

IDENTIFICATION DES MARQUEURS GÉNOMIQUES DE HAUT RISQUE DE PROGRESSION DU MYCOSIS FUNGOÏDE

Léa FLÉCHON

UNIVERSITÉ DE LILLE

ÉCOLE DOCTORALE BIOLOGIE SANTÉ DE LILLE (ED446)

THÈSE

POUR L'OBTENTION DU DIPLÔME D'ÉTAT

DE DOCTEUR EN BIOLOGIE-SCIENCE

DE LA VIE ET DE LA SANTÉ

Spécialité : Bioinformatique, Génétique, Génomique

Soutenue publiquement le 29 novembre 2024

Composition du jury :

<i>Présidente/Examinatrice :</i>	Pr Hélène TOUZET	DR – CRISTaL, Lille
<i>Rapporteuse :</i>	Pr Sandrine ROULLAND	DR – CIML, Aix-Marseille
<i>Rapporteur :</i>	Pr Pierre SUJOBERT	PU-PH – UCBL1, Lyon
<i>Examineur :</i>	Dr Philippe RUMINY	CR – CHB, Rouen
<i>Directeur de thèse :</i>	Pr Salomon MANIER	PU-PH – Canther, Lille
<i>Co-encadrant :</i>	Dr Thierry IDZIOREK	CR – Canther, Lille

Équipe « Facteurs de persistance des cellules leucémiques »

Laboratoire CANTHER « Hétérogénéité, Plasticité et Résistance aux thérapies des Cancers » (CNRS UMR9020 – INSERM U1277) - ONCOLILLE

« Mais les vrais voyageurs sont ceux-là seuls qui partent

Pour partir ; cœurs légers, semblables aux ballons,

De leur fatalité jamais ils ne s'écartent,

Et sans savoir pourquoi, disent toujours : Allons ! »

Charles Baudelaire, *Le Voyage*, *Les Fleurs du Mal*, 1861

« Courage, mon cœur ! Tu as souffert des douleurs
plus cruelles. »

Homère, *Odyssée*, Chant XX, vers 18

Remerciements

Mes premiers remerciements vont au Professeur Bruno Quesnel pour m'avoir accueillie au sein de l'équipe Canther « Facteurs de Persistance des Cellules Leucémiques » et pour la confiance qu'il m'a accordée dans le cadre de cette thèse.

Je tiens ensuite à exprimer ma gratitude envers mes encadrants. Le Professeur Salomon Manier de m'avoir choisie pour travailler avec lui sur ce projet ambitieux en hématologie et en bioinformatique. Je le remercie pour son partage d'expertise et de son savoir, ainsi que pour l'autonomie dont j'ai bénéficié. J'ai particulièrement apprécié son enseignement et les discussions constructives qui ont contribué au développement et à l'amélioration de mon sujet. Merci pour le soutien précieux et la disponibilité qu'il a apportés durant la préparation de cette thèse. Je lui suis reconnaissante de m'avoir permis de réaliser une partie de ma thèse au sein de l'équipe de la Professeure Irene Ghobrial à la *Harvard Medical School* de Boston, et d'avoir pu interagir avec les chercheurs du *Dana-Farber Cancer Institute* (DFCI) et du *Broad Institute* (BI) du MIT. Je considère que ces opportunités comme une chance inestimable et lui témoigne ma plus sincère reconnaissance. Je remercie également le Docteur Thierry Idziorek pour son accompagnement tout au long de la thèse, son partage de bureau durant les premières années, son aide précieuse, ses relectures pertinentes et ses conseils éclairés. Je lui suis reconnaissante de m'avoir parlé de ce sujet de thèse lors de ma recherche d'un doctorat. Merci à tous les deux pour votre contribution précieuse pour la rédaction de cette thèse.

Il est tout naturellement pour moi de remercier la Professeure Hélène Touzet pour l'honneur qu'elle m'a fait en acceptant d'être la présidente du jury et examinatrice de ma thèse et d'évaluer ce travail. Je remercie également la Professeure Sandrine Roulland, le Professeur Pierre Sujobert pour l'honneur qu'ils me font en acceptant d'être les rapporteurs, de lire et de discuter de mon travail et, notamment, leur aide inestimable pour la rédaction de mon second manuscrit de thèse. Je remercie le Docteur Philippe Ruminy pour l'honneur d'accepter d'être examinateur dans le jury de thèse. J'exprime tout mon respect pour leur savoir et leur expertise, et je les assure de ma plus sincère reconnaissance et de ma plus profonde estime.

Je remercie de plus, la Docteure Ségolène Caboche et le Docteur Philippe Ruminy d'avoir participé comme jury lors de mes comités de suivi individuels tout au long de cette thèse et de leurs encouragements pour mener à bien ce projet.

Plus personnellement, un immense merci à la Docteure Inès Arib du département d'hématologie du CHU de Lille, dont le travail sur la partie Mycosis fongoïde a été réalisé en collaboration. Je la remercie de m'avoir aidée à corréler les données cliniques avec mes échantillons séquencés pour les analyses biostatistiques, ainsi que de m'avoir accueillie chez elle pour la rédaction de l'article.

Je remercie l'Institut pour la Recherche sur le Cancer de Lille (IRCL), la Fondation Française pour la Recherche des Myélomes et les Gammopathies (FFRMG), la Société Française d'Hématologie (SFH) et le Centre Hospitalier Universitaire (CHU) de Lille pour le financement de ma thèse. Je tiens également à remercier la Fondation I-Site de l'Université de Lille pour la « Bourse de mobilité internationale de recherche » qui a contribué au financement de ma présence au sein de l'équipe Ghobrial à Boston.

Un grand merci au Docteur Martin Figeac, dont l'intervention a été indispensable, et ce même bien avant le début de ma thèse. Sa supervision dans la partie bioinformatique, son assistance dans la rédaction de scripts, son soutien pendant les moments difficiles et ses nombreuses interventions salvatrices face à des impasses ont été d'une valeur inestimable. Mes remerciements vont également à la Docteure Katarina Jovanovic pour son accueil dans l'équipe et pour m'avoir fait visiter le laboratoire et ses environs.

Je souhaite exprimer une profonde gratitude à la Professeure Irene Ghobrial pour son hospitalité chaleureuse, pour son accueil au DFCI pendant six mois, et pour l'opportunité de participer activement aux activités de l'équipe en tant que bioinformaticienne. Ce beau travail a abouti à une publication conjointe dans le journal « *Cancer Cell* » et je lui suis infiniment reconnaissante. Mes remerciements s'étendent également aux membres, actuels et anciens, de l'équipe Ghobrial, tels que Romanos Sklanvenitis-Pistofidis, Mark Bustoros, Ankit Dutta, Mike Agius, Jean-Baptiste Alberge, Elizabeth Lightbody, Cody Boehner, Mahshil Rahmat, Courtney Hamilton, Jihye Park, Tarek Mouhieddine, et les autres pour leur implication dans mon projet, pour la convivialité qu'ils ont instaurés dans leurs locaux, et pour ma pleine intégration à la vie du labo.

Je souhaite également à adresser mes remerciements au Professeur Gad Getz et à son équipe du BI du MIT pour avoir examiné mes données à plusieurs stades du projet. La possibilité de participer à leurs réunions d'équipe a été une expérience très formatrice. En particulier, un grand merci à Chip Stewart pour ses conseils inestimables sur l'utilisation des outils bioinformatiques et pour notre correspondance continue depuis mon retour en France.

Merci à Bénédicte pour m'avoir aidé à trouver un logement à Boston, ce qui a nécessité la visite trois appartements répartis dans la ville ! Sa gentillesse, sa générosité m'a permis de découvrir le meilleur de cette ville. Je retiens les week-ends touristiques où nous avons réussi à visiter quasiment tous les musées de la ville, les balades autour du lac de *Jamaica Plain*, la sortie théâtre et les après-midis-goûter au *Tatte bakery & Café*. Tu as vraiment contribué à rendre mon séjour au Massachussetts des plus agréables.

Un grand merci à l'ensemble du personnel de l'IRCL. Merci à Micheline Magdelon, Matthias Gerard, Corinne Merckx, et Michel Marissal pour leur gentillesse, leur bonne humeur, leur disponibilité, ainsi que pour les nombreuses discussions qui ont aidé à créer un environnement optimal pour la réalisation de ma thèse. Je tiens à exprimer mes remerciements également pour le Président de l'IRCL, Monsieur Pascal Boulanger, ainsi que les anciens et l'actuel Directeur de l'IRCL : le Professeur Pierre Formstecher, le Professeur Philippe Delannoy, et le Professeur Michel Salzet, pour le soutien qu'ils m'ont apporté.

Je remercie la directrice Isabelle Van Seuningen de l'unité Canther « Hétérogénéité, Plasticité et Résistance aux Thérapies des Cancers » pour son soutien. Je remercie également les autres membres des équipes de Canther pour leur investissement dans le bon fonctionnement du laboratoire.

Mes remerciements s'adressent également au Docteur Suman Mitra et à son équipe, pour leur écoute attentive et leurs nombreux conseils, à Silvia Gaggero pour sa générosité, à Tiziano Ingegnere pour son optimisme et à Malo Leprohon pour les bons moments partagés, il a de plus partagé, avec moi, la représentation de la bioinformatique dans nos équipes respectives. Je remercie aussi Leila Mohammadnezhad pour sa bonne humeur et son enthousiasme. Je salue Philippe

Mbabu et les autres membres du personnel de l'accueil d'OncoLille pour leur grande amabilité apportant une touche joyeuse à chaque journée.

Un remerciement spécial à Nathalie Jouy et Émilie Floquet de la plateforme de cytométrie pour leur bonne humeur et pour leur accueil dans leur bureau au tout début de la thèse.

À mes collègues et amies de la salle 12.29, Aurélie Guillemette, Clara Lewuillon, Sofia Titah, Marie-Océane Laguillaumie et Meriem Ben Khoud (sans oublier le beau petit Noah) avec qui nous avons refait le monde avec des discussions passionnantes autour de délicieux gâteaux. Même si notre rencontre a eu lieu durant la seconde partie de ma thèse, c'est avec un grand plaisir que j'ai fait votre connaissance. Mes remerciements vont également aux autres thésards qui ont terminé leur thèse avant moi, en particulier Raeeka Khamari et Quentin Fovez, pour leur gentillesse et leur bonne humeur. J'ai aimé les conversations échangées ensemble le temps d'une rencontre dans les couloirs. Je remercie aussi Paul Chauvet, l'un des co-auteurs de l'article de thèse, que j'ai rencontré en personne (trop) tardivement, et mon nouveau voisin de bureau pour sa gentillesse et nos conversations dans le bureau toujours amusantes, surprenantes et intéressantes. De même, j'aimerais remercier mon nouveau autre voisin de bureau bio-informaticien, Mojtaba Shekarkar. Je souhaitais également remercier les étudiants que j'ai croisés au fil de ma thèse et avec qui j'ai partagé d'agréables moments : Anisa, Benjamin, Émilie, Éva, Faruk, Louis, Mikail, Priscilla, Wenwen et Yanis. Je vous souhaite une bonne continuation dans la poursuite de vos études.

Un monumental merci à ma très chère collègue thésarde, Lama Hasan Bou Issa (« Hello my dear ! »), pour sa joie de vivre extraordinaire, son optimisme inébranlable, ses encouragements constants et pour ses magnifiques bracelets en résine dont l'un arbore mon nom ! J'ai été très heureuse de nos moments de solidarité durant les épreuves et nos fous rires et discussions animées resteront gravés dans ma mémoire. Travailler à tes côtes a été un véritable plaisir et j'espère sincèrement que nous garderons le contact.

Je tiens à exprimer ma reconnaissance envers François Delcroix, Marjorie Vandenhove, ainsi que le Professeur Patrick Vermersch, Directeur de l'école doctorale

EDBSL, pour leur précieuse aide et leur réactivité face aux imprévus rencontrés tout au long de ma thèse.

Mes remerciements vont également aux personnels hospitaliers, notamment la Professeure Marie-Christine Copin, le Docteur Romain Dubois, le Professeur Laurent Mortier, la Docteure Stéphanie Poulain, et Sylvie Janas de la tumorotheque pour leur confiance, leur énergie, et leur dévouement dans l'analyse des échantillons et des données cliniques.

Je souhaite exprimer ma gratitude envers l'équipe de biostatistiques du CHU de Lille, spécialement à Émeline Cailliau, pour son suivi et son travail biostatistique sur les données.

Un merci tout particulier aux anciens et actuels ingénieurs de l'équipe Bilille de l'Université de Lille, notamment à Jimmy Vandel, Clémentine Campart, Samuel Blanck, Jean-Pascal Meneboo et Marie. Je remercie également la Docteure Guillemette Marot, responsable scientifique de Bilille, pour son aide dans la recherche d'un sujet de thèse. Une triste pensée à mon ancien collègue et voisin de bureau, Maxime Brunin, avec qui j'ai partagé de nombreuses conversations enrichissantes et des repas conviviaux. Un merci chaleureux également tous les anciens et actuels membres de l'équipe Bonsai de CRISAL à Lille, en particulier le Professeur Jean-Stéphane Varré, pour m'avoir de nouveau donné goût au travail de chercheuse et pour m'avoir encouragé à débiter une thèse.

Je souhaite exprimer du fond du cœur ma reconnaissance envers ma famille, qui a été une présence essentielle tout au long de cette thèse. À mes parents, venus jusqu'à New-York pour célébrer mes 30 ans et à ma sœur, éternellement liée au mystère du quatrième crâne et au fameux « Dieu est mort ! », qui ont été présents à chaque étape de cette aventure. Mes pensées douces et affectueuses envers mes grands-parents, à ma grand-mère avec qui je partage une multitude de souvenirs heureux et à mon grand-père, ancien ingénieur des mines, qui aurait été fier d'avoir une petite-fille docteure en sciences.

Des remerciements affectueux vont à amis adorés de longue date, véritables piliers de mon parcours : Athénaïs Vaginay (Félicitation pour ton poste à Caen !), Alexandre Fruleux (Le grand chercheur lillois en biologie végétale !), Ambre (La

meilleure réalisatrice de la parodie de Star Wars), Angèle (Te voir remplie ma journée de joie), Anne-Charlotte (Merci pour ces séjours inoubliables passés), Ariane (Je t'envoie pleins d'énergies positives), Aymeric Antoine-Lorquin (« *Et de la Bretagne, il ne repartit jamais !* »), Émilie Allart (À nos futures beaux projets ensemble : le chant, la jungle et... le vélo), Stéphanie (Aux pompiers corses et à tes drifts maîtrisés dans les parkings) et son conjoint Maxime, Camille et Marc (Mon golden trio ave lon-lon, je n'aurais jamais réussi cette thèse sans vous trois pour me soutenir !), Marion Morel (À nos voyages à Boston et en Floride – et sa mare aux alligators ; et à nos futurs voyages !), Loïc Couderc (Le géomètre de génie de Poitier), Isabelle Guigon (À notre future bar à jeux, avec des chats et des papillons, et aux spécialités matcha !), Lieselot et sa fille Louenn, Maxime, Mylène et leur fille Léonce, Clotilde Lepers, Matt et leur fils Gaël. Adeline, son conjoint Axel et leur enfants Maxime et Eva, Théo (Le détective de ma future saga littéraire), Florian (le fidèle associé journaliste d'investigation), Pauline et Jérémy (À bientôt sur Rennes), Souhila (Merci pour tous ces bons repas ensemble et cette belle discussion sur Idir). Je tenais également à remercier mon amie Muriel qui a toujours été là pour moi, pour son soutien indéfectible et nos moments précieux. Merci également à Patrice et Nicole Nagel pour m'avoir soutenue tout au long de mes études.

Un merci particulier au groupe de théâtre universitaire de Géraldine qui m'a insufflé de la pêche, de la joie et de la vitalité dans la période post-confinement. À Arthur, Camille, Claire, Diana, Élise, Étienne, Esther, Jules, Laure, Lucile, Marco, Melvin, Missipssa, Samantha et Sylvain, votre énergie positive a rendu cette période plus agréable.

À Charles, Aude, Anne-Sophie et Cécile, un grand merci pour votre compassion, votre suivi durant des moments difficiles et votre aide précieuse qui m'a permis de trouver les ressources pour être résiliente.

Enfin, un éternel merci à Sophie Gallina, ma sensei de l'informatique, qui m'a fait découvrir le monde de la bioinformatique, elle m'a guidée tout au long de mes études et qui m'a inspirée une passion durable pour ce domaine.

Travaux Scientifiques

Articles

1. **Léa Fléchon***, Inès Arib*, Ankit K. Dutta*, Lama Hasan Bou Issa, Romanos Sklavenitis-Pistofidis, Remi Tilmont, Chip Stewart, Romain Dubois, Stéphanie Poulain, Marie-Christine Copin, Sahir Javed, Morgane Nudel, Doriane Cavalieri, Guillaume Escure, Nicolas Gower, Paul Chauvet, Nicolas Gazeau, Cynthia Saade, Marietou Binta Thiam, Aïcha Ouelkite-Oumouchal, Silvia Gaggero, Émeline Cailliau, Sarah Faiz, Olivier Carpentier, Nicolas Duployez, Thierry B. Idziorek, Laurent Mortier, Martin Figeac, Claude Preudhomme, Bruno Quesnel, Suman Mitra, Franck Morschhauser, Gad Getz, Irene M Ghobrial, Salomon Manier. Genomic profiling of Mycosis Fungoides identifies patients at high risk of disease progression. Blood Advances 2024; bloodadvances.2023012125. doi: <https://doi.org/10.1182/bloodadvances.2023012125>.
2. Lama Hasan Bou Issa*, **Léa Fléchon**, William Laine, Aïcha Ouelkdite, Silvia Gaggero, Adeline Cozzani, Rémi Tilmont, Paul Chauvet, Nicolas Gower, Romanos Sklavenitis-Pistofidis, Carine Brinster, Xavier Thuru, Yasmine Touil, Bruno Quesnel, Suman Mitra, Irene M. Ghobrial, Jérôme Kluza, Salomon Manier. MYC dependency in GLS1 and NAMPT is a therapeutic vulnerability in multiple myeloma. iScience. 2024 Mar 4;27(4):109417. doi: 10.1016/j.isci.2024.109417. PMID: 38510131; PMCID: PMC10952034.
3. Marie-Océane Laguillaumie*, Sofia Titah, Aurélie Guillemette, Bernadette Neve, Frederic Leprêtre, Pascaline Ségard, Faruk Azam Shaik, Dominique Collard, Jean-Claude Gerbedoen, **Léa Fléchon**, Lama Hasan Bou Issa, Audrey Vincent, Martin Figeac, Shéhérazade Sebda, Céline Villenet, Jérôme Kluza, William Laine, Isabelle Fournier, Jean-Pascal Gimeno, Maxence Wisztorski, Salomon Manier, Mehmet Cagatay Tarhan, Bruno Quesnel, Thierry Idziorek, Yasmine Touil. Deciphering genetic and nongenetic factors underlying tumour dormancy: insights from multiomics analysis of two syngeneic MRD models of melanoma and leukemia. Biological Research. 2024 Sep 3;57(1):59. doi: 10.1186/s40659-024-00540-y. PMID: 39223638, PMCID: PMC11370043.
4. Guillaume Escure*, Élise Fournier, Cynthia Saade, Lama Hasan Bou Issa, Inès Arib, Rémi Tilmont, Nicolas Gazeau, Binta M. Thiam, Morgane Chovet, Maxime Delforge, Nicolas Gower, **Léa Fléchon**, Doriane Cavalieri, Paul Chauvet, Morgane Nudel, Laure Goursaud, Céline Berthon, Bruno Quesnel, Thierry Facon, Claude Preudhomme, Nicolas Duployez, Salomon Manier. Small myeloid subclones are present at diagnosis of multiple myeloma in patients who develop secondary myelodysplastic syndromes. Haematologica 2024;109(4):1289-1292; <https://doi.org/10.3324/haematol.2023.284050>.

5. Romanos Sklavenitis-Pistofidis*, Michelle P. Aranha, Robert A. Redd, Joanna Baginska, Nicholas J. Haradhvala, Margaret Hallisey, Ankit K. Dutta, Alexandra Savell, Shohreh Varmeh, Daniel Heilpern-Mallory, Sylvia Ujwary, Oksana Zavidij, François Aguet, Nang K. Su, Elizabeth D. Lightbody, Mark Bustoros, Sabrin Tahri, Tarek H. Mouhieddine, Ting Wu, **Léa Fléchon**, Shankara Anand, Jacalyn M. Rosenblatt, Jeffrey Zonder, James J. Vredenburg, Adam Boruchov, Manisha Bhutani, Saad Z. Usmani, Jeffrey Matous, Andrew J. Yee, Andrzej Jakubowiak, Jacob Laubach, Salomon Manier, Omar Nadeem, Paul Richardson, Ashraf Z. Badros, Maria-Victoria Mateos, Lorenzo Trippa, Gad Getz, Irene M. Ghobrial. Immune biomarkers of response to immunotherapy in patients with high-risk smoldering myeloma. *Cancer Cell*. 2022 Nov 14;40(11):1358-1373.e8. doi: 10.1016/j.ccell.2022.10.017. PMID: 36379208; PMCID: PMC10019228.

