions is the underlying mechanism through which colonoscopy prevents CRC. Colorectal adenomas (CrADs) are recognized as the main precursor lesions for CRC in LS, where surgery such as total colectomy may be required due to the accelerated progression of the adenoma-carcinoma sequence, often occurring in less than three years. However, in LS, recent studies have reported high rates of CRC during colonoscopy surveillance with cumulative incidences of up to 45% in *MLH1* carriers, 35% in *MSH2* and 20% in *MSH6* carriers at 70 years despite colonoscopy surveillance. These data, together with the high rate of second CRC occurrence, have put into question the preventive benefit of colonoscopy in LS.

A recent simulation study supports intensive surveillance for *MLH1* or *MSH2* carriers but suggests a later initiation for *MSH6* and *PMS2* carriers at ages 35 and 40, respectively, with screenings every three years instead of annually<sup>18</sup>. Despite these

recommendations, patient adherence to frequent colonoscopy remains low, often due to the burden and cost associated with regular screenings<sup>19,20</sup>.

For EC prevention in LS patients, surveillance includes annual transvaginal gynaecological ultrasound for endometrial and annexal evaluation, endometrial biopsy (hysteroscopy with targeted biopsy), and tumour marker evaluations (CA 125 and HE4)<sup>16</sup>. At approximately age 40, prophylactic surgery, such as hysterectomy and double adnexectomy, is recommended, particularly for carriers of *MLH1*, *MSH2*, and *MSH6*<sup>21</sup>.

While there are effective surveillance and prophylactic strategies for CRC and EC, as mentioned before, there are no proven effective prophylactic mechanisms for other tumour types associated with LS<sup>19</sup>.

### **1.1.3 Molecular Mechanisms**

Understanding the molecular mechanisms involved in the DNA repair pathway that are affected by LS is crucial to comprehend how this disease can behave. This introduction will focus on the molecular mechanisms of the MMR pathway and how mutations in its constituent genes can disrupt its function.

#### *1.1.3.1 DNA damage and repair systems*

In any eukaryotic cell, genome stability is key to support their survival and reproduction. Whenever any disruption on the genetic material happens, leading to genomic instability, biological processes and cellular functions can be hampered and it can even lead to cell death or carcinogenesis due to a cascade response that continuously triggers proto-oncogenes<sup>22-24</sup>.

There are several types of DNA damage that can occur through direct pathways, where there is a chemical bond breakage in the DNA molecule due to direct contact with endogenous (reactive species released by metabolism or after cell damage) or exogenous (toxic heavy metals, ionizing radiation...) materials, or indirect pathways, where these materials interact and activate different products that then can damage the DNA<sup>25,26</sup>. Specifically, DNA damage consequences can be single-strand breaks, double-strand breaks (DSBs), base damage, sugar damage, DNA cross-linking, clustered damaged sites, as well as DNA mismatches, which is not DNA damage per se as we will

comment afterwards. Out of these mechanisms, DSBs have the most severe consequences<sup>27–29</sup>.

As DNA molecules cannot be replaced after being damaged, they must instead be repaired to avoid consequences such as permanent genomic defects or instability, malignant transformation, further damage to the DNA repair ability or even cell death. For this reason, there are several repair mechanisms associated to each of the DNA damage types that can remediate their effects, summarized in Figure 1.

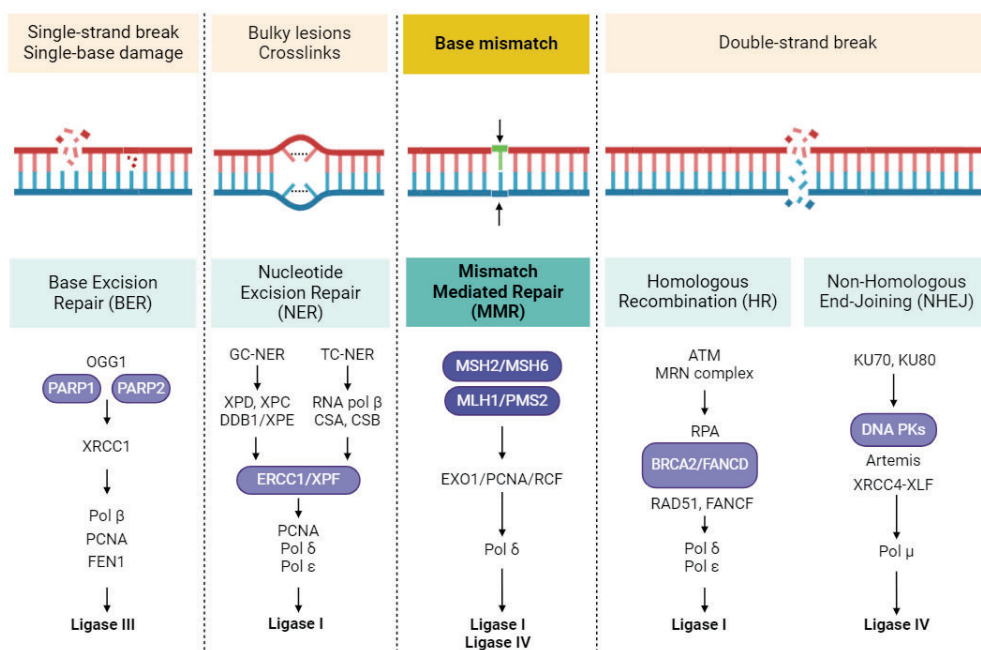


Figure 1. DNA damage types and repair mechanisms pathways associated. Made with Biorender.

Base excision repair (BER) addresses base damage caused primarily by ionizing radiation, replacing one to ten nucleotides using polymerase  $\beta$  and ligase III<sup>30,31</sup>. In contrast, nucleotide excision repair (NER) handles more complex lesions like bulky adducts and cross-links. NER operates through two pathways: transcription-coupled repair for actively transcribed genes and global genome repair for other regions. Once damage is recognized, the affected strand is excised and resynthesized by polymerases  $\delta$  and  $\epsilon$ , with ligase I sealing the strand<sup>32,33</sup>.

DSBs are repaired through either homologous recombination (HR) or non-homologous end-joining (NHEJ). HR involves recognition, strand annealing, and break-induced

replication, using the sister chromatid as a repair template for resynthesis by polymerases  $\delta$  and  $\epsilon$ , along with ligase I<sup>34,35</sup>. NHEJ, on the other hand, does not require a homologous template; instead, a protein complex (Ku70/Ku80 and DNA-PKcs) binds the broken ends, recruiting processing enzymes like Artemis and polymerase  $\mu$ . Ligase IV then joins the DNA ends, stabilized by XRCC4, to complete the repair<sup>36-38</sup>.

#### 1.1.3.2 Base mismatches and the MMR system

LS, as commented before, is caused by an inherited pathogenic variant in one of the MMR genes. The MMR system is one of the most important mechanisms to preserve genomic integrity and improve the fidelity of DNA replication by recognizing and excising spontaneous base-base mispairs and small InDel loops generated during DNA replication in the nascent DNA strand<sup>39</sup>.

Amongst the different errors in DNA replication, mismatches cannot be considered DNA damage as they exist only transiently while the DNA strand duplex remains annealed. It is important then to correct the error before the strands separate and give rise to 50% of DNA progeny containing a mutation. Replicative DNA polymerases incorporate errors in very low frequencies (approximately once per  $10^8$ – $10^{10}$  nucleotides polymerized)<sup>40</sup>, however there are repetitive nucleotide stretches (microsatellites) that are more prone to contain mismatches.

The MMR machinery is a highly conserved set of proteins called MSH and MLH/PMS that work in the heterodimers MutS $\alpha$  (*MSH2* and *MSH6*), MutS $\beta$  (*MSH2* and *MSH3*), MutL $\alpha$  (*MLH1* and *PMS2*) and MutL $\beta$  or  $\gamma$  (*MLH1* and *MLH3*). They have strong similarities to the prototypical *E. coli* MMR, including substrate specificity, bidirectionality and nick-directed strand specificity, but whereas MutS and MutL proteins in prokaryotes function as homo-oligomers, eukaryotic MMR proteins function as heterodimers<sup>41,42</sup>.

The recognition of mismatches is initiated by the MutS $\alpha$ /MutS $\beta$  complexes. The first have a preferential affinity to one base-pair insertion-deletion loops (i.e., mononucleotide mismatches), whereas MutS $\beta$  also recognizes larger base-pair insertion-deletion loops that can be up to  $\sim 13$  nucleotides<sup>43</sup>. Once the error is recognized, ATP-bound MutS $\alpha$  or  $\beta$  initiate the assembly of MutL dimers, which in turn manage the recruitment of additional proteins including proliferation-cell-nuclear antigen (PCNA), replication factor C (RFC), exonuclease 1 (Exo1), replication protein A

(RPA), and DNA ligase I<sup>44</sup>. Subsequently, RFC loads PCNA. The MLH proteins then connect the mismatch recognition signal to the distant DNA strand scission, where excision by Exo1 begins meanwhile RPA displaces the mismatch base and protects the DNA. Once done, there is a resynthesis of the excision gap by the replicative DNA polymerase  $\delta$  and ligases I and IV using the remaining DNA strand as a template<sup>45,46</sup>.

#### 1.1.3.3 MMR deficiency and microsatellite instability (MSI)

In LS, there is a loss of function, loss of expression or inactivation of the MMR system due to a heterozygous germline pathogenic variant, followed of a somatic inactivation of the wild-type allele.

Pathogenic variants in *MLH1* or *MSH2* result in the proteolytic degradation of both the primary and secondary partner proteins, inactivation of their corresponding complexes, *MLH1/PMS2* and *MSH2/MSH6*, resulting in the degradation of *PMS2* or *MSH6*. It has been seen that the loss of *MLH1* protein expression in tumours typically coincides with the loss of *PMS2* expression, while the loss of *MSH2* leads to the concurrent loss of *MSH6*. Conversely, mutations in the secondary genes *PMS2* or *MSH6* do not usually result in the loss of expression of the primary proteins *MLH1* or *MSH2*, likely due to interactions with other MMR components such as *MSH3*, *MLH3*, and *PMS1*<sup>47,48</sup>.

This loss of function causes an accumulation of mutations that cannot be repaired, primarily in repetitive DNA sequences called microsatellites<sup>39,49</sup>. Microsatellites may be mono-, di, tri-, tetra-, or penta-nucleotide DNA repeats that make up about 3% of the human genome, distributed non-randomly across coding and noncoding regions<sup>50,51</sup>. These sequences, by their repetitive nature, are prone to mismatch errors because of DNA polymerase slippage which primarily causes insertions and deletions to occur<sup>52</sup>. Because of this, MMR deficiency is often characterized by highly mutated microsatellites, a phenomenon that is named MSI. The terms MMRd (MMR deficient) and MSI-H (MSI-high) are often used interchangeably due to the high level of consistency between them, observed in about 90–95% of these tumours<sup>53</sup>.

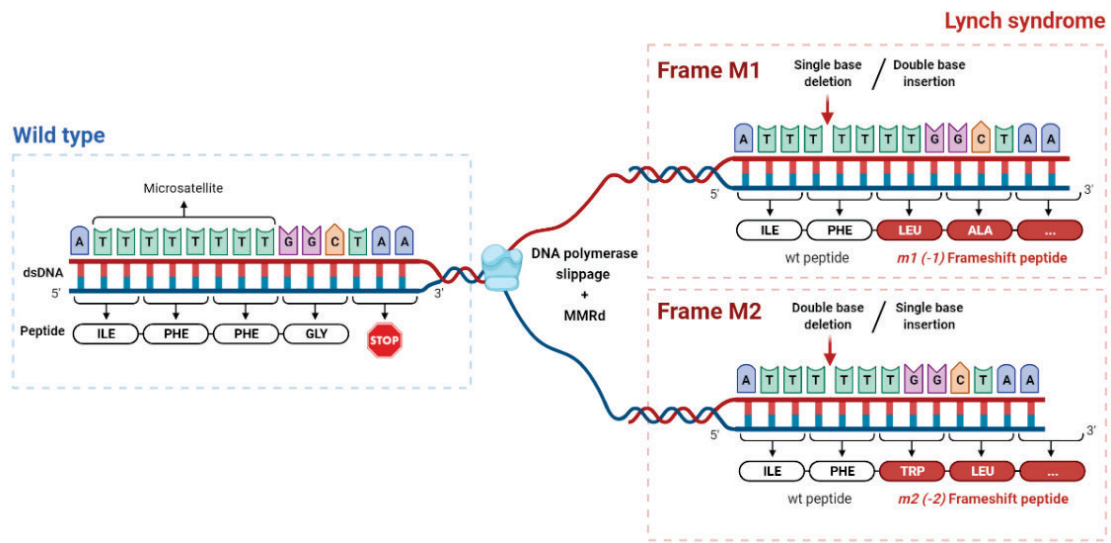


Figure 2. Schematic representation of the change on the grouping of codons or ORF due to MMRd in LS. Made with BioRender.

Most repetitive DNA sequences are located in non-coding regions. However, there is a limited number of genes that contain microsatellites in their coding sequence, where the unrepaired InDels can result in a frameshift - that is, a change on the grouping of codons or ORF (open reading frame) due to a modification of the DNA sequence length (Figure 2). These frameshift mutations frequently introduce a 'STOP' codon into the reading frame shortly after the mutation, likely rendering the protein inactive or preventing its expression entirely due to nonsense-mediated decay at the mRNA level<sup>54</sup>. However, some of these frameshift proteins end up being processed and presented by HLA molecules on the surface of malignant cells, where they will be able to be recognized by the immune system. These proteins are called neoantigens, and although they do not always derive from frameshift mutations (They may also derive from Single Nucleotide Variants (SNVs), Fusion Genes, or other non-genomic sources such as transcriptomic or proteomic variants, and even viral-derived neoantigens), frameshift-derived neoantigens (FSDN) are more frequently shared among MMRd cancers (e.g., endometrial, colorectal, and gastric) than SNV-derived neoantigens and others<sup>55-58</sup>.

The biomarkers of MSI-H, MMRd and tumoral mutation burden (TMB), which is defined as the total number of somatic missense mutations in the coding DNA sequence of the cancer cell genome, vary across different types of tumours, with some having more

prevalence of mutations than others. Table 2, adapted from Kang Y *et al.*<sup>59</sup>, summarizes the prevalence of MSI-H, TMB and MMRd across several tumour origins (Table 2, Figure 3).

Table 2. Prevalence of MSI, MMRd and TMB-H in different tumour origins. Adapted from<sup>59</sup>.

| Cancer type                       | Prevalence from random-effect model (95% CI) <sub>a</sub> |                                   |                                    |
|-----------------------------------|-----------------------------------------------------------|-----------------------------------|------------------------------------|
|                                   | MMRd                                                      | MSI                               | TMB-h                              |
| Gastrointestinal cancers          |                                                           |                                   |                                    |
| Anal cancer                       | –                                                         | 0.6% (0.1–1.4%), $I_2 = 0$        | –                                  |
| Appendiceal cancer                | 2.5% (0.0–7.4%)                                           | 2.1% (0.8–3.9%), $I_2 = 0.51$     | –                                  |
| Colon cancer <sup>c</sup>         | 8.9% (5.4–13.2%), $I_2 = 0.81^b$                          | 13.0% (9.5–17.0%)                 | –                                  |
| Colorectal cancer <sup>c</sup>    | 11.7% (9.3–14.4%), $I_2 = 0.98^b$                         | 10.2% (6.6–14.5%), $I_2 = 1^b$    | 8.5% (7.1–10.1%)                   |
| Esophageal cancer                 | 3.8% (1.1–7.8%), $I_2 = 0.77^b$                           | 2.4% (1.1–4.2%), $I_2 = 0.87^b$   | 32.9% (26.9–39.2%)                 |
| Gastric cancer                    | 8.7% (7.6–9.9%), $I_2 = 0.2$                              | 8.5% (6.4–10.9%), $I_2 = 0.95^b$  | 13.9% (10.9–17.1%)                 |
| Liver cancer                      | 0.0% (0.0–2.2%)                                           | 1.4% (0.4–2.9%), $I_2 = 0.23$     | 1.4% (0.2–3.6%)                    |
| Pancreatic cancer                 | 1.5% (0.6–2.7%), $I_2 = 0.15$                             | 0.9% (0.4–1.5%), $I_2 = 0.9^b$    | 0.0% (0.0–1.7%)                    |
| Small bowel cancer                | 21.0% (15.8–26.7%), $I_2 = 0.77^b$                        | 14.3% (5.4–26.3%), $I_2 = 0.95^b$ | 19.1% (11.8–27.8%)                 |
| Genitourinary tract cancers       |                                                           |                                   |                                    |
| Bladder/urothelial cancer         | 4.4% (1.6–8.3%), $I_2 = 0.88^b$                           | 2.9% (0.7–6.5%), $I_2 = 0.94^b$   | 38.1% (21.4–56.4%), $I_2 = 0.98^b$ |
| Kidney cancer                     | 0.8% (0.2–1.8%)                                           | 0.4% (0.2–0.6%), $I_2 = 0$        | –                                  |
| Penile cancer                     | 0.0% (0.0–2.4%)                                           | 0.0% (0.0–1.8%)                   | –                                  |
| Prostate cancer                   | 6.2% (0.7–16.2%), $I_2 = 0.97^b$                          | 2.3% (1.7–3.1%), $I_2 = 0.5$      | 4.0% (2.3–6.3%), $I_2 = 0.91^b$    |
| Breast and gynaecological cancers |                                                           |                                   |                                    |
| Breast cancer                     | 1.3% (0.7–2.0%), $I_2 = 0.71^b$                           | 0.6% (0.2–1.0%), $I_2 = 0.73^b$   | 7.2% (1.6–16.4%), $I_2 = 1^b$      |

|                                                                                                                                                                                                                                                                                                                                     |                                    |                                    |                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Cervical cancer                                                                                                                                                                                                                                                                                                                     | 1.9% (0.0–5.8%)                    | 1.5% (0.7–2.6%), $I_2 = 0.66$      | 23.7% (20.1–27.5%)                 |
| Endometrial cancer (EC)                                                                                                                                                                                                                                                                                                             | 26.8% (23.3–30.5%), $I_2 = 0.92^b$ | 21.9% (15.1–29.6%), $I_2 = 0.98^b$ | 43.0% (39.1–46.9%)                 |
| Ovarian cancer                                                                                                                                                                                                                                                                                                                      | 2.4% (0.5–5.5%), $I_2 = 0.92^b$    | 1.7% (0.0–5.4%), $I_2 = 0.99^b$    | –                                  |
| Uterine cancer (excl. EC)                                                                                                                                                                                                                                                                                                           | –                                  | 4.4% (3.2–5.7%)                    | –                                  |
| Thoracic cancers                                                                                                                                                                                                                                                                                                                    |                                    |                                    |                                    |
| Lung cancer                                                                                                                                                                                                                                                                                                                         | 1.6% (0.5–3.4%), $I_2 = 0.76^b$    | 0.4% (0.2–0.7%), $I_2 = 0.77^b$    | 27.5% (16.0–40.8%), $I_2 = 0.99^b$ |
| Thymic malignancy                                                                                                                                                                                                                                                                                                                   | 3.9% (0.1–11.5%)                   | 3.8% (0.1–11.2%)                   | –                                  |
| Bile duct/gallbladder                                                                                                                                                                                                                                                                                                               | 3.8% (1.5–7.0%), $I_2 = 0.66$      | 1.6% (1.0–2.3%), $I_2 = 0.69^b$    | 9.5% (3.3–18.2%)                   |
| Head and neck cancers, sarcomas, and skin cancers                                                                                                                                                                                                                                                                                   |                                    |                                    |                                    |
| Head and neck cancers                                                                                                                                                                                                                                                                                                               | 2.2% (0.1–6.1%), $I_2 = 0.86^b$    | 0.5% (0.3–0.7%), $I_2 = 0$         | 8.6% (1.8–19.4%), $I_2 = 0.95^b$   |
| Sarcomas                                                                                                                                                                                                                                                                                                                            | 0.5% (0.0–2.1%), $I_2 = 0.72$      | 1.8% (0.0–6.0%), $I_2 = 0.94^b$    | 1.7% (0.4–3.6%), $I_2 = 0.14$      |
| Skin cancers                                                                                                                                                                                                                                                                                                                        | 4.2% (0.3–11.9%), $I_2 = 0.94^b$   | 1.3% (0.0–4.1%), $I_2 = 0.92^b$    | 52.6% (49.7–55.5%)                 |
| Central nervous system tumours, endocrine tumours, neuroendocrine tumours, and other cancers                                                                                                                                                                                                                                        |                                    |                                    |                                    |
| Brain tumours                                                                                                                                                                                                                                                                                                                       | 3.8% (0.0–12.8%), $I_2 = 0.91^b$   | 0.6% (0.2–1.2%)                    | 2.8% (0.0–8.3%)                    |
| Endocrine tumours                                                                                                                                                                                                                                                                                                                   | 0.7% (0.1–1.7%)                    | 0.6% (0.0–2.2%), $I_2 = 0.53$      | 5.8% (3.8–8.2%)                    |
| Neuroendocrine tumours                                                                                                                                                                                                                                                                                                              | 0.0% (0.0–1.2%)                    | 1.4% (0.4–2.9%), $I_2 = 0.79^b$    | 24.8% (22.9–26.7%)                 |
| <sup>a</sup> Heterogeneity across studies was presented based on the point estimate of $I_2$ score and heterogeneity test, if $\geq 2$ records were available for each cancer type and stage group.                                                                                                                                 |                                    |                                    |                                    |
| <sup>b</sup> Statistically significant heterogeneity across studies ( $p < 0.05$ ).                                                                                                                                                                                                                                                 |                                    |                                    |                                    |
| <sup>c</sup> Pooled prevalence of pan-tumour biomarkers separately based on estimates including (1) colon cancer only, (2) rectal cancer only, and (3) tumours at any sites in the colon and the rectum, obtaining pooled prevalence estimates described as “colon cancer”, “rectal cancer”, and “colorectal cancer”, respectively. |                                    |                                    |                                    |

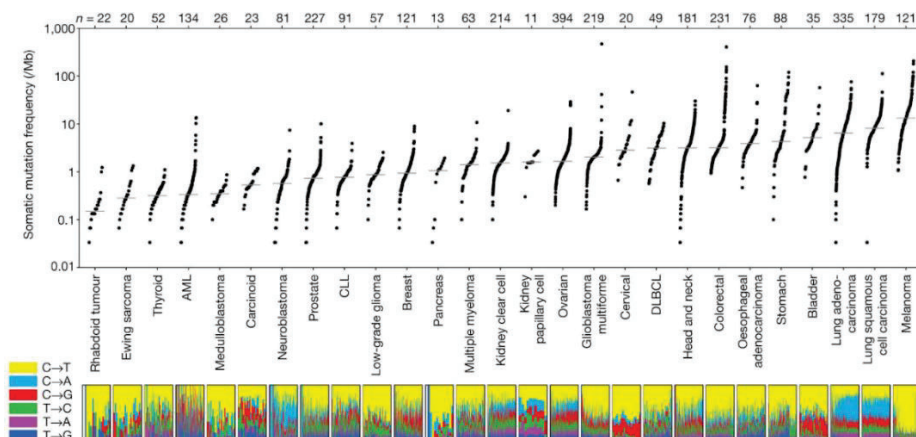


Figure 3. Somatic mutation frequencies observed from 3,083 tumour-normal pairs. Each dot represents a tumour-normal pair, y-axis indicates the total frequency of somatic mutations in the exome. Tumour types are arranged by their median somatic mutation frequency. The bottom panel illustrates the relative proportions of the six different possible base-pair substitutions, as indicated by the legend on the left. Adapted from<sup>60</sup>.

#### 1.1.4 Characteristics of tumours associated to LS

As mentioned, in LS most CRCs exhibit MSI and tend to develop proximal to the splenic flexure. These cancers often present as synchronous or metachronous tumours, with initial diagnoses typically occurring in individuals in their mid-40s<sup>61</sup>. The disease phenotype is highly variable and heavily influenced by the specific pathogenic variant present. Individuals with germline pathogenic variants in DNA MMR genes are not only predisposed to developing CRC at younger ages but also face elevated risks of extracolonic malignancies. These include cancers of the endometrium, stomach, ovaries, biliary tract, and urinary tract. Less commonly, LS has been associated with small intestinal and brain tumours, although these are substantially rarer (Table 1)<sup>11</sup>.

LS tumours are characterized by a heightened state of immune activation in the tumour microenvironment (TME), an accelerated rate of malignant transformation from adenoma to carcinoma (1-3 years, compared to a normal rate of 10-15 years), MSI and poor differentiation. Concretely, LS-associated CRCs display distinct histopathological features. They are commonly diploid, poorly differentiated, and exhibit a high density of tumour-infiltrating lymphocytes (TILs). Additional features include a medullary growth

pattern, mucinous or signet ring cell differentiation, and a characteristic "Crohn's-like reaction"<sup>62</sup>.

### **1.1.5 Challenges in LS management and psychological impact**

LS presents several challenges in its management, ranging from diagnosis to long-term patient outcomes.

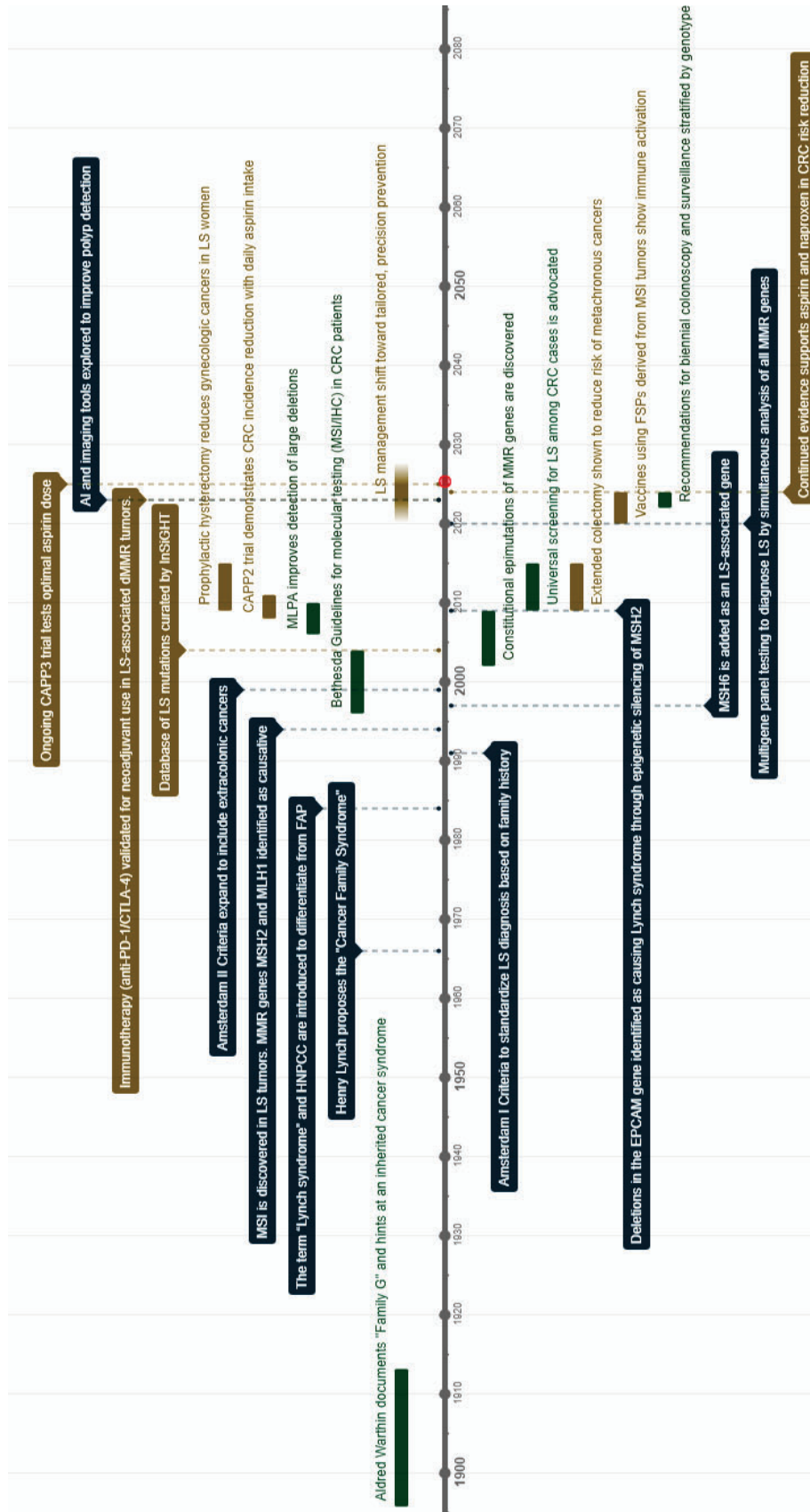
The identification of LS carriers involves genetic testing to detect pathogenic variants in MMR genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*). The identification of variants of uncertain significance complicates the diagnostic process, as they require extensive research to determine their pathogenicity, often leading to inconclusive results that hinder clinical decision-making<sup>63</sup>. This also leads to underdiagnosis or delayed diagnosis, especially in individuals without a strong family history of cancer which is very prominent in LS, both for lack of awareness and because the symptoms of LS-related cancers often overlap with those of sporadic cancers, making it difficult to distinguish based solely on clinical presentation<sup>18,64</sup>.

Added to this, the intensive surveillance protocols recommended for LS patients can be burdensome. For instance, frequent colonoscopies (every 1-2 years) are recommended<sup>11,18</sup>, but adherence to these schedules is low due to the discomfort and inconvenience associated with the procedure. The psychological burden of the threat of cancer development and of the constant surveillance, including regular medical procedures, significantly impacts the quality of life for patients and carriers, especially of young age, as they often experience significant anxiety, depression, and other mental health issues<sup>65,66</sup>.

### **1.1.6 Prevention and prophylaxis for LS-related cancers: History and current methods**

#### *1.1.6.1 Timeline of LS cancer prevention*

The timeline of LS cancer prevention (Figure 4) highlights over a century of progress, from the earliest familial observations by Warthin in 1913 to the current cutting-edge immunopreventive in the 2020s.



*Figure 4. Timeline of Key Milestones in Lynch Syndrome Cancer Prevention (1913–2025). This timeline illustrates the major scientific and clinical developments in the understanding, diagnosis, and prevention of LS over more than a century. Boxes with text inside indicate events (blue), Orange boxes indicate events related to LS cancer prevention advances. Boxes with text outside indicate a time period (green), Orange boxes indicate events related to LS cancer prevention advances.*

Key historical milestones include the clinical definition of "Cancer Family Syndrome" by Henry Lynch in 1966 and the genetic discovery of (MMR gene mutations in the 1990s, which established the molecular foundation for LS diagnosis<sup>67</sup>.

Preventive strategies evolved significantly with the introduction of genotype-based surveillance, colonoscopy at earlier ages, and chemoprevention using aspirin, notably validated by the CAPP2 trial<sup>68</sup>, which will be discussed below. More recently, efforts have shifted toward precision prevention, including the use of NSAIDs like naproxen and novel cancer vaccines, which have demonstrated immunogenicity and safety in early clinical trials<sup>69</sup>.

#### *1.1.6.2 Current cancer prevention strategies*

As mentioned previously, although current prevention strategies involving surveillance for CRC and (in a lesser way) EC are effective, many cancers still escape these methods<sup>70</sup>. Also, surveillance strategies for other LS-related tumours do not have enough beneficial evidence to be established. For example, the risk of developing central nervous system cancer throughout life ranges from 0.8 to 18.6% and there is currently no evidence to recommend systematic follow-up<sup>11</sup>.

However, there are other more general prophylactic measures with varying efficacies. Aspirin (ASA, acetylsalicylic acid) and naproxen have shown promise as chemoprophylactic agents for CRC, with studies like CAPP2 demonstrating its efficacy in 2011 after a 2-year vigilance, though optimal dosing is under investigation<sup>71</sup>. On another phase I clinical trial (NCT02052908), they explored the effects of naproxen, ASA or COX inhibitors in primary and secondary prevention for LS carriers<sup>72</sup>. Results showed that the levels of prostaglandin E2 (PGE2) in the colorectal mucosa was significantly decreased after treatment with naproxen, and in a later study evaluating the precise immune cell types enriched by naproxen, they showed that proliferating (Ki67<sup>+</sup>), activated (OX40<sup>+</sup>/GZMB<sup>+</sup>/CD137<sup>+</sup>), memory (CD45RO/CCR6<sup>+</sup>) CD8a<sup>+</sup> T-cells in the

colonic mucosa become activated upon naproxen treatment, which is likely maintained by reduction of PGE2 levels<sup>73</sup>.

Hormonal contraceptives, particularly progestogen-only options, may reduce gynaecologic cancer risks in pre-menopausal women before prophylactic surgeries. Immune-based therapies, including vaccines targeting MSI-induced FSDNs, represent a promising preventive approach, as LS-associated immune mechanisms play a critical role in carcinogenesis. Personalized care, considering gene-specific risks and prophylactic measures, remains central to LS management, with evolving strategies for genetic counselling, surveillance, and lifestyle modifications to mitigate cancer risks<sup>11</sup>.

#### *1.1.6.3 LS MSI tumour treatment strategies*

Management of MSI-H cancers in LS has increasingly shifted toward immunotherapies involving immune-checkpoint inhibitors (ICIs, see section 1.1.6.5.1) monoclonal antibodies over traditional chemotherapies, though treatment options, particularly for advanced-stage cancers, remain challenging. Systemic anti-cancer treatment options used in sporadic CRCs (fluorouracil, leucovorin, oxaliplatin and irinotecan) have conflicting efficacy results in MSI-H CRCs, with most reports not being specific to LS. For example, standard 5-fluorouracil-based regimens are ineffective in MMRd tumours due to their inability to recognize DNA damage<sup>74</sup>. The high TMB and robust immune infiltration characteristic of MMRd tumours make them optimal for ICI treatment (see sections 1.1.6.4 onward), although while they have proven effective in MSI cancer treatment, studies indicate they do not eliminate the risk of new neoplasia, needing continued LS-specific surveillance strategies<sup>75</sup>.

Advanced cancers often require interventions like surgical total colectomy, which significantly impact patients' quality of life, causing altered bowel habits and other complications. Adverse effects from chemotherapy and radiotherapy further challenge patient well-being. For CRC, surgery suffices for LS patients with stage II CRC, while the benefit of adjuvant chemotherapy in stage III remains uncertain<sup>11</sup>.

Emerging therapeutic strategies focus on synthetic lethal interactions, such as targeting Werner helicase (which helps unwind DNA during replication and repair) in *MSH2* or *PMS1*-mutated MMRd cancers, though clinical application is pending<sup>76</sup>.

#### 1.1.6.4 Introduction to immunotherapy

Immunotherapy is defined as *“The prevention or treatment of disease with substances that stimulate the immune response”*. Monoclonal antibodies stand out as the cornerstone of modern cancer therapy, widely employed in clinical settings for their ability to target specific molecular markers on cancer cells. They operate through various mechanisms, including blocking or stimulating cell signaling pathways crucial for tumour growth. Additionally, they can tag cancer cells for destruction by the immune system or deliver cytotoxic payloads directly to tumours.

Cancer vaccines play a dual role: preventive and therapeutic. Preventive vaccines aim to train the immune system to recognize and remember cancer-specific antigens, thereby preventing the development of cancer. On the other hand, therapeutic vaccines focus on stimulating an immune response against existing cancer cells, requiring precise identification of antigens unique to cancer cells, such as oncogenic viral peptides.

Adoptive cell therapy (ACT) represents a personalized approach where immune cells are harvested from patients, modified or activated *ex vivo*, and then reintroduced into the patient to enhance anti-tumour immunity. Techniques include transferring TILs, engineering T cells to express chimeric antigen receptors (CAR-T cells), modifying T cell receptors (TCRs), and employing natural killer (NK) cells to target cancer cells.

ICIs (see section 1.1.6.5.1) have revolutionized cancer treatment by targeting regulatory pathways that suppress immune responses against cancer. Key molecules like CTLA-4 and PD-1 are critical immune response inhibitory checkpoints that, when blocked with ICIs, re-activate the immune system's ability to attack cancerous cells. Since their initial approval, ICIs have shown remarkable efficacy across various cancers, particularly in melanoma and lung cancer, leading to significant improvements in patient outcomes.

The discovery and clinical application of ICIs, pioneered by researchers like James Allison and Tasuku Honjo, have not only transformed treatment strategies but also garnered international recognition, including the Nobel Prize in Medicine in 2018. Ongoing research continues to explore novel combinations of ICIs with other therapies and biomarkers to optimize treatment responses and broaden their applicability across different cancer types.

### 1.1.6.5 Immunotherapy strategies for LS (ICI, cancer vaccines)

#### 1.1.6.5.1 Immune checkpoint inhibitors (ICIs)

As mentioned in the previous section, ICIs are a class of drugs termed monoclonal antibodies designed to help the immune system recognize and attack cancer cells. Normally, immune checkpoints are regulatory pathways in the immune system that help maintain self-tolerance and modulate the duration and amplitude of immune responses in peripheral tissues, thereby preventing autoimmunity. Cancer cells can exploit these checkpoints to evade immune detection and destruction. ICIs work by blocking these immune checkpoints, thus activating exhausted immune cells.

Due to a high TMB and an associated immune infiltration in the TME, MMRd tumours are generally more responsive to ICIs than MMR proficient tumours<sup>77</sup>. In 2017 the FDA granted accelerated approval for the use of the anti-PD-1 drug pembrolizumab for adult and paediatric patients with unresectable or metastatic MMRd/MSI-H solid tumours that have progressed after previous treatment, independently of tumour type, marking the first FDA approval that does not account for cancer site<sup>78</sup>. Research indicates that poorly immunogenic tumours are partially sensitive to ICI therapy through the expansion of anti-tumour CD8+ T cells. Treatment with an oncolytic virus encoding multiple tumour-associated antigens (TAA) within the virus has shown to enhance the immune response by promoting anti-TAA CD8+ T cells, resulting in a synergistic effect between ICI therapy and oncolytic virotherapy<sup>79</sup>. This remarks the importance of a strong immune infiltrate within the tumour site for ICIs to be effective, which is why generally MMRd tumours that have a high TMB and thus neoantigens that increase the number of specific T-cells, are more responsive. Several clinical trials have evidenced their beneficial role in LS-related cancer treatment, and some are exploring their effectiveness on cancer prevention (Table 3).

#### 1.1.6.5.2 Cancer vaccines

Cancer vaccines represent a promising approach in immunotherapy for LS, taking advantage of the unique features of MMRd tumours. As previously mentioned, as these tumours accumulate frameshift mutations, they produce immunogenic neoantigens that can serve as vaccine targets. Shared frameshift mutations across different patients and tumour types have shown potential in enhancing immune surveillance and delaying

tumour progression in at-risk individuals<sup>80</sup>, thus aiding in both therapeutic and prevention settings. Research demonstrates that these vaccines can stimulate immune surveillance and cytotoxic T cell responses. Early preclinical studies and clinical trials have yielded encouraging results, highlighting their potential in reducing tumour burden<sup>81–83</sup>(Table 3). Despite challenges like tumour resistance mechanisms, including MHC-I (Major Histocompatibility complex class I) loss or immune evasion pathways, ongoing investigations aim to refine vaccine platforms and assess their long-term efficacy in LS populations.

*Table 3. List of registered clinical trials of cancer vaccines and/or CPI in MMRd/Lynch syndrome patients. Adapted from<sup>84</sup>.*

| Study ID                                                           | Description                                                                                                                     | Status                 | Location |
|--------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|------------------------|----------|
| <b>ICI and Other Immunotherapy Clinical Trials in MMRd Cancers</b> |                                                                                                                                 |                        |          |
| NCT04612309                                                        | Retrospective Study on the Use of Immunotherapy in Patients With MSI-H Metastatic Colorectal Cancer                             | Recruiting             | Italy    |
| NCT04795661                                                        | Immunotherapy in MSI/MMRd Tumours in Perioperative Setting                                                                      | Not yet recruiting     | France   |
| NCT03206073                                                        | A Phase I/II Study of Pexa-Vec Oncolytic Virus in Combination with Immune Checkpoint Inhibition in Refractory Colorectal Cancer | Completed              | USA      |
| NCT04019964                                                        | Nivolumab in Biochemically Recurrent MMRd Prostate Cancer                                                                       | Recruiting             | USA      |
| NCT02052908                                                        | Naproxen in Preventing DNA Mismatch Repair Deficient Colorectal Cancer in Patients with Lynch Syndrome                          | Completed              | USA      |
| NCT05411718                                                        | A Phase IIa Randomized, Double-Blinded Clinical Trial of Naproxen or Aspirin for Cancer Immune Interception in Lynch Syndrome   | Recruiting             | USA      |
| NCT04711434                                                        | PD-1 Antibody for The Prevention of Adenomatous Polyps and Second Primary Tumours in Lynch Syndrome Patients                    | Recruiting             | China    |
| NCT02912559                                                        | Combination Chemotherapy with or Without Atezolizumab in Treating Patients With Stage                                           | Active, not recruiting | USA      |

|                                                                |                                                                                                                                                                                                        |                        |             |
|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|-------------|
|                                                                | III Colon Cancer and Deficient DNA Mismatch Repair                                                                                                                                                     |                        |             |
| NCT02359565                                                    | Pembrolizumab in Treating Younger Patients with Recurrent, Progressive, or Refractory High-Grade Gliomas, Diffuse Intrinsic Pontine Gliomas, Hypermutated Brain Tumours, Ependymoma or Medulloblastoma | Active, not recruiting | USA         |
| <b>Clinical Trials Involving Off-the-Shelf Cancer Vaccines</b> |                                                                                                                                                                                                        |                        |             |
| NCT04117087                                                    | Pooled Mutant KRAS-Targeted Long Peptide Vaccine Combined with Nivolumab and Ipilimumab for Patients with Resected MMR-p Colorectal and Pancreatic Cancer                                              | Recruiting             | USA         |
| NCT01885702                                                    | Dendritic Cell Vaccination in Patients with Lynch Syndrome or Colorectal Cancer With MSI                                                                                                               | Active, not recruiting | Netherlands |
| NCT03152565                                                    | Avelumab Plus Autologous Dendritic Cell Vaccine in Pre-treated Metastatic Colorectal Cancer Patients                                                                                                   | Completed              | Spain       |
| NCT04041310                                                    | Nous-209 Genetic Vaccine for the Treatment of Microsatellite Unstable Solid Tumours                                                                                                                    | Active, not recruiting | USA         |
| NCT01461148                                                    | A Frameshift Peptide Neoantigen-Based Vaccine for Mismatch Repair-Deficient Cancers                                                                                                                    | Completed              | Germany     |
| NCT05419011                                                    | Tri-Ad5 Vaccine and N-803 in Preventing Cancer in Patients with Lynch Syndrome                                                                                                                         | Recruiting             | USA         |
| NCT05078866                                                    | Cancer Preventive Vaccine Nous-209 for Lynch Syndrome Patients                                                                                                                                         | Active, not recruiting | USA         |

## 1.2 Immunology: The adaptive cellular response

The immune system is a complex network of cells, tissues, and organs that work together to defend the body against infections and diseases through coordinated innate and adaptive responses. The innate immune response serves as the body's first line of defence, providing a rapid, nonspecific reaction to pathogens. It relies on physical barriers, such as the skin and mucosa, and cellular components, including macrophages, neutrophils, dendritic cells (DCs), and NK cells. These cells recognize pathogen-associated molecular patterns (PAMPs), which are conserved molecular structures found on pathogens and DAMPs (danger-associated molecular patterns), which are endogenous molecules released by damaged or stressed cells, through pattern recognition receptors (PRRs), triggering phagocytosis, inflammation, and cytokine release for a rapid mediation of the damage (Innate immunity) and an activation of an antigen-specific response that will also confer immunological memory, termed adaptive response.

Amongst the cells that intervene in the adaptive response we find lymphocytes, which can be divided into B lymphocytes (B cells), responsible for the humoral or antibody-mediated response, and T lymphocytes (T cells), responsible for the cellular response. Both types of lymphocytes have specific antigen receptors on their membranes, known as B cell receptors (BCRs) and T cell receptors (TCRs). BCRs are membrane-bound immunoglobulins (Ig) capable of recognizing native antigens by themselves, while TCRs, also part of the Ig protein family, recognize processed antigens presented by MHC molecules. Thus, TCRs require antigen-presenting cells (APCs) for activation. Among T cells, there are several subpopulations, with the two main groups being CD4<sup>+</sup> T helper cells (Th) and CD8<sup>+</sup> cytotoxic T lymphocytes (CTLs). On the other hand, NKT cells bridge innate and adaptive immunity by recognizing lipid antigens presented by CD1d, glycolipids, phospholipids, and sphingolipids derived from microbial pathogens, self-antigens, and tumour-associated lipids, among others. Both NK and NKT cells contribute to the "missing self" immune response by detecting cells that lack self-MHC class I molecules, a hallmark of many infected or malignant cells, leading to targeted cytotoxicity.

As mentioned, APCs are essential for initiating and regulating the adaptive immune response by processing and presenting antigens to T cells, with DCs being the most

potent and specialized APCs. DCs capture antigens, migrate to lymphoid tissues, and provide the necessary co-stimulatory signals to activate naive T cells, thus bridging innate and adaptive immunity. Activated T-cells then proliferate and differentiate into effector T cells that eliminate infected or abnormal cells, or into memory T cells that provide long-term immunity. These concepts will be further introduced in the following sections.

### 1.2.1 Dendritic cells

DCs are a specialized subset of lympho-myeloid bone marrow-derived cells first discovered by Ralph Steinman in the 1970s that play a crucial role in initiating and directing immune responses, thus bridging innate and adaptive immunity<sup>85</sup>. DCs originate from Hematopoietic Stem Cells (HSCs) in the bone marrow but can differentiate from both myeloid and lymphoid precursor cells. They are primarily found in tissues exposed to the external environment, such as the skin and mucosal surfaces, as well as in tissue stroma, while their presence in peripheral blood is low, and they are typically found in an immature state.

Classified as innate immune cells and APCs, DCs respond to PAMPs and DAMPs via specific PRRs, internalizing, processing, and presenting these antigens bound to MHC class I or II molecules to prime naive T-cells<sup>86</sup>. In addition to this, they can stimulate other innate immune cells, such as macrophages, NK cells, and NKTs (Natural Killer-T cells).

Under various micro-environmental conditions, immature DCs can mature and induce pro-inflammatory or anti-inflammatory responses, establishing immunogenicity or tolerance to specific antigens. This maturation process is influenced by multiple factors, including the DC subtype, the type of pathogen-related molecules, the local microenvironment, and T-cell signals, and is characterized by changes such as loss of phagocytic capacity due to receptor down-regulation, morphological alterations, distinct cytokine production patterns and up-regulation of co-stimulatory molecules, such as CD80, CD86, and CD40, which are critical for T-cell activation and the subsequent immune response<sup>87</sup>.

As pivotal regulators of immunity, DCs have become a focal point for immunotherapy strategies in cancer patients. These strategies aim to induce antigen-specific effector and

memory cells either in vivo, through DC vaccines, or in vitro, by expanding DCs with tumour antigens and reinfusing them into patients. Such approaches have shown promise in enhancing the efficacy of cancer immunotherapies, including checkpoint inhibitors and adoptive T-cell transfer, by improving the immune system's ability to recognize and attack cancer cells<sup>88</sup>.

#### *1.2.1.1 Dendritic cell Subtypes*

DCs are found in various tissues and are particularly abundant in lymphoid organs, skin, and the digestive and respiratory tracts. Most DCs derive from adult myeloid precursors, except for Langerhans cells in the skin, which originate from embryonic precursors.

DCs can be classified into different subtypes based on their phenotype, transcriptional profile, ontogeny, and functions (Figure 5). Initially, DCs were categorized into conventional DCs (cDCs) and plasmacytoid DCs (pDCs) primarily based on their ability to present antigens to T cells<sup>89</sup>. Recent advancements in understanding the transcription factors involved in DC differentiation, particularly in mice, have further refined these distinctions. Broadly, DCs are classified into three main subsets: Type 1 and Type 2 cDCs and pDCs, aside from monocyte-derived dendritic cells (moDCs)<sup>90</sup>.

In humans, DC development commences in the bone marrow from HSCs. Granulocyte, monocyte, and DC progenitors (GMDPs) give rise to monocyte and DC progenitors (MDPs), which can differentiate into all DC subsets. Upon becoming common DC progenitors (CDPs), they lose the ability to differentiate into monocytes and can further develop into either pDCs or a circulating pre-cDC progenitor<sup>91</sup> capable of maturing into either Type I or Type II cDCs. Additionally, monocytes can later derive into moDCs, representing a heterogeneous group of cells<sup>92</sup>. All these subsets will be discussed in the following sections.

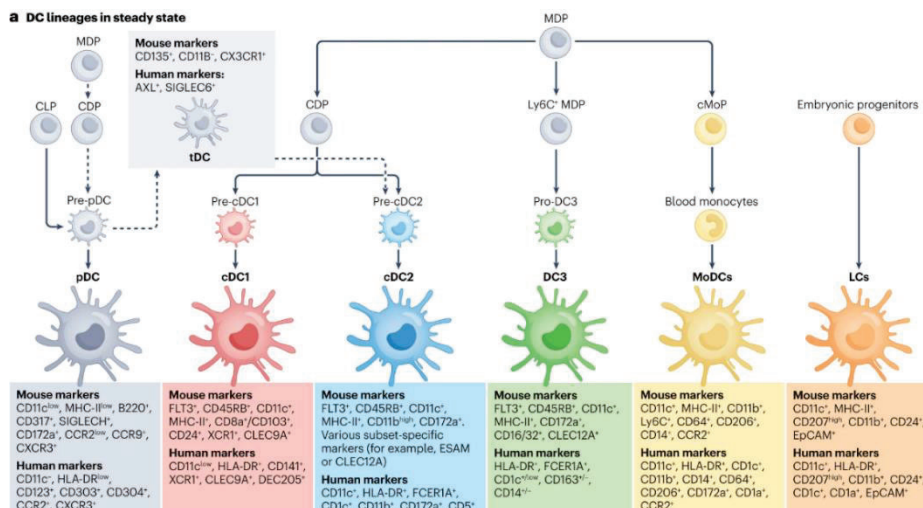


Figure 5. Heterogeneity of dendritic cell subsets. Adapted from<sup>93</sup>.

#### 1.2.1.1.1 Conventional Dendritic Cells

cDCs include all DCs except pDCs and function as key APCs within the immune system. They originate from common DC progenitors and differentiate into cDC1 and cDC2 subsets, which localize to different regions of lymphoid and non-lymphoid tissues based on their chemokine receptor expression<sup>94</sup>. cDC1s are rare but highly specialized for cross-presenting antigens to CD8<sup>+</sup> T cells and inducing Th1 responses<sup>95</sup>, playing a crucial role in anti-viral and anti-tumour immunity<sup>96</sup>. They are commonly identified by the CD141 (BDCA-3) marker in humans, are under the transcriptional control of *IRF8*, *ID2* and *BATF3*, and express the surface markers XCR1, CLEC9A, CADM1, BTLA, and CD26<sup>86,89,92</sup> (Figure 5). They are also marked by CD103 and TLR3 expression, enabling efficient uptake of cell-associated antigens and dead cells.

In contrast, cDC2s are the most abundant DC subset and are primarily responsible for polarizing CD4<sup>+</sup> T helper cells into Th2, Th17, and T-follicular helper (Tfh) subsets, depending on the specific immune environment. They are driven by the transcription factors *IRF4*, *ID2*, *ZEB2* and *NOTCH2/KLF4*, and are identified by the expression of

BDCA1 (CD1c), SIRP- $\alpha$  (CD172a), and CLEC10A (CD301)<sup>97,98</sup>

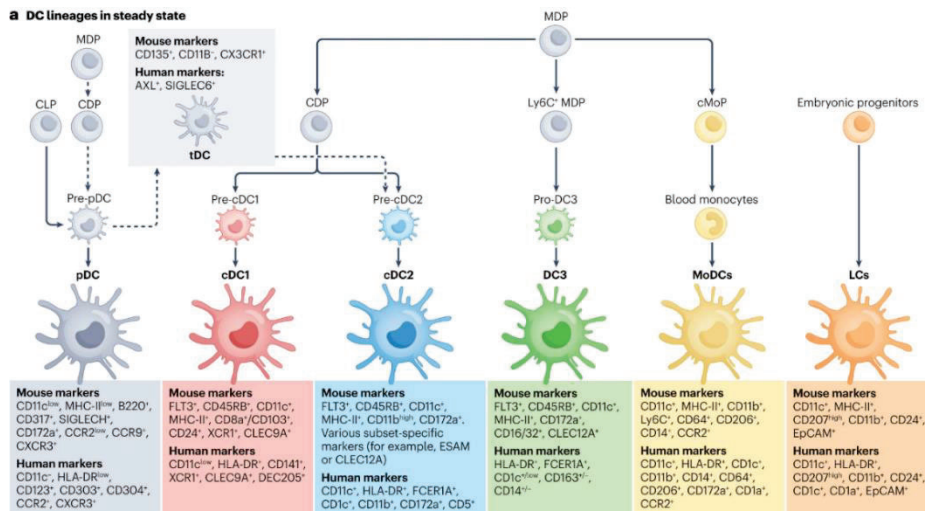


Figure 5). They are particularly effective at antigen uptake in the skin and mucosal tissues and are also involved in regulatory T cell induction<sup>96,98</sup>, especially in the intestine and liver<sup>99,100</sup>.

A newly identified subset, cDC3<sup>101</sup>, serves as circulating precursors of inflammatory DCs, developing under GM-CSF influence. These cells exhibit an inflammatory gene signature and have been implicated in tumour-associated immune responses, particularly in breast and ovarian cancer, suggesting a role in shaping immune dynamics within cancerous tissues<sup>102</sup>.

#### 1.2.1.1.2 Plasmacytoid Dendritic cells

pDCs were initially believed to have a myeloid origin<sup>103</sup> until studies suggested that lymphoid progenitors could also give rise to these cells, creating a debate about their ontogeny<sup>104</sup> and whether they develop predominantly from lymphoid progenitors, or if pDCs and cDCs share a common pool of Cx3cr1<sup>+</sup> hematopoietic progenitors<sup>105</sup>. Several features align pDCs more closely with cDCs than lymphocytes, such as their development being driven by *Flt3L* and independent of IL-2R $\gamma$  signaling. Additionally, the discovery of transitional DCs (tDCs), which share characteristics with both pDCs and cDC2s, adds complexity to understanding pDC development<sup>106</sup>.

pDCs exhibit a plasma cell-like morphology, featuring eccentric nuclei and a well-developed endoplasmic reticulum. They express specific markers such as CD123 and

CD45RA, which are retained from their GMDP progenitors<sup>107</sup>. As for their function, they are known for their ability to rapidly produce large amounts of IFN $\alpha$  (and other interferons such as IFN-I and IFN-III, as well as IL-6 and TNF- $\alpha$ ) as an antiviral response. This is activated through their TLR 7 and 9 receptors, which recognize single-stranded RNA and unmethylated CpG motif containing DNA, respectively, that upon viral recognition and signaling through the adaptor MyD88, up-regulate the interferon regulatory factor-7 (*IRF7*)<sup>106</sup>. pDCs can efficiently present endogenous antigens via MHC-I and MHC-II, but their ability to capture exogenous antigens is unclear.

#### 1.2.1.1.3 Monocyte-derived Dendritic cells

moDCs are associated with an inflammatory environment and share phenotypic and functional features with monocytes and cDC2s, relying on *IRF4* for differentiation and capable of cross-priming CD8+ T cells. They differentiate mostly within tissues from infiltrating monocytes in response to various stimuli, including cancer, and while generally less immunostimulatory than cDCs, can induce strong antigen-specific T cell responses<sup>93,108-110</sup>.

Advances in human DC biology have been facilitated by methods to derive moDC in vitro from CD34+ precursors using GM-CSF and TNF- $\alpha$ , or from monocytes using GM-CSF and IL-4. In mice, bone marrow precursors stimulated with GM-CSF generate cells resembling tissue DCs, known as bone marrow-derived dendritic cells (BMDC), which are heterogeneous, with GM-DCs resembling DCs and GM-Macs being more macrophage-like. Adding IL-4 reduces but does not eliminate GM-Macs<sup>111</sup>.

Functionally, they are similar to inflammatory DCs. They express CD1c, CD1a and high levels of MHC-II and co-stimulatory molecules, which increase after maturation stimuli that usually consists of a mix of pro-inflammatory cytokines and a TLR ligand. On the other hand, and illustrating their high plasticity, exposure of in vitro-derived moDCs to agents like vitamin D3, corticosteroids, or rapamycin induces a stable semi-mature phenotype. This is marked by a low MHC II and co-stimulatory molecule expression but high levels of indoleamine 2,3-dioxygenase, IL-10, TRAIL, and PDL-1. These cells are tolerogenic, promoting regulatory T cells and reducing adaptive immune responses<sup>112-</sup>

<sup>114</sup>.

#### 1.2.1.2 *Maturation and antigen processing*

DCs are naturally in an immature state, with a low surface expression of co-stimulatory molecules and chemokine receptors, also with decreased ability to produce pro-inflammatory cytokines and thus to prime naïve T-cells. This immature phenotype serves as a sentinel, with an increased capacity of antigen capture capacity via macropinocytosis and endocytosis, because of a high expression of receptors (including, but not limited to, Lectin receptors, TLRs, immunoglobulin constant fractions (Fc) and complement receptors)<sup>115,116</sup>. At barrier sites, they can sample antigens by projecting their dendrites through epithelial tight junctions. Once they capture an antigen (including PAMPs or DAMPs, as mentioned previously), a series of metabolic, cellular and genetic processes are triggered to enable DCs to mature and migrate to T-cell rich areas in secondary lymphoid organs, easing antigen presentation. These processes include reorganization of the cytoskeleton, higher motility, decrease in endocytic capacity and loss of adhesive structures, as well as an increase in the expression of MHC-II, co-stimulatory molecules and homing chemokine receptors such as CCR7. These changes promote the interaction, priming and activation of antigen-specific T-cells<sup>111</sup>.

##### 1.2.1.2.1 The MHC system

Antigen processing and presentation are the basis of adaptive immunity. For B cells to produce high-affinity antibodies, they require assistance from CD4+ T cells, and these T cells use antigen-specific receptors to recognize peptides presented by MHC molecules. Cytotoxic CD8+ T-cells also need to recognize peptide-MHC complexes to eliminate virus-infected cells, among others. The two main classes of MHC are named MHC class I and class II molecules (MHC-I or MHC-II), which present peptides to CD8+ and CD4+ T cells, respectively.

MHC-I molecules or HLA (Human Leukocyte Antigens) in humans are a cluster of co-dominantly expressed genes that codify for antigen processing pathways. The HLA loci are located on the short arm of human chromosome 6. These loci are divided into three regions: class I, class II, and class III. Each region contains classical and non-classical loci; the classical loci encode HLA molecules involved in antigen presentation, while the non-classical loci encode other HLA molecules. The class I region includes classical HLA loci such as HLA-A, -B, and -C, and non-classical loci such as HLA-E, -F, and -G. The class II

region contains genes encoding the  $\alpha$  and  $\beta$  chains of classical class II HLA molecules (HLA-DP, -DQ, and -DR) as well as non-classical class II molecules (HLA-DM and -DO). The  $\beta$  chains of HLA-DR molecules are encoded by DRB1 and sometimes linked with DRB3, DRB4, or DRB5, forming additional haplotypes<sup>117,118</sup>. We will only be exploring the classical HLA molecules here.

The class III region does not contain classical or non-classical HLA genes but includes genes encoding complement proteins and some cytokines<sup>119,120</sup> (Figure 6).

MHC- I and -II molecules share a similar structural fold, consisting of a peptide-binding platform formed by two domains that create a  $\beta$ -sheet base with two  $\alpha$ -helices on top. The MHC-I molecules are composed of a membrane-anchored, MHC-encoded glycoprotein heavy chain in association with  $\beta$ 2-microglobulin, binding smaller peptides typically around 8-10 amino acid residues, although longer peptides can also bind in some cases<sup>120</sup>. In contrast, MHC-II molecules consist of MHC-encoded, membrane-anchored glycosylated  $\alpha$  and  $\beta$  subunits and accommodate larger peptides, generally in the 10-15 amino acid range, but they can bind much longer peptides due to their more promiscuous nature<sup>121</sup>. Peptides bind within the groove between the helices through conserved hydrogen bonds and defined pockets. Human MHC-I (HLA-A, -B, -C) and II (HLA-DR, -DP, -DQ) proteins have variations in their peptide-binding grooves, influencing the repertoire of peptides presented to CD8+ and CD4+ T cells, respectively. The formation of peptide-MHC complexes is determined by factors like antigen availability, protease activity, and chaperones such as tapasin and HLA-DM, which facilitate the presentation of stable complexes (detailed below). Despite advances in bioinformatics for predicting peptide binding affinities, accurately predicting immunodominance remains challenging due to the dynamic nature of protein conformations.

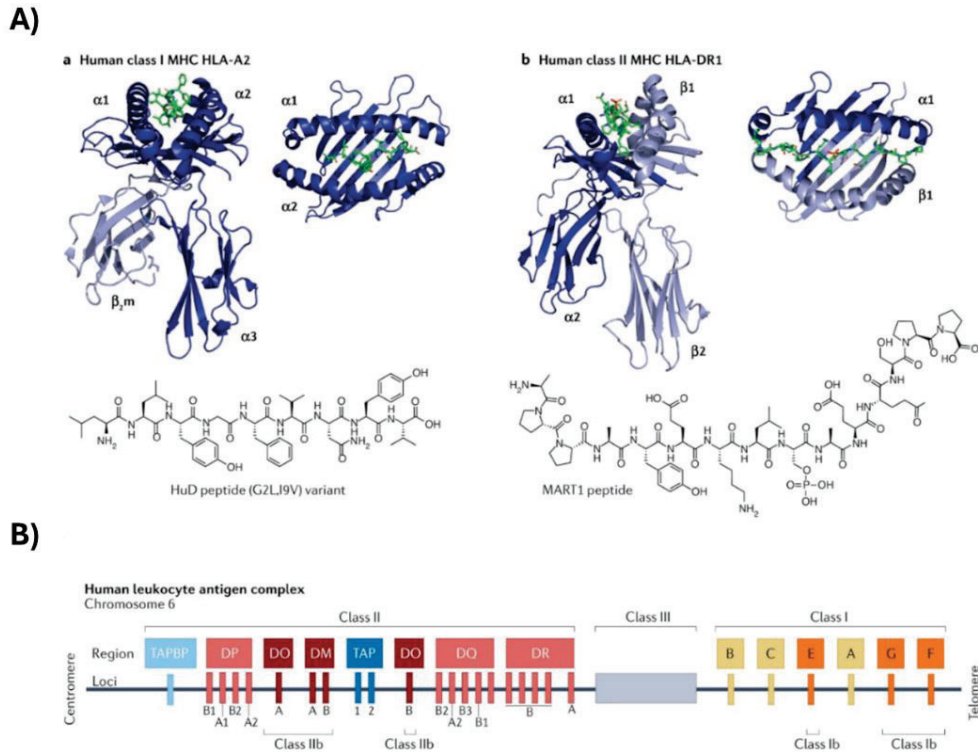


Figure 6. Human Leukocyte Antigen Complex and HLA-I and II molecular structures. A) Molecular models of classical MHC molecules and MHC-like molecules. a: Side view and top view (rotated by 90°) of a crystal structure of a human MHC-I–peptide complex (Protein Data Bank (PDB) ID 3PWJ). b: Side view and top view (rotated by 90°) of a crystal structure of a human MHC-II–peptide complex (PDB ID 3L6F). B) Simplified schematic maps of the MHC complex in humans. MHC genes, including genes for the transporter associated with antigen processing (TAP) and tapasin (also known as TAP-binding protein (TAPBP)). Adapted from<sup>119</sup>.

As commented, the HLA system is highly polygenic and polymorphic, it has thousands of allelic variants, making it the most diverse genetic system in the human genome. This diversity, especially in the antigen-binding region, defines each individual's peptidome<sup>122</sup> and contributes to the differences in susceptibility to diseases among genetically distinct groups, thus offering evolutionary advantages of a diverse immunological response to a wide range of infectious pathogenesis.

#### 1.2.1.2.2 Antigen processing

Processing of sampled antigens to ultimately load them onto classical MHC molecules differs between MHC-I and MHC-II, where MHC-I molecules associate to the products of cytosolic proteolysis (with fine-tuning in the ER), while MHC-II molecules mainly present extracellular antigens captured and processed in endolysosomal compartments<sup>119</sup>. The key parts of antigen processing are compartmentalization of proteolysis and the distinct intracellular distributions of MHC molecules, which will be explained in more detail (Figure 7).

Starting with antigen capture, cells engaged in MHC-II-restricted antigen presentation acquire antigens through phagocytosis and endocytosis, using various receptors like the BCR, mannose receptor, complement receptors, and Fc receptors<sup>123</sup>. The BCR is unique as it activates B cells and internalizes immune complexes for MHC-II presentation, calling for CD4+ T cell help<sup>124</sup>. Opsonization (covalent modification with components of the complement pathway) enhances antigen acquisition via Fc and complement receptors, while macropinocytosis and micropinocytosis deliver extracellular fluid and contents to endocytic compartments. Autophagy targets damaged organelles and viral replication machineries for degradation. The endolysosomal pathway, a network of interconnected compartments, continuously processes extracellular proteins and contributes peptide ligands for MHC-II. MHC-I molecules sample the cell proteome, including viral proteins in infected cells, via the ubiquitin–proteasome system and mechanisms like cross-presentation and ER-associated degradation. Non-classical MHC-I molecules, such as HLA-E, present ligands derived from signal peptides processed in the ER<sup>119,125</sup>.

Antigens then need to be tagged for proteolysis. MHC-II molecules primarily use lysosomal proteolysis to convert proteins into peptides for presentation, utilizing the acidic, protease-rich environment of endocytic compartments, which are less acidic in DCs than in macrophages probably favouring the generation of MHC-II ligands in the firsts<sup>126</sup>. These compartments are rich in reduced glutathione, promoting disulfide bond reduction, aided by the IFN- $\gamma$ -inducible oxidoreductase GILT<sup>127</sup>, facilitating protein unfolding for protease action. For MHC-I molecules, proteins are tagged for degradation by ubiquitylation, entering the ubiquitin–proteasome system due to translation errors or misfolding, with defective ribosomal products potentially serving as neoantigens<sup>128</sup>.

This system recognizes misfolded proteins, using a cascade of E1, E2, and numerous E3 ubiquitin ligases, ensuring both host and pathogen-derived proteins are processed for immune surveillance<sup>129</sup>.

Proteolysis by lysosomal proteases and proteasomes involves the formation of acyl-enzyme intermediates, typically lysed by hydrolysis, but peptides can also be processed by transpeptidation to form 'spliced' peptides<sup>130</sup>. The ubiquitin-proteasome system primarily generates peptides for MHC-I presentation, where proteins tagged for degradation by ubiquitin enter the proteasome, a complex with a core of four rings of several subunits, with catalytically active  $\beta$ -subunits sandwiched between  $\alpha$ -subunits<sup>131</sup>. This system degrades misfolded proteins, producing peptides via hydrolysis and transpeptidation, which are then potential MHC-I ligands. The immunoproteasome, induced by IFN- $\gamma$ , also enhances the production of MHC-I compatible peptides. ER-associated aminopeptidases further trim these peptides, and glycoproteins lose their N-linked glycans in the cytosol<sup>132,133</sup>.

Once peptides are lysed into potential epitopes, those that will be presented by MHC-I molecules, which are generated in the cytosol, must be transported across the membrane into the lumen of the ER where the peptide-binding portion of newly synthesized MHC-I molecules reside<sup>134</sup>. This transport is facilitated by the MHC-encoded TAP complex, a heterodimeric transporter that uses ATP to translocate peptides suitable for MHC-I binding. The COOH terminus of these peptides is crucial for TAP acceptance and fitting into the MHC-I groove, however further trimming by ERAAP may be needed<sup>135-137</sup>. In contrast, MHC-II molecules acquire peptides primarily in endocytic compartments. MHC-II proteins fold in the ER in complex with a protein called invariant chain (Ii) and are then transported to late endosomal compartments.

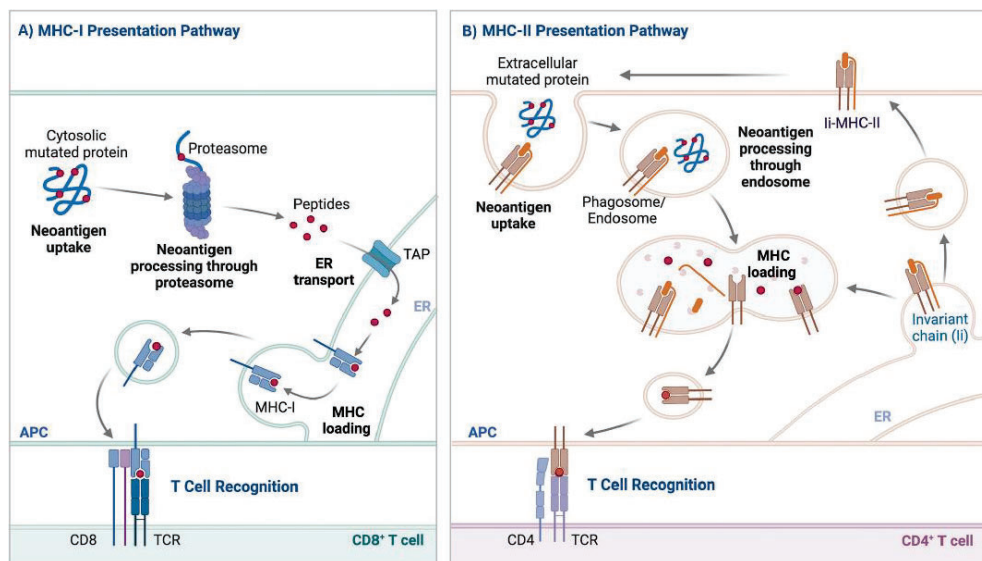


Figure 7. Overview of the MHC-I and MHC-II presentation pathways, focused on neoantigens. Neoantigens can develop at the genomic and at proteomic level; mutant peptides created by the proteasome from endogenous mutated proteins are loaded onto MHC-I molecules in the ER, and peptides derived from an exogenous mutant protein broken down in the endosomal compartment are loaded into MHC-II molecules in the endosomes. These peptide-MHC (pMHC) complexes then traffic to the cell surface where they are recognized by T cells. Made with Biorender.

Endocytic molecules are directed to these compartments by the MHC-II-Ii, whose NH<sub>2</sub>-terminal tail contains sorting information. MHC-II molecules that avoid this sorting end up on the cell surface but can be internalized into endosomal compartments. The invariant chain is eventually degraded to form the CLIP peptide, which blocks peptide binding until removed by HLA-DM, ensuring proper antigen presentation by MHC-II molecules<sup>138,139</sup>.

Finally, loading of peptides onto MHC molecules occurs as following: For MHC-I molecules, the peptide-loading complex (PLC, which includes TAP, tapasin, ERp57, calreticulin, and possibly calnexin) is essential. Tapasin is crucial for this process, preferentially interacting with MHC-I molecules yet to receive peptides, and its absence increases the display of peptide-receptive MHC-I molecules on the cell surface. For MHC-II molecules, the removal of CLIP requires HLA-DM, which promotes peptide exchange until a stable peptide-MHC-II complex is formed. HLA-DO modulates HLA-DM activity, affecting the peptide repertoire presented by MHC-II molecules<sup>140,141</sup>. Most steps of

antigen processing and presentation are constitutive in cells expressing MHC molecules, which affects T cell development and autoimmunity by displaying a diverse set of self-peptides for T cell repertoire testing in the thymus. DCs keep most of their MHC-II molecules intracellularly until activated by stimuli like TLRs, prompting these molecules to move to the cell surface for T cell interaction, a process influenced by cytokines such as IFN- $\gamma$  and IL-10.

#### 1.2.1.2.3 Cross-presentation

Cross-presentation is defined as extracellular antigens being presented on MHC I molecules, where normally they would be presented on MHC-II molecules. It is a key process carried out by some DC subsets (as previously stated) for collecting antigens from other cells in tissues, including cancerous or infected cells, and present them to CD8<sup>+</sup> T-cells to initiate immune responses. Two main pathways have been identified: the vacuolar and the endosome-to-cytosol pathways (Figure 8).

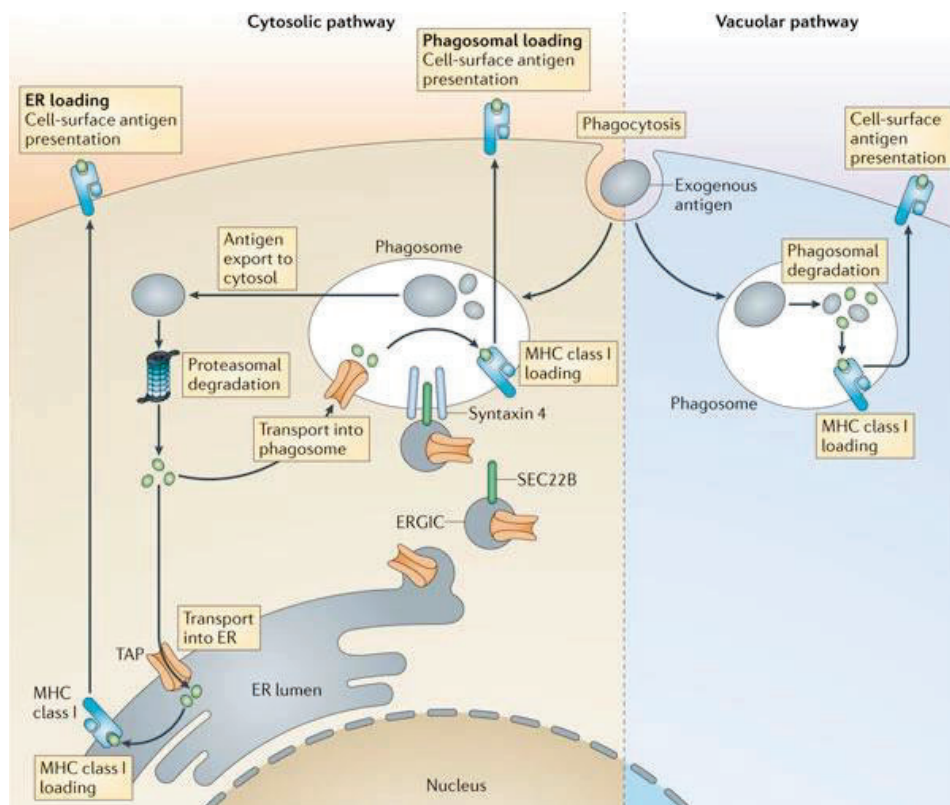


Figure 8. Antigen cross-presentation intracellular pathways. Adapted from<sup>142</sup>.

In the vacuolar pathway, antigens are processed and loaded onto MHC I molecules within the endo/lysosomal compartment, facilitated by lysosomal proteases such as Cathepsin S. This is also implied by the fact that the vacuolar pathway is resistant to proteasome inhibitors and independent of the protein TAP, but sensitive to lysosomal proteolysis inhibitors<sup>142,143</sup>.

The endosome-to-cytosol pathway involves transporting antigens to the cytosol for proteasomal degradation, with peptides then being transported by TAP to the endoplasmic reticulum for MHC-I loading or brought again to the phagosome for the same purpose. The ER-associated aminopeptidase 1 (ERAP1) and the endosomal insulin-responsive aminopeptidase (IRAP or cystinyl aminopeptidase) have been linked to cross-presentation for their ability to trim proteasome-generated polypeptides to make them suitable for loading onto MHC-I molecules<sup>144,145</sup>.

### *1.2.1.3 Antigen presentation and Lymphocyte activation*

DCs are the translators between innate and adaptive immunity. As explained previously, once an antigen is sensed and processed for presentation, DCs involve the adaptive immune system to mount an immunological response by transmitting what kind of danger are they facing. Specifically, DCs typically inform naïve T-cells circulating through the draining LNs, which upon contact will polarize to a functional subtype, dictated by the nature of their activation. T-cell activation and fate is crucial for an effective adaptive immune response.

#### *1.2.1.3.1 DC migration to LNs*

Antigen can reach LNs via two mechanisms: free draining via lymph (which requires a high antigenic dose)<sup>146</sup> or carried by cells, primarily migratory DCs.

Activated and antigen-loaded migratory DCs are guided by lymphatic-derived ligands such as CCR7 and the signalling lipid sphingosine-1-phosphate (S1P) to enter the LN parenchyma. Within the LN, these DCs can either present the antigen directly or transfer it to LN-resident DCs or B cells<sup>147,148</sup>. After presenting antigens for about 1–3 days, these activated DCs typically undergo cell death and do not migrate to other lymph nodes, requiring a new influx of DCs to continue antigen presentation<sup>149,150</sup>.

The composition and numbers of migratory DCs reaching a given LN (either cDC1s, cDC2s, or non-conventional DCs) depend on factors such as tissue type, timing, and the nature of the immunization. For example, because there are relatively fewer cDC1s in the skin, subcutaneous immunization results in a cDC2-dominated response in the LN, with antigen-carrying Langerhans cells (LCs) arriving a few days later. In contrast, in the lungs, an almost equal number of cDC1s and cDC2s transport antigens to mediastinal LNs within the first day of an immune response. After intramuscular immunization, both cDC1s and cDC2s are involved in antigen delivery to local LNs, but the number of DCs migrating to the draining LNs from muscle tissue is much lower than from the skin or lungs<sup>147,150</sup>. Therefore, the route of exposure, along with the antigen size and dose, influences which DC subsets present antigens and determines where this presentation occurs within the LN.

Once in the LN, migratory cDC1s and cDC2s enter the interfollicular zone and then move to the different LN regions depending on their subtype – generally, they organize in a central T-cell zone populated by migratory and LN-resident cDC1s along with LCs, and a peripheral ring of cDC2s at the T cell–B cell border.

### **1.2.2 T-lymphocytes**

Lymphocytes consist of two primary types: T cells and B cells. T cells are responsible for monitoring the body for foreign peptides, initiating an effective immune response, while B cells detect foreign antigens, leading to their activation, internalization of the antigen, and subsequent production of antibodies to fight infection. These lymphocytes identify cognate antigens presented on the surface of APCs through their specific receptors—TCRs for T cells and BCRs for B cells. When these receptors successfully engage with antigens, they trigger downstream signalling cascades that prompt functional responses. After receptor engagement, the lymphocytes undergo dynamic cytoskeletal reorganization, forming membrane protrusions that enable them to adhere to and spread over the surface of the APC. This interaction creates a specialized contact zone known as the immunological synapse. B cells can also act as APCs, especially at the T cell–B cell border in the LNs, promoting further B cell differentiation. However, we will only be discussing T-cells in this introduction.

### 1.2.2.1 T-cell ontogeny

T cells develop from common lymphoid progenitors (CLPs) derived from HSCs in the bone marrow and migrate to the thymus as pro-T cells. There, they undergo proliferation, TCR gene rearrangement, and selection. For  $\alpha\beta$  T-cells, in which we will centre this introduction, they initially start as double-negative (DN) cells for CD4 and CD8 molecules before rearranging their TCR  $\beta$ - and  $\alpha$ -chain genes and expressing both CD4 and CD8 as double-positive (DP) thymocytes. Positive selection by cortical thymic epithelial cells determines whether they become CD4 or CD8 single-positive (SP) thymocytes based on MHC engagement<sup>151,152</sup>. In the medulla, negative selection by medullary thymic epithelial cells (mTECs) and DCs eliminates strongly self-reactive cells, ensuring central tolerance. Regulatory T cells (Tregs), expressing Foxp3, emerge through a distinct selection process (see section 1.2.2.3.5)<sup>152,153</sup>.

Human T cell development begins around nine weeks of gestation, with mature T cells appearing in the thymus by 12–13 weeks and in peripheral lymphoid tissues by 24 weeks<sup>151</sup>. Thymic activity peaks at birth but declines with age, particularly after 40 years, due to structural changes like epithelial loss and fatty infiltration<sup>153</sup>.

### 1.2.2.2 T-cell activation: Antigen presentation and the TCR

In the periphery, naïve T cells remain inactive until their TCR specifically binds to an antigenic peptide presented by an APC. MHC class I molecules present peptides to CD8+ T cells, while MHC class II molecules are recognized by CD4+ T cells.

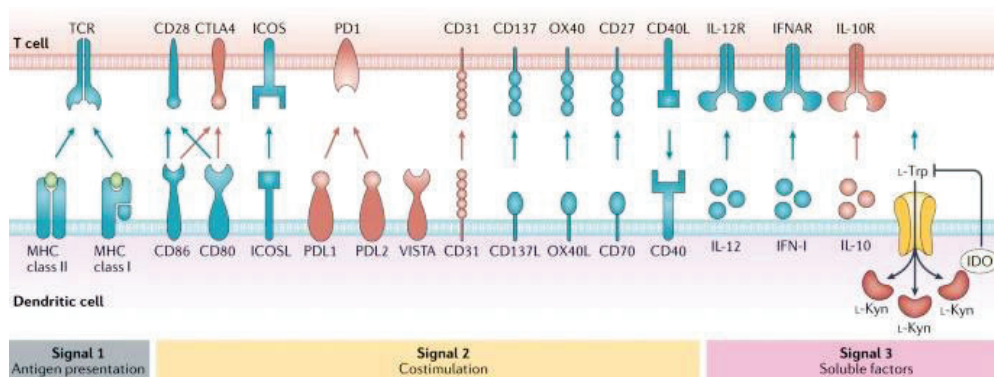


Figure 9. Induction of T cell-activation by DCs. Dendritic cells initiate T cell activity by presenting tumour-associated antigens on MHC class I and MHC class II molecules. Activation also requires

*additional positive signalling (represented by blue arrows and receptors). These include costimulatory molecules from the B7 and TNF families, as well as soluble factors such as IL-12 and IFN- $\gamma$ . In contrast, inhibitory mechanisms (depicted by red arrows and receptors) act to limit T cell activation. CTLA4 (cytotoxic T lymphocyte antigen 4), ICOS (inducible T cell co-stimulator), IDO (indoleamine 2,3-dioxygenase), IL-10R (IL-10 receptor), IL-12R (IL-12 receptor), Kyn (kynurenine), PD1 (programmed cell death protein 1), PDL1 (programmed cell death 1 ligand 1), and VISTA (V-domain immunoglobulin suppressor of T cell activation). Adapted from<sup>154</sup>.*

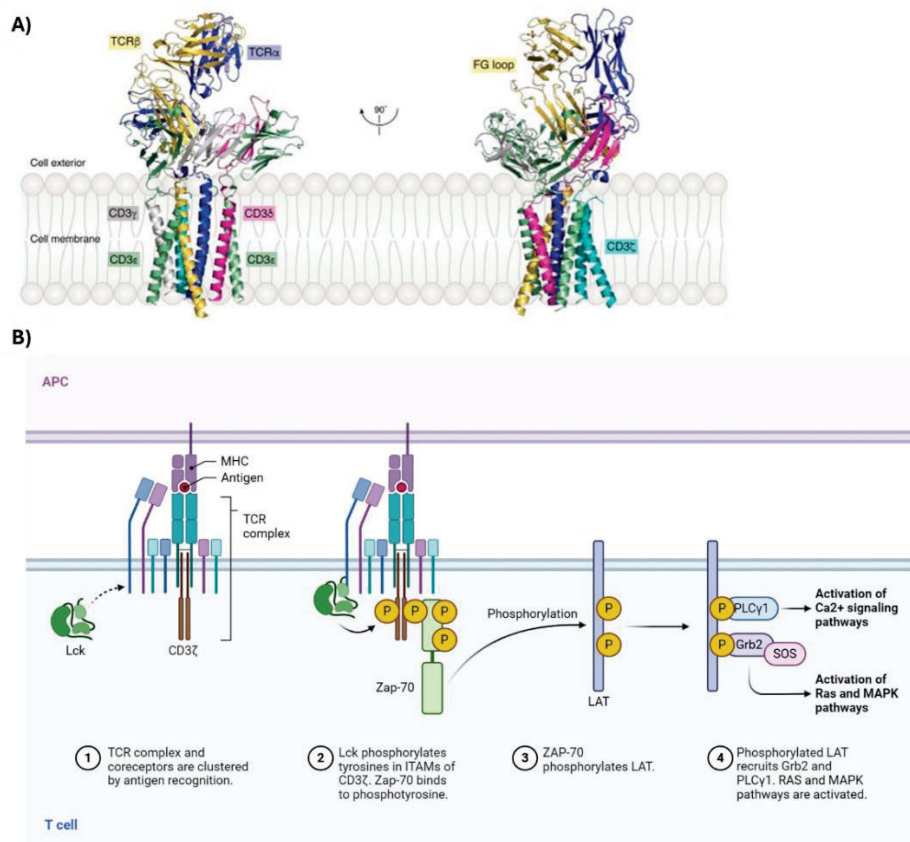
The TCR forms a complex with CD3 and is further associated with the CD4 or CD8 coreceptors. Activation requires three coordinated signals (Figure 9): (1) TCR engagement with pMHC complexes from APCs, with CD4 or CD8 coreceptors enhancing this interaction and initiating intracellular signaling by phosphorylating ITAM motifs on CD3; (2) costimulatory signals via CD80/CD86 on APCs interacting with the CD28 receptor on T cells, activating pathways like MAPK and transcription factors such as *NF $\kappa$ B*, *NFAT*, and *AP-1*<sup>155</sup>; and (3) cytokine-mediated signals from APCs that guide differentiation into effector T cell subsets. CD4<sup>+</sup> T cells differentiate into subtypes such as Th1, Th2, Tfh, Th17, or Tregs, each with specific functions, cytokine secretion patterns, and lineage-defining transcription factors<sup>156</sup>.

As mentioned, the recognition of p-MHC complexes by TCR induces upregulation of a series of activation markers (AIM, activation-induced markers) on the T cells, many of which are co-stimulatory molecules to activate the second signal, which can be detected by flow cytometry and are useful to perform functional analysis<sup>157</sup>. Amongst the most used are: CD25 (IL-2R $\alpha$ ), for antigen-specific CD4<sup>+</sup> T cells, but it must be used in combination with another marker due to its constitutive expression on Tregs. It is upregulated through TCR signaling and IL-2-induced *STAT5* signaling<sup>158</sup>. CD69, an activation inducer molecule, is mainly used to identify antigen-specific CD8<sup>+</sup> T cells, with its expression peaking around 16 hours post-activation, though it requires additional markers to ensure specificity due to TCR-independent induction<sup>159</sup>. CD40L, a member of the TNF superfamily, is transiently expressed on both CD4<sup>+</sup> and CD8<sup>+</sup> T cells after antigen stimulation, promoting cytokine production, but it is rapidly downregulated<sup>160</sup>. OX40 (CD134), another TNF receptor superfamily member, is expressed on CD4<sup>+</sup> T cells following TCR signaling and is used alongside CD25 to detect antigen-specific CD4<sup>+</sup> T cells. It plays a key role in T-cell survival and expansion<sup>161</sup>. 4-1BB (CD137), also part of the TNF receptor superfamily, is upregulated after TCR activation and promotes T-cell

proliferation and cytokine production. It is often used with CD69 to identify antigen-specific CD8<sup>+</sup> T cells<sup>162</sup>. Finally, PD-L1 is an early activation marker that plays a role in suppressing autoimmunity and antitumour immunity. When combined with CD25, PD-L1 is effective in detecting antigen-specific cells<sup>163</sup>.

#### 1.2.2.2.1 The TCR complex

The TCR-CD3 complex consists of a TCR heterodimer ( $\alpha\beta$  or  $\gamma\delta$ ) associated with invariant CD3 dimers: CD3 $\epsilon\gamma$ , CD3 $\epsilon\delta$ , and CD3 $\zeta\zeta$ <sup>164</sup>. The TCR's sensitivity enables it to recognize a single peptide among thousands of other MHC-bound peptides on an APC<sup>165</sup>. We will only be exploring the TCR  $\alpha\beta$  in this section. Each TCR subunit ( $\alpha$  and  $\beta$ ) includes an immunoglobulin-like extracellular domain, a connecting peptide, a transmembrane region, and a cytoplasmic tail. The CD3 dimers share a similar structure, except CD3 $\zeta$  has a shorter extracellular segment<sup>164,166</sup> (Figure 10).



*Figure 10. Structure and signalling of the TCR–CD3 complex. A) side views of the overall structure of the TCR–CD3 complex shown in ribbon representation (PDB code 6JXR), adapted from<sup>166</sup>. B) Simplified schematic representation of TCR-CD3 complex activation and signalling cascade. Lck (lymphocyte-specific protein tyrosine kinase); ITAM (Immunoreceptor tyrosine-based activation motif); LAT (Linker for Activation of T-cells); PLC $\gamma$ 1 (Phospholipase C Gamma 1); Grb2 (Growth factor receptor-bound protein 2); SOS (Son of Sevenless, set of genes encoding guanine nucleotide exchange factors that act on the Ras subfamily); MAPK (Mitogen-activated protein kinase). Made with Biorender.*

TCRs engage pMHC ligands via six complementarity-determining region (CDR) loops, three from each chain<sup>167</sup>. CDR1 and CDR2 are encoded within the TCR V segments, while CDR3 is generated by DNA recombination. Upon binding to pMHC, the CD3 molecules trigger activation signals to the T cell. CD3 $\epsilon$ ,  $\gamma$ ,  $\delta$ , and  $\zeta$  chains contain ITAMs that undergo phosphorylation by Src kinase Lck, initiating downstream signaling<sup>168</sup>.

Various models have been proposed for TCR activation, including aggregation of TCR–CD3 complexes for enhanced phosphorylation<sup>169</sup>, segregation of the inhibitory phosphatase CD45<sup>170</sup>, mechanosensing during T-cell scanning<sup>171</sup>, and allosteric changes at TCR binding sites<sup>172</sup>. Once Lck phosphorylates tyrosines in ITAMs of CD3 $\zeta$ , Zap-70 binds to phosphotyrosine, phosphorylates LAT (Linker for Activation of T-cells) and the latter recruits the downstream proteins Grb2 and PLC $\gamma$ 1, which end up activating the RAS and MAPK pathways, which play an important role in complex cellular programs like proliferation, differentiation, development, transformation, and apoptosis<sup>173</sup>(Figure 10).

### 1.2.2.3 CD4+ T-helper cells differentiation

As mentioned before, once activated, CD4+ T cells can differentiate into several effector subsets with a distinct expression of surface molecules, cytokines, and key transcription factors, such as Th1, Th2, Treg, Tfh, Th17, Th9, Th22, CD4+ cytotoxic T lymphocytes (CTLs), among others. Here we will introduce the main Th subsets involved in cellular immunity (Figure 11), including Th1, Th2, Treg, Th17 and Th9. As a brief note, Tfh cells localize in germinal centres, where they provide essential signals for B cell maturation, antibody affinity maturation, and class switching. Their function is critical for the development of high-affinity, long-lived humoral immunity.

#### 1.2.2.3.1 Th1 cells

Th1 cells aid in protecting hosts against intracellular bacteria and viruses by producing IFN- $\gamma$ , which is a potent inflammatory cytokine that activates macrophage destruction of phagocytosed materials, helps induce an antiviral state, and aids cytotoxic CD8+ T cell responses<sup>174</sup>. IL-12 and IFN- $\gamma$  are necessary cytokines for Th1 differentiation, which when combined with TCR stimulation and IFN- $\gamma$ -*STAT1* signaling, induce the expression of *T-bet* (the major transcription factor driving Th1 differentiation) while suppressing Th2/Th17 lineages. T-bet also binds to the *Ifng* gene to increase the expression of IFN- $\gamma$  and promotes the expression of *IL-12R $\beta$ 2*, allowing cells to respond to IL-12. IL-12 then promotes STAT4 activation, which maintains T-bet expression<sup>175</sup>. Other factors such as a strong TCR engagement have also proven to promote Th1 differentiation.

#### 1.2.2.3.2 Th2 cells

Th2 cells are defined by the expression of the transcription factor GATA-3, and produce the cytokines IL-4, IL-5, and IL-13, among others, which coordinate the accumulation of eosinophils and production of IgE by B cells and mucus by epithelial cells. Th2 cells are needed to protect the host against helminth infections and facilitate tissue repair, however they also play a role on chronic inflammation such as asthma and allergy. The signals from APCs that drive Th2 differentiation are still incompletely defined, however it is known that the IL-4 secreted by cells like innate lymphoid cell group 2 (ILC2), basophils or NKs leads to expression of GATA-3 through STAT6 phosphorylation, which initiates the production of Th2-related cytokines (such as IL-4, which further promotes Th2 differentiation) and the suppression of the Th1 lineage. IL-2 is also necessary to maintain Th2 lineage, as it promotes STAT5 priming and the expression of *IL-4RA*<sup>176</sup>. It has also been seen that low-affinity peptide-TCR interactions or the presence of low amounts of antigen promotes Th2 differentiation, as well as *Notch* signaling by DCs.

#### 1.2.2.3.3 Th9 cells

Th9 cells, identified as a distinct subset only since the last decade, are involved in infectious diseases, allergy, cancer, and autoimmune immunity. These cells can be generated in vitro by TCR stimulation in the presence of IL-4 and TGF- $\beta$ , produce high levels of IL-9, IL-10 and IL-21, and express key transcription factors such as IRF4 and PU-1. IL-4 signaling induces STAT6 phosphorylation, which drives the expression of

GATA-3, IRF4, and BATF, promoting *IL-9* transcription and Th9 differentiation. Also, TGF- $\beta$  signaling activates Smads (Smad2/3), PU.1, and IRF8, further enhancing Th9 cell differentiation<sup>177</sup>. Other factors influencing Th9 cell development include costimulatory signals (CD28, OX40, GITR, Notch, and DR3) and cytokines such as IL-1, IL-25, IL-7, and IL-21.

#### 1.2.2.3.4 Th17 cells

Th17 cells are defined by the production of cytokines such as IL-17A-F, IL-21, IL-10, IL-23, and IL-22, and their master transcription factor, ROR $\gamma$ t. These cells are crucial for defence against extracellular pathogens, particularly in mucosal tissues, and play a role in chronic inflammation and autoimmune diseases. IL-6, produced by DCs, induces ROR $\gamma$ t expression via STAT3 activation while suppressing Foxp3, a TGF- $\beta$ -induced factor. ROR $\gamma$ t directly promotes the expression of IL-17A, IL-17F, IL-22, and IL-23R. Smad2/3 signaling from TGF- $\beta$  can sustain STAT3 activity, while autocrine IL-21 further amplifies STAT3 activation via JAK1/3, enhancing IL-23R expression and responsiveness to IL-23. IL-23 subsequently strengthens STAT3 activation, stabilizing Th17 cells. Strong TCR stimulation also induces Th17 differentiation by enhancing CD40L expression, which in turn increases IL-6 production by DCs<sup>178</sup>.

#### 1.2.2.3.5 T-reg cells

Treg cells are an essential subset of CD4<sup>+</sup> T cells for maintaining immune tolerance, which act by suppressing immune responses. These cells are characterized by high expression of CD25 (IL-2 R $\alpha$ , which is a high-affinity receptor for IL-2), the production of inhibitory cytokines such as IL-10, TGF- $\beta$ , and IL-35, and their transcription factor Foxp3. Treg cells can have different origins: thymic Treg (tTreg) cells, which develop in the thymus and are promoted by increased-affinity interactions with self-peptide-MHC complexes, and induced Treg (iTreg) cells, differentiated from CD4<sup>+</sup> T cells in the periphery with the presence of TGF- $\beta$  and IL-2. Foxp3 expression is regulated by different mechanisms: TCR-CD28 signaling initiate Foxp3 expression by activating NF- $\kappa$ B, AP-1, and NFAT, which bind to its enhancer and promoter regions. TGF- $\beta$  signaling enhances Foxp3 transcription via Smad2, Smad3, and forkhead proteins FoxO1 and FoxO3. Additionally, STAT5, activated downstream of IL-2, stabilizes Foxp3 expression<sup>179</sup>.

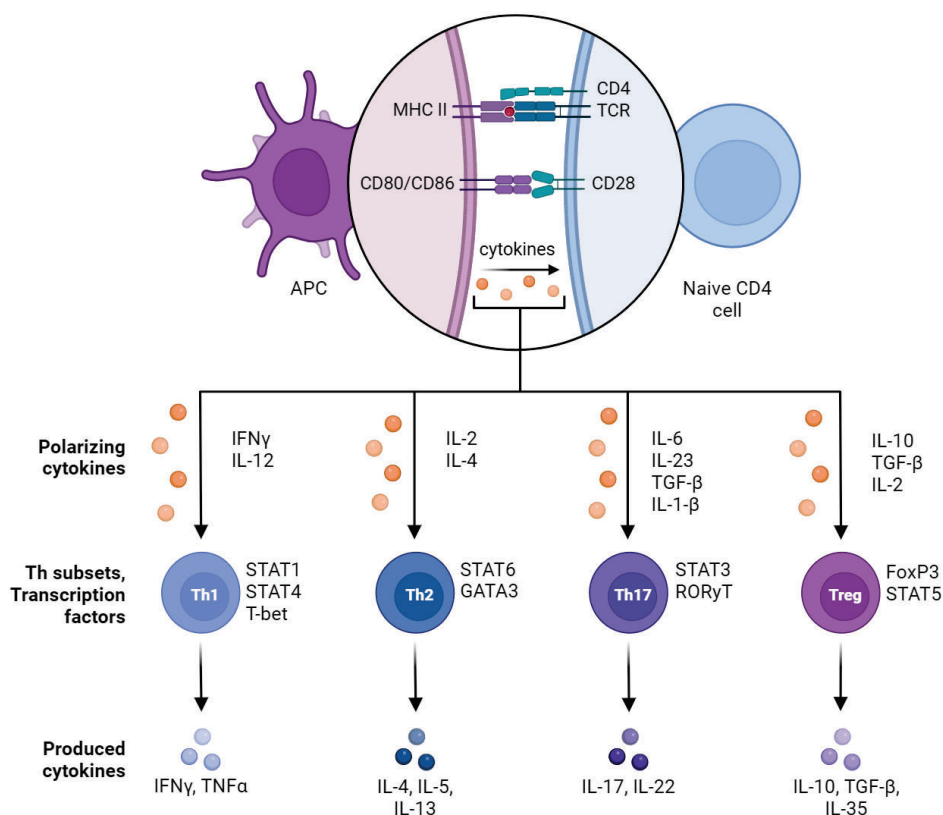


Figure 11. Schematic representation of naïve CD4<sup>+</sup> T-cell activation and differentiation into some distinct T-helper subpopulations based on the polarizing cytokines induced by the APC. Made with Biorender.

#### 1.2.2.4 CD4<sup>+</sup> and CD8<sup>+</sup> T-cell interaction

CD4<sup>+</sup> T cells are crucial for an optimal CD8<sup>+</sup> T cell response, and there is strong and consistent evidence that CD4<sup>+</sup> T cell support is essential for the generation and recall of memory CD8<sup>+</sup> T cells. Mechanistically, in a two-step priming process, CD4 and CD8 T cells initially activate independently by DC priming but later interact with the same cross-presenting cDC1 cells. Then CD4<sup>+</sup> T cells assist CD8<sup>+</sup> T cells by producing cytokines such as IL-2 and IL-21, engaging in CD40L signaling with DCs enhancing antigen presentation and costimulatory molecule expression, and modulating CD25 and PD-1 expression on CD8<sup>+</sup> T cells. They also produce cytokines like IL-2, IL-12, and type 1 IFNs to drive CD8<sup>+</sup> T cell differentiation and memory formation. Without CD4<sup>+</sup> T cell

help, CD8<sup>+</sup> T cells exhibit impaired memory responses, increased inhibitory molecule expression, and signs of exhaustion<sup>174,180–182</sup>.

#### 1.2.2.5 *T-cell Memory and effector differentiation*

Memory T cells, characterized by distinct surface markers, effector functions, and trafficking properties, are classified into subsets: central memory (TCM), effector memory (TEM), tissue-resident memory (TRM), and stem memory T cells (TSCM). Many disparate factors can serve to influence cell-fate specification in the first 24–72 hours following T cell activation; these include TCR signal strength, co-stimulation, inflammatory cytokines, tissue microenvironment, metabolic regulators and intermediates, transcription factors and the mode of cellular division (symmetric or asymmetric). These subsets, considered stable "fates," can occasionally interconvert, influenced by external signals like inflammation, TCR engagement, and cytokine levels<sup>183</sup>.

While differentiating (through activation) from TSCM to TCM, TTM (T-Transitional Memory), TEM and culminating in TTE (T-Terminal Effector) cells, memory T cells progressively lose or acquire specific surface protein markers. For example, CD197<sup>+</sup>CD62L<sup>+</sup> subsets, such as TN (T Naïve), TSCM, and TCM, migrate to lymphoid tissues to interact with APCs, while CD197<sup>−</sup>CD62L<sup>−</sup> subsets, such as TEM and TEFF (T-Effectors), patrol peripheral tissues and blood, rapidly producing effector molecules upon encountering antigens. Early memory subsets like TSCM and TCM are characterized by longevity and high proliferative capacity, which decrease with differentiation. Surface markers such as CD45RA/CD45RO, CD197 (CCR7), CD62L, CD95, CD28, CD103 and cytokine receptors IL-2R $\beta$  (CD122) and IL-7R $\alpha$  (CD127) are used to distinguish these subsets<sup>184,185</sup>.

##### 1.2.2.5.1 CD8<sup>+</sup> subsets

CD8<sup>+</sup> T lymphocytes, while less functionally diverse than CD4<sup>+</sup> T cells, exhibit considerable heterogeneity in cytokine production, cytotoxic molecule expression, and phenotypic markers. Effector CD8<sup>+</sup> T cells can differ in their ability to produce cytokines such as IFN- $\gamma$ , TNF, and IL-2 or to co-produce these with chemokines, enhancing their protective and anti-tumour capacities. While research has traditionally been focused on cytotoxic CD8 T cells, or Tc1 cells, which are the major population, known for producing

cytolytic cytokines like IFN- $\gamma$ , granzyme B, and TNF- $\alpha$ , more recent studies have uncovered additional subsets of effector CD8 T cells, each with unique characteristics distinct from those of Tc1 cells. These include subsets such as Tc2, Tc9, Tc17, Tc22, *etc.*, which closely resemble their CD4 T-cell subset counterparts in their signal 3 cytokine requirements, lineage-determining transcription factors, and effector cytokine profiles. This additional diversity is present on homeostatic conditions but may also depend on pathogen-specific inflammatory environments<sup>183,186</sup>.

Upon antigen interaction, naive T lymphocytes activate and proliferate, producing progeny with effector and memory fates. Effector cells mediate immediate protection, while memory cells ensure long-term immunity. Effector CD8+ T cells experience a sharp contraction phase during which most pathogen-specific cells undergo apoptosis. A small fraction (~5–10%) survives and matures into memory CD8+ T cells, a non-random process due to intrinsic CD8+ cell properties that make some more likely to persist and contribute to the memory T cell population<sup>187</sup>. It has also been demonstrated that repeated reactivation of memory CD8+ T cells through vaccines or infections enhances their effector properties and increases the proportion of TEM cells in the memory pool<sup>187,188</sup>.

#### 1.2.2.5.2 CD4+ subsets

Regarding CD4 T-cells, after differentiation into the Th subsets, also can transform into memory cells. As CD8+ T-cells, most effector CD4+ T cells are short-lived, but about 5–10% form long-lasting memory cells that persist after pathogen clearance. Memory CD4+ T cells, which are antigen-experienced, do not need MHC-II stimulation for survival, are more abundant for a given antigen, and mount a quicker, more robust response compared to naive cells. Although memory CD4+ T cell stability can vary by pathogen type, human studies show that memory CD4+ cells can be more stable than CD8+ cells after smallpox vaccination but less stable after measles vaccination. CD4+ TCM cells are defined by their ability to synthesize IL-2 and have greater proliferative potential, while TEM cells are more ready to secrete effector cytokines like IFN- $\gamma$ . Epigenetic studies show that memory Th1 cells have increased accessibility for IFN- $\gamma$  promoters, while memory Th2 cells show little plasticity. Memory Th17 cells are less epigenetically fixed, allowing them to differentiate into IFN- $\gamma$ -producing cells upon reinfection<sup>189</sup>.

There is another CD4+ T cell subset, known as CD4 CTLs, which play a role in immune responses by secreting granzyme B and perforin to kill target cells in an MHC class II-restricted manner. These cells are generated during chronic viral infections and anti-tumor responses, suggesting their involvement in host defense, though they may also contribute to autoimmune disease pathology. CD4 CTLs lack a unique surface marker, and their identification relies on a combination of surface markers and transcription factor expression patterns, indicating the existence of multiple subsets<sup>190</sup>.

### **1.2.3 Immunology in Lynch Syndrome**

Tumours associated with LS often demonstrate robust immune responses, as evidenced by dense infiltrates of TILs, especially CD8+ T-cells. These immune cells usually recognize FSDN that result from MSI foci and create memory cells that reside in the tissue, and can be detected not only in cancer survivors, but also as tissue-resident lymphocytes (TRL) in LS individuals without history of cancer<sup>57</sup>. RNA sequencing studies have shown that LS-associated adenomas exhibit elevated expression of immune-related genes, such as CD4, IFN- $\gamma$ , and TNF $\alpha$ , even in the early stages of tumorigenesis<sup>191</sup>. This immune activation is independent of mutational load, suggesting that immune surveillance in LS begins early and plays a significant role in delaying malignant transformation. High levels of TILs in LS-related CRCs have been associated with improved prognoses. However, this immune activation is not solely restricted to colorectal lesions; extracolonic tumours in LS patients also display similar immune characteristics.

#### *1.2.3.1 Immune surveillance in LS and antitumour immunity*

Several studies have shown that neoantigen-specific T-cells can be found in the blood and colorectal tissue of both healthy LS carriers without history of cancer and in LS cancer survivors<sup>56</sup>. Neoantigens derived from genes with high mutation frequencies, also exhibiting in vitro immunogenicity, are common to both LS-CRC and sporadic MSI-CRC, revealing a “shared landscape”. Among those are the neoantigens derived from mutations in the TGF $\beta$ R2 gene, which are present in over half of LS adenomas.

It is known that LS individuals have constant occurrence of MSI foci (or MMRd crypt focus)<sup>192</sup>, with FSDN being generated as early as haploinsufficiency of a single MMR gene, and CD8+ IELs being more abundant in these affected crypts. These observations,

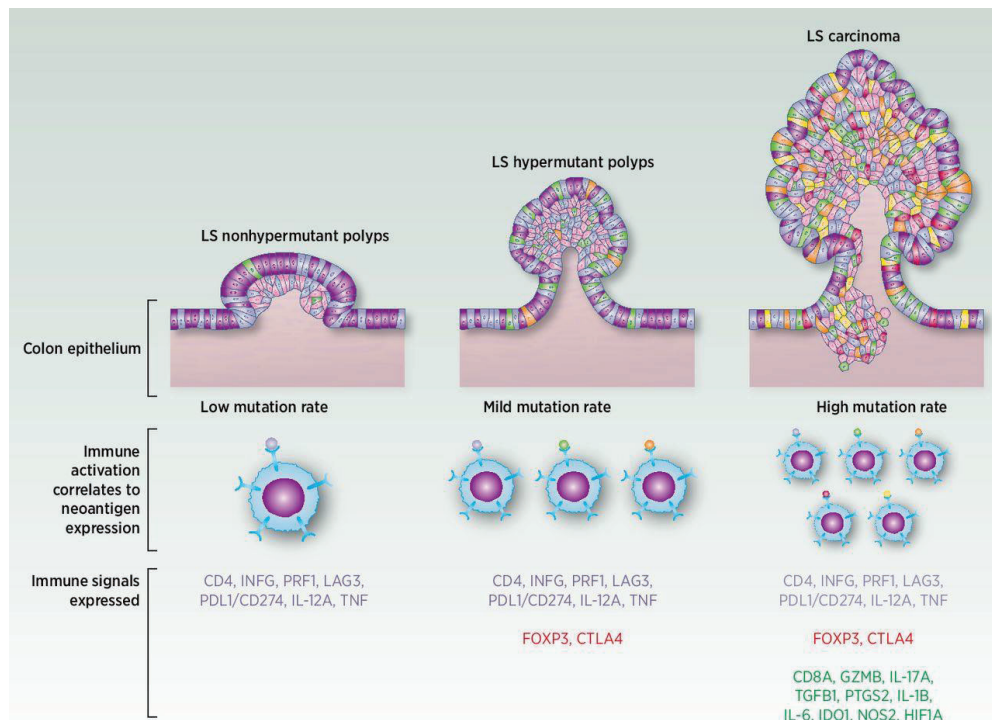
together with the sharp contrast with the low number of clinically manifest cancers in LS, suggest that there is a mechanism in constant play that controls these foci to avoid cancer progression. This mechanism is called immune surveillance, and it is part of the “elimination phase” theory, which explains the immune responses against neoantigens observed in healthy LS mutation carriers without tumours, proposing that immunoediting occurs early and consistently in LS, enabling efficient immune surveillance to eliminate many lesions at early stages.

Immune surveillance refers to the continuous monitoring of tissues by the immune system to detect and respond to potential threats, including malignant cells, and immunoediting is the process by which aberrant cell growth is halted and regressed by T-cell-mediated antitumour immunity. Mechanistically, the process involves the host immune system orchestrates dynamic interactions between tumour cells and non-tumour cells within the TME, initiating a cascade of events that drive this antitumour immunity: First, macrophages and neutrophils recognize and attack tumour cells, leading to cytokine-mediated tumour cell death. The release of cytokines then recruits circulating immune cells, enabling antigen processing by APCs, who then activate both naïve and primed CTLs, which further eliminate tumour cells through cell-mediated mechanisms. Afterwards, a small percentage of memory T-cells that stay in the previously affected tissues as TRLs<sup>81</sup>.

LS adenomas are not only distinguished by a higher number of infiltrating T lymphocytes compared to sporadic polyps, but they also show elevated expression levels of several immune-related genes, including CD4, IFN- $\gamma$ , LAG3, CD274/PDL1, IL12A, and TNF $\alpha$ , in comparison to non-LS polyps<sup>191</sup>, which occur independently of mutation rates and neoantigen load (Figure 12). Another study demonstrated that normal mucosa of LS carriers had increased immune cell density, including CD8+ T cells, NK cells, and B cells, which correlated to a longer time to advanced lesion<sup>193</sup>, altogether suggesting that immune activation is an early event in LS-associated tumorigenesis rather than a consequence of accumulating somatic mutations. This supports the idea of immunoprevention as a potential prophylactic approach for managing LS.

### 1.2.3.2 Immune escape

In cases where malignant cells are not fully eradicated, immunoediting is followed by the equilibrium and then immune escape phases, where the remaining cells are able to evade detection by the immune system. Immune editing exerts selection pressure on malignant cells, where those clones that are invisible to immune surveillance end up perseverating and ultimately forming pre-cancers or cancers. There are several mechanisms that contribute to immune evasion, such as the increase in Treg cells infiltration, particularly in hypermutated polyps, which may participate in immune suppression. A reduction in Treg density correlates with the presence of  $\beta$ 2M mutations, which are common in LS-CRC but rare in sporadic MSI-CRCs<sup>194</sup>.  $\beta$ 2M mutations lead to impaired HLA class I antigen expression, promoting tumour cell proliferation and immune evasion by reducing the immune system's ability to recognize tumour cells. Interestingly, despite contributing to immune evasion,  $\beta$ 2M mutations are associated with favourable outcomes, including lower rates of metastasis and relapse in LS-CRC. This may be due to the concurrent activation of NK cell-mediated apoptosis by “missing self” recognition, which helps control tumour growth<sup>195</sup>.



*Figure 12. Illustration of the relationship between immune activation (bottom track), neoantigen burden (middle track), and the adenoma-to-carcinoma progression in LS-related colorectal cancer (top track). In early adenomas (left column), markers of immune reactivity are present, even in the absence of significant somatic mutations or neoantigen load. As the lesions advance to more developed adenomas (middle column) and carcinomas (right column), there is a noticeable increase in mutation/neoantigen burden, alongside the emergence of immune tolerance markers. Adapted from<sup>81</sup>.*

Another immune evasion strategy involves the upregulation of PD-L1, a protein that suppresses immune responses. In LS-CRC, PD-L1 is more frequently expressed on immune cells at the tumour's invasive margin, particularly macrophages, but rarely on tumour cells themselves. This contrasts with sporadic MSI-H CRCs, where PD-L1 is expressed on both tumour cells and immune cells. The presence of PD-L1 on TILs in LS is linked to better prognosis and high levels of CD8+ T cells<sup>195</sup>.

In summary, immunogenic surveillance in LS-CRC can be divided into early and late stages. Early-stage surveillance is characterized by high CD3+ and CD8+ T cell infiltration, but frameshift mutations in the  $\beta 2M$  gene may lead to  $\beta 2M$  loss, impairing MHC class I presentation and reducing the effectiveness of T cell-directed therapies. In contrast, late-stage surveillance shows low baseline T cell infiltration, with the high mutation rate generating neoepitopes, leading to increased MHC class I presentation and T cell activation (Figure 12). However, this may also trigger PD-L1 upregulation on tumour cells, resulting in T cell exhaustion.

### 1.3 Cancer immunoprevention in LS

Preventive strategies can be categorized into different levels. Primary prevention aims to stop the onset of disease, encompassing both measures to protect healthy individuals and efforts to address pre-malignant conditions. Secondary prevention focuses on preventing recurrences or on detecting and managing early-stage non-invasive tumours. Tertiary prevention involves managing patients post-therapy to prevent the recurrence of their disease<sup>196</sup>. Here we will focus on primary prevention strategies.

In the early 20th century, Paul Ehrlich proposed that host immunity could prevent tumour formation, while William B. Coley observed tumour shrinkage in sarcoma patients after bacterial infections, leading him to develop Coley's Toxin, an early

immunotherapy. Observational studies also linked tuberculosis to reduced cancer incidence, prompting the use of Bacille Calmette-Guérin (BCG), a *Mycobacterium bovis* strain, as an attenuated vaccine for anti-cancer therapy<sup>197</sup>. BCG remains a clinical treatment for superficial bladder cancer and has shown a potential tumour-preventive effect in epidemiological studies. Experimental research in mouse sarcoma models provided early evidence of immune protection against tumours, demonstrating tumour-specific immunity mediated by lymphocytes.

With the advances in understanding immunoediting and immune escape, cancer immunoprevention in LS now focuses on two strategies, although chemoprevention by daily dosage of NSAIDs such as naproxen could be included as they have demonstrated to increase cytotoxic CD8+ infiltration in the colorectal mucosa, helping improve immune surveillance and reducing cancer incidence<sup>73</sup>. One approach involves targeting the adaptive immune resistance mechanisms that are implicated in MMRd colorectal adenomas, such as checkpoint molecules like PD-1, LAG3, and CTLA-4. Since inhibitors targeting the PD-1/PD-L1 axis have shown efficacy in advanced MMRd cancers, there is strong interest in evaluating their potential in preventing the progression of adenomas to carcinomas. Another promising strategy is the use of cancer vaccines, which could leverage the unique tumour-specific neoantigens generated in LS. Recent advancements in neoantigen identification, including the use of computational tools and the discovery of shared FSDN across various MSI tumours, have increased the interest in vaccine-based prevention.

### **1.3.1 Immunology of cancer prevention vaccines**

The main goal of cancer prevention vaccines (CPVs) is to enhance the immune system's ability to fight cancer by targeting antigens expressed by cancer cells that can be recognized as foreign, generating immune memory. Cancers progress from early pre-malignant lesions to advanced pre-malignant stages and, eventually, malignant tumours. Immune surveillance of cancer can result in three outcomes: elimination (complete immune system victory), equilibrium (a balance between the immune system and cancer), or escape (cancer evades the immune response). Cancer prevention vaccines mainly search to reinforce this immune surveillance by both boosting pre-existing peptide-specific memory T-cells and increasing their diversity, both systemically and resident on affected tissues (Figure 13). This may also occur by epitope spreading,

defined as the diversification of epitope specificity from the initial focused to others arising after malignant cell death.

Targeting both CD4<sup>+</sup> and CD8<sup>+</sup> T cells is crucial for effective cancer prevention vaccines due to their complementary roles in antitumour immunity. A major focus has been on MHC-I-restricted antigens; however, recent evidence underscores the importance of CD4<sup>+</sup> T cell responses in enhancing the efficacy of CD8<sup>+</sup> T cells. Although tumour immunity can occasionally occur without CD4<sup>+</sup> T cell help, most studies indicate that CD4<sup>+</sup> T cells are essential for priming and sustaining robust tumour-specific CD8<sup>+</sup> T cell responses and particularly important for generating long-term immune memory<sup>198</sup>.

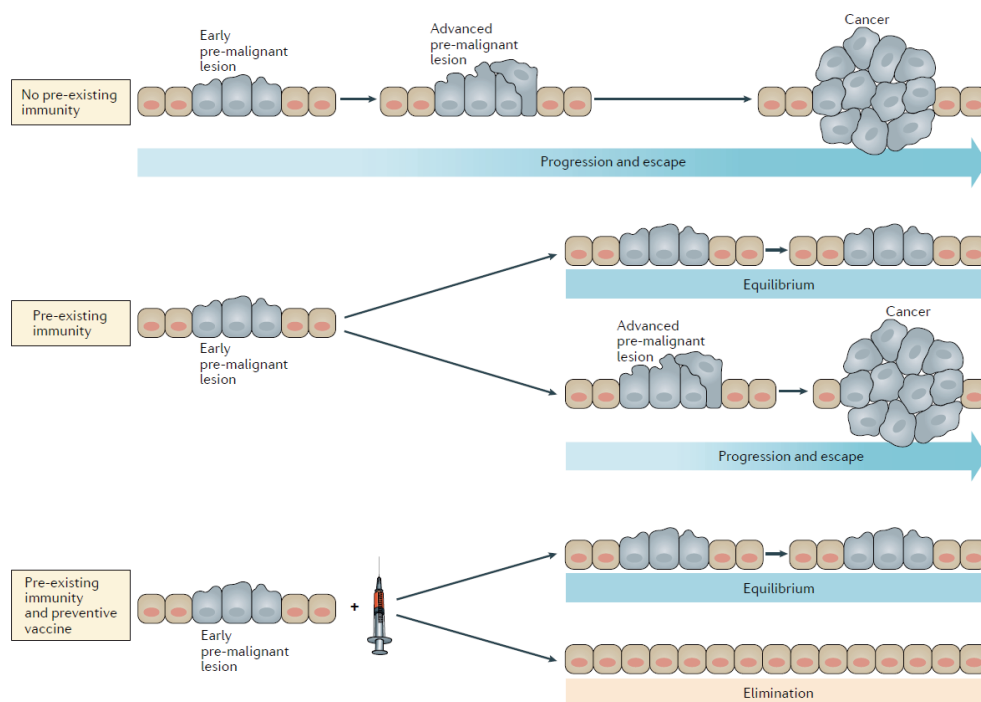


Figure 13. Effects of boosting antitumour immunosurveillance with a vaccine on disease outcome. Adapted from<sup>199</sup>.

Among others, there are two key factors that define how CPVs shape immune responses: A) The vaccine vector, which highly influences the proper priming of T-cells and the induction of long-lasting effector and memory CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses. B) The

antigen choice; For example, neoantigens are known to be more immunogenic than TAAs, as they are unique to cancer cells and less likely to cause autoimmunity<sup>200</sup>.

### 1.3.2 Vaccine vectors: Cellular vaccines

Selecting the appropriate vaccine vector or delivery system is pivotal for inducing effective immune responses in CPVs. Current delivery platforms include synthetic peptides, viral vectors, cellular vaccines, and nucleic acids (DNA and mRNA), each with distinct advantages and challenges (Figure 14).

|                         | Nucleic Acid vaccines                                                                                                                                                |                                                                                                                                                                                                      |                                                                                                                                                      | Peptides                                                                                                                                                                 |                                                                                                                    | Cellular                                                                                                                                                                   |                                                                                                                                                                    |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Formulation</b>      | <br>Naked DNA/mRNA                                                                  | <br>Nanoencapsulated DNA/mRNA                                                                                       | <br>Virus-Like particles (VLPs)                                     | <br>Epitopes                                                                            | <br>SLP (Synthetic Long Peptides) | <br>Dendritic cells                                                                     | <br>Cell lysate                                                                 |
| <b>Action mechanism</b> | Induction of host cells to express antigenic proteins/peptides that will be recognized by APCs and specific T-cells.                                                 |                                                                                                                                                                                                      |                                                                                                                                                      | In-vivo uptake and processing by APCs and subsequent priming of T-cells                                                                                                  |                                                                                                                    | Ex-vivo loading of autologous DCs with specific antigens (mRNA, DNA, peptides, cell lysates) and re-infusion to trigger specific immune responses                          |                                                                                                                                                                    |
| <b>Advantages</b>       | mRNA is Self-adjutant (Recognized by TLRs)<br><br>Easy purification and batch production<br><br>Can encompass a high numbers of antigens                             | Efficient packaging<br><br>Low toxicity profile<br><br>Protection against degradation<br><br>Facilitate endosomal escape<br><br>Can include selective ligands*                                       | Self-adjutant<br><br>Capacity of migration to lymph nodes<br><br>Induction of Th1 responses<br><br>Efficient packaging and uptake<br><br>Safety      | Can not require processing by APCs<br><br>Specificity<br><br>Easy batch production and storage<br><br>Safety, induction of CD4+ and CD8+ responses<br><br>Uptake by APCs | Not restricted to bind specific HLAs<br><br>Encompass several epitopes                                             | Safe, well tolerated, minimal side effects<br><br>Highly immunogenic, induction of complete immune responses<br><br>Can be loaded with various kinds and multiple antigens | Antigen accessibility<br><br>Target antigens do not have to be prospectively identified<br><br>Can be genetically modified to secrete immune stimulatory cytokines |
| <b>Disadvantages</b>    | Inefficient uptake<br><br>Highly susceptible to degradation<br><br>DNA: low immunogenicity, can integrate in the genome<br><br>Difficult and uncomfortable injection | Poor adjuvant activity<br><br>*Cell targeting still remains a challenge in human trials<br><br>Challenging manufacturing and scaling-up process<br><br>Aggregation, instability, and rapid clearance | Immune memory-driven neutralization of second doses<br><br>Structural instability due to lack of viral genome<br><br>Challenging manufacture process | Restriction to specific HLAs<br><br>Require adjuvants<br><br>Challenging manufacture process due to specific aminoacids<br><br>Inefficient delivery                      | Variability on induced immune responses                                                                            | Manufacture is costly and requires specialized infrastructures and personnel<br><br>Individualization, does not allow batch production                                     | Requires a lesion sample - not optimal for CPVs<br><br>Limited clinical efficacy                                                                                   |

Figure 14. Advantages and disadvantages of cancer prevention vaccine vectors. Created with BioRender, published in<sup>200</sup>.

While mRNA, DNA, and peptide vaccines offer flexibility and ease of design, their efficiency can be limited by rapid systemic dissemination and poor targeting of APCs. Encapsulation techniques (e.g., lipid nanoparticles) and the use of immunogenic adjuvants like poly ICLC or GM-CSF have improved their stability and immune activation.

Viral vectors, including virus-like particles, leverage their inherent immunogenicity to introduce genetic material into cells effectively. Although they show promise, challenges such as immune system neutralization of the vectors remain. Prime-boost strategies have been explored to mitigate these limitations, combining initial antigen priming with subsequent boosting phases to amplify immune responses. For instance, the PROSTVAC vaccine demonstrated potential in early-stage clinical trials but requires further optimization for broader efficacy.

Cellular-based vaccines encompass the use of whole-tumour-cell lysates, antigen-loaded APCs, or a combination of both, with DCs being the most commonly employed. In prophylactic settings except in pre-malignant lesions, secondary and tertiary prevention levels, tumour cell lysates are generally not available, making the use of autologous antigen-loaded cells more practical.

DCs have been widely explored in cancer immunotherapy, with over 200 clinical trials completed<sup>154</sup>. These strategies typically involve the *ex vivo* generation or isolation of autologous DCs, either from circulating blood or monocyte-derived sources obtained via leukapheresis. The DCs are then matured and primed with antigens, which may include mRNA, DNA, peptides, or tumour cell lysates, before being reinfused into the patient<sup>201</sup>. Extensive clinical experience with DC-based vaccines has demonstrated their safety, with minimal side effects, predominantly grade I and II adverse events and only 3–4% grade III adverse events. Sipuleucel-T (PROVENGE; Dendreon), a vaccine based on DCs pulsed with GM-CSF fused to prostatic-acid phosphatase (PAP), a protein expressed by prostate cancer cells, has been the first therapeutic cancer vaccine to be approved by the US Food and Drug Administration<sup>202</sup>.

These vaccines are highly immunogenic, capable of inducing robust CD4+ and CD8+ T cell responses and have shown notable response rates in therapeutic applications. However, their clinical effectiveness remains limited. Factors such as immune exhaustion in patients and the immunosuppressive *in vivo* environment that affects DC viability and functionality contribute to this limitation. Nevertheless, the full potential of DC-based vaccines is likely to be realized in preventive settings, where these limiting factors have not yet manifested.

### 1.3.3 Antigen choice

#### 1.3.3.1 *Tumour-Associated Antigens*

TAAAs are molecules derived from self-antigens that are selectively or overexpressed in tumour cells, such as MUC1 and HER2<sup>203,204</sup>, which are often aberrantly expressed in cancers. While TAAAs have shown promise in therapeutic cancer vaccines, their application in primary prevention is more challenging due to potential risks of autoimmunity and limited immunogenicity against healthy tissues. For example, MUC1, overexpressed in epithelial cancers, has been explored in therapeutic settings, demonstrating immune responses but limited clinical efficacy due to tumour immunosuppression. Preventive trials, like the MUC1 peptide-poly-ICLC vaccine for CRC prevention, have shown immune activation in high-risk patients, indicating potential in earlier disease stages<sup>205</sup>. HER2 is overexpressed in early breast cancer lesions, and vaccines targeting it have demonstrated safety and immune activation, though clinical efficacy remains variable<sup>206–208</sup>.

#### 1.3.3.2 *Cancer-Testis Antigens*

CTAs are a subset of TAA with restricted expression in healthy tissues, primarily germ and placental cells, but are aberrantly expressed in various cancers, making them promising targets for cancer prevention vaccines. Key CTAs such as MAGE-A, PRAME, and NY-ESO-1 are widely studied due to their immunogenicity. NY-ESO-1 vaccines have shown significant potential by eliciting strong CD4+ and CD8+ immune responses in esophageal, lung, and ovarian cancers, among others<sup>209</sup>. PRAME-targeting immunotherapies also demonstrated specific antibody and T-cell responses in melanoma and lung cancer trials<sup>210</sup>. While MAGE-A vaccines initially faced limited clinical success<sup>211,212</sup>, new combinatorial approaches, such as pairing them with immune checkpoint inhibitors, are restoring their antitumour efficacy<sup>213,214</sup>. These vaccines hold particular promise for high-risk individuals, as their restricted expression pattern minimizes off-target effects.

#### 1.3.3.3 *Neoantigens*

Neoantigens are identified as foreign proteins that are absent in normal tissues, but can arise from malignant cells through various mechanisms, such as genomic mutations

(including InDels, which is the case on LS), aberrant transcriptomic variants, post-translational modifications (PTMs), and viral ORFs. These peptides provide a distinct advantage as therapeutic targets due to their unique tumour-specific nature and absence from normal tissues. Moreover, neoantigen-specific T cells can evade negative selection in the thymus, as the somatic tumour mutations generating these highly antigenic neoantigens are not present during T cell development. This ability to bypass central T cell tolerance expands the pool of neoantigen-specific T cells, thereby enhancing tumour-specific immune responses. Additionally, the durable and robust neoantigen-specific T cell responses stimulated by immunotherapy, coupled with the development of posttreatment immunological memory, hold promise for long-term protection against disease recurrence<sup>58</sup>.

Neoantigen-based vaccines can be either personalized or off-the-shelf. Personalized vaccines are developed using tumour RNA and DNA sequencing, combined with *in silico* prediction, to target neoantigens unique to an individual's mutations and HLA haplotype. In contrast, off-the-shelf vaccines are designed to target recurrent, shared mutations—often in oncogenes—allowing broader applicability across diverse patient populations. MSI-H tumours, characterized by a high mutational burden and the generation of FSDNs (as seen in previous sections), are particularly suitable for such approaches<sup>200</sup>.

Early clinical trials have demonstrated the feasibility, safety, and immunogenicity of neoantigen vaccines<sup>57,215–218</sup>, however, tumours with low mutational loads (MSS or MSI-low, which encompass around 85% of global tumours) are usually incompatible with this strategy.

Neoantigen-based CPVs are particularly promising in hereditary cancer syndromes like LS, where patients have a high risk of developing MSI-H cancers. Vaccines targeting shared FSPs in this population can complement existing prevention measures and address cancers for which current preventive strategies are inadequate.

### 1.3.4 Identification of candidate neoantigens

The workflow of neoantigen identification usually starts with bioinformatics prediction based on tumoral or pre-cancerous sequences. *In silico* prediction strategies consider multiple biological factors, from identifying somatic mutations to assessing MHC-TCR

interactions. Typically, tumour-normal DNA sequencing (via whole-exome or whole-genome sequencing) is analysed using a variant caller software to detect mutations. Simultaneously, the patient's HLA alleles are determined by HLA haplotyping. Ideally RNA-sequencing from the same sample is performed to aid in validating gene expression, identifying alternative splicing isoforms, gene fusions, and RNA editing events that DNA sequencing alone cannot detect<sup>219</sup> (Figure 15).

After mutation analysis, potential neoantigens are identified through bioinformatics pipelines incorporating binding prediction algorithms for MHC class I and II molecules. Additional predictors assess gene expression patterns, RNA splicing, peptide processing (e.g., proteasomal cleavage and TAP transporter efficiency), and secondary structure degradation for MHC-II peptides. Notably, including predictors of mRNA decay escape has been linked to enhanced antitumour immunogenicity<sup>219,220</sup>.

Most studies emphasize MHC-I-restricted predictions due to the limited accuracy of MHC-II prediction algorithms. The most effective current methods for predicting MHC-II restricted neoantigens depend on tools like netMHCII-2.3 and netMHCIIpan-3.2, which exhibit inaccuracies partly due to the open structure of the MHC-II binding groove, resulting in considerable epitope length and anchor points variability. New models are being developed to solve these challenges, such as hmMHC, which inherently accommodate peptide sequences of variable length and are trained on recent Immune Epitope Database (IEDB) content<sup>198</sup>.

Mass spectrometry has also significantly improved prediction accuracy by directly identifying peptides bound to MHC proteins. However, these enhancements are still limited to certain alleles, necessitating further research to expand allele coverage and refine antigen presentation and immunogenicity predictions<sup>221</sup>.

While *in silico* predictions are essential for identifying and prioritizing neoantigen candidates, they do not confirm the ability to induce specific T-cell responses. Typically, an immunogenic T-cell priming results in a Th1-like TNF- $\alpha$  and/or IFN- $\gamma$  producing responses, IL-2 release, T cell proliferation, and the acquisition of cytolytic activity in the case of CD8+ T cells.

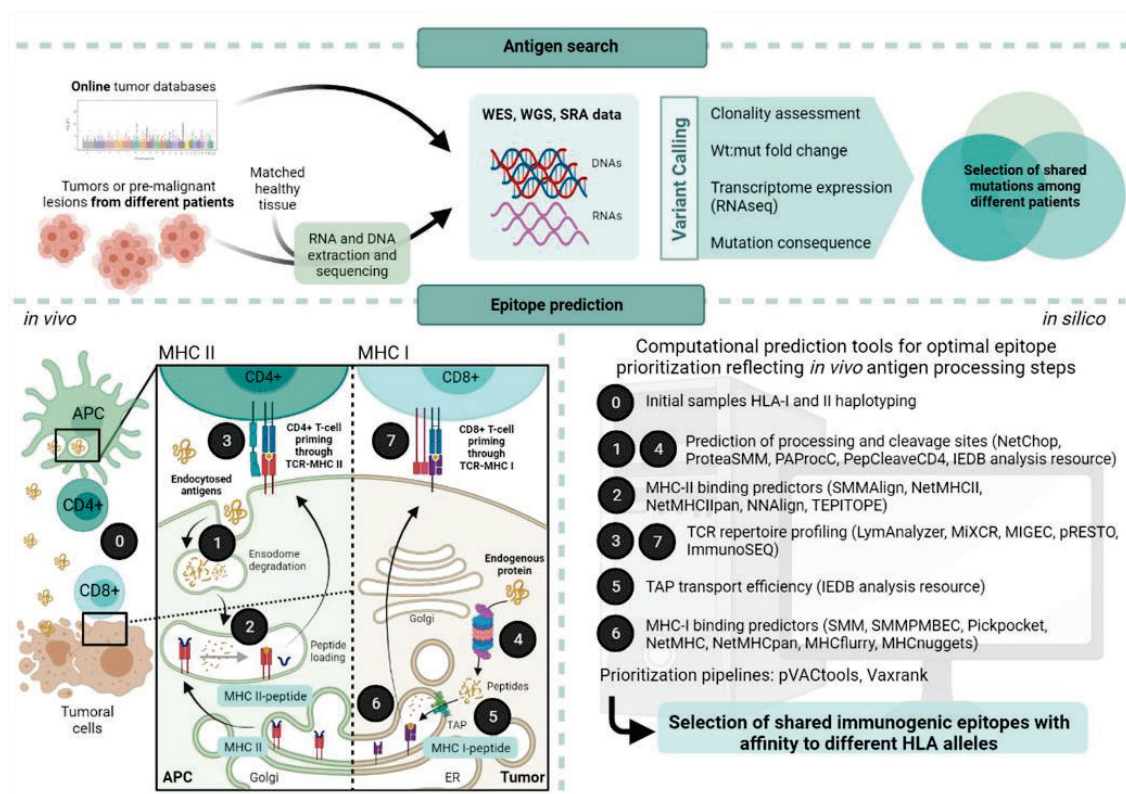


Figure 15. Schematic representation of antigen prediction pipeline for the design of CPVs. Above: Workflow to determine shared mutations among different tumours and patients for a preventive and off-the-shelf approach. Below: Biological key steps in antigen presentation pathways (MHC-I and MHC-II) to consider when performing *in silico* epitope predictions. Abbreviations: WES, whole exome sequencing; WGS, whole genome sequencing, SRA, short read alignment; Wt, wildtype. Created with BioRender, published in<sup>200</sup>.

To correctly identify neoantigens that elicit such responses, functional validation is required, typically through *in vitro* assays assessing several aspects such as cytokine production (e.g., ELISA, ELISPOT, Luminex, intracellular staining), peptide-MHC binding using multimer complexes, T-cell activation via flow cytometry using common AIMs, immunohistochemistry, and RNA analysis<sup>222</sup>.

### 1.3.5 State of the art in neoantigen-based cancer prevention vaccines

Immunization with neoantigens derived from recurrent FS mutations has recently gained interest as an approach to boost immune surveillance of precancerous MMRd cell

clones. Recently, it has been shown that vaccination against only 4 FSDN prevents tumours in a mouse model with *VCMSH2* mice, who have a conditional deactivation of the *MSH2* gene in the intestinal tract and spontaneously develop intestinal cancer<sup>83</sup>. This prevention potency was enhanced by daily naproxen treatment, and at a lesser rate ASA (acetylsalicylic acid).

In humans, the phase I/IIa clinical trial of the MicOryx vaccine (NCT01461148) demonstrated that their vaccine based on three FSDNs in LS, combined with an adjuvant (Montanide), induced primarily CD4+ specific responses after vaccination in patients with a history of advanced MMRd CRC (19 out of 22 patients entered the study without active disease)<sup>57</sup>. No grade 3 or 4 adverse effects were observed, and grade 1-2 effects were limited to injection site reactions and flu-like symptoms. To date, only the evaluation of tumour response in 3 patients has been published, with 2 of them showing stable disease. No metastasis was detected in any of the patients during the study period.

The National Cancer Institute (NCI), in collaboration with the company NousCom, is conducting an active recruitment phase clinical trial (NCT05078866). This Phase Ib/II study is vaccinating healthy individuals with LS with an adenoviral vaccine loaded with 209 FSDNs in mRNA format (Gad-209-FSP / Gad20-209-FSP / MVA-209-FSP) for cancer prevention, with the primary objectives of determining vaccine safety and assessing immune responses to the vaccine antigens<sup>82</sup>.

The research group led by Dr. Jolanda M. de Vries at Radboud University Medical Center has developed a vaccine of autologous DCs loaded with neoantigens in the form of HLA-A\*02:01-restricted epitopes derived from alterations in the MMR pathway (TGF- $\beta$ R2 and Caspase-5) and related to LS. This open-label phase I/II clinical trial (NCT01885702) administered the product to patients with CRC with MMRd and healthy LS carriers, with the main objectives of determining safety and induction of peptide-specific immune responses. Preliminary results from this trial, which are in the process of being published, show that out of 20 individuals with LS, only 1 patient experienced a grade 4 adverse effect (fever of  $>40^{\circ}\text{C}$  for more than 24h) and immune responses were observed in all vaccinated patients, with neoantigen-specific T-cells being detected in the dermis near the injection site in 15 patients. Moreover, those individuals that had

responses to the TGFBR2 neoantigen, have shown significant increase in disease-free survival compared to non-responders.

## **2 *HYPOTHESIS AND OBJECTIVES***

### **2.1 Hypothesis**

The hypothesis of this thesis is that DC vaccination targeting shared LS tumour neopeptides in healthy individuals will enhance immune surveillance by diversifying and amplifying pre-existing or de novo induced neoantigen-specific memory T-cell responses, thereby preventing the development of cancerous lesions. By offering broader protection against malignancies, this approach has the potential to improve the quality of life for individuals at risk.

Contextualizing our hypothesis, recent advances in cancer immunotherapy have shown that DC-based vaccines can effectively stimulate the immune system to recognize and target TAAs. LS, characterized by mutations in MMR genes, leads to MSI and the generation of FSDNs, which are highly immunogenic. While current prophylactic measures for LS primarily involve regular surveillance and surgical interventions, these are only effective for CRC and in less measure, for EC. There are no current prevention strategies for all other LS-related tumours and for those that escape current measures. The development of a vaccine targeting LS-related FSDNs represents a less invasive strategy that harnesses the host's immune system. This approach aligns with the growing body of evidence suggesting that neoantigen-specific immune responses can provide effective prophylaxis or therapy in individuals predisposed to malignancies. Importantly, these immunizations will be administered prior to the transformation of cells into a tumorigenic state, thereby circumventing the tolerogenic and immunosuppressive effects exerted by solid tumours on T lymphocytes.

## 2.2 Objectives

The primary objective of this thesis is to develop a LS-related FSDN-loaded autologous dendritic-cell based vaccine to immunize LS individuals against common FSDN, and to test the hypothesis that vaccination with DCs may be as effective as, or superior to, current prophylactic measures for cancer prevention in LS.

Specifically, the objectives of this thesis are:

- To identify specific FSDNs associated with the deficient DNA repair pathway, based on literature review and bioinformatic prediction of their interactions with common HLA class I and II alleles.
- To validate the predicted FSDNs in silico by assessing the presence of FSDN-associated mutations in the DNA and RNA of LS-derived cancers, pre-cancers, non-LS MSI cancers, and tumor cell lines.
- To functionally validate and evaluate in vitro the antigenic recognition of FSDNs by T-lymphocytes isolated from normal and adenomatous colorectal mucosa biopsies of patients with LS.
- To prioritize an optimal set of FSDNs based on their frequency in evaluated samples and their ability to elicit antigenic responses in LS-derived lymphocytes.
- To validate GMP-compliant manufacturing of the vaccine product in preparation for a clinical trial in individuals with LS, aimed at evaluating the safety and immunogenicity of autologous DCs loaded with the selected FSDN peptides.

## 3 MATERIALS AND METHODS

### 3.1 Samples and sequences sources

#### 3.1.1 Bioinformatic sequences

Sequences for the bioinformatic analysis have been obtained both from public and private informatic databases, as well as from sequencing of patients' samples obtained from the Gastroenterology Department of Hospital Clinic of Barcelona (HCB). Below are the sourced from all sequences processed, as well as the methodology for their analysis.

##### 3.1.1.1 Data collection for FSDN predictions

Neoantigens prediction was performed using 531 frameshift peptide sequences derived from *m1* (deletion of 1 nucleotide) and *m2* (deletions of 2 nucleotides) mutations in 269 coding microsatellite(cMS)-containing sequences retrieved from the Selective Targets in Human MSI-H Tumorigenesis Database (Seltarbase<sup>223</sup>, [www.seltarbase.org](http://www.seltarbase.org), accessed September of 2021) and different bibliography sources (see references<sup>56,80</sup>). As the Seltarbase database is no longer available and cannot be accessed, the initial sequences selected from all sources used are provided in Supplementary table 1. We included cMS with a length of at least eight bases, cMS in putative tumour-driver genes and oncogenes (according to recent standardized lists of cancer driver genes<sup>224–226</sup>) and cMS known to produce frameshift peptides with predicted high affinity ligands according to literature. Eight wild-type amino acids upstream adjacent to the beginning of the frameshift peptide sequence were included in all sequences to also assess possible epitopes located at the junction between N-terminal wild type and C-terminal mutant peptide sequences.