Présentation Orales

- « Study of the Clonal Evolution in Multiple Myeloma WGS samples », Getz-Ghobrial meeting in Broad Institute (2019), Cambridge, États-Unis (présentation orale).
- « Introduction to bash (Linux) training » in Dana Farber Cancer Institute (2020), Boston, États-Unis (présentation orale).
- « Molecular Characterization and Clonal Evolution of Mycosis Fungoides », European Hematology Association (2022), Vienne, Autriche (présentation poster).

Prix

- Bourse Fondation I-SITE ULNE de mobilité internationale de recherche : jeunes chercheur.e.s et doctorant.e.s (2019).
- EHA2022 travel grant (2022).

Table des Matières

Remerciements	2
Travaux Scientifiques	8
Table des Matières	10
Liste des Figures	13
Liste des Tableaux	15
Liste des Abréviations.....	16
Résumé	18
Abstract	20
1 Introduction	22
1.1 Lymphomes T Cutanés Primitifs	24
1.1.1 Lymphomes Cutanés Primitifs	24
1.1.2 Classification.....	25
1.2 Mycosis Fongoïde	26
1.2.1 Historique et Description.....	26
1.2.2 Mycosis Fongoïde Classique	26
1.2.2.1 Manifestations Cutanées	26
1.2.2.2 Histologie.....	27
1.2.2.3 Diagnostic.....	29
1.2.2.4 Classification TNMB	31
1.2.2.5 Évolution de la Pathologie	33
1.2.3 Formes Variantes du Mycosis Fongoïde.....	33
1.2.3.1 Mycosis Fongoïde Folliculotrope	34
1.2.3.2 Réticulose Pagétoïde	34
1.2.3.3 Chalazodermie Granulomateuse	34
1.2.4 Mycosis Fongoïde Transformé et Syndrome de Sézary	35
1.2.4.1 Cellule de Mycosis Fongoïde Transformée	35
1.2.4.2 Infiltrats de Lymphocytes B	36
1.2.4.3 Cellule de Sézary	37
1.2.4.4 Distinctions Mycosis Fongoïde et Syndrome de Sézary	38
1.2.5 Pronostic.....	40
1.2.5.1 Mycosis Fongoïde Classique.....	40
1.2.5.2 Mycosis Fongoïde Folliculotrope	41
1.2.5.3 Mycosis Fongoïde Transformé	41
1.2.6 Traitements.....	42
1.3 Physiopathologie.....	43
1.3.1 Mutations Somatiques	43
1.3.2 Mutations <i>Drivers</i> , <i>Passengers</i> et Biomarqueurs	44
1.3.3 Variant Nucléotidique Simple	44
1.3.4 Variation du Nombre de Copies Somatiques	47
1.3.5 Voies de Signalisation.....	48
1.3.6 Évolution Clonale	51
1.3.6.1 Architecture Clonale	51
1.3.6.2 Hétérogénéité Clonale.....	52
1.3.6.3 Clonalité des Lymphocytes T dans le MF	53
1.3.7 Signatures Mutationnelles.....	54

1.3.8 Épigénétique	55
2 Objectifs de la Thèse.....	59
2.1 Identification des Biomarqueurs Impliqués dans la Progression du MF	60
3 Résultat	62
3.1 Échantillons Tumoraux, Normaux et Données Cliniques	64
3.1.1 Patients et Échantillons.....	64
3.1.2 Panel d'Échantillons Normaux	66
3.1.3 Données Cliniques.....	67
3.2 Séquençage, Alignement, Contrôle de Qualité et Contamination	69
3.2.1 Construction des Bibliothèques et Séquençage	69
3.2.2 Plateforme Terra du Broad Institute et Langage de Workflow	70
3.2.3 Contrôle de Qualité et Alignement	71
3.2.4 Contamination.....	77
3.2.4.1 Contamination entre Échantillons	77
3.2.4.2 Inversion des Échantillons	78
3.3 Identification des Marqueurs de Haut Risque de Progression	81
3.3.1 Détection des SCNVS	81
3.3.2 Détection des SNVs et des InDels	82
3.3.3 Filtrage avec les PoNs	84
3.3.4 Contamination <i>Tumor in Normal</i>	85
3.3.5 Annotation des Mutations Somatiques	85
3.3.6 Identification des Gènes Fréquemment Mutés	86
3.3.7 Retrait des <i>Fishy Genes</i>	87
3.3.8 Estimation des Signatures Mutationnelles	88
3.4 Pureté, Ploïdie et Fraction de Cellules Cancéreuses et Filtres Internes	89
3.4.1 Estimation de la Pureté, la Ploïdie et la Fraction de Cellules Cancéreuses	89
3.4.2 Filtres Internes	90
3.4.2.1 Filtrage avec GnomAD	90
3.4.2.2 Filtrage avec les Fichiers FASTQ Normaux	90
3.5 Analyses Statistiques	91
3.6 Paysage Génomique.....	92
3.6.1 Mutations Nucléotidiques Somatiques Putatives	92
3.6.1.1 Nombre de Mutations Somatiques à Chaque Étape du Pipeline	92
3.6.1.2 Charge Tumorale Mutationnelle	95
3.6.1.3 Gènes Fréquemment Mutés	97
3.6.2 Variations du Nombre de Copies Somatiques Récurrentes	101
3.6.2.1 Identification des Altérations Significativement Récurrentes	101
3.6.2.2 Gènes <i>Drivers</i> Identifiés dans les SCNVS Focaux.....	102
3.6.2.3 État des connaissances des Bandes Cytogénétiques dans le MF	104
3.6.3 Signatures Mutationnelles.....	108
3.6.4 Biomarqueurs Impactant la Survie Globale et la Progression.....	112
3.6.4.1 Survie Globale de la Cohorte	112
3.6.4.2 Association entre Biomarqueurs avec Stade et Pronostic	113
3.6.5 L'Architecture Clonale et la Phylogénie du MF	118
3.6.6 Modifications des Voies de Signalisations	123

4 Discussion	125
4.1 Cohorte Mycosis Fongoïde	126
4.1.1 Vue d'Ensemble sur la Cohorte	126
4.1.2 Échantillons <i>Tumor-Only</i> et <i>Tumor-Normal</i>	127
4.1.3 Cohorte PoN de Dix Échantillons de Sang	128
4.1.4 Amélioration de la Cohorte	129
4.2 Séquençage	130
4.2.1 Whole Exome Sequencing et Whole Génome Séquençage	130
4.2.2 Whole Exome Sequencing et le Séquençage Long-Read	132
4.2.3 Régions de Faible Complexité	133
4.2.4 Nombre de Lectures Séquencées et Profondeur de Séquençage	133
4.3 Filtres Supplémentaires dans la Réduction de Faux Positifs	135
4.3.1 Panel of Normal du Broad Institute et Filtre <i>Tumor-Only</i>	135
4.3.2 Retrait des Variants Synonymes	136
4.3.3 Filtres Internes	136
4.4 Paysage mutationnel	138
4.4.1 Survie Globale	138
4.4.2 Évolution Clonale	140
4.4.3 Épigénétique	141
4.4.4 Signatures Mutationnelles	142
4.4.5 Potentiels Biomarqueurs Identifiés	144
4.5 Perspectives	145
5 Conclusion	147
Références	151
6 ANNEXE : Articles	174

Liste des Figures

Figure 1 : Répartition des LCPs selon le type (T/NK ou B).....	24
Figure 2 : Différentes lésions cutanées provoquées par le MF	26
Figure 3 : Stades précoces du MF.....	29
Figure 4 : Histologies de MF avancés montrant plus de 25% de grandes cellules ..	36
Figure 5 : Cellules de Sézary de petite à moyenne taille	37
Figure 6 : Sous-types de lymphocytes T circulant dans la peau	39
Figure 7 : Exemples de SNVs et InDels affectant la séquence nucléotidique.....	45
Figure 8 : Différents types de SCNVs sur un bras chromosomique.....	47
Figure 9 : Gènes <i>drivers</i> connus et voies de signalisations chez les CTCLs.....	50
Figure 10 : Composition de la cohorte de patients	64
Figure 11 : Répartition des groupes de patients et des échantillons	66
Figure 12 : Échantillons tumoraux et normaux sélectionnés pour le plan d'étude ...	67
Figure 13 : Page de configuration des méthodes sur la plateforme FireCloud	71
Figure 14 : Intégration de l'outil CutAdapt dans la section QC du pipeline	72
Figure 15 : Q Score sur l'intégralité des lectures séquencées.....	74
Figure 16 : Alignement des lectures séquencées	75
Figure 17 : Profondeur de couverture des échantillons MF après séquençage.....	76
Figure 18 : Matrice d'empreinte des échantillons de MF	81
Figure 19 : Nombre de mutations somatiques par échantillon après chaque étape	92
Figure 20 : Corrélation entre la TMB et le stade du MF et le genre du patient	96
Figure 21 : Paysage des altérations génomiques les plus récurrentes dans le MF .	98
Figure 22 : Lollipop plots des mutations non-synonymes pour dix gènes.....	99
Figure 23 : Diagramme de mutation pour les gènes <i>EPB41L3</i> , <i>MAGI1</i> et <i>PRKCB</i>	101
Figure 24 : SCNVs récurrentes dans le MF	102

Figure 25 : Extraction des signatures mutationnelles	109
Figure 26 : Signatures mutationnelles identifiées dans les échantillons de MF	111
Figure 27 : Corrélation entre les signatures mutationnelles et le stade du MF	112
Figure 28 : Courbe Kaplan-Meier pour l'analyse de l'OS entre les patients LR et HR chez les 48 patients au diagnostic.....	113
Figure 29 : Corrélation entre les événements génomiques et le stade du MF	114
Figure 30 : Association du pronostic avec les différents facteurs génétiques.....	116
Figure 31: Analyses univariées et multivariées des biomarqueurs du MF	117
Figure 32 : Hétérogénéité clonale dans le MF	118
Figure 33 : Co-occurrences et des exclusions mutuelles des altérations génomiques récurrentes chez les 48 patients atteints de MF	119
Figure 34 : Chronologie inférée des altérations génétiques dans le MF	120
Figure 35 : Évolution clonale des échantillons séquentiels.....	122
Figure 36 : Altérations récurrentes affectant les principales voies de signalisation	123

Liste des Tableaux

Tableau 1 : Classification OMS-EORTC 2018 des LCPs T/NK et B	25
Tableau 2 : Panel d'anticorps pour l'analyse immunohistologique du MF	28
Tableau 3 : Critères pour le diagnostic précoce du MF.	30
Tableau 4 : Classification TMNB des différents stades du MF selon l'ISCL, l'USCLC et l'EORTC	32
Tableau 5 : Publications rapportant des mutations SNVs dans les CTCLs, en particulier le MF, et les gènes impactés.	46
Tableau 6 : Informations cliniques et démographiques des patients étudiés atteints d'un MF diagnostiqué	68

Liste des Abréviations

0-9

18F-FDG : Glucose marqué par le radio-nucléide émetteur de positrons ^{18}F

A

AF : *Allelic frequency*

B

BAM : Binary Alignment Map

BdD : Base de données

BI : Broad Institute

Bp : Base pair

C

CBCL : *Cutaneous B-Cell Lymphomas*

CBP : Centre Biologie Pathologie de Lille.

CCF : *Cancer Cell Fraction*

chr : Chromosome

CHU : Centre Hospitalier Universitaire

CN : *Copy-Number*

COSMIC : *Catalogue of Somatic Mutations in Cancer*

CTCL : Cutaneous T-cell lymphoma

Cytoband : bande qui décrit la localisation approximative chromosomique

D

del : Délétion d'un locus ou d'un bras de chromosome

DFCI : *Dana-Farber Cancer Institute*

Driver : Gène conducteur

E

EORTC : European Organization for Research and Treatment of Cancer

F

FDR : *False Discovery Rate*

FISH : Fluorescence in situ hybridization

fMF : folliculotropic MF

FP : Faux positif

G

gain : Gain d'un locus ou d'un bras de chromosome

gnomAD : Genome Aggregation Database

H

hg19 : Génome de référence humain 19 proposé par l'Université de Californie à Santa Cruz.

HGVS : Human Genome Variation Society

HR : *High-Risk*

I

IFN : Interferon

Ig : Immunoglobuline

IHC : Immunohistochimie

InDel : Insertion-Délétion

IRB : Institutional Review Board

IRM : Imagerie par Résonance Magnétique

ISCL : *International Society for Cutaneous Lym-phoma*

L

LCP : Lymphome cutané primitif

LOD : *Logarithm of odds*

LR : *Low-Risk*

LR progressor : LR qui progresse ensuite en HR

M

M : Million

MAF : *Mutation Annotation Format*

Mb : Mégabase

MF : Mycosis Fongoïde

MSI : *Microsatellite instability*

mOS : *median of Overall Survival*

N

NCBI : National Center for Biotechnology Information

NGS : Next Generation Sequencing

NK : Natural Killer cell

NMF : Non-negative Matrix Factorization

NR : median Not Reached

O

OMS : Organisation Mondiale de la Santé

OR : Odds Ratio

OS : Overall Survival

8-oxoG : Oxidation of guanine to 8-oxo-guanine contamination

P

P : Probabilité d'erreur du séquenceur dans l'appel de la base

PCR : Polymerase Chain Reaction

PD-1 : Programmed Death-1

PET/CT : Positron Emission Tomography/Computed Tomography

PF : Pass Filters

PFS : Progression-Free Survival

PoN : Panel of Normal

Protéine M : Protéine mono-clonale

PUVA : Psoralène et UVA

Q

QC : Quality Control

RT-PCR : Reverse Transcription PCR

Q score : Score de qualité Phred pour l'appel de la base par le séquenceur

R

RNASeq : RNA-Sequencing

S

SBS : Single Base Substitution signature dans COSMIC

SCNV : Somatic Copy-Number Variation

SNP : Single-Nucleotide Polymorphism

SNV : Single-Nucleotide Variant

SS : Syndrome de Sézary

T

TCGA : The Cancer Genome Atlas

TCR : T-Cell Receptor

TF : Transcription Factor

TiN : Tumor-in-Normal

TMB : Tumor Mutational Burden

tMF : transformed MF

TN : Échantillon tumoral apparié à un échantillon normal (*Tumor-Normal*)

TNMB : Classification clinique du MF (*Tumor-lymph Node-Metatase-Blood*)

TO : Échantillon tumoral non apparié à un échantillon normal (*Tumor-Only*)

TSEB : total-skin electron-beam radiation therapy

TTP : Time to progression

U

USCLC : United States Cutaneous Lymphoma Consortium

UV : Ultraviolet

V

VAF : Variant Allele Frequency

VCF : Variant Call Format

W

WES : Whole Exome Sequencing

WGS : Whole Genome Sequencing

X

X : Profondeur de séquençage

Résumé

Titre : Identification des Marqueurs Génomiques de Haut Risque de Progression du Mycosis Fongoïde.

Les hémopathies lymphoïdes sont des cancers qui se développent à partir de cellules lymphocytaires du système immunitaire. Le lymphome non hodgkinien fait partie des hémopathies lymphoïdes les plus fréquentes. L'histoire naturelle de ces cancers est souvent constituée de stades précurseurs évoluant vers des stades plus agressifs. Les facteurs pronostiques sont essentiels en clinique afin de stratifier le risque évolutif des patients. Le profilage génomique de ces tumeurs à partir de séquençage exomique complet (WES) offre ainsi une approche intéressante pour identifier les patients présentant un haut risque (HR) de progression. En effet, la détection d'anomalies génomiques, dont les mutations de variants nucléotidiques simples (SNVs) et les variations somatiques du nombre de copies (SCNVs), permettent d'analyser l'hétérogénéité et l'évolution clonales de nombreux cancers. Mes travaux de thèse portent sur l'identification des marqueurs génomiques indiquant un HR de progression dans le mycosis fongoïde (MF).

Mon travail porte sur les mécanismes moléculaires impliqués dans la progression du MF, le lymphome T cutané primitif le plus fréquent. Celui-ci évolue souvent de stades peu avancés vers des formes agressives de la maladie. Les anomalies génomiques responsables de cette progression sont encore peu connues. L'objectif de ce travail était de caractériser le profil génétique du MF par le séquençage WES de 67 échantillons tumoraux provenant de 48 patients du Centre Hospitalier Universitaire (CHU) de Lille. L'analyse visait à identifier les SNVs et les SCNVs récurrents afin de définir les signatures mutationnelles du MF et à déterminer l'hétérogénéité et l'évolution clonale à partir de dix-huit échantillons séquentiels prélevés à différents stades de la maladie. Les données génomiques ont été analysées sur la plateforme bioinformatique Terra du Broad Institute (BI). Les résultats ont montré que les mutations somatiques les plus fréquentes au moment du diagnostic étaient *JUNB*, *TET2* et *MAPK1*. Nous avons constaté que le gain7q, le gain10p15.1 (*IL2RA* et *IL15RA*), la del10p11.22 (*ZEB1*) ou les mutations de *JUNB* et *TET2* sont associés à des stades avancés de la maladie, suggérant leur rôle dans la progression du MF.

De plus, le gain7q, le gain10p15.1 (*IL2RA* et *IL15RA*), la del10p11.22 (*ZEB1*) et la del6q16.3 (*TNFAIP3*) sont associés à un mauvais pronostic, avec une survie globale plus courte. Le del6q16.3 (*TNFAIP3*) est également un facteur de risque de progression chez les patients présentant un stade précoce de la maladie. L'analyse de l'hétérogénéité et de l'évolution clonale de la cohorte nous a permis de définir plusieurs profils phylogéniques du MF, avec l'acquisition de *JUNB*, du gain10p15.1 (*IL2RA* et *IL15RA*) ou de la del12p13.1 (*CDKN1B*) lors de la progression. Ces résultats fournissent des caractéristiques génomiques et clonales afin d'aider à identifier des patients avec un risque de progression du MF.

Mots clés : Mycosis Fongoïde, *Whole-Exome Sequencing*, Pipeline Bioinformatique, Mutations Somatiques, Évolution Clonale.

Abstract

Title: Identification of Genomic Markers of High-Risk Progression in Mycosis Fungoides.

Lymphoid hemopathies are cancers arising from lymphocytic cells of the immune system. Non-Hodgkin's lymphoma is among the most common of these malignancies. The natural history often consists of precursor stages progressing to more aggressive stages. Prognostic factors are essential in clinical practice to stratify patients' risk of progression. Genomic profiling of these tumors using whole-exome sequencing (WES) thus offers an interesting approach to identifying patients at high risk (HR) of progression. Indeed, the detection of genomic abnormalities, such as single nucleotide variant mutations (SNVs) and somatic copy number variations (SCNVs), makes it possible to analyze the clonal heterogeneity and evolution of many cancers. My PhD work focuses on the identification of genomic markers indicative of HR of progression in two lymphoid hemopathies: mycosis fungoides (MF).

My work focuses on the molecular mechanisms involved in the progression of MF, the most common primary cutaneous T-cell lymphoma. It often progresses from early stages to aggressive forms of the disease. Little is known about the genomic abnormalities responsible for this progression. The aim of this study was to characterize the genetic profile of MF by WES sequencing of 67 tumor samples from 48 patients from the Lille University Hospital (CHU). The analysis aimed to identify recurrent SNVs and SCNVs in order to define the mutational signatures of MF, and also determined heterogeneity and clonal evolution from 18 sequential samples taken at different stages of the disease. Genomic data were analyzed on the Broad Institute's (BI) Terra bioinformatics platform. Results showed that the most frequent somatic mutations at diagnosis were: *JUNB*, *TET2* and *MAPK1*. We found that gain7q, gain10p15.1 (*IL2RA* and *IL15RA*), del10p11.22 (*ZEB1*) or mutations in *JUNB* and *TET2* are associated with advanced disease stages, suggesting their role in MF progression. Furthermore, the gain7q, gain10p15.1 (*IL2RA* and *IL15RA*), del10p11.22 (*ZEB1*) and del6q16.3 (*TNFAIP3*) are associated with a poor prognosis, with shorter overall survival. The del6q16.3 (*TNFAIP3*) is also a risk factor for progression in patients with early-stages disease. Analysis of the heterogeneity and clonal evolution

of the cohort enabled us to define several phylogenic profiles of MF, with the acquisition of *JUNB*, gain10p15.1 (*IL2RA* and *IL15RA*) or del12p13.1 (*CDKN1B*) during progression. These results provide genomic and clonal features to help identify patient at risk of MF progression.

Keywords: Mycosis Fungoides, Whole-Exome Sequencing, Bioinformatics Pipeline, Somatic Mutations, Clonal Evolution.