##### 3.1.1.2 Colorectal adenomas (CrAD) and ECs FFPE samples source

Formalin-fixed paraffined-embedded (FFPE) samples from 17 advanced CrADs and 8 ECs from 8 patients with LS were obtained from the Gastroenterology and Gynaecology Department of HCB (Table 5). All CrAD and ECs displayed MMR deficiency by immunohistochemistry, as reported by the anatomic pathology lab from the HCB, showing loss of MMR proteins expression. In addition, the HLA haplotype of the patients included in the study has been

determined by high-resolution NGS typing for the HLA-A, HLA-B, HLA-C, HLA-DR1, HLA-DQ1, and HLA-DP1 loci.

#### 3.1.1.3 *Data source for CRC and CrAD DNA sequences*

We obtained access to WGS data from the NIH study HNPCC-Sys: Molecular Characterization of LS (phs001407.v1.p1; dbGaP, TCGA) carried out by the HNPCC-Sys consortium. This study was supported by the German Federal Ministry of Education and Research with the PTJ grant HNPCC-Sys 031 6065A. The study gave access to 98 samples from CrADs (n=10), CRC (n=40), and paired healthy samples (n=48) of 11 patients with LS (The metadata of this distribution set can be accessed freely under the accession code phs001407.v1.p1.c1, link: [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001407.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001407.v1.p1)). The FASTQ files obtained for WGS sequences from HNPCC-Sys were transferred to the Barcelona Supercomputing Centre (BSC) central storage for read pre-processing and quality control.

#### 3.1.1.4 *Data source for sporadic MMRd genetic variants*

Annotated frameshift variant files from MSI-H, MSI-Low (MSI-L) and MSS exome sequencing samples were obtained from different projects with high prevalence of MSI in the TCGA database (genomics data portal database, NCI), including TCGA-BRCA (Breast Cancer, n=15), TCGA-CESC (Cervix Uteri cancer, n=8), TCGA-COAD (Colorectal Adenocarcinoma, n=84), TCGA-READ (Rectum neoplasms and adenomas, n=11), TCGA-STAD (Stomach cancer, n=84) and TCGA-UCEC (Corpus Uteri cancer, n=164), either by direct download from their domain or from the study in reference<sup>227</sup>. Status of MSI, if not provided by the source, was established by the MSI detection tool MANTIS<sup>228</sup>.

#### 3.1.1.5 *Data source for MSS and MSI cell lines*

We selected 141 cell lines (CL) and organoids from CRCs, ECs and normal tissue with available variant calling files at the Cell Model Passports website (<https://cellmodelpassports.sanger.ac.uk>; version 2.7.0) from the Sanger Institute<sup>229</sup> (Supplementary table 2). 43 CL are MSI, 91 are MSS and 7 are from normal tissue. All CL had mutation data available, and we retrieved RNA-seq (Illumina 2000 or 6000) data from 27 of them to analyse FSDN expression from several projects from the public functional genomics data repository Sequence Read Archive (NCBI, NIH).

### 3.1.2 Biopsies and Peripheral blood

Individuals with LS or healthy controls followed in the high-risk clinic for colorectal cancer at the Gastroenterology Department of the HCB have been recruited based on their availability for colonoscopy and germline mutation: Patients with germline variants other than *MLH1*, *MSH2* or *MSH6* have been excluded, with one exception on a *PMS2* pathogenic variant carrier. The study was approved by the HCB Institutional Review Board (HCB/2019/0213), and all patients signed informed consent.

During surveillance colonoscopies, 40 mL of peripheral blood and biopsies of healthy mucosa, adenomatous lesions and CRC (if available) from each individual have been retrieved. Samples were processed fresh, biopsies were transported in RPMI media (Gibco, Spain) supplemented with antibiotics (1% Penicillin/Streptomycin, Gibco, Spain). In addition, the HLA haplotype of the patients included in the study has been determined by high-resolution NGS typing for the HLA-A, HLA-B, HLA-C, HLA-DR1, HLA-DQ1, and HLA-DP1 loci.

### 3.1.3 Cell lines

K562-HLA null or GLE799 pTRPE eGFP-T2A-b2m-transfected cell lines with different HLA alleles were kindly gifted from the Linette/Carreño Lab, Center for Cellular Immunotherapies, University of Pennsylvania. Table 4 summarizes all K562-based cell lines available.

Table 4. K562 cell lines and derived cell lines available in the HCB.

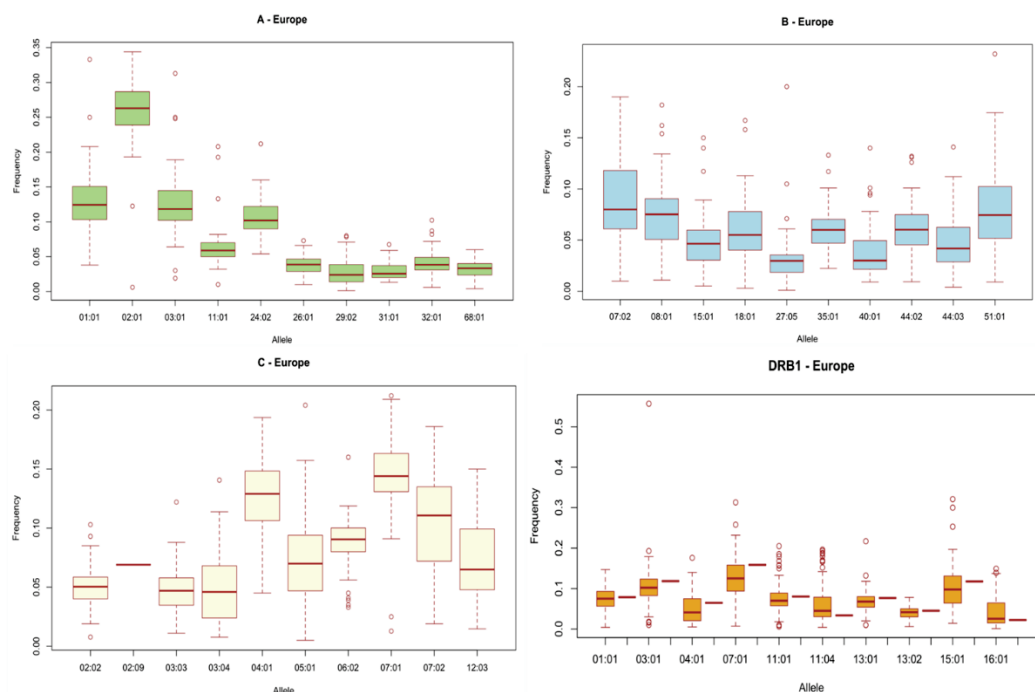
| K562 derived cell line | HLA class I                 |
|------------------------|-----------------------------|
| K562                   | HLA class I null            |
| K799                   | GFP-T2A-b2m-HLA-A*03:01 SCD |
| K809                   | GFP-T2A-b2m-HLA-A*11:01 SCD |
| K810                   | GFP-T2A-b2m-HLA-A*24:02 SCD |
| K811                   | GFP-T2A-b2m-HLA-B*07:02 SCD |
| K814                   | GFP-T2A-b2m-HLA-B*08:01 SCD |
| K815                   | GFP-T2A-b2m-HLA-A*02:01 SCD |

Jurkat and HEK 293T cell lines were purchased from the American Type Culture Collection (ATCC).

## 3.2 Bioinformatic analysis

### 3.2.1 Neoantigen prediction

Entries derived from selected cMS (see section 3.1.1.1) were processed by the pVACbind prediction pipeline (pVACtools v3.1.0<sup>230</sup>) for neoantigen prediction against the 40 most common HLA class I (30) and II (10) alleles in the European population<sup>231</sup> (**Figure 16**). The following options were enabled: Epitope length from 8 to 11 aa for MHC-I predictions and 15 aa for MHC-II predictions; binding threshold: IC50 <5000. The remaining options were left by default. All available binding-prediction algorithms were used (MHCflurry, MHCnuggetsI, NetMHC, NetMHCcons, PickPocket and NetMHCpan for HLA-I, and MHCnuggetsII, NetMHCIIpan and NNalign for HLA-II) and the final binding score was considered as the median score from all predictors.



**Figure 16.** Overall estimation of frequencies for the 10 most common HLA-A, HLA-B, HLA-C and HLA-DRB1 alleles in Europe, adapted from the Allele Frequency Net Database (<https://www.allelefrequencies.net/>).

### 3.2.1.1 Data filtering and neoantigens prioritization

The output results of the pVACbind pipeline were classified as strong ( $IC_{50} < 50nM$ ), weak ( $IC_{50} < 500nM$ ) and very weak ( $IC_{50} < 5000nM$ ) affinity binders to HLA and filtered by the following criteria, which aimed to integrate a high population coverage and a high mutation frequency on tumours: A) *m1* or *m2* mutations from each gene were chosen based on mutation frequency reports from Ballhausen *et. al.*<sup>56</sup> and TCGA database. In the case of genes without frameshift mutation frequency reported, we selected the epitopes derived from *m1* mutations since they are the most frequent events in MSI tumours; B) For each gene, we chose the most representative epitopes in terms of HLA-allele coverage (one for HLA-I and one for HLA-II) that also fulfilled the other stated criteria. For epitopes that tied the number of covered alleles, we prioritized those with an overall lower  $IC_{50}$  value. Finally, pathogen and self-similarity, as well as nonsense-mediated RNA decay (NMD) filters were applied (see following sections).

### 3.2.1.2 NMD

The R package *masonmd*<sup>232</sup> (MAke Sense Of NMD) was used to predict the effect of nonsense-mediated decay on the mutated genes. Some mutations can introduce premature termination codons (PTCs) into genes, NMD can detect these PTCs and then eliminate the abnormal transcripts, thus playing an important role in expressed cancer neoantigens. *Masonmd* uses three rules to predict whether a PTC-generating mutation is NMD-elicited or NMD-escape: A) PTC is more than 50-54bp upstream of the last-exon-exon junction; B) targeted gene is not intronless; C) PTC is more than 200bp downstream of the start codon. Using these rules, we predicted whether a called mutation will elicit NMD on the mRNA from the mutated gene, i.e. the NMD-elicited mutations, and corroborated this through public databases with transcriptomic data (SeltarBase, Version 201307, [www.seltarbase.org](http://www.seltarbase.org)).

### 3.2.1.3 Pathogen and self-similarity

Peptide sequences from prioritized neoantigens were compared to the human reference proteome (NCBI Protein Reference Sequences) or pathogen-related peptide sequences with positive immune assays derived from the Immune Epitope Database (IEDB) through a protein BLAST for short sequences (BLASTp-short). A normal distribution was assessed to determine a cutoff e-value and bit-score. Considering the length of the queries and the input sequences, we determined a cut-off for the e-value of 0.5 for pathogen-similarity and 0.1 for self-similarity. Cutoff for bit-score was established above 15. The cut-off was established by

plotting the density distribution of the scores using the *density* function in the R package “stats” and by applying the critical value approach (Figure 17). For self-similarity, neoepitope sequences that included 6 or more *wt* aminoacids were not considered.

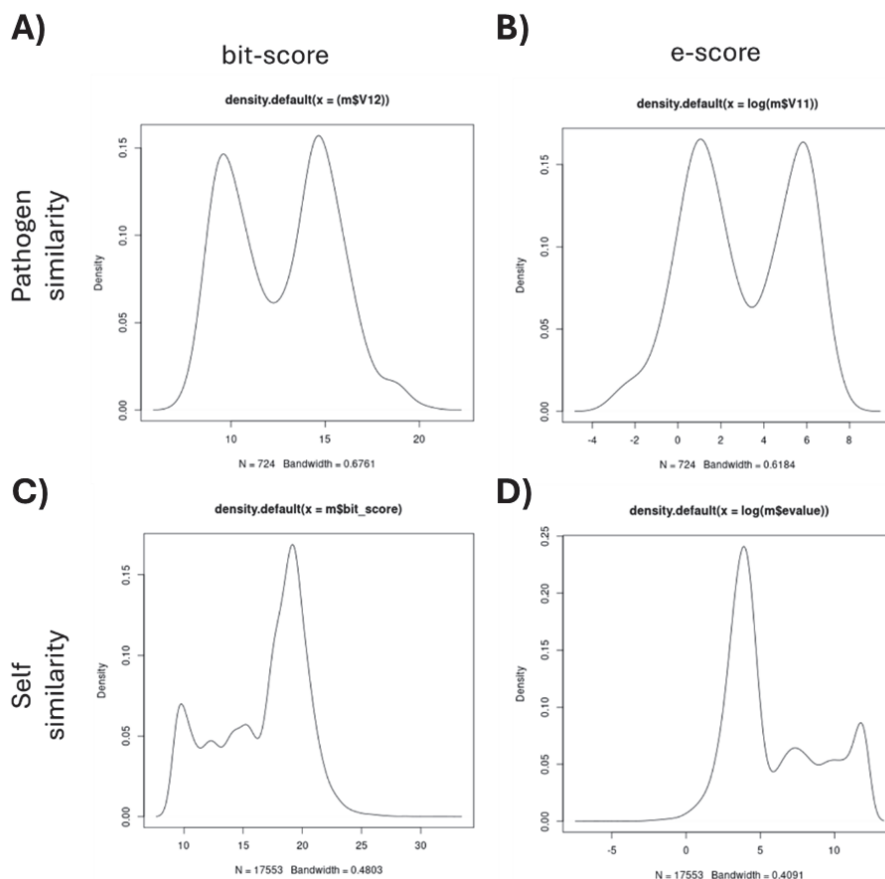


Figure 17. Plots of density distribution of pathogen (A and B) and self-similarity (C and D) bit (A and C) and e scores (B and D) with the density function in the R package “stats”.

### 3.2.2 Library preparations and sequencing samples from LS cancer patients

Deparaffinization and DNA extraction of FFPE CrAD and EC samples from LS patients was performed with the QIAamp DNA FFPE Tissue Kit (Qiagen) following provider directions. Quality control for NGS libraries was performed through High Sensitivity DNA electrophoresis with TapeStation HS D1000, where gDNA DIN (DNA Integrity Number) was analysed with a control of a good quality sample (DIN=6.7).

A custom panel targeting 192 microsatellites was designed using HaloPlex HS technology (Agilent Technologies, CA, USA) for highly sensitive MSI (hs-MSI) assessment<sup>233,234</sup>. The hs-MSI panel included 93 cMS, 27 of which matched our pFSDNs (corresponding to 40 pFSDN). Libraries were prepared according to the manufacturer's instructions by using 150-300 ng from FFPE samples. Library quantification and quality control assessment were performed by using TapeStation 4200 (Agilent Technologies, CA, USA). Pooled libraries with a minimum concentration of 20 nM were sequenced in a HiSeq platform at high coverage, reaching a mean depth of around 932 reads/sample after deduplication. Only individual markers reaching a minimum coverage of 100X were evaluated.

The obtained targeted sequence data was computed through a high-sensitivity software for the detection of insertions and deletions<sup>233,234</sup>. The hs-MSI score, representing the percentage of unstable microsatellites out of the total number of analysed microsatellite markers, was calculated per case sample with the formula: MSI score= total unstable MS/(total sequenced MS)\*100. An hs-MSI score above the previously established threshold of 4.576, was considered indicative of MSI. The HLA haplotype from each included patient was determined by high-resolution NGS typing for the HLA-A, HLA-B, HLA-C, HLA-DR1, HLA-DQ1, and HLA-DP1 loci in the molecular biology Core from the HCB.

*Table 5. Characteristics of the EC and CrAD samples from the HCB biobank included for the targeted cMS sequencing.*

| Sample ID    | Sample Type | Histology | DNA MMR proteins expression loss | Total MS sequenced | n° Instable MS | MSI score <sup>1</sup> | % Neoplastic Representation |
|--------------|-------------|-----------|----------------------------------|--------------------|----------------|------------------------|-----------------------------|
| B05-20234_EC | EC          | EA        | <i>MSH2</i>                      | 32                 | 15             | 46,88                  | 40%                         |
| B06-13452_EC | EC          | EA        | <i>MLH1</i> and <i>PMS2</i>      | 33                 | 13             | 39,39                  | 60%                         |
| B06-21517_EC | EC          | EA        | <i>MLH1</i> and <i>PMS2</i>      | 71                 | 35             | 49,3                   | 35%                         |
| B08-12823_EC | EC          | EA        | <i>MSH6</i>                      | 34                 | 8              | 23,53                  | 25%                         |
| B09-19814_EC | EC          | EA        | <i>MSH2</i> and <i>MSH6</i>      | 15                 | 4              | 26,67                  | 30%                         |
| B12-06384_PC | CrAD        | TA + LGD  | <i>MSH6</i>                      | 95                 | 26             | 27,37                  | 30%                         |
| B12-35459_PC | CrAD        | TVA + HGD | <i>MSH2</i> and <i>MSH6</i>      | 38                 | 25             | 65,79                  | 20%                         |
| B13-22339_PC | CrAD        | TA + LGD  | <i>MSH2</i> and <i>MSH6</i>      | 92                 | 60             | 65,22                  | 30%                         |
| B13-33741_PC | CrAD        | TA + LGD  | <i>MSH2</i> and <i>MSH6</i>      | 149                | 112            | 75,17                  | 10%                         |
| B17-06323_EC | EC          | IEA       | <i>MSH2</i> and <i>MSH6</i>      | 146                | 66             | 45,21                  | <20%                        |
| B17-12278_EC | EC          | EA        | <i>MSH6</i>                      | 80                 | 30             | 37,5                   | 40%                         |
| B19-03895_EC | EC          | EA        | <i>MLH1</i> and <i>PMS2</i>      | 172                | 70             | 40,7                   | <10%                        |

|              |      |                        |                             |     |     |       |             |
|--------------|------|------------------------|-----------------------------|-----|-----|-------|-------------|
| B99-10652_PC | CrAD | TA + LGD               | <i>MSH2</i> and <i>MSH6</i> | 75  | 53  | 70,67 | 20%         |
| B99-10653_PC | CrAD | TA + LGD               | <i>MSH2</i> and <i>MSH6</i> | 120 | 86  | 71,67 | 30%         |
| BP-FB-1_PC   | CrAD | TA + LGD               | <i>MSH6</i>                 | 179 | 54  | 30,17 | <10%        |
| BP-FB-10_PC  | CrAD | TA + LGD               | <i>MSH2</i>                 | 184 | 138 | 75    | 10%/5-10%   |
| BP-FB-12_PC  | CrAD | TA + LGD               | <i>MSH2</i>                 | 20  | 2   | 10    | 5%          |
| BP-FB-2_PC   | CrAD | TA + LGD               | <i>MSH6</i>                 | 176 | 40  | 22,73 | 10%         |
| BP-FB-3_PC   | CrAD | TA + LGD               | <i>MSH6</i>                 | 138 | 69  | 50    | 40%/10%     |
| BP-FB-4_PC   | CrAD | TA + LGD               | <i>MSH2</i>                 | 145 | 117 | 80,69 | 10%/10%/10% |
| BP-FB-5_PC   | CrAD | TA + LGD               | <i>MSH2</i>                 | 175 | 123 | 70,29 | 30%         |
| BP-FB-6_PC   | CrAD | TA + LGD               | <i>MSH6</i>                 | 161 | 49  | 30,43 | <10%        |
| BP-FB-7_PC   | CrAD | TA + LGD               | <i>MSH2</i>                 | 157 | 66  | 42,04 | 10%         |
| BP-FB-8_PC   | CrAD | TA + LGD +<br>HGD foci | <i>MSH2</i>                 | 127 | 56  | 44,09 | 10%/10%     |
| BP-FB-9_PC   | CrAD | TA + LGD               | <i>MLH1</i>                 | 160 | 102 | 63,75 | 20%/20%     |

<sup>1</sup>MSI score: number of instable MS/total MS sequenced. **Abbreviations:** CrAD, colorectal adenoma; EC, Endometrial Cancer; TA, Tubular Adenoma; TVA, Tubulovillous Adenoma; EA, Endometrioid Adenocarcinoma; LGD, Low Grade Dysplasia; HGD, High Grade Dysplasia

### 3.2.3 DNA sequencing analysis and variant calling

The FASTQ files obtained for WGS sequences from HNPCC-Sys were transferred to the BSC central storage for read pre-processing and quality control. FASTQC was run on raw reads to access base call quality, adapter content, G/C content, sequencing length and duplication level. The sequencing reads were trimmed of adapters and low-quality bases using Cutadapt<sup>235</sup>. The trimmed reads were mapped against the reference genome GChr37 (hg19) using BWA-mem<sup>236</sup> in paired-end mode. Post alignment quality control (QC) was performed based on BWA alignment BAM files, generating genome mapped percentages, coverage and duplication level. A multiQC<sup>237</sup> run report was generated for the whole sample set. In the case of WGS data, mapped reads were filtered to keep only the coordinates containing mononucleotide repeats (MNR) in coding sequences, which were obtained through the MSIseq software by searching for 7 to 25 nucleotide base repeats in the human reference genome GChr37. We then used three different somatic variant callers Mutect<sup>238</sup>, SamTools<sup>239</sup> and VarScan<sup>240</sup>, and ran each of them using the default parameters. The VEP tool<sup>241</sup> has been used for the annotation and prioritization of genomic variants in coding and non-coding regions from both HNPCC-Sys data and previously sequenced CrAD and EC LS samples data (see section 3.2.2). Only the

variants annotated as frameshift and common to all three variant callers were considered in the following analysis.

The annotated variant files from human sporadic MMRd tumour sequences and from MSI and MSS cell lines were converted from GChr38 to GChr37 using the LiftOver conversion tool from the UCSC Genome Browser. We then identified the presence of pFSDN associated mutations in these sequences by comparing the genome coordinates of each pFSDN associated mutations with the retrieved sequence files in GChr37 format.

### **3.2.4 pVACseq analysis**

The variant calling files from our CrAD and EC cohorts annotated with the Ensemble VEP tool were subsequently used for the pVACseq neoantigen prediction analysis to predict the binding of frameshift-resulting neoantigens to the HLA alleles of each patient. The program was run with default settings and with the predictors MHCnuggetsI, MHCnuggetsII, NetMHC, NetMHCcons, NetMHCIpan, NetMHCpan, NNalign and PickPocket. The resulting files were filtered to exclude neoantigens with an HLA binding threshold greater than an IC50 of 5000nM, and remaining epitopes were characterized as high (IC50<50nM), low (IC50<500nM) and very low (IC50<5000nM) affinity binders based on the median binding score from all used softwares. We then filtered for our potential FSDN by their peptide sequence.

### **3.2.5 RNA sequencing analysis**

mRNA sequencing data from tumoral human cell lines was analysed by removing the adaptor with Trimalore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) and the sequences with repeats and low complexity. Subsequently, the sequences were mapped to the GRCh37 Genome Reference Consortium Human Build 37 (GRCh37) Organism Homo sapiens (human) using the TopHat<sup>242</sup> fast splice junction mapper for both RNA-seq single- and paired-end reads to identify known RNAs transcripts. The expression level of mRNAs was normalized by RPKM (transcript per million) method using the *limma* package from Bioconductor.

### **3.2.6 Statistic tests**

Statistic tests were performed with GraphPad Prism v.8.0.1 or R. All data was evaluated using normality tests prior to analysis, by the Shapiro-Wilk test or the Kolmogorov-Smirnov test

depending on sample size. Paired T-tests (for gaussian distributed cohorts) or Kolmogorov-Smirnov test (for non-gaussian cohorts), Brown-Forsythe and Welch's ANOVA or ordinary One-Way Anova were used to determine significance when comparing cohorts. For categorical variables, analysis was conducted using Pearson's Chi Square ( $\chi^2$ ). Significance shown in APA style (p value: ".123" ns, ".012" \*, ".001" \*\*, "<.001" \*\*\*).

R was used to determine the correlation coefficient (linear correlation) for cell lines DNA data, using the Pearson correlation coefficient (PCC), and to perform clustered graphs for mutation data using the ggplot2, complexHeatmap and Oncoplot packages.

### 3.3 *In-vitro* validation of FSDN

#### 3.3.1 FSDN synthesis

The top 98 pFSDN obtained from *in silico* validations were synthesized by Proteogenix (France) at >80% purity and 4 mg per peptide. Synthetic pFSDNs were reconstituted with either DMSO (Werfen, Spain) or ACN (Sigma-Aldrich, Germany): H<sub>2</sub>O (1:3), following providers recommendations, at a stock concentration of 5 mM per peptide.

#### 3.3.2 Primary cultures

All media cultures compositions used can be found at Supplementary table 3.

##### 3.3.2.1 PBMCs, PBLs and moDCs

To isolate peripheral blood mononuclear cells (PBMCs), peripheral blood from recruited individuals was diluted 1:2 with PBS (Gibco, ThermoFisher scientific, UK) and then slowly poured into two 50 or 15mL falcon tubes containing each 13mL or 5 mL of ficoll lymphoprep<sup>TM</sup> (STEMCELL technologies, Grenoble, France), respectively. Tubes were centrifuged at 800 g for 20 min (RT) without acceleration or brake. Once separated, the PBMC ring located at the interphase was carefully recovered with Pasteur pipettes. Cells were then washed with PBS (RT) and spun at 500 g for 10 min. Additionally, two washes were performed adding cold PBS (4°) and centrifuging at 400 g for 5 min and at 200 g for 10 min to remove cell debris. PBMCs were counted and either used for subsequent analysis or further divided into PBLs and monocytes by resuspending cells in 8 mL of X-VIVO medium (BioWhittaker, Lonza, Belgium) and culture them in a 75 cm<sup>2</sup> culture flask for 2 h at 37° and 5% CO<sub>2</sub> to promote monocyte adherence. After incubation, flasks were carefully washed with PBS (RT) to remove non-

adherent cells, mainly Peripheral blood lymphocytes (PBLs). PBLs were cryopreserved at 10-25 million cells per cryovial, in 1 mL FBS (Gibco, Spain) with 10% DMSO (Sigma-Aldrich, Spain) per cryovial, and 12 mL of X-VIVO medium supplemented with 2% HS and IL-4 (300 U/mL) and GM-CSF (450 U/mL) (Miltenyi Biotec, Madrid, Spain) were added to the adhered monocytes to promote mo-DC differentiation. Cells were incubated for 6 days, changing the media and adding IL-4 and GM-CSF on day 3. On occasion, day 3 mo-DCs were cryopreserved for further use. At day 6, a maturation cocktail composed of IL-1 $\beta$  (1000 U/mL), IL-6 (1000 U/mL), TNF- $\alpha$  (500 U/mL) (CellGenix, Freiburg, German), PGE2 (10  $\mu$ g/mL) (Sigma-Aldrich, Spain), and the TLR-3 agonist Poly I:C (10  $\mu$ g/mL) (InvivoGen, Europe) was added and incubated for 24h to promote mo-DCs maturation. At the same time and when indicated, synthetic peptides were added to mo-DCs cultures at a final concentration of 5  $\mu$ M.

### 3.3.2.2 *Tumour-infiltrating lymphocytes (TILs) from Colorectal Biopsies*

#### 3.3.2.2.1 Colorectal Biopsies processing

Patients' biopsies have been obtained from both macroscopically normal colonic mucosa and from CrADs or CRCs, if available.

Biopsies have been washed twice with XVIVO-15 media supplemented with 2%HS (Human Serum) 2%PS (Penicillin/Streptomycin) before being mechanically digested with a scalpel until achieving approximately 1mm<sup>2</sup> tissue cubes if the pieces were bigger than that.

Once dissected, the tissue has been enzymatically digested with the Miltenyi human tumour dissociation kit with the gentleMACS™ Dissociator (Miltenyi biotec) following provider's instructions.

Single cell suspensions (CS) have been filtered with 70  $\mu$ m strainers to remove debris, counted and cultured to achieve expansion of specific cells, either polyclonally expanded TILs or TRLs (section 3.3.2.2.2) or FSDN-specific T-cells by co-cultures with autologous moDCs or PBMCs (sections 3.3.5.2 and 3.3.5.3).

#### 3.3.2.2.2 Polyclonal expansion of TILs and TRLs

CS from disaggregated biopsies have been cultured in an adequate plate (48 well plates to 12 well plates, depending on recovered cell number) in T-cell media (XVIVO-15 5%HS, 1%PS, 1%GlutaMax) at 10<sup>6</sup> cells per mL, supplemented with Normocin (1:500), IL-2 (2000 IU/mL)

and 10  $\mu$ L of anti-CD3/CD28 beads or TransAct (Miltenyi Biotech, Madrid, Spain) to stimulate polyclonal T-cell growth.

Media change and IL-2 addition (2000U/mL of final well volume) has been performed every 48h and depending on cell growth. Once the desired number of cells has been achieved, cultures have been recovered and cryopreserved in albumin at 10%DMSO in liquid nitrogen (-180°C). Each cryovial contained around 10 to 25 million cells. When available, an aliquot of cells from early expansion (0.5-1-week post culture, depending on cell availability) has been cryopreserved.

### 3.3.3 Cell lines

K-562 derivatives, and Jurkat cell lines were cultured at 200.000 cells/mL in RPMI 10% FBS, 1% P/S and 1% Glutamax, HEK 293T were cultured in DMEM 10% FBS, 1% P/S and 1% Glutamax, and all lines were regularly checked for mycoplasma contamination by PCR.

#### 3.3.3.1 *Generation and culture of monoallelic K562-based Cell Lines expressing vaccine peptides*

HLA class I monoallelic cell lines were generated in the Linette-Carreño lab (University of Pennsylvania, Philadelphia, USA) by transduction of K562 cell lines using lentiviral particles encoding different alleles of HLA-SCD/eGFP52 (Figure 18) and sorted for GFP positivity to achieve >99% pure cell lines (Table 4), as explained in previous works<sup>243</sup>. Upon arrival at our institution, cell lines were subsequently transduced with lentiviral particles encoding a tandem minigene construct (TMC)/mCherry containing all vaccine FSDNs (see section 3.3.3.1.1 for TMG construction and 3.3.3.1.2 onward for TMG plasmid synthesis, TMG-lentiviral production and transduction details), then sorted to >99% purity by flow cytometry based on co-expression of GFP, mCherry, and HLA class expression as measured by APC-conjugated anti-human HLA-ABC (clone W6/32) staining. Sorted cell lines were cultured in RPMI media supplemented with 10% FBS, 1% Penicillin/Streptomycin and 1% Glutamax and maintained at a concentration of 200.000 cells/mL by regular passaging.



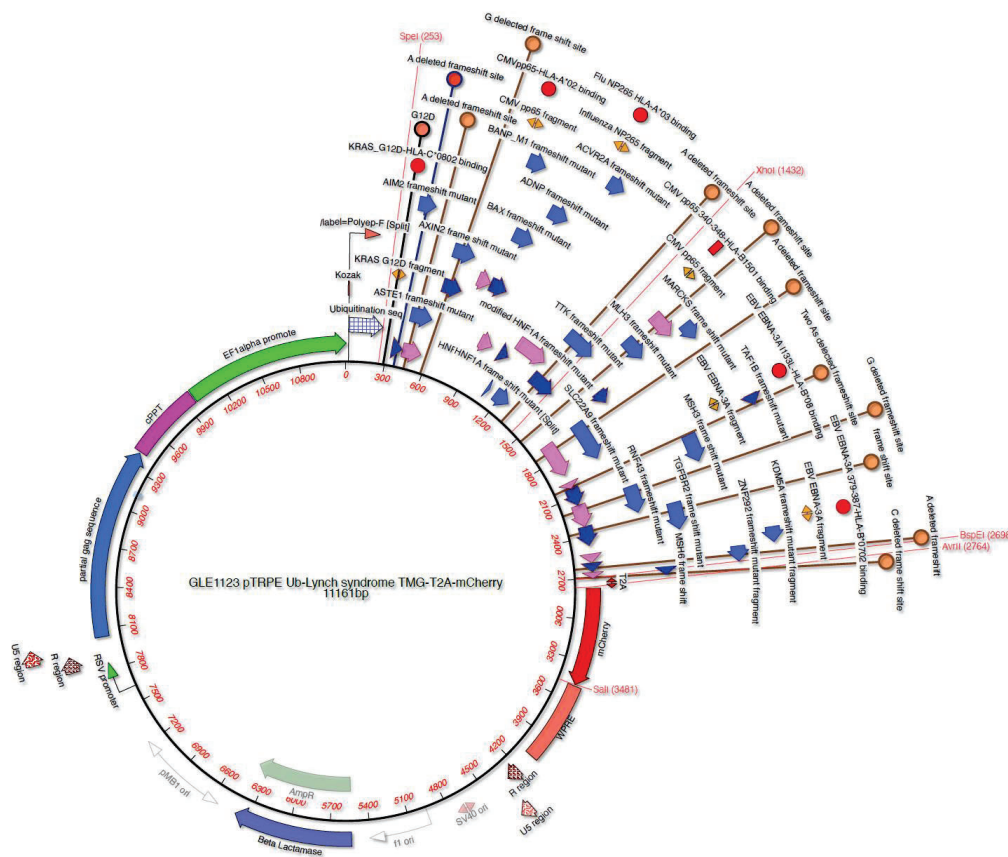


Figure 19. Tandem Minigene map. Vector map for the GLE1123 pTRPE Ub-Lynch syndrome TMG-T2A-mCherry lentiviral construct encoding the 24 peptides included in the LS vaccine.

#### 3.3.3.1.2 Bacterial transformation

One Shot™ Stbl3™ Chemically Competent *Escherichia coli* (ThermoFisher, USA) was used to incorporate cloned DNA. Working on ice, a vial of *E. coli* was thawed and 1 µL of TMG DNA was added. The mix was incubated for 30 min, followed by a 45 second heat shock at 42°C and immediately returning the vial on ice. Afterwards, 250 µL of pre-warmed S.O.C. (ThermoFisher, USA) were added to the vial, which was shaken for 1h at 37°C and 230 rpm. Finally, 25 µL from the vial were spread in a LB agar plate and incubated at 37°C ON.

Transformed *E. coli* colonies were selected, transposed into LB medium with ampicillin and left shaking ON at 37°C. Finally, DNA extraction was performed by using the HiPureLink™ Plasmid MaxiPrep Kit (ThermoFisher, USA) and plasmid concentration was assessed using a

NanoDrop® ND-1000 spectrophotometer and adjusted at 1000ng/μL. Plasmids were kept at -80°C until use.

#### 3.3.3.1.3 Lentiviral production

Third generation TMC-expressing lentiviral particles were generated using HEK 293T cell line as a packaging cell line and a standard polyethylenimine (PEI) transfection protocol. Briefly, HEK 293T cells were plated in one or more T-150 cell culture plates to reach a sufficient cell number. Afterwards, cells were collected by trypsinization (TripLE, ThermoFisher, USA) and re-plated in P15 plates at  $10 \times 10^6$  cells per plate, at a final volume of 20 mL per plate of supplemented DMEM. After a 24h incubation at 37°C, plasmid and PEI mixes were prepared to perform the transfection.

The plasmid mix consisted of 18 μL of the GLE1123 pTRPE Ub-Lynch syndrome TMG-T2A-mCherry plasmid, 15 μL of pMDLg-pRRE (Addgene, 12251) which encodes Gag and Pol, 15 μL of pRSV-Rev (Addgene, 12259) which encodes REV, 7 μL of pMD2.G (Addgene, 12253) which encodes VSV-G, and 675 μL of non-supplemented DMEM media. The PEI mix consisted of 95 μL of PEI solution plus 675 μL of non-supplemented DMEM. Both preparations were carefully mixed and incubated for 20 minutes at RT before adding it to culture plates (1.5 mL of mix per plate). After a 4-6 h incubation, a media change was performed in all plates. The supernatant from transfected HEK-293 plates was collected after 48 and 72h, adding fresh media after recollections and keeping the gathered media at 4°C.

HEK-293-T supernatants were centrifuged at 500 g for 10 minutes, filtered with 0.45 μm filters (Millipore, Merck, Germany) and mixed with LentiX solution (Takara Bio, Europe) at a 1:3 ratio (Supernatant : LentiX). Tubes were then centrifuged at 1500 g for 45 minutes at 4°C, and pellets were resuspended in 200 μL per collected plate in non-supplemented DMEM. Stock samples of 50 to 100 μL were kept at -80°Cs until use.

#### 3.3.3.1.4 Lentiviral titration

A Jurkat cell line was used for lentiviral titration. Jurkat cells were seeded on the first column of a U-bottom 96 well plate at 20.000 cells and 100 μL of supplemented X-VIVO per well. Separately, a 1:100 pre-dilution of the vector was first created by mixing 2 μL of vector with 18 μL of medium, followed by the addition of 180 μL of medium. 3-fold serial dilutions were prepared in rows A through H, starting with the pre-dilution in row A and ending with a 1:6561

dilution in row H. Then, 50 µL of each vector dilution were transferred to Jurkat wells, leaving one with no viral particles. The plate was incubated for 24 to 72 h at 37°C before analysing mCherry fluorescence by flow cytometry.

Lentiviral titer was calculated using the following formula:

$$\text{Titer} \left( \frac{\text{TU}}{\text{mL}} \right) = \left( \frac{\% \text{ mCherry cells}}{100} \right) * N^{\circ} \text{ Jurkat cells} * \text{dilution}$$

The first dilution at which the percentage of positive Jurkat cells is less than 20% was chosen as the titer calculated from this dilution represents the most accurate limiting-dilution titer for the vector sample. To calculate the µL of lentiviral particles to add to one million cells to achieve a standard multiplicity of infection (MOI) 5, the following formula was used:

$$\text{Lentiviral particles (}\mu\text{L)} = \left( \frac{1.000.000 \text{ cells} * 5 \text{ (MOI)}}{\text{Limiting dil titer TU/mL}} \right) * 1000$$

### 3.3.4 Cell phenotyping and characterization assays

#### 3.3.4.1 Staining and detection of surface cellular markers by flow cytometry

Cells were stained either on cytometry tubes or on U-bottom 96 well plates. Cytometry analysis was performed on an Attune™ NxT (ThermoFisher) cytometer, and resulting data was analysed by the FlowJo software (Tree Star Inc.). All monoclonal antibodies used can be found in Supplementary table 4, and all panels used can be found in Supplementary table 5.

For the detection of antigens on the cell surface, a minimum of 50.000 cells were taken and washed with PBS by centrifuging at 600 g for 5 minutes. Subsequently, the supernatant was discarded and 2 µl of each of the monoclonal antibodies of interest labelled with different fluorochromes were added to 100 µl of the cell suspension and incubated for 20 minutes at room temperature. Next, the cells were washed twice with excess PBS by centrifuging at 600 g for 5 minutes and finally they were passed through the flow cytometer. Dextramer (Immudex, Denmark) and Pentamer (Proimmune, UK) stainings were performed following providers directions.

#### 3.3.4.2 Single cell multiplex cytokine profiling

Single cell secretome profiles of PBMCs or Cell Suspensions from healthy donors and Lynch Syndrome carriers were analysed. Briefly, cells were thawed and left to rest ON in RPMI

supplemented with 10% FBS and 1% P/S (Sigma-Aldrich), 1% GlutaMax (ThermoFisher) and 50 U/mL of IL-2 (Biolegend), at a concentration of  $1.5 \times 10^6$  cells/mL. After incubation, cells were recovered and enriched for CD3+ cells with the Miltenyi Pan T-cell Isolation Kit (Human) using magnetic separation with LS columns (Miltenyi), following providers instructions. After separation, the negative fraction was kept and cells were immediately stained with a cell tracer (405nm Cell stain, violet, Bruker) by incubating them with the freshly prepared reagent for 10 minutes at 37°C. Afterwards, cells were counted and distributed in a U-bottom 96 well plate, at a concentration of 100.000 cells/well in 100 µL/well of complete RPMI media plus 1:500 of the Cell Activation Cocktail (without Brefeldin A, BioLegend). The plate was incubated at 37°C for 1.5h to stimulate T-cells. Meanwhile, IsoCode Chips were thawed at RT and liquid reagents and setup in the Bruker IsoPlexis Instrument were prepared. After incubation, cells were recovered, pooled by patients and stained with an anti-human CD8-AF647 antibody. After washing the cells, they were counted, resuspended at  $10^6$  cells/mL, and loaded onto the Single Cell adaptive immune secretome IsoCode chips (IsoPlexis, Haven, Connecticut, USA) by adding 30 µL of each sample to the inlet ports. Chips were then loaded onto the instrument to perform the run. Results were exported and analysed with the Data Visualization Software for IsoLight (IsoSpeak). This approach is based on single-cell analysis, providing measures in relative fluorescence units (RFUs). The background RFUs from empty microchambers or cell debris are measured and subtracted from test chambers.

#### 3.3.4.3 *TCRVβ clonogram analysis*

The TCR Vβ repertoire was determined using the flow cytometry-based IOTest® Beta Mark TCR Vβ repertoire kit (Beckman Coulter) for LS and Healthy control T cells obtained following PBMC isolation and biopsy disaggregation protocols (see sections 3.3.2.1 and 3.3.2.2.1). Cells were thawed and left to rest 2h in supplemented X-VIVO media. Afterwards, cells were collected, washed 2x with PBS and the cell suspension was divided into ten wells of a U-bottom 96 well plate (100 µL/well) labelled C+, C- and A-H, corresponding to the different antibody cocktails to stain for the 24 TCR Vβ-families (IOTest® Beta Mark TCR Vβ repertoire kit, Beckman Coulter B.V.) plus the controls. Each cocktail contains antibodies directed to three different Vβ-families, i.e., one is fluorescein isothiocyanate (FITC-), one is PE-labeled and one is labeled with both FITC and PE. Table 6 shows the description of the antibodies directed to the different TCR Vβ-families in wells A to H. Afterwards, cells were subsequently stained with anti-CD3-APC-H7 to distinguish T-cells.

Table 6. TCR V $\beta$ -families in tube A-H from the IOTest ® Beta Mark TCR Vbeta repertoire kit (Beckman Coulter).

| Tube | Volume / No. of tests (20 $\mu$ L / test) | V $\beta$ * (%)                                                                       | Fluorochrome          |
|------|-------------------------------------------|---------------------------------------------------------------------------------------|-----------------------|
| A    | 1.0 mL / 50 tests                         | Vb 5.3 (TRBV5-5)<br>Vb 7.1 (TRBV4-1, TRBV4-2, TRBV4-3)<br>Vb 3 (TRBV28)               | PE<br>PE+FITC<br>FITC |
| B    | 0.5 mL / 25 tests                         | Vb 9 (TRBV3-1)<br>Vb 17 (TRBV19)<br>Vb 16 (TRBV14)                                    | PE<br>PE+FITC<br>FITC |
| C    | 0.5 mL / 25 tests                         | Vb 18 (TRBV18)<br>Vb 5.1 (TRBV5-1)<br>Vb 20 (TRBV30)                                  | PE<br>PE+FITC<br>FITC |
| D    | 0.5 mL / 25 tests                         | Vb 13.1 (TRBV6-5, TRBV6-6, TRBV6-9)<br>Vb 13.6 (TRBV6-6)<br>Vb 8 (TRBV12-3, TRBV12-4) | PE<br>PE+FITC<br>FITC |
| E    | 0.5 mL / 25 tests                         | Vb 5.2 (TRBV5-6)<br>Vb 2 (TRBV20-1)<br>Vb 12 (TRBV10-3)                               | PE<br>PE+FITC<br>FITC |
| F    | 0.5 mL / 25 tests                         | Vb 23 (TRBV13)<br>Vb 1 (TRBV9)<br>Vb 21.3 (TRBV11-2)                                  | PE<br>PE+FITC<br>FITC |
| G    | 0.5 mL / 25 tests                         | Vb 11 (TRBV25-1)<br>Vb 22 (TRBV2)<br>Vb 14 (TRBV27)                                   | PE<br>PE+FITC<br>FITC |
| H    | 0.5 mL / 25 tests                         | Vb 13.2 (TRBV6-2)<br>Vb 4 (TRBV29-1)<br>Vb 7.2 (TRBV4-3)                              | PE<br>PE+FITC<br>FITC |

#### 3.3.4.3.1 Cluster and expansions analysis of TCR V $\beta$ clonogram

For TCR V $\beta$  data analysis we used Clustergrammer<sup>245</sup>, a web-based tool for visualizing and analysing high-dimensional data as hierarchically clustered heatmaps. We uploaded data matrixes as tab delimited .txt files and used the default cosine distance and average linkage options.

Expansions of specific V $\beta$  families were defined as a frequency above the mean percentage + 2 times the SD of a certain TCR V $\beta$ -family obtained using the healthy controls as reference population. By using this approach, finding an expansion by chance is lower than 2.5%.

### 3.3.5 Co-cultures and cell reactivity assays

#### 3.3.5.1 *TILs and moDCs co-cultures for bulk IFN- $\gamma$ ELISPOTs*

Previously cryopreserved day 3 DCs were thawed and cultured as stated above. Briefly, cells were cultured for 4 days in XVIVO-15 2%HS plus IL-4 (300 U/mL) and GM-CSF (450 U/mL)

(Miltenyi Biotec, Madrid, Spain), and matured in a pro-inflammatory environment (IL-1 $\beta$  10ng/mL, IL-6 10ng/mL, TNF $\alpha$  10ng/mL, PGE2 10mg/mL, and Poly(I:C) at 20 $\mu$ g/mL). At the same time, they were loaded with synthetic FSDNs (5 $\mu$ M per peptide) and incubated for 24h to promote maturation and antigen presentation. Autologous TILs or TRLs were then thawed and cultured for 48h in T-cell media.

The corresponding number of wells of a commercially pre-coated IFN- $\gamma$  ELISPOT plate (Mabtech, Sweden) were blocked with XVIVO-15 10% FBS (Merck, Germany) for 30 minutes. Then, the mature DCs loaded with pFSDNs were co-cultured at a 1:10 ratio with autologous T cells in a final volume of 200  $\mu$ L per well (10.000 DCs : 100.000 TILs). The negative control consisted of mature non-loaded DCs co-cultured with T cells plus DMSO, and for the positive control the ELISPOT kit's CD3 stimulation via CD3-2 monoclonal antibody (1:1000) was used on T-cells alone. The plate was incubated at 37°C and 5% CO<sub>2</sub> for 16 to 24h, and the ELISPOT was revealed following providers instructions. Spot forming units (SFUs) were measured using an AID ELISpot reader. Tests were considered positive when the number of spots was 1.5 times higher than the negative control. In this case, we used only a 1.5-fold threshold for ELISPOT positivity due to an expected high background noise from the co-culture of non-loaded mature moDCs plus autologous TILs, which usually induces an unspecific pro-inflammatory stimulus. The supernatants were collected after incubation to analyse cytokine production by IL-10 ELISA kits (ThermoFisher, USA).

#### 3.3.5.2 *Specific T-cell expansion with moDCs*

Autologous TILs or TRLs and moDCs were thawed in an adequate flask, and let set for 24 to 72h, depending on the DCs state of development. DCs were fully matured with the cytokine cocktail IL-6, IL-1 $\beta$ , TNF $\alpha$ , PGE2 and PolyI:C as described previously. At the same time, DCs were pulsed with 5  $\mu$ M of the corresponding peptide and incubated 24h at 37°C. T-cells were cultured without IL-2 for at least 48h.

On the co-culture day 0, DCs and T-cells were recovered and resuspended at 10<sup>5</sup> and 10<sup>6</sup> cells/mL in T-cell media, respectively. If enough cells were available, an aliquot of TILs or TRLs was analysed by flow cytometry to determine the CD4 and CD8 population percentages. Depending on T-CD8+ percentages and on cell number, CD8+ T cells were purified by negative selection using the Miltenyi CD8+ negative selection kit (cat 130-091-154) following providers instructions.

To a minimum of 4 wells per peptide of a 48 well tray, cells were added in this order, 0.125 mL T-cell media, 0.125 mL ( $1.25 \times 10^4$  cells) of the DC cell suspension and 0.25 mL ( $2.5 \times 10^5$  cells) of the T cell suspension. T cells to pulsed DC ratio was 20:1. On days 2 and every 48hr afterwards IL-2 (50U/100mL/well) was added. The culture was continued until day 14, where cells were re-stimulated with freshly pulsed DCs at a 10:1 ratio. If at any time wells were confluent, each well from a 48 well plate was transferred to a 24 well plate, the media volume doubled, and IL-2 concentration adapted to the media volume for subsequent addition.

Once the cultures were at days 26 to 28, or whenever a high T-cell growth was observed, cells were stained with peptide-MHC-I-PE dextramers (Immudex) plus a CD8-AF700 (BD) antibody to determine peptide specific T-cell percentages. A negative control consisting of an irrelevant peptide-HLA-A\*03:01-PE multimer was used for all stainings.

### 3.3.5.3 *Specific T-cell expansion with autologous PBMCs*

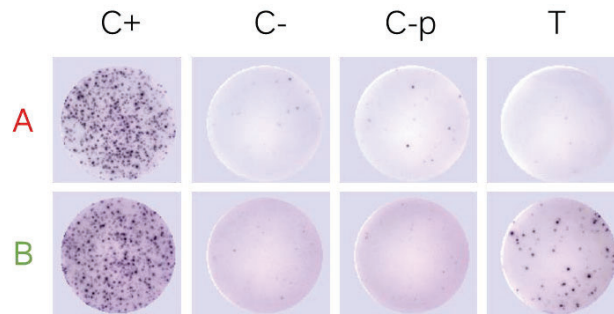
On day 0, fresh CrAD or normal mucosa biopsies and 40mL of blood were obtained from LS patients. Blood was processed as stated on the section 3.3.2.1 to obtain PBMCs, and biopsies were parallelly processed as stated on the section 3.3.2.2 to obtain a CS that contains TRLs. Fresh PBMCs were irradiated at 40 Gy for 30 minutes before their use. Afterwards, cells derived from the biopsies were co-cultured with irradiated PBMCs at a ratio of 1:5 (CS:PBMCs), to ensure at least 5 APCs for each responding cell, on a minimum of 5 wells of a p48 and a minimum of 100.000 CS cells per well, in T-cell media without IL-2. To each well, a peptide mix pooling all 18 HLA-I peptides included in the DELAY vaccine (DELAY-I pool, see chapter I results section 4.5) was added at 5  $\mu$ M concentration per peptide. Cells were incubated for 48 to 96h before adding IL-2 (20 to 40 U/mL) and changing the media. From then on, media was changed every 48h until day 14.

On day 14, T-cells were re-stimulated with frozen PBMCs at the same ratio (1:5 T-cells to PBMCs) and tested for peptide reactivity by IFN- $\gamma$  ELISPOT (Thermofisher). Briefly, 8 wells per patient of a commercially pre-coated IFN- $\gamma$  ELISPOT plate were blocked with XVIVO-15 10% FBS for 30 minutes. Then, 400.000 to 800.000 T-cells were obtained from cell cultures, resuspended in 800  $\mu$ L of non-supplemented XVIVO-15 media and distributed in the blocked wells, adding 100  $\mu$ L per well (50.000 to 100.000 cells per well). In duplicates, the following stimuli were added to the corresponding wells:

| Well ID                           | Reagent added per well                                            |
|-----------------------------------|-------------------------------------------------------------------|
| Positive control                  | 100 $\mu$ L XVIVO-15 + 1:1000 aCD3 ELISPOT +control (0.2 $\mu$ L) |
| Neg. control                      | 100 $\mu$ L XVIVO-15                                              |
| Neg. control + irrelevant peptide | 100 $\mu$ L XVIVO-15 + Irrelevant peptide (3.6 $\mu$ L)           |
| Test                              | 100 $\mu$ L XVIVO-15 + 5 $\mu$ M peptide pool (3.6 $\mu$ L)       |

The irrelevant peptide consisted in an H<sub>2</sub>O:ACN (3:1) resuspended peptide that had not been added to the stimulation cocktail. Once added, the plate was incubated ON (for approximately 16h) at 37°C and 5% CO<sub>2</sub>.

On the following day, cells were collected and duplicates pooled for flow cytometry analysis of surface population and activation markers. ELISPOT was revealed as stated in the commercial kit instructions and read using an AID ELISpot reader. Tests were considered positive when the number of spots was 3 times higher than the negative control (Figure 20).



*Figure 20. IFN- $\gamma$  ELISPOT analysis criteria. Randomly picked representative IFN- $\gamma$  ELISPOT plate reading. Row A (red) shows a negative result, row B (green) shows a positive result ( $n^{\circ}$  spots Test  $\geq$  3 times  $n^{\circ}$  spots C-). First column corresponds to the positive control, second column to the negative control (no added stimuli), third column to the negative control plus an irrelevant peptide at the same concentration as the ones added to test wells, and last column corresponds to the test wells.*

### 3.3.6 Statistical analysis

Statistical analyses were performed using PRISM8 or R. The Shapiro-Wilk test was used to assess normal distribution of the data in every analysis. Statistical significance of performed analysis was inferred with the paired or unpaired T-test for normally distributed data, and Mann-Whitney test for or paired Wilcoxon sign-ranked test for non-normally distributed data.

Chi-square test was used to determine significance for categorical variables. For every test, significance was defined by a P-value<0.05.

R was used to perform association analysis on HLA data using the MiDAS package, using a conventional linear regression model or a multinomial logistic regression model for and the Bonferroni correction for correlations to different phenotypic categorical variables.

### 3.4 CLINICAL TRIAL

Once the vaccine neoantigens had been prioritized and validated, we started a phase Ib, preventive, first-in-human, non-randomized clinical trial aiming to induce cellular memory against the selected neoantigens in LS carriers with no previous history of cancer.

#### 3.4.1 Drug substance

The product consists of autologous human DCs derived from mature monocytes (moDCs), immunogenic, generated ex vivo and loaded with 4 pools of 6 class I and II HLA-restricted FSDN each (8-9 amino acids and 15 amino acids, respectively) with high frequency in tumours related to LS. These antigens are expressed as epitopes on HLA-I and HLA-II molecules on the surface of DCs. The product vial composition consists of a suspension of mature moDCs loaded with neo-antigens in albumin with 10% DMSO as a cryoprotective.

##### 3.4.1.1 Manufacturing process

The obtaining of the DCs takes place in class B clean rooms and class A laminar flow cabins and is part of a leukapheresis of the receiving individual. The neo-peptides that are incubated with the moDCs are synthetic and of GMP grade, reconstituted in DMSO.

The vaccine is entirely produced in the Advanced Therapies Laboratory of the HCB, which is currently accredited by the Spanish Agency of Medicines and Medical Devices (AEMPS) for the production of DCs complying with the Good Manufacturing Practices (GMP), as stated in the certificate issued by AEMPS no. ES/079HVI/23.

The preparation of autologous DCs loaded with FSDN, also schematically represented in Figure 21, takes 14 days consisting of 8 days of manufacturing plus an additional week for the quality controls. The leukapheresis bag is fully processed on the day or day after of extraction

to obtain in a single procedure the necessary amount of DCs to be able to inoculate all the doses aimed in the trial, cryopreserving the different doses until the day of inoculation.

The process of monocyte adherence and differentiation into DCs spans several days and is fully carried out in grade A and B white rooms. Briefly, the initial apheresis product is diluted with PBS and layered over ficoll, followed by centrifugation to isolate PBMCs, which are then washed, counted, and resuspended in XVIVO-15 medium with patient serum. Monocytes adhere to culture flasks and non-adherent cells are separated and cryopreserved as controls. Over the next few days, the adhered monocytes are cultured with IL-4 and GM-CSF to promote their differentiation into moDCs.

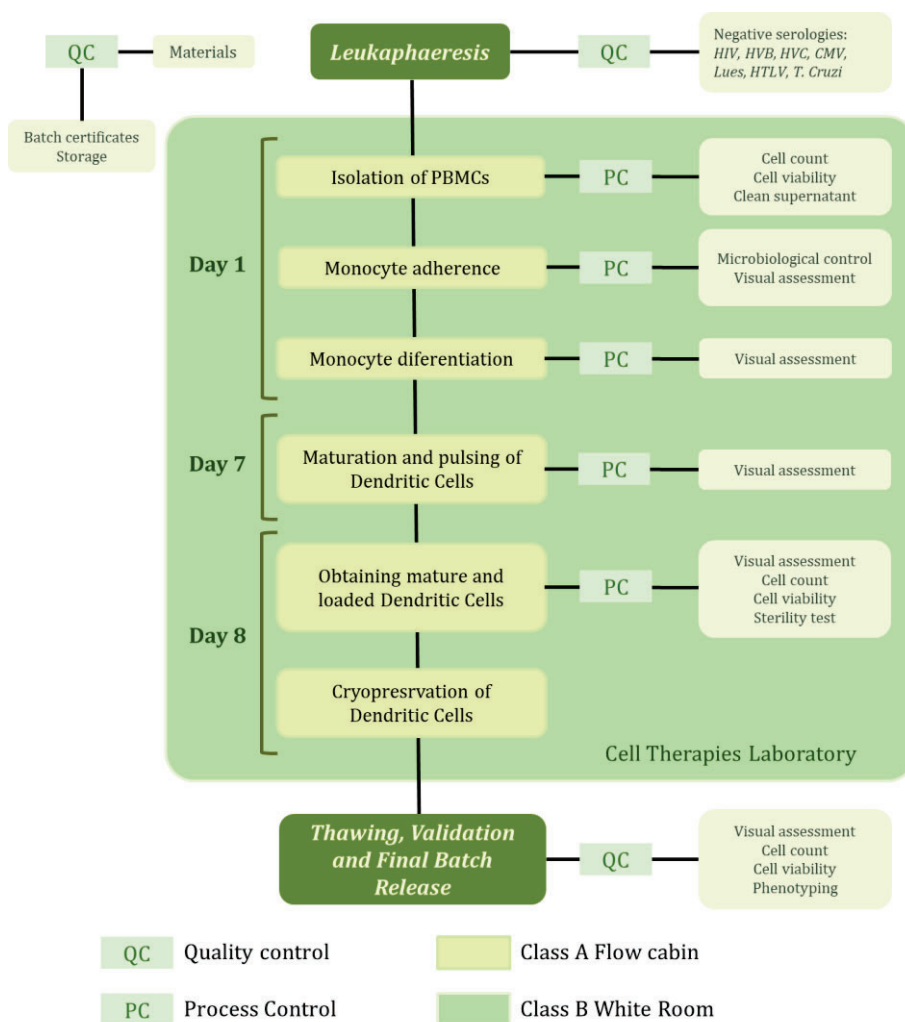


Figure 21. Detailed diagram of the manufacturing process of the active substance.

On the seventh day, maturation cytokines and GMP synthetic peptides are added by peptide pools to enhance their immune-stimulating properties. Finally, on the eighth day, the matured moDCs are collected, washed, resuspended in human albumin, mixed with 10% DMSO, ensuring an homogenous mix of all added pools, and cryopreserved for clinical use. Each cryovial contains a DELAY vaccine dose ( $10 \times 10^6$  DCs) in 0.4 mL total.

### 3.4.2 Design

Briefly, the study design will be as follows: 20 individuals with LS treated at the High-Risk CRC Clinic at the HCB will be recruited, and after signing the IC, a leukapheresis will be performed. Once the product is generated, it will be administered, and clinical and immunological follow-up (immunomonitoring) will be conducted for 36 months. The cohort included in the study will continue to undergo standard prevention measures for CRC and gynaecological cancer, consisting of surveillance colonoscopies for CRC and periodic gynaecological ultrasounds +/- endometrial biopsies.

#### 3.4.2.1 Study objectives and endpoints

##### *Primary Objectives*

- To evaluate the safety and tolerability of autologous peripheral blood differentiated and matured adult dendritic cells loaded with frameshift derived neopeptides (DC-DELAY) in LS carriers.
- To evaluate specific frameshift-derived neoantigens immunogenicity of DC-DELAY in LS carriers at week 12.

##### *Secondary Objectives*

- To evaluate the early-term specific frameshift-derived neoantigens immunogenicity of DC-DELAY in LS carriers at week 6.
- To evaluate the long-term specific frameshift-derived neoantigens immunogenicity of DC-DELAY in LS carriers at 6, 12 and 24 months.
- To evaluate the specific frameshift-derived neoantigens immunogenicity of DC-DELAY in the normal colonic mucosa in LS carriers at month 12.
- To evaluate the effect of DC-DELAY in LS carriers at week 12 on the burden of mismatch repair deficient colorectal neoplasia throughout the study duration.

- To evaluate the effect of DC-DELAY immunization in LS carriers at week 12 on the burden of LS-related carcinomas throughout the study duration.
- To evaluate the safety and tolerability of DC-DELAY throughout the study duration.

#### *Primary Endpoints*

- Proportion of participants with grade 3-4 related adverse events for 12 months following the first immunization.
- Proportion of participants with specific frameshift-derived neoantigens immune response induced by DC-DELAY as measured in peripheral blood by enzyme-linked immune absorbent spot (ELISpot) assay at week 12.

NOTE: Immunogenicity will be defined as T-cell reactivity to at least 1 out of 4 of the pools using an enzyme-linked immune absorbent spot (ELISpot) assay.

#### *Secondary Endpoints*

- Proportion of participants with early immune response induced by DC- DELAY as measured in peripheral blood by ELISpot assay at week 6.
- Proportion of participants with long term immune response induced by DC-DELAY as measured in peripheral blood by ELISpot assay at months 6, 12 and 24.
- Proportion of participants with immune response induced by DC-DELAY as measured in the normal colonic mucosa at month 12.
- Proportion of participants with mismatch repair deficient colorectal adenomas, advanced neoplasia, and/or carcinoma throughout the study duration (36 months).
- Proportion of participants with LS-related carcinomas throughout the study duration (36 months).
- Number and proportion of participants with related adverse events at 7 days after each DC-DELAY immunization.
- Number and proportion of participants with grade 3-4 related adverse events throughout the study duration (36 months).

### 3.4.3 Patients

#### 3.4.3.1 Inclusion, exclusion and discontinuation

The cohort of individuals to be included in the study consists of individuals with Lynch Syndrome without a previous or active history of cancer. There is no placebo cohort included. The inclusion and exclusion criteria are specified in Supplementary table 7.

Treatment will be suspended for any of the following reasons: patient refusal or non-compliance; unacceptable toxicity (any grade 4 toxicity or other potentially life-threatening toxicity); any change in the patient's condition that, in the opinion of the investigator, justifies the interruption of treatment.

#### 3.4.4 Treatment and clinical schedule

The treatment to be administered is DC-DELAY (Autologous peripheral blood differentiated and matured adult dendritic cells loaded with frameshift derived neopeptides). DC-DELAY will be administered intradermal under a split regime of 6 vaccinations once every 2 weeks (Figure 22). A total dose 60,000,000 (+/- 12,000,000) will be administered on each vaccination. The volume of immunization intradermal administration is 0.4 mL.

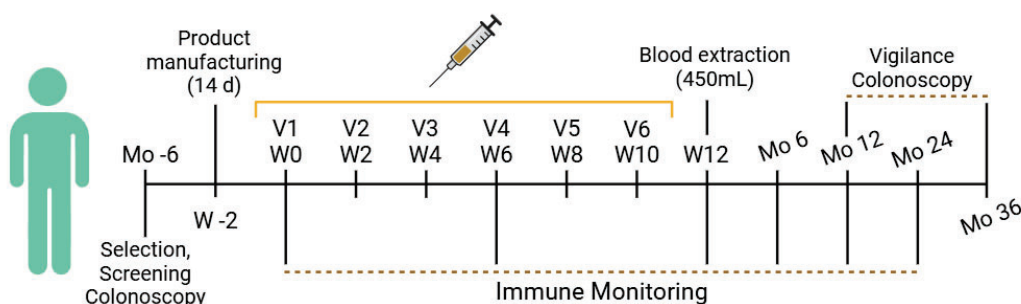


Figure 22. Schematic representation of the treatment and clinical schedule for the DC-DELAY vaccine.

Initial participant selection will be conducted prior to the initial apheresis, but the final eligibility assessment will be made after confirming the isolation of a favourable cellular count needed for the development of the product. Screening colonoscopies are normally performed every 12-24 months, and during the trial they will be arranged between 1-6 months prior the vaccine administration, and the next colonoscopy 12 months after the first dose. Colonic mucosa samples will be obtained at the screening colonoscopy on month 12. Additionally,

adverse events will be monitored and registered for the full duration of the trial. For immune monitoring, 50 mL in each extraction point (see Figure 22) are required. Additionally, a high-volume blood extraction of 450mL will be performed on week 12. This volume includes blood for immune monitoring, the complete metabolic panel, and viral serologies.

### **3.4.5 Immune monitoring**

The immunomonitoring component of this study is designed to assess cellular immunity in response to the inoculated peptides, with the goal of determining whether they elicit an immunogenic response, establish long-term memory, and influence clinical outcomes in the context of cancer prevention. Blood samples (50 mL per extraction) are collected from all enrolled patients at the screening stage (month -6) and subsequently on days 28, 70, and at months 6, 12, 24, and 36 (Figure 22). Colonic biopsies are obtained during the screening phase (month -6) and again at month 12 following vaccination. The additional high-volume blood extraction of 450mL will be used to obtain peptide-specific T-cells for functional analysis.

Immunogenic reactivity will be evaluated using IFN- $\gamma$  ELISpot assays on isolated blood and colorectal T cells against the peptide pools, alongside flow cytometry analysis. ELISpots will be performed as previously specified, and the immunogenicity rate will be defined as T-cell reactivity (defined by a positive ELISpot) to at least one of these four FSDN pools. The treatment will be considered to generate an immune response if at least 25% of participants show a positive response to the neoantigens.

Furthermore, a panel of PE-labelled pentamers will be employed to track peptide-specific T cells in patients whose HLA haplotypes align with the available pentamers and who demonstrate a response to a specific peptide recognized by those pentamers.

Isolated FSDN-specific T-cells will be tested for cytotoxic capacity against the Lynch-TMG-transduced K-562 cell line by a basic cytotoxicity assay. Cell death will be measured as loss of mCherry and GFP fluorescence of the cell lines, normalized to a control of only target cells.

### **3.4.6 Clinical efficacy and safety evaluation**

#### **3.4.6.1 Safety**

The clinical trial will monitor and evaluate participant safety through the systematic collection, classification, and reporting of adverse events (AEs) and serious adverse events

(SAEs). Adverse events will be defined as any medical occurrence during the study, regardless of its relationship to the investigational product, while SAEs will include events such as hospitalization, life-threatening conditions, or death. Causality and severity will be assessed using standardized criteria (e.g., CTCAE v5.0), and all AEs will be documented in the clinical trial records. Each event's intensity will be classified as mild, moderate, or severe, and its relationship to the study drug will be determined.

Immediate reporting of SAEs to the sponsor will be required within 24 hours of awareness. Suspected Unexpected Serious Adverse Reactions (SUSARs) will be reported to regulatory authorities within defined timelines (7 days for fatal or life-threatening cases; 15 days for others). All AEs and SAEs will be followed until resolution or stabilization, and data will be collected in compliance with regulatory requirements. The proportion of participants with grade  $\geq 2$  adverse events must be less than 40% to be considered safe.

#### 3.4.6.2 *Clinical efficacy*

The clinical efficacy of the DC-DELAY immunization will be assessed over a 3-year follow-up period in 20 treated patients, focusing on the incidence of neoplasms and their relationship with immune response. The primary outcome will be the presence and total number of neoplasms, while the main independent variable will be a positive immune response, defined as T-cell reactivity to FSDN at day 84. Covariates such as age, sex, mutation type, and prior history of neoplasms will also be considered.

The analysis will compare neoplasm incidence, the average number of neoplasms per patient, and time to the first neoplasm between patients with and without an immune response to the peptide pools. Statistical methods, including Chi-square, Fisher's exact test, t-tests, Kaplan-Meier survival analysis, and multivariate models (e.g., Cox regression), will assess these relationships while adjusting for covariates. Regression models (Poisson or negative binomial) will be applied to evaluate the total neoplasm count per patient.

A significance level of  $p < 0.05$  will be used, supported by confidence intervals for key parameters and hypothesis tests to examine expected differences in clinical outcomes and immunogenicity.





**CHAPTER 1:**  
*NEOANTIGENS PREDICTION,  
VALIDATION AND PRIORITIZATION*



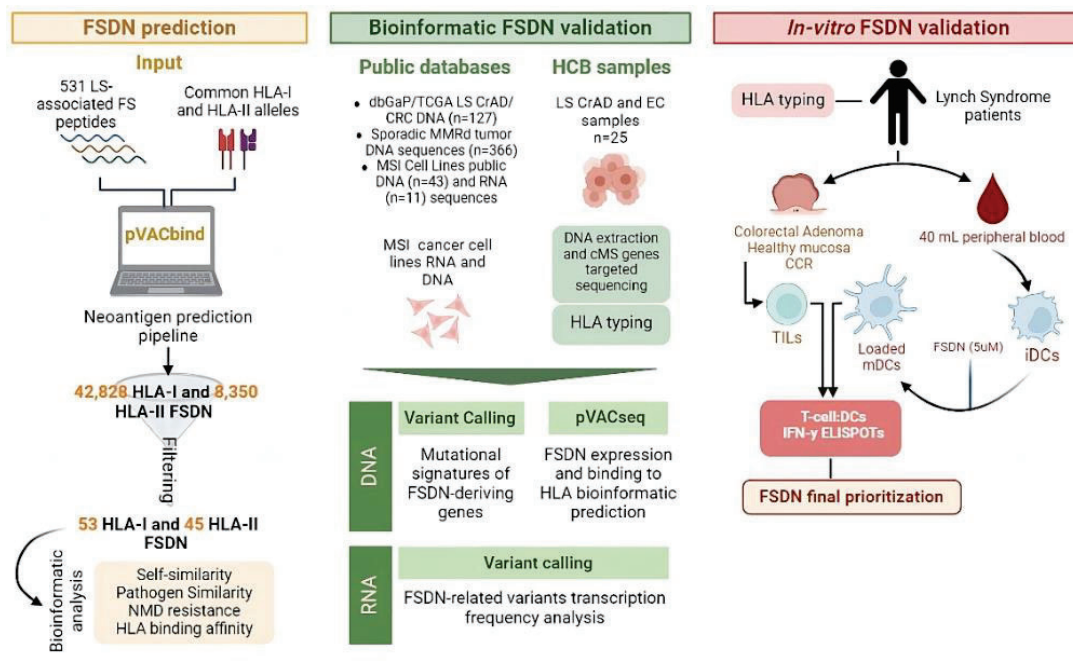


## 4 CHAPTER I: NEOANTIGENS PREDICTION, VALIDATION AND PRIORITIZATION

The work from this chapter and from Chapter II section 7.2.2. has been published<sup>246</sup> in the *Journal for ImmunoTherapy of Cancer* (JITC) on April 20<sup>th</sup>, 2025, under the citation: Bayó C, Castellano G, *et al.* Discovery and validation of frameshift-derived neopeptides in Lynch syndrome: paving the way for novel cancer prevention strategies. *Journal for ImmunoTherapy of Cancer*, 2025; 13:e011177. doi: 10.1136/jitc-2024-011177; with Cristina Bayó as the first author.

### 4.1 Prediction and characterization of shared neoantigens in LS-related tumours

To predict shared neoantigens amongst different LS and MMR deficient tumours we recovered commonly mutated coding MS sequences in these tumours derived from m1 (deletion of 1 nucleotide) and m2 (deletions of 2 nucleotides) variants. These sequences were primarily derived from tumour driver genes and were inputted as the full mutated protein plus 8 *wt* aminoacids.



*Figure 23. Summary of the methods for bioinformatic FSDN prediction and validation. A first phase of bioinformatic FSDN prediction was carried out by using the pVACbind pipeline on commonly mutated coding MS sequences in MSI tumours. Once an optimal pFSDN set was filtered out, an in-silico validation was performed by searching for the prioritized FSDN associated mutations in DNA sequences from LS-related tumours, sporadic MMRd, as well as DNA and RNA sequences from human tumoral CLs. A final phase of in vitro analysis was carried out to determine which of the pFSDN induced an inflammatory response on LS patients' mucosa infiltrating T lymphocytes. Abbreviations: HCB, Hospital Clinic of Barcelona; NMD, non-mediated decay; CrAD, colorectal adenomas; iDCs: immature dendritic cells; mDC, mature dendritic cells.*

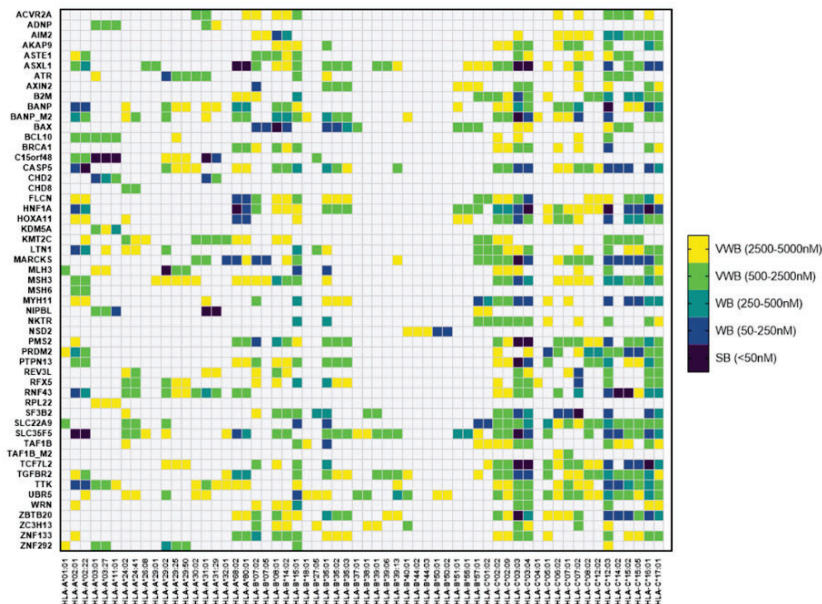
After computing the sequences through the pVACbind pipeline with all binding predictors enabled, we identified a total of 42,828 HLA-I-restricted and 8,350 HLA-II-restricted epitopes which were filtered and reduced to 53 FSDN restricted in length to HLA-I and 45 to HLA-II (Figure 23), which derived from a total of 53 different frameshift mutations, following the filters mentioned in the methods section 3.2.1.1. The final set of genes associated with the pFSDN is detailed in Supplementary table 6.

Selected FSDN were classified as strong (SB, IC<sub>50</sub><50nM), weak (WB, IC<sub>50</sub><500nM) and very weak (VWB, IC<sub>50</sub><5000nM) affinity binders to HLA molecules (Figure 24) taking into account the median score from all binding predictors used. A significantly higher proportion of FSDN with an IC<sub>50</sub> of 500-2500 to HLA-I alleles was noticeable, whereas HLA-II-restricted FSDN had more similar proportions of SB, WB and VWB which can be due to the promiscuous nature of the HLA-II molecule or to the less reliable prediction algorithms for class II alleles.

For further characterization of the selected epitopes, we carried out a self-similarity analysis against the human proteome to assess the possibility of generating autoimmunity. Results determined that 8 of our prioritized FSDN (8.16%) had significant similarity to a self-protein (Table 7) with e-values ranging from 0.02 as the lowest to the established cutoff of 0.1. We also performed a prediction of FSDN sequence similarity to protein sequences from known human pathogens, which has been demonstrated to increase the immune response generated against neoepitopes due to cross-reactivity by molecular mimicry. Twenty-two FSDN (22.4%) were reported to have aminoacid sequence similarity to at least one of the human infectious pathogens' aminoacid sequences with positive immune assays (IEDB) analysed (Table 8), of which

40% had similarities to more than one pathogenic sequence. In this case, the cutoff for the bit-score was established at 15 and e-score at 0.5 due to a smaller sample size.

**A)**



**B)**

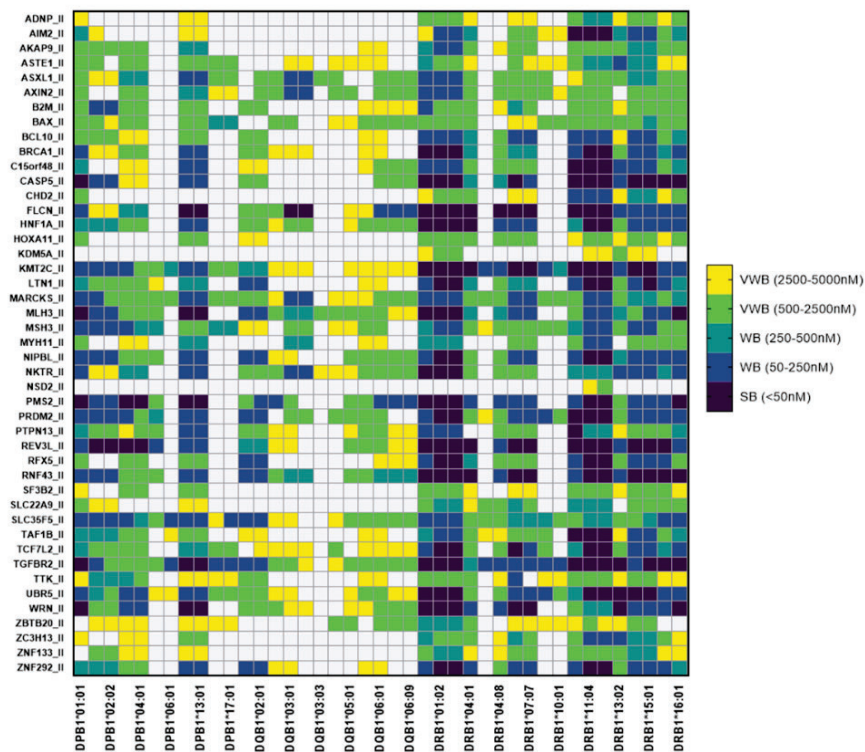


Figure 24. FSDN binding affinity to HLA-I (A) and HLA-II (B) tested alleles. Epitopes are classified and colour coded in strong ( $IC_{50} < 50$  nM), weak (tier 50 nM <  $IC_{50} < 250$  nM and tier 250 nM <  $IC_{50} < 500$  nM), and very weak (tier 500 nM <  $IC_{50} < 2500$  nM and tier 2500 nM <  $IC_{50} < 5000$  nM) binders.

Table 7. Self-similarity prediction results.

| FSDN            | Human Protein | e-value* | Bit-score** |
|-----------------|---------------|----------|-------------|
| RNF43_HLA-II    | COL5A3        | 0.02     | 29.1        |
| NKTR_HLA-I      | CAND1         | 0.051    | 26.5        |
| SF3B2_HLA-II    | ULK2          | 0.058    | 27.8        |
| ASXL1_HLA-II    | CACNA1        | 0.082    | 27.4        |
| C15orf48_HLA-II | AARS          | 0.082    | 27.4        |
| MLH3_HLA-II     | DCLK2         | 0.082    | 27.4        |
| REV3L_HLA-II    | FSHR          | 0.082    | 27.4        |
| SF3B2_HLA-I     | CDH3          | 0.1      | 25.7        |

Epitopes with a pairing e-value lower than 0.1 are shown. \*E-Value: Expect value (E); Parameter that describes the number of hits expected by chance when searching a database of a particular size. \*\*Bit-Score: Required size of a sequence database in which the current match could be found just by chance. Log2 scaled and normalized raw-score in which each increase by one doubles the required database size (2bit-score).

Table 8. Pathogen similarity prediction results.

| Epitopes                                              | Pathogen                 | Median Bit-score | Median e-value |
|-------------------------------------------------------|--------------------------|------------------|----------------|
| AKAP9_HLA-I, C15orf48_HLA-I, CHD2_HLA-II, FLCN_HLA-II | Dengue virus 2           | 17,8             | 0,3            |
| LTN1_HLA-II                                           | Dengue virus 4           | 17,2             | 0,5            |
| SLC22A9_HLA-I, TCF7L2_HLA-II                          | Hantaan virus            | 18,1             | 0,2            |
| HOXA11_HLA-I, HOXA11_HLA-II                           | Hepatitis C virus        | 18,9             | 0,1            |
| C15orf48_HLA-I                                        | Human betaherpesvirus 5  | 16,3             | 0,5            |
| C15orf48_HLA-I                                        | Human gammaherpesvirus 4 | 17,6             | 0,2            |
| SLC35F5_HLA-I, SLC35F5_HLA-II, TCF7L2_HLA-II          | Human gammaherpesvirus 8 | 19,4             | 0,1            |
| BANP_M2_HLA-I, RNF43_HLA-I                            | Human hepatitis A virus  | 17,1             | 0,3            |

|                                                                                     |                                   |      |     |
|-------------------------------------------------------------------------------------|-----------------------------------|------|-----|
| ATR_HLA-I, AXIN2_HLA-I,<br>RPL22_HLA-I, TAF1B_HLA-II,<br>TGFB2_HLA-II, ZBTB20_HLA-I | Human immunodeficiency<br>virus 1 | 17,6 | 0,2 |
| CHD2_HLA-I, LTN1_HLA-I                                                              | Human papillomavirus type<br>16   | 17,4 | 0,3 |
| BAX_HLA-I, BAX_HLA-II                                                               | Human papillomavirus type<br>18   | 17,2 | 0,3 |
| FLCN_HLA-I                                                                          | Mycobacterium tuberculosis        | 16,8 | 0,5 |
| MSH3_HLA-I                                                                          | Plasmodium falciparum             | 16,3 | 0,5 |
| C15orf48_HLA-I                                                                      | Trypanosoma cruzi                 | 16,3 | 0,5 |
| CHD2_HLA-I                                                                          | West Nile virus                   | 16,8 | 0,3 |

Lastly, 44.23% of all FSDN have been predicted to be resistant to nonsense-mediated RNA decay, a mechanism that degrades nonsense mRNA usually derived from frameshift mutations that give rise to immunogenic neopeptides<sup>220,247</sup>. Resistance to this process increases the probability of epitopes being presented in MMRd cells.

## 4.2 Occurrence of FSDN variants in LS pre-cancers and cancers

### 4.2.1 Analysis of variants from targeted cMS NGS sequencing from LS patients CrAD and EC samples

#### 4.2.1.1 NGS sequencing data analysis

Sequencing FFPE samples from 17 CrAD and 8 EC (Table 5) from patients with LS was done using a molecular barcoding-based NGS method high sensitivity MSI panel consisting of 288 cMS-containing genes and covering 27 of the selected coding MS corresponding to 40 pFSDN<sup>234</sup>. The resulting reads had >71.17% specificity, a mean coverage >932,88 and a minimum average of unique reads of 84.09%. Studying their mutational patterns revealed that 71% of the coding MS readings had m1 deletions and 1% had 2p insertions, which alter the reading frame to M1 (72%); in contrast, 15% have two base deletions and 11% have one base insertions, which change the reading frame to M2 (26%). 2% of the mutations were synonymous (M0) (Figure 25A).

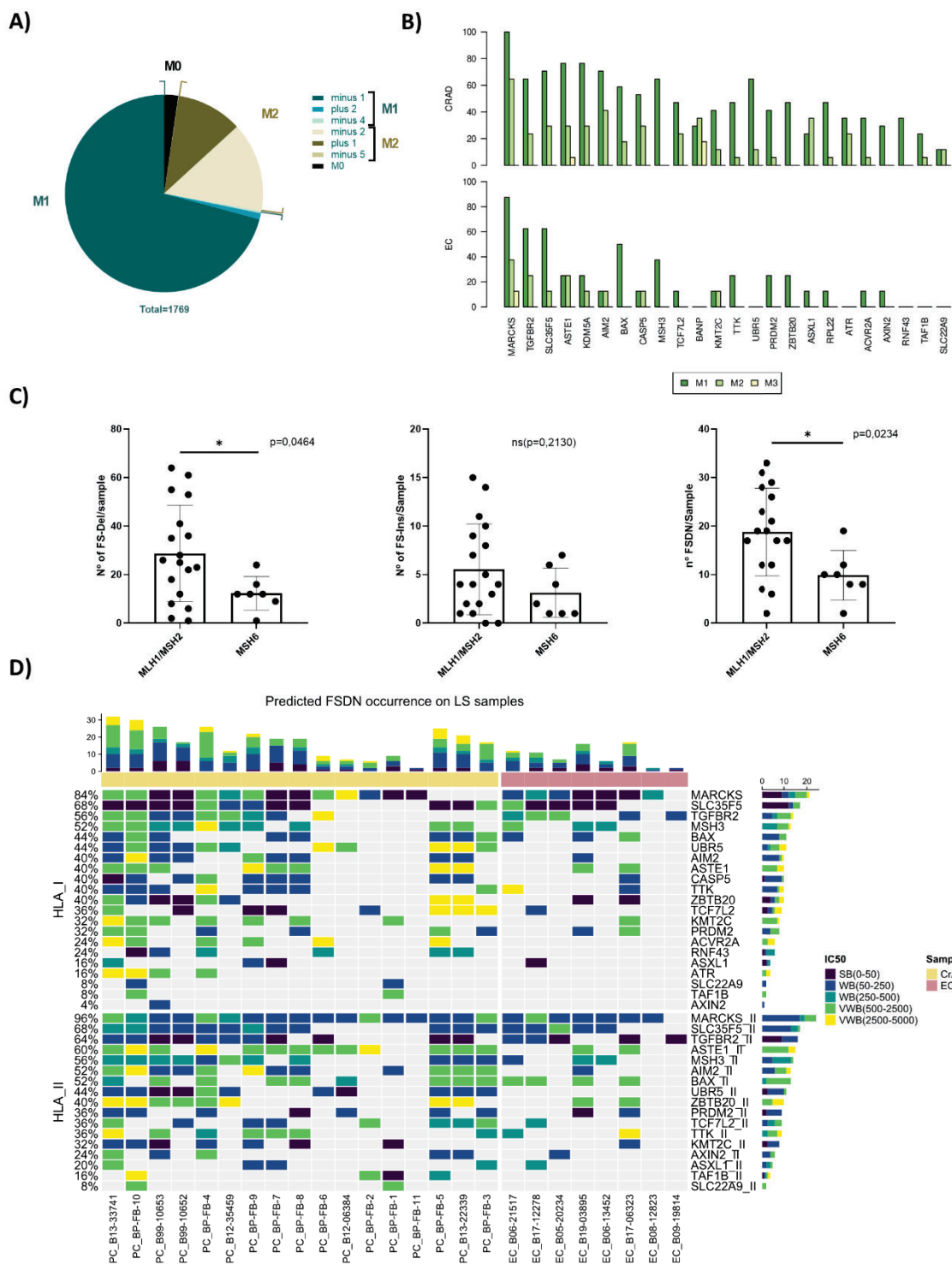


Figure 25. Analysis of LS-related colorectal adenoma (n=17) and endometrial cancer (n=8) samples. A) Pie-chart showing the distribution of all coding MS frameshift mutations (minus (m) 5 – plus (p) 2) and their corresponding frames (M0, black; M1, aquamarine; M2, beige). B) Mutational profile of LS EC (bottom) and CrAD (top) samples from coding MS targeted DNA sequencing classified by the mutation type (1bp deletion or 2 bp insertion, M1; 2bp deletion or 1 bp insertion, M2; synonymous mutation, M3). C) Barplots (mean with SD) showing the number of frameshift deletions (left), insertions (center) and number of predicted pFSDN expression (right) are represented by germline MMRd mutations. Unpaired T-tests were performed to determine significance (p value in APA style: ".123" ns, ".012" \*, ".001" \*\*, "<.001" \*\*\*). D) Analysis of LS-related colorectal adenoma (n=17) and endometrial cancer (n=8) samples. Heatmap of the pVACseq neoepitope prediction against each patient HLA haplotype. The color-code corresponds to the IC50 values for the strongest HLA allele binder from each patient classified as strong (SB), weak (WB), and very weak (VWB) binders.

The majority of coding MS alterations favour a frameshift to the M1 reading frame, which corresponds to the origin of 97% of our potential FSDN and is consistent with reports from other groups<sup>56</sup>.

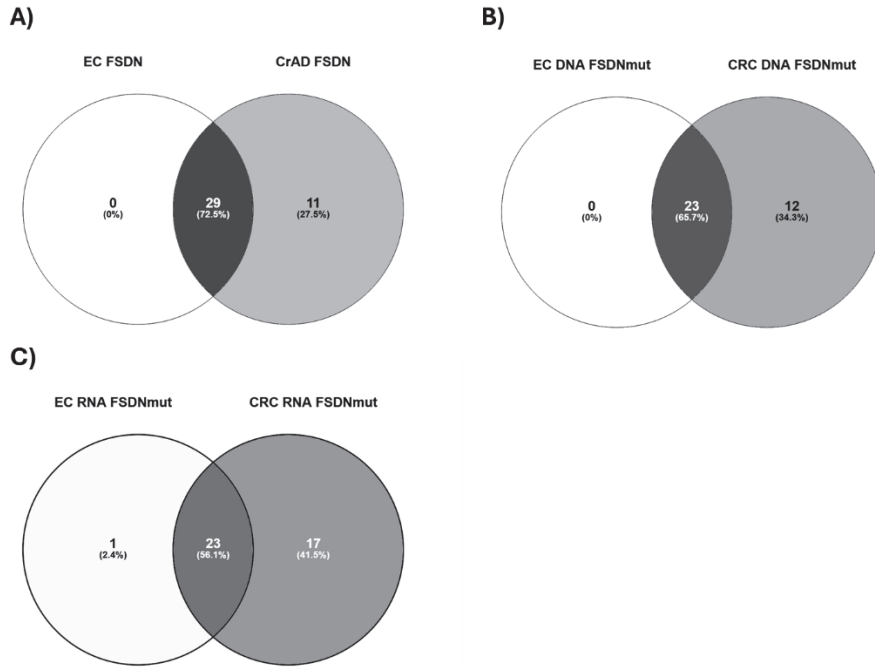
Mutational prevalence analysis showed that all CrAD and EC samples sequenced harboured pFSDN-associated mutations. Notably, *MARCKS*, *SLC35F5*, *TGFBR2*, and *BAX* variants appeared in over 50% of both CrAD and EC samples, underscoring the high prevalence of these shared mutations. Additionally, *ASTE1*, *KDM5A*, *AIM2*, *MSH3*, *URB5*, and *CASP5* mutations were frequently observed in more than 50% of CrAD samples (Figure 25B), but not EC samples.

We studied the mutational patterns based on the included patient's germline mutations. Interestingly, *MLH1* and *MSH2* pathogenic variant carriers had a significantly higher occurrence of pFSDN-associated mutations in CrADs and ECs compared to those with *MSH6* variants (p=0.0234) (Figure 25C).

#### 4.2.1.2 FSDN HLA-binding prediction for analysed variants

Computing sequenced data through the pVACseq neoantigen prediction software against each patient's HLA haplotype determined that 38/40 (95%) pFSDN had a high predicted probability of being presented on HLA-I or II molecules in CrAD samples, and 31/40 (77.5%) in ECs, with a median 4.92(2-7) predicted strong binding (<100 nM) [3.625(2-4.5) for ECs and 5.53(3-8) for CrADs] and 9.96(5-14) mid to very weak binding (100 to 5000 nM) [5.5(2.5-8.5) for ECs and 12.06(6-17) for CrADs] pFSDN per sample.

Twenty-four out of 40 (60%) pFSDN appeared in  $\geq 50\%$  of CrAD samples and 12 (30%) in  $\geq 50\%$  of EC samples. Moreover, 72.5% of analysed FSDN were shared among both sample types (Figure 26), underscoring the shared nature of these peptides. pFSDNs restricted to class I and II derived from *MARCKS*, *SLC35F5*, *TGFBR2* and *BAX* were the most prevalent considering both sample types (Figure 25D).



*Figure 26. Venn Diagrams of shared mutations between endometrial cancers and colorectal cancers or adenomas, derived from human samples or cell lines. A) Shared predicted pFSDN expression between sequenced ECs and CrADs from LS patients. B) Shared FSDN associated mutations on EC and CRC cell lines DNA. C) Shared FSDN associated protein mutations on EC and CRC cell lines RNA.*

#### 4.2.2 Analysis of WGS data from LS CrADs and CRCs sequences

The mutational pattern of WGS data from study HNPCC-Sys (phs001407.v1.p1, dbGaP), which included 127 CrAD, CRC, and paired normal distal colorectal mucosa samples from 11 LS patients, revealed a high proportion of frameshift deletions in MNR sequences in both CrAD and CRC samples (Figure 27). The average number of variants and InDel (median(IQR)) variants per sample was increased in CrADs (101.5, 60.5(21.13-69.04)) and in CRCs (50.5, 17.1(11.13-40.17)) samples compared to paired

control samples (38.5, 6.5(3.58-9.10)). No significant differences were found in the VAF of these mutations across sample types.

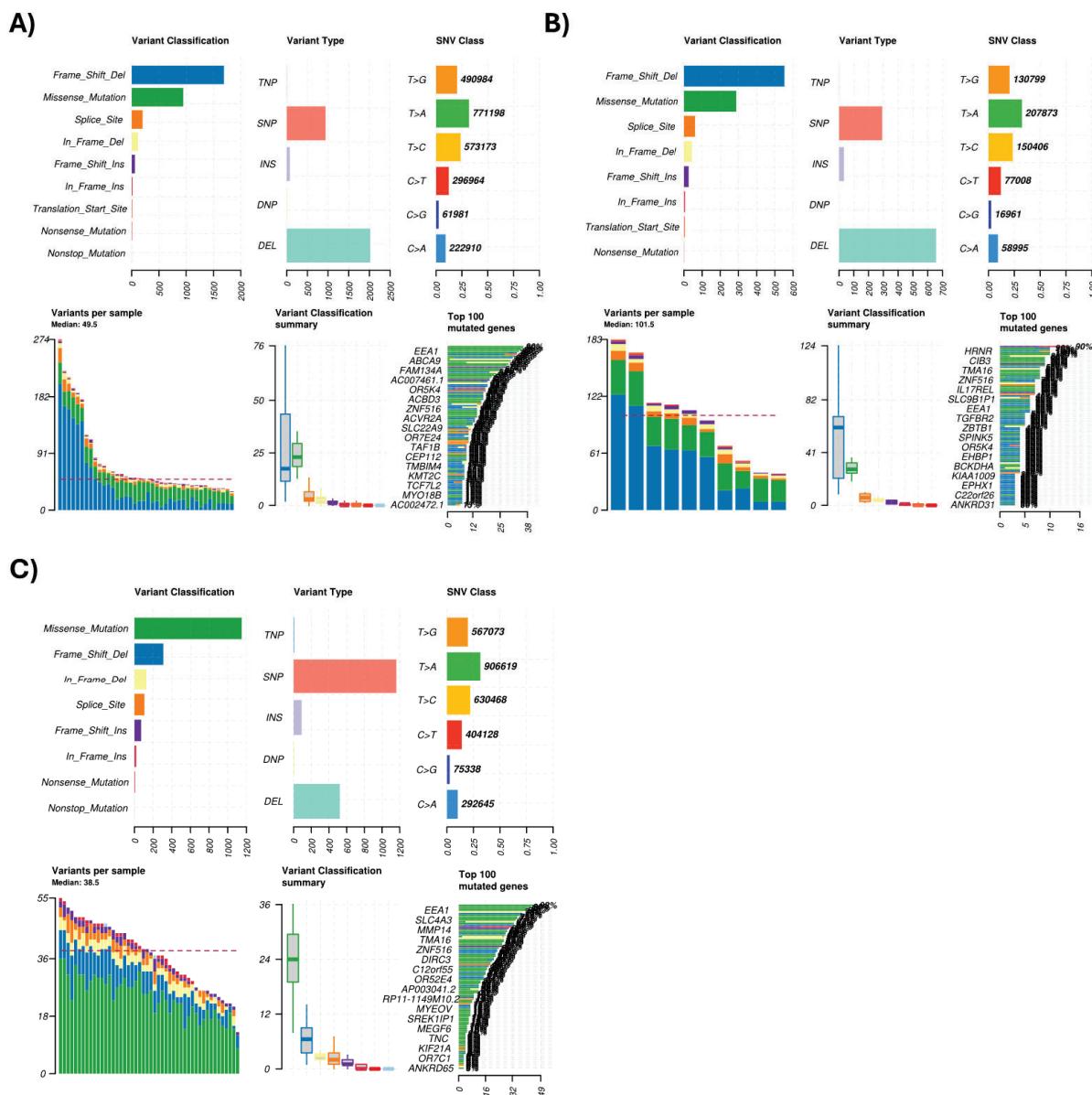


Figure 27. Analysis of mutational burden in HNPCC-sys colorectal tumour (A), adenoma (B) and normal mucosa paired samples (C). Oncoplots showing mutation type distribution.

Focusing on our list of pFSDN associated mutations, we found that 46 out of 53 (86.7%) were present on CRC samples. Specifically, *TGFBR2*, *SLC35F5*, *ACVR2A*, *ASTE1*, *TTK* and *ZNF292* were mutated in >25% of CRC samples, and *TGFBR2*, *SLC35F5*, *ACVR2A*, *ASTE1*, *LTN1* and *RPL22* were mutated in >35% of CrAD samples. The median of pFSDN associated mutations per sample type was 3.05(1-4) in control paired normal mucosa samples, 7.67(6.5-9) in CrAD samples and 6.02(2-10) in CRC samples (Figure 28A). Interestingly, *MLH1/MSH2* pathogenic variant carriers' samples were associated with a higher number of pFSDN associated mutations compared to *MSH6* carriers (5.39 (2-8) vs 2.11(1-3.5), respectively;  $p=0.0351^*$ ). Due to HLA typing data being unavailable for the patients included in this study, FSDN binding predictions could not be performed. (Figure 28B).

Data from HNPCC-Sys was first analysed by H. Binder et. Al.<sup>248</sup>, where they defined two patient groups: The group G1 (patients 1 to 6), with mainly germline mutations in *MLH1*, exhibited a higher percentage of InDels, MSI, and immune infiltrates. Meanwhile group G2 (patients 7 to 11), where germline mutations were mostly *MSH2* and *MSH6*, resembled MSS tumours. We also noticed that FSDN variants are considerably higher occurrence in tumour samples from group G1 ( $p=0.015$ , Figure 28C). This heterogeneity amongst LS patients' profiles could be important when assessing both FSDN variants occurrences and FSDN immunological responses.

In summary, the cMS that precede our selected FSDN are highly mutated in LS-related CRC, CrAD and EC samples which in turn increases the probability of these epitopes to be presented in the studied samples cells. A higher mutational rate on obliged oncogenes is related to increased related epitopes presence on tumours.

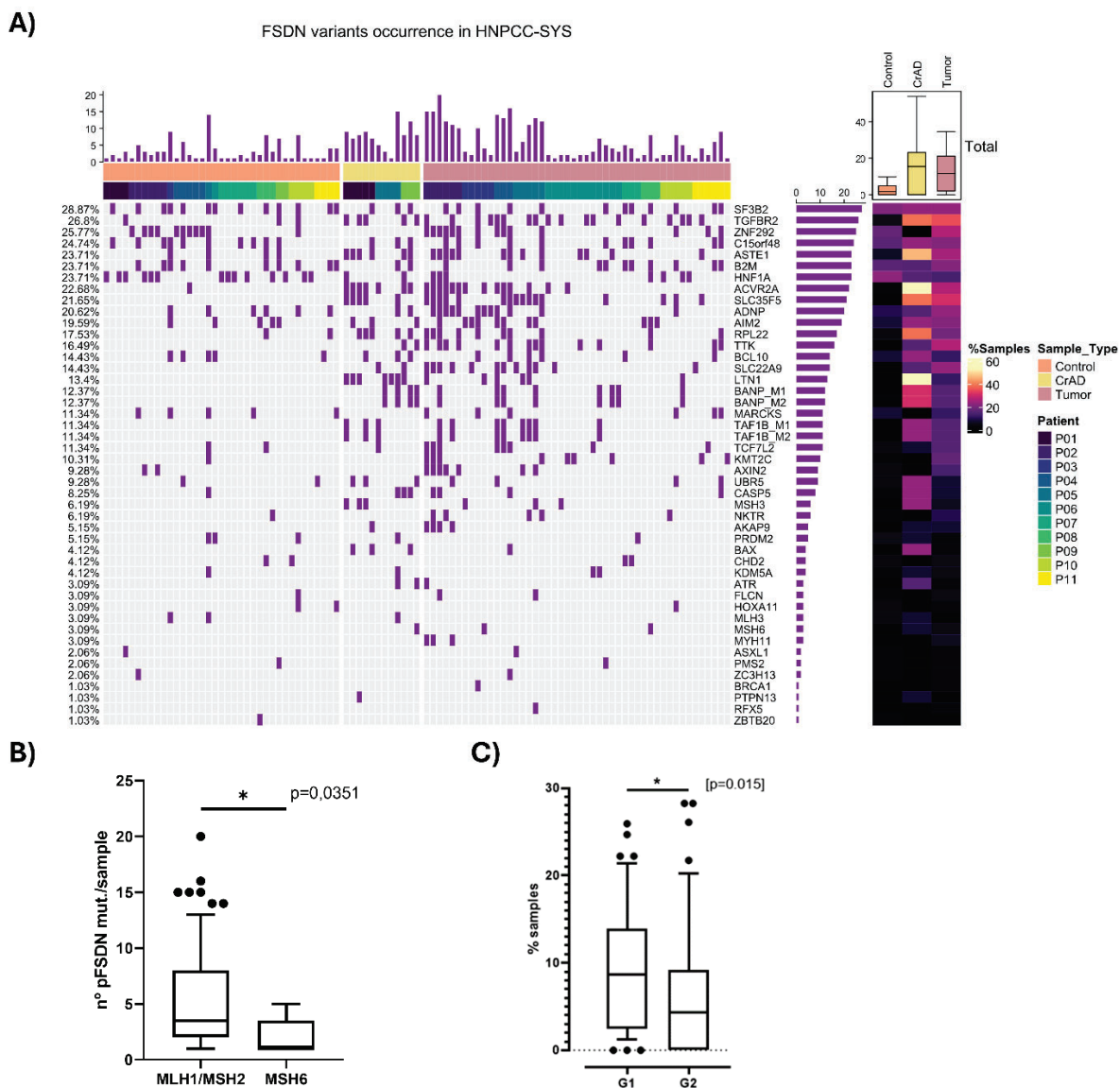


Figure 28. Occurrence of FSDN-associated variants in LS samples from the study HNPCC-Sys. Oncoplot depicting the occurrence of prioritized FSDN associated mutations in Control, CrAD and CRC samples. Top panel shows the number of variants detected in each sample. The right panel shows a heatmap (bottom) and a boxplot (top) of the relative percentages of mutated genes in Control, CrAD and Tumour samples. B) Boxplots (mean with SD and 95% percentile) showing the number of pFSDN associated mutations per sample by germline MMRd mutations. C) Boxplot showing the distribution of the occurrences of FSDN-associated variants in HNPCC-Sys G1 (patients 1 to 6) and

*G2 (patients 7 to 11) samples. Significance determined by paired T-test (p value in APA style: ".123" ns, ".012" \*, ".001" \*\*, "<.001" \*\*\*).*

### **4.3 Occurrence of FSDN variants in non-LS MSI-h cancers**

The presence of FSDN-deriving mutations in other tumours than those related to LS could increase the target options for the vaccine, including adjuvancy for MSI cancer treatment or relapse prevention. For this reason, we analysed public sequences from different types of non-LS, MMR deficient or proficient tumours, as well as human tumoral cell lines, in search of the mutations in coding MS that derive to our potential FSDN.

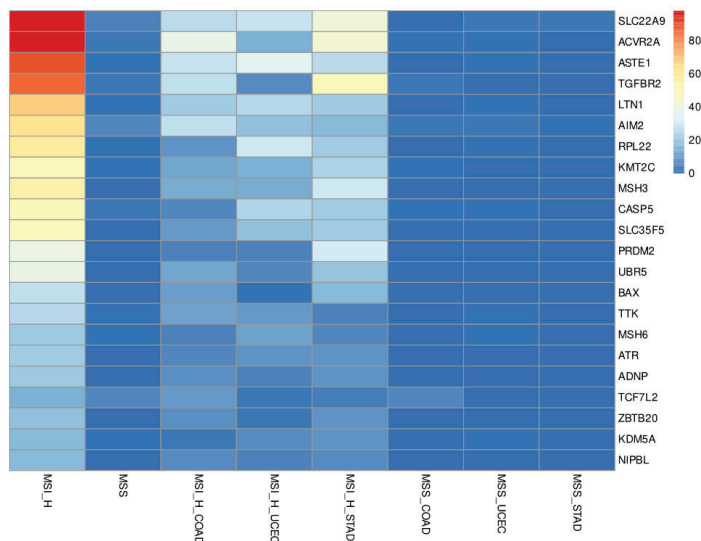
#### **4.3.1 pFSDN associated variants in non-LS MMRd human tumours**

To first assess if the pFSDN associated mutations were highly ranked in recurrently targeted MS loci and whether they were specific for MSI-H tumours, the mutations were searched in the top 1,000 most recurrently altered MS loci from 190 MSI-H and 640 MSS exomes from the TCGA datasets COAD, UCEC and STAD obtained from a study that focused on molecular portraits across different MSI cancer types<sup>227</sup> by comparing mutational coordinates in coding MNR sequences. As expected, we found that MSS and MSI-L samples contain little to no pFSDN associated mutations (mean  $0.2\% \pm 0.24$  samples/pFSDN mutation for all analysed genes), in contrast to a high prevalence on MSI-H samples, with a mean of  $23.15\% \pm 14.8$  samples per each pFSDN mutation in the top 1000 altered MS loci (Figure 29A,B). This is consistent with the reports from the mentioned study, where they demonstrate that the total numbers of MSI and frameshift MSI events are significantly higher in MSI-H tumours than in MSI-L or MSS tumours (Po10E-15; Kolmogorov-Smirnov test). We observed that 22 of our pFSDN mutations (41.5%) are highly ranked in the most altered MS loci on non-LS MMRd, MSI-h tumours. Focusing on the datasets, we observed a higher frequency of pFSDN mutations on STAD sequences than in COAD and UCEC, however this difference was not significant ( $p=0.08$ ).

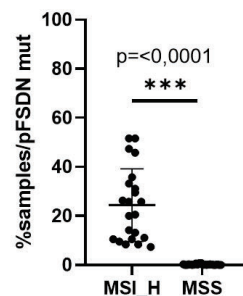
Afterwards, we deepened the analysis to assess the frequency of all pFSDN-associated mutations in MMRd, MSI-H sequences, independently on the high recurrence of their MS loci. To do so, we analysed the full whole exome sequences from both the already included and additional TCGA tumour datasets (COAD, UCEC and STAD, plus READ, CESC

and BRCA) and searched for pFSDN mutational coordinates in the coding MNR of these sequences.

A)



B)



C)



*Figure 29. Prioritized FSDN associated mutations presence in non-LS MSS and MSI-H samples. Relative number of samples (from the total samples in each set) containing the exact prioritized FSDN associated mutations in coding MS from TCGA datasets colon adenocarcinoma (COAD), uterine corpus endometrial carcinoma (UCEC), rectum adenocarcinoma (READ), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), stomach adenocarcinoma (STAD), and breast invasive carcinoma (BRCA). A) Heatmap representing the total counts of samples containing the pFSDNm in cMS from TCGA datasets COAD, UCEC and STAD, categorized by MSI-H (n=190) or MSS (n=640) status. B) Scatterplot (mean with SD) showing the differences in pFSDN associated mutations occurrence between MSI-H and MSS samples from COAD, UCEC and STAD datasets. Significance determined by T-test (p value in APA style: ".123" ns, "0.012" \*, ".001" \*\*, "<.001" \*\*\*). C) Oncoplot showing the occurrences of prioritized FSDN associated mutations on TCGA datasets. The frequency (percentages on the left and right barplot) is calculated by dividing the number of variants in the specific gene by the number of patients with at least one mutation identified. Top panel shows the number of aberrations detected in each sample. The heatmap on the right panel shows the relative percentages of pFSDN associated mutations in each dataset.*

We observed that 48 out of the 53 (90.76%) pFSDN associated mutations were present in the analysed sequences, with each of these mutations appearing in a median of  $16 \pm 33.03$  samples. From all datasets, TCGA-CESC, COAD, STAD and UCEC had more than 55% samples containing pFSDN associated mutations, with UCEC having the highest representation (80.35% of samples from n=168), and READ the least (18.18% samples from n=11). From all pFSDN mutations, *ACVR2A*, *RPL22*, *SLC22A9*, *ASTE1*, *RNF43*, *TGFBR2* and *ZBTB20* were mutated in most sets analysed (>25% samples) (Figure 29C).

These analyses demonstrate that MSI-h tumours that are not related to LS patients also contain pFSDN associated mutations, which implies that a high percentage of these mutations are important to the carcinogenesis and maintenance of the tumour, and that increasing immunogenicity against peptides derived from them may benefit not only LS patients, but also MSI-h tumour patients.

#### **4.3.2 pFSDN associated variants in MSI and MSS cancer cell lines**

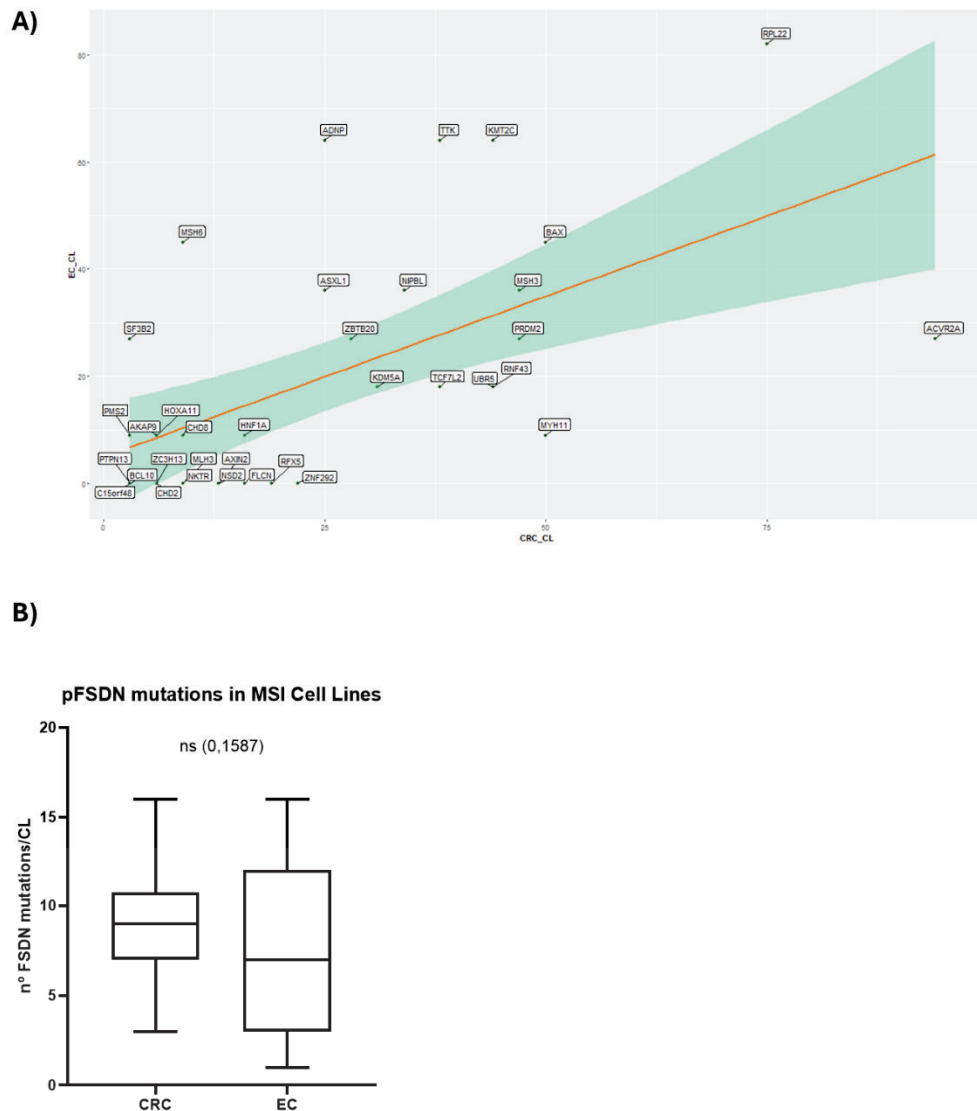
To broaden the range of our analysis, we included both DNA and RNA public sequences from CCR and EC cancer cell lines, either classified as MSS or MSI. Analysing the presence of FSDN variants in these cell lines can uncover useful models for the study of neoantigen-specific T-cells reactivity against a FSDN-expressing tumour target, as well as confirming the presence of FSDN variants in other tumours outside those related to

LS. Moreover, analysis of FSDN mutations on mRNA provides the crucial information of whether these variants have the capability of being translated into mRNA and not degraded, increasing the probability of being presented as neoepitopes in the cancer cell surface.

pFSDN associated variants were analysed in both DNA and RNA from a total of 141 MSI and MSS human tumoral cell lines (Supplementary table 2). DNA data was accessed through the public database Cell Model Passports (Sanger institute, <https://cellmodelpassports.sanger.ac.uk>), whereas RNA data has been downloaded from the Sequence Read Archive (SRA) from different groups that posted cell lines transcriptomics information. The differences in sequencing timepoints and passaging numbers can alter the variants found in the RNA compared to DNA.

#### 4.3.2.1 *Analysis of FSDN mutations occurrence on CL DNA sequences*

Cell lines from ECs and CRCs with MSI harboured up to 35 out of the 53 (66%) pFSDN associated variants in DNA sequences, with a median of 8.88 (7-10.25) mutations per MSI CRC CL and 7.1(3-10.5) per MSI EC CL. No statistically significant differences were observed amongst the number of pFSDN associated mutations between CRC and EC MSI CLs, but the correlation and trend line between sample types was significant (p-value = 0.0002441) (Pearson correlation and Pval) (Figure 30A). Despite this, we found that 65.7% of the FSDN associated mutations were shared between both sample types (Figure 26). When comparing variants between MSI-CL and MSS-CL we observed that little to no mutations in FSDN-cMS were found in MSS samples (Figure 30B), similarly to what we saw in TCGA datasets. Importantly, we found just one FSDN-cMS altered in normal tissue-derived CL (KMT2C), which indicates that our FSDN appear only in non-healthy tissues and relates to the previously observed absence of epitope sequence similarity to self-proteins.



#### 4.3.2.2 Analysis of FSDN mutations occurrence on CL mRNA sequences

To ascertain that FSDN-related variants are transcribed into mRNA, we accessed public mRNA (cDNA) sequences from 25 of the previously analysed cell lines, 12 with established MSI and 13 with an MSS phenotype (Supplementary table 2).

Out of the 53 prioritized FSDN-related variants, 37 (69,8%) were transcribed in MSS or MSI cell lines, with a median (IQR) of 7(6-9.5) variants per MSI cell line and 8(6-9) per MSS cell line. Although we find a similar proportion of altered genes amongst MSS and MSI cell lines, there is a statistically significant increase in the allele frequency (AF) of each variant ( $p=0.0003$ ) amongst MSI CLs, with a median of 0.53(0.264-0.741) in MSI CLs versus a median of 0.12(0.058-0.199) in MSS CLs (Figure 31A,B).

Deepening the RNA analysis, we corroborated that the genes and exons that contain our variants are also expressed in different levels on cell lines (Figure 31C, D, E). At gene level, RPKMs for the genes associated to FSDN were counted in MSI and MSS CLs separately. Interestingly we observe a higher gene RPKM homogeneity between MSS cell lines than in MSI cell lines and an overall higher gene expression, however this may be due to parameters that we did not have access to, such as the quality of RNA samples, the passaging status of cell lines at the point of sequencing, and the sequencing methods used. No major differences are distinguished when looking at the expression of genes containing FSDNs, however we can observe a higher expression of RPL22 and B2M ( $>12$  RPKM median) in comparison to other genes. Looking a step further at the exon level, we confirm that most genes are expressed at high frequencies both in MSS and MSI cell lines, with an overall higher Log Count in MSS CLs. Moreover, a higher exon count is observed on ADNP, ASXL1, B2M, MARCKS, NSD2, RPL22 and ZNF292 (Log Count  $>7.5$ ).

Altogether, this analysis demonstrates that the variants associated to FSDN are expressed at DNA level and transcribed to RNA at higher frequencies in MSI EC and CRC tumour cell lines compared to MSS cell lines.

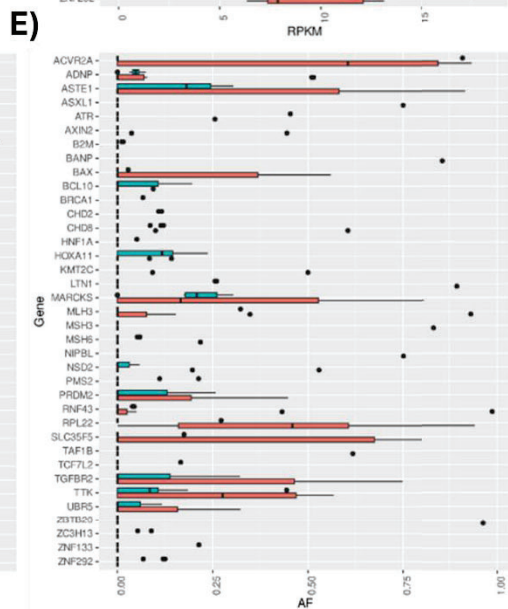
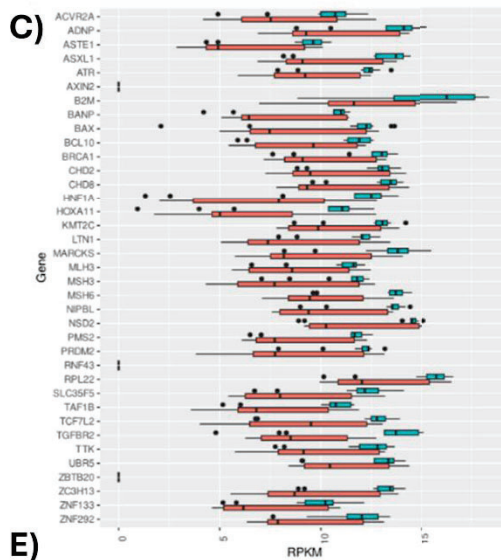
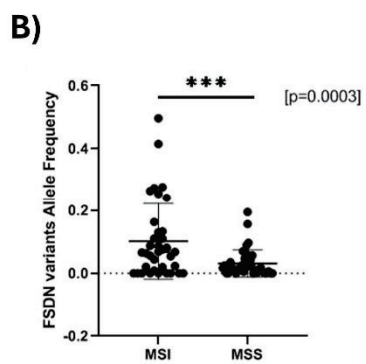
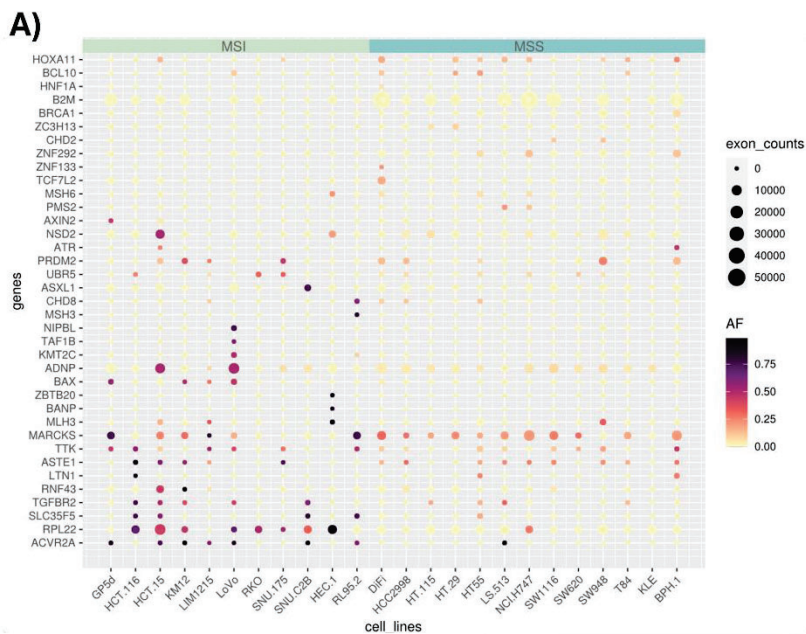


Figure 31. FSDN-related variants expression on MSI and MSS cell lines mRNA transcripts. A) Bubbleplot-heatmap depicting allele frequencies (0 to 100%) for all FSDN-related variants analysed as a colour scale, and exon-count as bubble size. Cell lines are separated by MS status (MSI: left, green; MSS: right, blue). B) Scatterplot (mean with SD) showing the mean AF for all the FSDN-related variants on MSI or MSS cell lines, significance determined by a paired T-test (p value in APA style: ".123" ns, ".012" \*, ".001" \*\*, "<.001" \*\*\*). C, D, E) Boxplot showing the distribution of the FSDN-related expression at gene level (C, RPKMs), at exon level (D, Log Counts) and at AF level (E, Allele frequency percentage).

#### 4.4 In-vitro bulk FSDN-loaded moDCs recognition assays

One of the primary challenges of FSDN evaluation is accurately identifying peptides capable of eliciting a robust immunogenic response. Many of the peptides that have a predicted high binding affinity end up performing poorly in experimental assays<sup>249–251</sup>; knowing that the capacity of a FSDN to be recognized in an immunogenic way is key to a successful vaccine and to enhance the accuracy of our pFSDN set, we stimulated previously expanded (polyclonally and non-specifically) TILs from CRCs and CrADs from LS patients with the 98 synthetic pFSDNs, loaded individually on autologous mature moDCs. We then quantified IFN- $\gamma$  and IL-10 release using bulk ELISPOT assays, establishing a low positivity threshold to discard those pFSDNs that do not induce a response in any of the tested patients. The characteristics from the included patients in this first assessment are summarized in Table 9.

Table 9. Characteristics of patients included in the assays for the prioritization of FSDN.

| Characteristics                          | Total (n=8)   |
|------------------------------------------|---------------|
| <b>Sex, n (%)</b>                        |               |
| Female                                   | 3 (37.5)      |
| Male                                     | 5 (62.5)      |
| <b>Age, median (range), years</b>        | 54.5 (40 -78) |
| <b>MMR gene germline mutation, n (%)</b> |               |
| MLH1                                     | 3 (37.5)      |
| MSH2                                     | 3 (37.5)      |
| MSH6                                     | 2 (25)        |
| <b>Previous cancer, type, n (%)</b>      |               |
| No                                       | 1 (12.5)      |

|                                                                  |          |
|------------------------------------------------------------------|----------|
| Yes                                                              | 7 (87.5) |
| <b>Types of previous cancer (among those with cancer), n (%)</b> |          |
| Colorectal Cancer (CRC)                                          | 6 (75)   |
| Endometrial Cancer (EC)                                          | 2 (25)   |
| Breast Cancer (BC)                                               | 1 (12.5) |
| Squamous cell carcinoma (SCC)                                    | 1 (12.5) |
| Small Intestine Cancer (SIC)                                     | 2 (25)   |
| Renal Cell Carcinoma (RCC)                                       | 1 (12.5) |
| Lymphoma                                                         | 1 (12.5) |
| <b>Sample type, n (%)</b>                                        |          |
| CRC                                                              | 4 (50)   |
| CrAD                                                             | 4 (50)   |

The heatmap in Figure 32A shows IFN- $\gamma$  ELISPOT results (SFUs/peptide) which varied greatly among patients. These differences likely reflect the hypermutator phenotype of LS patients and different HLA haplotypes. Our analysis revealed that 70 out of 98 (71%) pFSDNs elicited a positive IFN- $\gamma$  ELISPOT response (exemplified in Figure 32B), where seven out of 8 (87.5%) patients responded to FSDNs, with a median of 7.37(1-10.75) responding HLA-I pFSDN per patient [10.5(3.25-14.25) in CrADs and 4.25(0.75-5.5) in CRCs], and 6(2-5.75) for HLA-II pFSDNs [4(2.75-4.25) in CrADs and 8(1.5-10) in CRCs].

To confirm that no neoantigens were inducing anti-inflammatory cytokine release, we analysed IL-10 production from ELISPOT supernatants of peptides with high response rates. As shown in Figure 32C, these peptides did not induce IL-10 secretion by either mo-DCs or TILs, supporting their potential as immunogenic candidates.

We acknowledge several limitations of this approach, such as the low number of tested samples, the high background of ELISPOT cultures, which is likely due to co-culturing matured moDCs and T-cells on the same plate, and the lack of responses confirmation by flow cytometry measure of T-cell activation markers, which is due to the high sample numbers. However, it is worth noting that this analysis was primarily to entirely discard those peptides that were not recognized by any of the tested patients.

We did not find any significant differences in the number of peptides that induced an IFN- $\gamma$  response on any of the studied parameters when analysing the patients based on their clinical and sample type, probably due to a small cohort number, however a clear tendency on higher peptide response on patients with CrAD that had previously had cancer (excluding those patients with active cancer) was observed when compared to healthy LS carriers without previous cancer.

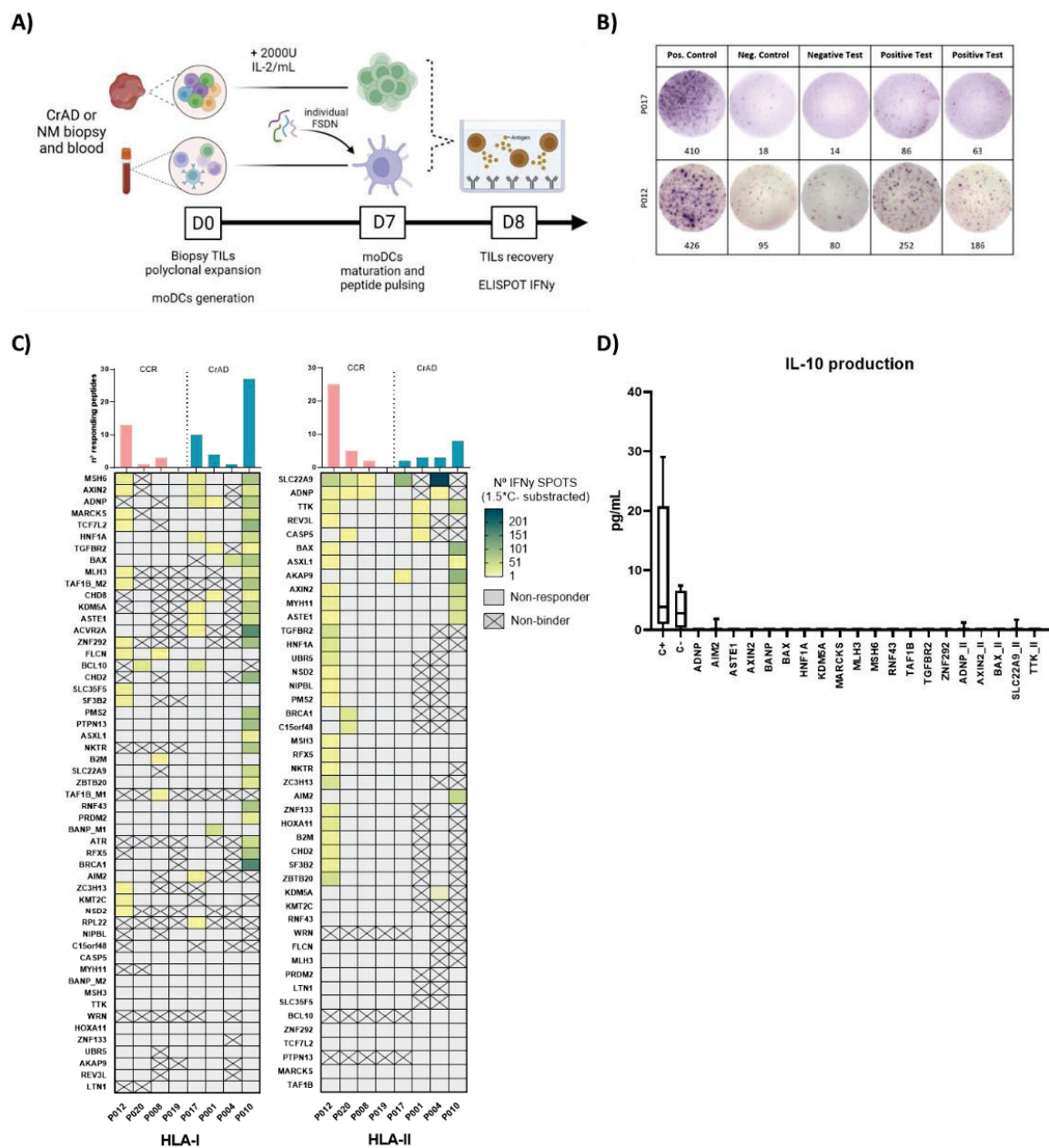


Figure 32. IFN-  $\gamma$  ELISPOTs in response to synthetic FSDNs. A) Schematic representation of the protocol used. B) Randomly picked examples of IFN- $\gamma$  ELISPOTs positive control, negative control, positive test and negative test wells. C) Heatmap depicting the number of spots per peptide (y axis) and patient (x axis) after subtraction of the negative control, separated by HLA-I restricted pFSDNs (left panel) and HLA-II restricted (right panel). Cells defined as “non-responder” in grey represent negative ELISPOTs for peptides predicted to bind to any HLA alleles from the corresponding patient ( $IC_{50} < 5000nM$ ), whereas crossed-out cells defined as “non-Binder” correspond to negative ELISPOTs for pFSDNs without any predicted compatible HLA alleles in the corresponding patient ( $IC_{50} > 5000nM$ ). Barplots on top of the plots represent the total number of peptides that induce a positive IFN-  $\gamma$  response for each patient. D) Boxplot showing pg/mL of IL-10 in TiLs and moDCs co-cultures supernatants recovered from IFN- $\gamma$  ELISPOTs. Graphic shows n=9 patients.

## 4.5 Prioritization of optimal FSDN

With the information gathered in both *in silico* and *in vitro* bulk analysis we searched to prioritize a smaller peptide set to conform the vaccine, which was named “DELAY” (DEndritic cell-based vAccine for LYnch syndrome). Additionally, reducing the number of FSDN to a more manageable size allows us to deepen *in vitro* studies and to find, isolate and characterize neoantigen-specific T-cells.

Considering all parameters studied, we prioritized the FSDN that met the following criteria (represented in Table 10):

- Good antigenic response, defined by a positive ELISPOT.
- Lack of aminoacidic sequence similarity to self-proteins.
- High coverage for HLA alleles
- Prediction of high occurrence in tumours related to LS or derived from sporadic MMRd (see Table 10).

Table 10. Details of the prioritized peptides. SB, Strong Binder; WB, Weak Binder; VWB, Very Weak Binder; NM-RNA Decay, Nonsense-mediated RNA Decay; LS, Lynch Syndrome; CrAD, Colorectal Adenoma; EC, Endometrial Cancer; CRC, Colorectal Cancer; sMMRd, sporadic MisMatch Repair deficiency; MSI, Microsatellite Instability; CL, Cell Line; AF, Allele Frequency.

| Peptide ID | Sequence            | IFN-γ<br>Response <sup>a</sup> (%<br>patients n=9,<br>>1.5°C) | HLA alleles binding affinities (IC50 nM) <sup>b</sup> |             |              |                |                 | Pathogen<br>sequences<br>similarity <sup>c</sup> | NM-RNA<br>Decay <sup>d</sup> | % Predicted expression in<br>sequenced LS samples <sup>e</sup> |          |             | % Mutated LS HNPCC-sys<br>samples <sup>f</sup> |       | % Mutated<br>sMMRD<br>samples <sup>g</sup><br>(n=290) | CL DNA <sup>h</sup><br>(n=43) | MSI CL RNA<br>(median AF,<br>n=11) <sup>i</sup> |
|------------|---------------------|---------------------------------------------------------------|-------------------------------------------------------|-------------|--------------|----------------|-----------------|--------------------------------------------------|------------------------------|----------------------------------------------------------------|----------|-------------|------------------------------------------------|-------|-------------------------------------------------------|-------------------------------|-------------------------------------------------|
|            |                     |                                                               | HLA-I                                                 |             |              |                |                 |                                                  |                              | HLA-II                                                         |          |             |                                                |       |                                                       |                               |                                                 |
|            |                     |                                                               | SB (0-50)                                             | WB (50-250) | WB (250-500) | VWB (500-2500) | VWB (2500-5000) |                                                  |                              | C/AD (n=17)                                                    | EC (n=8) | O/AD (n=13) | CRC (n=52)                                     |       |                                                       |                               |                                                 |
| HLA-I      |                     |                                                               |                                                       |             |              |                |                 |                                                  |                              |                                                                |          |             |                                                |       |                                                       |                               |                                                 |
| ACVR2A     | VVHKRGLF            | 40                                                            | 0                                                     | 0           | 0            | 6              | 7               | 0                                                | NO                           | 35.29                                                          | 12.5     | 53.8        | 26.9                                           | 45.17 | 76.74                                                 | 0.83                          |                                                 |
| ADNP       | AAKKRIPOK           | 50                                                            | 0                                                     | 0           | 0            | 4              | 1               | 0                                                | NO                           | N/A                                                            | N/A      | 15.38       | 23.08                                          | 6.55  | 34.88                                                 | 0.08                          |                                                 |
| AIM2       | KAKKHREV            | 20                                                            | 0                                                     | 1           | 3            | 4              | 7               | 0                                                | NO                           | 70.59                                                          | 12.5     | 23.08       | 21.15                                          | 20.00 | 0                                                     | 0                             |                                                 |
| ASTE1      | VLRTGEPL            | 33.3                                                          | 0                                                     | 0           | 0            | 8              | 6               | 0                                                | NO                           | 76.47                                                          | 25       | 46.15       | 25                                             | 29.31 | 0                                                     | 0.61                          |                                                 |
| AXIN2      | VPTCSPTL            | 42.9                                                          | 0                                                     | 1           | 0            | 6              | 6               | 3                                                | YES                          | 29.41                                                          | 12.5     | 0           | 13.46                                          | 4.83  | 9.30                                                  | 0.24                          |                                                 |
| BANP_M1    | FFPFCSV             | 11.1                                                          | 1                                                     | 3           | 5            | 6              | 11              | 0                                                | NO                           | N/A                                                            | N/A      | 30.77       | 15.38                                          | N/A   | 0                                                     | 0.86                          |                                                 |
| BAX        | WPWTRCLRM           | 25                                                            | 1                                                     | 5           | 1            | 6              | 3               | 1                                                | NO                           | 58.82                                                          | 50       | 23.08       | 1.92                                           | 8.97  | 48.84                                                 | 0.48                          |                                                 |
| HNF1A      | CAMDSLRPV           | 22.2                                                          | 4                                                     | 6           | 3            | 8              | 8               | 0                                                | YES                          | N/A                                                            | N/A      | 15.38       | 13.46                                          | 2.76  | 13.95                                                 | 0                             |                                                 |
| KDM5A      | SSQKKDPAGK          | 40                                                            | 0                                                     | 0           | 1            | 1              | 1               | 0                                                | YES                          | N/A                                                            | N/A      | 7.69        | 3.85                                           | 6.55  | 27.91                                                 | 0                             |                                                 |
| MARCKS     | RSAPFSRL            | 25                                                            | 1                                                     | 10          | 0            | 11             | 4               | 0                                                | NO                           | 100.00                                                         | 87.5     | 0           | 11.54                                          | 4.48  | 0                                                     | 0.53                          |                                                 |
| MLH3       | FLVMGELNY           | 40                                                            | 1                                                     | 2           | 0            | 5              | 7               | 0                                                | YES                          | N/A                                                            | N/A      | 7.69        | 0                                              | 2.76  | 6.98                                                  | 0.35                          |                                                 |
| MSH3       | FLALWECSL           | N/A                                                           | 2                                                     | 1           | 3            | 5              | 6               | 1                                                | YES                          | 64.71                                                          | 37.5     | 23.08       | 5.77                                           | 23.45 | 44.19                                                 | 0.83                          |                                                 |
| MSH6       | LLPEDTPPS           | 37.5                                                          | 0                                                     | 0           | 0            | 2              | 0               | 0                                                | YES                          | N/A                                                            | N/A      | 7.69        | 3.85                                           | 6.90  | 18.60                                                 | 0.22                          |                                                 |
| RNF43      | RFPITPPV            | 11.1                                                          | 2                                                     | 2           | 5            | 10             | 5               | 2                                                | NO                           | 35.29                                                          | 0        | 0           | 0                                              | 28.62 | 37.21                                                 | 0.43                          |                                                 |
| TAF1B_M2   | NRGLKKQY            | 66.7                                                          | 0                                                     | 1           | 0            | 5              | 9               | 0                                                | YES                          | 5.88                                                           | 0        | 23.08       | 15.38                                          | 3.10  | 0                                                     | 0.62                          |                                                 |
| TGFB2      | LSSCPVAL            | 25                                                            | 0                                                     | 2           | 4            | 14             | 9               | 0                                                | YES                          | 64.71                                                          | 62.5     | 38.46       | 34.62                                          | 28.28 | 0                                                     | 0.51                          |                                                 |
| TGFB2_B    | RLSSCPVA            | N/A                                                           | 1                                                     | 1           | 0            | 2              | 0               | 0                                                | YES                          | N/A                                                            | N/A      | 38.46       | 34.62                                          | 28.28 | 0                                                     | 0.51                          |                                                 |
| ZNF292     | VLTKQVIV            | 50                                                            | 0                                                     | 0           | 1            | 5              | 2               | 0                                                | NO                           | N/A                                                            | N/A      | 0           | 26.92                                          | 2.41  | 16.28                                                 | 0                             |                                                 |
| HLA-II     |                     |                                                               |                                                       |             |              |                |                 |                                                  |                              |                                                                |          |             |                                                |       |                                                       |                               |                                                 |
| BAX_II     | MGGRHPSWP<br>WTRCLR | 22.2                                                          | 0                                                     | 0           | 3            | 25             | 5               | 1                                                | NO                           | 58.82                                                          | 50       | 23.08       | 1.92                                           | 8.97  | 48.84                                                 | 0.48                          |                                                 |
| TTK_II     | FEKGEKNDLQ<br>LFWM  | 33.3                                                          | 0                                                     | 1           | 3            | 12             | 13              | 0                                                | NO                           | 47.06                                                          | 25       | 15.38       | 26.92                                          | 12.76 | 44.19                                                 | 0.44                          |                                                 |
| AXIN2_II   | ASRHLWLWGT<br>AGTPA | 22.2                                                          | 0                                                     | 5           | 2            | 22             | 3               | 0                                                | YES                          | 29.41                                                          | 12.5     | 0           | 13.46                                          | 4.83  | 9.30                                                  | 0.24                          |                                                 |
| ADNP_II    | SSKPAKKRLPC<br>KVT  | 66.7                                                          | 0                                                     | 0           | 2            | 6              | 7               | 0                                                | NO                           | N/A                                                            | N/A      | 15.38       | 23.08                                          | 6.55  | 34.88                                                 | 0.08                          |                                                 |
| SLC22A9_II | AAQKNLLVK<br>CSTC   | 71.4                                                          | 0                                                     | 0           | 6            | 9              | 5               | 0                                                | YES                          | 11.76                                                          | 0        | 15.38       | 23.08                                          | 32.07 | 0                                                     | 0                             |                                                 |
| ACVR2A_II  | EDMQEVVHK<br>KRGIF  | N/A                                                           | 0                                                     | 3           | 1            | 12             | 6               | 0                                                | NO                           | 35.29                                                          | 12.5     | 53.85       | 26.92                                          | 45.17 | 76.74                                                 | 0.83                          |                                                 |

- a. IFN- $\gamma$  Response: Percentage of patients (n=9) with a positive ELISPOT for a specific FSDN, defined as a number of SPOTS 1.5 times greater than the negative control.
- b. HLA alleles binding affinities (IC50 nM): Number of HLA alleles to which a specific FSDN can bind, classified by their affinity from most affine (SB, strong binding) to least (VWB, very weak binding).
- c. Pathogen sequences similarity: Number of known pathogen protein sequences to which a specific FSDN has significant similarity.
- d. NM-RNA Decay: Sensitivity of a specific FSDN to be degraded by the nonsense RNA mechanism, reducing its likelihood of being presented in a cell. Variants: YES = sensitive, NO = insensitive.
- e. Predicted expression in LS samples: Relative percentage of CrAD or EC patient samples with LS in which a specific FSDN is predicted to be presented, based on the HLA haplotype of each included patient. N/A: FSDN not included in the sequencing panel.
- f. % Mutated LS HNPCC-sys samples: Relative percentage of total CrAD or CRC patient samples with LS from the study HNPCC-sys analysed in which the mutation that produces a specific FSDN is found.
- g. Mutated sMMRd samples (n=366): Relative percentage of sequences from samples with sporadic MMR alteration (not LS) in which the mutation that produces a specific FSDN is found.
- h. Mutated MSI CL DNA (n=43): Relative percentage of DNA sequences from cell lines with microsatellite instability (MSI) in which the mutation that produces a specific FSDN is found.
- i. MSI CL RNA (median AF, n=11): Median allelic frequency of a specific variant associated with a FSDN. Median of 11 RNA sequences derived from cell lines with MSI.