Chapitre 1

Introduction

Sommaire

1.1	Lymphomes T Cutanés Primitifs	24
1.1.1	Lymphomes Cutanés Primitifs	24
1.1.2	Classification.....	25
1.2	Mycosis Fongoïde	26
1.2.1	Historique et Description	26
1.2.2	Mycosis Fongoïde Classique	26
1.2.2.1	Manifestations Cutanées	26
1.2.2.2	Histologie.....	27
1.2.2.3	Diagnostic.....	29
1.2.2.4	Classification TNMB	31
1.2.2.5	Évolution de la Pathologie	33
1.2.3	Formes Variantes du Mycosis Fongoïde	33
1.2.3.1	Mycosis Fongoïde Folliculotrope	34
1.2.3.2	Réticulose Pagétoïde	34
1.2.3.3	Chalazodermie Granulomateuse	34
1.2.4	Mycosis Fongoïde Transformé et Syndrome de Sézary	35
1.2.4.1	Cellule de Mycosis Fongoïde Transformée	35
1.2.4.2	Infiltrats de Lymphocytes B	36
1.2.4.3	Cellule de Sézary	37
1.2.4.4	Distinctions Mycosis Fongoïde et Syndrome de Sézary	38
1.2.5	Pronostic.....	40
1.2.5.1	Mycosis Fongoïde Classique.....	40
1.2.5.2	Mycosis Fongoïde Folliculotrope	41
1.2.5.3	Mycosis Fongoïde Transformé	41
1.2.6	Traitements.....	42
1.3	Physiopathologie.....	43
1.3.1	Mutations Somatiques	43
1.3.2	Mutations <i>Drivers</i> , <i>Passengers</i> et Biomarqueurs	44
1.3.3	Variant Nucléotidique Simple	44

1.3.4	Variation du Nombre de Copies Somatiques	47
1.3.5	Voies de Signalisation	48
1.3.6	Évolution Clonale	51
1.3.6.1	Architecture Clonale	51
1.3.6.2	Hétérogénéité Clonale	52
1.3.6.3	Clonalité des Lymphocytes T dans le MF	53
1.3.7	Signatures Mutationnelles	54
1.3.8	Épigénétique	55

1.1 Lymphomes T Cutanés Primitifs

1.1.1 Lymphomes Cutanés Primitifs

Les lymphomes cutanés primitifs (LCPs) constituent un groupe hétérogène de lymphomes non hodgkiniens qui se développent à partir de lymphocytes cancéreux au niveau de la peau. Ils sont prédominants chez les hommes, avec un âge médian de 60 ans au diagnostic¹. Contrairement à la majorité des lymphomes non hodgkiniens, qui trouvent leur origine dans les ganglions lymphatiques, les LCPs proviennent de proliférations de lymphocytes clonaux dans des lésions cutanées^{2,3}. Les LCPs se divisent en deux catégories principales : les lymphomes T cutanés primitifs (CTCLs pour *Cutaneous T-Cell lymphoma*), qui représentent entre 75 et 80% des cas, et les lymphomes B cutanés primitifs (CBCLs pour *Cutaneous B-Cell lymphoma*), qui comptent pour environ 20 à 25% des cas (Figure 1)¹. Plus rarement, les LCPs peuvent également provenir des lymphocytes *Natural Killer* (NK). Les CTCLs sont considérés comme des maladies rares avec un taux d'incidence en France de 0,96 pour 100 000 cas en 2018⁴.

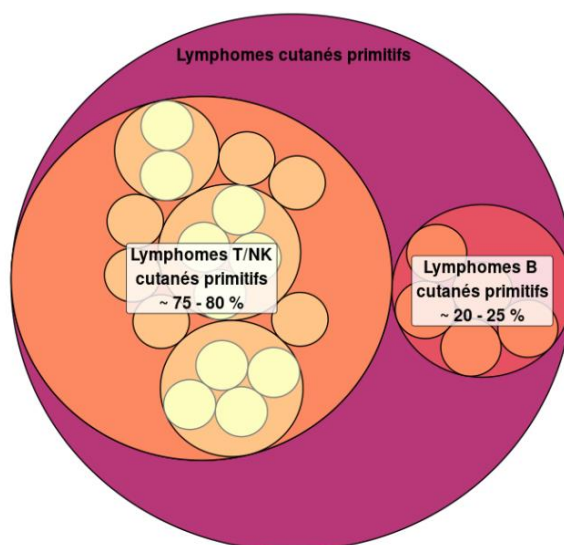


Figure 1 : Répartition des LCPs selon le type (T/NK ou B). Les CTCLs sont plus nombreux et plus fréquemment rencontrés que les CBCLs. Les cercles à l'intérieur de chaque catégorie indiquent le nombre de lymphomes cutanés décrits. Les noms des lymphomes sont détaillés dans le Tableau 1.

1.1.2 Classification

Les CTCLs constituent un groupe cliniquement varié de lymphomes incurables qui concerne les lymphocytes T matures résidant dans la peau. Une première classification a été établie en 2005 par l'Organisation Mondiale de la Santé (OMS) et de l'Organisation Européenne pour la Recherche et le Traitement du Cancer (EORTC)¹. Cette classification a connu quatre révisions, dont la dernière a été publiée en 2018³ (Tableau 1) avec une mise-à-jour récente en 2022^{5,6}. Les formes les plus fréquentes sont les CTCLs de faible malignité, tels que le mycosis fongoïde (MF), les lymphomes anaplasiques cutanés primitifs à grandes cellules (pcALCL) et le Syndrome de Sézary (SS), par rapport aux formes plus graves et plus rares, comme les lymphomes T/NK extranodaux et les CTCLs γ/δ .

Classification	Fréquence (%)	Survie spécifique à 5 ans (%)
Lymphomes T et NK cutanés primitifs		
Mycosis fongoïde (MF)	39	88
Variants de MF :		
- MF folliculotrope	5	75
- Réticulose pagétoïde	<1	100
- Chalazodermie granulomateuse	<1	100
Syndrome de Sézary	2	36
Lymphome/leucémie T de l'adulte	<1	Pas de données
Lymphoproliférations T cutanées CD30+ :		
- Lymphome T anaplasique cutané primitif	8	95
- Papulose lymphomatoïde	12	99
Lymphome T sous-cutané type panniculite	1	87
Lymphome T/NK extranodal	<1	16
Maladie systémique chronique active EBV+	<1	Pas de données
Lymphome T périphériques cutanées primitifs, sous-types rares		
- Lymphome T γ/δ cutané primitif	<1	11
- Lymphome T cutané primitif CD8+ épidermotrope cytotoxique agressif	<1	31
- Lymphoprolifération T cutanée primitive CD4+ à petites ou moyennes cellules	6	100
- Lymphome T acral CD8+ cutané primitif	<1	100
Lymphome T périphérique cutané primitif, sans précision	2	15
Lymphomes B cutanés primitifs		
Lymphome de la zone marginal cutané primitif	9	99
Lymphome centro-folliculaire cutané primitif	12	95
Lymphome B diffus à grandes cellules cutané primitif, type jambe	4	56
Ulcère cutané-muqueux EBV+	<1	100
Lymphome B à grandes cellules intravasculaire	<1	72

Tableau 1 : Classification OMS-EORTC 2018 des LCPs T/NK et B, (d'après la quatrième édition de 2018³ et la cinquième édition de 2022⁶). EBV pour Epstein–Barr virus.

1.2 Mycosis Fongoïde

1.2.1 Historique et Description

Décrit pour la première fois en 1806 par le dermatologue français Jean-Louis Alibert⁷, le MF illustre la représentation clinique des tumeurs dont la texture se rapprocherait de celle d'un champignon.

Le MF est le LCP le plus commun, représentant plus de 50% de tous les cas de CTCLs⁸. Son incidence est estimée à en moyenne 0,66 cas pour 100 000 habitants par an aux États-Unis⁹. Cette maladie affecte principalement la peau dans les zones du corps protégées du soleil durant les stades précoces chez les individus de plus de 55 ans^{1,10}. Il y a une prédominance masculine avec un sex-ratio de 2:1¹.

1.2.2 Mycosis Fongoïde Classique

1.2.2.1 Manifestations Cutanées

Les premiers stades du MF classique se manifestent par des lésions cutanées sous forme de taches appelées « macules » ou de plaques rougeâtres appelées « papules ». La taille et la forme des macules et des papules sont variables. Les stades avancés sont caractérisés par le développement de tumeurs, d'ulcérations et souvent d'une érythrodermie, une coloration généralisée de la peau en rouge (Figure 2).

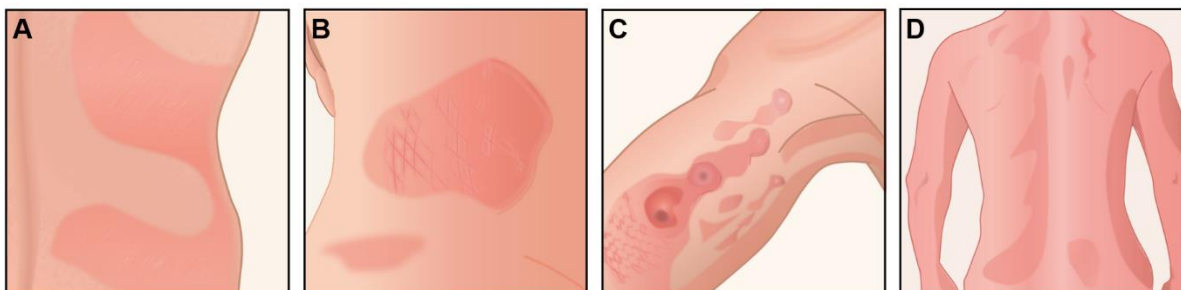


Figure 2 : Différentes lésions cutanées provoquées par le MF. A. Les taches ou macules cutanées modifient la coloration de la peau (rouge ou rose) sans changer le relief. **B.** Les plaques ou papules cutanées sont généralement surélevées, rougeâtres et provoquent des démangeaisons. **C.** Les tumeurs se présentent sous la forme de nodules, de petites masses

bien délimitées au niveau de la peau avec des lésions ouvertes ou fermées. **D.**
L'érythrodermie se caractérise par la coloration rouge sur une grande surface corporelle (~90%) accompagnée d'une desquamation des couches superficielles de l'épiderme.

Illustrations adaptées à partir des illustrations scientifiques d'Onyx Visual.

Ces quatre catégories de lésions morphologiques (macules, papules et tumeurs, érythrodermie) reflètent généralement bien l'évolution de la maladie. La présence de macules et de papules très délimitées sur la peau est généralement associée à un MF précoce. Dans le MF avancé, les tumeurs peuvent être localisées ou généralisées et sont souvent associées à des papules. Dans certains cas, des tumeurs plates et ulcérées peuvent être confondues avec des papules.

1.2.2.2 Histologie

L'association entre le MF et le lymphocyte T a été confirmée par l'étude immunologique de Lutzner et de son équipe en 1975¹¹. L'histologie classique s'avère être un infiltrat de lymphocytes T mémoires effecteurs CD2⁺, CD3⁺, CD4⁺, CD5⁺, CD8⁻, CD45RO⁺ atypiques de petite à moyenne taille¹ (Tableau 2). Les lymphocytes T tumoraux, présents dans le tissu cutané, expriment les marqueurs *C-C chemokine Receptor 4* (CCR4, CD194) et *Cutaneous Lymphocyte Antigen* (CLA) avec une migration entre les couches du derme et de l'épiderme (épidermotropisme) (Figure 3A).

Anticorps	Immunomarquage et/ou spécificité
CD2/CD3	Marqueur des lymphocytes T.
CD4	Positif dans les lymphocytes T auxiliaires et les lymphomes à cellules T apparentés.
CD5	Marqueur des lymphocytes T. Également positif chez les lymphocytes B dans certains lymphomes/leucémies propres aux lymphocytes B.
CD7	Marqueur des lymphocytes T. Perte d'expression observée dans certains CTCLs et des dermatoses inflammatoires, ce qui ne constitue pas un critère de diagnostic pour les CTCLs.
CD8	Positif dans les lymphocytes T cytotoxiques.

CD20	Positif dans les lymphocytes B. La combinaison CD20/PAX5 identifie les lymphocytes B dans le MF.
CD25	Récepteur de l'IL-2, exprimé sur les lymphocytes activés. Positif dans le lymphome ATLL et dans certains cas de MF ¹² .
CD30	Lymphocytes T et B. Positif dans un groupe de CTCLs comprenant l'ALCL et la LyP. Un nombre variable de cellules positives dans la plupart des troubles lymphoprolifératifs cutanés.
CD45RO	Lymphocytes T mémoires.
CD56	Positif dans les cellules NK et un marqueur pour la majorité des lymphomes NK/T extranodal.
CCR4	Récepteur de la chimiokine C-C de type 4. S'exprime notamment dans les lymphocytes T tumoraux de la peau et du sang périphérique des patients atteints de MF/SS.
Ki-67	Marqueur de prolifération. Utile pour détecter la faible prolifération typique du MF au stade précoce. Proposé avec p53 comme marqueurs pour identifier les cellules tMF ¹³ .
KIR3DL2	Récepteur <i>killer immunoglobuline-like</i> normalement exprimé de façon aberrante dans le tMF et le SS par des NKs et une minorité de lymphocytes T.
p53	Suppresseur de tumeur décrit dans une grande variété de carcinomes, de sarcomes et de leucémies/lymphomes. Proposé avec Ki-67 pour identifier les cellules tMF ¹³ .
PAX5	<i>Paired box gene 5</i> , lymphocytes B immatures et matures. La combinaison CD20/PAX5 identifie les lymphocytes B dans le MF ¹² .
PD-1	Positif dans les lymphocytes T auxiliaires folliculaires ainsi que dans un sous-ensemble de lymphocytes activés.

Tableau 2 : Panel d'anticorps pour l'analyse immunohistologique du MF (d'après Cerroni, Édition Wiley-Blackwell, 2020). ALCL pour *Anaplastic Large-Cell Lymphoma*, ATLL pour *Adult T-cell Leukemia/Lymphoma*, IHC pour Immunohistochimie, LyP pour *Lymphomatoid Papulosis*, tMF pour *transformed MF* et SS pour *Sézary Syndrome*.

Une minorité de cas présente un phénotype CD4⁺/CD8⁺ sans montrer de différences cliniques ou pronostiques significatives^{14,15}. Plus rarement, des cas CD4⁺ / CD8⁺ ont été observés exprimant *Programmed Death-1* (PD-1, CD279), encodé par le gène *PDCD1*^{16,17}. PD-1 est un inhibiteur de l'activation et de la prolifération des lymphocytes T, il module ainsi l'activité immunitaire en interagissant avec ses deux ligands PD-L1 et PD-L2. Inhabituellement, des cas présentant un phénotype cytotoxique CD4⁺ ou CD8⁺ peuvent exprimer le marqueur NK CD56^{18,19} (Tableau 2).

Les cellules MF du derme ont la capacité de migrer vers le compartiment épidermique, formant des petites cavités caractéristiques de la maladie appelées « nids de Darier » ou « micro-abcès de Pautrier »²⁰ (Figure 3B).

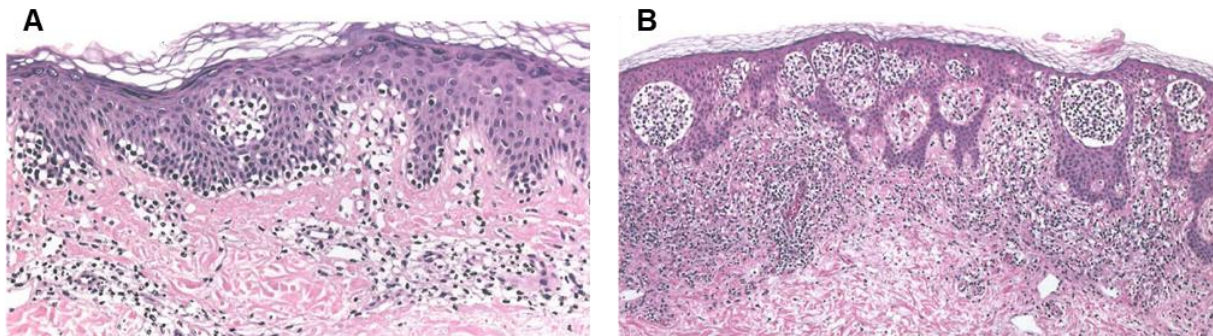


Figure 3 : Stades précoces du MF. **A.** Lymphocytes épidermotropes alignés le long de la couche basale (épidermotropisme basilaire). **B.** Infiltrats en forme de bande de lymphocytes épidermotropes avec présence de cavités appelées micro-abcès de Pautrier. Photographies des coupes histologiques issues de *WHO classification of skin tumours*, 4ème édition (2018), IARC²¹.

1.2.2.3 Diagnostic

Le diagnostic du MF peut s'avérer complexe, car les premiers symptômes peuvent ressembler à d'autres dermatoses inflammatoires d'origine indéterminée^{22,23}. Les lésions ne sont pas toujours caractéristiques du MF lors des stades initiaux de la maladie, ou être peu visibles qu'après l'application de traitements topiques, comme les dermocorticoïdes. Le diagnostic histologique peut également être difficile, notamment pendant la phase érythrodermique où les cellules inflammatoires sont prédominantes, nécessitant parfois plusieurs biopsies. Dans de nombreux cas, le diagnostic définitif requiert l'analyse minutieuse des caractéristiques cliniques, histologiques et biologiques de la maladie¹⁴.

En 2005, l'*International Society for Cutaneous Lymphoma* (ISCL) a proposé un système de notation comprenant des critères cliniques pour déterminer le stade précoce du MF¹⁰. Le diagnostic repose principalement sur la perte des antigènes

associés aux lymphocytes T (CD2, CD3, CD5 et CD7) (Tableau 2, Tableau 3) et la présence de lymphocytes T clonaux dans les couches du derme et de l'épiderme. Un score minimal de 4 points sur un maximum de 8 points est nécessaire pour poser le diagnostic d'un MF précoce. Si ces critères ne sont pas atteints (score < 4 points), le diagnostic du MF ne peut pas être confirmé (Tableau 3).

Critères	Système de points
Clinique	
Critère de base : Macule/papule persistante ou progressive. Critères additionnels : 1) Localisation non exposée au soleil. 2) Variation de la taille/forme. 3) Poikilodermie*	2 points pour le critère de base et deux critères additionnels. ET 1 point pour le critère de base et un critère additionnel.
Histologie	
Critère de base : Infiltrat superficiel lymphoïde. Critères additionnels : 1) Épidermotropisme sans œdème épidermique intercellulaire (spongiose). 2) Lymphocytes cérébriformes	2 points pour le critère de base et deux critères additionnels. ET 1 point pour le critère de base et un critère additionnel.
Biologie moléculaire	
Réarrangement clonal du gène <i>TCR</i>	1 point pour la clonalité.
Phénotype aberrant	
1) < 50% lymphocytes T CD2 ⁺ et CD3 ⁺ et/ou CD5 ⁺ 2) < 10% Lymphocytes T CD7 ⁺ 3) Différence d'expression de CD2, CD3, CD5 ou CD7 entre le derme et l'épiderme**	1 point si un critère ou plus.

Tableau 3 : Critères pour le diagnostic précoce du MF (adapté de Pimpinelli *et al.*, JAAD 2005)¹⁰. Un score total de 4 points (sur 8) est nécessaire pour établir un diagnostic d'un MF, en tenant compte de divers critères cliniques, histologiques, moléculaires et d'immunophénotypage. TCR pour le récepteur des lymphocytes T. * La poikilodermie est relativement spécifique au MF précoce et correspond cliniquement à une pigmentation ou à une dépigmentation tachetée accompagnée d'une légère infiltration. ** Perte d'antigènes associés aux lymphocytes T confinée dans l'épiderme.

1.2.2.4 Classification TNMB

La classification TNM (*Tumor, lymph Node, Metatasis*) a été établie en 1979²⁴ et modifiée en 2007²⁵ avec l'ajout de la classification de l'envahissement sanguin parmi les critères (TNMB, B pour *Blood*). Plus récemment, en 2022, l'ISCL, l'EORTC et l'*United States Cutaneous Lymphoma Consortium* (USCLC) ont proposé une mise à jour qui prend en compte la clonalité des sites cutanés et extracutanés, ainsi que l'implication de la moelle osseuse²⁶. Les clones de lymphocytes T peuvent être précisés par réaction en chaîne par polymérase (PCR pour *Polymerase Chain Reaction*) ou par *Southern blot* du récepteur spécifique des lymphocytes T (TCR pour *T-Cell Receptor*) qui leur permet de reconnaître les antigènes. Cette classification s'applique à la fois au MF et au SS et offre la possibilité de différencier les stades cliniques avec des pronostics divergents (Tableau 4).