Included in this prioritization are 3 new peptides (MSH3 and TGFBR2\_B, restricted to HLA-I, and ACVR2A\_II, restricted to HLA-II), which do not stem from our prediction and validation process. These peptides were kindly shared by Dr. Beatriz Carreño from Linette-Carreno Lab (Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia) and Dr. Jolanda M. De Vries from the Tumour Immunology Lab at the Radboud Institute for Molecular Life Sciences (Radboud, Nijmegen, The Netherlands), respectively, given their good immunogenic responses in their respective clinical trials (NCT01885702).

To corroborate that the selected peptide set has an optimal coverage on the population in terms of HLA alleles, we used the IEDB analysis resource<sup>252</sup> to calculate the fraction of individuals predicted to respond to the epitope set based on HLA genotypic frequencies and MHC binding. We analysed epitope coverage for both IC50 < 5000 nM binding alleles and for only stronger binding IC50 < 500 nM alleles, on the Spanish, European and global (World) populations.

As shown in Table 11 the coverage on Europe, Spanish and global populations is very high (>90%) for both HLA-I, HLA-II and the combined sets in all different analysis (Barplots for all coverage categories shown in Supplementary figure 1).

*Table 11. Prioritized epitope set population coverage in Spanish, European and World populations.*

| population/area                                                                                    | IC50 nM max | Class I            |                           |                   | Class II |              |       | Class combined |              |       |
|----------------------------------------------------------------------------------------------------|-------------|--------------------|---------------------------|-------------------|----------|--------------|-------|----------------|--------------|-------|
|                                                                                                    |             | cover <sup>a</sup> | average _hit <sup>b</sup> | pc90 <sup>c</sup> | cover    | average _hit | pc90  | cover          | average _hit | pc90  |
| Spain                                                                                              | 500         | 92.28%             | 6.94                      | 1.52              | 96.16%   | 3.33         | 1.61  | 99.70%         | 10.26        | 4.91  |
| Europe                                                                                             | 500         | 96.46%             | 7.26                      | 2.46              | 96.63%   | 3.29         | 1.62  | 99.88%         | 10.55        | 5.45  |
| World                                                                                              | 500         | 93.01%             | 6.15                      | 1.49              | 93.20%   | 2.78         | 1.2   | 99.52%         | 8.93         | 4.07  |
| Spain                                                                                              | 5000        | 99.88%             | 21.85                     | 11.65             | 99.98%   | 15.76        | 10.49 | 100.00%        | 37.61        | 26.02 |
| Europe                                                                                             | 5000        | 99.98%             | 25.48                     | 15.09             | 99.99%   | 16.47        | 11.4  | 100.00%        | 41.96        | 30.29 |
| World                                                                                              | 5000        | 99.77%             | 21.33                     | 10.74             | 99.92%   | 14.11        | 8.44  | 100.00%        | 35.44        | 23.35 |
| <sup>a</sup> projected population coverage                                                         |             |                    |                           |                   |          |              |       |                |              |       |
| <sup>b</sup> average number of epitope hits / HLA combinations recognized by the population        |             |                    |                           |                   |          |              |       |                |              |       |
| <sup>c</sup> minimum number of epitope hits / HLA combinations recognized by 90% of the population |             |                    |                           |                   |          |              |       |                |              |       |

#### 4.6 Peptide pooling for the vaccine product

Once we had the final prioritized peptides, we optimized the manufacturing process of the DELAY vaccine to ensure that the final product contains a homogenous mix of peptide-pulsed DCs. To that aim and considering that it was not feasible to pulse DCs with each of the individual peptides per separate for technical reasons, we divided the DELAY FSDNs into different pools taking into account their affinity to different HLA alleles to avoid peptide competition and over-representation.

Thus, the criteria followed for the FSDN division into pools was as follows:

- A. Each pool should have representation of different high-binding alleles for the most common alleles in the Spanish population.
- B. Each pool should avoid having 2+ peptides with similar HLA binding spectrum.

- C. Peptides should be divided in no more than 4 pools for technical reasons (cell number availability, efficient cell recovery, manufacturing time optimization).
- D. Each pool should contain an equal number of HLA-I and HLA-II restricted peptides.

Based on this, we separated peptides in 4 pools of 6 peptides per pool as follows:

| PEPTIDE | SEQUENCE        | POOL | PEPTIDE   | SEQUENCE        | POOL |
|---------|-----------------|------|-----------|-----------------|------|
| MSH3    | FLLALWECSL      | A    | MLH3      | FLVMGELNY       | B    |
| HNF1A   | CAMDSLRPV       | A    | BANP      | FFFPFFCSV       | B    |
| ZNF292  | VLTKQKVLY       | A    | AXIN2     | VPTCSPRTL       | B    |
| ADNP    | AAKKRLPCK       | A    | ACVR2A    | VVHKKRGLF       | B    |
| ASTE1   | VLRTEGEPL       | A    | ACVR2A_II | EDMQEVVVHKKRGLF | B    |
| BAX_II  | MGGRHPSWPWTRCLR | A    | AXIN2_II  | ASRHHLWGATAGTPA | B    |

| PEPTIDE    | SEQUENCE      | POOL | PEPTIDE  | SEQUENCE        | POOL |
|------------|---------------|------|----------|-----------------|------|
| RNF43      | RFFPITPPV     | C    | TGFBR2_B | RLSSCPVA        | D    |
| MARCKS     | RSAFPSSL      | C    | BAX      | WPWTRCLRM       | D    |
| TAF1B      | NRGLKKKQY     | C    | TGFBR2   | LSSCPVAL        | D    |
| AIM2       | KAKKKHREV     | C    | KDM5A    | SSQKDPAGK       | D    |
| MSH6       | LLPEDTPPS     | C    | ADNP_II  | SSKPAKKRLPCKVT  | D    |
| SLC22A9_II | AAQKNLLVCKSTC | C    | TTK_II   | FEKKGEKNDLQLFVM | D    |

The highest affinities in nM to each of the most common HLA alleles from each pool are summarized in the following table.

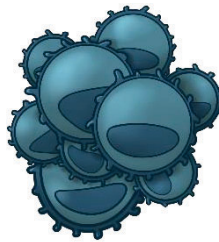
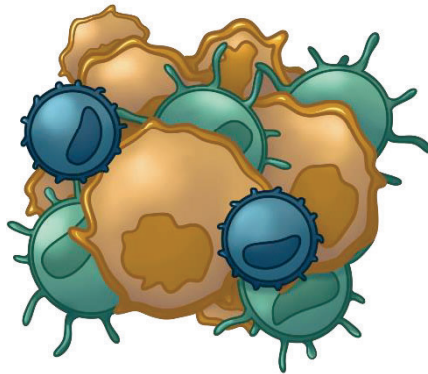
| Allele frequency<br>(%, Spain) | HLA Allele  | POOL (highest affinity, nM) |        |        |        |
|--------------------------------|-------------|-----------------------------|--------|--------|--------|
|                                |             | A                           | B      | C      | D      |
| 41.7                           | HLA-A*02:01 | 9.2                         | 57.5   | 160.5  | 71.6   |
| 31                             | DRB1*07:01  | 2793.9                      | 523.0  | 444.4  | 157.5  |
| 24.2                           | HLA-C*07:01 | 2772.9                      | 1645.6 | 1941.3 | 4688.0 |
| 23.7                           | DRB1*03:01  | 891.3                       | 191.6  | 342.6  | 512.2  |
| 20                             | HLA-C*05:01 | 3478.6                      |        |        | 1146.4 |
| 19.5                           | HLA-A*03:01 | 1028.7                      | 3579.5 |        | 2434.2 |
| 18.8                           | HLA-A*01:01 | 4307.5                      | 1799.2 |        |        |
| 17.4                           | HLA-A*24:02 |                             | 4903.4 | 897.0  |        |

|      |             |        |        |        |        |
|------|-------------|--------|--------|--------|--------|
| 17.2 | HLA-C*07:02 | 2693.1 | 312.9  | 283.3  | 4142.7 |
| 15.9 | HLA-B*18:01 |        |        | 3909.5 |        |
| 15.5 | DRB1*15:01  | 288.1  | 776.0  | 686.6  | 782.5  |
| 15.2 | HLA-C*16:01 | 46.3   | 151.7  | 232.5  | 450.3  |
| 15.1 | HLA-B*07:02 | 1828.6 | 62.2   | 113.0  | 92.6   |
| 15   | HLA-A*29:02 | 310.8  | 13.9   | 1718.9 |        |
| 14.5 | HLA-B*51:01 | 1375.4 | 2943.3 |        | 740.8  |
| 14.2 | HLA-C*06:02 | 1220.1 | 898.8  | 2440.7 |        |
| 14.1 | DRB1*01:01  | 515.7  | 152.8  | 1231.5 | 1287.4 |
| 13.8 | DRB1*13:01  | 692.3  | 56.1   | 291.8  | 348.6  |
| 13.5 | HLA-A*11:01 | 560.6  |        |        | 498.9  |
| 13.2 | HLA-C*12:03 | 11.9   | 14.1   | 132.8  | 345.8  |
| 12.2 | DRB1*11:01  | 1100.1 | 168.2  | 421.2  | 938.9  |
| 10.4 | HLA-B*08:01 | 449.3  | 897.3  | 207.3  | 39.9   |
| 7.9  | DRB1*11:04  | 713.5  | 116.9  | 291.8  | 348.6  |
| 7.7  | DRB1*13:02  | 1943.5 | 1279.8 |        | 3085.0 |
| 5.2  | DRB1*04:01  | 1959.2 | 766.0  | 2815.4 | 2460.4 |
| 3.4  | DRB1*16:01  | 1295.7 | 991.3  | 1417.4 | 2457.6 |



# CHAPTER 2:

## *IDENTIFICATION AND CHARACTERIZATION OF DELAY FSDN-SPECIFIC T-CELLS*





## 5 CHAPTER II: IDENTIFICATION AND CHARACTERIZATION OF DELAY FSDN-SPECIFIC T-CELLS

### 5.1 Patients and samples characteristics

#### 5.1.1 Demographics and HLA allele distribution analysis

For the preclinical analysis, a total of 82 individuals with LS have been recruited through the CRC-high risk vigilance clinic of the HCB. All individuals have available clinical history and those with available DNA samples have been HLA-typed by targeted NGS for the HLA-A, B, C, DRB1, DQB1 and DPB1 alleles. This was used both to study the Spanish LS population and to assess whether cells derived from these patients would be predicted to bind to the selected FSDN. The characteristics of included patients are summarized in Table 12.

Table 12. Characteristics of included patients in the pre-clinical studies from the DELAY vaccine.

| Characteristics                   | Total (n=82) |
|-----------------------------------|--------------|
| <b>Sex, n (%)</b>                 |              |
| Female                            | 50 (60.97)   |
| Male                              | 32 (39.02)   |
| <b>Age, median (range), years</b> | 50 (20 - 79) |
| <b>Germline mutation, n (%)</b>   |              |
| <i>MLH1</i>                       | 27 (32.93)   |
| <i>MSH2</i>                       | 29 (35.37)   |
| <i>MSH6</i>                       | 26 (31.7)    |
| <b>Previous cancers, n (%)</b>    |              |
| No (Previvors)                    | 40 (48.78)   |
| Yes (Survivors)                   | 42 (51.22)   |

As an initial analysis, we examined the relationship between previvor/survivor status and various phenotypic factors (Figure 33). As anticipated, we observed a significant positive correlation between age and cancer history ( $p = 0.0004$ ), with the incidence of malignancies increasing as individuals age. No significant correlations were found between the occurrence of malignancies and sex or germline mutations, however when

we investigated the occurrence of malignancies in relation to specific germline mutations, segregated by age range, we noted a distinct pattern in the behaviour of *MSH6* compared to *MLH1* and *MSH2* ( $p = 0.0115$ ) where individuals with *MSH6* mutations did not experience cancer cases until the age of 40, as also reported by others<sup>18</sup>.

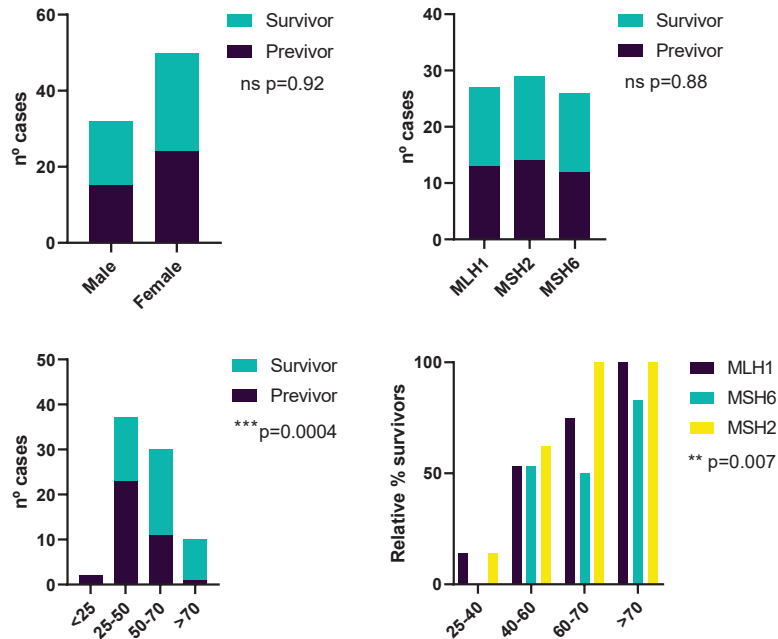


Figure 33. Analysis of cancer incidence among LS individuals in the preclinical DELAY studies. The bar plots display the number of survivor or previvor cases grouped by (from left to right) sex, germline mutation, and age range. The final bar plot shows the proportion of survivor cases relative to the total number of individuals with a specific germline mutation, further segmented by age range. Statistical significance was assessed using the Chi-square test.

Next, we analysed the HLA haplotype data with the R package MiDAS for immunogenetics data transformation and statistical analysis, to evaluate HLA frequencies in our cohort and the associations to different clinical factors such as germline mutations, sex and the previvor/survivor status.

HLA frequencies varied slightly from the expected pattern available at the HLA allele frequency net database (AFND)<sup>231</sup>, for example there is a very noticeable decrease on

HLA-A\*02:01 and HLA-C\*07:02 frequencies and a high representation of HLA-B\*18:01 and B\*51:01 in comparison to reference values, among others (Figure 34).



Figure 34. HLA allele frequencies from the individuals included in the preclinical DELAY studies. Barplots representing the frequency (0 to 1) of a determinate HLA allele in our LS cohort (n=82), separated by gene. Red dots indicate the frequency of a given allele on the HLA allele frequency net database (AFND), only used for available alleles.

The under-representation of HLA-A\*02:01, one of the most used alleles for the development of cancer vaccines, supports the hypothesis that a clinical trial where an inclusion criteria is to have this allele is potentially leaving out a very significant part of the population depending on their characteristics.

When searching for HLA alleles associations to different characteristics, we discovered several statistically significant correlations:

- The allele A\*01:01 was significantly associated with the male phenotype (p=0.01), with an estimated effect size (EES) of 1.45 (log-odds) and a 95% confidence interval (CI) ranging from 0.33 to 2.56. Conversely, the allele DRB1\*07:01 showed a significant association with the female phenotype (p=0.038, EES -1.07 (log-odds), 95% CI -2.07 to -0.06).
- The allele C\*07:02 was significantly associated to cancer previvors (p=0.031, EES -2.36 (log-odds) 95% CI -4.52 to -0.21), and the allele B\*44:03 exhibited a significant positive association with the survivor phenotype (p=0.046, EES 1.59 (log-odds), 95% CI 0.03 to 3.15).
- The allele DRB1\*11:01 was significantly negatively associated with the presence of the *MSH2* mutation (p=0.0068, EES -2.30 (log-odds), 95% CI -3.97 to -0.63), on the other hand, the allele A\*26:01 had a significant positive association with the *MSH6* mutation (p=0.031, EES 2.38 (log-odds), 95% CI 0.21 to 4.55).

### 5.1.2 Lynch syndrome patients T-cell populations characterization

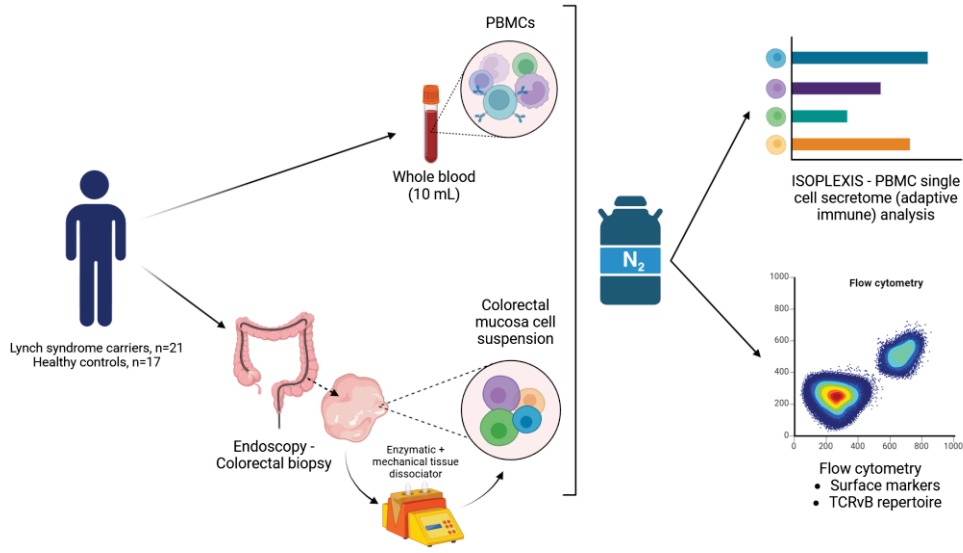
Starting to work with LS individuals, we first sought to investigate differences in T-cell subpopulations in the periphery and intestinal mucosa from healthy LS carriers compared to healthy controls. Several studies have reported differences in immune populations based on LS stage and status, but no deep analysis have been performed yet for healthy LS carriers, which may give insights on LS prevention strategies.

To that end, we recruited 21 LS individuals and 17 healthy controls from similar age ranges and with an equal proportion of male vs female and obtained 10 mL of peripheral blood and colorectal biopsies of healthy mucosa. Table 13 summarizes the characteristics of included patients in this study:

*Table 13. Characteristics of included individuals in the LS T-cell characterization study.*

| <b>Characteristics</b>                                       | <b>Total (n=38)</b> |
|--------------------------------------------------------------|---------------------|
| <b>Sex, n (%)</b>                                            |                     |
| Female                                                       | 18 (47.37)          |
| Male                                                         | 20 (52.63)          |
| <b>Status, n (%)</b>                                         |                     |
| Healthy control                                              | 17 (44.74)          |
| Lynch                                                        | 21 (55.26)          |
| <b>Age, median (range), years</b>                            | 53 (20 - 78)        |
| <b>Germline mutation, n (%)</b>                              |                     |
| <i>MLH1</i>                                                  | 5 (13.16)           |
| <i>MSH2</i>                                                  | 7 (18.42)           |
| <i>MSH6</i>                                                  | 9 (23.68)           |
| N/A                                                          | 17 (44.74)          |
| <b>Previous cancer (LS individuals), n (%)</b>               |                     |
| No                                                           | 11 (52.38)          |
| Yes                                                          | 10 (47.62)          |
| <b>Types of previous cancer (among those with cancer), n</b> |                     |
| Endometrial Cancer (EC)                                      | 4                   |
| Colorectal Cancer (CCR)                                      | 5                   |
| Urinary tract Cancer (CU)                                    | 1                   |
| Breast Cancer (BC)                                           | 1                   |
| Gastric Cancer (GC)                                          | 2                   |
| Ovarian Cancer (OC)                                          | 1                   |
| Papillary Renal Cell Carcinoma (PRCC)                        | 1                   |

Obtained samples were processed to obtain PBMCs from whole blood and a cell suspension from colorectal biopsies, following the methods described in sections 3.3.2.1 and 3.3.2.2.1, respectively. Once processed, samples were frozen in liquid nitrogen (-180°C) until use. The following scheme summarizes the protocol used in this section:



#### 5.1.2.1 Surface phenotype markers analysis

To first study the T-cell populations of LS carrier's healthy mucosa and peripheral blood T-cells compared to healthy donors, we phenotyped healthy mucosa cell suspensions and PBMCs from these cohorts by flow cytometry. In the literature, it has been demonstrated that LS carriers have a higher CD3+ CD8+ and FoxP3+ cell infiltration in colorectal mucosa compared to non-LS<sup>193</sup>. We defined CD8+ and CD4+ populations, and amongst them we differentiated between the subsets shown in Table 14 (TCMs, TN, TSCM, TEMRA, TEM, TTM and TRM based on different bibliography sources<sup>185,253,254</sup>.

Table 14. T-cell subsets defined by surface markers.

| Population   | Definition                                        | Phenotype                 |
|--------------|---------------------------------------------------|---------------------------|
| <b>TCM</b>   | T Central Memory                                  | CD45RA- CCR7+ CD28+ CD95+ |
| <b>TN</b>    | T Naïve                                           | CD45RA+ CCR7+ CD28+ CD95- |
| <b>TSCM</b>  | T Stem Cell Memory                                | CD45RA+ CCR7+ CD28+ CD95+ |
| <b>TEMRA</b> | Terminally differentiated Effector Memory T-cells | CD45RA+ CCR7- CD28-       |
| <b>TEM</b>   | T Effector Memory                                 | CD45RA- CCR7- CD28-       |
| <b>TTM</b>   | T Transitional Memory                             | CD45RA- CCR7- CD28+       |
| <b>TRM</b>   | Tissue Resident Memory                            | CD103+                    |

As shown in Figure 35, the percentage of CD4+ T-cells was increased in colorectal mucosa samples compared to CD8+ T-cells (mean 61.75% ± 8.77 for CD4+ and 22.95%

$\pm 7.42$  for CD8+,  $p < 0.0001$ , a difference which was not observed in peripheral blood (mean  $44.59\% \pm 14.81$  for CD4+ and  $42.35\% \pm 13.78$  for CD8+,  $p = 0.509$ ).

When comparing lymphocyte populations between mucosa and blood, regardless of LS status, TTM accounted for a larger portion of the total CD4+ T-cells in the mucosa ( $72.58\% \pm 8.35$ ) compared to peripheral blood ( $37.29\% \pm 10.50$ ). Notably, TEMRA were more prevalent in CD8+ T-cells than CD4+ T-cells and showed increased presence in peripheral blood. Predictably, TEM were highly represented in CD8+ T-cells, with a greater presence in mucosa than in blood. Finally, and as expected, TRM were only highly represented in the CD8+ mucosa compartment ( $27.34\% \pm 6.55$ ), with a very minimal representation of CD4+ in mucosa ( $3.74\% \pm 2.20$ ) and of circulating CD8+ T-cells ( $3.17\% \pm 8.08$ ).

Comparing healthy controls to LS carriers, we found a significant increase in the CD3+ proportion from total lymphocytes, normalized for the number of events analysed, in LS mucosa ( $p = 0.03$ ); however, no differences were found in the CD4/CD8 T-cell ratios.

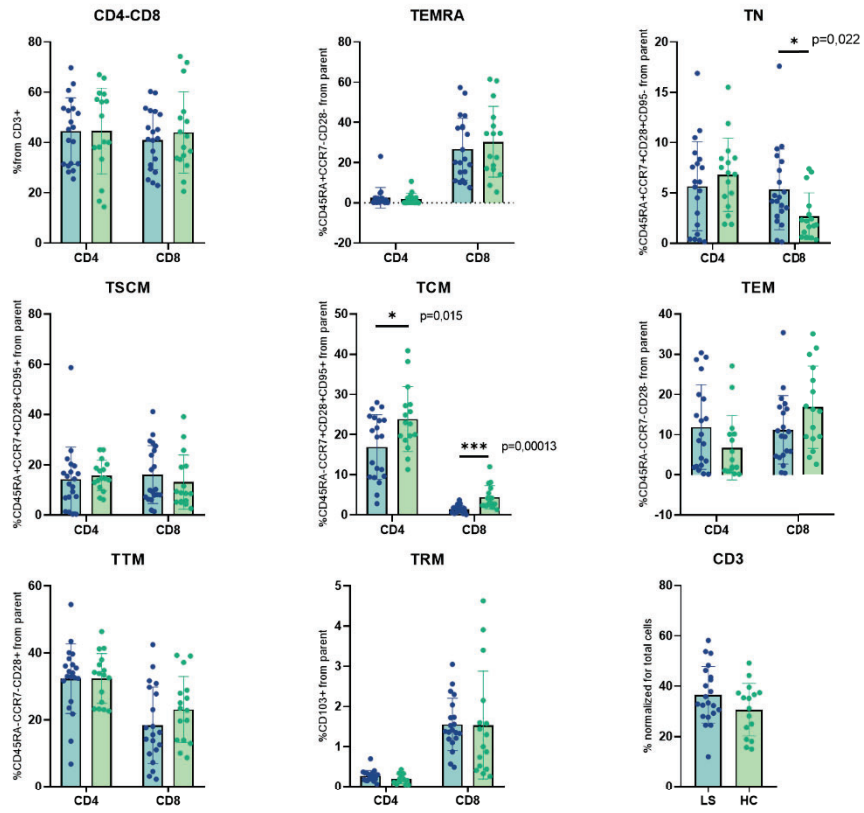
Regarding specific populations, TCM cells were decreased in CD4+ and CD8+ compartments from LS samples in both mucosa ( $p = 0.00009$  and  $p = 0.042$ , respectively) and in circulation ( $p = 0.015$  and  $p = 0.0001$ , respectively), compared to healthy controls.

Importantly, CD8+ TEM cell percentages were significantly elevated in the colorectal mucosa of LS carriers ( $p = 0.007$ ). Lastly, in PBMCs we found increased TN CD8+ cells ( $p = 0.022$ ) and decreased CD4+ T-reg cells (as defined by  $CD25^+CD127^{\dim}FOXP3^+$ ,  $p = 0.032$ ) in LS samples compared to healthy controls.

Regarding the expression of exhaustion markers, no significant differences were found in mucosa infiltrating lymphocytes. However, in LS PBMC samples compared to healthy controls, there was a mild but statistically significant decrease in CD8+ PD-1+ cells ( $p = 0.009$ ).

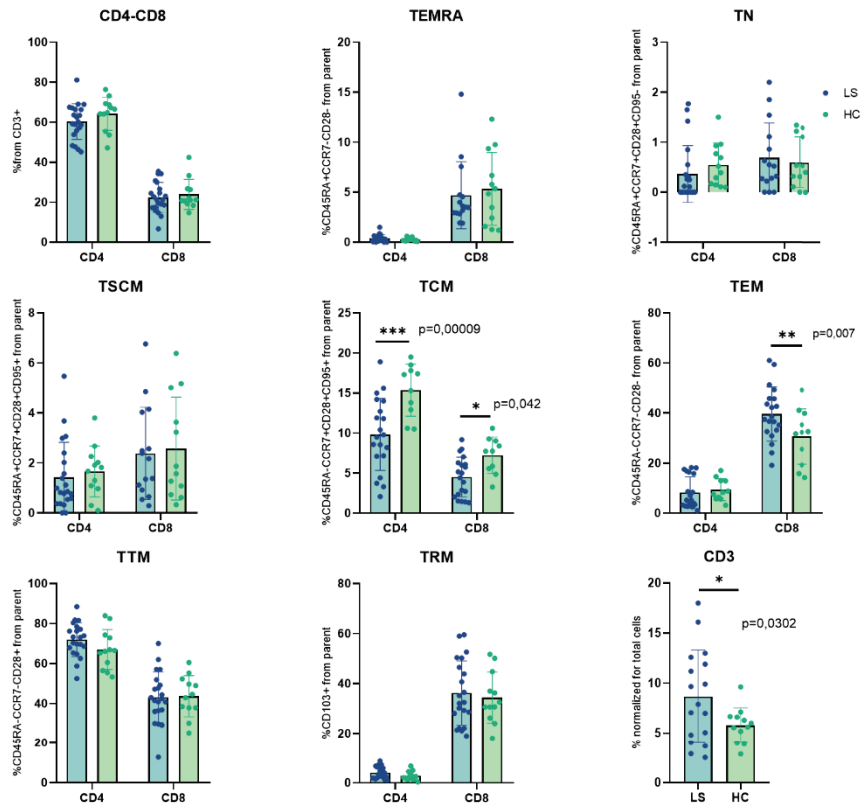
A)

Populations PBMCs

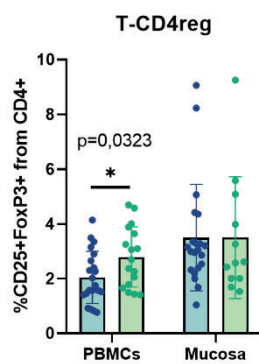


B)

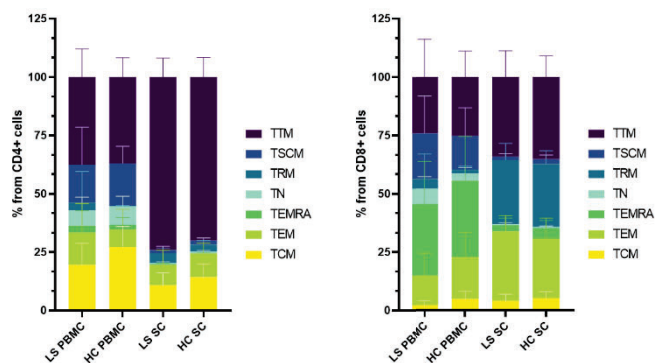
Populations - Colorectal Mucosa



C)

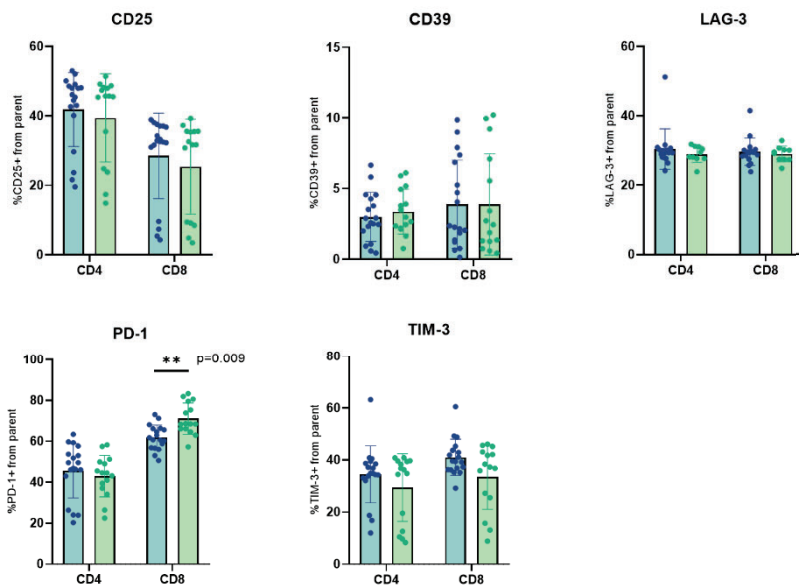


D)



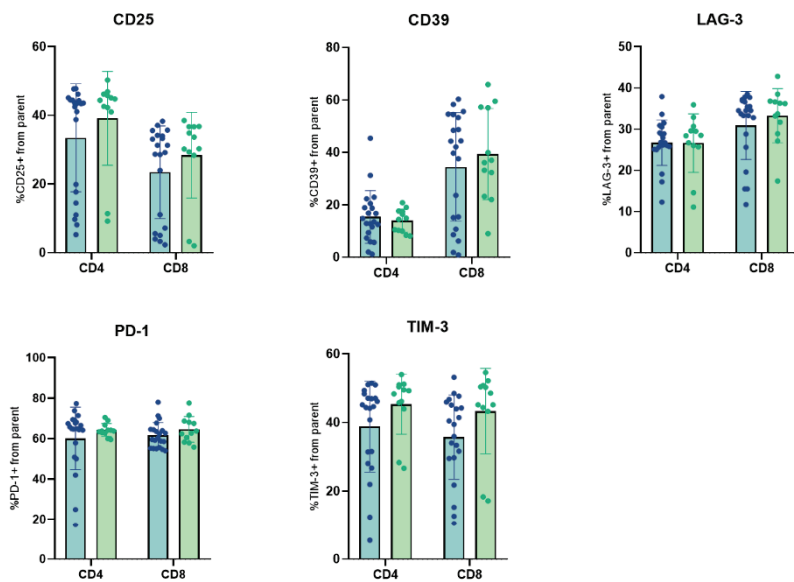
E)

**Exhaustion markers - PBMCs**



F)

**Exhaustion markers - Colorectal Mucosa**



*Figure 35. Phenotype of LS (blue) and healthy control (green) T-cells from peripheral blood (PBMCs) or colorectal mucosa. A, B) Barplots comparing Colorectal mucosa infiltrating T-cells (A) and PBMCs (B) population surface markers analysis. Populations are defined as a combination of surface markers as established in Table 14. C) Stacked barplot with SD for all PBMC and mucosa population compartments in CD4+ T-cells (left) or CD8+ T-cells (right). D) Barplot comparing CD4+ T-reg populations on mucosa and PBMC T-cells from LS (blue) and Healthy controls (green) samples. E, F) Barplots comparing exhaustion surface markers on PBMCs (E) and colorectal mucosa infiltrating T-cells (F). Normality and lognormality were assessed and statistical significance determined by multiple T-tests using the Holm-Sidak method.*

Lastly, some limitations must be addressed. First, we did not include markers to exclude non-lymphocyte populations and our antibodies for CD8+ molecules targeted CD8 $\alpha$ , which could pertain to a CD8 $\alpha\alpha$  receptor from NKT cells or IELs rather than the conventional CD8 $\alpha\beta$  from T-cells, and thus that the decrease in TCM cells may pertain to a higher number of NKT cells. Moreover, T-cells have not been separated by  $\alpha\beta$  or  $\gamma\delta$  lineages. However, the differences between LS and healthy controls in this analysis are robust and make room for deeper analysis on T-cell populations, as well as further studies on the presence of NKT or IEL cells in relation to LS.

#### *5.1.2.2 TCR V $\beta$ repertoire analysis*

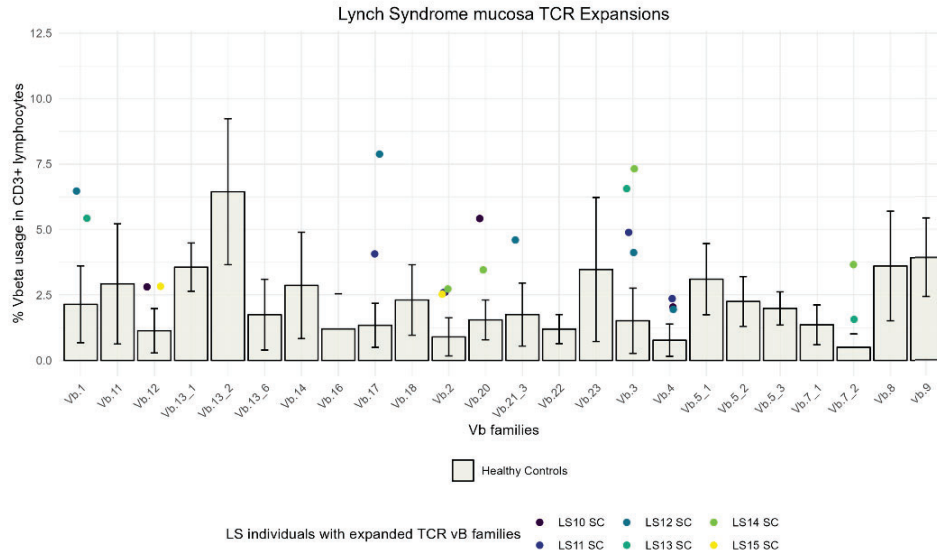
Next, we analysed the TCR V $\beta$  repertoire from PBMCs and CS samples using the IOTest Beta Mark Kit (Beckman Coulter), a multi parametric analysis tool designed for quantitative determination by flow cytometry. There are several limitations to this analysis, which is why this should be understood as a complementary analysis where results obtained need of further investigation to corroborate. First, the cohort of analysed samples is very limited, which decreases the reliability of taking the healthy control cohort as a reference. Second, TCR V $\beta$  family analysis should be corroborated by PCR, which we could not include due to the small sample size; however, scRNA sequencing for these samples will be performed in the future.

A standard comparison analysis of the raw data, averaging all samples from the same cohort, showed that the T cell population from PBMCs remained polyclonal with only a significant outgrowth of the V $\beta$ 8 clone ( $p=0.0040$ ) on Lynch Syndrome T-cells, in comparison to the healthy controls. However, in mucosa-infiltrating T-cells we see several differences, such as a decrease in V $\beta$ 13.2 ( $p<0.0001$ ) and V $\beta$ 9 ( $p=0.0089$ )

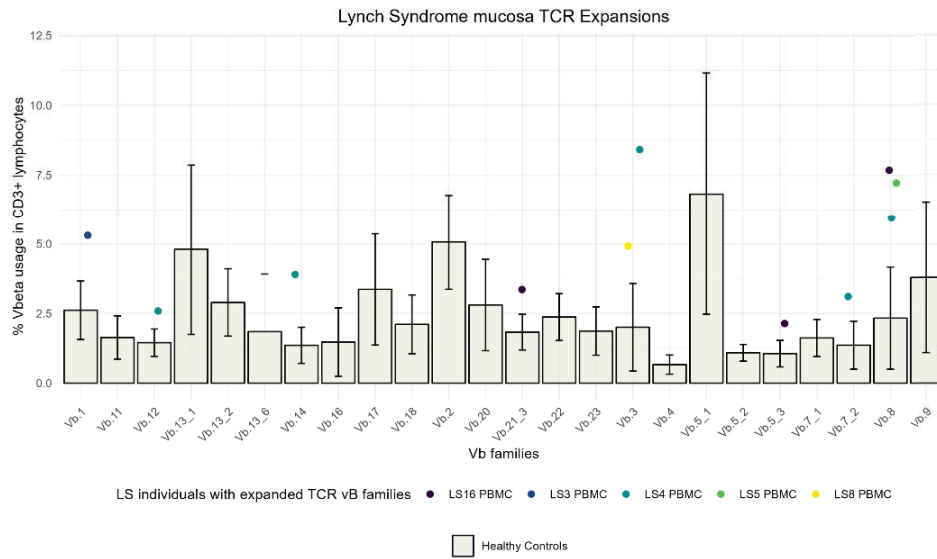
representation as well as an outgrow of Vb17 ( $p=0.0132$ ) on LS samples compared to healthy controls.

We also evaluated clonal expansions of TCR V $\beta$ -families by comparing frequencies on LS samples to the average + 2SD obtained using healthy controls as reference population. Clonal expansions of TCR V $\beta$ -families were observed within CD3<sup>+</sup> T cells in all LS individuals (Figure 36 A, B), with a more differentiated profile in mucosa infiltrating T-cells than in PBMCs. This can also be observed when comparing the number of patients with expansions per V $\beta$  family, which is significantly higher in LS colorectal mucosa than in PBMCs ( $p=0.027$ ), as well as the number of expanded V $\beta$  families per individual, that while not statistically significant a clear tendency of higher values on mucosa samples can be observed ( $p = 0.12$ ).

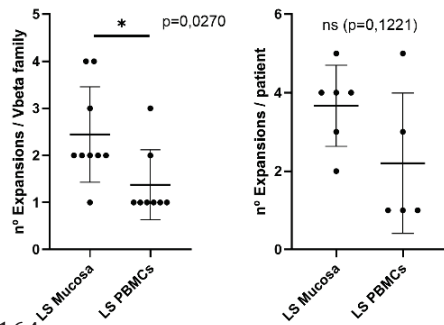
**A)**



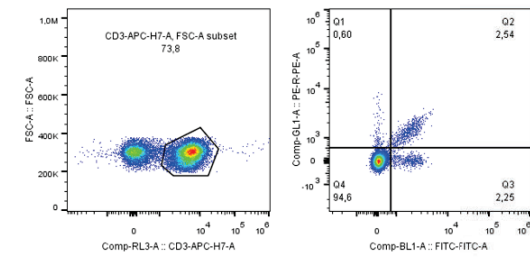
**B)**



**C)**



**D)**



*Figure 36. Expansions of T cell receptor (TCR) V $\beta$ -families within CD3<sup>+</sup> T cells from Lynch syndrome colorectal mucosa (A, n=6) or peripheral blood (B, n=5). Boxes and error bars represent the mean and SD of 24 TCR-V $\beta$  families from 5 (A) and 10 (B) healthy controls (HC). Dots correspond to expanded TCR V $\beta$ -families from LS individuals (frequencies > mean + 2SD from HC). C) Scatter plot comparing number of expansions per TCR V $\beta$  family (left) or per patient (right) between LS mucosa or PBMC samples. Normality and lognormality were assessed, and statistical significance was inferred with an unpaired T-test. D) Randomly picked example of the flow cytometry analysis for TCR V $\beta$  populations. CD3<sup>+</sup> lymphocytes are gated from total lymphocytes (left), from where a dot-plot gating PE and FITC (right) is derived. Quadrants Q1, Q2 and Q3 from that dot-plot determine different TCR V $\beta$  families depending on the kit staining (see methods).*

Next, we used Clustergrammer<sup>245</sup>, a web-based tool for visualizing and analysing high-dimensional data as hierarchically clustered heatmaps, to determine if the TCR V $\beta$  clonogram could differentiate LS T-cells from control cells. Using cosine distance and average linkage, the analysis identified two distinct clusters for mucosa samples, separating LS T-cells from controls (Figure 37). These clusters were evident in the heatmap and supported by the sample similarity matrix, which showed clear intragroup similarity within LS and control mucosa samples.

In PBMC samples, no distinct clustering was observed between LS and control groups. The heatmap and similarity matrices for PBMC samples displayed a lack of clear separation, indicating more overlapping TCR V $\beta$  repertoire profiles between these groups.

The clustering in mucosa samples was primarily driven by differences in specific TCR V $\beta$  families, as visualized in the heatmap. This data underscores the distinction in TCR V $\beta$  profiles between LS and control mucosa samples, which was not observed in the systemic PBMC samples.

As mentioned previously, further analysis needs to be performed in order to corroborate these results, and an amplified healthy control reference library should be considered.

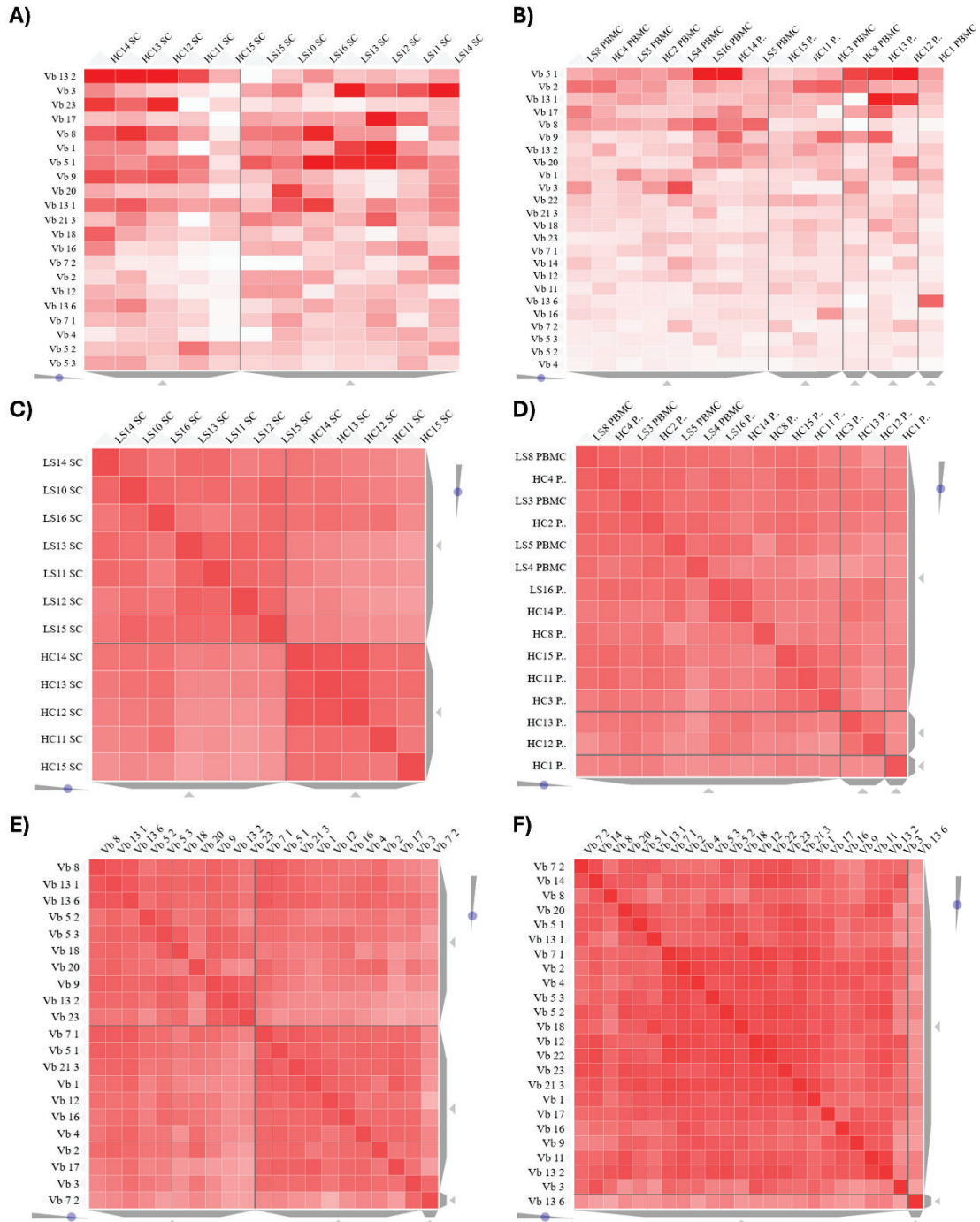


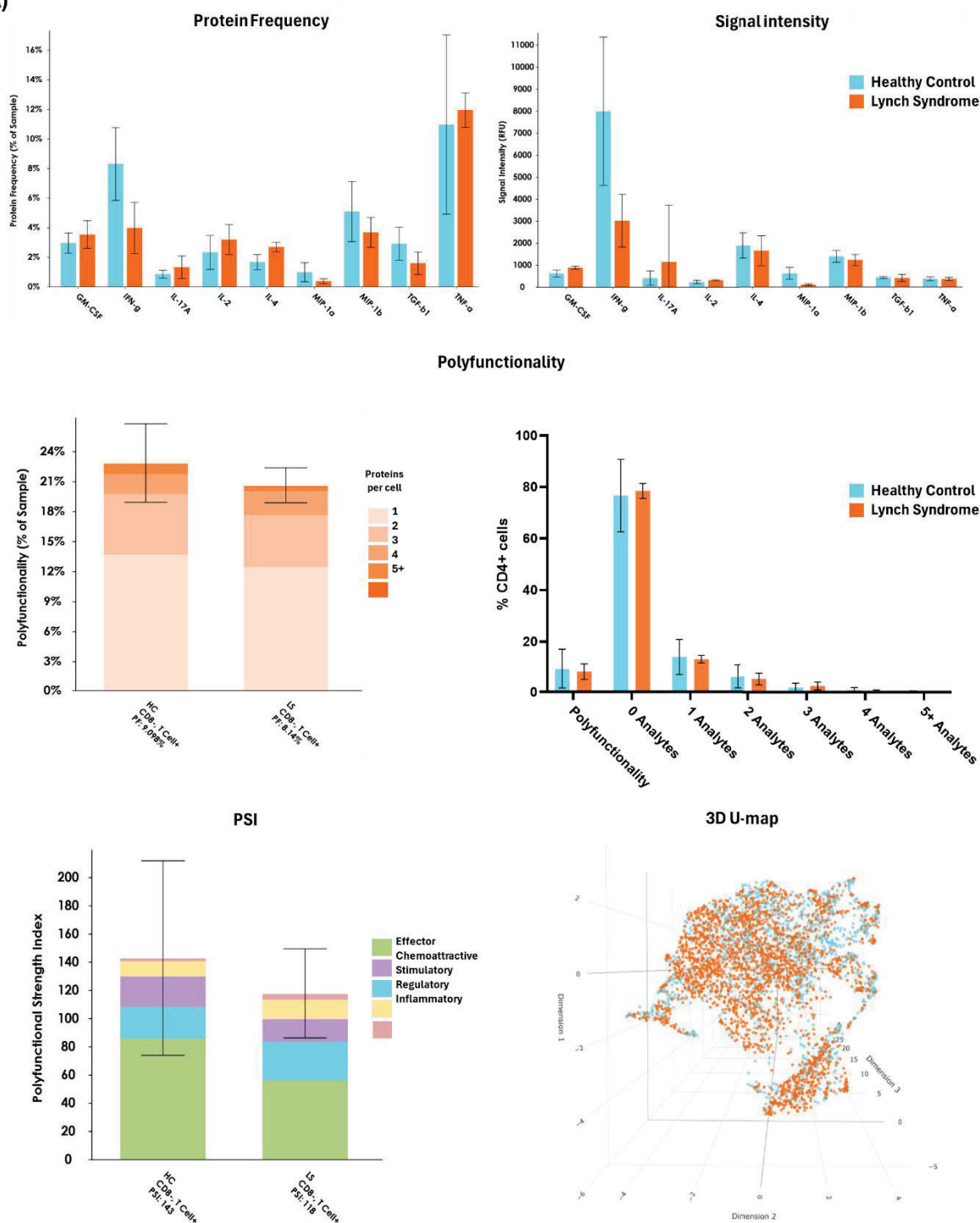
Figure 37. TCR Vβ clonogram analysis through clustered heatmaps by Clustergrammer. A, B) Clustergram/heatmap visualization of the TCR Vβ clonogram from mucosa (A) or PBMC (B) samples. The columns have been hierarchically clustered, using the Scipy library in Python, using cosine

distance and average linkage. The rows have been ranked by sum values. Matrix values are scaled from 0 (white) to 10 (red). C, D) Similarity matrix of the TCR V $\beta$  clonogram from mucosa (C) or PBMC (D) samples, iterated by the included individuals. The cells in the matrix represent the positive similarity between columns (measured as  $1 - \text{cosine-distance}$ ). E, F) Similarity matrix of TCR V $\beta$  clonogram from mucosa (E) or PBMC (F) samples, iterated by the TCRs. The cells in the matrix represent the positive similarity between rows (measured as  $1 - \text{cosine-distance}$ ).

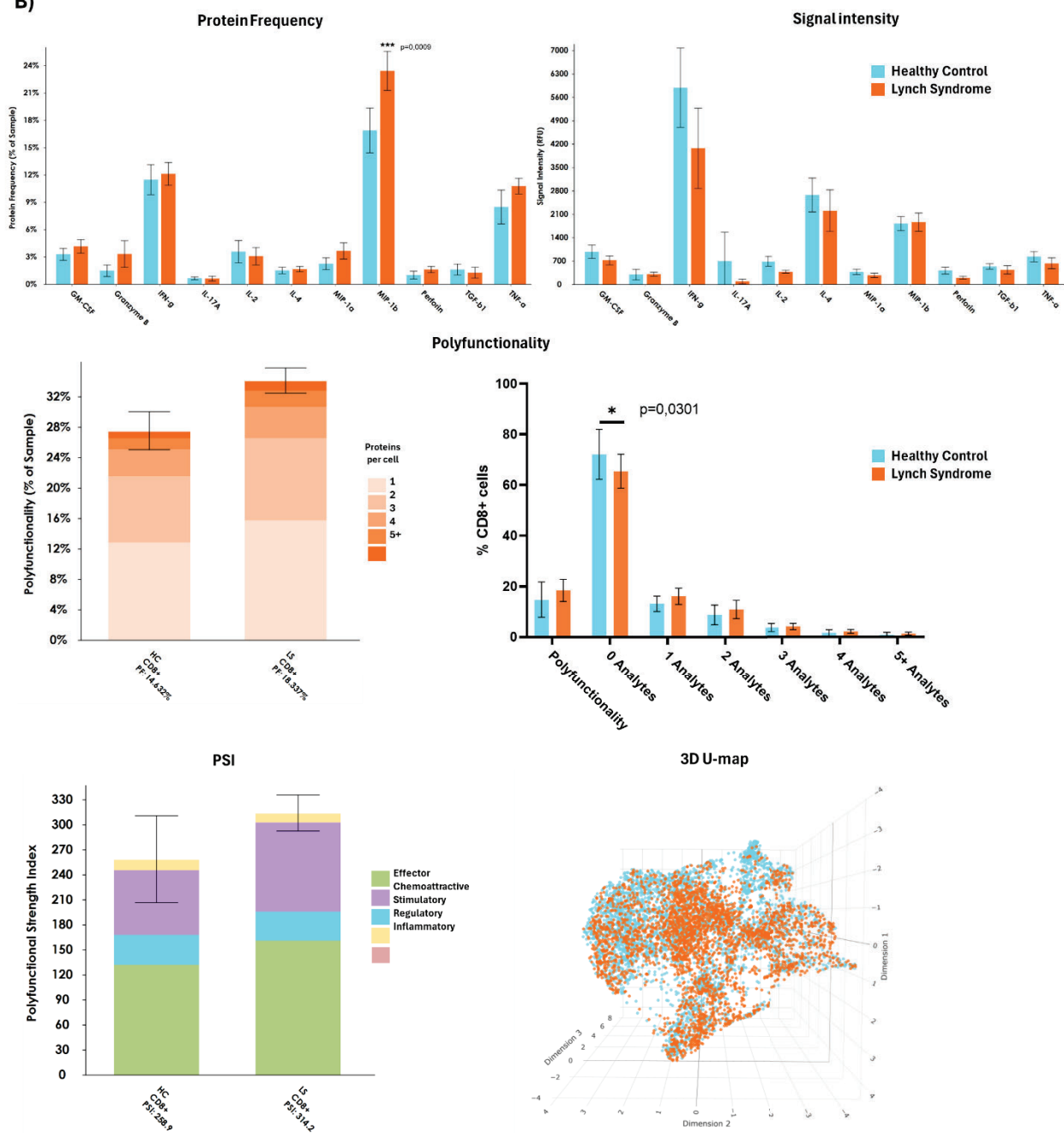
### 5.1.2.3 Single cell secretome analysis

Based on previous studies showing differences in cytokines from the immune microenvironment in LS compared to sporadic CRC patients<sup>255</sup>—where specific sets of differentially expressed cytokines were predictive of LS status—we aimed to apply this approach to our cohort to screen for biomarkers that may influence the local immune environment. To that end, we used IsoPlexis single cell multiplex cytokine profiling for CD3+ cells from PBMCs derived from seven LS individuals and eight healthy controls. Unfortunately, the cell tracer stain used to distinguish CD4+ cells failed in eight patients (specifically, four healthy controls and four LS carriers), limiting the CD4+ cell analysis to four healthy controls and three LS individuals. CD8+ T-cells were analysed for all individuals after confirming that background RFU from non-stained CD4+ cells did not influence CD8+ cytokine signals (Supplementary figure 2).

A)



B)



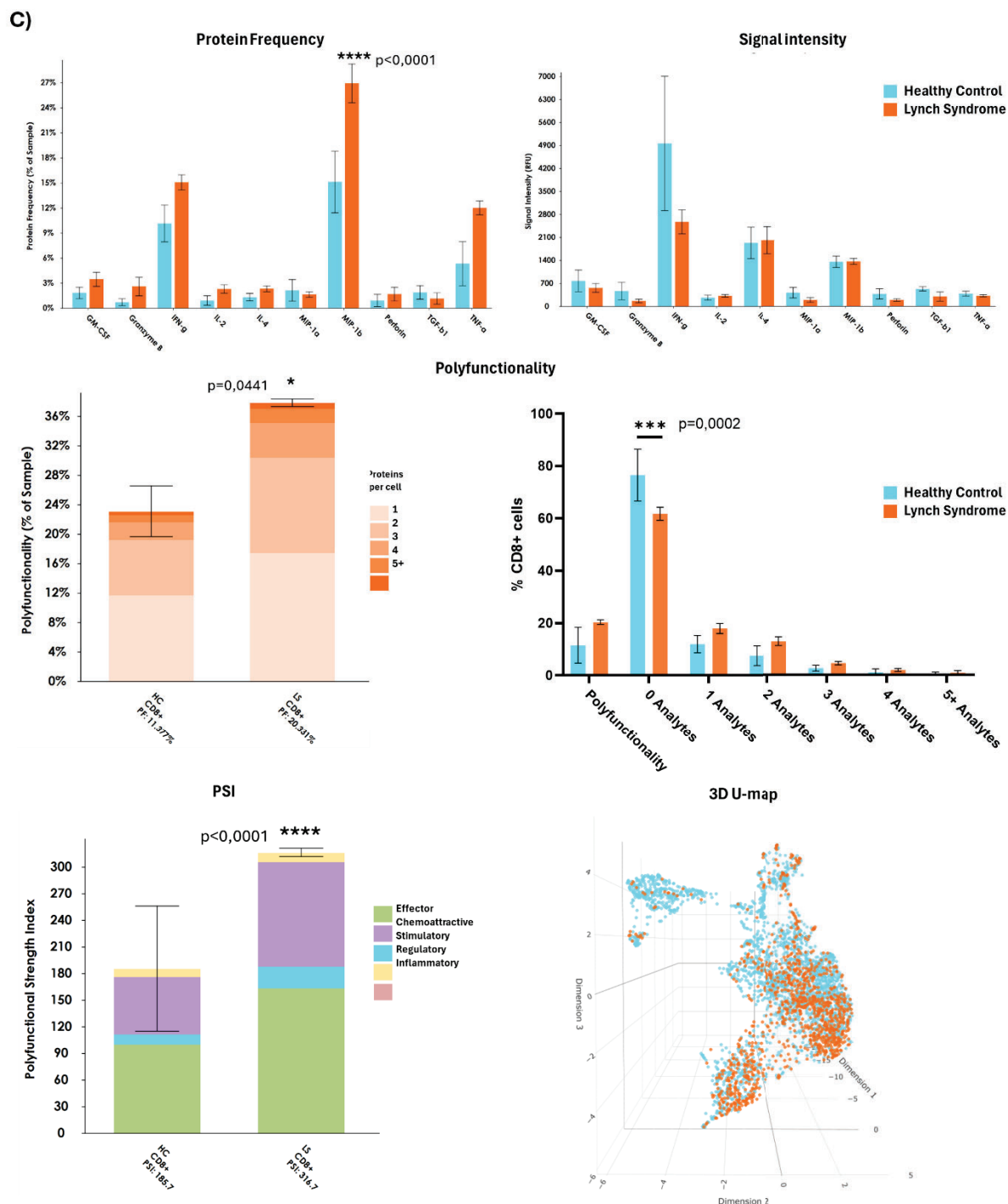


Figure 38. Single-cell multiplexed functional cytokine profiling of CD4 and CD8 T-cells from HCs and LS carriers. Plots show the average protein frequency (% from sample secreting a cytokine), average

*signal intensity, average polyfunctionality (defined as a cell co-secreting two or more cytokines, plots show 1 to 5+ analytes) as a comparative barplot and a stacked barplot, average PSI (Polyfunctionality Strength Index, which is generated by identifying the polyfunctional single cells and measuring the intensity of the cytokines they secrete) for the effector, chemoattractive, stimulatory, regulatory and inflammatory cytokine groups, and 3D U-maps showing clusters based on the greatest cytokine-based functional differences. A) CD4+ cells from LS carriers (orange) or HCs (blue), B) CD8+ cells from all available LS carriers (orange) or HCs (blue), and C) CD8+ cells from only LS carriers (orange) or HCs (blue) with available CD4+ and CD8+ data. Significance determined by 2-way ANOVA applying a Sidak's multiple comparisons test.*

No differences were observed in the polyfunctionality strength index (PSI), polyfunctionality, protein frequencies, or protein signals of CD4+ cells, likely due to the smaller cohort. However, a trend toward a less polyfunctional (defined as a cell co-secreting two or more cytokines) CD4+ profile and a lower IFN- $\gamma$  protein frequency and signal intensity in LS carriers was noted (Figure 38A).

In contrast, CD8+ T cells in LS carriers showed a tendency toward higher PSI (Polyfunctionality Strength Index, which is generated by identifying the polyfunctional single cells and measuring the intensity of the cytokines they secrete) and polyfunctionality, with significantly fewer cells producing zero analytes ( $p = 0.0301$ ). Additionally, LS carrier CD8+ cells exhibited a significantly higher frequency of MIP-1 $\beta$  production ( $p = 0.0009$ ), while intensity signals remained unchanged, suggesting a greater proportion of cytokine-producing cells rather than increased cytokine levels per cell. A trend toward higher TNF- $\alpha$  protein frequency in LS carriers was also observed, though it did not reach statistical significance (Figure 38B).

However, when analysing only patients with available CD4+ and CD8+ data, these trends became statistically significant. LS carriers demonstrated a higher PSI ( $p < 0.0001$ ) and greater polyfunctionality ( $p = 0.0441$ ), while maintaining a lower percentage of cells producing zero analytes ( $p = 0.0002$ ) and a higher MIP-1 $\beta$  protein frequency ( $p < 0.0001$ ). Additionally, the previously observed trend in TNF- $\alpha$  protein frequency became significant ( $p = 0.0274$ ) (Figure 38C).

3D U-maps dependent on donor groups did not show any notable differences based on the greatest cytokine-based functional differences, except for CD8+ cells from only patients with available CD4+ and CD8+ data, where a missing cluster on LS carriers can

be observed. A deeper analysis examining the different polyfunctional groups from each sample type can be found on Supplementary figure 3 and Supplementary figure 4.

### **5.1.3 FSDNs do not impair Dendritic Cell maturation capacity**

Before initiating DELAY-FSDN *in vitro* analyses, first we wanted to determine possible FSDN cytotoxicity to moDCs. To that end we generated and pulsed mature moDCs from 5 healthy donors with the 4 peptide pools defined in chapter I, phenotyped them and assessed their capacity to stimulate T-cell proliferation by an MLR, in comparison to non-loaded mature and immature moDCs from the same individuals.

All DCs had >85% viability, which did not decrease after loading them with the FSDN pools. HLA-I and HLA-DR were expressed by >85% in all conditions, CD80 expression was increased in all mature DCs (>70%) compared to iDCs (mean 53%(39.45-62.18)), CD83 followed a similar pattern with >92% expression on all mDCs conditions and a mean of 48%(36.75-58.75) in iDCs, and CD86 remained with a high expression (>84%) in all conditions, however a significant difference in the MFI is observed on all mDCs conditions compared to iDCs (Figure 39A, B).

With the MLR we confirmed that all pulsed mDCs maintained the capacity of stimulating allogenic T-cells, with a mean negative CFSE population (which indicates the rate of proliferation of T-cells) of 27.08%(IQR 11.15-44.63) for the peptide pools, which increased compared to iDCs with a mean of 16.88%(7.8-30.3) and was similar to the mature non-loaded control (29.32%(11.96-50.43) (Figure 39C).

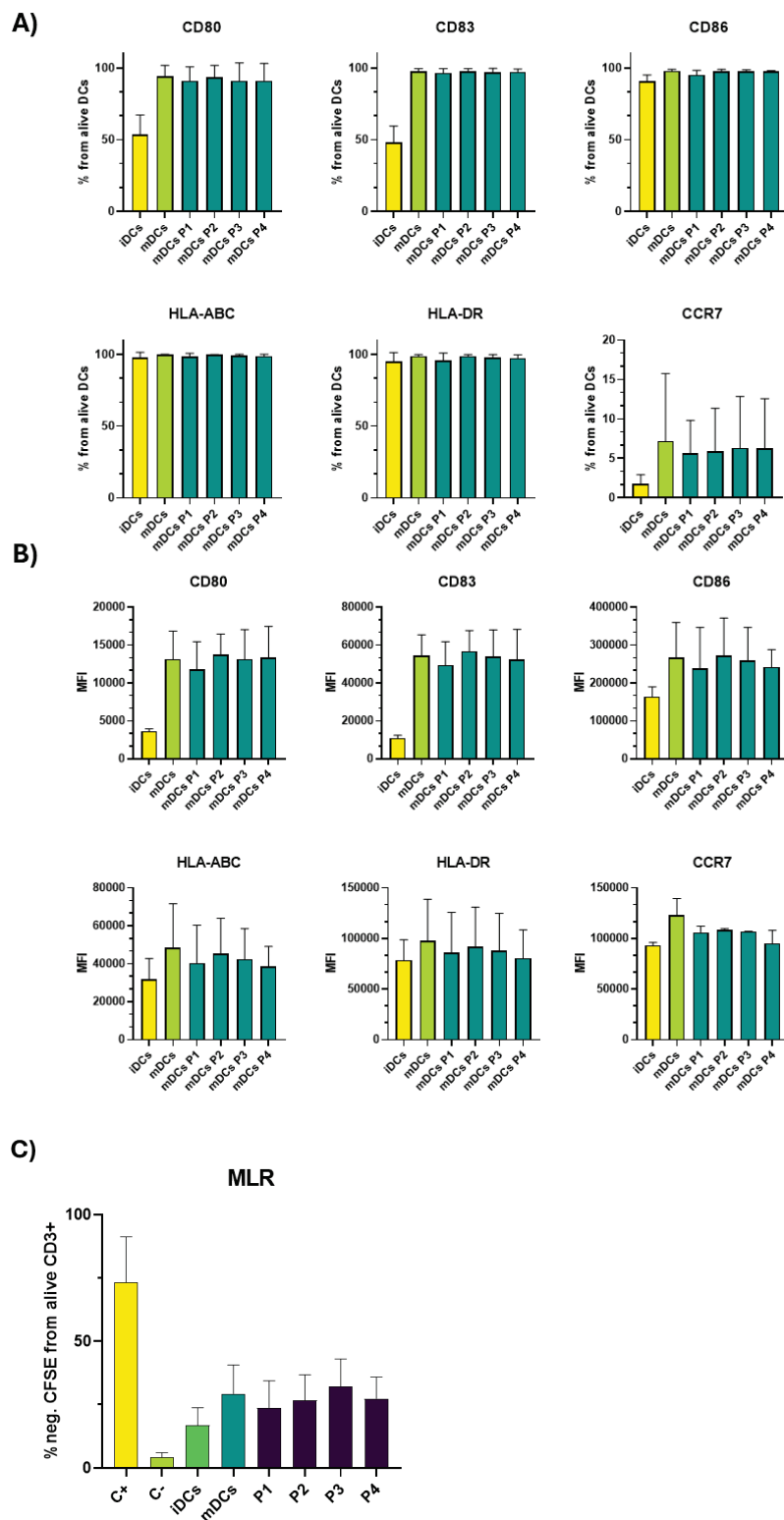


Figure 39. *moDCs* cultures phenotype by flow cytometry. N=5 mature *moDCs* were generated from healthy donors and compared immature, matured with pro-inflammatory stimuli or matured and loaded with the different peptide pools (1 to 4) conditions. A, B) Barplots with SD showing flow cytometry analysis of surface maturation markers for all conditions, represented as positive percentage from alive DCs (A) and mean fluorescence intensity (MFI) of analysed molecules (B). C) Barplot with SD showing results from a mixed lymphocyte reaction (MLR) with allogenic PBMCs stained with CFSE cell tracer. Results represented as loss of CFSE (Percentage of dim to negative population) from total alive CD3+ cells.

## 5.2 DELAY-FSDN-specific T-cells are present on normal mucosa and pre-cancers of LS carriers

Once we identified which peptides were optimal to induce an immunogenic response on LS patients and corroborated that they did not affect DCs viability or maturation capacity, we sought to detect and isolate FSDN-specific T-cells to test whether they are cytotoxic against target cells expressing those neoantigens. Moreover, these experiments also serve to further confirm the presence of memory T-cells against the prioritized neoantigens in LS carriers with or without cancer history, which would prove that neoantigens arise in pre-malignant cells on MSI foci of LS colorectal mucosa and are controlled by immune surveillance, generating memory against these peptides.

### 5.2.1 Detection of mucosa-infiltrating neoantigen specific cells by dextramers

As a first approach we searched for FSDN-specific cells in healthy colorectal mucosa and CrADs from 7 LS patients of known HLA haplotype by dextramer staining. The characteristics of these patients are summarized in Table 15. We had dextramers for 5 peptides and 3 different HLA-I alleles available: HLA-A\*02:01, HLA-A\*29:02 and HLA-B\*07:02. Patients that did not meet these haplotype criteria were excluded.

Table 15. *Characteristics of the included individuals in the assays for the detection of FSDN-specific T-cells by dextramer staining.*

| Characteristics                   | Total (n=7)  |
|-----------------------------------|--------------|
| <b>Sex, n (%)</b>                 |              |
| Female                            | 6 (85.7)     |
| Male                              | 1 (14.3)     |
| <b>Age, median (range), years</b> | 54 (47 - 79) |

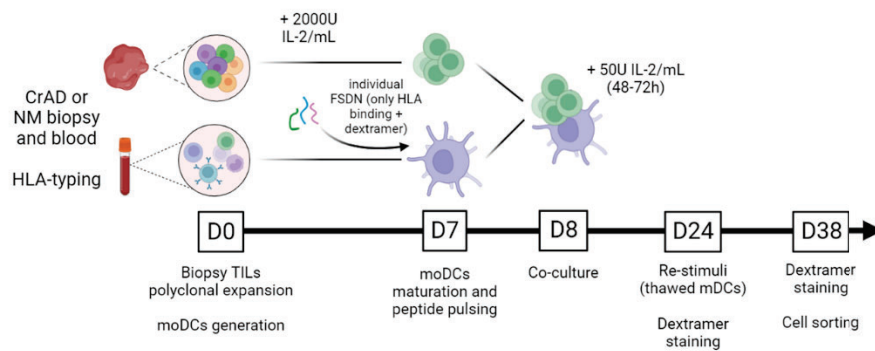
|                                                              |          |
|--------------------------------------------------------------|----------|
| <b>Germline mutation, n (%)</b>                              |          |
| <i>MLH1</i>                                                  | 4 (57.1) |
| <i>MSH2</i>                                                  | 2 (28.6) |
| <i>MSH6</i>                                                  | 1 (14.3) |
| <b>Previous cancer, n (%)</b>                                |          |
| No                                                           | 3 (42.9) |
| Yes                                                          | 4 (57.1) |
| <b>Types of previous cancer (among those with cancer), n</b> |          |
| Endometrial Cancer (EC)                                      | 2        |
| Colorectal Cancer (CCR)                                      | 1        |
| Glioblastoma (GB)                                            | 1        |
| <b>Sample type, n (%)</b>                                    |          |
| NM                                                           | 4 (57.1) |
| NM and CrAD                                                  | 3 (42.9) |

From the obtained biopsy CSs, we expanded infiltrating CD8+ T lymphocytes by co-culturing them with autologous mature DCs after an unspecific pre-expansion of 1 week (which is the necessary time to generate autologous moDCs), pulsed with 5 of our neoantigens which had high affinity to HLA-A\*02:01, HLA-A\*29:02 and HLA-B\*07:02 and with dextramers available. After 2 weeks of culture, we stained cells with CD8+-APC and the different dextramers labelled with PE fluorochrome searching for double positive T-cells.

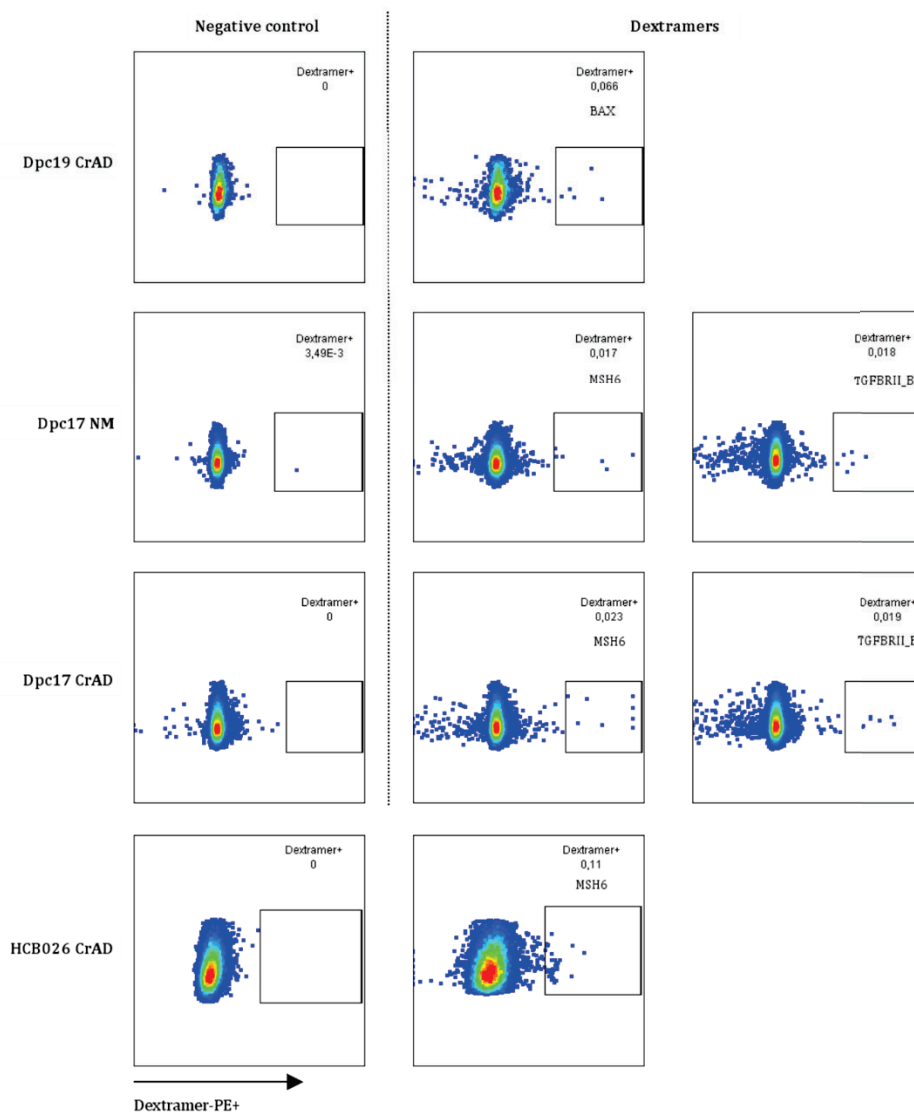
Figure 40 displays cytometry plots for 4 different samples from 3 patients which had specific CD8+ T-cells compared to the negative control (an HLA-A\*03:01 multimer presenting an irrelevant peptide and conjugated with PE). While the significant detection in positive cells compared to the negative control could be attributed to the dispersion caused by the staining, the negative control of an irrelevant multimer should also cause this dispersion. However, further analyses should be carried out to confirm positivity.

Despite these results, due to the low frequency of specific cells after the pre-expansion (0.01 to 0.07% from total alive CD8+ cells), we were not able to enrich or sort them by FACS sorting.

**A)**



**B)**



*Figure 40. Flow cytometry detection of DELAY-FSDN-specific T-cells by dextramer staining. A) Schematic representation of the protocol used for the detection of FSDN-specific T-cells by dextramer staining. LS mucosa and CrAD-derived T-cells were pre-expanded with autologous peptide-loaded moDCs for two weeks, then stained with the corresponding dextramers-PE accounting for each patient's HLA haplotype. B) FACS dot plots showing the percentage of positive multimer cells gated from total alive CD8+ cells for n=3 patients and for different peptides (right) compared to the negative control (left), which consists of a PE-labelled HLA-A\*03:01 multimer presenting an irrelevant peptide.*

### 5.2.2 Detection of neoantigen-specific T-cells by IFN- $\gamma$ production

Taking into account the low number of specific cells that we found on the dextramer assays and the inability to sort them, we decided to opt for a PBMC-based T-cell expansion where mucosa-derived T-cells were put into contact with autologous PBMCs (irradiated at 40Gy for 25 minutes and non-irradiated) and a cocktail with the 18 HLA-I peptides included in the vaccine since day 0. Thus, the principal advantages of this approach are that TILs are not pre-expanded non-specifically but rather selected by specificity pressure since day 0. Unlike when we used dextramers to detect specificity, we were not restricted by the patient's HLA alleles, which increased the likelihood of identifying neoantigen-responding cells. These results have been published in<sup>246</sup>. The characteristics of included patients in this study are summarized in the following table:

| Characteristics                                              | Total (n=13) |
|--------------------------------------------------------------|--------------|
| <b>Sex, n (%)</b>                                            |              |
| Female                                                       | 6 (46.2)     |
| Male                                                         | 7 (53.8)     |
| <b>Age, median (range), years</b>                            | 59 (32 - 76) |
| <b>Germline mutation, n (%)</b>                              |              |
| MLH1                                                         | 5 (38.4)     |
| MSH2                                                         | 4 (30.7)     |
| MSH6                                                         | 3 (23.1)     |
| PMS2                                                         | 1 (7.7)      |
| <b>Previous cancer, n (%)</b>                                |              |
| No                                                           | 4 (30.8)     |
| Yes                                                          | 9 (69.2)     |
| <b>Types of previous cancer (among those with cancer), n</b> |              |

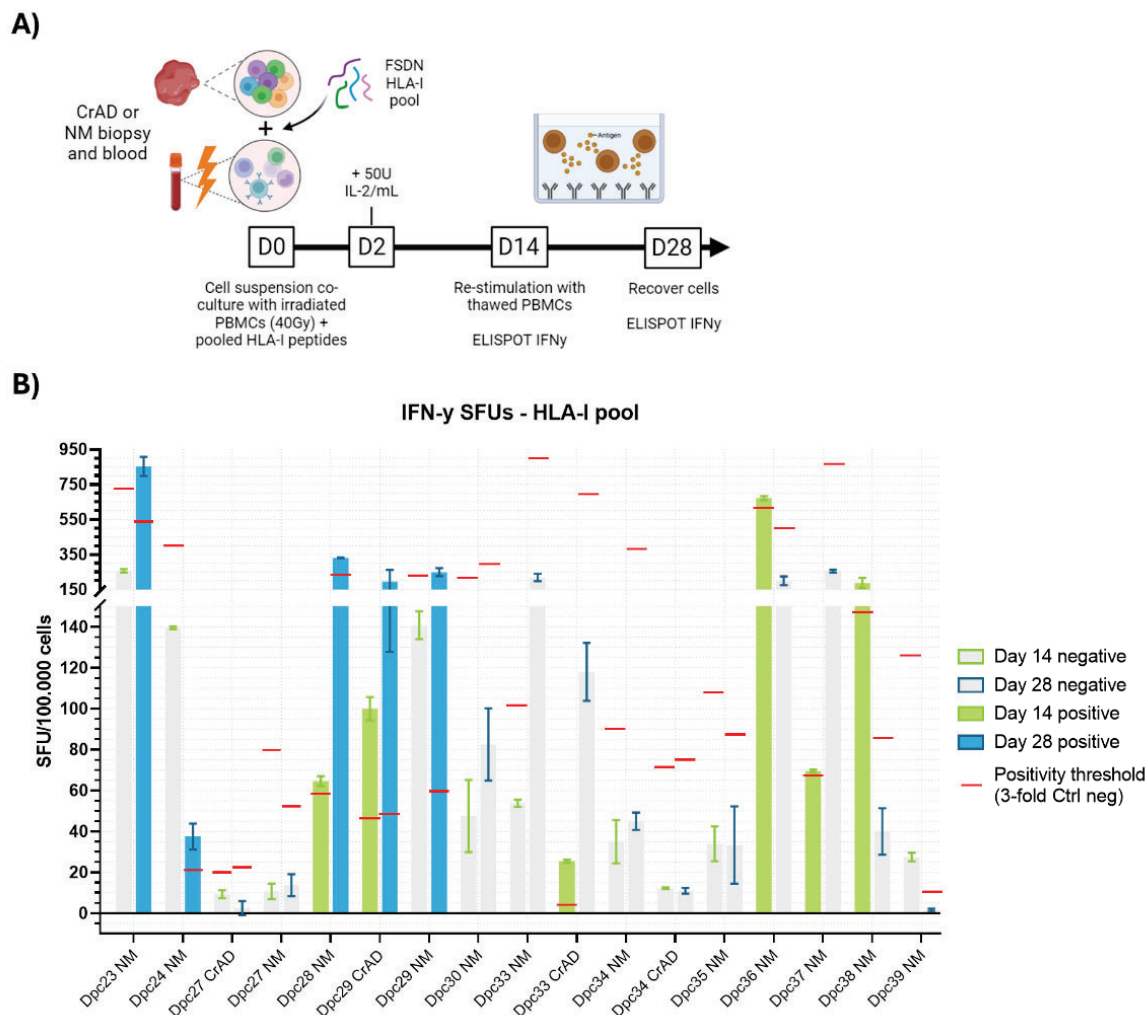
|                                        |          |
|----------------------------------------|----------|
| Endometrial Cancer (EC)                | 2        |
| Colorectal Cancer (CCR)                | 5        |
| Intrahepatic Cholangiocarcinoma (IHCC) | 1        |
| Testicular Cancer (TC)                 | 1        |
| Synovial Sarcoma (SA)                  | 1        |
| Breast Cancer (BC)                     | 1        |
| <b>Sample type, n (%)</b>              |          |
| NM                                     | 9 (69.2) |
| NM and CrAD                            | 4 (30.8) |

Out of 17 samples (4 CrAD and 13 NM samples) from 13 different patients with LS, we detected T-cells reactive to an HLA-I neoantigen pool (defined by a positive IFN- $\gamma$  ELISPOT) in 9 samples (53%) from 8 different patients (62%). Peptide-specific T-cells either increased in number after second stimulation with autologous PBMCs (patients Dpc23, Dpc24, Dpc28, Dpc29) (Figure 41) or lost specificity due to non-specific cell overgrowth (patients Dpc33, Dpc36, Dpc37, Dpc38).

Out of all eight responding individuals, three of them had no previous cancer history and cells were derived from healthy mucosa (Dpc36, Dpc37) or pre-cancerous lesions (Dpc33), which strengthens the immune surveillance hypothesis. On the other hand, from responding patients with previous cancer history, only two of them had CRC, while the rest of manifestations were extracolonic.

The T-cells from these patients were also characterized by flow cytometry to define changes in CD4-CD8 populations, in the T-cell activation markers CD137 (to define activated CD8 cells) and CD154 (to define activated CD4 cells) combined with CD69.

Comparing the pre-expanded samples with or without HLA-I peptide specific cells detected after expansion (Figure 42; Responder, R, and Non-Responder, NR), we observed that a higher proportion of CD8+ cells seems to correlate to R samples, although without statistical significance ( $p=0.07$ ), probably due to a small cohort size.



*Figure 41. IFN- $\gamma$  ELISPOTs for the HLA-I peptide pool. A) Schematic representation of the protocol used for the analysis. B) Barplots showing SFUs per test, an  $n=2$  technical duplicates with the SD are plotted, values are normalized for 100.000 cells. Two ELISPOT tests per patient are plotted for the first (Day 14, green) and second (Day 28, blue) stimulation with PBMCs and  $5\mu\text{M}$  of the HLA-I peptide pool. Positive tests, with values above their threshold, are colour-filled, and negative tests bars remain grey. The positivity threshold for each ELISPOT, determined as 3-fold the negative control with an irrelevant peptide, is marked as a red line.*

All samples increased their CD8<sup>+</sup> cell proportion when exposed to the HLA-I neoantigen pool, from day 14 to day 28 ( $p=0.0431$ ). When looking at activation markers, we find that only R samples gradually increase the CD8<sup>+</sup>CD137<sup>+</sup> cell percentage after expansion

with the HLA-I peptide pool (measured after an HLA-I peptide pool ELISPOT at day 14 and 28), which is significantly higher at the end of the expansion (day 28,  $p=0.037$ ) in comparison to NR samples. This is specific only to CD8+ cells, as CD4+CD154+ cells do not increase after exposure to the peptide pool.

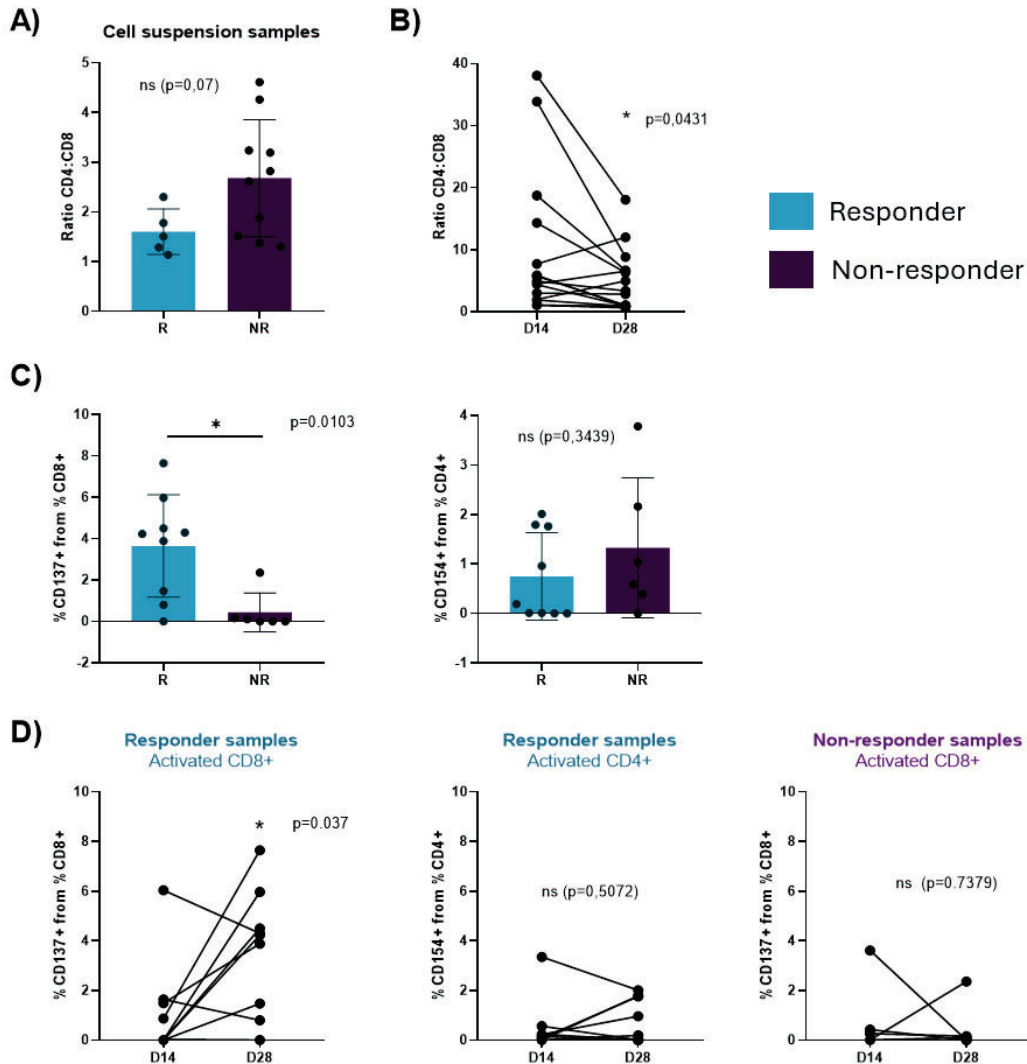
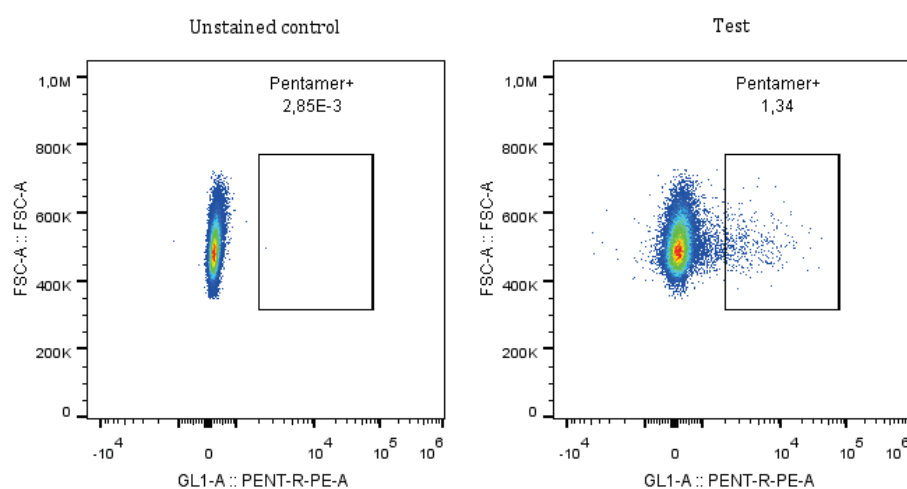


Figure 42. Flow cytometry analysis of LS cells expanded with the HLA-I vaccine pool. A) Comparison of CD4:CD8 ratio in Responder (R, blue) and Non-Responder (NR, purple) initial cell suspension samples, before cell expansion, from parent alive CD3+ cells. Barplot shows  $n=14$  independent experiments. B) Evolution of CD8+ cells percentage from expanded samples with the vaccine HLA-I

pool, at day 14 and 28, from alive CD3+ parent cells. Plot shows n=12 independent experiments. C) CD8+CD137+ (left) and CD4+CD154+ (right) cells percentage in R and NR samples after a 28-day expansion with the vaccine HLA-I pool. Plots show n=12 independent experiments. D) Evolution of CD8+CD137+ and CD4+CD154+ cells percentage from expanded R (two left, respectively) and CD8+CD137+ in NR (right) samples with the vaccine HLA-I pool, at day 14 and 28, from alive CD3+ parent cells. Plots show n=7 and n=5 independent experiments for R and NR samples, respectively. Normality and lognormality were assessed for all analysis and statistical significance was inferred with a paired or unpaired T-test.

We did not find any significant differences in the number of responding patients or the increase of activation markers when comparing individuals with or without history of cancer.

For patient Dpc24, we thawed T cells that had been frozen after the second IFN- $\gamma$  ELISPOT to confirm whether these specific cells were responding to peptides for which we have corresponding pentamers (ProImmune, UK) available for cytometry staining. Knowing the patient's HLA haplotype, we stained cells for the multimer constructs from peptides that are predicted to bind to their HLA-I alleles, which were HLA-A\*02:01-MSH3 and HLA-A\*02:01-TGFBRII\_B. Due to the low number of cells, we pooled the staining in one single test and added an unstained control. We found that Dpc24 had positivity to the tested multimers (Figure 43), although at low levels, which is consistent with the ELISPOT from the same patient.



*Figure 43. Flow cytometry analysis of HLA-I DELAY-FSDN specific T-cells using pentamers conjugated to PE for the peptides HLA-A\*02:01-MSH3 and HLA-A\*02:01-TGFBR1I\_B. Dot plots show cells gated from total alive CD8+ cells.*

Most importantly, with these results we confirm that the prioritized peptides are recognized by specific T-cells from LS carriers both with or without history of cancer, which indicates that our FSDN are presented in intestinal MSI foci and can be recognized as pathogenic by resident T-cells.

### **5.3 Creating a DELAY-FSDN-expressing target cell line**

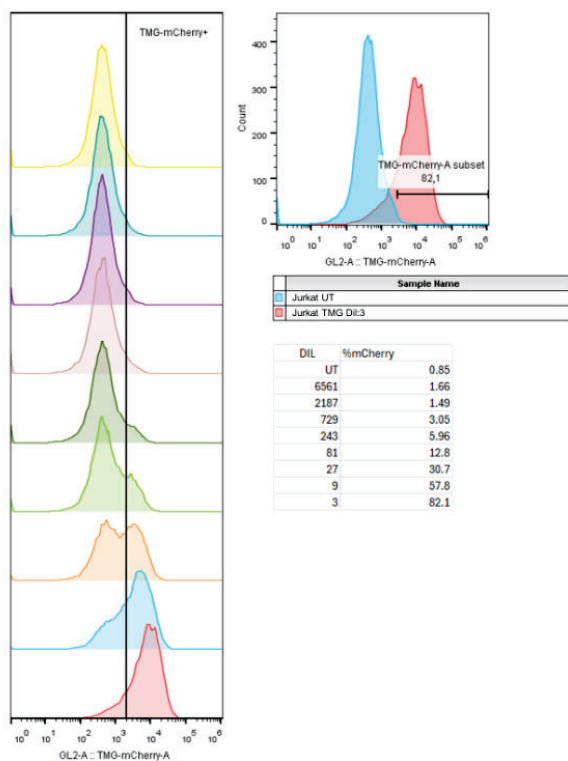
The GLE1123 pTRPE Ub-Lynch syndrome TMG-T2A-mCherry plasmid was packed on a VSV G glycoprotein-pseudotyped lentiviral vector and tittered with a JURKAT cell line. Limiting dilution titer was  $4.15 \times 10^8$  TU/mL (dilution 1:81) which gave 12.8% of positive mCherry+ cells (Figure 44A).

Titer was established for 12,06  $\mu$ L/million cells for a MOI 5. Viral particles at this concentration were then used to infect two distinct cell lines: K562 and K815. K815 cells are a modified version of K562, engineered to express the plasmid GFP-T2A-b2m-HLA-A\*02:01 SCD. Following infection, flow cytometric analysis at 72 hours post-transduction showed mCherry positivity in 25.6% of K562 cells and 36.1% of K815 cells, indicating successful transduction (Figure 44B).

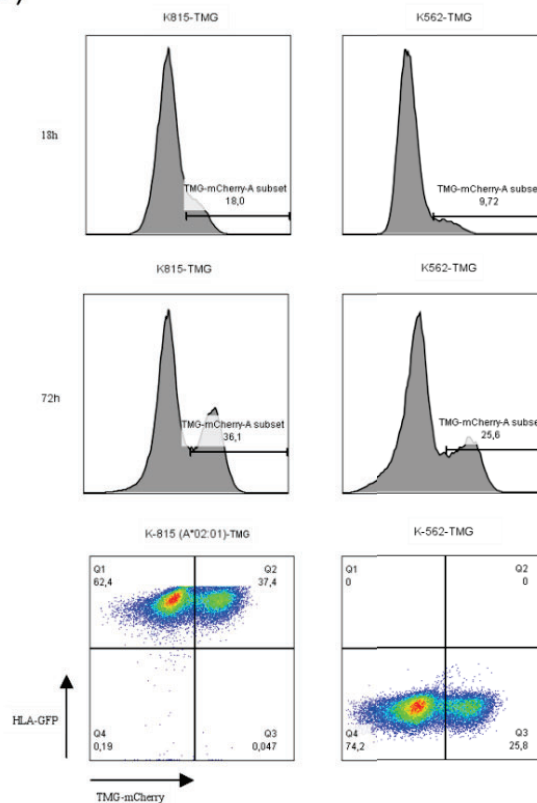
To establish purified cell lines with stable plasmid expression, infected cells were sorted using FACS. For the K-562-TMG line, cells were gated based on mCherry fluorescence, while for K-815-TMG, dual fluorescence for both mCherry and GFP was used as sorting criteria. The sorting process resulted in highly enriched populations with >98% purity for both cell lines, as confirmed by post-sorting analysis (Figure 44C).

These cell lines provide a target model for downstream applications such as the testing of specific cytotoxicity of DELAY-FSDN-specific T-cells, which can be applied as a routine potency test on vaccinated and responding individuals on the DELAY phase Ib clinical trial. However, verifying proper FSDN peptide presentation on the CL membranes should be performed prior to functional assays to avoid false-negative or false-positive results.

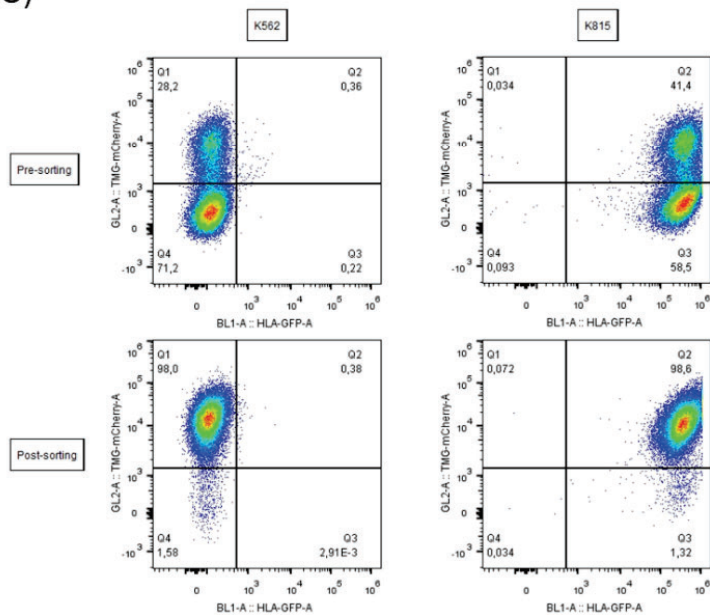
A)



B)



C)



*Figure 44. Flow cytometry plots for TMG-transduced Jurkat and K-562 cell lines. A) Histogram plots for the lentiviral titer analysis on Jurkat cells. Left histogram overlay shows all titering dilutions, indicated in the right table from the highest (1:6561) to the lowest (1:3). Left histogram shows the lowest dilution overlayed to the untransduced cells. B, C) TMG-transduced K-560 (HLA-null) and K-815 (HLA-A\*02:01-GFP positive) cell lines. Transduction efficiency after 18h and 72 hours (B), and cell sorting results for TMG+ K-562 or K-815 cell lines (C) (pre- and post-sorting cytometry dot plots).*

**CHAPTER 3:**  
*THE DELAY VACCINE - A PHASE  
IB FIRST-IN-HUMAN  
CLINICAL TRIAL*





## 6 CHAPTER III: A PHASE Ib FIRST-IN-HUMAN CLINICAL TRIAL

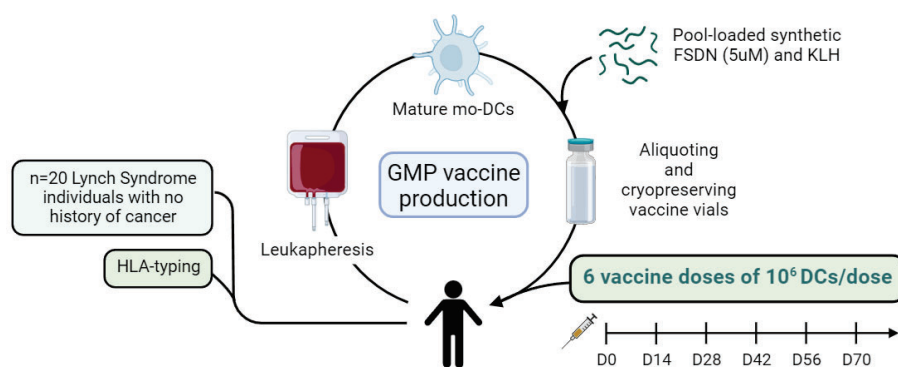
The phase Ib clinical trial was funded by an independent clinical investigation grant (ICI) from the Instituto de Salud Carlos III and the science and innovation ministry (Spain), under the application number ICI22/00063. Documentation necessary for the approval of the Spanish Medical Agency (AEMPs) has been sent and is under review.

Product validations under GMP conditions according to the regulatory agencies (AEMPS) have shown that the final product complies release criteria established by these products.

### 6.1 Validation results

#### 6.1.1 Product generation and characterization

The product was produced according to the methodology in grade A and B cleanrooms at the Advanced Therapies Unit, Immunotherapy Section, Hospital Clinic of Barcelona. The initial manufacturing process used leukocytic residue. The subsequent two manufacturing processes focused on obtaining a viable final product, and for that purpose two LS carriers were recruited.



The final product is defined as a cell suspension containing 10 million (+/- 2 million) autologous DCs loaded with FSDN, with mature phenotypic characteristics (CD80+, CD83+, CD86+, HLA Class II+), with a purity of over 85% from all cells (CD3

(lymphocytes) and CD14 (monocytes) < 15%) and contained in sterile cryotubes. Each vial should contain the mentioned product resuspended in 0.4 ml of human serum albumin solution (GRIFOLS®) with a 10% DMSO used as a cell cryoprotectant.

#### 6.1.1.1 Product specification compliance

All processes and final products from the three validations complied with the specified requirements. The results are summarized in the following table:

| Day | Phase                                                                   | Control                                                               | Acceptance Criteria                          | DC-DELAY-VAL-02 | DC-DELAY-VAL-03                               | DC-DELAY-VAL-04                               |
|-----|-------------------------------------------------------------------------|-----------------------------------------------------------------------|----------------------------------------------|-----------------|-----------------------------------------------|-----------------------------------------------|
|     |                                                                         |                                                                       |                                              | Healthy donor   | Lynch Syndrome carrier                        | Lynch Syndrome carrier                        |
| 1   | Leukapheresis                                                           | Serologies (detection of viral and bacterial pathogens)               | Negative                                     | Complies        | Complies                                      | Complies                                      |
|     | Blood processing, isolation of PBMCs, adherence, and monocyte isolation | Clean environment/surface, control for the presence of foreign bodies | Negative                                     | Complies        | Complies                                      | Complies                                      |
|     |                                                                         | Non-viable particles in the environment                               | Negative                                     | Complies        | Complies                                      | Complies                                      |
|     |                                                                         | Microbiology (product)                                                | Negative                                     | Complies        | Complies                                      | Complies                                      |
|     |                                                                         | Visual (supernatant)                                                  | Clean                                        | Complies        | Complies                                      | Complies                                      |
|     |                                                                         | Viability                                                             | >80%                                         | Complies - 95%  | Complies - 95%                                | Complies - 95%                                |
|     |                                                                         | Cell count                                                            | >1 x 10 <sup>9</sup> total mononuclear cells | N/A             | 3.8 x 10 <sup>9</sup> total mononuclear cells | 1.8 x 10 <sup>9</sup> total mononuclear cells |
| 3   | Obtaining semi-differentiated dendritic cells,                          | Microbiological environment/surface                                   | Negative                                     | Complies        | Complies                                      | Complies                                      |
|     |                                                                         | Non-viable particles in the environment                               | Negative                                     | Complies        | Complies                                      | Complies                                      |

|   |                                                                               |                                         |                                              |                |                |                |
|---|-------------------------------------------------------------------------------|-----------------------------------------|----------------------------------------------|----------------|----------------|----------------|
|   | counting, and seeding cells into corresponding neoantigen groups (A, B, C, D) | Microbiology (product)                  | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Visual (supernatant)                    | Clean                                        | Complies       | Complies       | Complies       |
|   |                                                                               | Viability                               | >80%                                         | Complies - 92% | Complies - 86% | Complies - 93% |
|   |                                                                               | Cell count                              | >1 x 10 <sup>8</sup> total mononuclear cells | N/A            | Complies       | Complies       |
| 7 | Maturation of DCs, pulsing with corresponding peptide pools                   | Microbiological environment/surface     | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Non-viable particles in the environment | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Visual (culture flasks)                 | No contamination                             | Complies       | Complies       | Complies       |
| 8 | Final product preparation, dosing, and storage                                | Microbiological environment/surface     | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Non-viable particles in the environment | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Visual (culture flasks)                 | No contamination                             | Complies       | Complies       | Complies       |
|   |                                                                               | Final cell count                        | Sufficient for 6 doses: 60x10 <sup>6</sup>   | N/A            | Complies       | Complies       |
|   |                                                                               | Sterility test                          | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Mycoplasma test                         | Negative                                     | Complies       | Complies       | Complies       |
|   |                                                                               | Viability                               | >80%                                         | Complies       | Complies       | Complies       |
|   |                                                                               | Phenotype                               | >85% DCs                                     | Complies       | Complies       | Complies       |

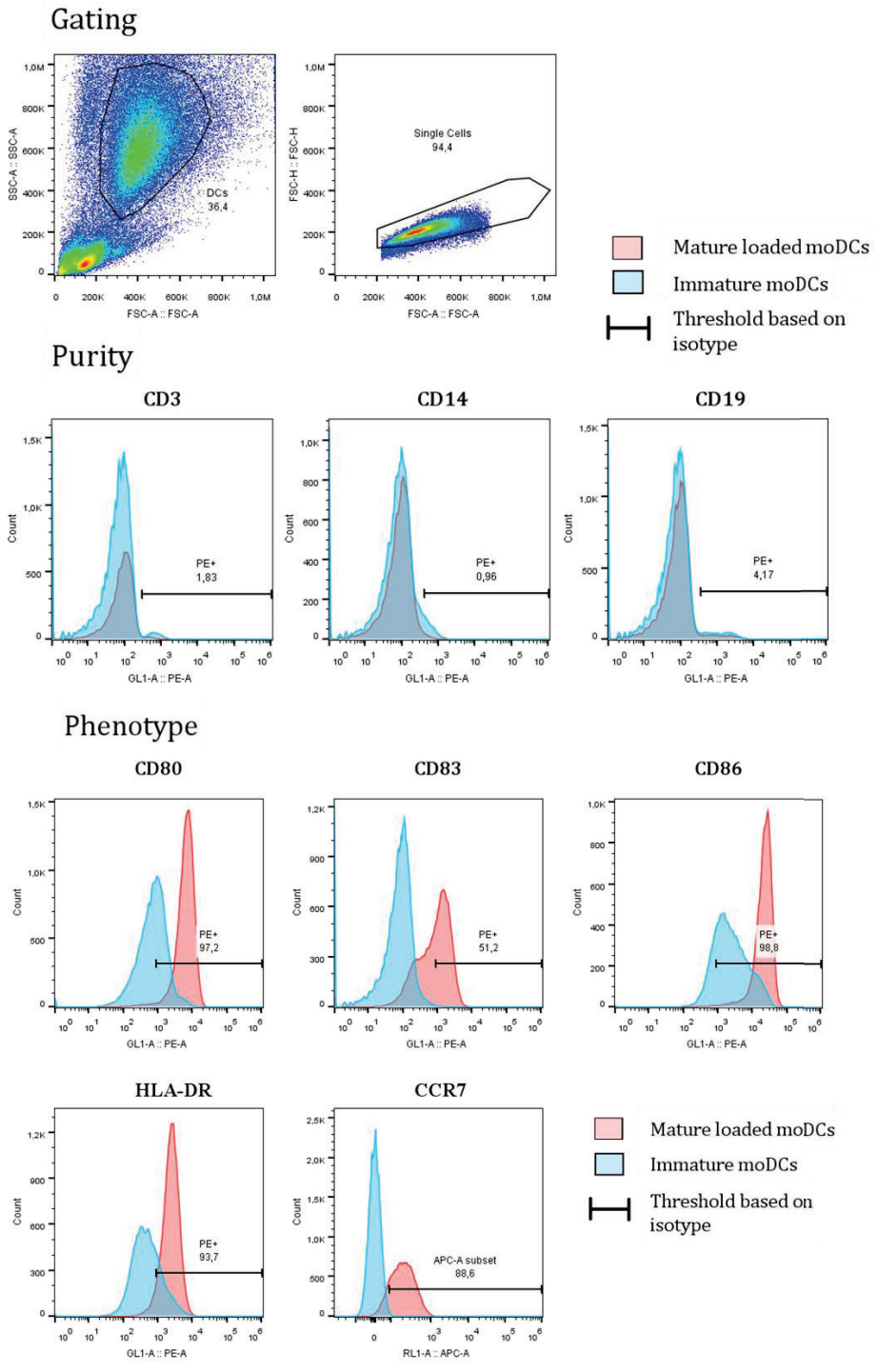
#### 6.1.1.2 Product phenotype and potency

We analysed the final products phenotype as mature immunogenic moDCs by measuring expression of the surface activation markers CD80, CD83 and CD86, the homing marker CCR7 and the percentage of expression of HLA-II molecules (HLA-DR). We also assessed product purity by measuring the proportion of other cell types: CD3 for T-cells, CD19 for B-cells and CD14 for residual monocytes.

All products complied with the established thresholds and increased activation and homing markers upon maturation compared to an immature control (see DC-DELAY-VAL-04 example on Figure 45). All products were highly pure on moDCs, without significant contaminations of other cell types.

We also performed an MLR for all products, confirming that all pulsed mDCs maintained the capacity of stimulating allogenic T-cells. As a representative example, in DC-DELAY-VAL-04 we see a negative CFSE population of >38% on all mature conditions (mean(Range) of 41.3%(38.4-41.9)) compared to iDCs (12.5%(9.1-15.1)) (Figure 45).

A)



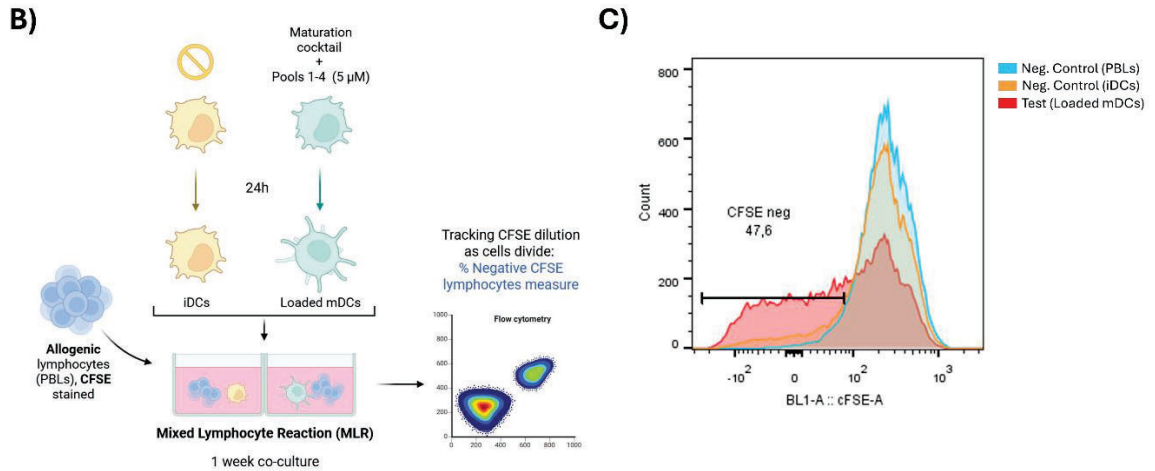


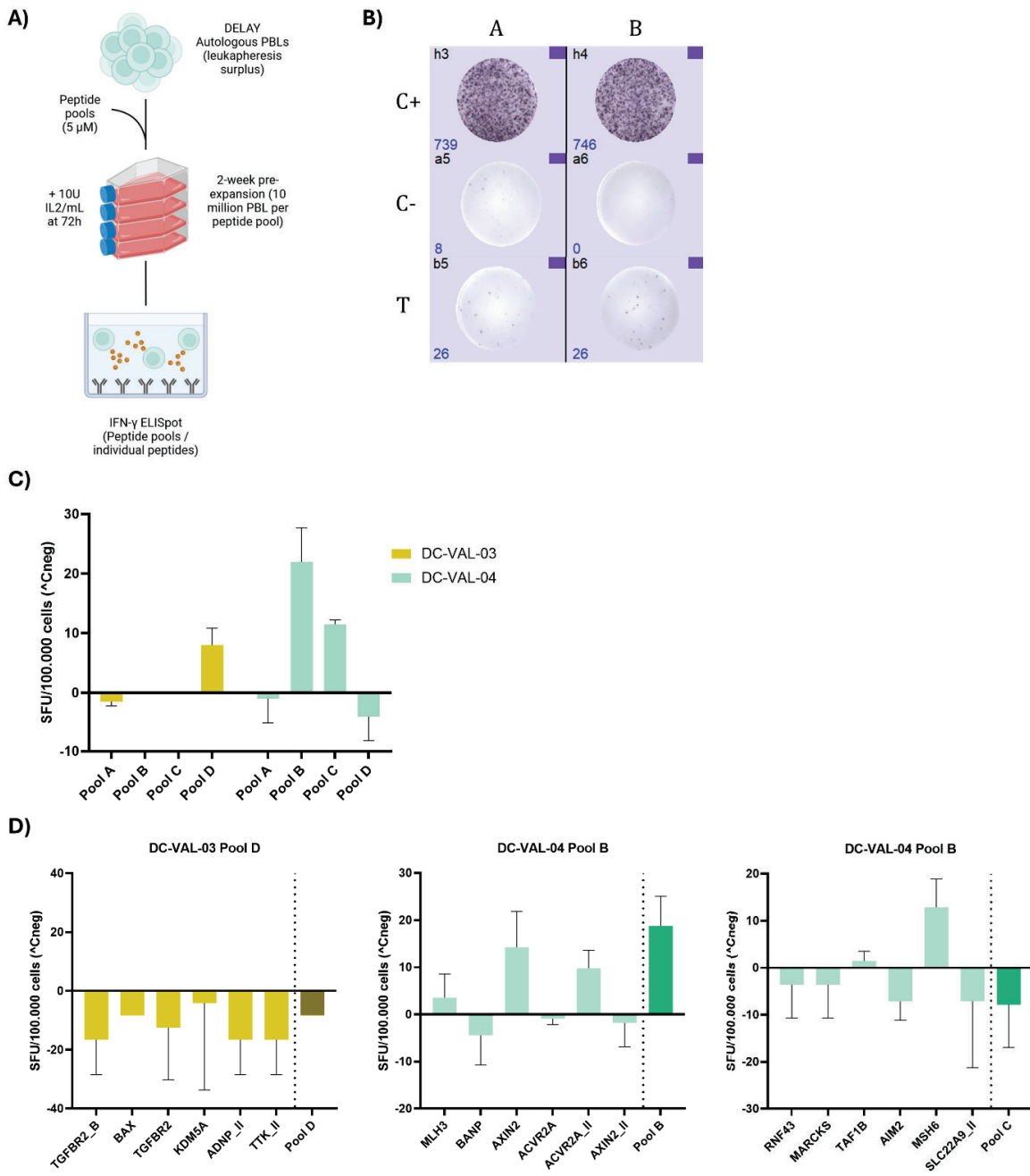
Figure 45. Cytometry analysis of purity, maturation markers and potency in the DC-DELAY-VAL-04 validation. A) Gating strategy, purity and maturation markers of the DC-DELAY-VAL-04 final product (in red, Mature loaded moDCs) compared to an immature control (in blue). Percentages of PE-positive DCs from alive DCs from each marker, threshold established from an isotype control. B) Schematic representation of the mixed lymphocyte reaction (MLR) protocol used. C) Results from a MLR with allogeneic PBLs stained with CFSE cell tracer. Results represented as loss of CFSE (Percentage of dim to negative population) from total alive CD3+ cells from control PBLs (blue), immature DCs (orange) or the DC-DELAY-VAL-04 final product (red).

### 6.1.2 Immunomonitoring – basal vaccine immunogenicity assessment

To evaluate the baseline immune response to vaccine peptides in PBLs from the two recruited LS individuals, we pre-expanded peptide-specific T cells. For each individual, 40 million PBLs were cultured for two weeks in the presence of peptide pools (5 µM per pool), with 10 million PBLs allocated per peptide condition. Following expansion, antigen-specific reactivity was assessed using an IFN-γ ELISPOT assay targeting the same peptide pool used during pre-expansion. SFUs were normalized to 100,000 cells, with background subtraction based on the DMSO negative control. Responses were considered positive if SFU counts were at least three times higher than those of the negative control. (Figure 46A, B).

Preliminary results suggested the presence of peptide-specific T cells in three conditions: DC-DELAY-VAL-03 pool D, and DC-DELAY-VAL-04 pools B and C, where test wells had 3-fold SFUs compared to the negative control (Figure 46C).

To further confirm these observations and identify specific peptides being recognized, we maintained the cultures for an additional week and performed a second IFN- $\gamma$  ELISPOT assay. This time, individual peptides from each pool were tested alongside a control with the full pool.



*Figure 46. IFN- $\gamma$  ELISPOTs for the DC-DELAY-VAL-03 and DC-DELAY-VAL-04 cells. A) Schematic representation of the protocol used. B) Example of the ELISPOT results, duplicates (termed A and B) are shown for the positive control (C+), negative control (C-) and test (T) for one of the positive test samples (DC-VAL-04 pool B). C, D) IFN- $\gamma$  ELISPOT results for the peptide pools (C) and for the individual peptides (D). An n=2 technical duplicates with the SD are plotted, values are normalized for 100.000 cells and the negative control (wells with T-cells and DMSO) has been subtracted from individual values.*

In patient DC-DELAY-VAL-04 pool B, we confirmed the presence of peptide-specific T cells recognizing AXIN2 (HLA-I restricted) and ACVR2A\_II (HLA-II restricted). For DC-DELAY-VAL-04 pool C, a positive ELISPOT response was observed for peptide *MSH6*, but no positivity was detected in the full pool control, indicating a likely false positive. No positive ELISPOT responses were detected for DC-DELAY-VAL-03 pool D in the follow-up assay (Figure 46D).

From these analyses, we can conclude that patient DC-DELAY-VAL-04 had a pre-existing immunity to the peptides AXIN2, a tumour suppressor in the Wnt pathway (restricted to HLA-I) and ACVR2A\_II, a tumour suppressor from the TGF- $\beta$  pathway that inhibits the growth and differentiation of cells (restricted to HLA-II), which could be detected in their peripheral blood. Both of these genes are considered to be tumour drivers, which may imply that this patient had a pre-malignancy which got recognized and eliminated by immune surveillance.

Further evaluation of diversification and maintenance of already existing FSDN-specific T-cells will be made during monitorization of the patients after vaccination.

## 7 DISCUSSION AND CONCLUSIONS

### 7.1 Discussion

#### 7.1.1 Chapter I

In this thesis, we identified LS-associated FSDNs using a custom filtering strategy. We validated the presence of FSDN-associated mutations across multiple datasets, including cancer and CrAD samples from LS carriers and sporadic MMRd patients. Our findings demonstrated that many of these mutations are transcribed into RNA in tumour cell lines and escape nonsense-mediated decay. Importantly, we showed that individuals with LS mount specific immune responses against these FSDN, as evidenced by IFN- $\gamma$  ELISPOT assays performed on infiltrating lymphocytes from CrAD and CRC samples. This sequential approach has effectively identified a set of immunogenic peptides that would be valid for a cancer prevention vaccine<sup>246</sup>.

This study adds new data to the growing evidence on the importance of selecting neoantigens for LS prevention considering their cancer specificity, immunogenicity, and safety<sup>57,256–259</sup>. We prioritized FSDNs recognized by common HLA alleles in the European population to enhance the immune response across diverse individuals. By including both HLA Class I and II epitopes, we ensured a comprehensive immune response involving both CD8+ and CD4+ T-cells, crucial for effective tumour prevention and control<sup>198</sup>. We further refined FSDN selection by filtering for self-similarity and pathogen-derived peptide similarity, enhancing cross-reactivity<sup>56,80,260</sup> (22.4% of potential FSDNs showed similarity to pathogen peptides) while minimizing the risk of autoimmunity by selecting peptides dissimilar to self, which is also associated with higher immunogenicity<sup>261</sup>. Targeting neoantigens from driver mutations reduces the risk of immune evasion by focusing on clinically significant tumour cell populations, reducing the risk of immune evasion through antigen loss<sup>262</sup>. Our analysis across datasets demonstrated that MSI-H tumours unrelated to LS patients also harbour pFSDN-associated mutations, suggesting that a significant proportion of these mutations play a crucial role in tumour initiation and maintenance. Enhancing immunogenic responses against peptides derived from these mutations could provide therapeutic and

preventive benefits not only for LS patients but also for individuals with MSI-H tumours.<sup>263,264</sup>

Our study also considered mid- and low-affinity binding peptides, often overlooked in favour of high-affinity ones. Recent research shows that mid- and low-affinity peptides play a crucial role in generating progenitor-like, multi-epitope targeting T cells that maintain immune surveillance. Incorporating peptides with varying HLA affinities could enhance immune vigilance<sup>80,265–267</sup>, and some studies indicate that the commonly used threshold for predicted affinity of 500 nM is by far too low for some HLA allotypes, especially in rare ones<sup>251</sup>.

In this regard, during the prioritization step, we made sure to include a sufficient number of classically defined binding epitopes ( $IC_{50} < 500$  nM) to cover more than 90% of the European, Spanish, and global populations. Additionally, when dividing peptides into sub-groups (pools A to D), we ensured that higher binding neoantigens were distributed across different alleles in each pool. This approach was designed to minimize competitive binding during DC antigen presentation, ensuring that strong-binding and weak-binding peptides for the same HLA allele were not placed in the same pool.

Recently, Bolivar et al. published a study analysing the mutational and neoantigen landscape of LS colorectal lesions, including crADs<sup>263</sup>. In that study, 65% of the top 100 predicted FSDNs were immunogenic *in vitro*. These were present in the majority of CRCs as well as in a significant proportion of CrADs. Their findings highlighted a significant number of MMRd CrADs bearing high rate of neoantigens and immune activation. Even though our initial prediction of FSDN originated from cancer sequences, we incorporated CrADs in the analysis, recognizing them as the main CRC precursor in LS. In our comparative analysis of sequenced ECs and CrADs, despite inherent differences—such as CrADs being precursor lesions rather than fully developed cancers—we observed shared mutations between these samples. Interestingly, we identified more predicted FSDNs in CrADs than in ECs, a recurrent finding that we also see in the analysed HNPCC-sys data, likely due to the immune pressure exerted on cancers. This pressure may eliminate clones with high mutational loads and FSDN expression, favouring those capable of immune evasion. Distinct characteristics between these samples were further observed when validating the selected FSDNs in CRC and EC tumour cell lines, where shared mutations were also evident. This is consistent with our

*in vitro* results, which indicate that FSDNs are generated at the initial stages of LS carcinogenesis and can induce T-cell responses. Interestingly, Bohaumilitzky L et al. observed dense CD8+ T-cell infiltration in the normal mucosa of LS patients, indicating early recognition of FSDNs<sup>193</sup>. This finding supports our focus on targeting FSDN in precursor lesions to enhance immunoprevention strategies in LS.

The present work suggests that carriers of *MLH1* and *MSH2* pathogenic variants exhibit a higher frequency of FSDN-associated mutations than those found in *MSH6* carriers, aligning with established data that *MLH1* and *MSH2* mutations are associated with a greater mutation burden and higher rates of neoplastic occurrences<sup>1,268,269</sup>. The differential in mutation load is significant regarding FSDN-based vaccines as it may directly impact their efficacy: *MLH1* and *MSH2* pathogenic variant carriers may present a more diverse and abundant pool of immunogenic targets, potentially leading to a stronger and more sustained vaccine-induced immune response. Conversely, *MSH6* carriers, with fewer FSDNs, could have a less robust response to the same vaccine formulations. Therefore, tailored approaches based on the specific MMR mutation may be required to optimize vaccine efficacy.

We acknowledge several limitations in our study. First, although low-affinity (IC<sub>50</sub> < 5000 nM) HLA binding predictions may raise concerns, we believe weak binders are key to preventing immune evasion by boosting recognition of overlooked antigens. Second, while HLA Class II predictions are less robust due to the promiscuous nature of these molecules, the tools we utilized remain the best available for this purpose. The inclusion of class II epitopes is crucial for eliciting CD4+ T cell responses, which are vital in supporting long-lasting antitumour immunity. Third, our *in vitro* validation was conducted on a small cohort of patients. Despite this, the immune responses observed in these individuals suggest the potential efficacy of our approach and highlight the need for larger studies to validate these preliminary findings. We recognize that the limited cohort of sequenced LS samples in our study (comprising 17 CrADs and 8 ECs) constrains the generalizability of our findings. Nevertheless, the overlap of the presence of FSDN associated mutations is clinically significant.

### 7.1.2 Chapter II

In this chapter, we explored the characteristics of peripheral blood and colorectal mucosa T cells in healthy LS carriers, confirming that the lymphocyte populations in this group differ from those found in healthy donors, both in circulation and within infiltrating tissues. After determining that the prioritized neoantigens from Chapter I do not influence the maturation or stimulatory capacity of DCs, we confirmed that these peptides can elicit immune responses in colorectal mucosa T cells from LS previvors and survivors. These findings indicate the presence of memory T cells in the mucosa that had previously recognized these neoantigens. Knowing that our selected peptides can induce immune responses and that some appear to be present in precancerous lesions, we proceeded with the clinical trial using this set of candidates.

When looking into our entire LS cohort based on phenotypic data (e.g., sex, age, germline mutation, clinical history), we observed that while the overall cohort behaves similarly in absolute numbers, differences emerge when stratified by age. Notably, *MSH6* mutation carriers typically do not develop malignancies until around the age of 40, whereas *MLH1* and *MSH2* carriers often present with tumours between the ages of 25 and 40. This finding is consistent with previous reports<sup>18</sup> and has important implications for cancer prevention. These patients should be considered when designing vaccination strategies and statistically evaluating the impact of immunoprevention.

By analysing the HLA haplotype distribution in our LS cohort, we observed that it differs from the European population as reported in the HLA allele frequencies database<sup>231</sup>. Specifically, we found both over- and under-representation of certain alleles, including a lower representation of HLA-A\*02:01 and a higher representation of HLA-A\*29:02. When correlating HLA data with phenotypic variables such as sex and cancer incidence, we identified significant allele associations. Some alleles were more prevalent in males or females, and notably, DRB\*11:01 was negatively associated with cancer risk, whereas A\*26:01 was positively associated. This analysis has been expanded on a larger scale by research groups from the Heidelberg Institute for Theoretical Studies (HITS) and the German Cancer Research Centre (DKFZ) through the INDICATE project—an international collaborative initiative aimed at understanding the role of HLA type as a risk modifier in individuals with a genetic predisposition to cancer. Our group has also contributed to this initiative as understanding the association between HLA alleles and

cancer risk is crucial, particularly when considering LS preferential cohorts for vaccination strategies.

We then sought to understand LS immune responses behaviour and deepen the knowledge on the shaping effect of LS on the immune system, and whether the high production on neoantigens induced changes on populations on both circulating and mucosa T-cells. This may impact on immune responses generated to vaccination. Some studies have previously analysed these cell populations in FFPE specimens of tumour-distant normal colonic mucosa<sup>193</sup>, as well as the local immune signaling microenvironment from *ex-vivo* explanted colorectal biopsies<sup>255</sup>. The first study found that LS normal mucosa showed significantly elevated CD3-, FOXP3-, and CD8-positive T-cell densities compared with non-LS cancer control specimens. On the other hand, Brand *et al.* observed that cytokine based local immune microenvironment profiling could identify sets of cytokines whose differential expression was predictive of LS-status compared to sporadic CRC patients, as well as a LS and CRC status-dependent constitutive inflammatory state of the normal mucosa. Based on these indications that normal mucosa immune populations of LS individuals are altered, we wanted to corroborate it and compare it to that of healthy subjects, which had not been done previously. Moreover, this is a field which has very little information, and a deeper profiling of T-cell populations and TCR repertoire had also not been done before, so we aimed to fill that gap.

To that end we characterized peripheral blood and normal mucosa infiltrating lymphocytes from these individuals and correlated it with a cohort of healthy controls matched in age and sex. Similar to what L. Bohaumilitzky *et al.* observed<sup>193</sup>, we found increased infiltration of CD3+ T-cells in mucosa. Interestingly, we found that the TCM subset was significantly underrepresented in both peripheral blood and mucosa CD8+ and CD4+ T-cells from LS individuals but also found a higher proportion of CD8+ TEM in mucosa. These proportions did not change after stratification of patients in previvors and survivors. TCM cells primarily reside in lymphoid organs and circulate in the blood, providing long-term immunosurveillance against known pathogens. They have a high capacity for persistence, self-renewal, and proliferation, making them essential for sustained immune memory. Although studies suggest they may contribute to tissue immune surveillance due to the expression of peripheral tissue-homing receptors at

levels similar to their effector memory counterparts, their role is still not well characterized<sup>270</sup>. In contrast, TEM cells are more widely distributed, circulating in the blood and trafficking through peripheral organs. TEM exhibit a rapid effector response, readily differentiating into effector T cells (TEFF) that produce high levels of IFN- $\gamma$  and display strong cytotoxic activity. Meanwhile, TCM are less differentiated, acquire effector functions more gradually, and predominantly produce IL-2<sup>270-272</sup>. An increased population of CD8+ TEM cells on LS mucosa was expected due to immune surveillance-based detection of neoantigens and memory formation in MSI foci on the colorectal mucosa<sup>273,274</sup>. However, the observation of decreased TCM representation was more surprising, and corroborated in Dr. Beatriz Carreno group in the University of Pennsylvania, who had similar results in their cohort of LS individuals (preliminary data from unpublished results). The increase in TEM CD8+ T-cell compartment may be at the expense of CD8+ TCM cells, however this increase is observed in both peripheral blood and mucosa of either CD8+ and CD4+ T-cells, suggesting a broader systemic impact.

Some limitations should be considered for this analysis. Our antibodies targeted CD8 $\alpha$ , which may recognize CD8 $\alpha\alpha$  receptors from NKT cells or IELs<sup>275</sup> rather than the conventional CD8 $\alpha\beta$  from T-cells, offering another explanation for the observed differences, as it has been demonstrated that some NKT cells can express all the markers used here to identify T-cell subsets<sup>276</sup>. Additionally, T-cells were not distinguished by  $\alpha\beta$  or  $\gamma\delta$  lineages. Despite these limitations, the observed robust differences between LS and healthy controls highlight the need for further analysis of T-cell populations and the potential role of NKT and IEL cells in LS.

Given that no previous studies have analysed the TCR repertoire in LS individuals compared to healthy controls, we conducted an initial screening in a sub-cohort of the previously phenotyped patients to identify any significant changes that might justify expanding our analysis to a larger cohort and employing more specialized techniques such as NGS or single-cell RNA sequencing. Preliminary results from our analysis of the TCR V $\beta$  repertoire in PBMCs and colorectal mucosa samples revealed significant differences in T-cell clonality between LS patients and healthy controls, particularly in mucosa-infiltrating T-cells. These findings suggest that systemic immune alterations in LS may be relatively subtle, with more pronounced changes occurring in the colorectal mucosa. This is consistent with previous cytometry phenotyping results. We also

observed a significantly higher number of expanded TCR V $\beta$  families in LS mucosa samples compared to LS PBMCs, and automated hierarchical clustering identified a distinct V $\beta$  family signature that differentiated LS T-cells from those of healthy controls exclusively in mucosal samples.

Lastly, following the findings of other groups on the differential microenvironment profiles depending on LS status and cancer type<sup>255</sup>, we analysed the cytokine production profile of LS individuals and healthy controls' PBMCs at the single-cell level and found that LS carriers exhibited a significantly higher PSI in CD8+ T cells, along with increased polyfunctionality, with fewer non-secreting cells and an increased frequency of MIP-1 $\beta$  and TNF- $\alpha$  production. In contrast, CD4+ cells showed no significant differences, though a trend toward reduced polyfunctionality and lower IFN- $\gamma$  production was observed. This suggests a distinct cytokine profile in LS carriers, characterized by a more functionally active CD8+ T-cell response. Further characterization through single-cell RNA-seq or in vitro functional studies could provide deeper insights into the underlying mechanisms and potential implications for immune surveillance in LS.

While these preliminary findings provide valuable insights to deepen the analysis in the future, several limitations must be considered: The small sample size restricts the robustness of the analysis, necessitating further studies with larger cohorts. Additionally, the lack of PCR validation limits the depth of our characterization of TCR V $\beta$  repertoire changes, although future single-cell RNA sequencing may provide further clarity. Moreover, expanding the healthy control reference cohort would also enhance statistical power and the reliability of comparative analyses.

Regarding DELAY-FSDN immune responses, the results presented in this chapter provide significant insights into the immune surveillance mechanisms in LS by demonstrating the presence of prioritized neoantigen-specific T-cells in normal mucosa and premalignant lesions. These findings also contribute to the growing body of evidence that LS carriers exhibit pre-existing immune recognition of neoantigens, supporting the hypothesis that early immune activation plays a crucial role in cancer prevention<sup>193,263</sup>.

Previous work by Roudko *et al.* and Ballhausen *et al.*, among others, demonstrated the frequent presence and high immunogenicity of tumour-specific antigens resulting from

common frameshift mutations in MSI-H cancer and Lynch syndrome patients<sup>80</sup>, and that frameshift mutation frequency inversely correlated with peptide immunogenicity, indicating counterselection of highly immunogenic clones, except for some driver mutations which persist in MSI cancers<sup>56</sup>. With a focus on premalignancies, the work from Bolivar *et al.* demonstrated that mutation burden and neoantigen load increased with disease progression starting at the pre-cancer stage, and that most of their top predicted peptides were immunogenic in vitro and shared within several cancer types<sup>263</sup>, indicating that these immune responses are initiated before malignant transformation occurs due to immune surveillance.

In this context, our data demonstrated that both in adenomas and healthy colorectal mucosa of LS previvors and survivors, there exist tissue-infiltrating T-cells that can recognize our prioritized neoantigens, indicating that they are presented in intestinal MSI foci, can be recognized as pathogenic and form long-term memory responses. Concretely, we found IFN- $\gamma$  producing responses in 53% of the tested LS samples, with some increasing the strength of the response after a longer exposure (4 weeks of culture) and some losing specificity. These percentages are in line with other studies testing neoantigens in personalized vaccines. Bolivar *et al.* found that 75% of samples generated up to 8 immunogenic neoantigens predicted from MSI-H pre-cancers<sup>263</sup> (which, interestingly, were validated in T-cells from healthy donors); Zhang *et al.* found that from 198 predicted neoantigens for triple-negative breast cancer, 47 (23%) of them were found immunogenic by ELISPOT<sup>277</sup>.

The peptides we selected, as commented on the previous chapter discussion, had a wide range of affinities and included HLA-I and HLA-II restriction. Although the majority of our experiments focused on HLA-I neoantigens in the pursuit of isolating specific T-cells to test their cytotoxicity on the TMG-transduced cell line, which only contains HLA-I DELAY-FSDNs, we demonstrated that both CD4+ and CD8+ T-cells can recognize their respective neoantigens. Ensuring that this representation is translated in vivo after vaccination is essential for inducing a robust immune response against FSDN involving all components of the adaptive cellular response, and a long-term memory afterwards<sup>198,278</sup>. Furthermore, studies have demonstrated that instead of non-specific helper antigens, tumour-specific MHC class II-restricted neoantigens are crucial for eliciting effective CD4+ T cell help, which enhances vaccine-induced CD8+ T cell

responses and improves tumour control<sup>279,280</sup>. On the other hand, and as also commented before, the inclusion of lower-binding neoantigens is encouraged by accumulating evidence to promote immune surveillance against overlooked antigens and to diversify the progenitor-like, multi-epitope targeting T cells<sup>80,265–267</sup>. Regarding this, we found IFN- $\gamma$  producing responses in patients lacking predicted high-affinity binding epitopes (all predicted IC<sub>50</sub> >500 nM), which indicates that these FSDN can still be recognized as pathogenic and mount a response on T-cells.

Another important factor is the targeting of driver genes, which is important to avoid immune evasion by antigenic loss, and to understand the pathways involved with the peptides that we are immunizing against. In our FSDN set we included both tumour driver genes and tumour-related genes (or Oncogenes). So far, with the patients that we were able to deepen the analysis, we found T-cells specific for the *TGFBRII*, *MSH3*, *AXIN2*, *MSH6* and *BAX* genes. *TGFBRII* loss disrupts the TGF- $\beta$  signaling pathway, a key regulator of cell proliferation, differentiation, and apoptosis. Its inactivation promotes tumour growth and immune evasion. *MSH3* and *MSH6* are both part of the DNA MMR system, altered in LS. As specified before, mutations in these genes contribute to genomic instability, leading to the accumulation of mutations that drive tumorigenesis. *AXIN2* is a crucial regulator of the Wnt/ $\beta$ -catenin signaling pathway, when mutated it results in aberrant pathway activation, promoting uncontrolled cell proliferation and tumour development. Finally, *BAX* is a key pro-apoptotic gene in the intrinsic apoptosis pathway and its loss allows cancer cells to evade programmed cell death, contributing to tumour survival and progression<sup>281</sup>. These results indicate that all responsive peptides originate from genes that contribute to tumour development and survival. Finding memory cells against these antigens suggests that there has been an active process of immune surveillance and immune editing in the mucosa from these LS individuals, potentially leading to lesion elimination (in the case of healthy mucosa samples) or to the formation of adenomas.

One of the challenges in translational projects in preventive settings is the demonstration of treatment efficacy. To demonstrate that prevention is effective in a pre-clinical setting, a model simulating a cancer high-risk individual is necessary. So far, the group of Gebert and Kloor *et al.* demonstrated the potential of FSDN vaccines for LS cancer prevention using *VCMSH2* mice, which have a conditional knockout of *MSH2* in

the intestinal tract and develop intestinal tumours. They identified four shared immunogenic frameshift peptides in these mice—FSDN-1 Nacad, Maz, Senp6, and Xirp1—and showed that vaccination with these neoantigens enhanced adaptive immunity, reduced tumour burden, and prolonged survival, and that these outcomes were potentiated by daily administration of naproxen, and at lesser extent with ASA<sup>83</sup>. However, this approach cannot be applied to human neoantigen vaccines as it depends on the tumoral proteins developed by mice.

Other strategies that can be used as a proof of concept of the potential of neoantigen immunized T-cells are organoids and cell lines which express the interest peptides. To that end, we developed K562 monoallelic HLA-I cell lines, which were then transduced with a tandem minigene that encompasses all HLA-I restricted peptides included in the DELAY vaccine, to then use these cells as targets for DELAY-immunized T-cells. This strategy can be implemented as routine immune monitoring in the clinical trial. The viability of this approach has been previously assessed by the group of B. Carreño *et al.*, who validated the mKRAS antigen processing and presentation with targeted mass spectrometry on peptides isolated from single HLA class I allele/TMG-expressing cell lines<sup>244</sup>. On the other hand, our group has also developed and implemented the generation of colorectal organoids from healthy mucosa and MSI and MSS tumours, where two of the generated samples were from LS individuals. Unpublished data from these models demonstrates that they retain their original tumour-associated mutations, have high expression of LGR5 (an intestinal stem cell marker) and EPCAM, meaning they are defined as epithelial human colorectal cancer stem cells<sup>282</sup>, and can be maintained by long periods in culture and used in tumour recognition assays with autologous TILs. This approach could be implemented in the LS cancer prevention setting by generating organoids from pre-vaccinated patients, ensuring loss of the affected MMR protein or knocking out the previously healthy allele, allowing neoantigen generation through accumulated mutations during serial passaging and using them as targets for cytotoxicity assays against autologous DELAY-immunized T-cells. This would be, even though still not fully accurate, a more similar representation of an *in vivo* situation after vaccination.

Despite these promising findings, several limitations must be acknowledged. The study cohort was relatively small, and further validation in larger, independent populations is

necessary. Additionally, the extent to which these immune responses translate into effective tumour prevention remains to be fully understood. Functional assays to determine the cytotoxic capacity of these T-cells against LS-related tumours will be pursued in future studies.

### 7.1.3 Chapter III

The Phase Ib clinical trial outlined in this chapter represents a crucial step toward translating the immunological findings from Chapter II into a potential preventive intervention for LS carriers. These preliminary results demonstrate the feasibility of a GMP-produced dendritic cell-based vaccine targeting FSDN neoantigens.

Kloor *et al.* were the first to explore neoantigen vaccination in LS patients, conducting a Phase I/IIa trial in 22 individuals with stage III/IV MMRd CRC using three neoantigens (TAF1B, HT001, and AIM2), which enhanced CD8+ and CD4+ T-cell immunity<sup>57</sup>. More recently, a Phase I trial tested a neoantigen-based adenoviral vaccine encoding 209 frameshift peptides in metastatic MMRd cancers alongside PD-1 inhibitor pembrolizumab, leading to increased tumour-infiltrating T cells with an effector memory signature<sup>283</sup>. The same vaccine was tested as an immunopreventive strategy on healthy LS carriers, demonstrating safety and the capacity of eliciting potent and broad immune responses in both LS carriers previvors and survivors<sup>82</sup>. Other studies such as E. Vilar's work in the University of Texas MD Anderson Cancer Centre, are being used to develop vaccine strategies based on their discovery of immunogenic shared neoantigens within unstable microsatellite loci at initial stages of LS carcinogenesis<sup>263</sup>.

Focusing on DC-based vaccines, a seminal study led by de Vries and colleagues at Radboud University Medical Centre in Nijmegen investigated a DC-based vaccine targeting 2 FSDN (*TGFBR2* and *CASP-5*; NCT01885702). Unpublished data demonstrated strong T-cell responses against the *TGFBR2* peptide with preliminary data suggesting a potential long-term decrease in cancer incidence among LS carriers who responded to this FSDN. In addition, NOUSCOM, a biotechnology company, is conducting a Phase 1b clinical trial (NCT05078866) evaluating the efficacy of an off-the-shelf viral vector-based cancer vaccine called NOUS-209 in LS carriers that encodes 209 shared FSDN in LS-associated tumours<sup>283</sup>. Preliminary results from the first 10 LS individuals showed that the vaccine was safe, well-tolerated, and capable of generating potent and broad

immunogenic T-cell responses<sup>82</sup>. On the other hand, a recent study by Inges *et al.* focused on the use of a DC vaccine targeting patient-specific neoantigens in resected NSCLC. In a Phase I clinical trial, the vaccine had only mild adverse events (grade 1–2) and systemic T cell responses were observed in five of the six vaccinated patients and remained detectable for up to 19 months. Despite disease recurrence in three patients over a two-year follow-up, the study demonstrated the feasibility, safety, and immunogenicity of this approach<sup>256</sup>.