Stades cliniques		T	N	M	B	OS (en année, médiane)
Précoces	IA	1	0	0	0-1	35,5
		Macules/Papules < 10% SCT	Ø envahissement ganglionnaire	Ø envahissement viscéral	< 1000 cellules atypiques	
	IB	2	0	0	0-1	21,5
		Macules/Papules ≥ 10% SCT	Ø envahissement ganglionnaire	Ø envahissement viscéral	< 1000 cellules atypiques	
	IIA	1-2	1-2	0	0-1	15,8
		Macules/Papules	Infiltrat de cellules inflammatoires	Ø envahissement viscéral	< 1000 cellules atypiques	
Avancés	IIIB	3	0-2	0	0-1	4,7
		Lésion tumorale	Ø envahissement ou infiltrat de cellules inflammatoires	Ø envahissement viscéral	< 1000 cellules atypiques	
	IIIA	4	0-2	0	0	4,7
		Érythrodermie	Ø envahissement ou infiltrat de cellules inflammatoires	Ø envahissement viscéral	< 250 cellules atypiques	
	IIIB	4	0-2	0	1	3,4
		Érythrodermie	Ø envahissement ou infiltrat de cellules inflammatoires	Ø envahissement viscéral	250 à 1000 cellules atypiques	
	IVA1	1-4	0-2	0	2	3,8
		Toute atteinte cutanée	Ø envahissement ou infiltrat de cellules inflammatoires	Ø envahissement viscéral	> 1000 cellules atypiques avec présence d'une clonalité	
	IVA2	1-4	3	0	0-2	2,1
		Toute atteinte cutanée	Envahissement ganglionnaire avec destruction partielle ou complète de l'architecture	Ø envahissement viscéral	Tout envahissement circulant	
	IVB	1-4	0-3	1	0-2	1,4
		Toute atteinte cutanée	Toute atteinte ganglionnaire	Envahissement viscéral	Tout envahissement circulant	

Tableau 4 : Classification TMNB des différents stades du MF selon l'ISCL, l'USCLC et l'EORTC. La ligne rouge délimite les stades précoces et avancés. OS : survie globale, SCT : Surface Corporelle Totale, TMB : charge tumorale mutationnelle. Les médianes pour la survie globale (mOS) proviennent de la publication d'Agar *et al.*, JCO 2010²⁷).

1.2.2.5 Évolution de la Pathologie

Le MF évolue de manière indolente sur plusieurs années, traversant différents stades, et les patients présentent un large éventail de symptômes cliniques et d'évolution de la maladie¹. Aux premiers stades, le MF se manifeste généralement par des macules et des papules érythémateuses sur la peau, ce qui rend le diagnostic souvent difficile en raison du manque de spécificité des lésions. Éventuellement, les symptômes évoluent de ces stades cliniques à faible risque (LR pour *Low-Risk*), associés à une survie favorable (IA à IIA dans la classification TNMB) aux stades à haut risque (HR pour *High-Risk*) très agressifs et résistants aux traitements (IIB à IVB), caractérisés par la présence de tumeurs, d'ulcères ou d'une érythrodermie généralisée^{1,25}.

Au niveau cellulaire, environ 25% des patients qui risquent de développer une histologie avec des cellules MF transformées de grande taille²⁸ qui sont associées à un pronostic très défavorable²⁹⁻³¹. Comme la détection du MF n'est pas toujours évidente, la vérification histologique du diagnostic est nécessaire avant de commencer le traitement par des médicaments immunomodulateurs (méthotrexate, interféron, *etc.*). Cependant, certains cas de MF précoces peuvent encore passer inaperçus, même après une biopsie.

1.2.3 Formes Variantes du Mycosis Fongoïde

En plus de sa forme classique, le MF possède plusieurs variantes histologiques reconnues par l'OMS en raison de leurs caractéristiques cliniques et/ou pronostiques. Parmi elles, trois sont bien établies : le MF folliculotrope (fMF pour *folliculotropic MF*), la réticulose pagétoïde et la chalazodermie granulomateuse. Le fMF est la forme variante la plus courante, représentant jusqu'à 10% de tous les cas de MF, tandis que

la réticulose pagétoïde et la chalazodermie granulomateuse sont extrêmement rares^{1,3,27}.

1.2.3.1 Mycosis Fongoïde Folliculotrope

Le fMF survient généralement entre 45 et 60 ans, avec une prédominance masculine³². Il a été catégorisé comme une entité distincte du MF classique en raison de ses particularités cliniques, histologiques et de sa résistance aux traitements. Du point de vue histologique, l'infiltrat est composé principalement de lymphocytes atypiques CD4⁺, avec une infiltration plus prononcée des follicules (folliculotropisme) par rapport à l'épiderme³³.

1.2.3.2 Réticulose Pagétoïde

La réticulose pagétoïde localisée ou maladie de Woringer-Kolopp affecte des patients de tout âge. L'immunophénotype classique est CD3⁺, CD4⁺, CD8⁺³⁴. Sur le plan clinique, les lésions se présentent sous forme de plaques psoriasiformes bien délimitées avec un éclaircissement central et une bordure surélevée. La pathologie présente un épidermotropisme intense de lymphocytes cérébriformes, bien supérieur à celui du MF classique (30% contre 10%). Cependant, la réticulose pagétoïde progresse lentement et a un bon pronostic³⁵.

1.2.3.3 Chalazodermie Granulomateuse

La chalazodermie granulomateuse est une variante extrêmement rare du MF qui survient généralement chez les sujets masculins entre 30 et 50 ans. Cette maladie est caractérisée cliniquement par des papules flasques érythémateuses qui conduisent à la perte de l'élasticité cutanée, principalement au niveau des aisselles et de l'aîne. Sur le plan histologique, il y a présence d'un infiltrat dermique de lymphocytes

tumoraux clonaux CD4⁺, de cellules géantes multinucléées et un grand nombre de cellules macrophages, responsables de la destruction du tissu élastique. La progression de la chalazodermie granulomateuse est lente et l'extension extra-cutanée reste rare³⁶.

1.2.4 Mycosis Fongoïde Transformé et Syndrome de Sézary

1.2.4.1 Cellule de Mycosis Fongoïde Transformée

À différents stades du MF, les patients peuvent développer des lésions où les lymphocytes T augmentent en nombre avec une perte d'épidermotropisme (Figure 4A), mais aussi en taille (environ 4 fois la taille d'un lymphocyte normal)^{37,38}. Ces grandes cellules atypiques sont dites transformées (tMF pour *transformed Mycosis Fungoides*) (Figure 4B)³⁹. Cette transformation est associée à une agressivité de la maladie élevée, accompagnée d'une diminution de l'OS des patients à généralement quelques mois^{40,41}.

La transformation histologique est définie par la présence de cellules tMF représentant plus de 25% de l'infiltrat de la lésion du derme ou par la formation de nodules constitués de ces grands lymphocytes^{28,29,42}. Bien que la transformation puisse survenir à un stade précoce, elle est généralement associée aux stades avancés, en particulier le stade tumoral^{31,43–45}. La prévalence du tMF varie grandement dans la littérature, avec des rapports entre 3 % et 34 % des cas de MF entre 1987 et 2022, dans des études portant sur des cohortes CTCLs d'au moins 50 patients^{27–29,38,41,43–55}.

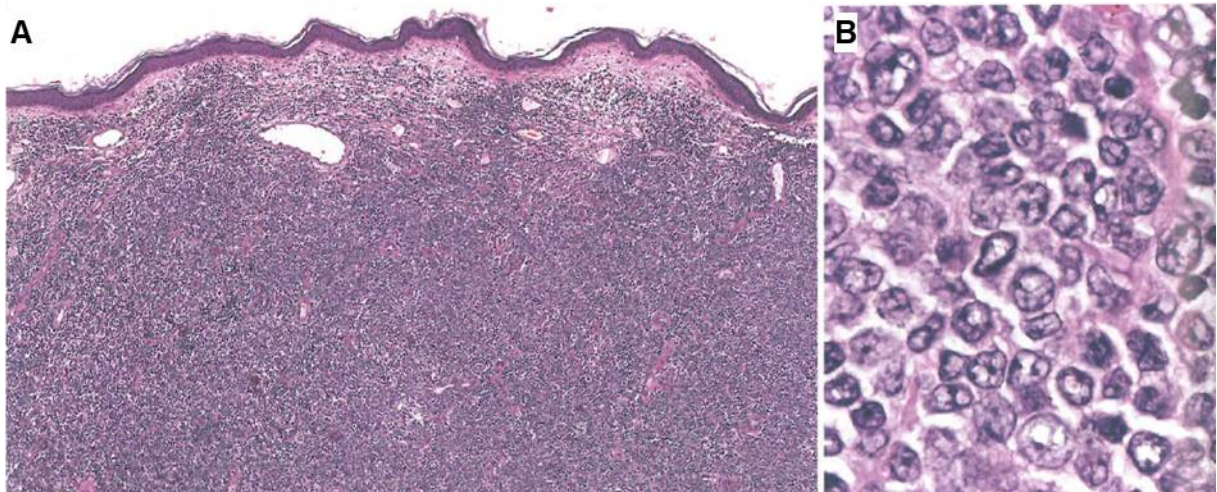


Figure 4 : Histologies de MF avancés montrant plus de 25% de grandes cellules. A

Lésion tumorale avec une accumulation de lymphocytes dans le derme sans épidermotropisme. **B.** Détail de grands tMF.

Généralement, le tMF présente davantage de grandes cellules CD30⁺ ou CD30⁻ (récepteur du facteur de nécrose tumorale 8, TNFRSF8) par rapport aux cellules MF non transformées. De plus, les cellules tMF expriment significativement plus les marqueurs immunohistochimiques (IHC) Ki-67 et p53¹³, deux protéines clés impliquées dans la prolifération, le cycle cellulaire et l'apoptose (Tableau 2). Ces marqueurs sont proposés comme indicateurs diagnostiques de la transformation du MF¹³. Parmi les autres marqueurs diagnostiques identifiables par IHC, figurent CD25 (chaîne *alpha* du récepteur interleukine 2)⁴⁵, un récepteur souvent utilisé pour identifier divers lymphomes T, ainsi que KIR3DL2 (CD158k), un récepteur *killer immunoglobulin-like* présent dans un sous-groupe de cellules NK^{56,57}. L'expression de ces deux marqueurs est significativement plus élevée chez les patients présentant des cellules tMF par rapport à ceux sans transformation^{45,57}.

1.2.4.2 Infiltrats de Lymphocytes B

Il existe des cas avec des infiltrats de lymphocytes B exprimant les marqueurs PAX5⁺ et/ou CD20⁺ dans les stades avancés¹² (Tableau 2). Le nombre de

lymphocytes B peut être proéminent, cependant ces cas ne doivent pas être interprétés à tort comme des CBCLs. D'autre part, il a déjà été recensé des observations de « lymphomes composites » où tMF et CBCLs cohabitent dans la même lésion⁵⁸⁻⁶¹.

1.2.4.3 Cellule de Sézary

Lorsque la circulation sanguine est atteinte, il y a fréquemment des lymphocytes T malins appelés « cellules de Sézary » dans la peau, le sang et les ganglions. Ces cellules de Sézary possèdent un noyau cérébriforme qui occupe environ 80% du volume cellulaire avec une hétérochromatine condensée à la périphérie et est entouré d'un mince anneau de cytoplasme (Figure 5A). Cet aspect nucléaire plissé, de forme serpentine (Figure 5B et C), n'est pas spécifique et peut également être observé dans de petits lymphocytes normaux, notamment dans certaines maladies inflammatoires, ce qui contribue à la complexité du diagnostic des lymphomes.

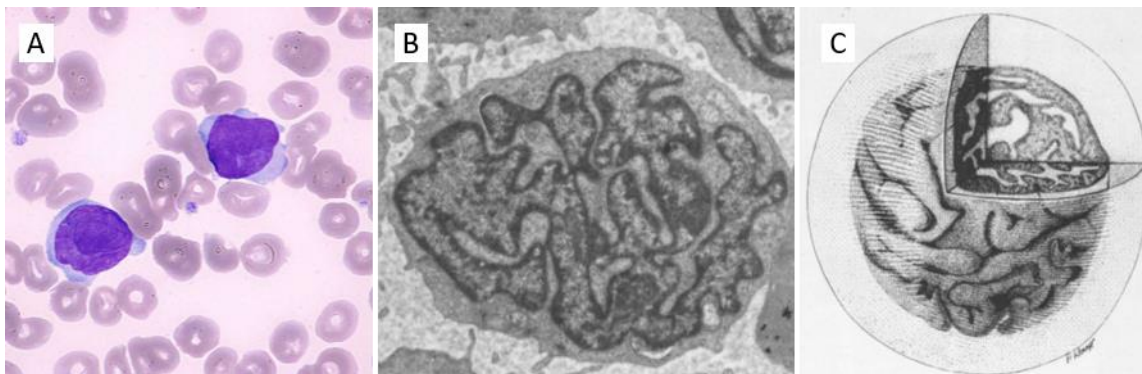


Figure 5 : Cellules de Sézary de petite à moyenne taille avec un noyau cérébriforme. A. Frottis sanguin présentant deux cellules anormales dans le sang périphérique d'un patient avec un SS. **B.** Observation par microscopie électronique d'une cellule de Sézary (Grossissement x3000). Le noyau est très imposant et présente un aspect serpentin. **C.** Diagramme d'une cellule de Sézary avec un noyau cérébriforme, mais avec une apparence serpentine dans la section coupée. La photographie du frottis sanguin provient de l'*Atlas Hematological Cytology*⁶². La photographie de microscopie électronique et le diagramme sont issus de Lutzner *et al.* Ann Intern Med 1975¹¹.

Sur le plan phénotypique, la cellule de Sézary est classiquement un profil de lymphocyte T (CD2⁺, CD3⁺, CD5⁺, CD28⁺) auxiliaire (CD4⁺, CD8⁻) mémoire centrale (CD27⁺, CD45RO⁺)⁶³, avec l'expression des récepteurs à chimiokines nodaux CCR7⁺ (CD197) et présentant un homing particulier pour la peau (CCR4⁺ et CCR10⁺)⁶⁴. Il existe des pertes variables des antigènes CD2, CD7 et CD26^{41,63,65}.

1.2.4.4 Distinctions Mycosis Fongoïde et Syndrome de Sézary

Le Syndrome de Sézary (SS), un CTCL érythrodermique très agressif et rare, a longtemps été considéré comme la forme leucémique du MF en raison de leurs similitudes dermatologiques et histologiques¹. La difficulté à distinguer les symptômes entre le MF et le SS suggère qu'il existe un continuum de manifestations qui englobe les deux entités. Cependant, de nombreuses différences sur les plans immunophénotypiques, cytologiques, pronostiques et thérapeutiques ont rendu nécessaire de distinguer les deux maladies.

Depuis 2002, l'ISCL reconnaît le SS comme un CTCL avec une évolution clinique agressive, caractérisé par une érythrodermie, une atteinte des ganglions lymphatiques et la présence de cellules de Sézary dans le sang périphérique qui le différencie du MF^{27,66}. Cette distinction a été intégrée dans la classification de 2005, où l'EORTC et l'OMS ont également différencié le SS des cas de MF qui développent ultérieurement des caractéristiques du SS. Dans ces derniers cas, la maladie est désignée comme un « MF leucémique », ou un « SS précédé de MF » ou encore un « SS secondaire »^{1,67}.

En 2010, l'équipe de Campbell a montré que le MF provient de lymphocytes T mémoires effecteurs qui expriment fortement les marqueurs cutanés CCR4 et CLA, qui orientent les lymphocytes vers la peau. Ces lymphocytes T sont appelés les

lymphocytes mémoires résidents. En revanche, le SS possède des lymphocytes qui expriment les marqueurs ganglionnaires CCR7 et L-Sélectine, ainsi que CD27, un marqueur des lymphocytes T mémoires centrales. Le SS exprime également d'autres marqueurs de homing cutanés, tels que CCR4, CCR6, CCR10 et CLA. Ces résultats suggèrent que le MF et le SS sont des maladies distinctes, ayant pour origine des sous-types différents de lymphocytes T⁶³ (Figure 6).

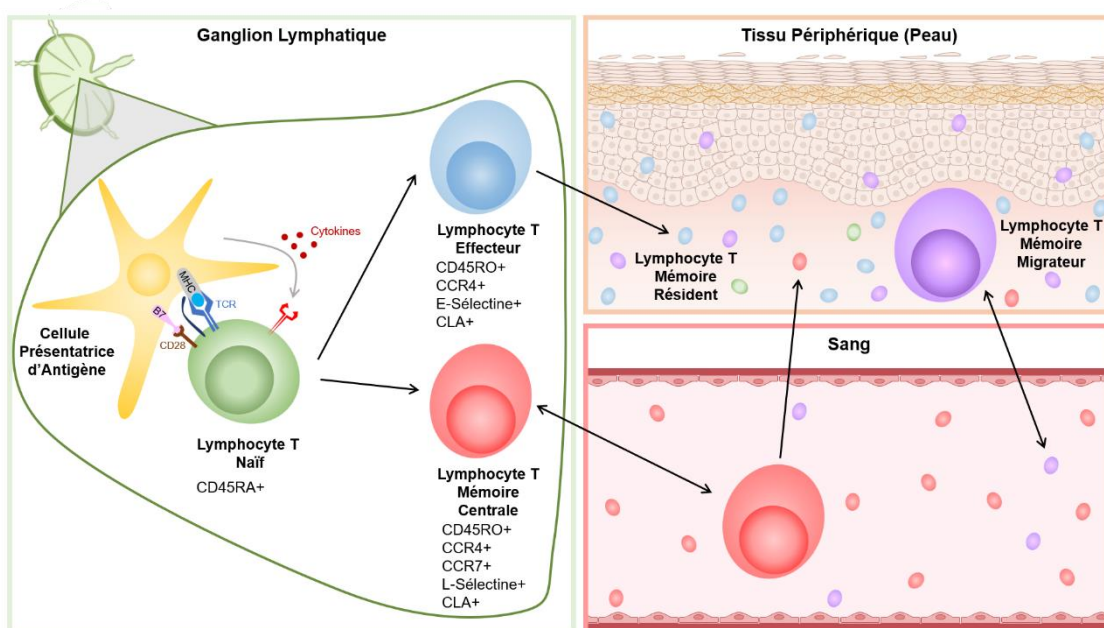


Figure 6 : Sous-types de lymphocytes T circulant dans la peau présentant l'origine différente de la cellule de MF et celle du SS. Les Lymphocytes T Naïfs se différencient en Lymphocytes T Mémoires Effecteurs et en Lymphocytes T Mémoires Centrales après s'être liés à leur antigène sur les Cellules Présentatrices d'Antigènes à l'intérieur des ganglions lymphatiques. L'expression du ligand de surface CCR4 détermine la capacité des lymphocytes à se diriger vers la peau. Le Lymphocyte T Mémoire Résident dans la peau représente la cellule d'origine du MF. L'expression de CCR7/L-sélectine détermine la capacité des lymphocytes à circuler entre le sang et les ganglions lymphatiques. Le Lymphocyte T Mémoire Centrale représente la cellule d'origine du SS. Figure d'après Larocca et Kupper, Hematol Oncol Clin North Am, 2019⁶⁸.

Une autre différence importante reste la localisation des cellules concernées. L'immunomarquage sélectif de la chaîne β du TCR sur les lymphocytes T tumoraux dans les échantillons de peau et de sang des patients atteints de MF révèle que ces

cellules sont localisées dans la jonction dermo-épidermique. En revanche, chez les patients atteints de CTCLs leucémiques, dont le SS, les lymphocytes T tumoraux se trouvent principalement dans le derme⁶⁹.

Au niveau génomique, deux études de 2009 et 2010 renforcent la distinction entre le MF/tMF et le SS en utilisant la technique d'hybridation génomique comparative sur puce à ADN (array-CGH) pour en identifiant des altérations chromosomiques spécifiques à chaque sous-types de CTCLs^{70,71}. Les gènes présents dans les régions concernées ont été relevés^{70,71} et l'impact des altérations sur l'expression de certains de ces gènes a été validé par les techniques de microarrays, de qPCR et d'IHC⁷⁰. De même, l'identification de l'expression des gènes *KIR3DL2*, *NKp46* (CD335), *PLS3* et *Twist* dans le sang à l'aide de la PCR par transcription inverse (RT-PCR) peut servir dans le diagnostic du SS, en raison de leur spécificité et leur expression préférentielle dans les cellules tumorales CD4⁺ de SS^{56,72}.

1.2.5 Pronostic

1.2.5.1 Mycosis Fongoïde Classique

Le pronostic du patient atteint de MF va grandement varier en fonction du stade, du type et de l'étendue des lésions cutanées, ainsi que de la présence d'une atteinte extra-cutanée de la maladie¹. Les patients possédant un MF précoce, ont généralement un pronostic favorable similaire à la population générale, avec une médiane de la survie globale mOS (*median of Overall Survival*) à 10 ans entre 97% et 98%¹. Au contraire, les patients qui possèdent un MF à un stade avancé avec des symptômes comprenant une augmentation anormale du volume des ganglions lymphatiques (adénopathie) ou une atteinte viscérale, ont un pronostic plus

défavorable avec une mOS à 5 ans comprise entre 18% et 67%^{27,43}, voire même 10,6% pour le stade IV⁴⁸.

Il est à noter que dans la plupart des classifications cliniques actuelles, la présence de macules et celle de papules sont généralement groupées ensemble pour évaluer la surface corporelle totale impactée par les lésions. Cependant, des études ont montré que les papules ont un pronostic clairement plus défavorable comparé à la présence de macules uniquement^{27,73}.

1.2.5.2 Mycosis Fongoïde Folliculotrope

En raison de son infiltration préférentiel des follicules pileux que de l'épiderme, le fMF est moins accessible aux traitements cutanés. En absence d'épidermotropisme significatif, les cellules tumorales restent majoritairement dans le derme, ce qui diminue leur exposition aux traitement topiques. De plus, des phénomènes de fibrose périfolliculaire (accumulation excessive de tissu conjonctif autour des follicules pileux) et de mucinose folliculaire (accumulation anormale de mucine) sont souvent observés lors de fMF³³. Ces phénomènes peuvent jouer le rôle de barrière physique protectrice empêchant les traitements d'atteindre les cellules tumorales. Il y a également un taux de progression plus élevé vers une atteinte extracutanée et un pronostic moins favorable dans le fMF comparé à la forme classique³³. La survie globale (OS) est faible, avec un taux de survie de 26% sur 10 ans, même dans les stades précoces. Quel que soit l'aspect clinique de la lésion au moment du diagnostic, le fMF implique un diagnostic de stade III^{25,33,35}.

1.2.5.3 Mycosis Fongoïde Transformé

La présence de cellules tMF est liée à un pronostic très défavorable pour les patients⁴⁸. La mOS est généralement comprise entre 2 et 8,3 années^{27,28,44}. Dans le

contexte d'un tMF, les facteurs ayant un impact négatif significatif sur l'OS comprennent le stade avancé lors de la transformation (stades IIB-IVB)^{28,29,31,38,44,47}, la transformation extra-cutanée (stade IV)^{28,31,47}, l'extension des lésions cutanées et la présence d'un fMF^{29,31,44,48,74}. Parmi les autres facteurs dont la significativité varie dans la littérature, se trouvent l'âge avancé et un court délai entre le diagnostic et la transformation^{38,44,48}. Curieusement, bien que l'expression de CD30 soit plus commune dans les stades avancés et/ou le tMF, ce facteur a été reporté comme ayant un pronostic favorable^{45,48}.

1.2.6 Traitements

Le traitement du MF et du SS repose sur les recommandations thérapeutiques européennes issues du consensus de l'EORTC, actualisé en 2023, et dépend du stade diagnostiqué de la maladie⁷⁵. L'approche thérapeutique est principalement « palliative », c'est-à-dire qu'elle vise à atténuer les symptômes visibles de la maladie afin d'améliorer la qualité de vie du patient dans sa dimension psychologique et sociale. À ce jour, le seul traitement curatif est l'allogreffe de cellules souches hématopoïétiques (CSH). La surveillance des symptômes seule est recommandée pour les patients au stade T1, leur espérance de vie étant similaire à celle de la population générale.

Le traitement initial (stratégie de première ligne) consiste en des thérapies locales dirigées vers la peau. Une liste non exhaustive de ces traitements comprend les solutions et gels appliqués directement sur la peau (corticostéroïdes, chimiothérapie topique), la photothérapie aux ultraviolets (UVs) accompagnée d'un médicament photosensibilisateur (PUVA, UVB) et la radiothérapie aux rayons X⁷⁵.

Cependant, si la maladie progresse vers un MF avancé ou un SS, des traitements systémiques (stratégie de deuxième ligne) sont envisagés et sont administrés par voie orale et/ou intraveineuse afin de contrôler les symptômes et la charge tumorale dans tout l'organisme^{26,75}. Cette stratégie peut également être utilisée si le patient avec un MF précoce est réfractaire aux thérapies topiques ou présente des contre-indications. Les traitements systémiques comprennent la photophérèse extracorporelle (photothérapie du sang qui est réinjecté dans le corps), l'utilisation de composés ayant des propriétés antitumorales (interféron peg-IFN- α 2a, rétinoïdes, romidepsine), les chimiothérapies (Doxorubicine liposomale pégylée, Gemcitabine), les thérapies ciblées utilisant des anticorps monoclonaux et la greffe allogénique de cellules souches^{24,26,76–83}.

1.3 Physiopathologie

1.3.1 Mutations Somatiques

Les mutations somatiques, également appelées « acquises » par opposition aux mutations héréditaires, sont des anomalies génétiques qui surviennent durant le développement embryonnaire et/ou au cours de la vie, à mesure que les cellules normales se divisent. Ces mutations apparaissent fréquemment en réponse à des facteurs externes, tels que des expositions environnementales (rayons UVs solaires, particules fines, certaines infections, *etc.*), le mode de vie (tabagisme, consommation d'alcool, comportements alimentaires, *etc.*) ou encore des traitements cytotoxiques ou impliquant des UVs (chimiothérapie, photothérapie, *etc.*). Bien que le nombre de mutations somatiques soit habituellement diminué par des mécanismes de réparation de l'ADN, un dysfonctionnement de ces systèmes peut entraîner une augmentation significative du taux de mutations⁸⁴.

1.3.2 Mutations *Drivers*, *Passengers* et Biomarqueurs

Dans les études génomiques des cancers, le nombre de mutations détectées peut varier d'une dizaine à plusieurs centaines de milliers. La grande majorité d'entre elles ne confère pas d'avantages sélectifs dans le développement du cancer, et ces mutations, dites « *passengers* » (passagères ou neutres), n'ont que peu d'influences sur l'évolution de la tumeur⁸⁴. Un nombre plus restreint de mutations, dites « *drivers* » (conductrices), vont impacter des gènes clés impliqués dans l'oncogenèse ou la progression du cancer⁸⁵.

En clinique, ces mutations *drivers* peuvent être utilisées comme biomarqueurs pour faciliter le diagnostic ou servir d'indicateur du pronostic de la maladie. Cependant, ces biomarqueurs doivent être validés scientifiquement avant d'être incorporés dans la pratique clinique. Idéalement, ces biomarqueurs pourraient également permettre de mieux comprendre le mécanisme de développement du MF⁸⁶.

1.3.3 Variant Nucléotidique Simple

Au niveau moléculaire, le MF est une maladie génétiquement hétérogène, caractérisée en partie par la présence de multiples variants nucléotidiques simples (SNV pour *Single-Nucleotide Variant*). Ces mutations comprennent les substitutions et les Insertions-Délétions (InDels) (Figure 7).

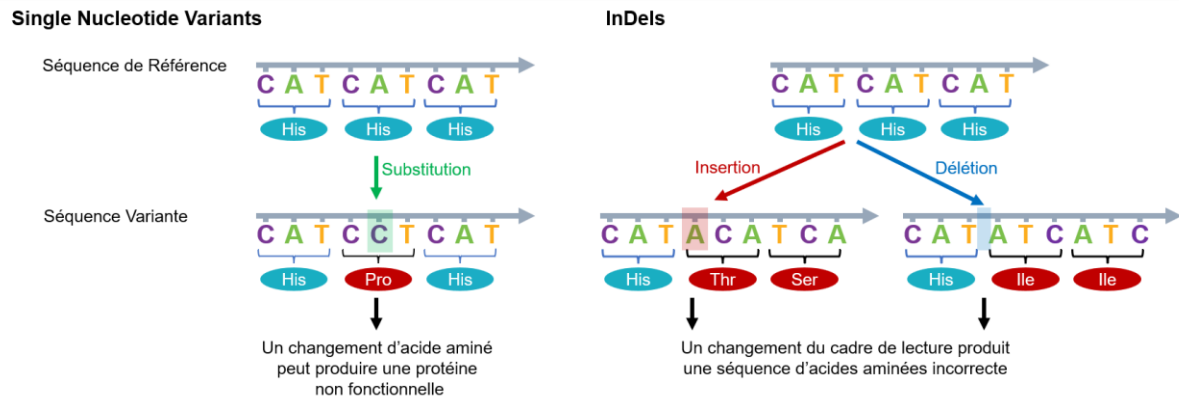


Figure 7 : Exemples de SNVs et InDels affectant la séquence nucléotidique et leurs conséquences sur la séquence protéique.

Bien que plusieurs études génomiques, utilisant des techniques, telles que l'hybridation génomique comparative, la PCR ou le séquençage à haut débit (NGS pour *Next Generation Sequencing*), ont recherché des SNVs dans le MF entre 2002 et 2022^{17,63,87-100}, les connaissances sur la pathogenèse de la maladie restent limitées. Toutefois, ces études sont souvent restreintes par plusieurs facteurs :

- Une taille réduite de leur cohorte de patients (souvent inférieure à 12)^{63,90,92}.
- Une focalisation sur les stades HR du MF^{63,88-90,95,96}
- Un manque de distinction claire entre le terme « CTCL », le MF, le SS et d'autres LCPs, ainsi que leur stade clinique associé^{87-89,91,100}.

En conséquence, les analyses génétiques du MF, en particulier au stade LR, restent insuffisamment explorées dans la littérature sur les CTCLs^{17,92,98,99,101}.

Les études sur les CTCLs, et plus spécifiquement sur le MF, ont identifié des SNVs présents sur plusieurs gènes clés tumoraux (Tableau 5). Parmi eux, *TP53*^{86,89,90,94,96,97} qui est fréquemment plus muté aux stades HR^{92,98}. D'autres gènes récurrents incluent *ATM*^{89,92,98} et *CCR4/CCR7* qui codent pour des récepteurs cutanés aux chémokines^{63,90,102}, *CD28* codant pour un récepteur des lymphocytes T^{89,94,102},

FAS^{87,89,98}, *PDCD1*^{90,97,98}, ainsi que *TNFRSF1B* qui code pour le récepteur du *tumor necrosis factor* *TNFR2*^{94,97} (voir Figure 9A).

Publication	Gènes impactés par des mutations SNVs
Vaqué <i>et al.</i> Blood 2014	<i>PLCG1 TP53</i>
Ungewickell <i>et al.</i> Nat Genet 2015	<i>CARD11, CD28, KMT2C, KRAS, PLCG1, TNFRSF1B, TP53</i>
Choi <i>et al.</i> Nat Genet 2015	<i>ATM, BRAF, CD28, CTCF, DNMT3A, FAS, IRF4, PLCG1, RHOA, STAT5B, TP53, ZEB1</i>
McGirt <i>et al.</i> Blood 2015	<i>ATM, CDKN2A/B, JAK3, KMT2D, NCOR1, TP53, ZEB1</i>
Argyropoulos <i>et al.</i> Blood Cancer J 2020	<i>ATM, CDKN2A/B, DNMT3A, FAT1, FAT3, JAK3, KMT2C, NOTCH3, TNFAIP3, TP53</i>

Tableau 5 : Publications rapportant des mutations SNVs dans les CTCLs, en particulier le MF, et les gènes impactés.

De manière générale, le nombre de gènes candidats impactés par des SNVs, susceptibles d'être des *drivers*, est relativement élevé dans les études consacrées au MF. Une analyse de Park *et al.* en 2021, portant sur le séquençage génomique (18 échantillons de MF et 7 échantillons de MF leucémiques), exomique (19 échantillons de MF) et transcriptomique (22 échantillons de MF) sur des données MF/SS a révélé 86 gènes *drivers*. Parmi eux, les gènes *NFKB1* et *KLF2*, deux facteurs de transcription (TFs) impliqués en aval de la voie de signalisation du TCR. De plus, certains gènes, bien que déjà identifiés dans d'autres cancers, apparaissent pour la première fois dans le MF, comme *JUNB* (p.A282V) (Figure 9A) et *TBL1XR1*. Auparavant, *JUNB* avait été identifié seulement dans des études transcriptomiques centrées sur le SS, avec une augmentation d'expression observée à partir de puces à ADN^{86,103,104}. En outre, des SNVs inédits ont été découverts dans les gènes suppresseurs de tumeurs, *FUBP1* et *ANO6*¹⁷.

1.3.4 Variation du Nombre de Copies Somatiques

Outre les SNVs, plusieurs variations du nombre de copies somatiques (SCNV pour *Somatic Copy-Number Variation*) (Figure 8) ont été identifiées dans les CTCLs, y compris le MF (Figure 9A)^{70,86,89,91,93–96,98,102,105–111}.

Somatique Copy Number Variants

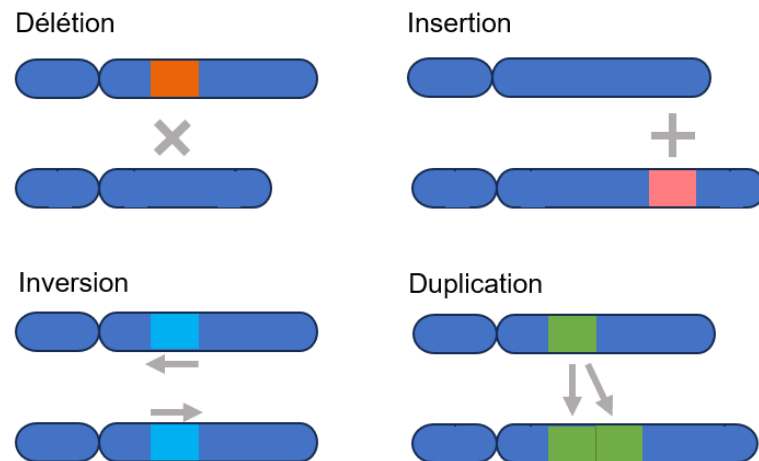


Figure 8 : Différents types de SCNVs sur un bras chromosomique.

Les premières études génomiques sur les CTCLs ont révélé de nombreuses SCNVs, très répandues dans les cohortes de patients et associées à une réduction de l'OS, qui affectent des gènes suppresseurs de tumeurs : le gain8q (dont *MYC*), les délétions del6q, del10q23.31 (dont *PTEN* et *FAS*), del13q14.2 (dont *RB1*), del17p13.1 (dont *TP53*), del9p21.2 (dont *CDKN2A/CDKN2B*) et del12p13.1 (*CDKN1B*)^{90,91,106,107,109,110,112}. De plus, la del9p21.3 est particulièrement plus fréquente chez les patients atteints de tMF^{95,105,113,114}. Certaines de ces SCNVs sont communes au MF et au SS^{86,110}, tandis que d'autres sont plus spécifiques au MF, tels que le gain7 et la del9p21^{86,97,106}. En revanche, certaines SCNVs sont propres au SS le gain8q, le gain17q, la del10q et la del17p^{91,107,109,112,112,115}. De plus, une augmentation du nombre de copies (CN pour *Copy Number*) du gène *JUNB* a été observé dans le SS par la technique FISH^{86,103,104,116}.

1.3.5 Voies de Signalisation

Parmi les nombreux SNVs et CNVs relevés dans les études sur les CTCLs, plusieurs gènes clés sont impliqués dans l'activité des voies de signalisation. Le ciblage de ces voies contribue à la progression tumorale en favorisant la prolifération cellulaire, la résistance à l'apoptose et la différenciation cellulaire (Figure 9A).

Les mutations dans les gènes *BRAF*, *KRAS*, *NRAS*, *MAP2K1* et *MAPK1* retrouvées dans les CTCLs^{17,89,91,97,98,111}, impactent la voie MAPK dont la perturbation favorise la prolifération incontrôlée des cellules tumorales. Les gènes *BRAF*, *KRAS* et *NRAS* sont notamment plus mutés dans les cas de MF avancés^{94,100}.

Les gènes mutés de la voie JAK/STAT : *JAK1*, *JAK2*, *JAK3*, *STAT3* et *STAT5B* influencent la survie et la prolifération des cellules tumorales. Une délétion des gènes régulateurs négatifs *SOCS1* et *HNRNPK*, entraîne une activation continue de la voie, favorisant ainsi la progression du MF⁹⁵.

La voie NF- κ B est un facteur crucial dans la prolifération cellulaire, l'inflammation et l'apoptose. Des mutations impactant les gènes *TNFRSF1B*, *NFKB1*, *NFKB2* et *TNFAIP3* perturbent le contrôle normal de cette voie et contribuent à rendre les cellules tumorales résistantes à l'apoptose induite par les cytokines^{94,95,117}. De plus, les mutations du gène *TNFRSF1B*, codant pour le récepteur TNFR2, sont fréquemment retrouvées dans le MF et le SS^{17,70,94,97,102}.

De très nombreux SNVs et SCNVs ont été relevés dans la voie de signalisation du TCR, essentielle dans la prolifération des lymphocytes T, dans les gènes *PLCG1*, *PRKCB*, *PRKCQ*, *CARD11*, *CD28* et *VAV1*^{88–90,94,96–98}. Les mutations qui affectent cette voie entraînent une activation accrue de la signalisation en aval, notamment via la voie NF- κ B, et participent à la résistance des cellules tumorales^{90,94}.

Des mutations oncogéniques récurrentes ont été identifiées dans les CTCLs, en particulier dans le gène *ZEB1*, un TF clé dans la différenciation des lymphocytes Th2 et un régulateur majeur dans la transition épithélio-mésenchymateuse (EMT), un processus souvent réactivé dans les cancers pour faciliter l'invasion des cellules tumorales^{102,118}. De plus, les gènes *CCR4/CCR7*, impliqués dans l'épidermotropisme des lymphocytes T, sont régulièrement impactés dans le cas de CTCLs^{102,118}. D'autres gènes sont affectés par les mutations, dont *IRF4*, un TF nécessaire à l'expansion des lymphocytes T activés par la signalisation du TCR^{89,119}, et *PRKG1*, codant pour une protéine kinase dépendante du GMP cyclique (cGMP), un régulateur de la voie cGMP/PKG. Cette voie joue un rôle essentiel dans le contrôle de la prolifération, l'apoptose et la différenciation cellulaire. La dérégulation de cette voie *via PRKG1* peut entraîner une prolifération cellulaire accrue et la survie des cellules tumorales⁹¹.

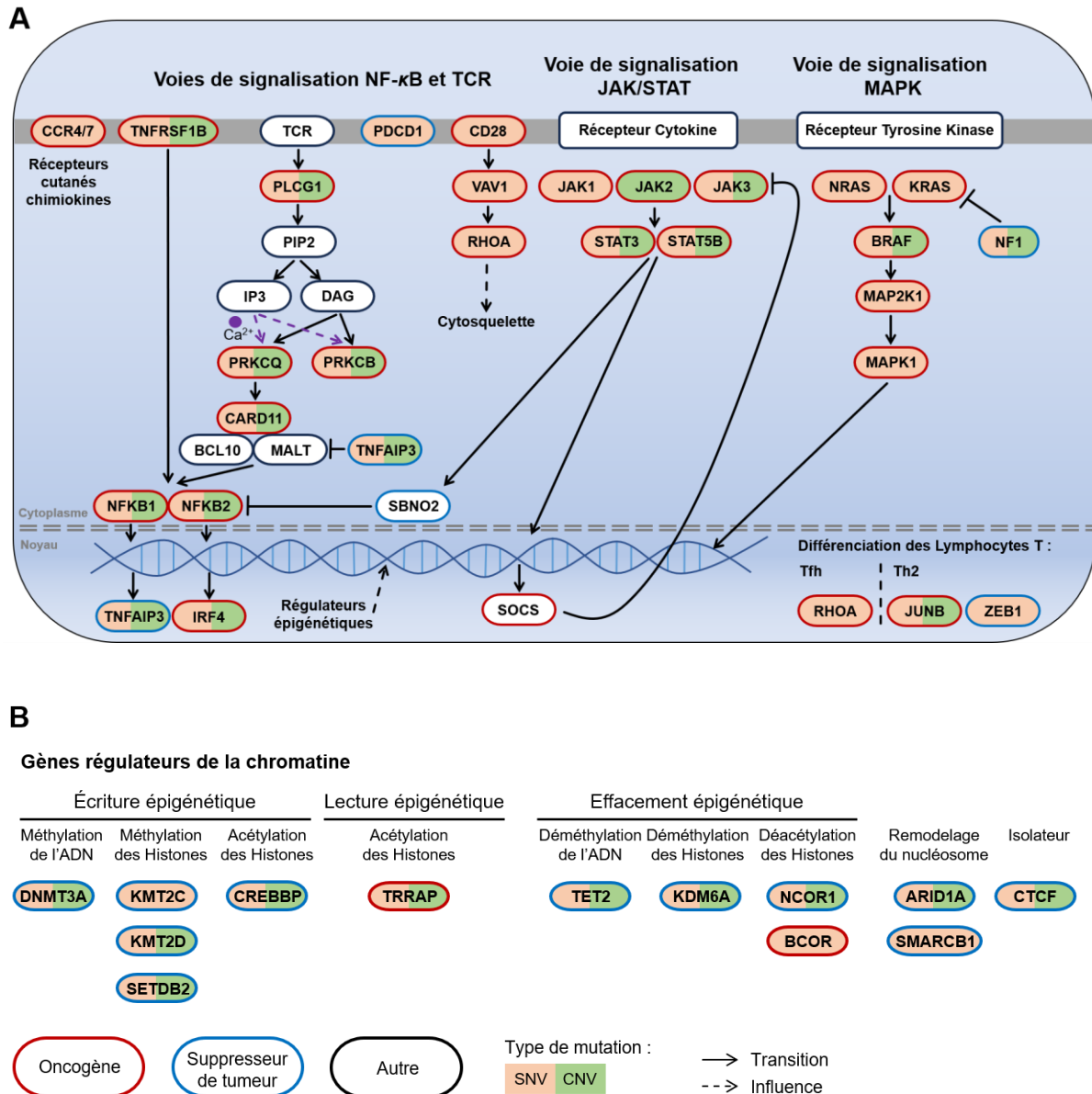


Figure 9 : Gènes *drivers* connus et voies de signalisations chez les CTCLs. A. Gènes *drivers* présents dans les voies de signalisation du NF-κB, TCR, JAK/STAT, MAPK et de différenciation des lymphocytes T. **B.** Gènes *drivers* connus impliqués dans les modifications épigénétiques et de la chromatine dans les CTCLs. Les onco-gènes putatifs et les supresseurs de tumeurs sont indiqués par des encadrés rouges et bleus respectivement. Figure d'après Ungewickell *et al.* Nat Genet 2015, Damsky *et al.* Curr Treat Options Oncol 2016, Park *et al.* Blood 2017 et 2021 et García-Díaz *et al.* Cancers 2021^{17,94,99,118,120}. Le type de mutation (SNV/CNV) détecté dans le gène proviennent de Park *et al.* Blood 2017⁹⁹.