All these studies have demonstrated the safety of targeting neoantigens, the latter being particularly interesting as they not only confirm safety but also induce robust immune responses through DC-based vaccination. In this regard, our centre has extensive experience in producing DC-based therapies, including eight clinical trials (three Phase I, one Phase Ib, one Phase I-II, and three Phase II). Although these trials showed limited clinical responses, they demonstrated our centre's capacity for GMP-grade production, as well as the safety and immunogenicity of the treatment in patients. Notably, while several FSDNs in the referenced studies overlap with ours, we have specifically focused on selected epitopes with varying affinities rather than the entire *frameshift* protein. While the mutated protein may be shared across studies, the specific epitopes targeted in our study are distinct. Based on these findings, we anticipated minimal challenges in vaccine manufacturing and limited adverse events.

Indeed, we have confirmed that the GMP manufacturing of our product meets all required criteria, ensuring sterility and yielding a sufficient number of cells from the starting material for six doses of 10 million peptide-loaded moDCs per dose. Additionally, we have demonstrated that these cells exhibit a mature phenotype with high co-stimulatory molecule expression while retaining their ability to stimulate allogeneic T cells.

Basal immunomonitoring results show that one of the recruited LS carriers had pre-existing FSDN-specific T-cell responses pre-vaccination, further validating our findings from Chapter II. Long-term follow-up will be necessary to determine whether these immune responses are sustained and whether their existence correlates with a reduced incidence of colorectal and endometrial cancer in LS patients, compared to carriers without pre-existing immunity.

A key limitation of this study is the small cohort size, which restricts the statistical power of our findings. While the absence of a placebo group limits the ability to draw definitive conclusions about vaccine efficacy, this is consistent with the primary objectives of a Phase I study, which focuses on safety and immunogenicity. Evaluation of efficacy should be addressed in future phases through randomized controlled trials with larger cohorts. Another challenge is the heterogeneity of LS mutations, which may impact vaccine responsiveness across different patients. Another limitation is the scalability of our vaccine production. The manufacturing of an autologous cellular vaccine is costly and takes a long time and a high number of resources to be produced. At a small scale, this does not present a challenge. However, if needing to increase production, for example for a phase II or III trial, this would be challenging. This is why we are currently exploring other formulations for the vaccine that would better translate to the clinic, such as the use of peptide- or mRNA-loaded nanoparticles targeted to DCs. We have previously validated the encapsulation of bioactive compounds in water-dispersible polyurethane-polyurea hybrid nanoparticles, demonstrating their efficient internalization in moDCs and selective accumulation in tumour tissues<sup>284</sup>. These nanoparticles have also been loaded with immunogenic (R848, a TLR-7 and 8 agonist) and tolerogenic compounds (Budesonide). When internalized by moDCs, these nanoparticles induced a mature and pro-inflammatory phenotype (unpublished results) or a tolerogenic phenotype characterized by downregulation of costimulatory molecules, increased IL-10 secretion, and suppression of T-cell activation<sup>114</sup>, respectively. We believe that this strategy would enable more efficient vaccine production and distribution, transitioning from an autologous to a universal approach.

Despite these limitations, this trial represents an important first step in exploring preventive immunotherapy in LS. The results support further investigation into FSDN-based vaccines as a means of enhancing immune surveillance and potentially reducing cancer incidence in LS carriers. If validated in larger trials, this vaccine approach could offer a non-invasive, immunologically driven alternative to current LS management strategies, which rely heavily on frequent colonoscopy surveillance and prophylactic surgery and are lacking for many LS-associated extracolonic cancers. We also expect that this trial will yield insights into antigen-specific responses to further refine the antigen selection and vaccine formulation in the future.



## 7.2 Conclusions

The conclusions for this thesis are the following:

### *Chapter I: Bioinformatic neoantigen prediction and prioritization*

1. Using publicly available tumour sequences, current neoantigen prediction pipelines can effectively identify shared MHC-I and MHC-II restricted neoantigens associated with MMRd.
2. One-base deletion mutations are the most prevalent type of mutations in Lynch Syndrome colorectal adenomas and endometrial cancers.
3. The identified FSDN-associated mutations are highly recurrent in colorectal pre-cancers, colorectal cancers, and endometrial cancers linked to Lynch Syndrome, with most being shared across different cancer types.
4. MLH1 and MSH2 pathogenic variant carriers exhibit a significantly higher mutation burden and a greater frequency of FSDN-associated mutations compared to MSH6 variant carriers.
5. The prioritized FSDN-associated mutations are frequent in sporadic MMRd MSI-H tumours of diverse origins, suggesting that immunization strategies could also benefit non-LS MSI-H tumour patients.
6. The variants linked to selected FSDN are expressed at the DNA level and transcribed at high allele frequencies in MSI endometrial and colorectal cancer cell lines.

### *Chapter II: Prioritized neoantigens functional validation*

7. Peripheral and mucosa-infiltrating T-cells in healthy Lynch Syndrome carriers display distinct subpopulations, TCR repertoires, and secretome profiles compared to healthy controls, indicating a systemic immune modulation due to the syndrome.
8. The selected FSDN do not impair dendritic cell viability, maturation, or ability to induce T-cell proliferation; once loaded, these cells elicit T-lymphocytes

specific IFN- $\gamma$ -producing responses in Lynch Syndrome colorectal mucosa lymphocytes.

9. FSDN-specific T-cells are present in colorectal adenomas and normal mucosa of Lynch Syndrome patients, even in the absence of cancer history, confirming immune surveillance and recognition of these neoantigens in early MSI foci.
10. Both high- and low-binding affinity prioritized FSDN can trigger IFN- $\gamma$ -producing immunogenic responses in these T-cells, which could be relevant to overcome immune tolerance to low immunogenic peptides.

### *Chapter III: Clinical trial validation*

11. The DELAY vaccine's manufacturing under Good Manufacturing Practice conditions is feasible and meets all established quality criteria.

## BIBLIOGRAPHY

1. Dominguez-Valentin, M. *et al.* Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. *Genet Med* **22**, 15–25 (2020).
2. Daca Alvarez, M. *et al.* The Inherited and Familial Component of Early-Onset Colorectal Cancer. *Cells* **10**, 710 (2021).
3. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895–2015. *Nat Rev Cancer* **15**, 181–194 (2015).
4. Vasen, H. F. A. Progress in Genetic Testing, Classification, and Identification of Lynch Syndrome. *JAMA* **293**, 2028 (2005).
5. Pećina-Šlaus, N., Kafka, A., Salamon, I. & Bukovac, A. Mismatch Repair Pathway, Genome Stability and Cancer. *Front Mol Biosci* **7**, 1–12 (2020).
6. Moreira, L. *et al.* Identification of Lynch Syndrome Among Patients With Colorectal Cancer. *JAMA* **308**, 1555 (2012).
7. Kravochuck, S. E., Kalady, M. F., Burke, C. A., Heald, B. & Church, J. M. Defining HNPCC and Lynch syndrome: what's in a name?: Table 1. *Gut* **63**, 1525–1526 (2014).
8. Jass, J. R. Hereditary non-polyposis colorectal cancer: The rise and fall of a confusing term. *World J Gastroenterol* **12**, 4943 (2006).
9. Win, A. K. *et al.* Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. *Cancer Epidemiology, Biomarkers & Prevention* **26**, 404–412 (2017).
10. Møller, P. *et al.* Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: First report from the prospective Lynch syndrome database. *Gut* **66**, 464–472 (2017).
11. Llach, J., Pellisé, M. & Monahan, K. Lynch syndrome; towards more personalized management? *Best Pract Res Clin Gastroenterol* **58–59**, 101790 (2022).

12. de Jong, A. E. *et al.* Decrease in Mortality in Lynch Syndrome Families Because of Surveillance. *Gastroenterology* **130**, 665–671 (2006).
13. Sánchez, A. *et al.* Quality of Colonoscopy Is Associated With Adenoma Detection and Postcolonoscopy Colorectal Cancer Prevention in Lynch Syndrome. *Clin Gastroenterol Hepatol* (2020) doi:10.1016/j.cgh.2020.11.002.
14. Stjepanovic, N. *et al.* Hereditary gastrointestinal cancers: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* 1558–1571 (2019) doi:10.1093/annonc/mdz233.
15. Balmaña, J., Balaguer, F., Cervantes, A. & Arnold, D. Familial risk-colorectal cancer: ESMO clinical practice guidelines. *Annals of Oncology* **24**, vi73–vi80 (2013).
16. Concin, N. *et al.* ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *International Journal of Gynecologic Cancer* **31**, 12–39 (2021).
17. Provenzale, D. *et al.* Genetic/Familial High-Risk Assessment: Colorectal Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network* **14**, 1010–1030 (2016).
18. Kastrinos, F. *et al.* Gene-Specific Variation in Colorectal Cancer Surveillance Strategies for Lynch Syndrome. *Gastroenterology* **161**, 453-462.e15 (2021).
19. D’Andrea, E., Ahnen, D. J., Sussman, D. A. & Najafzadeh, M. Quantifying the impact of adherence to screening strategies on colorectal cancer incidence and mortality. *Cancer Med* **9**, 824–836 (2020).
20. Singal, A. G. *et al.* Effect of Colonoscopy Outreach vs Fecal Immunochemical Test Outreach on Colorectal Cancer Screening Completion. *JAMA* **318**, 806 (2017).
21. Matthews, C. A. Management Strategies for the Ovaries at the Time of Hysterectomy for Benign Disease. *Obstet Gynecol Clin North Am* **43**, 539–549 (2016).
22. Burgess, J. T. *et al.* The Therapeutic Potential of DNA Damage Repair Pathways and Genomic Stability in Lung Cancer. *Front Oncol* **10**, (2020).

23. Ma, Q. *et al.* Increased chromosomal instability characterizes metastatic renal cell carcinoma. *Transl Oncol* **14**, 100929 (2021).
24. Lasolle, H. *et al.* Chromosomal instability in the prediction of pituitary neuroendocrine tumors prognosis. *Acta Neuropathol Commun* **8**, 190 (2020).
25. Sharma, D. *et al.* DNA binding and antiradical potential of ethyl pyruvate: Key to the DNA radioprotection. *Chem Biol Interact* **332**, 109313 (2020).
26. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**, 235–263 (2017).
27. Ide, H., Nakano, T., Salem, A. M. H. & Shoulkamy, M. I. DNA–protein cross-links: Formidable challenges to maintaining genome integrity. *DNA Repair (Amst)* **71**, 190–197 (2018).
28. Evans, M. D., Dizdaroglu, M. & Cooke, M. S. Oxidative DNA damage and disease: induction, repair and significance. *Mutation Research/Reviews in Mutation Research* **567**, 1–61 (2004).
29. Li, J. *et al.* Pathways and assays for DNA double-strand break repair by homologous recombination. *Acta Biochim Biophys Sin (Shanghai)* **51**, 879–889 (2019).
30. Forster, J. C., Douglass, M. J. J., Phillips, W. M. & Bezak, E. Stochastic multicellular modeling of x-ray irradiation, DNA damage induction, DNA free-end misrejoining and cell death. *Sci Rep* **9**, 18888 (2019).
31. Wityk, P., Piątek, R., Nowak, R. & Kostrzewa-Nowak, D. Generation and Characterization of a DNA-GCN4 Oligonucleotide-Peptide Conjugate: The Impact DNA/Protein Interactions on the Sensitization of DNA. *Molecules* **25**, 3630 (2020).
32. Cai, Y., Geacintov, N. E. & Broyde, S. Variable impact of conformationally distinct DNA lesions on nucleosome structure and dynamics: Implications for nucleotide excision repair. *DNA Repair (Amst)* **87**, 102768 (2020).
33. Kajitani, G. S. *et al.* Transcription blockage by DNA damage in nucleotide excision repair-related neurological dysfunctions. *Semin Cell Dev Biol* **114**, 20–35 (2021).

34. Gavande, N. S. *et al.* Structure-Guided Optimization of Replication Protein A (RPA)–DNA Interaction Inhibitors. *ACS Med Chem Lett* **11**, 1118–1124 (2020).
35. Sinha, A. *et al.* RAD51-Mediated DNA Homologous Recombination Is Independent of PTEN Mutational Status. *Cancers (Basel)* **12**, 3178 (2020).
36. Shibata, A. & Jeggo, P. A. Roles for the DNA-PK complex and 53BP1 in protecting ends from resection during DNA double-strand break repair. *J Radiat Res* **61**, 718–726 (2020).
37. Inagawa, T. *et al.* C-Terminal Extensions of Ku70 and Ku80 Differentially Influence DNA End Binding Properties. *Int J Mol Sci* **21**, 6725 (2020).
38. Koike, M. & Koike, A. The Ku70-binding site of Ku80 is required for the stabilization of Ku70 in the cytoplasm, for the nuclear translocation of Ku80, and for Ku80-dependent DNA repair. *Exp Cell Res* **305**, 266–276 (2005).
39. Jiricny, J. Postreplicative Mismatch Repair. *Cold Spring Harb Perspect Biol* **5**, a012633–a012633 (2013).
40. Bębenek, A. & Ziuzia-Graczyk, I. Fidelity of DNA replication—a matter of proofreading. *Curr Genet* **64**, 985–996 (2018).
41. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**, 85–98 (2008).
42. Kunkel, T. A. & Erie, D. A. DNA MISMATCH REPAIR. *Annu Rev Biochem* **74**, 681–710 (2005).
43. Sharma, M., Predeus, A. V., Kovacs, N. & Feig, M. Differential Mismatch Recognition Specificities of Eukaryotic MutS Homologs, MutS $\alpha$  and MutS $\beta$ . *Biophys J* **106**, 2483–2492 (2014).
44. Hewish, M., Lord, C. J., Martin, S. A., Cunningham, D. & Ashworth, A. Mismatch repair deficient colorectal cancer in the era of personalized treatment. *Nat Rev Clin Oncol* **7**, 197–208 (2010).
45. Olave, M. C. & Graham, R. P. Mismatch repair deficiency: The what, how and why it is important. *Genes Chromosomes Cancer* **61**, 314–321 (2022).

46. Martín-López, J. V. & Fishel, R. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. *Fam Cancer* **12**, 159–168 (2013).
47. Acharya, S. *et al.* hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proceedings of the National Academy of Sciences* **93**, 13629–13634 (1996).
48. Kadyrov, F. A., Dzantiev, L., Constantin, N. & Modrich, P. Endonucleolytic Function of MutL $\alpha$  in Human Mismatch Repair. *Cell* **126**, 297–308 (2006).
49. Taieb, J. *et al.* Deficient mismatch repair/microsatellite unstable colorectal cancer: Diagnosis, prognosis and treatment. *Eur J Cancer* **175**, 136–157 (2022).
50. Woerner, S. M., Kloor, M., Von Knebel Doeberitz, M. & Gebert, J. F. Microsatellite instability in the development of DNA mismatch repair deficient tumors. *Cancer Biomarkers* **2**, 69–86 (2006).
51. De' Angelis, G. L. *et al.* Microsatellite instability in colorectal cancer. *Acta Biomed* **89**, 97–101 (2018).
52. Poulogiannis, G., Frayling, I. M. & Arends, M. J. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* **56**, 167–179 (2010).
53. Cicek, M. S. *et al.* Quality Assessment and Correlation of Microsatellite Instability and Immunohistochemical Markers among Population- and Clinic-Based Colorectal Tumors. *The Journal of Molecular Diagnostics* **13**, 271–281 (2011).
54. Zhao, P., Li, L., Jiang, X. & Li, Q. Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. *J Hematol Oncol* **12**, 54 (2019).
55. Thol, K. & McGranahan, N. Potential Use of Shared Frameshift Mutations in 'Off-the-Shelf' Neoantigen Vaccines. *Trends Cancer* **7**, 175–177 (2021).
56. Ballhausen, A. *et al.* The shared frameshift mutation landscape of microsatellite-unstable cancers suggests immunoediting during tumor evolution. *Nat Commun* **11**, 1–13 (2020).

57. Kloor, M. *et al.* A Frameshift Peptide Neoantigen-Based Vaccine for Mismatch Repair-Deficient Cancers: A Phase I/IIa Clinical Trial. *Clinical Cancer Research* **26**, 4503–4510 (2020).
58. Xie, N. *et al.* Neoantigens: promising targets for cancer therapy. *Signal Transduct Target Ther* **8**, 9 (2023).
59. Kang, Y.-J. *et al.* A scoping review and meta-analysis on the prevalence of pan-tumour biomarkers (dMMR, MSI, high TMB) in different solid tumours. *Sci Rep* **12**, 20495 (2022).
60. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
61. Vasen, H. F. A. Clinical Description of the Lynch Syndrome [Hereditary Nonpolyposis Colorectal Cancer (HNPCC)]. *Fam Cancer* **4**, 219–225 (2005).
62. Pastor, D. M. & Schlom, J. Immunology of Lynch Syndrome. *Curr Oncol Rep* **23**, 96 (2021).
63. Adam, F., Fluri, M., Scherz, A. & Rabaglio, M. Occurrence of variants of unknown clinical significance in genetic testing for hereditary breast and ovarian cancer syndrome and Lynch syndrome: a literature review and analytical observational retrospective cohort study. *BMC Med Genomics* **16**, 7 (2023).
64. Williams, M. H., Hadjinicolaou, A. V., Norton, B. C., Kader, R. & Lovat, L. B. Lynch syndrome: from detection to treatment. *Front Oncol* **13**, (2023).
65. Khalid, A. “Lynch Syndrome”—From Healer to a Vulnerable Patient: A Transformative Odyssey. *J Patient Exp* **10**, (2023).
66. Barbosa Oliveira, F. F. *et al.* Evaluation of the psychological state of patients suspected of Lynch syndrome after the germline genetic testing result. *Journal of Clinical Oncology* **41**, e24201–e24201 (2023).
67. Kastrinos, F. & Stoffel, E. M. History, Genetics, and Strategies for Cancer Prevention in Lynch Syndrome. *Clinical Gastroenterology and Hepatology* **12**, 715–727 (2014).

68. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895–2015. *Nat Rev Cancer* **15**, 181–194 (2015).
69. Dal Buono, A. *et al.* Lynch Syndrome: From Multidisciplinary Management to Precision Prevention. *Cancers (Basel)* **16**, 849 (2024).
70. Castillo-Iturra, J., Sánchez, A. & Balaguer, F. Colonoscopic surveillance in Lynch syndrome: guidelines in perspective. *Fam Cancer* **23**, 459–468 (2024).
71. Monahan, K. J. *et al.* Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). *Gut* **69**, 411–444 (2020).
72. Reyes-Urbe, L. *et al.* Naproxen chemoprevention promotes immune activation in Lynch syndrome colorectal mucosa. *Gut* **70**, 555–566 (2021).
73. Bowen, C. M. *et al.* Naproxen chemoprevention induces proliferation of cytotoxic lymphocytes in Lynch Syndrome colorectal mucosa. *Front Immunol* **14**, (2023).
74. Hewish, M., Lord, C. J., Martin, S. A., Cunningham, D. & Ashworth, A. Mismatch repair deficient colorectal cancer in the era of personalized treatment. *Nat Rev Clin Oncol* **7**, 197–208 (2010).
75. Harrold, E. *et al.* The frequency of second primary malignancies and colonic polyps in Lynch syndrome with MSI tumors following immune checkpoint blockade. *Journal of Clinical Oncology* **40**, 10505–10505 (2022).
76. Picco, G. *et al.* Werner Helicase Is a Synthetic-Lethal Vulnerability in Mismatch Repair-Deficient Colorectal Cancer Refractory to Targeted Therapies, Chemotherapy, and Immunotherapy. *Cancer Discov* **11**, 1923–1937 (2021).
77. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine* **372**, 2509–2520 (2015).
78. Mastrodomenico, L. *et al.* Personalized Systemic Therapies in Hereditary Cancer Syndromes. *Genes (Basel)* **14**, 684 (2023).

79. Webb, M. J. *et al.* Expression of tumor antigens within an oncolytic virus enhances the anti-tumor T cell response. *Nat Commun* **15**, 5442 (2024).
80. Roudko, V. *et al.* Shared Immunogenic Poly-Epitope Frameshift Mutations in Microsatellite Unstable Tumors. *Cell* **183**, 1634-1649.e17 (2020).
81. Willis, J. A., Reyes-Urbe, L., Chang, K., Lipkin, S. M. & Vilar, E. Immune Activation in Mismatch Repair-Deficient Carcinogenesis: More Than Just Mutational Rate. *Clinical Cancer Research* **26**, 11-17 (2020).
82. D'Alise, M. *et al.* 1526 Nous-209 genetic vaccine encoding shared cancer neoantigens is safe and elicits robust immune response in healthy Lynch syndrome carriers: interim results from Phase 1 cancer interception trial. in *Late-Breaking Abstracts* A1751-A1751 (BMJ Publishing Group Ltd, 2023). doi:10.1136/jitc-2023-SITC2023.1526.
83. Gebert, J. *et al.* Recurrent frameshift neoantigen vaccine elicits protective immunity with reduced tumor burden and improved overall survival in a Lynch syndrome mouse model. *Gastroenterology* (2021) doi:10.1053/j.gastro.2021.06.073.
84. Roudko, V. *et al.* Lynch Syndrome and MSI-H Cancers: From Mechanisms to "Off-The-Shelf" Cancer Vaccines. *Front Immunol* **12**, (2021).
85. Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245-252 (1998).
86. Collin, M. & Bigley, V. Human dendritic cell subsets: An update. *Immunology* (2018) doi:10.1111/imm.12888.
87. Banchereau, J. *et al.* Immunobiology of dendritic cells. *Annu Rev Immunol* **18**, 767-811 (2000).
88. Tacken, P. J., De Vries, I. J. M., Torensma, R. & Figdor, C. G. Dendritic-cell immunotherapy: From ex vivo loading to in vivo targeting. *Nature Reviews Immunology* vol. 7 790-802 Preprint at <https://doi.org/10.1038/nri2173> (2007).

89. Sichien, D. *et al.* IRF8 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and Plasmacytoid Dendritic Cells, Respectively. *Immunity* **45**, 626–640 (2016).
90. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* (1979) **356**, (2017).
91. Breton, G. *et al.* Circulating precursors of human CD1c+ and CD141+ dendritic cells. *Journal of Experimental Medicine* **212**, 401–413 (2015).
92. Chen, M. Y., Zhang, F., Goedegebuure, S. P. & Gillanders, W. E. Dendritic cell subsets and implications for cancer immunotherapy. *Front Immunol* **15**, (2024).
93. Heras-Murillo, I., Adán-Barrientos, I., Galán, M., Wculek, S. K. & Sancho, D. Dendritic cells as orchestrators of anticancer immunity and immunotherapy. *Nat Rev Clin Oncol* **21**, 257–277 (2024).
94. Eisenbarth, S. C. Dendritic cell subsets in T cell programming: location dictates function. *Nat Rev Immunol* **19**, 89–103 (2019).
95. Granot, T. *et al.* Dendritic Cells Display Subset and Tissue-Specific Maturation Dynamics over Human Life. *Immunity* **46**, 504–515 (2017).
96. Kvedaraite, E. & Ginhoux, F. Human dendritic cells in cancer. *Sci Immunol* **7**, (2022).
97. Schlitzer, A. *et al.* IRF4 Transcription Factor-Dependent CD11b+ Dendritic Cells in Human and Mouse Control Mucosal IL-17 Cytokine Responses. *Immunity* **38**, 970–983 (2013).
98. Del Prete, A. *et al.* Dendritic cell subsets in cancer immunity and tumor antigen sensing. *Cell Mol Immunol* **20**, 432–447 (2023).
99. Sato, T. *et al.* Human CD1c+ Myeloid Dendritic Cells Acquire a High Level of Retinoic Acid–Producing Capacity in Response to Vitamin D3. *The Journal of Immunology* **191**, 3152–3160 (2013).

100. Watchmaker, P. B. *et al.* Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. *Nat Immunol* **15**, 98–108 (2014).
101. Ginhoux, F., Guilliams, M. & Merad, M. Expanding dendritic cell nomenclature in the single-cell era. *Nat Rev Immunol* **22**, 67–68 (2022).
102. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* (1979) **356**, (2017).
103. Olweus, J. *et al.* Dendritic cell ontogeny: A human dendritic cell lineage of myeloid origin. *Proceedings of the National Academy of Sciences* **94**, 12551–12556 (1997).
104. Chicha, L., Jarrossay, D. & Manz, M. G. Clonal Type I Interferon–producing and Dendritic Cell Precursors Are Contained in Both Human Lymphoid and Myeloid Progenitor Populations. *J Exp Med* **200**, 1519–1524 (2004).
105. Feng, J. *et al.* Clonal lineage tracing reveals shared origin of conventional and plasmacytoid dendritic cells. *Immunity* **55**, 405–422.e11 (2022).
106. Arroyo Hornero, R. & Idoyaga, J. Plasmacytoid dendritic cells: A dendritic cell in disguise. *Mol Immunol* **159**, 38–45 (2023).
107. Zhou, B., Lawrence, T. & Liang, Y. The Role of Plasmacytoid Dendritic Cells in Cancers. *Front Immunol* **12**, (2021).
108. Briseño, C. G. *et al.* Distinct Transcriptional Programs Control Cross-Priming in Classical and Monocyte-Derived Dendritic Cells. *Cell Rep* **15**, 2462–2474 (2016).
109. Ko, H.-J. *et al.* GM-CSF–Responsive Monocyte-Derived Dendritic Cells Are Pivotal in Th17 Pathogenesis. *The Journal of Immunology* **192**, 2202–2209 (2014).
110. Sharma, M. D. *et al.* Activation of p53 in Immature Myeloid Precursor Cells Controls Differentiation into Ly6c+CD103+ Monocytic Antigen-Presenting Cells in Tumors. *Immunity* **48**, 91–106.e6 (2018).
111. Patente, T. A. *et al.* Human Dendritic Cells: Their Heterogeneity and Clinical Application Potential in Cancer Immunotherapy. *Front Immunol* **9**, (2019).

112. Chow, K. V, Sutherland, R. M., Zhan, Y. & Lew, A. M. Heterogeneity, functional specialization and differentiation of monocyte-derived dendritic cells. *Immunol Cell Biol* **95**, 244–251 (2017).
113. Bader, P. *et al.* Immune response Generated With the administration of autologous Dendritic Cells Pulsed With an allogenic Tumoral Cell-lines lysate in Patients With newly Diagnosed Diffuse intrinsic Pontine Glioma. *Front. Oncol* (2018) doi:10.3389/fonc.2018.00127.
114. Flórez-Grau, G. *et al.* Nanoencapsulated budesonide in self-stratified polyurethane-polyurea nanoparticles is highly effective in inducing human tolerogenic dendritic cells. *Int J Pharm* **511**, 785–793 (2016).
115. Wilson, N. S., El-Sukkari, D. & Villadangos, J. A. Dendritic cells constitutively present self antigens in their immature state in vivo and regulate antigen presentation by controlling the rates of MHC class II synthesis and endocytosis. *Blood* **103**, 2187–2195 (2004).
116. Valladeau, J. *et al.* Langerin, a Novel C-Type Lectin Specific to Langerhans Cells, Is an Endocytic Receptor that Induces the Formation of Birbeck Granules. *Immunity* **12**, 71–81 (2000).
117. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of Antigen Processing. *Annu Rev Immunol* **31**, 443–473 (2013).
118. Horton, R. *et al.* Gene map of the extended human MHC. *Nat Rev Genet* **5**, 889–899 (2004).
119. Pishesha, N., Harmand, T. J. & Ploegh, H. L. A guide to antigen processing and presentation. *Nat Rev Immunol* **22**, 751–764 (2022).
120. Zacharias, M. & Springer, S. Conformational Flexibility of the MHC Class I  $\alpha 1$ - $\alpha 2$  Domain in Peptide Bound and Free States: A Molecular Dynamics Simulation Study. *Biophys J* **87**, 2203–2214 (2004).
121. Chicz, R. M. *et al.* Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* **358**, 764–768 (1992).

122. Sette, A. & Sidney, J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* **50**, 201–212 (1999).
123. Tse, S. M. L. *et al.* Differential Role of Actin, Clathrin, and Dynamin in Fcγ Receptor-mediated Endocytosis and Phagocytosis. *Journal of Biological Chemistry* **278**, 3331–3338 (2003).
124. Lankar, D. *et al.* Dynamics of Major Histocompatibility Complex Class II Compartments during B Cell Receptor-mediated Cell Activation. *J Exp Med* **195**, 461–472 (2002).
125. Lim, J. P. & Gleeson, P. A. Macropinocytosis: an endocytic pathway for internalising large gulps. *Immunol Cell Biol* **89**, 836–843 (2011).
126. Delamarre, L., Pack, M., Chang, H., Mellman, I. & Trombetta, E. S. Differential Lysosomal Proteolysis in Antigen-Presenting Cells Determines Antigen Fate. *Science* (1979) **307**, 1630–1634 (2005).
127. West, L. C. & Cresswell, P. Expanding roles for GILT in immunity. *Curr Opin Immunol* **25**, 103–108 (2013).
128. Dolan, B. P. *et al.* Distinct Pathways Generate Peptides from Defective Ribosomal Products for CD8+ T Cell Immunosurveillance. *The Journal of Immunology* **186**, 2065–2072 (2011).
129. Reits, E. A. J., Vos, J. C., Grommé, M. & Neefjes, J. The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature* **404**, 774–778 (2000).
130. Vigneron, N., Ferrari, V., Stroobant, V., Abi Habib, J. & Van den Eynde, B. J. Peptide splicing by the proteasome. *Journal of Biological Chemistry* **292**, 21170–21179 (2017).
131. Bard, J. A. M. *et al.* Structure and Function of the 26S Proteasome. *Annu Rev Biochem* **87**, 697–724 (2018).
132. Suzuki, T., Huang, C. & Fujihira, H. The cytoplasmic peptide:N-glycanase (NGLY1) — Structure, expression and cellular functions. *Gene* **577**, 1–7 (2016).

133. Hattori, A. & Tsujimoto, M. Endoplasmic reticulum aminopeptidases: biochemistry, physiology and pathology. *J Biochem* **154**, 219–228 (2013).
134. Jensen, P. E. Recent advances in antigen processing and presentation. *Nat Immunol* **8**, 1041–1048 (2007).
135. Uebel, S. *et al.* Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proceedings of the National Academy of Sciences* **94**, 8976–8981 (1997).
136. Serwold, T., Gonzalez, F., Kim, J., Jacob, R. & Shastri, N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419**, 480–483 (2002).
137. Thomas, C. & Tampé, R. Structural and Mechanistic Principles of ABC Transporters. *Annu Rev Biochem* **89**, 605–636 (2020).
138. Reich, M. *et al.* Invariant chain processing is independent of cathepsin variation between primary human B cells/dendritic cells and B-lymphoblastoid cells. *Cell Immunol* **269**, 96–103 (2011).
139. Roche, P. A. & Cresswell, P. Invariant chain association with HLA-DR molecules inhibits immunogenic peptide binding. *Nature* **345**, 615–618 (1990).
140. Poluektov, Y. O., Kim, A. & Sadegh-Nasseri, S. HLA-DO and Its Role in MHC Class II Antigen Presentation. *Front Immunol* **4**, (2013).
141. Raghavan, M., Del Cid, N., Rizvi, S. M. & Peters, L. R. MHC class I assembly: out and about. *Trends Immunol* **29**, 436–443 (2008).
142. Joffre, O. P., Segura, E., Savina, A. & Amigorena, S. Cross-presentation by dendritic cells. *Nat Rev Immunol* **12**, 557–569 (2012).
143. Embgenbroich, M. & Burgdorf, S. Current Concepts of Antigen Cross-Presentation. *Front Immunol* **9**, (2018).
144. Saveanu, L. *et al.* IRAP Identifies an Endosomal Compartment Required for MHC Class I Cross-Presentation. *Science (1979)* **325**, 213–217 (2009).

145. Firat, E. *et al.* The Role of Endoplasmic Reticulum-Associated Aminopeptidase 1 in Immunity to Infection and in Cross-Presentation. *The Journal of Immunology* **178**, 2241–2248 (2007).
146. Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol* **10**, 787–96 (2010).
147. Eisenbarth, S. C. Dendritic cell subsets in T cell programming: location dictates function. *Nat Rev Immunol* **19**, 89–103 (2019).
148. Gurevich, I. *et al.* Active dissemination of cellular antigens by DCs facilitates CD8+ T-cell priming in lymph nodes. *Eur J Immunol* **47**, 1802–1818 (2017).
149. Braun, A. *et al.* Afferent lymph-derived T cells and DCs use different chemokine receptor CCR7-dependent routes for entry into the lymph node and intranodal migration. *Nat Immunol* **12**, 879–87 (2011).
150. Tomura, M. *et al.* Tracking and quantification of dendritic cell migration and antigen trafficking between the skin and lymph nodes. *Sci Rep* **4**, 6030 (2014).
151. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T Cell Development, Localization, and Function throughout Life. *Immunity* **48**, 202–213 (2018).
152. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–61 (2003).
153. Stritesky, G. L., Jameson, S. C. & Hogquist, K. A. Selection of self-reactive T cells in the thymus. *Annu Rev Immunol* **30**, 95–114 (2012).
154. Wculek, S. K. *et al.* Dendritic cells in cancer immunology and immunotherapy. *Nat Rev Immunol* **20**, 7–24 (2020).
155. Lotze, M. T., Olejniczak, S. H. & Skokos, D. CD28 co-stimulation: novel insights and applications in cancer immunotherapy. *Nat Rev Immunol* **24**, 878–895 (2024).
156. Sun, L., Su, Y., Jiao, A., Wang, X. & Zhang, B. T cells in health and disease. *Signal Transduct Target Ther* **8**, 235 (2023).
157. Poloni, C. *et al.* T-cell activation-induced marker assays in health and disease. *Immunol Cell Biol* **101**, 491–503 (2023).

158. Shatrova, A. N. *et al.* Time-Dependent Regulation of IL-2R  $\alpha$ -Chain (CD25) Expression by TCR Signal Strength and IL-2-Induced STAT5 Signaling in Activated Human Blood T Lymphocytes. *PLoS One* **11**, e0167215 (2016).
159. Wu, H., Witzl, A. & Ueno, H. Assessment of TCR signal strength of antigen-specific memory CD8<sup>+</sup> T cells in human blood. *Blood Adv* **3**, 2153–2163 (2019).
160. Stark, R., Hartung, A., Zehn, D., Frentsch, M. & Thiel, A. IL-12-mediated STAT4 signaling and TCR signal strength cooperate in the induction of CD40L in human and mouse CD8<sup>+</sup> T cells. *Eur J Immunol* **43**, 1511–1517 (2013).
161. Takeda, I. *et al.* Distinct Roles for the OX40-OX40 Ligand Interaction in Regulatory and Nonregulatory T Cells. *The Journal of Immunology* **172**, 3580–3589 (2004).
162. Wölfl, M., Kuball, J., Eyrich, M., Schlegel, P. G. & Greenberg, P. D. Use of CD137 to study the full repertoire of CD8<sup>+</sup> T cells without the need to know epitope specificities. *Cytometry Part A* **73A**, 1043–1049 (2008).
163. Han, Y., Liu, D. & Li, L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res* **10**, 727–742 (2020).
164. Wucherpfennig, K. W., Gagnon, E., Call, M. J., Huseby, E. S. & Call, M. E. Structural Biology of the T-cell Receptor: Insights into Receptor Assembly, Ligand Recognition, and Initiation of Signaling. *Cold Spring Harb Perspect Biol* **2**, a005140–a005140 (2010).
165. Brameshuber, M. *et al.* Monomeric TCRs drive T cell antigen recognition. *Nat Immunol* **19**, 487–496 (2018).
166. Mariuzza, R. A., Agnihotri, P. & Orban, J. The structural basis of T-cell receptor (TCR) activation: An enduring enigma. *J Biol Chem* **295**, 914–925 (2020).
167. Rossjohn, J. *et al.* T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules. *Annu Rev Immunol* **33**, 169–200 (2015).
168. Samelson, L. E. Signal Transduction Mediated by the T Cell Antigen Receptor: The Role of Adapter Proteins. *Annu Rev Immunol* **20**, 371–394 (2002).

169. Malissen, B. & Bongrand, P. Early T Cell Activation: Integrating Biochemical, Structural, and Biophysical Cues. *Annu Rev Immunol* **33**, 539–561 (2015).
170. Chang, V. T. *et al.* Initiation of T cell signaling by CD45 segregation at ‘close contacts’. *Nat Immunol* **17**, 574–582 (2016).
171. Brazin, K. N. *et al.* Structural Features of the  $\alpha\beta$ TCR Mechanotransduction Apparatus That Promote pMHC Discrimination. *Front Immunol* **6**, (2015).
172. Rangarajan, S. *et al.* Peptide–MHC (pMHC) binding to a human antiviral T cell receptor induces long-range allosteric communication between pMHC- and CD3-binding sites. *Journal of Biological Chemistry* **293**, 15991–16005 (2018).
173. ZHANG, W. & LIU, H. T. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res* **12**, 9–18 (2002).
174. Yin, X., Chen, S. & Eisenbarth, S. C. Dendritic Cell Regulation of T Helper Cells. *Annu Rev Immunol* **39**, 759–790 (2021).
175. Afkarian, M. *et al.* T-bet is a STAT1-induced regulator of IL-12R expression in naïve CD4<sup>+</sup> T cells. *Nat Immunol* **3**, 549–557 (2002).
176. Walker, J. A. & McKenzie, A. N. J. TH2 cell development and function. *Nat Rev Immunol* **18**, 121–133 (2018).
177. Angkasekwinai, P. & Dong, C. IL-9-producing T cells: potential players in allergy and cancer. *Nat Rev Immunol* **21**, 37–48 (2021).
178. Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 Cells. *Annu Rev Immunol* **27**, 485–517 (2009).
179. Hori, S. FOXP3 as a master regulator of Treg cells. *Nat Rev Immunol* **21**, 618–619 (2021).
180. Shedlock, D. J. & Shen, H. Requirement for CD4 T Cell Help in Generating Functional CD8 T Cell Memory. *Science (1979)* **300**, 337–339 (2003).
181. Zander, R. *et al.* CD4<sup>+</sup> T Cell Help Is Required for the Formation of a Cytolytic CD8<sup>+</sup> T Cell Subset that Protects against Chronic Infection and Cancer. *Immunity* **51**, 1028–1042.e4 (2019).

182. Snell, L. M. *et al.* CD8+ T Cell Priming in Established Chronic Viral Infection Preferentially Directs Differentiation of Memory-like Cells for Sustained Immunity. *Immunity* **49**, 678-694.e5 (2018).
183. Chang, J. T., Wherry, E. J. & Goldrath, A. W. Molecular regulation of effector and memory T cell differentiation. *Nat Immunol* **15**, 1104–1115 (2014).
184. Mousset, C. M. *et al.* Comprehensive Phenotyping of T Cells Using Flow Cytometry. *Cytometry Part A* **95**, 647–654 (2019).
185. Saxena, A., Dagur, P. K. & Biancotto, A. Multiparametric Flow Cytometry Analysis of Naïve, Memory, and Effector T Cells. in 129–140 (2019). doi:10.1007/978-1-4939-9650-6\_8.
186. Koh, C.-H., Lee, S., Kwak, M., Kim, B.-S. & Chung, Y. CD8 T-cell subsets: heterogeneity, functions, and therapeutic potential. *Exp Mol Med* **55**, 2287–2299 (2023).
187. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* **12**, 749–761 (2012).
188. Nolz, J. C. & Harty, J. T. Protective Capacity of Memory CD8+ T Cells Is Dictated by Antigen Exposure History and Nature of the Infection. *Immunity* **34**, 781–793 (2011).
189. Künzli, M. & Masopust, D. CD4+ T cell memory. *Nat Immunol* **24**, 903–914 (2023).
190. Preglej, T. & Ellmeier, W. CD4+ Cytotoxic T cells – Phenotype, Function and Transcriptional Networks Controlling Their Differentiation Pathways. *Immunol Lett* **247**, 27–42 (2022).
191. Chang, K. *et al.* Immune Profiling of Premalignant Lesions in Patients With Lynch Syndrome. *JAMA Oncol* **4**, 1085 (2018).
192. Kloor, M. *et al.* Prevalence of mismatch repair-deficient crypt foci in Lynch syndrome: a pathological study. *Lancet Oncol* **13**, 598–606 (2012).

193. Bohaumilitzky, L. *et al.* The Different Immune Profiles of Normal Colonic Mucosa in Cancer-Free Lynch Syndrome Carriers and Lynch Syndrome Colorectal Cancer Patients. *Gastroenterology* **162**, 907-919.e10 (2022).
194. Echterdiek, F. *et al.* Low density of FOXP3-positive T cells in normal colonic mucosa is related to the presence of beta2-microglobulin mutations in Lynch syndrome-associated colorectal cancer. *Oncoimmunology* **5**, e1075692 (2016).
195. Walkowska, J. *et al.* Immunoprofiles of colorectal cancer from Lynch syndrome. *Oncoimmunology* **8**, (2019).
196. Wojtowicz, M. E., Dunn, B. K. & Umar, A. Immunologic approaches to cancer prevention - Current status, challenges, and future perspectives. *Semin Oncol* **43**, 161–172 (2016).
197. Ribatti, D. The concept of immune surveillance against tumors: The first theories. *Oncotarget* **8**, 7175–7180 (2017).
198. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* (2019) doi:10.1038/s41586-019-1671-8.
199. Finn, O. J. The dawn of vaccines for cancer prevention. *Nat Rev Immunol* **18**, 183–194 (2018).
200. Bayó, C., Jung, G., Español-Rego, M., Balaguer, F. & Benitez-Ribas, D. Vaccines for Non-Viral Cancer Prevention. *Int J Mol Sci* **22**, 10900 (2021).
201. Sabado, R. L., Balan, S. & Bhardwaj, N. Dendritic cell-based immunotherapy. *Cell Res* **27**, 74–95 (2017).
202. Kantoff, P. W. *et al.* Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *New England Journal of Medicine* **363**, 411–422 (2010).
203. Gao, T., Cen, Q. & Lei, H. A review on development of MUC1-based cancer vaccine. *Biomedicine and Pharmacotherapy* **132**, 110888 (2020).
204. Arab, A., Yazdian-Robati, R. & Behravan, J. HER2-Positive Breast Cancer Immunotherapy: A Focus on Vaccine Development. *Arch Immunol Ther Exp (Warsz)* **68**, 2 (2020).

205. Kimura, T. *et al.* MUC1 Vaccine for Individuals with Advanced Adenoma of the Colon: A Cancer Immunoprevention Feasibility Study. *Cancer Prevention Research* **6**, 18–26 (2013).
206. Mittendorf, E. A. *et al.* Efficacy and safety analysis of nelipepimut-S vaccine to prevent breast cancer recurrence: A randomized, multicenter, phase III clinical trial. *Clinical Cancer Research* **25**, 4248–4254 (2019).
207. McCarthy, P. M. *et al.* AE37: a HER2-targeted vaccine for the prevention of breast cancer recurrence. *Expert Opin Investig Drugs* **30**, 5–11 (2021).
208. Brown, T. A. *et al.* Prospective, randomized, single-blinded, multi-center phase II trial of two HER2 peptide vaccines, GP2 and AE37, in breast cancer patients to prevent recurrence. *Breast Cancer Res Treat* **181**, 391–401 (2020).
209. Raza, A. *et al.* Unleashing the immune response to NY-ESO-1 cancer testis antigen as a potential target for cancer immunotherapy. *J Transl Med* **18**, 140 (2020).
210. Xu, Y., Zou, R., Wang, J., Wang, Z. & Zhu, X. The role of the cancer testis antigen PRAME in tumorigenesis and immunotherapy in human cancer. *Cell Prolif* **53**, (2020).
211. Vansteenkiste, J. F. *et al.* Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* **17**, 822–835 (2016).
212. Dreno, B. *et al.* MAGE-A3 immunotherapeutic as adjuvant therapy for patients with resected, MAGE-A3-positive, stage III melanoma (DERMA): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol* **19**, 916–929 (2018).
213. Krishnadas, D. K. *et al.* A phase I trial combining decitabine/dendritic cell vaccine targeting MAGE-A1, MAGE-A3 and NY-ESO-1 for children with relapsed or therapy-refractory neuroblastoma and sarcoma. *Cancer Immunology, Immunotherapy* **64**, 1251–1260 (2015).

214. Chen, X. *et al.* Dual TGF- $\beta$  and PD-1 blockade synergistically enhances MAGE-A3-specific CD8 + T cell response in esophageal squamous cell carcinoma. *Int J Cancer* **143**, 2561–2574 (2018).
215. Hu, Z. *et al.* Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat Med* **27**, 515–525 (2021).
216. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
217. Keskin, D. B. *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234–239 (2019).
218. Gross, S. *et al.* Twelve-year survival and immune correlates in dendritic cell-vaccinated melanoma patients. *JCI Insight* **2**, (2017).
219. Richters, M. M. *et al.* Best practices for bioinformatic characterization of neoantigens for clinical utility. *Genome Med* **11**, 1–21 (2019).
220. Litchfield, K. *et al.* Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun* **11**, 1–11 (2020).
221. Blass, E. & Ott, P. A. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat Rev Clin Oncol* (2021) doi:10.1038/s41571-020-00460-2.
222. Roudko, V., Greenbaum, B. & Bhardwaj, N. Computational Prediction and Validation of Tumor-Associated Neoantigens. *Front Immunol* **11**, 1–11 (2020).
223. Woerner, S. M. *et al.* SelTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. *Nucleic Acids Res* **38**, D682–D689 (2010).
224. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat Rev Cancer* **20**, 555–572 (2020).
225. Dietlein, F. *et al.* Identification of cancer driver genes based on nucleotide context. *Nat Genet* **52**, 208–218 (2020).

226. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174**, 1034–1035 (2018).
227. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* **8**, 15180 (2017).
228. Kautto, E. A. *et al.* Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* **8**, 7452–7463 (2017).
229. van der Meer, D. *et al.* Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res* **47**, D923–D929 (2019).
230. Hundal, J. *et al.* pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol Res* **8**, 409–420 (2020).
231. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* (2019) doi:10.1093/nar/gkz1029.
232. Hu, Z., Yau, C. & Ahmed, A. A. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat Commun* **8**, 15943 (2017).
233. Canet-Hermida, J. *et al.* Highly Sensitive Microsatellite Instability and Immunohistochemistry Assessment in Endometrial Aspirates as a Tool for Cancer Risk Individualization in Lynch Syndrome. *Mod Pathol* **36**, 100158 (2023).
234. González-Acosta, M. *et al.* High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers. *J Med Genet* **57**, 269–273 (2020).
235. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10 (2011).
236. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

237. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
238. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219 (2013).
239. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
240. Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).
241. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
242. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
243. Carreno, B. M. *et al.* Amino-Terminal Extended Peptide Single-Chain Trimers Are Potent Synthetic Agonists for Memory Human CD8<sup>+</sup> T Cells. *The Journal of Immunology* **188**, 5839–5849 (2012).
244. Bear, A. S. *et al.* Biochemical and functional characterization of mutant KRAS epitopes validates this oncoprotein for immunological targeting. *Nat Commun* **12**, 4365 (2021).
245. Fernandez, N. F. *et al.* Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. *Sci Data* **4**, 170151 (2017).
246. Bayó, C. *et al.* Discovery and validation of frameshift-derived neopeptides in Lynch syndrome: paving the way for novel cancer prevention strategies. *J Immunother Cancer* **13**, e011177 (2025).
247. Jaffrey, S. R. & Wilkinson, M. F. Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. *Nat Rev Neurosci* **19**, 715–728 (2018).

248. Binder, H. *et al.* Genomic and transcriptomic heterogeneity of colorectal tumours arising in Lynch syndrome. *J Pathol* **243**, 242–254 (2017).
249. Strønen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science (1979)* **352**, 1337–1341 (2016).
250. Kalaora, S. *et al.* Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* **8**, 1366–1375 (2018).
251. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404 (2016).
252. Bui, H.-H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* **7**, 153 (2006).
253. Jameson, S. C. & Masopust, D. Understanding Subset Diversity in T Cell Memory. *Immunity* **48**, 214–226 (2018).
254. Mousset, C. M. *et al.* Comprehensive Phenotyping of T Cells Using Flow Cytometry. *Cytometry Part A* **95**, 647–654 (2019).
255. Brand, R. M. *et al.* Immune microenvironment profiling of normal appearing colorectal mucosa biopsied over repeat patient visits reproducibly separates lynch syndrome patients based on their history of colon cancer. *Front Oncol* **13**, (2023).
256. Ingels, J. *et al.* Neoantigen-targeted dendritic cell vaccination in lung cancer patients induces long-lived T cells exhibiting the full differentiation spectrum. *Cell Rep Med* **5**, 101516 (2024).
257. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
258. Carreno, B. M. *et al.* A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science (1979)* **348**, 803–808 (2015).

259. Kloor, M. *et al.* Vaccination of MSI-H colorectal cancer patients with frameshift peptide antigens: A phase I/IIa clinical trial. *Journal of Clinical Oncology* **33**, 3020–3020 (2015).
260. Bassani-Sternberg, M. & Harari, A. Immune pressure sculps tumor cells and trims high-quality mutations. *Cancer Cell* **40**, 717–719 (2022).
261. Richman, L. P., Vonderheide, R. H. & Rech, A. J. Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. *Cell Syst* **9**, 375–382.e4 (2019).
262. Dersh, D., Hollý, J. & Yewdell, J. W. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat Rev Immunol* **21**, 116–128 (2021).
263. Bolivar, A. M. *et al.* Genomic Landscape of Lynch Syndrome Colorectal Neoplasia Identifies Shared Mutated Neoantigens for Immunoprevention. *Gastroenterology* **166**, 787–801.e11 (2024).
264. Mestrallet, G., Brown, M., Bozkus, C. C. & Bhardwaj, N. Immune escape and resistance to immunotherapy in mismatch repair deficient tumors. *Front Immunol* **14**, (2023).
265. Dolton, G. *et al.* Targeting of multiple tumor-associated antigens by individual T cell receptors during successful cancer immunotherapy. *Cell* **186**, 3333–3349.e27 (2023).
266. Hay, Z. L. Z., Knapp, J. R., Magallon, R. E., O'Connor, B. P. & Slansky, J. E. Low TCR Binding Strength Results in Increased Progenitor-like CD8<sup>+</sup> Tumor-Infiltrating Lymphocytes. *Cancer Immunol Res* **11**, 570–582 (2023).
267. Martinez, R. J. & Evavold, B. D. Lower Affinity T Cells are Critical Components and Active Participants of the Immune Response. *Front Immunol* **6**, (2015).
268. Bonadona, V. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* **305**, 2304 (2011).
269. Helderman, N. c. *et al.* Molecular Profile of MSH6-Associated Colorectal Carcinomas Shows Distinct Features From Other Lynch Syndrome–Associated Colorectal Carcinomas. *Gastroenterology* **165**, 271–274.e2 (2023).

270. Gehad, A. *et al.* A primary role for human central memory cells in tissue immunosurveillance. *Blood Adv* **2**, 292–298 (2018).
271. Gray, J. I., Westerhof, L. M. & MacLeod, M. K. L. The roles of resident, central and effector memory <scp>CD</scp> 4 T-cells in protective immunity following infection or vaccination. *Immunology* **154**, 574–581 (2018).
272. Reiser, J. & Banerjee, A. Effector, Memory, and Dysfunctional CD8 + T Cell Fates in the Antitumor Immune Response. *J Immunol Res* **2016**, 1–14 (2016).
273. Kloor, M. *et al.* Prevalence of mismatch repair-deficient crypt foci in Lynch syndrome: a pathological study. *Lancet Oncol* **13**, 598–606 (2012).
274. Staffa, L. *et al.* Mismatch Repair-Deficient Crypt Foci in Lynch Syndrome – Molecular Alterations and Association with Clinical Parameters. *PLoS One* **10**, e0121980 (2015).
275. Cheroutre, H. & Lambolez, F. Doubting the TCR Coreceptor Function of CD8 $\alpha\alpha$ . *Immunity* **28**, 149–159 (2008).
276. Liu, J. *et al.* The peripheral differentiation of human natural killer T cells. *Immunol Cell Biol* **97**, 586–596 (2019).
277. Zhang, X. *et al.* Neoantigen DNA vaccines are safe, feasible, and induce neoantigen-specific immune responses in triple-negative breast cancer patients. *Genome Med* **16**, 131 (2024).
278. Linnemann, C. *et al.* High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med* **21**, 81–85 (2015).
279. de Graaf, J. F. *et al.* Neoantigen-specific T cell help outperforms non-specific help in multi-antigen DNA vaccination against cancer. *Molecular Therapy: Oncology* **32**, 200835 (2024).
280. Garzia, I. *et al.* Tumor Burden Dictates the Neoantigen Features Required to Generate an Effective Cancer Vaccine. *Cancer Immunol Res* **12**, 440–452 (2024).

281. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, (2016).
282. Leng, Z. *et al.* Lgr5+CD44+EpCAM+ Strictly Defines Cancer Stem Cells in Human Colorectal Cancer. *Cellular Physiology and Biochemistry* **46**, 860–872 (2018).
283. Leoni, G. *et al.* A Genetic Vaccine Encoding Shared Cancer Neoantigens to Treat Tumors with Microsatellite Instability. *Cancer Res* **80**, 3972–3982 (2020).
284. Bonelli, J. *et al.* Novel Tumor-Targeted Self-Nanostructured and Compartmentalized Water-in-Oil-in-Water Polyurethane-Polyurea Nanocapsules for Cancer Theragnosis. *Pharmaceutics* **15**, 58 (2022).



## 8 SUPPLEMENTARY MATERIAL

Supplementary table 1. Selective Targets in Human MSI-High Tumourigenesis Database (Seltarbase, [www.seltarbase.org](http://www.seltarbase.org)) and two additional sources, used as input for FSDN prediction.

| Gene ID   | cMS (aa) | m1 peptide sequence (+ 8 N-terminal wt aa) | m1 (aa) | m2 peptide sequence (+ 8 N-terminal wt aa) | m2 (aa) |
|-----------|----------|--------------------------------------------|---------|--------------------------------------------|---------|
| ABCC2     | A8       | LEDVEKKKRLGPKKMFQNPQ                       | 13      | LEDVEKKKEVWDQKRCISKILVDEGSVQN              | 35      |
| ABCC5     | T9       | DDNQAPFFCLRVRCGGWLCGWTSASPS                | 25      | DDNQAPFFVYVCDVAGCAAGPHQHRPH                | 61      |
| ABCF1     | A10      | QQQQQQKKSEIPEKAGGRMMWMMMEKRK               | 59      | QQQQQQKKARYPKRQAEEGCG                      | 13      |
| ACVR2A    | A8       | QEVVVHKKRGLF                               | 4       | QEVVVHKKACFKRLLAETCWNGNAL                  | 18      |
| ADNP      | A8       | SSKPAAKKRLPCKVTESS                         | 10      | SSKPAAKKGYHAR                              | 5       |
| AIM2      | A10      | KVIKAKKKHREVKRTNSSLV                       | 13      | IKVIKAKKNIEK                               | 4       |
| AKAP9     | A8       | SPSKKQKKREKRQAVNMMCQHTMI                   | 16      | SPSKKQKKKEKNVKQ                            | 6       |
| ANGPT2    | A9       | IIEELEKK                                   | 0       | SIIEELEKNSDCHGE                            | 7       |
| ANKRD49   | A9       | WYRLQEKKKWKTQADCFGLKKIGLPQ                 | 31      | EWYRLQEKNGKRPKQIASLGC                      | 13      |
| APAF1     | A8       | YLEWINKKTSRIFPA                            | 7       | YLEWINKKHESFPLSCPPPHRCCLPCL                | 22      |
| ASH1L     | A9       | NVSEGEKKQRKVVKNPSQPVPLRNDGIT               | 43      | ANVSEGEKNRGK                               | 4       |
| ASTE1     | A11      | GRSNSKKKGRRNRIPAVLRTEGEPLHTP               | 32      | GRSNSKKKAEETEQYLF                          | 9       |
| ASXL1     | G8       | ATTAIGGGVARVEVAAGPPMREVAEAAA               | 57      | ATTAIGGGWPGWRWRRGHR                        | 11      |
| ATG2B     | A9       | NYRSRRKKKN                                 | 1       | NYRSRRKKIRLSEQELSEFSLSSSEY                 | 18      |
| ATG3      | T10      | KWLLTIF                                    | 0       | KWLLTIFFEKFSVTLMIQ                         | 10      |
| ATR       | A10      | FLFLKKKKYLQ                                | 4       | PFLFLKKKNT                                 | 2       |
| B3GLCT    | A8       | KRAEQLKKAS                                 | 2       | KRAEQLKKHLKAGCRSYTGAPQCPPSS                | 22      |
| B4GALNT1  | G8       | CSCSSGGGASPSRKNSELTSRPLT                   | 23      | CSCSSGGGPPPLPETSPSY                        | 12      |
| BAG6      | C8       | TGNGTRPPQLPMQRHLPVLGRPHWLR                 | 60      | TGNGTRPPNSQCRGTPSPWSWAGLIRGSV              | 28      |
| BANP      | T12      | CFFFFPFFCSVGA                              | 5       | EALPCFFFSLSFAL                             | 6       |
| BAX [BAX] | G8       | DRAGRMGGRHPSWPWTRCLMRPPRS                  | 18      | DRAGRMGGGTRAGPGGASGCVHQEAER                | 31      |
| BCL10     | A8       | FDHLRAKKYSVEKTLKKFLVEHQVEKGL               | 23      | HFDHLRAKNTQ                                | 3       |
| BEND5     | A9       | VTGVATKKRKMQLNHP SRLTN                     | 14      | VTGVATKKERCSP                              | 5       |
| BLM       | A9       | ETPRLGKKMKALISQEMFSQALL                    | 15      | ETPRLGKK                                   | 0       |
| BRCA1     | A8       | SSEEIKKKSTTKCQSGTAETYSWKVKV                | 46      | SSEEIKKKVQPNASQAQKPTTHGR                   | 17      |
| C15orf48  | A9       | DVILDRKKIQNLGKLWTLTYLKL                    | 16      | DVILDRKKSRTLGNCGPYCTSKAYNNQP               | 25      |
| CAAP1     | A9       | LELLSEKKF                                  | 1       | QLELLSEKNFEDS                              | 5       |
| CAMKK2    | A13      | LKGTKKKKDLTP                               | 4       | ELKGTKKKRT                                 | 2       |
| CANX      | T8       | WSMTSDIFLTLSFVLIEE                         | 11      | WSMTSDIF                                   | 0       |
| CASD1     | A8       | WPQIIKKQTEIVSGILAYC                        | 12      | LWPQIIQKSKRKLFLAFWLTVETRLFAV               | 32      |
| CASP1     | A8       | GYSVDVKKISLLRT                             | 6       | GYSVDVKKSHCFGHYRAGGICTPPRAQ                | 22      |
| CASP5     | A10      | VKKDNHKKKQLRCWNTWAKMFFMVFLII               | 26      | VKKDNHKKNS                                 | 2       |
| CASP5     | A8       | EFLRLCKKIMMRSIQ                            | 7       | EFLRLCKKS                                  | 1       |
| CCDC168   | A9       | KEERNQKK                                   | 0       | QKEERNQKNNRYERHRYNFEVKEVTSFL               | 39      |
| C18orf34  | A10      | LKNHYKKKWR                                 | 2       | DLKNHYKKNGNGPTQKS                          | 9       |
| CCDC28A   | G8       | TQVVNAKKMPFQ                               | 4       | TQVVNAKKCHSSE                              | 5       |
| CCDC43    | A9       | ERKEKEKKGHREGSESDNLGLLYSFP                 | 19      | ERKEKEKKDTERGAKAITLAYSILSPHE               | 59      |
| CCKBR     | T8       | LVIVVLFVCGCQFIVPTRGAPLMARVH                | 34      | LLVIVVLFVSLVASL                            | 7       |

|         |     |                              |    |                              |     |
|---------|-----|------------------------------|----|------------------------------|-----|
| CDC25C  | A8  | TIPDKVKKSIFLAKESSGRAYV       | 14 | TIPDKVKKVFFWPRKAQEGLMFKEDSLS | 21  |
| CDC7    | A9  | QAEGSLKKTSRILNLQVLKKILRSFMKL | 50 | QAEGSLKKRAEF                 | 4   |
| CDKL2   | A9  | LLEVCKKKNDGT                 | 4  | LLEVCKKKTMVPSL               | 6   |
| CEBPZ   | A9  | FRTCVMKKMLNQKCLAPF           | 10 | FFRTCVMKKRC                  | 2   |
| CENPF   | A8  | ELDTMSKKPRHWISCLKK           | 10 | GELDTMSKNHGTGSVV             | 8   |
| CENPH   | A8  | RLSTALKKTWRKLADSLVCSWIT      | 15 | RLSTALKKPGEN                 | 4   |
| CEP164  | A11 | AIKKKKKKRKRKTRRTETPPKVRWPWVP | 21 | AIKKKKKKGKERQEQRPQKFAGLGFL   | 87  |
| CEP290  | A10 | ERLDLKKKFVKWLKKEEKEVQLQD     | 16 | EERLDLKKNSSNGSRKRKKCNFRINH   | 19  |
| CEP57   | A8  | PNARRIKKRSQNHQKRKVLGTILVHNHI | 30 | PNARRIKKEVKTTTRKEKF          | 10  |
| CHD2    | A10 | PVDPEEKNTA                   | 3  | PVDPEEKILLNLSCQSCTF          | 12  |
| CHD2    | A9  | LEKSPMKKNRRRKRTRRTRRNK       | 14 | LEKSPMKKTEEREQGEQGETNEF      | 16  |
| CHD2    | A8  | ESEPEQKK                     | 0  | AESEPEQKSKSQKTCPOK           | 10  |
| CHD8    | A8  | QKWQRCKKIIQRN                | 5  | QKWQRCKK                     | 0   |
| CHEK1   | A9  | YSDWKEKKHTSTLGKKSILL         | 13 | EYSDWKEKNIPQPLEKNRFCSSSSAA   | 18  |
| CIR1    | A9  | ESNNKEKKYKGRKERKTSVQGITTILK  | 50 | SESNNKEKNTKEEKKQVFRA         | 14  |
| CLOCK   | T9  | LEALDGFF                     | 0  | LEALDGFFSNHDRWKHNICV         | 12  |
| COBLL1  | T9  | QAKPSSFFRCRCRREYRVTM         | 11 | QAKPSSFFADAEEISGLCDICSCQCEP  | 23  |
| COIL    | A8  | TDQTVSKKTRKIKQPAQWVMITKRPK   | 46 | TDQTVSKKQEKK                 | 4   |
| CTSC    | A8  | VMGPQEKK                     | 0  | SVMGPQEKSSGVPSEAGYSI         | 12  |
| CUL2    | A8  | NIIRQEKKMTWQICTSYSVLCPLVYLI  | 19 | NIIRQEKK                     | 0   |
| CYHR1   | G9  | NTYGMWGGRCVWVQAS             | 8  | NTYGMWGGDAGCGRPEPAGGGQAAAAT  | 142 |
| CYSLTR1 | A8  | LLKKSMMKICQVIKRL             | 8  | LLKKSMMKKS VKS               | 4   |
| DDX39B  | T8  | PNVKVAVFLVCLSRMRKRC          | 12 | PNVKVAVFWWSVYQEG             | 8   |
| DIAPH3  | A9  | TSAGEEKKLTDFVLWKASGTIQFNCK   | 19 | LTSAGEEKN                    | 1   |
| DNAH2   | T10 | LSSFYFFFLLLFLSLPTSHPAINILIQ  | 30 | LSSFYFFFFFYFCLSPPTLLSTF      | 16  |
| DPAGT1  | T8  | LYFMIPFFSPLWDCSTTTGTHHGCLWEI | 27 | LYFMIPFFHHFGIALPQLVPITGVCGRY | 115 |
| DSTN    | T8  | RKEELMFFCGHQ                 | 5  | RKEELMFFVGTSTSEK             | 10  |
| EAPP    | A8  | AVQVTKKRRNNTFRQMQMTNYCMILKKI | 54 | AVQVTKKKEETTQDSNK            | 9   |
| EBPL    | T9  | YCWLYLFFLTVCGF               | 6  | YCWLYLFF                     | 0   |
| ELAVL3  | G9  | AMESQVGGARPAPCPTGHSVLQMEPLT  | 34 | AMESQVGGPRGPAPQATPWYKWSH     | 18  |
| ELOA    | A8  | LGDKGLKMTLALVKTWTQFRNYPR     | 18 | LGDKGLKK                     | 0   |
| EPHB2   | A9  | CNSNDGKKREWEKRQILGGGKYKEYF   | 62 | CNSNDGKKGNGKKENRSWEGAGNTRNIF | 20  |
| ERCC5   | A9  | PHDTKVKKNYGHCSPLAFLTQLPRPT   | 64 | PHDTKVKKITDIATHPWLS          | 11  |
| ESRP1   | A8  | AAQKCHKKT                    | 1  | AAQKCHKKHEGQIC               | 6   |
| F8      | A8  | HFLQGAKKITFL                 | 4  | HFLQGAKK                     | 0   |
| FAM111B | A10 | EMDISKKKHYNRKISIKLNRMKVPLMK  | 25 | LEMDISKKSITTERYP             | 8   |
| FBXL3   | T9  | VNLVMYFFYMKNLTPSFAMKYLPICT   | 24 | VNLVMYFFI                    | 1   |
| FERMT2  | A8  | PRDPTKKKRRS                  | 3  | PRDPTKKKEAR                  | 4   |
| FGFBP1  | A9  | LLVEGKKK                     | 0  | VLLVEGKKSEEWTSQSGRLTKHSGQH   | 22  |
| FILIP1L | A8  | GDHSLVLQKN                   | 1  | GDHSLVLQKTKSTRKQEQFRFRKD     | 15  |
| FLCN    | C8  | SPHVQIPPTCSPQSLSSWRSTQPHVPP  | 38 | SPHVQIPPRALLRVCHRGGRSRTFHP   | 25  |
| FLT3LG  | C9  | KCAFQPPPAVFASSRPTSPASCRRPPSS | 23 | TKCAFQPPQLSSLRPDQHLPPPAGDLRA | 188 |
| FXR1    | A8  | NQNKTVKKIPSLNAQWMFLRI        | 13 | NQNKTVKKYLL                  | 3   |
| G2E3    | A10 | LGFQIKKKLKDCISTKPISGIVP      | 15 | ELGFQIKKN                    | 1   |
| GART    | A10 | NHFSFEKKRPEWLS               | 6  | NHFSFEKKGGSGCLNIWNRIEPASTYRQ | 25  |

|           |     |                               |    |                                |     |
|-----------|-----|-------------------------------|----|--------------------------------|-----|
| GBP3      | A10 | LQKTLKKKPRDICRIS              | 8  | KLQKTLKKNQEIYVA                | 7   |
| GCC2      | A8  | NNKLSSEKNS                    | 2  | NNKLSSEKTVE                    | 3   |
| GIN51     | A9  | GTSVLLKKIASTFYLDGNVSS         | 13 | GTSVLLKK                       | 0   |
| GLYR1     | G8  | QVIVSRGGAFWKPPSQGISSCLMTGCW   | 19 | AQVIVSRGALSGSPRLRESAAV         | 14  |
| GLYR1     | G8  | QVIVSRGGAFWKPPSQGISSCLMTGCW   | 19 | AQVIVSRGALSGSPRLRESAAV         | 14  |
| GOLGA4    | A8  | KEELKEKKRLAVLVNN              | 9  | MKEELKEKS                      | 1   |
| GPATCH4   | A8  | SESKPPKKRRKRGGRKKRKLQHLKGMM   | 72 | SESKPPKKEKEKEAERGGRSYSI        | 15  |
| GRB14     | A9  | YVSLAGKKNMEHRLTMSALSCLKREGP   | 22 | YVSLAGKKTWSTD                  | 5   |
| GRK4      | A9  | ACKKLQKKE                     | 1  | ACKKLQKKNKEEER                 | 6   |
| HDAC2     | A9  | SQGGGKKKSATTTTTVILEIIIMDRVIP  | 19 | YSQGGGKSLLLLR                  | 7   |
| HMGXB4    | A10 | TDGHSEKKRKK                   | 3  | TDGHSEKKEKR                    | 3   |
| HMMR      | A9  | LRCQLAKKNKVRQNFKRN            | 10 | LRCQLAKKTK                     | 2   |
| HNFI1A    | C8  | MDTYSGPPQGQARDLRCPLTAPLACLHL  | 50 | MDTYSGPPRARPGTCAARSQLPWPASTC   | 24  |
| HOXA11    | A9  | NRRMKEKKLTETVYSTTQIHSSKTPAA   | 29 | QNRRMKEKN                      | 1   |
| HOXB7     | G8  | QGLYPGGGAWRARARAPASTRPAMGSSRV | 98 | QGLYPGGGGHGGPERGRRRLRGRLWARAEF | 28  |
| HPS1      | C8  | WLEGGTPPWMPFR                 | 5  | IWLEGGTPHGCPSDSRGHPNPTGSPLPC   | 127 |
| ICA1      | A9  | TQVRLAKKTLTN                  | 4  | TQVRLAKKL                      | 1   |
| IGF2R     | G8  | LKMNFTGGTLAIRFISAPQSSSTVTAA   | 26 | LKMNFTGGHLP                    | 3   |
| INO80E    | C9  | PPPKMPPPRS                    | 2  | LPPPKMPPHDEHGPSADVQRCR         | 15  |
| INTS8     | A8  | LAAQRRKKSFQKQWQNTFKQLNFFNFY   | 43 | LAAQRRKKVSPSNGKTLLLSS          | 13  |
| KCNC1     | A8  | KQKLPKKRSIFRGHRSWDLPIIVNLS    | 19 | KQKLPKKKEAYSAAATAAGISQLL       | 15  |
| KCTD16    | A9  | SNMSSKKLLKKSSQLRRSWRNVSRIS    | 19 | ESNMSSKKSC                     | 2   |
| KDM5A     | A8  | PKSSSQKDPAGKLLK               | 8  | LPKSSSQKRIQLAS                 | 6   |
| KIAA0391  | T9  | VYSVPHFF                      | 0  | VYSVPHFFSWSS                   | 4   |
| KIF22     | C8  | PCTLLAPPSPCWAUTPGGWAAS        | 14 | FPCTLLAPLPRAGPLPGGGRPHELLPA    | 33  |
| KIR2DL4   | A9  | HRWCSSKKCCCNEPRACGTQNSEQGGL   | 19 | LHRWCSSKKNAVMNQEPAGHRTVNREDS   | 107 |
| KMT2C     | A9  | CVSVEPKKRNKTKLWFSLNIHHRKNP    | 25 | CVSVEPKKGRKQNSGSL              | 10  |
| KNTC1     | A8  | TAKHVFEKN                     | 1  | TAKHVFEKTEAKAPEVNTS            | 11  |
| LARP7     | A8  | ASMASLKKQYPK                  | 4  | KASMASLKNIPNPKIRVRNGNRQWSTSK   | 25  |
| LENG8     | T9  | PVFLLIFFPLSLPPSPHSSLGSAQVKRF  | 30 | PVFLLIFFFPPLSLPLRILPLVQHR      | 16  |
| LIG3      | A8  | RARATTKKSRTSQSWKAGKSWKIMRRNR  | 20 | ERARATTKNRGPHRAGRLGRAGR        | 15  |
| LIMK2     | G8  | PESRAPPGPQERARAARMMAQ         | 14 | LEPESRAPRRRGPGLRG              | 11  |
| LINC01587 | T10 | NQFFTIFFSQQLNLGRKEHAKIFTFFFQ  | 58 | NQFFTIFFPVS                    | 3   |
| LMAN1     | A9  | QQELDKKKRNSRRATPTSKGSLRRKYLR  | 21 | FQQELDKKRGIEGPPRPPRAACGGNI     | 19  |
| LTN1      | A11 | SLKSSKKMVRDLMLRYLKAIRKMKNV    | 32 | SLKSSKKKW                      | 1   |
| MAPRE3    | C8  | RLSNVAPPFSGRILHQPEMAAMRLMPK   | 30 | GRLSNVAPLHSPEESSISPKWRP        | 15  |
| MARCKS    | A11 | SNETPKKKRSAPPSRLSS            | 11 | SNETPKKKEALFLQEVFQAERLLLQEEQ   | 27  |
| MARS      | T9  | SAICRYFFCYLAGSKMTSLTSGWNGKRQ  | 32 | SAICRYFFVIWLGAR                | 7   |
| MBD4      | A10 | ENSLVKKKKDH                   | 3  | EENSLVKKRKIIEFRIKFLF           | 12  |
| MCPH1     | A9  | HLSSLIKKNVNVCSPKILILKHQKMIRD  | 38 | HLSSLIKKT                      | 1   |
| MED23     | A8  | ESRFNLKKMWRRSIGSGSQ           | 11 | ESRFNLKKCGGV                   | 5   |
| MED8      | A9  | LLGQEGKKMPTCSSGEAWACSPFHPLRS  | 61 | LLGQEGKKCPPAALVKLGPALLSIL      | 17  |
| MGAT4C    | A8  | ENPIIIKKLK                    | 2  | FENPIIIKN                      | 1   |
| MIS18BP1  | A11 | TPNRVEKKNCISPTIYVKKRKTTHSSQM  | 23 | TPNRVEKKIAAYLPM                | 7   |
| MKI67     | A8  | QCLDNSKKMNLPGFSMSQ            | 11 | QCLDNSKK                       | 0   |