1.3.6 Évolution Clonale

1.3.6.1 Architecture Clonale

Une cellule tumorale qui héberge une mutation somatique peut donner naissance à une nouvelle lignée de cellules identiques, appelée « clone », qui partage cette mutation. La sélection darwinienne peut maintenir ces clones et leur permettre d'acquérir des mutations *drivers* qui favorisent la prolifération tumorale, l'échappement à l'apoptose et au système immunitaire, ainsi que la résistance aux traitements⁸⁵.

L'évolution clonale dans le cancer repose sur une architecture complexe où les cellules tumorales se diversifient en formant une hiérarchie, souvent représentée sous la forme d'un arbre phylogénétique. À la base de cet arbre, un clone fondateur émerge d'une cellule initialement normale par l'acquisition de mutations précoces dites « clonales ». Ces mutations, présentes dès le début du cancer, sont partagées par toutes les cellules tumorales. Au fil du temps, certains groupes de cellules tumorales vont connaître des mutations additionnelles appelées « sous-clonales ». Ces mutations vont créer des sous-clones qui coexistent avec le clone fondateur et peuvent évoluer indépendamment^{121,122}.

L'évolution clonale des cancers n'est la plupart du temps pas « linéaire », c'est-à-dire qu'elle ne repose pas sur accumulation successive de mutations dans un clone unique (architecture monoclonale). Cependant, elle suit généralement un modèle « branché » où plusieurs sous-clones émergent d'un même clone fondateur et évoluent indépendamment en parallèle et pouvant entrer en compétition (architecture polyclonale). Ce phénomène conduit à une hétérogénéité clonale, où différentes populations de cellules tumorales coexistent avec des profils génétiques distincts⁸⁵.

1.3.6.2 Hétérogénéité Clonale

L'hétérogénéité clonale se traduit par la variabilité génétique et phénotypique des cellules tumorales dans la tumeur. Cette hétérogénéité fournit aux tumeurs une meilleure capacité d'adaptation face aux traitements, favorisant l'apparition de résistances thérapeutiques⁸⁵. Elle peut être influencée par des pressions sélectives (traitements, microenvironnement tumoral), l'évolution dynamique des sous-clones où certains deviennent dominant tandis que d'autres disparaissent et le taux de mutation qui diffère selon le cancer.

La charge tumorale mutationnelle (TMB pour *Tumor Mutational Burden*) est définie comme le nombre total de mutations somatiques présentes par mégabase (Mb) d'ADN tumoral. L'exome de cancers comme le mélanome malin, le carcinome épidermoïde du poumon ou le cancer de la vessie présente une TMB élevée (fréquence médiane des mutations non-synonymes d'environ 100 par Mb), souvent due à une exposition à des agents mutagènes comme les rayons UV solaires ou la fumée de tabac^{121,122}. En revanche, d'autres cancers comme la tumeur rhabdoïde, le sarcome d'Ewing et le cancer de la thyroïde présentent une TMB beaucoup plus faible (proche de 0,1 par Mb, soit environ une mutation sur l'ensemble de l'exome)^{122,123}. Cependant, ces fréquences peuvent varier énormément d'un patient à l'autre au sein d'un même cancer. Par exemple, la fréquence de clone par patient est comprise entre 0,1 à 100 Mb dans le mélanome et le cancer du poumon.

Cependant, bien qu'une TMB élevée puisse parfois générer une plus grande diversité clonale, elle ne reflète pas nécessairement l'hétérogénéité clonale de la tumeur. Par exemple, deux tumeurs peuvent avoir une TMB similaire, mais l'une peut

être dominée par un seul clone, tandis que l'autre présente plusieurs sous-clones génétiquement distincts.

1.3.6.3 Clonalité des Lymphocytes T dans le MF

Chez les CTCLs, la clonalité des lymphocytes T est un marqueur clé de la progression tumorale. Elle est le résultat de réarrangements somatiques au niveau des gènes du TCR, un processus unique à chaque lymphocyte T. Ces réarrangements clonaux spécifiques permettent de suivre l'évolution clonale tout au long de la progression de la maladie.

Actuellement, la clonalité des lymphocytes T dans le MF est évaluée *via* la PCR des chaînes *bêta* (TCR β) et *gamma* (TCR γ) du TCR, suivie d'une analyse par électrophorèse capillaire GeneScan¹²⁴. Cependant, cette méthode présente certaines limites : elle ne permet de détecter des clones monoclonaux que chez un sous-ensemble de patients atteints de CTCLs et présente un taux de faux positifs (FPs) significatif. Par conséquent, le diagnostic définitif du MF est souvent retardé, et il faut un temps médian de 4 années après l'apparition des premières lésions cutanées pour poser un diagnostic^{125,126}.

Une méthode plus fiable pour l'analyse de la clonalité des lymphocytes T est le séquençage NGS du domaine *Complementary Determining Region 3 (CDR3)* des chaînes TCR β et TCR γ . L'intérêt de cette technique est une meilleure sensibilité dans la détection des réarrangements clonaux, mais aussi des différents sous-clones présents dans la tumeur, ce qui permet d'évaluer l'hétérogénéité clonale. Il devient donc possible de poser un diagnostic plus précoce et plus précis du MF afin de le distinguer des autres dermatoses^{69,126}.

Le réarrangement monoclonal du TCR est communément rencontré dans les MF de type plaque ou tumoral. Par exemple, une étude de séquençage exomique complet (WES pour *Whole Exome Sequencing*) ayant analysé 33 échantillons issus de 27 patients MF atteints de plaques ou de tumeurs, a révélé la présence multiples clonotypes TCR *alpha* (TCR α) , TCR β et TCR γ . Cependant, chez 9 des patients, leur échantillon était composé d'un clonotype TCR γ unique, entre 2 et 7 clonotypes pour TCR β et de plusieurs clonotypes TCR α . Cette observation suggère que l'événement malin initiateur du MF ne se produit pas au niveau des lymphocytes matures résidents dans la peau, mais après le réarrangement du TCR γ , au cours du développement des lymphocytes. Il a aussi lieu avant le début des réarrangements du TCR α ou du TCR β ¹²⁶.

Une étude de 1995, ayant effectué une analyse moléculaire IHC et PCR du TCR β chez un patient atteint de CTCL, a montré que les cellules transformées et les petites cellules cérébriformes partageaient le même réarrangement du TCR⁴⁰. Cette découverte soutient l'hypothèse selon laquelle les cellules transformées proviennent d'un clone unique de lymphocytes T malins dans les CTCLs, et que la transformation n'est pas une entité clinicopathologique distincte, mais bien une progression évolutive du même clone tumoral^{40,41}.

1.3.7 Signatures Mutationnelles

Les signatures mutationnelles sont des empreintes caractéristiques laissées dans le génome des cellules tumorales et sont le résultat de processus mutationnels endogènes et exogènes variés. Chaque processus mutationnel peut impliquer des dommages ou des mécanismes de modification, réparation et réplication de l'ADN qui comporte des SNVs, des InDels, des SCNVs et des réarrangements du génome. Un cancer peut avoir été généré par plusieurs processus mutationnels et donc incorporer

de multiples signatures. L'étude d'Alexandrov *et al.* en 2020, dans le cadre de l'initiative Pan-Cancer Analysis of Whole Genomes (PCAWG), a permis d'identifier un large répertoire de signatures mutationnelles associées à divers cancers et cause étiologiques¹²⁷.

Les cancers cutanés, incluant les CTCLs et le MF, ont montré des associations avec les signatures liées aux dommages causés par les UVs^{17,92} et aux processus de vieillissement¹⁷. En 2015, un séquençage génomique complet de cinq patients présentant un MF a montré des mutations reliées à trois signatures mutationnelles induites par la lumière UV. Il a été proposé que l'exposition aux UVs puisse accélérer la progression du MF, avec des mutations *TP53* liées à la signature UV dans le MF avancé^{92,128}. Cette recherche a également révélé des mutations récurrentes dans la voie de signalisation du récepteur IL-2 impliquant les gènes *JAK3* et *STAT5*, suggérant leur rôle clé dans le MF. Des SCNVs sont observées pour le gène *ZEB1*, un inhibiteur de la transcription de l'IL-2 et intervenant dans le remodelage du cytosquelette, chez deux patients avec une perte d'hétérozygotie^{92,129}.

Une analyse, plus récente de 2021, ayant étudié plusieurs échantillons MF à des stades LR et HR, a aussi illustré l'implication des signatures mutationnelles reliées à l'âge, les UVs et des agents cytotoxiques alkylants. Cette observation confirme que plusieurs processus mutationnels interagissent dans l'évolution de la maladie¹⁷.

1.3.8 Épigénétique

Le microenvironnement tumoral et les modifications épigénétiques influencent fortement l'évolution des CTCLs et sur le MF⁹⁷.

Les processus épigénétiques interviennent dans la régulation génique, la progression tumorale sans modification de la séquence d'ADN et sont donc réversibles. Ces processus se résument en cinq groupes¹³⁰ :

- La méthylation de l'ADN : ajout d'un groupement méthyle CH₃ sur les cytosines (C) situées dans les îlots dinucléotides CpG. Elle cible souvent la région promotrice des gènes. En général, l'hyperméthylation conduit à la répression des gènes suppresseurs de tumeur, tandis que l'hypométhylation peut activer des oncogènes⁶⁸.
- L'acétylation et la désacétylation des histones : l'acétylation ouvre la chromatine par l'ajout d'un groupement acétyle sur les résidus lysines des histones, les protéines autour desquelles l'ADN est enroulé. L'ouverture de la chromatine facilite l'expression des gènes. À l'inverse, la désacétylation referme la chromatine et réprime l'expression des gènes.
- La méthylation des histones : la méthylation principalement des résidus lysines des histones affecte l'état de la chromatine et déclenche l'activation et la répression de la transcription des gènes.
- La modification du remodelage de la chromatine/nucléosome : participe à la désorganisation des histones, ce qui influence l'accès à l'ADN des TFs et de l'ARN polymérase.
- Les microARNs : correspondent à de petits fragments d'ARN non codants qui régulent l'expression des gènes de manière post-transcriptionnelle, sans altérer directement la séquence d'ADN. Ils ciblent les ARNm et inhibent leur traduction ou favoriser leur dégradation. Ils jouent également un rôle important dans la régulation d'autres mécanismes épigénétiques en influençant l'expression des

enzymes impliquées dans la méthylation de l'ADN et les modifications des histones.

L'étude de l'épigénétique dans les CTCLs a pris de l'importance au fil des années. En 2005, van Doorn *et al.* avaient identifié à l'aide de puces à îlots CpG chez 28 patients atteints de CTCLs (MF avancés et lymphoproliférations T CD30⁺) des hyperméthylations dans les promoteurs et à l'intérieur des gènes suppresseurs de tumeurs suivants : *BCL7a* (48%), *PTPRG* (27%), *THBS4* (52%), *p73* (48%), *p16* (33%), *CHFR* (19%), *p15* (10%) et *TMS1* (10%)¹³¹.

Plusieurs mutations impactant des gènes régulateurs de l'épigénétique ont été reportées dans la littérature comme intervenant dans les CTCLs (Figure 9B). Ces gènes impactent plusieurs processus cités précédemment : la méthylation et hyperméthylation de l'ADN (*DNMT3A*^{89,91,98,126}), la méthylation des histones (*KMT2C/KMT2D*^{68,91,92,94,98,118,126} et *SETDB2*¹²⁶), l'acétylation des histones (*CREBBP*^{91,98,126} et *TRRAP*^{98,126}), la déméthylation de l'ADN (*TET2*^{91,98,126}), la déméthylation des histones (*KDM6A*^{98,126}), la désacétylation des histones (*NCOR1*^{92,98,126} et *BCOR*^{90,98,126}), remodelage de la chromatine (*ARID1A*^{91,92,118} et *SMARCB1*^{92,126}) et isolateur des interactions entre les régions régulatrices et leurs gènes cibles (*CTCF*^{89,98,126}).

Un lien important entre l'instabilité des microsatellites (MSI pour *microsatellite instability*) et les altérations épigénétiques a été observé dans les CTCLs. En effet, la méthylation de promoteurs de gènes de réparation, tels que *hMLH1* et *hMLH2*, entraîne une perte de fonction de ces gènes. Par exemple, l'hyperméthylation de *hMLH1* est liée à une réduction de son expression, ce qui perturbe la réparation des mésappariements et favorise l'accumulation de mutations¹³². Dans l'étude de

Scarisbrick *et al.* 2000 une MSI a été observée dans 13 des 54 patients (24%), dont 12 atteints d'un MF et 1 avec un SS. De plus, la MSI est plus associée au MF de stade HR (6/15, 40% des patients HR possédaient une MSI) par rapport au stade LR (6/29, 21% des patients LR avaient une MSI)¹⁰¹. Cette augmentation de la MSI dans les stades HR pourrait être un facteur contribuant à l'agressivité et à la progression rapide de la maladie.

Chapitre 2

Objectifs de la Thèse

Sommaire

2.1	Identification des Biomarqueurs Impliqués dans la Progression du MF	60
-----	---	----

2.1 Identification des Biomarqueurs Impliqués dans la Progression du MF

Au niveau moléculaire, les facteurs expliquant la variabilité de l'évolution clinique des patients atteints de MF restent insuffisamment étudiés. De plus, le paysage génomique et la clonalité du MF restent encore mal compris, tandis que les mécanismes qui déterminent la progression de la forme LR vers la forme HR avec ou sans cellules tMF, demeurent inconnus. Le diagnostic précoce du MF n'est pas toujours évident, car les premiers symptômes peuvent être confondus avec ceux d'autres dermatoses plus courantes telles que l'eczéma ou le psoriasis^{22,23,133}. Repérer au plus tôt les potentiels marqueurs de progression est essentiel pour une meilleure prise en charge des patients et améliorer leur pronostic. Déterminer ces facteurs de risque permettrait d'identifier de nouvelles cibles thérapeutiques afin d'intervenir avant que la maladie n'atteigne des stades avancés.

Hypothèse : Le séquençage par WES à différentes étapes de la maladie et l'étude de leur évolution clonale nous permettra d'identifier les marqueurs de progression clés et les cibles thérapeutiques du MF.

Afin de détecter les moteurs moléculaires de la progression et de l'agressivité de la maladie, nous avons sélectionné des échantillons de patients atteints de MF diagnostiqués à différents stades de la maladie. Nous avons réalisé un séquençage WES des séquences ADN de 67 échantillons tumoraux provenant de 48 patients, dont 18 échantillons séquentiels provenant de 13 patients, y compris 9 échantillons de 7 patients dont la maladie a progressé du stade LR au stade HR. Les échantillons séquentiels nous ont permis d'analyser l'évolution clonale des altérations génétiques liées à la progression du MF. De plus, ces échantillons ont été utilisés pour caractériser le paysage génomique des gènes potentiellement *drivers* qui contribuent à la

progression des patients atteints de LR vers le stade HR. L'objectif principal de cette étude est d'identifier les mutations impliquées dans le MF, d'examiner leur rôle à la fois dans les stades LR et HR, et d'évaluer leur impact sur la progression et la survie des patients. En outre, nous examineront l'influence de l'évolution clonale et des signatures mutationnelles de ces gènes.

Cette étude a été approuvée par l'*Institutional Review Board* (IRB) du CHU de Lille et a reçu un avis favorable du Comité de protection des personnes du Nord-Ouest IV (protocole #ECH18/03) conformément à la Déclaration d'Helsinki. Les échantillons WES utilisés dans cette étude ont été déposés dans la BdD SRA (*Sequence Read Archive*, NCBI, *National Institutes of Health*, *National Library of Medicine*) avec le numéro d'accès : PRJNA925900.

Chapitre 3

Résultats

Sommaire

3.1	Échantillons Tumoraux, Normaux et Données Cliniques	64
3.1.1	Patients et Échantillons	64
3.1.2	Panel d'Échantillons Normaux	66
3.1.3	Données Cliniques.....	67
3.2	Séquençage, Alignement, Contrôle de Qualité et Contamination	69
3.2.1	Construction des Bibliothèques et Séquençage	69
3.2.2	Plateforme Terra du Broad Institute et Langage de Workflow	70
3.2.3	Contrôle de Qualité et Alignement	71
3.2.4	Contamination.....	77
3.2.4.1	Contamination entre Échantillons.....	77
3.2.4.2	Inversion des Échantillons.....	78
3.3	Identification des Marqueurs de Haut Risque de Progression	81
3.3.1	Détection des SCNVs	81
3.3.2	Détection des SNVs et des InDels	82
3.3.3	Filtrage avec les PoNs.....	84
3.3.4	Contamination <i>Tumor in Normal</i>	85
3.3.5	Annotation des Mutations Somatiques	85
3.3.6	Identification des Gènes Fréquemment Mutés	86
3.3.7	Retrait des <i>Fishy Genes</i>	87
3.3.8	Estimation des Signatures Mutationnelles	88
3.4	Pureté, Ploïdie et Fraction de Cellules Cancéreuses et Filtres Internes	89
3.4.1	Estimation de la Pureté, la Ploïdie et la Fraction de Cellules Cancéreuses.....	89
3.4.2	Filtres Internes	90
3.4.2.1	Filtrage avec GnomAD	90

3.4.2.2 Filtrage avec les Fichiers FASTQ Normaux	90
3.5 Analyses Statistiques	91
3.6 Paysage Génomique.....	92
3.6.1 Mutations Nucléotidiques Somatiques Putatives	92
3.6.1.1 Nombre de Mutations Somatiques à Chaque Étape du Pipeline	92
3.6.1.2 Charge Tumorale Mutationnelle	95
3.6.1.3 Gènes Fréquemment Mutés.....	97
3.6.2 Variations du Nombre de Copies Somatiques Récurrentes.....	101
3.6.2.1 Identification des Altérations Significativement Récurrentes	101
3.6.2.2 Gènes <i>Drivers</i> Identifiés dans les SCNVs Focaux.....	102
3.6.2.3 État des connaissances des Bandes Cytogénétiques dans le MF	104
3.6.3 Signatures Mutationnelles.....	108
3.6.4 Biomarqueurs Impactant la Survie Globale et la Progression.....	112
3.6.4.1 Survie Globale de la Cohorte	112
3.6.4.2 Association entre Biomarqueurs avec Stade et Pronostic	113
3.6.5 L'Architecture Clonale et la Phylogénie du MF	118
3.6.6 Modifications des Voies de Signalisations	123

3.1 Échantillons Tumoraux, Normaux et Données Cliniques

3.1.1 Patients et Échantillons

Nous avons examiné 67 échantillons tumoraux provenant d'une cohorte de 48 patients diagnostiqués avec un MF au CHU de Lille entre 2003 et 2017. Ces patients étaient suivis soit dans le service de dermatologie du Professeur Mortier, soit dans celui d'hématologie du Professeur Facon. L'âge médian des patients était de 62,5 ans (de 19 à 94 ans), avec une proportion de 60% d'hommes (29 hommes contre 19 femmes) (Figure 10A-C).

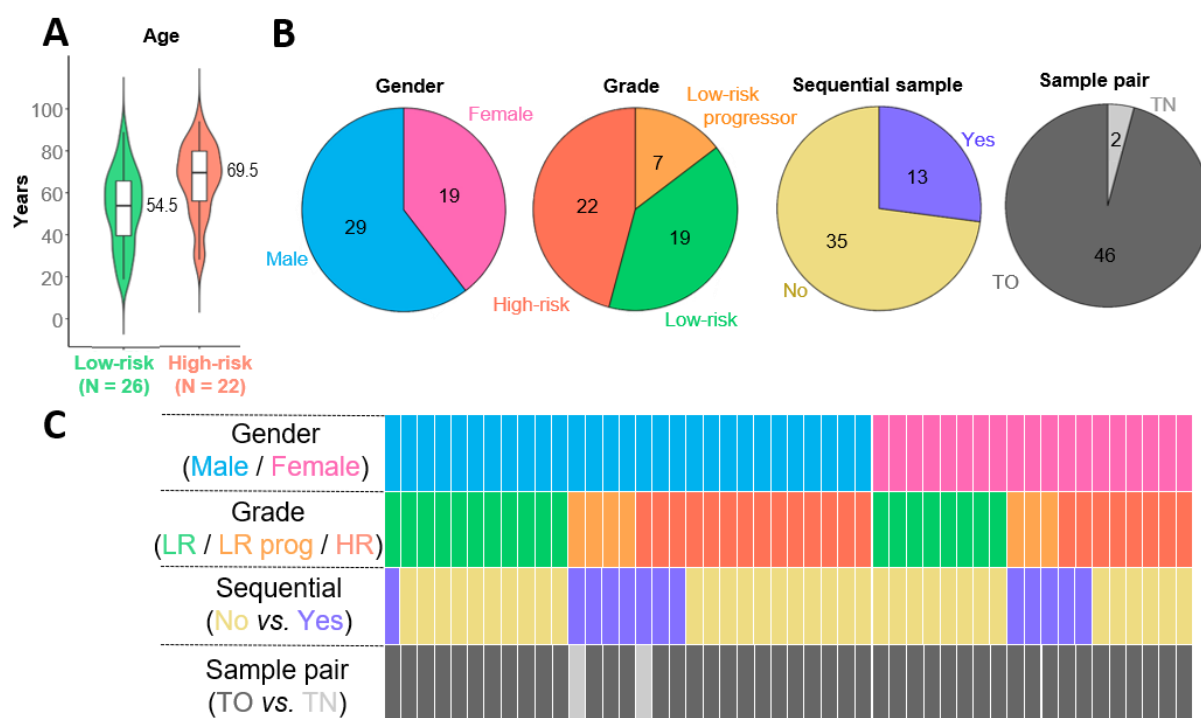


Figure 10 : Composition de la cohorte de patients. **A.** Diagramme en violon illustrant la distribution de l'âge au sein de la cohorte des patients atteints de MF. L'âge médian de cette cohorte est de 62,5 ans. Les patients de LR présentent un âge médian de 54,5 ans, tandis que ceux de haut risque (HR) possèdent un âge médian de 69,5 ans. **B.** La cohorte est caractérisée par une prédominance masculine. Parmi les 48 patients, 22 appartiennent au groupe HR et 26 au groupe LR (incluant les LR et les LR *progressors*). De plus, 27% des patients (13/35) ont fourni plusieurs échantillons séquentiels au cours de leur suivi. La grande majorité des échantillons tumoraux (96%) n'ont pas d'échantillons normaux appariés (TO pour *Tumor-Only* ; TN pour *Tumor-Normal*). **C.** Distribution des échantillons selon le genre, le grade, si l'échantillon est séquentiel et le type d'analyse par patient.

Tous les cas ont été réévalués par des spécialistes du réseau GFLEC, qui ont effectué une analyse histologique approfondie pour diagnostiquer le MF, complétée par une PCR visant à détecter la recombinaison clonale du gène du TCR. L'IHC a été réalisée en utilisant un panel comprenant les marqueurs suivants : CD20, PAX5, CD2, CD3, CD4, CD5, CD7, CD8, CD30, PD1 et Ki-67 (Tableau 2). Une analyse IHC supplémentaire du CD25 a été effectuée sur des échantillons provenant de patients présentant ou non un gain10p15.1 (impliquant les gènes *IL2RA* et *IL15RA*).

Nous avons défini deux groupes de pronostic : le groupe LR, comprenant les stades TNMB IA, IB et IIA, et le groupe HR, comprenant les stades TNMB élevés (IIB, III et IV) ou une histologie transformée (Figure 10B-C, Tableau 6). Les échantillons ont été extraits de biopsies congelées et conservés dans la tumorotheque du C2R2 de l'unité de Biologie-Pathologie-Génétique (CBP) du CHU de Lille (certification NF 96900-2014/65453-1). Pour affiner la pureté tumorale, les échantillons ont été macrodisséqués par la Pr. Marie-Christine Copin et le pourcentage de cellules tumorales a été estimé visuellement par observation microscopique. L'extraction de l'ADN a été supervisée par la Pr. Stéphanie Poulain au CBP. Au total, la cohorte comprenait 70 échantillons (67 cutanés et 3 ganglionnaires) provenant de 49 patients :

- 29 échantillons provenaient de 26 patients LR, dont 9 patients avaient ultérieurement progressé en HR. Pour 7 de ces patients, des échantillons séquentiels ont été prélevés à différents stades de progression de la maladie dans le cadre de notre étude (Figure 11, Figure 12).
- 38 échantillons provenaient de 22 patients HR. Parmi ces échantillons, 7 avaient une maladie limitée, 15 une maladie étendue et 18 une maladie transformée, avec 5 échantillons présentant un SS.

- 3 échantillons cutanés étaient normaux. Deux des 49 patients avaient fourni un échantillon normal en plus de leur échantillon tumoral, tandis qu'un patient avait fourni un échantillon normal sans échantillon tumoral disponible pour ce dernier patient (Figure 11, Figure 12).

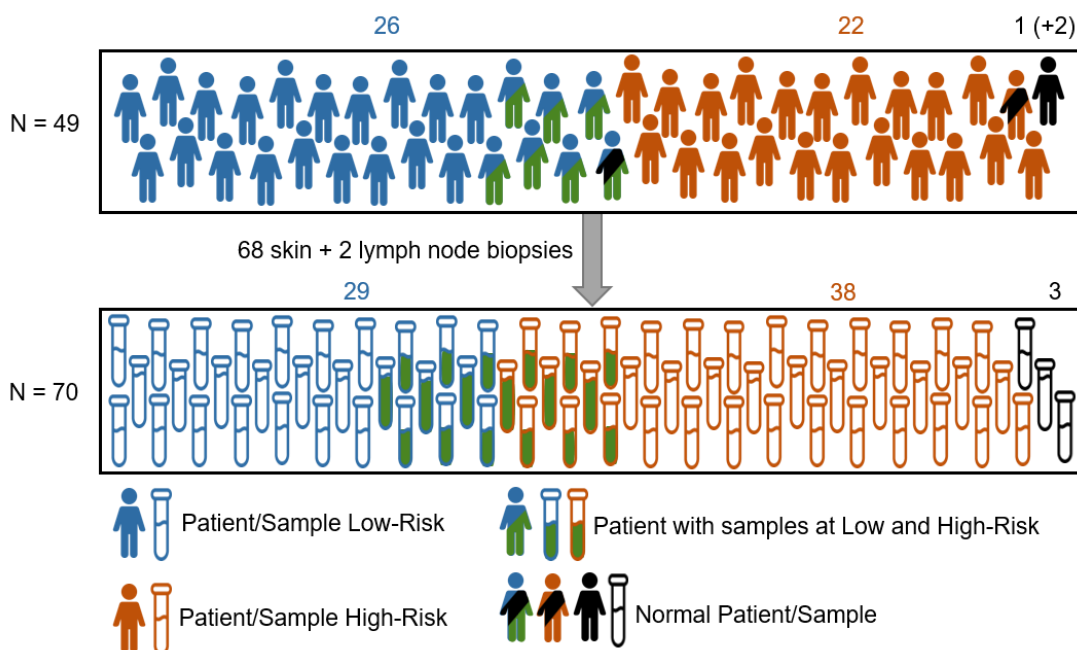


Figure 11 : Répartition des groupes de patients et des échantillons. Ce schéma présente la répartition détaillée des patients et des échantillons selon les deux groupes de risque : LR, en bleu et HR, en orange. Les patients ayant évolué de LR à HR au cours de leur suivi sont indiqués en vert (LR *progressors*), et les échantillons associés à ces patients sont également marqués en vert (ce qui signifie qu'au moins un échantillon est classé dans le groupe LR et au moins un dans le groupe HR par patient). En noir sont représentés les patients pour lesquels un échantillon normal a été obtenu.

3.1.2 Panel d'Échantillons Normaux

En plus des trois échantillons normaux de notre cohorte, dix échantillons de sang humain supplémentaires ont été fournis par le *Research Blood Components* (Watertown, Massachusetts, États-Unis). Ces treize échantillons ont servi à construire un panel d'échantillons normaux (PoN, pour *Panel of Normal*) afin de filtrer les artefacts éventuels survenus lors des étapes de préparation des échantillons en paillasse ou lors du séquençage (Figure 12).

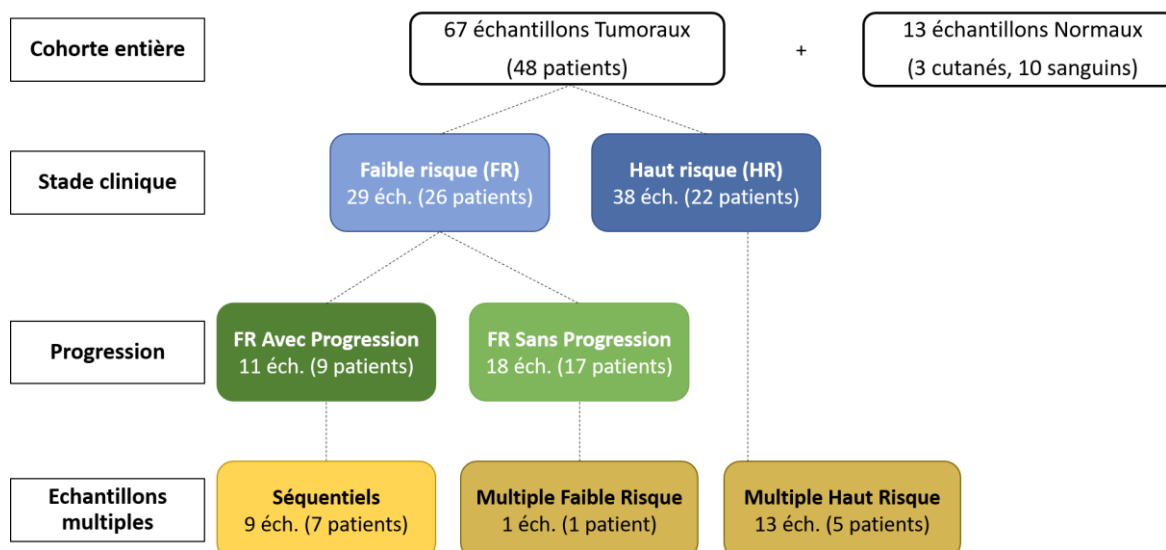


Figure 12 : Échantillons tumoraux et normaux sélectionnés pour le plan d'étude du MF. Le nombre de patients est indiqué entre parenthèses. Les échantillons tumoraux sont classés en fonction de leur stade clinique (LR ou HR). Pour les échantillons LR, leur classement est en fonction de la progression du patient vers le stade HR au cours de son suivi.

3.1.3 Données Cliniques

Pour chaque patient, les données cliniques sont disponibles, rassemblées et organisées par la Dr. Ines Arib dans le cadre de sa thèse de médecine en octobre 2021 (Tableau 6). Pour chacun des 48 patients, le temps du diagnostic correspond à la date de prélèvement du premier échantillon analysé. Cette date est utilisée comme point de départ pour les analyses de progression de la maladie et de survie des patients. La période de suivi s'étend du diagnostic à la date des dernières nouvelles.

	Population globale	MF de faible risque	MF de haut risque
Nombre de patients avec un échantillon tumoral (%)	48 (100)	26 (54)	22 (46)
Genre			
Homme (%)	29 (60)	15 (58)	14 (64)
Femme (%)	19 (40)	11 (42)	8 (36)
Age en années : Médiane (Minimum-Maximum)	63 (19-94)	54 (19-89)	70 (29-94)
< 60 ans (%)	21 (44)	15 (58)	6 (27)
> 60 ans (%)	27 (56)	11 (47)	16 (72)
Stade			
IA	11	10	1
IB	20	16	4
IIA	2	0	2
IIB	7	0	7
IIIA	3	0	3
IIIB	2	0	2
IVA1	2	0	2
IVA2	1	0	1
IVB	0	0	0
Histologie			
Réticulose pagétoïde	0	0	0
Chalazodermie granulomateuse	0	0	0
MF classique	24	21	3
Mycosis folliculotrope	6	5	1
MF transformé	18	0	18
Nombre d'échantillons tumoraux	67	29	38

Tableau 6 : Informations cliniques et démographiques des patients étudiés atteints d'un MF diagnostiqué. Les caractéristiques cliniques des patients comprennent le genre, l'âge, le stade de la maladie, l'histologie et le nombre d'échantillons tumoraux. Les nombres sont détaillés pour la population générale et pour les sous-groupes de LR et de HR.

3.2 Séquençage, Alignement, Contrôle de Qualité et Contamination

3.2.1 Construction des Bibliothèques et Séquençage

La construction des bibliothèques et le séquençage ont été réalisés par le Dr. Ankit Dutta en octobre 2019 au sein de l'équipe Ghobrial au DFCI. La quantité d'ADN extrait a été mesurée par spectrophotométrie en utilisant le kit "Quant-iT PicoGreen dsDNA Assay" (Thermo Fisher Scientific).

Pour le WES, la fragmentation de l'ADN et la sélection hybride sont utilisées pour capturer et enrichir des régions génomiques ciblées avant le séquençage. Dans un premier temps, 100 ng d'ADN génomique ont été fragmentés à l'aide d'un sonicateur à ultrasons de la marque Covaris. Nous obtenons des fragments d'environ 150 à 200 paires de bases (bp pour *base pair*) pour faciliter les étapes de capture et d'amplification.

Dans un deuxième temps, la sélection hybride est réalisée pour isoler les régions d'intérêt à partir ces fragments. La capture des régions cibles (*targets*) a été effectuée à l'aide du SureSelect Human All Exon V7 d'Agilent (Agilent Technologies), ainsi que du kit d'amorces adaptateurs Dual-Index d'Illumina. Cette technologie repose sur l'utilisation de sondes biotinylées (*baits*) d'environ 120 pb, conçues pour être complémentaires des séquences *targets* dans le génome. Les fragments d'ADN contenant les *targets* s'hybrident spécifiquement aux baits lors de l'hybridation, puis sont capturés via des billes magnétiques recouvertes de streptavidine qui se lient aux baits biotinylés, permettant d'isoler les fragments d'intérêt du reste du génome.

Les séquences capturées sont ensuite amplifiées par PCR pour enrichir les *targets*, créant des bibliothèques prêtes pour le séquençage. Ces bibliothèques d'ADN ont été préparées avec le kit SureSelect XT low input d'Agilent. La génération des

bibliothèques a été automatisée grâce au robot de manipulation de liquides Bravo NGS, conformément aux recommandations du fournisseur. Les bibliothèques finales ont été évaluées avec le kit Agilent High Sensitivity DNA Analysis sur le Bioanalyzer 2100 et quantifiées par PCR quantitative à l'aide du kit Kapa Library Quantification (Roche). Les bibliothèques ont ensuite été regroupées et séquencées en mode *paired-end* (2 × 100 bp) sur une plateforme Illumina NovaSeq 6000 sur une flowcell S4.

3.2.2 Plateforme Terra du Broad Institute et Langage de Workflow

Les données ont été déposées sur la plateforme cloud « Terra » du BI (anciennement FireCloud), qui propose une large gamme d'outils bioinformatiques pour l'analyse des données génomiques. Cet environnement de calcul permet de stocker et gérer des jeux de données génomiques volumineux et de pouvoir exécuter des analyses bioinformatiques complexes à l'aide de ressources de calcul intensif.

Les pipelines bioinformatiques sur Terra utilisent un langage de script standardisé appelé *Workflow Description Language* (WDL) qui automatise et relie les différents logiciels et étapes d'analyses. Ce langage permet également la reproductibilité des analyses et facilite le partage entre utilisateurs, dont des échanges de codes modifiés. Ainsi, Terra intègre directement les pipelines écrits en WDL, permettant ainsi de lancer, gérer et surveiller à distance les analyses de données (Figure 13).

Summary	Data	Analysis	Notebooks	BETA	Method Configurations	Monitor
---------	------	----------	-----------	------	-----------------------	---------

Method Configuration Name

Hotspots_by_file

Referenced Method

Namespace: lfelechon
Name: Hotspots
Snapshot ID: 21
Entity Type: Workflow
Source: FireCloud

Created: 15 mars 2022 à 14:42
Owners:
Synopsis: Counts number of time a variant appears in fastq files
Snapshot Comment:

▼ Documentation:
No documentation provided

WDL: [Collapse](#)

```

7
8 # WORKFLOW DEFINITION
9
10 workflow HotCountFastqWorkflow {
11   String group_id
12   File fastq1_gz_files
13   File fastq2_gz_files
14   File VCF_with_tumor_variants
15   File referenceFasta
16   File referenceFastaIndex
17   Int memoryGb
18   Int diskSpaceGb
19   Int cpu
20   Int maxRetries
21   Int preemptible_tries
22
23   call vcf2design as vcf2designTumorAltVariants {
24     input:
25       id = group_id,
26       inputVCF = VCF_with_tumor_variants,
27       ref = referenceFasta,
28       index = referenceFastaIndex,
29       memoryGb = memoryGb,
30       diskSpaceGb = diskSpaceGb
31   }

```

Figure 13 : Page de configuration des méthodes sur l'espace de travail personnel dédié au MF sur la plateforme FireCloud (Terra). Les onglets permettent de naviguer entre les différentes fonctions principales de la plateforme : « Summary » fournit une description personnalisable de l'étude, « Data » permet d'accéder aux jeux d'échantillons et aux résultats des analyses effectuées. L'onglet « Method Configurations » affiche la liste des pipelines utilisés pour l'analyse, offre des options d'édition et permet de parcourir les pipelines publics partagés par autres utilisateurs. À l'écran, un extrait de code personnel rédigé en WDL permet de compter les occurrences d'un variant dans un fichier FASTQ. Enfin, l'onglet « Monitor » permet de suivre l'avancement des analyses en cours et aide à identifier certains bugs pour les analyses qui ont échouées.

3.2.3 Contrôle de Qualité et Alignement

L'analyse s'est initialement appuyée sur l'articulation des logiciels proposés dans le pipeline *CGA_WES_Characterization_Pipeline*, développé par l'équipe du

Professeur Gad Getz du BI. Cependant, ce pipeline présentait des limitations, notamment dans les outils de contrôle de qualité (QC pour *Quality Control*), une étape essentielle pour garantir que les données générées sont fiables et utilisables pour des analyses bioinformatiques ultérieures. De plus, le pipeline avait d'abord été conçu pour fonctionner avec des échantillons *Tumor-Normal* appariés (TN), puis ensuite adapté pour des échantillons *Tumor-Only* (TO), ce qui fait que certaines parties ne sont pas optimisées pour fonctionner en mode TO et ont demandé des ajustements de notre part. Les modifications ont nécessité des changements dans les paramètres de chaque logiciel, dans le code du pipeline articulant les logiciels entre eux et dans l'ajout de certains outils externes QC.

Par exemple, pour améliorer la qualité des lectures issues du séquenceur, nous avons intégré l'outil CutAdapt¹³⁴ afin de supprimer les amorces classiques Illumina, les lectures très courtes (< 30 bp), ainsi que celles de qualité médiocre (Figure 14).

```
command <<<
set -euxo pipefail

awk 'NR==2 {print length(); exit}' <(zcat ${read1_fastq}) > read1_length.txt
awk 'NR==2 {print length(); exit}' <(zcat ${read2_fastq}) > read2_length.txt

mkdir output
cutadapt \
  -m 30 \
  -q 20,20 \
  -a ${read1_adapter} \
  -A ${read2_adapter} \
  -o output/${read1_name} \
  -p output/${read2_name} \
  -j ${runtime_cpu} \
  ${read1_fastq} \
  ${read2_fastq}
>>>

output {
  File clipped_read1_fastq = 'output/' + read1_name
  File clipped_read2_fastq = 'output/' + read2_name
  Int unclipped_read1_length = read_int('read1_length.txt')
  Int unclipped_read2_length = read_int('read2_length.txt')
}
```

Figure 14 : Intégration de l'outil CutAdapt dans la section *Quality Control* du pipeline.

Afin d'améliorer l'étape l'alignement, le logiciel va supprimer les adaptateurs Illumina connus aux l'extrémités 3' des deux paires de lectures (-a et -A). Il élimine également les lectures trop petites inférieures à 30bp (-m 30) et tronque les extrémité des lectures de moins bonne qualité dans les deux sens, en fonction du Q score (voir plus loin) (-q 20,20).

La qualité des données brutes de séquençage au format FASTQ a été évaluée avec le logiciel FastQC¹³⁵ et visualisée au niveau de la cohorte entière avec MultiQC¹³⁶. Le séquençage WES a généré une moyenne de 152M (millions) de lectures par échantillon.

FastQC fournit une vue d'ensemble des données au format FASTQ sous forme de graphiques et des tableaux récapitulatifs permettant d'évaluer la qualité des lectures et de détecter d'éventuels problèmes. *FastQC* génère d'abord des statistiques de base, puis il encode chaque base nucléotidique avec un symbole ASCII (Code américain standardisé pour l'échange d'information) pour attribuer un score de qualité Phred de séquençage, également appelé Q score.

Le Q score, généralement compris entre 0 et 40, est une mesure logarithmique couramment utilisée pour connaître la précision du séquenceur. Il indique la probabilité qu'une base ait été incorrectement appelée¹³⁷. Mathématiquement, ce score est défini par la relation suivante, où P est la probabilité d'une erreur dans l'appel de la base¹³⁸ :

$$Q = -10\log_{10}P$$

Par exemple, si le Q score d'une base est de 30 (Q30), cela signifie qu'il y a une erreur toutes les 1000 bases ($P = 0,001$), soit une précision de l'appel de la base de 99,9%. Ainsi, un Q Score de Q30 indique un taux erreur et d'ambiguïté faible.

Pour nos données, le Q score est excellent pour l'intégralité des lectures, aux alentours de Q35 (99,97% de précision) pour la majorité de la longueur des séquences. Les erreurs de séquençage sont plus fréquentes aux extrémités ce qui est un problème inhérent au séquenceur dû à la baisse du signal ou au phasage durant le séquençage. Cependant, le Q score ne descend pas en dessous de Q25, témoignant d'un séquençage réussi et d'un appel des bases fiable (Figure 15).