|         |     |                               |     |                              |     |
|---------|-----|-------------------------------|-----|------------------------------|-----|
| MLH3    | A9  | QTEKEKKKNLAIVEEEMFLVMGELNYVP  | 23  | AQTEKEKKRI                   | 2   |
| MLH3    | A8  | SGQLPNKKIAERI                 | 5   | SGQLPNKKLQNEYKLWARE          | 11  |
| MNS1    | A10 | QLEEQEKKSRKLMSSC              | 8   | QLEEQEKKAGSL                 | 4   |
| MSH3    | A8  | KENVRDKKRATFLLALWECSLPQARLCL  | 31  | KENVRDKKGQHFYWHGCSAACHRRGCV  | 19  |
| MSH6    | C8  | LLPEDTPPS                     | 1   | ILLPEDTPLLRA                 | 4   |
| MYBL1   | A8  | SLVLEGKKTLMVATVKLFL           | 12  | SLVLEGKKHL                   | 2   |
| MYH11   | C8  | KSKLRGPPHRKLRSAPGEETRPLSFL    | 90  | KSKLRGPPTGNFAVMHQARKRDLFRSF  | 19  |
| MYO10   | G8  | GWLHKKGGAPPRCPEIGRSAGLSSASP   | 21  | GWLHKKGGLLHAVQEKLEELVCPPPVQ  | 24  |
| MYO1A   | A8  | NSKLRYKKRGVIAWR               | 7   | NSKLRYKKGESLLGGCAVRRGHHHEMA  | 24  |
| NDUFC2  | T9  | LLYITAFFLLDIIL                | 6   | LLYITAFFCWLSCKT              | 8   |
| NFRKB   | A8  | YDLAVLKKRLRKKRKRKR            | 12  | YDLAVLKKG                    | 1   |
| NIPBL   | A8  | NHPETPKKSLISFQVRK             | 10  | NHPETPKKV                    | 1   |
| NKTR    | C8  | IPLSDSPPLQDGSLDRLNGSPLMSEFRK  | 20  | IPLSDSPFKMEAWTETLEALL        | 14  |
| NRIP1   | A8  | NKGKSEKKRKL                   | 4   | PNKGKSEKRENSLKR              | 7   |
| NSD2    | C8  | STKTEKPPQSQSGRRGRGGGGAGGESQ   | 25  | STKTEKPPRAREAEGEAAAEGLAESHR  | 25  |
| NTAN1   | A8  | NKALLYKKMKMACGKRSLQEVKKHELP   | 62  | NKALLYKK                     | 0   |
| OR51E2  | T8  | VVRGSLFFSHCLC                 | 5   | VVRGSLFFPTASADQAAGLLPLQCPLAL | 120 |
| OR7C1   | T10 | GCLSQIFFSLHLDAWTIYS           | 11  | GCLSQIFFHFIWMPGQFTLDRDGL     | 16  |
| OR7E24  | T11 | YFPILFFFSSKGVSRATQSHRISQVSQNS | 88  | YFPILFFFPQKVSELHRATESHRLRIP  | 58  |
| OSGEPL1 | T8  | LTKTAGVFLNHQKGKFMNF           | 11  | LTKTAGVF                     | 0   |
| P4HB    | A8  | DVAFDEKKTSLWSSMPHGVVTANSWLFP  | 117 | DVAFDEKKRLCGVLCPMVWSLQTVGSHL | 21  |
| PA2G4   | A8  | RKTQKKKKRRPPRLQRMPPVGKH       | 15  | RKTQKKKKEGLQDCRECHQWGNIRRK   | 18  |
| PIGB    | T8  | SLMIDRIFLANGLWFLIF            | 11  | SLMIDRIFWPMDSGSI             | 8   |
| PLXNA3  | G8  | LLFLAVGGPWATGPSVPSW           | 11  | LLFLAVGGPGQALPCLGRDRHHAYPPG  | 37  |
| PMS2    | A8  | LRTGEEKTKCPFPDCERPFLFVTQQR    | 33  | SLRTGEEKRRVHFQTARGLFSSSHNREQ | 43  |
| POLD3   | A9  | KVLQKEKKGGSE                  | 4   | KVLQKEKKGEASSII              | 7   |
| PPRC1   | C8  | SGYPCLPPHQRC                  | 5   | SGYPCLPPTNGAPSVWYWSLCLASHLQ  | 74  |
| PRCC    | C8  | QDPALVPPRKLPQMPPSSMTKHLSCRA   | 46  | QDPALVPPGNCPRCLLHR           | 10  |
| PRDM2   | A9  | PLSPPKKKFLIHLRKVDTHHLQVVT     | 73  | KPLSPPKKSFSFI                | 5   |
| PRDM2   | A8  | KLNALKKKIS                    | 2   | KLNALKKKSASTESNSSEKQICKAEGRL | 24  |
| PRKCI   | A8  | VLLVRLKKQIVFMQ                | 6   | KVLLVRLKNRSYLCNESCEKRAC      | 15  |
| PRRC2C  | A10 | RRERDEKKMLT                   | 3   | RRERDEKKC                    | 1   |
| PRRG1   | C8  | LNHITPPPHQMKCLTAVDCLQAFWDM    | 18  | LNHITPPPTR                   | 2   |
| PTHLH   | A11 | LLWGLKKKRKTTEEHIICN           | 11  | LLWGLKKKGKQKNTSYATNDLII      | 16  |
| PTPN13  | A8  | FKTSGPEKNPSALMCFLLRRRSGLHPWT  | 44  | FKTSGPEKTHPWH                | 5   |
| PTPN21  | A8  | KTRVDAAKLVLNLWP               | 8   | KKTRVDANKWSS                 | 4   |
| RAB40C  | C8  | IRPPQSPPTARGVTARSPSGDGRGLC    | 57  | IRPPQSPELLAE                 | 5   |
| RAD50   | A9  | STESLKKRKSVMKCDLCP            | 13  | STESLKKGKAA                  | 4   |
| RAD50   | A8  | KEELINKKIQATK                 | 5   | KEELINKKYQQNSTG              | 8   |
| RAPGEF6 | A9  | IPKIAEKKVIAILSSMCQEILNRHRRK   | 58  | IPKIAEKK                     | 0   |
| RAPH1   | C8  | QVAPPTPPHLLLSLHPSLPKLPQNPL    | 18  | QVAPPTPPTSSYPCTPPSPSPKTPCDH  | 156 |
| RARRES3 | A8  | AIRRYQKKRQPEAATKSCVRSSCGGSPG  | 55  | FAIRRYQKSDSLKQPQNPVLEAAVGVPV | 54  |
| RBBP8   | A9  | SLLPQGGKNI                    | 2   | SLLPQGGKTSENTPF              | 7   |
| RECQL   | A9  | QELIQKKKS                     | 1   | QQELIQKKSPDKENKAVFRGF        | 13  |
| REV1    | A8  | PGREKLKK                      | 0   | FPGREKLKNENRQVCTCCN          | 11  |

|          |     |                              |     |                              |     |
|----------|-----|------------------------------|-----|------------------------------|-----|
| REV3L    | A8  | VPATESEKNFRDRFRKFSNRMISL     | 16  | VPATESEKISEETSGNSQTE         | 12  |
| RFC3     | A10 | TITTPSKKN                    | 1   | ITTPSKKKLKLAPLQVTTTLKLILVMLE | 24  |
| RGS12    | A9  | SIKIGKKKNIRKLIWTKQRSFLSLFPKL | 31  | SIKIGKKKISEN                 | 4   |
| RGS12    | C8  | PPGTTTPGRSLPAGPSALPSPSPSRRR  | 48  | PGPPGTTTPRAEVSQRALLHSPVRLPRA | 44  |
| RGS22    | A9  | QEYNKQKKNWQS                 | 4   | QEYNKQKKIGSPRRRKIWKGWNQTICKY | 31  |
| RIF1     | A8  | AETEQNKKRQTLRTLRLGRMVPRTL    | 17  | AETEQNKKGRP                  | 3   |
| RNASEH2B | A10 | FGVKNKKKLERFETLKIKSSKNICFLHV | 25  | FFGVKNKKKNWKGKLKL            | 7   |
| RNF43    | C6  | KLESPRRAPAPACHWLARLGWRVSEEPV | 41  | KLESPRRAPLPVTG               | 7   |
| RNF43    | G7  | HPQRRRGVPPSPPLALGPRMQLCTQLA  | 40  | HPQRRRGSLRAHPWLSAPGCNCAPSLP  | 86  |
| RNPC3    | A12 | TSGSEKKKGLMTLSKMIKKKKNLVI    | 17  | TSGSEKKKV                    | 1   |
| ROCK1    | A8  | ACSQDEQKNG                   | 2   | ACSQDEQKMGNSFSKENP           | 10  |
| RPL22    | A8  | LVVKGKKRSKF                  | 4   | LVVKGKKKEASSEVHS             | 8   |
| RTN3     | A8  | TPWNRQKKA                    | 2   | KTPWNRQKRQNKYMETRNATVTKTPFNS | 66  |
| RUFY2    | A9  | GLEVPRRPGARTGPATKDPTAVERANLL | 129 | GLEVPRRPGARTGPATKDPTAVERANLL | 137 |
| RXFP2    | A10 | SIFIKKKKVYLHPLCG             | 8   | SIFIKKKKFIYIHCVDRLGLLPETWGF  | 94  |
| SAFB2    | C8  | PHPHPHPPRTPTSPAATKSHSLRVFGWA | 48  | PHPHPHPPVPLHPLPLSPTRCEFSGGQ  | 23  |
| SEC31A   | A9  | FINYCQKKMLLRLNLRKMC          | 15  | GFINYCQKN                    | 1   |
| SEC62    | A9  | KKEKEKKKMVKRKNPKRRKLQELLKRRK | 57  | TKKEKEKKRW                   | 2   |
| SEC63    | A10 | KKKKPLKKNLHLCCYHQSNNRNSRQME  | 27  | KKKKPLKKTYTCAITTVKATETKAGKWS | 23  |
| SF3B2    | A8  | HAAKQKQKNGKLSPTAVGAARNIRSS   | 58  | HAAKQKQKTESSAPGQPWGQEI       | 15  |
| SG02     | T9  | GLKKGNFPSKPKRIKNSLKT         | 13  | GLKKGNFQNPGR                 | 5   |
| SHC1     | G8  | GKEPPLGGW                    | 1   | GKEPPLGGGRHEASGRSRGCSHTCTQ   | 59  |
| SLAMF1   | A9  | YQTTVEKKALRSMPSKRNQVLFRRNLTP | 53  | YQTTVEKKPYDLCSPETRSSSEET     | 17  |
| SLC16A4  | T9  | LLSSVSFFLYHWPDKGKT           | 11  | LLSSVSFFCTIGRMEKQSDLKERRLQS  | 23  |
| SLC17A2  | A8  | AGIFGAKKCLVLVC               | 6   | LAGIFGAKNAWCWFADLFPSHPLYTTGC | 20  |
| SLC22A9  | A11 | ELEAAQKKNLCCVKCSTCPTYVKGSPSC | 66  | ELEAAQKKTFV                  | 4   |
| SLC23A2  | C9  | LSCAPPPSTQ                   | 3   | RLSCAPPPHPRNKQGNFRGRPLCS     | 17  |
| SLC35F5  | T10 | VAKISFFALCGFWQICHKKHFQTHKL   | 21  | QVAKISFFLLCVVFGKFVISRSTFRHTS | 23  |
| SMAP1    | A10 | TNKEKEKKRKRKREKRSQKSRQNLQLK  | 47  | TNKEKEKKGREKERKGARKAGKTTYS   | 18  |
| SMAP2    | A8  | GSEPVPEKNWNLLFLRR            | 9   | GSEPVPEKIGTCCF               | 6   |
| SMARCC2  | C8  | AVPPGVPLDPMAPHRSPTNKLLPQ     | 17  | AVPPGVPPWTPWPLTVPQPTNSSNDAR  | 116 |
| SMC1B    | A8  | QFNFNKKKI                    | 1   | QFNFNKKKYSGRAQTSKIREGRGRTLPE | 22  |
| SMG7     | A9  | QQQPLEKK                     | 0   | MQQQPLEKNEAFSHGAI            | 9   |
| SP100    | A8  | IPPKGEKKRSSRIMHPRGLLWPFSCSA  | 39  | IPPKGEKKEVQGSQCTQEASFGLFPVLL | 20  |
| SPAG9    | A9  | NPAMQEKKGQAFGSFADFSAPQVTRLR  | 27  | NPAMQEKKVKHLAVFQPTFQLLK      | 15  |
| SPINK5   | A10 | EREAEEKRKRMTGAIQEKGAIEKGA    | 125 | EREAEEKKERG                  | 3   |
| SREK1IP1 | A10 | EKIKLKKKGKGLTHPVPLKRTLQNKRNK | 66  | EKIKLKKKEKVLIIQFH            | 9   |
| SRRT     | G8  | YEMPYAGVGGAQLMAPLSPGATLTSTSC | 44  | YEMPYAGGWGNLWPPSALGPP        | 14  |
| ST8SIA6  | A8  | YGNLKEKKPYFWRTLQPMEMHFFFCQH  | 52  | KYGNLKEKSPISGGHCNLWRCIFSSASI | 39  |
| SYCP1    | A10 | IAKMDRKKN                    | 1   | IAKMDRKKTKRS                 | 4   |
| TAF1B    | A11 | LNRGLKKKTLKKAGIGMCVKVSSIFFI  | 25  | LNRGLKKKQY                   | 2   |
| TBC1D23  | A9  | VVKITSKKNILNSLPSSMEIAVLQE    | 17  | VVKITSKKTS                   | 2   |
| TCERG1   | A8  | STWKEVKKSLRKILDVLSPPVTGKNKE  | 40  | TSTWKEVKNH                   | 2   |
| TCF7L2   | A9  | CGPCRRKKSATFYKVKAAASAHPLQME  | 22  | CGPCRRKKVRS LHTR             | 7   |
| TEAD2    | C8  | ELYDRGPPMPSSWSSSGRT          | 11  | ELYDRGPPCLLPQVGLGGPELGPKW    | 17  |

|         |     |                               |     |                              |     |
|---------|-----|-------------------------------|-----|------------------------------|-----|
| TFAM    | A10 | RKAMTKKKS                     | 1   | KRKAMTKKRVNTAWTKKTSFSL       | 15  |
| TFDP2   | A8  | NIISKEKKSSGLACLPIILLRNVIWR    | 19  | MNIISKEKRNQVDWPAYQFCSGMSESGD | 56  |
| TFE3    | G8  | AATFHVGGDLPRMLPISSLPHRPMPFW   | 43  | AATFHVGGTCPECSPSAAPCTALRCPSG | 137 |
| TGFBR2  | A10 | KCIMKEKKSRLSSCVPVALMSAMTTS    | 34  | KCIMKEKKAW                   | 2   |
| TMBIM4  | T10 | SGFLKFFIVR                    | 3   | LSGFLKFLL                    | 1   |
| TMEM60  | A9  | GSHNIKKKPGTSLQCYLN            | 10  | HGSHNIKKSLVPHCNVT            | 9   |
| TMEM97  | A10 | KYEEKRKKNEGNNHWPRVEMPTGWLLVG  | 187 | KYEEKRKKMKETTTGPG            | 9   |
| TNKS2   | A9  | HGADPTKKTGMEILLWILLKMEIQIFI   | 28  | HGADPTKKQGWKYSFGSC           | 10  |
| TNRC6C  | A8  | GHLGDGKKMDLGMILTIVIGQLVGMTPR  | 44  | GHLGDGKKWIWMGC               | 6   |
| TOPBP1  | A8  | AAMRWKSSKDKNLALLREHLVGGRLF    | 60  | LAAMRWKKNPAKTRIWHC           | 10  |
| TOPORS  | A8  | ERTNARKKIITVRGSITTMKGTQDGACL  | 33  | ERTNARKK                     | 0   |
| TRIM59  | A10 | DTLEQKKKVS                    | 2   | DTLEQKKKFPNGSL               | 6   |
| TRMT10C | A10 | CVSNTAKKNI                    | 2   | CVSNTAKKIFIFIYEGKSEKS        | 14  |
| TTC39A  | G10 | RIDWGRGGAKREPTREQQERPFWQSSL   | 34  | RIDWGRGGPRENQPESRGRRGPSGRAAW | 38  |
| TTK     | A9  | SSKTFEKKGEKNLQLFVMSDTTYKIYW   | 38  | SSKTFEKKGKKMICSYS            | 9   |
| UBE4B   | A8  | RVGIEEKKHQRCAASQQSASF         | 13  | DRVGIEEKSTKDVQPASSQPASEQHPLT | 51  |
| UBR5    | A8  | NMQNRQKKKGKNSPCCQKKLRVQNQGHL  | 27  | MNMQNRQKRRGRRTARAARRN        | 12  |
| UGDH    | T9  | CRGKNLFFLPILMMPSSKKLILYFL     | 17  | CRGKNLFFYQY                  | 3   |
| USF3    | A9  | PSDQVQKKLLFIPTGISLVETAREQLFR  | 21  | IPSDQVQKNYCLFQRESAWWKQPGNSCS | 24  |
| USF3    | A11 | LSDLAKKKTLRNHLFLIRWIILTFLLQKI | 22  | LSDLAKKKPSEIISF              | 7   |
| USP1    | A8  | PSKVLNKKM                     | 1   | PSKVLNKKCRSYWTSWRTKEQSRL     | 16  |
| UVRAG   | A10 | TSNELKKKVNAACS                | 5   | TSNELKKK                     | 0   |
| VDAC1   | A8  | FSPNTGKKMLKSRQGTSGSTLTWAATWI  | 33  | FSPNTGKKC                    | 1   |
| VEZT    | T9  | IKSWIFFSVQ                    | 2   | TIKSWIFFLSAIRKMTYFTSWILDSDST | 37  |
| VPS13A  | A8  | QYKELYKKS                     | 1   | QYKELYKKVNK                  | 3   |
| VPS13C  | A8  | TAKAYLKKLVCSALISLTLKGNLFTLLT  | 23  | MTAKAYLKN                    | 1   |
| VPS33A  | T8  | ADVKNIIFLSDPG                 | 5   | ADVKNIIFCQTQARVDGYNR         | 12  |
| VPS37B  | C9  | VAPRRIPPHHPRLRDA              | 9   | VAPRRIPPTTPGACGTLSPVYCGHEFG  | 122 |
| VPS54   | A8  | GKIALPEKNQEPQKGNQLKFLSRDNS    | 29  | GKIALPEKIRSHRRKETS           | 10  |
| VRK2    | A8  | HNRLIEKSTVREALSPVQHGGCRKRRN   | 20  | AHNRLIEKSPQ                  | 3   |
| WASF3   | C8  | NRPQPPPLRLPRQGRPRPLHRWLQQT    | 27  | NRPQPPPPASPSGPRGVPLCTDGSSRL  | 28  |
| WDTC1   | G8  | ELLVNMGGNRSICLT               | 7   | ELLVNMGGTGLFV                | 5   |
| WISP3   | A9  | NLPLIWKKNVLCQKQNGLPAPEHVGWEY  | 23  | NLPLIWKKMSCASNKMDSLQNMWDGNI  | 20  |
| WNK1    | A10 | VREEQEKKSRKRAVSNSR            | 10  | VREEQEKKAGREQSQTAGRTIQCFPDRN | 26  |
| WRN     | A8  | MSEKNWKQLHSSGNVLNG            | 14  | MSEKIGNNCTAAEMS              | 11  |
| XPOT    | T9  | LTKWPKFFLTFQ                  | 5   | LTKWPKFF                     | 0   |
| XRCC2   | T8  | KYCLGRFFWCTAVVAPTYFLHFTH      | 16  | KYCLGRFFGVLQ                 | 4   |
| YBX1    | C8  | QQPPAAPPPPPSPAPTPSPALRAAAQG   | 47  | TQQPPAAPRRPRPQRRRHQARHYGQRRR | 63  |
| ZC3H13  | A8  | QKKKSIEKNVKNPKVILFLMKKQPSKV   | 30  | QKKKSIEKT                    | 1   |
| ZMYND8  | A8  | KEPSAVKKSPLQTQWRLKRS          | 13  | KEPSAVKKAQAYKPSGD            | 9   |
| ZNF133  | A8  | KETWREEKNVHRQPVQIQSSTSILSA    | 120 | KETWREEKMFTGNLSRSRALARPRSLP  | 35  |
| ZNF292  | A9  | EKQLVEKKVLTKQKVLKY            | 10  | KQLVEKKS                     | 0   |
| ZNF365  | A8  | IRPKMAKKSQQPL                 | 5   | IRPKMAKKANSHCEHHLKGWVVLQSSSL | 31  |
| ZNF609  | A8  | DKKKDKKRRNLQRNLKVL            | 11  | DKKKDKKEGIFKGT               | 7   |
| ZNF638  | A8  | KKNISEKVTWMKRRRRNLIRKPEWIF    | 21  | DKKNISEKR                    | 1   |

|        |    |                              |    |                           |    |
|--------|----|------------------------------|----|---------------------------|----|
| ZNF644 | A9 | SKIGLEKKESEKWM               | 6  | SKIGLEKKKAKNGCKQDNSLYRGLL | 17 |
| AXIN2  | G7 | ASRHHLWGATAGTPAPPPVPTCSPRTL  | 20 |                           |    |
| B2M    | A5 | KNGERIEKWSIQTCLSARTGLSISCTTL | 20 |                           |    |
| RFX5   | C7 | PKTGGRAPQVVRLGNLPPFFRGSEVPF  | 20 |                           |    |
| SLC4A3 | C7 | EAEPVEPPTQGPARRQSSPLEVTRMTVQ | 20 |                           |    |
| XPO5   | A7 | QWEAMTLFWKVLSPRCFEH          | 11 |                           |    |
| XYLT2  | C7 | HLYGSYPARQPSRPTGRTPTRLMAPV   | 20 |                           |    |
| ZBTB20 | C7 | HSASNGTPLQAHPQVPALAPQAWWPARR | 20 |                           |    |

*Supplementary table 2. Data table for cell lines (CL) and organoids from CRC, EC with available variant calling files from the Cell Model Passports website selected for variant analysis.*

| Model Name        | Tissue          | Cancer type           | Model type | Tissue Status      | MSI status | RNA |
|-------------------|-----------------|-----------------------|------------|--------------------|------------|-----|
| AN3-CA            | Endometrium     | Endometrial Carcinoma | Cell Line  | Tumour             | MSI        | NO  |
| BB30-PBL          | Unknown         | Non-Cancerous         | Cell Line  | Normal             | n/a        | NO  |
| BB49-EBV          | Unknown         | Non-Cancerous         | Cell Line  | Normal             | n/a        | NO  |
| BB65-EBV          | Unknown         | Non-Cancerous         | Cell Line  | Normal             | n/a        | NO  |
| BPH-1             | Prostate        | Non-Cancerous         | Cell Line  | Benign hyperplasia | MSS        | YES |
| C2BBe1            | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | NO  |
| CaR-1             | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | NO  |
| CCK-81            | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSI        | NO  |
| CL-11             | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | NO  |
| CL-34             | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSI        | NO  |
| CL-40             | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | NO  |
| COLO-205          | Large Intestine | Colorectal Carcinoma  | Cell Line  | Metastasis         | MSS        | NO  |
| COLO-320-HSR      | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | YES |
| COLO-678          | Large Intestine | Colorectal Carcinoma  | Cell Line  | Metastasis         | MSS        | NO  |
| COLO-684          | Endometrium     | Endometrial Carcinoma | Cell Line  | Tumour             | MSI        | NO  |
| CP50-EBV          | Unknown         | Non-Cancerous         | Cell Line  | Normal             | n/a        | NO  |
| CW-2              | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSI        | NO  |
| DiFi              | Large Intestine | Colorectal Carcinoma  | Cell Line  | Metastasis         | MSS        | YES |
| EN                | Endometrium     | Endometrial Carcinoma | Cell Line  | Tumour             | MSI        | NO  |
| ESS-1             | Endometrium     | Other Sarcomas        | Cell Line  | Tumour             | MSS        | NO  |
| Geo               | Large Intestine | Colorectal Carcinoma  | Cell Line  | Unknown            | MSS        | YES |
| GP5d              | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSI        | YES |
| HA7-EBV           | Unknown         | Non-Cancerous         | Cell Line  | Normal             | n/a        | NO  |
| HCC2998           | Large Intestine | Colorectal Carcinoma  | Cell Line  | Tumour             | MSS        | YES |
| HCM-SANG-0265-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Metastasis         | MSS        | NO  |
| HCM-SANG-0266-C20 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0267-D12 | Large Intestine | Non-Cancerous         | Organoid   | Precancerous       | MSS        | NO  |
| HCM-SANG-0268-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0269-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0270-C20 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0271-D12 | Large Intestine | Non-Cancerous         | Organoid   | Precancerous       | MSS        | NO  |
| HCM-SANG-0272-C20 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0273-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSI        | NO  |
| HCM-SANG-0274-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0275-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0276-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSI        | NO  |
| HCM-SANG-0277-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0278-C20 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0280-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0281-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSS        | NO  |
| HCM-SANG-0282-C18 | Large Intestine | Colorectal Carcinoma  | Organoid   | Tumour             | MSI        | NO  |



|                   |                 |                       |           |            |     |     |
|-------------------|-----------------|-----------------------|-----------|------------|-----|-----|
| HCM-SANG-1332-C18 | Large Intestine | Colorectal Carcinoma  | Organoid  | Tumour     | MSS | NO  |
| HCM-SANG-1333-C20 | Large Intestine | Colorectal Carcinoma  | Organoid  | Tumour     | MSS | NO  |
| HCM-SANG-1334-C18 | Large Intestine | Colorectal Carcinoma  | Organoid  | Tumour     | MSI | NO  |
| HCM-SANG-1338-C20 | Large Intestine | Colorectal Carcinoma  | Organoid  | Tumour     | MSS | NO  |
| HCM-SANG-1339-C18 | Large Intestine | Colorectal Carcinoma  | Organoid  | Tumour     | MSS | NO  |
| HCT-116           | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| HCT-15            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| HEC-1             | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | YES |
| HEC-108           | Endometrium     | Endometrial Carcinoma | Cell Line | Unknown    | MSI | NO  |
| HEC-151           | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | NO  |
| HT-115            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |
| HT-29             | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |
| HT55              | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |
| KLE               | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSS | YES |
| KM12              | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| LIM1215           | Large Intestine | Colorectal Carcinoma  | Cell Line | Unknown    | MSI | YES |
| LoVo              | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSI | YES |
| LS1034            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| LS-123            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| LS-180            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | NO  |
| LS-411N           | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | NO  |
| LS-513            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |
| MDST8             | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| MFE-280           | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSS | NO  |
| MFE-296           | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | NO  |
| MFE-319           | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | NO  |
| NCI-H508          | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| NCI-H716          | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| NCI-H747          | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | YES |
| PWR-1E            | Prostate        | Non-Cancerous         | Cell Line | Normal     | n/a | NO  |
| RCM-1             | Large Intestine | Colorectal Carcinoma  | Cell Line | Unknown    | MSS | NO  |
| RK0               | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| RL95-2            | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | YES |
| SK-CO-1           | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| SNG-M             | Endometrium     | Endometrial Carcinoma | Cell Line | Metastasis | MSI | NO  |
| SNU-1040          | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | NO  |
| SNU-175           | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSI | YES |
| SNU-283           | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| SNU-407           | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSI | NO  |
| SNU-61            | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| SNU-81            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| SNU-C1            | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| SNU-C2B           | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| SNU-C5            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | NO  |
| SW1116            | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |

|        |                 |                       |           |            |     |     |
|--------|-----------------|-----------------------|-----------|------------|-----|-----|
| SW1417 | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| SW1463 | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| SW403  | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| SW48   | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSI | YES |
| SW620  | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | YES |
| SW626  | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | NO  |
| SW837  | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | NO  |
| SW948  | Large Intestine | Colorectal Carcinoma  | Cell Line | Tumour     | MSS | YES |
| T84    | Large Intestine | Colorectal Carcinoma  | Cell Line | Metastasis | MSS | YES |
| TEN    | Endometrium     | Endometrial Carcinoma | Cell Line | Tumour     | MSI | NO  |

Supplementary table 3. Cell media compositions used in the study.

| Reagent                              | Concentration | Company      | Reference   |
|--------------------------------------|---------------|--------------|-------------|
| <b><i>T-cell media</i></b>           |               |              |             |
| X-VIVO 15                            | -             | Lonza        | BE02-060F   |
| Human Serum                          | 5%            | Sigma        | H4522-100mL |
| GlutaMax                             | 1%            | Gibco        | 35050061    |
| P/S (Penicillin/Streptomycin)        | 1%            | Gibco        | 15140122    |
| <b><i>moDCs generation media</i></b> |               |              |             |
| X-VIVO 15                            | -             | Lonza        | BE02-060F   |
| Human Serum                          | 2%            | Sigma        | H4522-100mL |
| IL-4                                 | 300 U/mL      | Miltenyi     | 130-093-924 |
| GM-CSF                               | 450 U/mL      | Miltenyi     | 130-093-867 |
| <b><i>K-562 cell lines media</i></b> |               |              |             |
| RPMI 1640                            | -             | ThermoFisher | 21875091    |
| FBS                                  | 10%           | Gibco        | A5256801    |
| GlutaMax                             | 1%            | Gibco        | 35050061    |
| P/S (Penicillin/Streptomycin)        | 1%            | Gibco        | 15140122    |

*Supplementary table 4. Monoclonal antibodies used in the DELAY preclinical analysis and available multimers.*

| Target                | Fluorochrome   | Reference   | Company      |
|-----------------------|----------------|-------------|--------------|
| Monoclonal Antibodies |                |             |              |
| CCR7                  | PerCP-Cy5.5    | 561144      | BD           |
| CD103                 | BV421          | 563882      | BD           |
| CD127                 | PE             | 130-122-994 | Miltenyi     |
| CD137                 | PE             | 555956      | BD           |
| CD14                  | FITC           | 555397      | BD           |
| CD154                 | BV421          | 563886      | BD           |
| CD25                  | PE-Cy7         | 557741      | BD           |
| CD25                  | VioBright B515 | 130-122-994 | Miltenyi     |
| CD28                  | PE             | 555729      | BD           |
| CD3                   | BV421          | 563798      | BD           |
| CD3                   | FITC           | 555339      | BD           |
| CD39                  | PerCP Cy5.5    | 567534      | BD           |
| CD4                   | V500           | 560768      | BD           |
| CD4                   | VioGreen       | 130-122-994 | Miltenyi     |
| CD45                  | VioBlue        | 130-122-994 | Miltenyi     |
| CD45-RA               | PE-Cy7         | 337186      | BD           |
| CD69                  | APC-Cy7        | 557756      | BD           |
| CD8                   | APC            | 555369      | BD           |
| CD8                   | APC-H7         | 560179      | BD           |
| CD80                  | APC-H7         | 561134      | BD           |
| CD80                  | PE             | 557227      | BD           |
| CD83                  | PE             | 556855      | BD           |
| CD83                  | PE-Cy7         | 561132      | BD           |
| CD86                  | PE             | 555658      | BD           |
| CD95                  | APC-Cy7        | 305636      | Biolegend    |
| FoxP3                 | VioR667        | 130-122-994 | Miltenyi     |
| HLA-DR                | PE             | 555812      | BD           |
| HLA-DR                | V500           | 561224      | BD           |
| HLA-I                 | APC            | 555555      | BD           |
| LAG-3 (CD223)         | BV421          | 565720      | BD           |
| PD1                   | APC-Cy7        | 329922      | Biolegend    |
| TIM-3 (CD366)         | PE             | 563422      | BD           |
| Live/Dead             | BV421          | L34955      | ThermoFisher |
| Pentamers             |                |             |              |
| MSH3-A*02:01          | PE             | Custom      | ProImmune    |

|                       |    |        |           |
|-----------------------|----|--------|-----------|
| TGFBRII_B-A*02:01     | PE | Custom | ProImmune |
| ADNP-A*11:01          | PE | Custom | ProImmune |
| KDM5A-A*11:01         | PE | Custom | ProImmune |
| MLH3-A*29:02          | PE | Custom | ProImmune |
| ACVR2A-B*08:01        | PE | Custom | ProImmune |
| AIM2-B*08:01          | PE | Custom | ProImmune |
| ASTE1-B*08:01         | PE | Custom | ProImmune |
| MLH3-B*15:01          | PE | Custom | ProImmune |
| ZNF292-B*15:01        | PE | Custom | ProImmune |
| BAX-B*35:01           | PE | Custom | ProImmune |
| ADNP_II-DRB1*01:01    | PE | Custom | ProImmune |
| AXIN2_II-DRB1*01:01   | PE | Custom | ProImmune |
| SLC22A9_II-DRB1*01:01 | PE | Custom | ProImmune |
| ACVR2A_II-DRB1*03:01  | PE | Custom | ProImmune |
| ACVR2A_II-DRB1*04:01  | PE | Custom | ProImmune |
| ACVR2A_II-DRB1*07:01  | PE | Custom | ProImmune |
| ADNP_II-DRB1*07:01    | PE | Custom | ProImmune |
| SLC22A9_II-DRB1*07:01 | PE | Custom | ProImmune |
| ACVR2A_II-DRB1*11:01  | PE | Custom | ProImmune |
| ADNP_II-DRB1*11:01    | PE | Custom | ProImmune |
| SLC22A9_II-DRB1*11:01 | PE | Custom | ProImmune |
| ACVR2A_II-DRB1*15:01  | PE | Custom | ProImmune |
| TTK_II-DRB1*15:01     | PE | Custom | ProImmune |
| Dextramers            |    |        |           |
| MSH6-A*02:01          | PE | Custom | Immudex   |
| AXIN2-B*07:02         | PE | Custom | Immudex   |
| BAX-B*07:02           | PE | Custom | Immudex   |
| MLH3-A*29:02          | PE | Custom | Immudex   |
| TGFBR2_B-A*02:01      | PE | Custom | Immudex   |
| Control-A*02:01       | PE | Custom | Immudex   |

Supplementary table 5. Flow cytometry panels used in the DELAY preclinical analysis.

| Target                    | FC             | Abs. max (nm) | Emission max (nm) | Laser          |
|---------------------------|----------------|---------------|-------------------|----------------|
| T-reg cells phenotyping   |                |               |                   |                |
| CD45                      | VioBlue        | 400           | 452               | Violet (405nm) |
| CD4                       | VioGreen       | 388           | 520               | Violet (405nm) |
| CD25                      | VioBright B515 | 488           | 514               | Blue (488nm)   |
| CD127                     | PE             | -             | -                 | Green (532 nm) |
| FoxP3                     | VioR667        | 645           | 668               | Red (640nm)    |
| T-cell subpopulations     |                |               |                   |                |
| CD3                       | FITC           | 494           | 518               | Blue (488nm)   |
| CD4                       | V500           | 415           | 500               | Violet (405nm) |
| CD8                       | APC            | 651           | 660               | Red (640nm)    |
| CD45-RA                   | PE-Cy7         | 496 / 566     | 781               | Red (640nm)    |
| CCR7                      | PerCP Cy5.5    | 482           | 676               | Blue (488nm)   |
| CD103                     | BV421          | 407           | 423               | Violet (405nm) |
| CD28                      | PE             | 496 / 566     | 576               | Green (532 nm) |
| CD95                      | APC-Cy7        | 651           | 779               | Red (640nm)    |
| T-cell Exhaustion markers |                |               |                   |                |
| CD3                       | FITC           | 494           | 518               | Blue (488nm)   |
| CD4                       | V500           | 415           | 500               | Violet (405nm) |
| CD8                       | APC            | 651           | 660               | Red (640nm)    |
| CD25                      | PE-Cy7         | 496 / 566     | 781               | Red (640nm)    |
| CD39                      | PerCP Cy5.5    | 482           | 676               | Blue (488nm)   |
| LAG-3 (CD223)             | BV421          | 407           | 423               | Violet (405nm) |
| TIM-3 (CD366)             | PE             | 496 / 566     | 576               | Green (532 nm) |
| PD1                       | APC-Cy7        | 651           | 779               | Red (640nm)    |
| DCs phenotype and purity  |                |               |                   |                |
| CD14                      | FITC           | 494           | 518               | Blue (488nm)   |
| HLA-DR                    | V500           | 415           | 500               | Violet (405nm) |
| HLA-I                     | APC            | 651           | 660               | Red (640nm)    |
| CD83                      | PE-Cy7         | 496 / 566     | 781               | Red (640nm)    |
| CCR7                      | PerCP Cy5.5    | 482           | 676               | Blue (488nm)   |
| CD3                       | BV421          | 407           | 423               | Violet (405nm) |
| CD86                      | PE             | 496 / 566     | 576               | Green (532 nm) |
| CD80                      | APC-H7         | 659           | 782               | Red (640nm)    |
| T-cell activation markers |                |               |                   |                |
| CD3                       | FITC           | 494           | 518               | Blue (488nm)   |
| CD4                       | V500           | 415           | 500               | Violet (405nm) |
| CD8                       | APC            | 651           | 660               | Red (640nm)    |

|                          |         |           |     |                |
|--------------------------|---------|-----------|-----|----------------|
| CD154                    | BV421   | 407       | 423 | Violet (405nm) |
| CD137                    | PE      | 496 / 566 | 576 | Green (532 nm) |
| CD69                     | APC-Cy7 | 651       | 779 | Red (640nm)    |
| K-562 cell sorting       |         |           |     |                |
| Live/Dead                | BV421   | 407       | 423 | Violet (405nm) |
| TMG-plasmid              | mCherry | 590       | 650 | Green (532 nm) |
| HLA-plasmid              | GFP     | 395 / 475 | 508 | Blue (488nm)   |
| T-cell multimer staining |         |           |     |                |
| Live/Dead                | BV421   | 407       | 423 | Violet (405nm) |
| CD8                      | APC-H7  | 659       | 782 | Red (640nm)    |
| Multimer                 | PE      | 496 / 566 | 576 | Green (532 nm) |

Supplementary table 6. List of prioritized FSDN genes, their HLA class restriction in length, Frame-Shift of origin, classification as potential tumour-driver or tumour-escape genes, and number of HLA allele binders (IC50<5000nM).

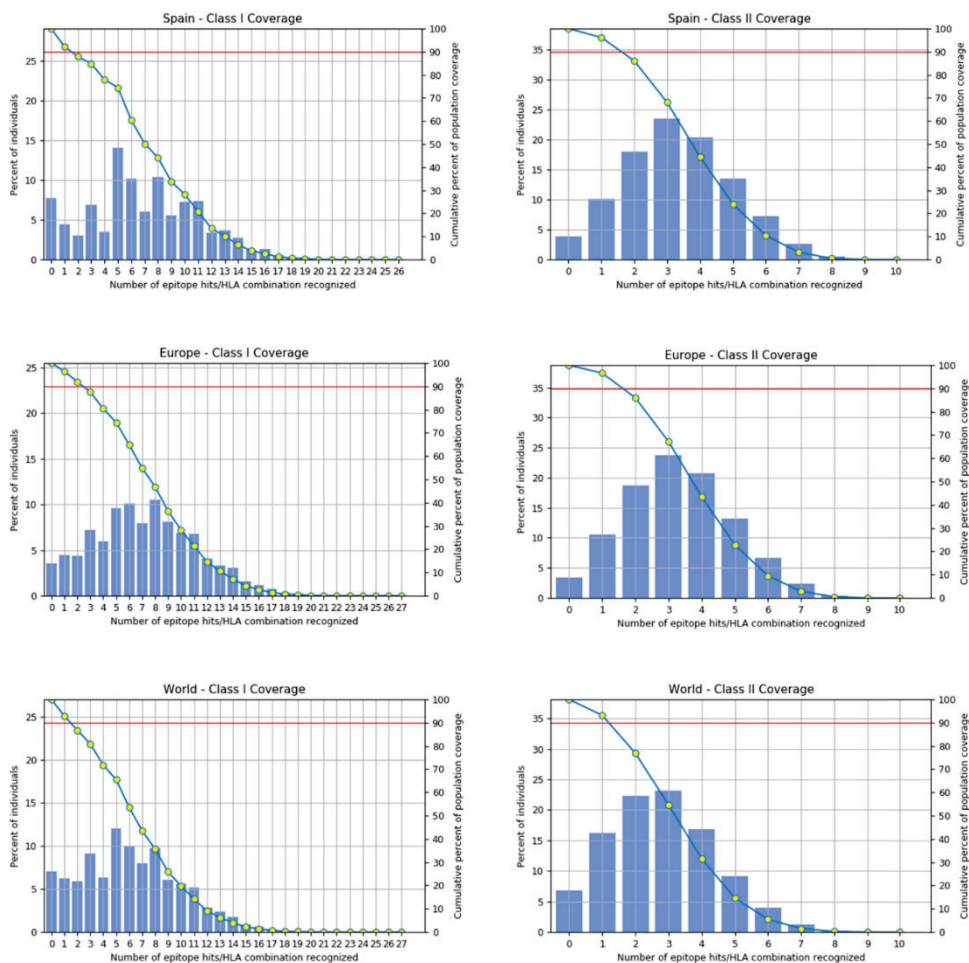
| ID       | Sequence        | HLA restriction | Frame-shift | Tumor_Driver | N binding HLA alleles |
|----------|-----------------|-----------------|-------------|--------------|-----------------------|
| ACVR2A   | VVHKKRGLF       | HLA-I           | M1          | Yes          | 13                    |
| ADNP     | AAKKRLPCK       | HLA-I           | M1          | No           | 5                     |
| ADNP     | SSKPAAKKRLPCKVT | HLA-II          | M1          | No           | 17                    |
| AIM2     | KAKKKHREV       | HLA-I           | M1          | Yes          | 15                    |
| AIM2     | KVIKAKKKHREVKRT | HLA-II          | M1          | Yes          | 20                    |
| AKAP9    | VNMMCQHTM       | HLA-I           | M1          | Yes          | 14                    |
| AKAP9    | KQKKREKRQAVNMMC | HLA-II          | M1          | Yes          | 24                    |
| ASTE1    | VLRTGEPL        | HLA-I           | M1          | No           | 14                    |
| ASTE1    | SNSKKKGRRNRIPAV | HLA-II          | M1          | No           | 30                    |
| ASXL1    | MVVRPVATL       | HLA-I           | M1          | Yes          | 31                    |
| ASXL1    | GGGVARVEVAAGPPM | HLA-II          | M1          | Yes          | 34                    |
| ATR      | FLFLKKKY        | HLA-I           | M1          | Yes          | 12                    |
| AXIN2    | VPTCSPRTL       | HLA-I           | M1          | Yes          | 13                    |
| AXIN2    | ASRHHLWGATAGTPA | HLA-II          | M1          | Yes          | 32                    |
| B2M      | LSARTGLSI       | HLA-I           | M1          | Yes          | 18                    |
| B2M      | KNGERIEKWSIQTCL | HLA-II          | M1          | Yes          | 28                    |
| BANP     | FFFPFFCSV       | HLA-I           | M1          | No           | 26                    |
| BANP_M2  | FFSFLFSAL       | HLA-I           | M2          | No           | 27                    |
| BAX      | WPWTRCLRM       | HLA-I           | M1          | Yes          | 16                    |
| BAX      | MGGRHSPWPTRCLR  | HLA-II          | M1          | Yes          | 33                    |
| BCL10    | FLVEHQVEK       | HLA-I           | M1          | No           | 10                    |
| BCL10    | HLRAKKYSVEKTLKK | HLA-II          | M1          | No           | 26                    |
| BRCA1    | QVKDMTAIL       | HLA-I           | M1          | Yes          | 15                    |
| BRCA1    | SWKVKNLQLEPRRV  | HLA-II          | M1          | Yes          | 29                    |
| C15orf48 | KLWTLLYLK       | HLA-I           | M1          | No           | 11                    |
| C15orf48 | ILDRKKIQNLGKLWT | HLA-II          | M1          | No           | 26                    |
| CASP5    | FLIHWQNTM       | HLA-I           | M1          | Yes          | 29                    |
| CASP5    | EFLRLCKKIMMRSIQ | HLA-II          | M1          | Yes          | 27                    |
| CHD2     | RTRRTRRNK       | HLA-I           | M1          | Yes          | 5                     |
| CHD2     | LEKSPMKKNRRRKRT | HLA-II          | M1          | Yes          | 15                    |
| CHD8     | KWQRCKKII       | HLA-I           | M1          | Yes          | 2                     |
| FLCN     | MTSLSASTSL      | HLA-I           | M1          | No           | 24                    |
| FLCN     | LWGVMTSLSASTSL  | HLA-II          | M1          | No           | 32                    |
| HNF1A    | CAMDSLRPV       | HLA-I           | M1          | Yes          | 29                    |
| HNF1A    | RSTVCAMDSLRPVRL | HLA-II          | M1          | Yes          | 32                    |
| HOXA11   | TVYSTTQOI       | HLA-I           | M1          | No           | 21                    |
| HOXA11   | RMKEKLTETVYSTT  | HLA-II          | M1          | No           | 20                    |
| KDM5A    | SSQKKDPAGK      | HLA-I           | M1          | No           | 3                     |

|                 |                 |        |    |     |    |
|-----------------|-----------------|--------|----|-----|----|
| <b>KDM5A</b>    | KSSSQKKDPAGKLKK | HLA-II | M1 | No  | 8  |
| <b>KMT2C</b>    | KTKLWFSLI       | HLA-I  | M1 | Yes | 20 |
| <b>KMT2C</b>    | LWFSLINIHRKNPL  | HLA-II | M1 | Yes | 39 |
| <b>LTN1</b>     | KMVRDLDM        | HLA-I  | M1 | No  | 21 |
| <b>LTN1</b>     | KSSKKKMVRDLDMR  | HLA-II | M1 | No  | 27 |
| <b>MARCKS</b>   | RSAPPSRSL       | HLA-I  | M1 | No  | 26 |
| <b>MARCKS</b>   | TPKKKRSAPPSRSLS | HLA-II | M1 | No  | 40 |
| <b>MLH3</b>     | FLVMGELNY       | HLA-I  | M1 | Yes | 15 |
| <b>MLH3</b>     | VEEEMFLVMGELNYV | HLA-II | M1 | Yes | 35 |
| <b>MSH3</b>     | LIVSRTL         | HLA-I  | M1 | Yes | 29 |
| <b>MSH3</b>     | KENVRDKKRATFLA  | HLA-II | M1 | Yes | 36 |
| <b>MSH6</b>     | LLPEDTPPS       | HLA-I  | M1 | Yes | 2  |
| <b>MYH11</b>    | RTLETQTSM       | HLA-I  | M1 | No  | 21 |
| <b>MYH11</b>    | KSKLRGPPHRKLRS  | HLA-II | M1 | No  | 21 |
| <b>NIPBL</b>    | KSLILSFQR       | HLA-I  | M1 | Yes | 7  |
| <b>NIPBL</b>    | ETPKKSLILSFQRVK | HLA-II | M1 | Yes | 32 |
| <b>NKTR</b>     | LGSPMLSEF       | HLA-I  | M1 | No  | 13 |
| <b>NKTR</b>     | GSLDRNLGSPMLSEF | HLA-II | M1 | No  | 34 |
| <b>NSD2</b>     | GESQRANSA       | HLA-I  | M1 | Yes | 5  |
| <b>NSD2</b>     | EKPPQSQSRRGRGG  | HLA-II | M1 | Yes | 2  |
| <b>PMS2</b>     | FVTQQRSTL       | HLA-I  | M1 | Yes | 24 |
| <b>PMS2</b>     | FLFVTQQRSTLTAQR | HLA-II | M1 | Yes | 32 |
| <b>PRDM2</b>    | KVDTHHLQV       | HLA-I  | M1 | Yes | 16 |
| <b>PRDM2</b>    | SPPKKKFLIHLRKVD | HLA-II | M1 | Yes | 34 |
| <b>PTPN13</b>   | FVQLTETSL       | HLA-I  | M1 | No  | 23 |
| <b>PTPN13</b>   | GPEKNPSLALMCFLR | HLA-II | M1 | No  | 31 |
| <b>REV3L</b>    | KFSNRMISL       | HLA-I  | M1 | Yes | 14 |
| <b>REV3L</b>    | KNFRDRFRKFSNRMI | HLA-II | M1 | Yes | 32 |
| <b>RFX5</b>     | FFRGSEVPF       | HLA-I  | M1 | Yes | 19 |
| <b>RFX5</b>     | GRAPQVVLRLGNLPP | HLA-II | M1 | Yes | 25 |
| <b>RNF43</b>    | RFFPITPPV       | HLA-I  | M1 | Yes | 24 |
| <b>RNF43</b>    | EPVLSSTLSLRIEL  | HLA-II | M1 | Yes | 33 |
| <b>RPL22</b>    | LVVKGKKRSK      | HLA-I  | M1 | Yes | 3  |
| <b>SF3B2</b>    | FRSPHTSPF       | HLA-I  | M1 | Yes | 19 |
| <b>SF3B2</b>    | KQKQKNGKLSPTAV  | HLA-II | M1 | Yes | 19 |
| <b>SLC22A9</b>  | MSSIWGTMF       | HLA-I  | M1 | No  | 24 |
| <b>SLC22A9</b>  | AAQKKNLCKVKSTC  | HLA-II | M1 | No  | 20 |
| <b>SLC35F5</b>  | FALCGFWQI       | HLA-I  | M1 | No  | 34 |
| <b>SLC35F5</b>  | AKISFFALCGFWQI  | HLA-II | M1 | No  | 39 |
| <b>TAF1B</b>    | MCVKVSSIF       | HLA-I  | M1 | No  | 15 |
| <b>TAF1B</b>    | LNRLGKKKTILKKAG | HLA-II | M2 | No  | 29 |
| <b>TAF1B_M2</b> | NRGLKKKQY       | HLA-I  | M1 | No  | 2  |
| <b>TCF7L2</b>   | AASAHPLQM       | HLA-I  | M1 | Yes | 23 |

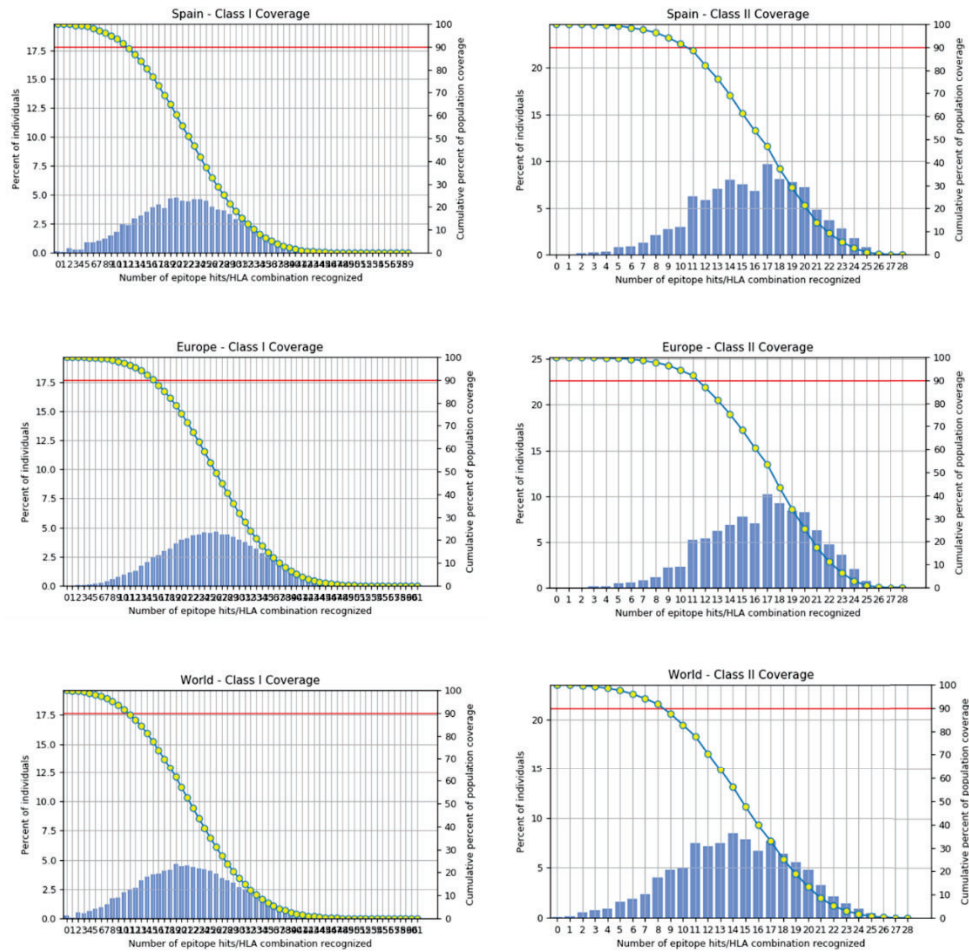
|               |                  |        |    |     |    |
|---------------|------------------|--------|----|-----|----|
| <b>TCF7L2</b> | GPCRRKKSASFATYKV | HLA-II | M1 | Yes | 34 |
| <b>TGFBR2</b> | LSSCVPVAL        | HLA-I  | M1 | Yes | 29 |
| <b>TGFBR2</b> | KEKKSLVRLSSCVPV  | HLA-II | M1 | Yes | 40 |
| <b>TTK</b>    | KIYWTVILL        | HLA-I  | M1 | Yes | 27 |
| <b>TTK</b>    | FEKKGEKNDLQLFVM  | HLA-II | M1 | Yes | 29 |
| <b>UBR5</b>   | VQNQGHLLM        | HLA-I  | M1 | Yes | 32 |
| <b>UBR5</b>   | QKKLRVQNQGHLLMI  | HLA-II | M1 | Yes | 37 |
| <b>WRN</b>    | QLHSSGNVL        | HLA-I  | M1 | No  | 10 |
| <b>WRN</b>    | EKNWKQLHSSGNVLN  | HLA-II | M1 | No  | 32 |
| <b>ZBTB20</b> | QAHPQVPAL        | HLA-I  | M1 | Yes | 25 |
| <b>ZBTB20</b> | ASNGTPLQAHPQVPA  | HLA-II | M1 | Yes | 27 |
| <b>ZC3H13</b> | RKEAHGLPL        | HLA-I  | M1 | Yes | 11 |
| <b>ZC3H13</b> | KKKSIEKNVKNPKVI  | HLA-II | M1 | Yes | 20 |
| <b>ZNF133</b> | CVQKVTSSL        | HLA-I  | M1 | Yes | 20 |
| <b>ZNF133</b> | WREEKNVHRQPVQIQ  | HLA-II | M1 | Yes | 21 |
| <b>ZNF292</b> | VLTKQKVLY        | HLA-I  | M1 | Yes | 8  |
| <b>ZNF292</b> | LVEKKVLTKQKVLYK  | HLA-II | M1 | Yes | 28 |

Supplementary figure 1. Coverage of the DELAY vaccine set on Spain, Europe and World populations based on the covered HLA alleles. Barplots depict the number of epitope hits / HLA combinations recognized by the population. A) Barplots for epitope-HLA combinations with IC-50 <500 nM. B) Barplots for epitope-HLA combinations with IC-50 <5000 nM.

A)

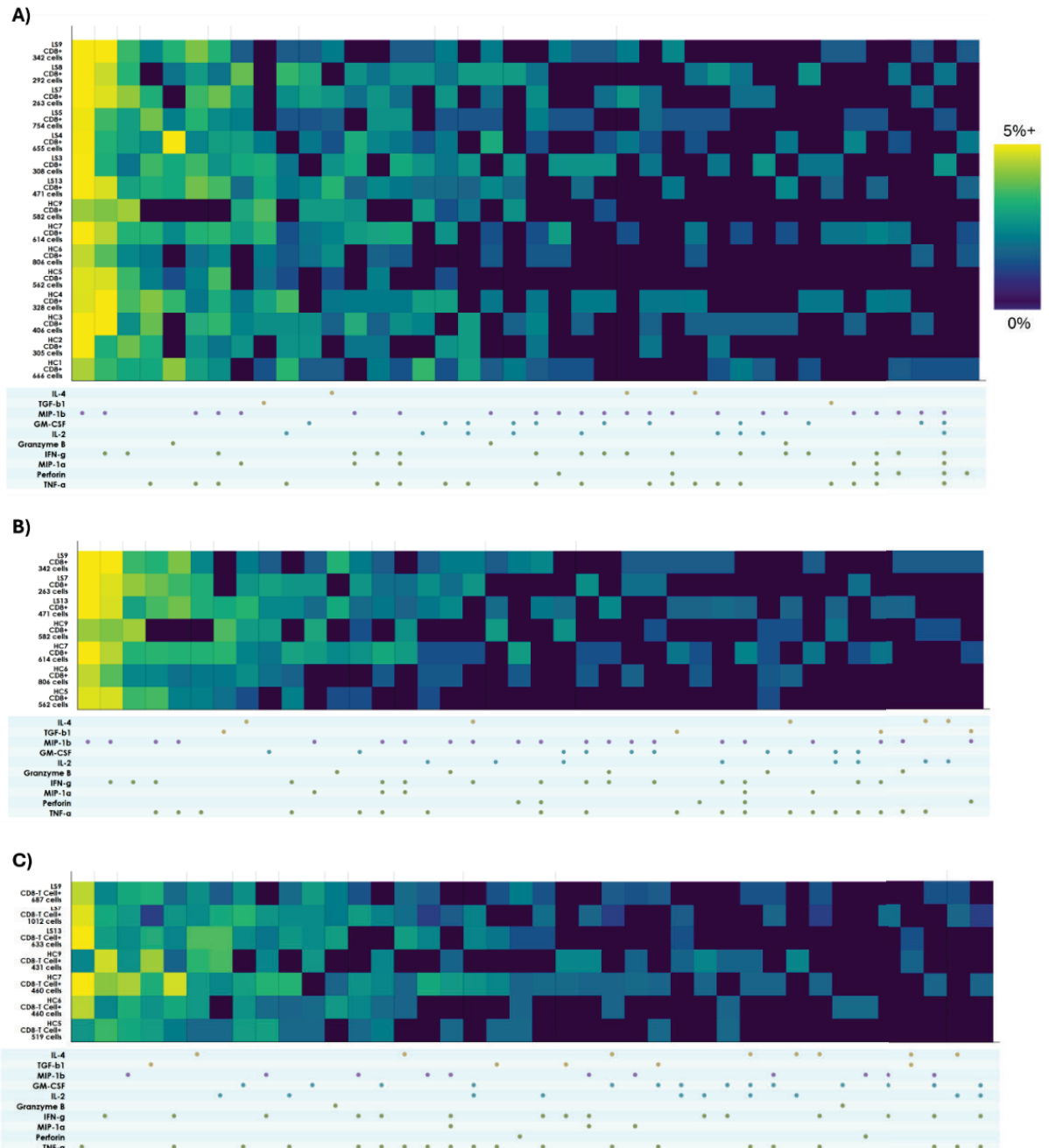


B)



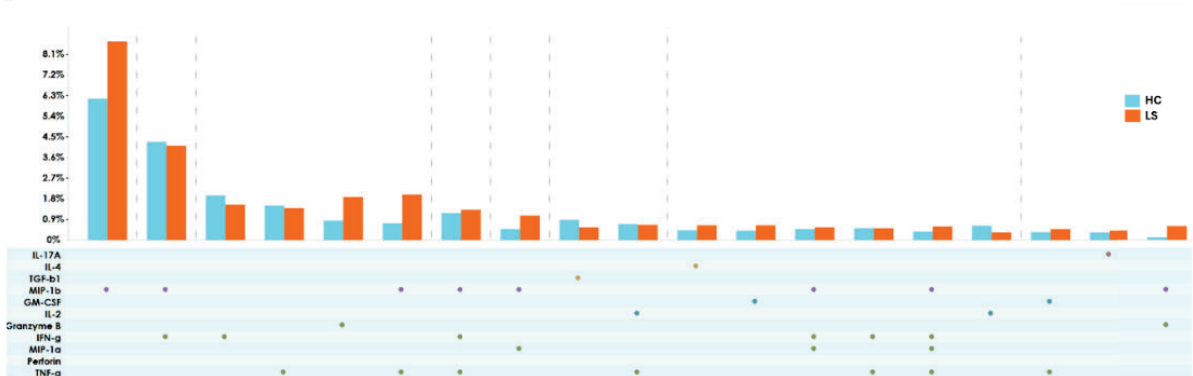


Supplementary figure 3. Single-cell multiplexed functional cytokine profiling of CD4+ and CD8+ T-cells from HCs and LS carriers. Polyfunctional groups for HCs and LS carrier samples, heatmap of the % of the sample expressing the dot-marked cytokine groups in each column, from blue (0%) to yellow (>5%). A) CD8+ cells from all available samples, B) CD8+ cells from only LS carriers or HCs with available CD4+ and CD8+ data, C) CD4+ T-cells from LS carriers and HCs.

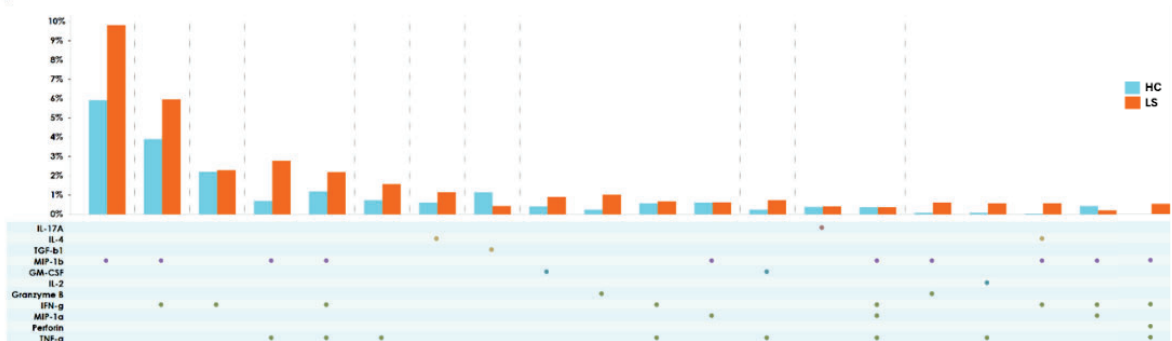


Supplementary figure 4. Single-cell multiplexed functional cytokine profiling of CD4+ and CD8+ T-cells from HCs and LS carriers. Polyfunctional groups for HCs (blue) and LS carrier (orange) samples as the % of the sample expressing the dot-marked cytokine groups in each column. A) CD8+ cells from all available samples, B) CD8+ cells from only LS carriers or HCs with available CD4+ and CD8+ data, C) CD4+ T-cells from LS carriers and HCs.

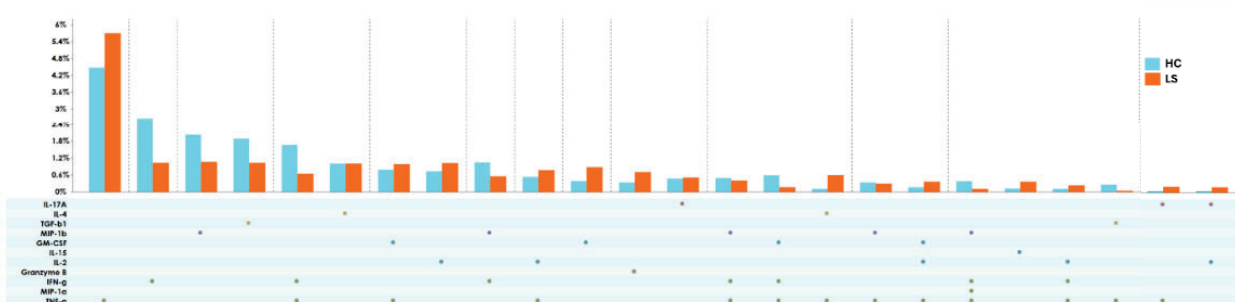
A)



B)



C)



Supplementary table 7. Inclusion and exclusion criteria from the phase Ib clinical trial.

| <b><i>Inclusion criteria</i></b>                                                                                                                                                                                                                           | <b><i>Exclusion criteria</i></b>                                                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Individuals carrying a pathogenic or likely pathogenic germ line variant in one of the DNA repair genes (MMR) ( <i>MLH1</i> , <i>MSH2</i> / <i>EPCAM</i> , <i>MSH6</i> ).                                                                                  | Individuals carrying a pathogenic or likely pathogenic germ line variant in <i>PMS2</i> .                                                                                                                                                                                         |
| No evidence of active or previous invasive cancer.                                                                                                                                                                                                         | Active malignancy or previous malignancy (excluding non-melanoma skin cancer).                                                                                                                                                                                                    |
| Endoscopically accessible colon                                                                                                                                                                                                                            | Distal colon not endoscopically accessible and/or without rectal mucosa.                                                                                                                                                                                                          |
| Consent to standard surveillance with colonoscopy and biopsies every 1-2 years.                                                                                                                                                                            | Participants who cannot be removed from their baseline medication for the duration of the trial to administer the investigational treatment. This includes the daily use of >100 mg aspirin or non-steroidal anti-inflammatory drugs (NSAIDs) or cyclooxygenase (COX) inhibitors. |
| Age $\geq$ 18 years.                                                                                                                                                                                                                                       | Serious pre-existing non-controlled and/or unstable medical disorder, psychiatric disorder, or other conditions that could interfere with participant safety, obtaining informed consent, or compliance with study procedures.                                                    |
| Eastern Cooperative Oncology Group (ECOG) functional status $\leq$ 1 (Karnofsky $\geq$ 70%).                                                                                                                                                               | Positive hepatitis B surface antigen or hepatitis C antibody tests at screening.                                                                                                                                                                                                  |
| Hemoglobin $\geq$ 10 g/dL or hematocrit $\geq$ 30%; Leukocyte count $\geq$ $3.0 \times 10^9$ /l; Platelet count $\geq$ $100 \times 10^9$ /l; Absolute neutrophil count $\geq$ $1.5 \times 10^9$ /l; Absolute lymphocyte count $\geq$ $0.8 \times 10^9$ /l. | Active systemic bacterial, viral or fungal infections or known to have human immunodeficiency virus (HIV) or to test positive for HIV antibody at screening.                                                                                                                      |
| Creatinine clearance (calculated if measurement is not available) $\geq$ 60 ml/min/1.73 m <sup>2</sup> .                                                                                                                                                   | History of organ allograft or other immunodeficiency history.                                                                                                                                                                                                                     |
| Aspartate aminotransferase (AST) (serum glutamic oxaloacetic transaminase [SGOT])/alanine aminotransferase (ALT)                                                                                                                                           | Individuals with a condition requiring systemic treatment with corticosteroids or other immunosuppressive drugs.                                                                                                                                                                  |

|                                                                                                                                                                   |                                                                                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (serum glutamic pyruvic transaminase [SGPT]) $\leq$ 2 times the institutional upper limit of normal (ULN).                                                        |                                                                                                                                                                                                                                          |
| Total bilirubin $\leq$ 1.5 ULN; participants with Gilbert's disease may enroll with higher total bilirubin if their direct bilirubin is $\leq$ 1.5 times the ULN. | Individuals who cannot, for medical reasons, stop taking aspirin or non-steroidal anti-inflammatory drugs (NSAIDs) or cyclooxygenase (COX) inhibitors during the trial duration, except for cardioprotective aspirin ( $<100$ mg daily). |
| Ability to understand and willingness to sign a written informed consent document.                                                                                | Pregnancy or breastfeeding or planning to become pregnant within 6 months after the end of the study.                                                                                                                                    |
|                                                                                                                                                                   | Men who are attempting or planning to father children during the study or within 6 months.                                                                                                                                               |
|                                                                                                                                                                   | Participants who are receiving other agents in research                                                                                                                                                                                  |