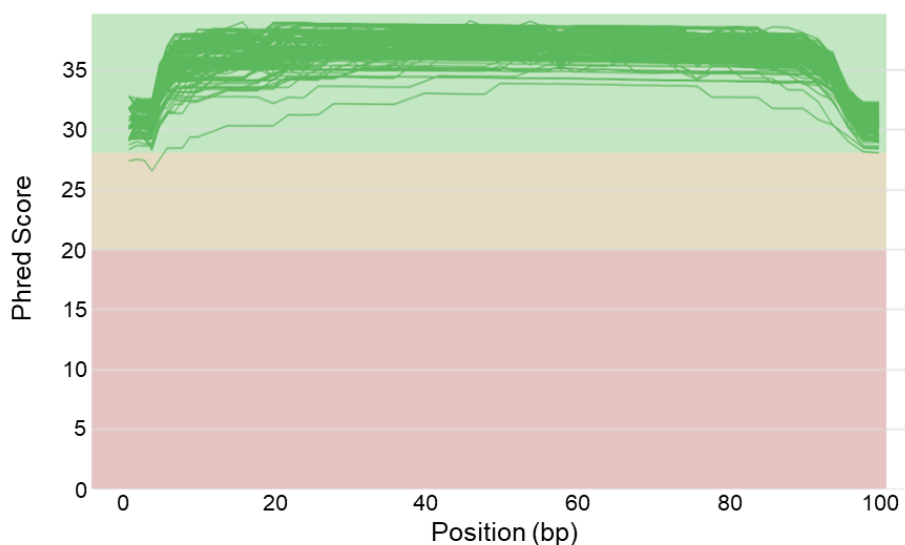


Figure 15 : Q Score sur l'intégralité des lectures séquencées.

Les séquences ont été alignées avec BWA-MEM (*Burrows-Wheeler Aligner-Maximal Exact Matches*)¹³⁹, et la qualité de l'alignement a été estimée à l'aide des outils de la suite logicielle Picard du BI. En moyenne, les échantillons tumoraux possèdent une couverture de 99,72% sur le génome de référence hg19. La médiane du nombre de lectures séquencées par échantillon est de 145M (en millions, avec le nombre de lectures divisé par 2 pour les données *paired-end*). Nous pouvons noter que 9 échantillons possèdent un nombre de lectures inférieur à 100M (36_MF, 48_MF, 42_MF, 9_MF, 21_MF, 40_MF, 67_MF, 33_MF et 51_MF), ce qui peut être expliqué soit par une qualité moindre de l'ADN pour ces échantillons ou alors des problèmes survenant lors de la préparation ou le séquençage qui affecterait certaines bibliothèques (Figure 16).

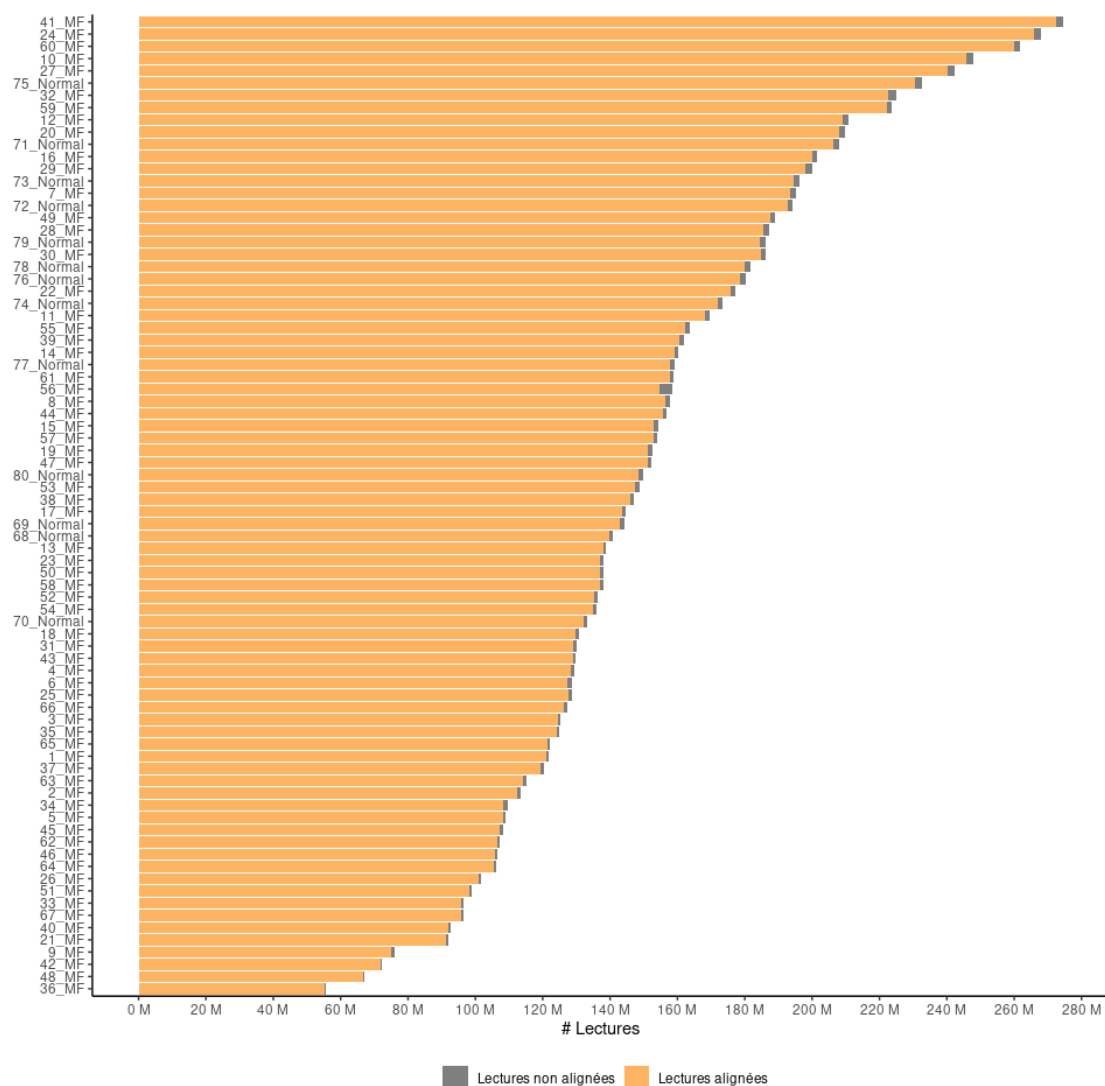


Figure 16 : Alignement des lectures séquencées. Le nombre de lectures est divisé par 2 pour les données *paired-end* et est indiqué en millions (M). Le nom des échantillons est indiqué en ordonnée. Les lectures en orange ont passé les filtres de qualité du séquenceur Illumina (PF pour *Pass Filters*) et qui ont pu être alignées par l'aligneur BWA-MEM contre le génome de référence hg19. Les lectures en gris sont n'ont soit pas passé les PFs, soit étaient du bruit (séquence que d'une seule base nucléotide ou N, etc.) et qui n'ont, de ce fait, pas pu être alignées.

Le processus de nettoyage de l'alignement inclus l'élimination des doublons de PCR avec MarkDuplicate, le réalignement local autour des InDels avec IndelRealigner, et le recalibrage de la qualité des bases avec BaseRecalibrator et ApplyBQSR¹⁴⁰. Tous les échantillons ont satisfait les seuils de QC et présentaient une profondeur de

séquençage moyenne de 231,85X. Nous avons deux échantillons avec une profondeur inférieure à 100X : 9_MF (66X) et 51_MF (76X) (Figure 17).

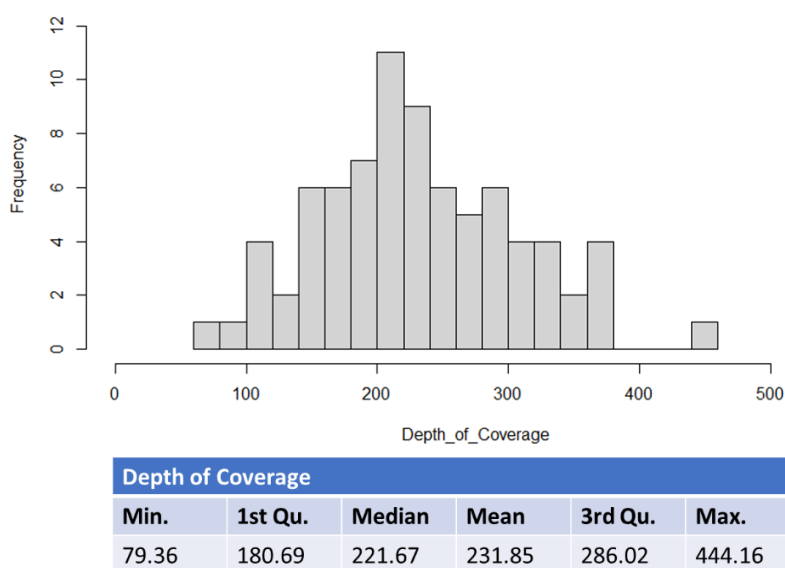


Figure 17 : Profondeur de couverture des échantillons MF après séquençage.

Enfin, les artefacts d'oxydation de la guanine en 8-oxoguanine (8-oxoG), qui apparaissent souvent à de faibles fréquences alléliques (AF pour *allelic frequency*), peuvent survenir lors de la préparation des bibliothèques sous l'effet combiné de la chaleur, de la fragmentation de l'ADN et de la présence de contaminants métalliques. Ces artefacts ont été détectés et éliminés avec l'outil CollectOxoGMetrics de la suite Picard. L'oxydation des guanines (G) sur le brin de référence va modifier ses propriétés d'appariement et augmenter leurs affinités de liaison avec les bases adénines (A) plutôt qu'avec les cytosines (C). Ce processus entraîne des artefacts G>T pour la lecture 1 et C>A pour la lecture 2 lors de l'étape d'enrichissement des fragments par PCR¹⁴¹. CollectOxoGMetrics a permis de mesurer la potentielle présence inhabituellement élevée du taux G>T/C>A dans les données de séquençage et a pu filtrer les lectures endommagées pour éviter les erreurs d'interprétations.

3.2.4 Contamination

Lors de la préparation des échantillons en laboratoire, divers types de contamination peuvent survenir :

- Contamination croisée inter-échantillons : contamination des cellules d'un individu par les cellules d'autres individus.
- Inversion d'échantillons lors de la préparation : erreur dans la correspondance entre l'identité du patient et celle de l'échantillon.
- Contamination *Tumor in Normal* : contamination de l'échantillon normal par des cellules tumorales.
- Pureté de la tumeur : présence de cellules normales du patient dans l'échantillon tumoral.

L'estimation de la contamination *Tumor in Normal* est effectuée après la détection des mutations somatiques (voir section 3.3.4). La pureté tumorale a été évaluée ultérieurement dans le pipeline à l'aide du logiciel ABSOLUTE (voir section 3.4).

3.2.4.1 Contamination entre Échantillons

La contamination croisée inter-échantillons correspond aux cas où l'ADN provenant d'un individu extérieur s'est mélangé avec celui de l'échantillon analysé. Dans le pipeline, cette contamination est estimée par défaut avec le logiciel du GATK ContEst¹⁴² qui compare des génotypes dans l'échantillon normal avec ceux alternatifs dans le tumoral pour y repérer les discordances. Cependant, ContEst utilise un modèle statistique bayésien qui suppose un génotype diploïde ne supportant pas les variations du CN et qui fonctionne uniquement qu'avec des paires d'échantillons TN appariés. En raison de ces limitations, le pipeline a été modifié pour utiliser un autre outil plus

récent du GATK, CalculateContamination¹⁴³. Ce dernier fonctionne sur un principe similaire à ContEst, mais avec une meilleure performance en présence de variations du CN et qui s'adapte correctement en l'absence de données normales appariées (TO pour *Tumor-Only*). Pour résumer, l'outil analyse les écarts entre les AFs attendues et observées, ainsi que la distribution des allèles dans les marqueurs d'hétérozygotie dans l'échantillon tumoral pour estimer le degré de contamination par un autre ADN. L'utilisation de CalculateContamination a révélé que la fraction de lectures provenant de contamination inter-échantillons dans nos données était très faible, avec une médiane de 0,002. L'équipe du Pr. Gad Getz estime que la contamination est trop importante si elle dépasse un seuil de 0,04 et recommande un nouveau séquençage des échantillons.

3.2.4.2 Inversion des Échantillons

L'analyse d'empreinte, également appelée *Fingerprints*, a été effectuée avec l'outil CrossCheckLane Fingerprints pour détecter d'éventuelles inversions entre les échantillons de patients. Cette méthode compare les génotypes sur des sites de substitution d'un seul nucléotide (SNP pour *Single-Nucleotide Polymorphism*) présents chez plus de 1% de la population générale, afin d'identifier d'éventuels échanges d'échantillons ou de lignes de séquençage entre les patients. Un logarithme des probabilités (LOD pour *logarithm of odds*) est calculé pour chaque paire d'échantillons tumoraux, offrant une mesure de la probabilité relative que deux échantillons proviennent du même individu. Un score $\text{LOD} \geq 15$ indique que la paire d'échantillons provient du même individu. Si $1 \leq \text{LOD} < 15$, il y a une concordance faible entre les deux échantillons, mais pas assez faible pour être identifiée comme problématique. Enfin, si le $\text{LOD} < 1$ entre deux échantillons liés, alors ils appartiennent à des patients différents.

L'analyse a comparé nos échantillons sur 50 sites SNPs clés, sélectionnés pour leur capacité à différencier une personne dans les populations d'Europe du Nord-Ouest¹⁴⁴. Ces SNPs, répartis sur l'ensemble du génome, permettent de distinguer des individus au sein d'une population donnée. Les 50 sites SNPs ont été définis dans une étude de référence à partir d'échantillons de sang provenaient de 250 familles néerlandaises, une cohorte représentative des populations d'Europe du Nord-Ouest¹⁴⁴. Ces SNPs ont été appliqués à nos échantillons pour vérifier leur cohérence génétique et pour détecter d'éventuels échanges accidentels entre individus dans notre plan expérimental.

Au final, cette analyse a confirmé l'absence d'inversion entre les échantillons et n'a identifié aucune erreur inter-échantillons. De plus, les échantillons séquentiels correspondaient bien aux mêmes patients (Figure 18). Bien que quelques valeurs LOD se situent entre 1 et 2,5 pour les échantillons 14_MF, 15_MF et 16_MF (patient 6) et entre 7 et 9 pour les échantillons 37_MF, 38_MF et 39_MF (patient 22), ces valeurs positives indiquent une probabilité élevée que ces échantillons proviennent du même patient. Il est également intéressant de noter que l'échantillon normal 68_Normal est associé aux échantillons tumoraux 17_MF et 18_MF, de même que 69_Normal à l'échantillon 50_MF, confirmant qu'ils proviennent des mêmes patients (respectivement patient 7 et patient 32).

Sample 1

[illegible]

Figure 18 : Matrice d’empreinte des échantillons de MF. Le nom des échantillons est indiqué en abscisse et en ordonnée. A l’intérieur de chaque case se trouve la valeur du logarithme des probabilités (LOD) pour chaque paire d’échantillons. Une échelle de couleurs indique les valeurs positives (vers le jaune), proches de zéro (orange vif) et négatives (vers l’orange foncé et le rouge).

3.3 Identification des Marqueurs de Haut Risque de Progression

3.3.1 Détection des SCNVs

Les SCNVs et les variations alléliques ont été détectées à l’aide du logiciel ModelSegments du BI¹⁴³. En résumé, le logiciel exploite notre PoN de 13 échantillons normaux (10 sanguins et 3 cutanés) pour distinguer les variations somatiques des variations constitutionnelles. Il segmente le génome dans chaque échantillon tumoral pour repérer les régions altérées et représente les résultats sous la forme de « *Genome-wide Copy ratio plot* » (Graphique du ratio de copies sur l’ensemble du génome). Le logiciel a été exécuté en mode TO pour l’ensemble des échantillons et en mode TN pour les trois échantillons tumoraux disposant d’un échantillon normal apparié.

Les SCNVs fréquents, larges et focaux, ont été identifiés par GISTIC2.0¹⁴⁵. Ce logiciel est particulièrement utile pour détecter les SCNVs statistiquement significatives en différenciant les gains, qui peuvent impliquer des oncogènes, et les délétions, qui affectent généralement des gènes suppresseurs de tumeur. Dans un premier temps, GISTIC utilise un modèle de fond pour estimer le taux d’altérations attendu dans chaque région. Dans un second temps, il va calculer pour chaque segment génomique un score de significativité, converti en q-value, en s’appuyant sur la fréquence et l’amplitude des altérations observées. Nous avons utilisé un seuil de q-value par défaut de 0,01 qui est efficace pour détecter les SCNVs les plus significatifs dans notre cohorte et pour éviter le bruit causé par les altérations *passengers*.

3.3.2 Détection des SNVs et des InDels

Les SNVs ont été détectés par MuTect¹⁴⁶, tandis que les InDels ont été identifiés par Strelka¹⁴⁷ et Mutect2¹⁴⁸. MuTect est reconnu pour sa sensibilité élevée et son faible taux d'artefacts dans la détection des mutations somatiques sur une large gamme de AFs. Habituellement, ce logiciel nécessite un échantillon normal apparié pour détecter les SNVs somatiques et les distinguer des mutations germinales. Toutefois, dans le cas d'analyses en mode TO, où l'échantillon normal est absent, un échantillon extérieur peut être utilisé comme référence pour permettre à MuTect de fonctionner. Bien que cela puisse entraîner une légère diminution de la spécificité du logiciel dans la distinction entre mutations somatiques et constitutionnelles, il reste efficace pour identifier les mutations somatiques. Ainsi, pour nos échantillons sans normaux appariés, nous avons choisi d'utiliser le premier des dix échantillons sanguins normaux comme référence « normale » pour tous les échantillons tumoraux. La première raison de ce choix est qu'un échantillon sanguin est susceptible d'être moins influencé par le microenvironnement tumoral. En effet, les tissus cutanés peuvent être infiltrés par des cellules malignes ou affectés par des modifications épigénétiques liées à l'environnement d'une tumeur avoisinante. La seconde raison est que les tissus cutanés sains peuvent être pollué par les mutations somatiques non liées à la pathologie étudiée qui sont produites par des processus comme l'exposition aux UVs ou le vieillissement. En revanche, le sang, en tant que tissu plus homogène et moins exposé aux facteurs environnementaux, constitue une meilleure référence pour filtrer les mutations constitutionnelles avec Mutect.

Strelka a été utilisé pour la détection des InDels somatiques en réalignant des lectures après la détection des InDels candidats et en utilisant un modèle de vraisemblance bayésien pour appeler les InDels. Par mesure de précaution, un double

contrôle est effectué par Mutect2 qui va appeler indépendamment les InDels, et seuls ceux détectés par les deux logiciels ont été retenus, afin d'assurer une plus grande fiabilité des résultats.

Les fichiers de sortie de ces logiciels sont au format VCF (*Variant Call Format*) et est utilisé pour lister l'ensemble des variants identifiés comme SNVs ou InDels. Chaque variant est accompagné d'informations supplémentaires, comme la position génomique, le type de mutation et des annotations de qualité. Pour la suite de l'analyse, ces fichiers VCF issus de ces différentes analyses sont combinés dans un unique fichier au format MAF (*Mutation Annotation Format*). Ce format est largement utilisé dans les analyses de cancer pour regrouper en un seul fichier les informations sur les mutations pour l'ensemble de la cohorte. Il correspond en un tableau standardisé où chaque ligne est un variant associé à l'échantillon dans lequel il a été identifié et chaque colonne contient des informations identifiées par divers outils de détection de variants et d'autres logiciels d'annotation importants. Parmi les colonnes les plus utiles se trouvent les suivantes :

- Hugo_Symbol : Nom du gène impacté par la mutation.
- Chromosome, Start_Position et End_Position : Localisations génomiques du variant.
- Variant_Classification : Nature du variant (*Missense, Nonsense, Frameshift, etc.*) pour aider à repérer les mutations synonymes et non-synonymes.
- Variant_Type : Type de mutation. SNV ou INS (Insertion) ou DEL (Délétion).
- Tumor_Sample_Barcode et Matched_Norm_Sample_Barcode : Identifiants de la paire d'échantillons tumoral et normal utilisée.

- `t_alt_count` et `t_ref_count` : Nombre de lecture pour les allèles mutant et de référence.
- `Protein_Change` : Changement au niveau de la protéine codée dans la séquence protéique.

3.3.3 Filtrage avec les PoNs

Les résultats de MuTect, Strelka et Mutect2 ont été filtrés à l'aide de MAFPoNFilter¹⁴⁹. Cet outil utilise deux PoNs comme une liste noire pour éliminer les variants constitutionnels et les polymorphismes propres à chaque individu. Le premier PoN spécifique à notre cohorte, est composé de treize échantillons normaux WES. Le second est un PoN à accès contrôlé, issu du *The Cancer Genome Atlas Program* – (TCGA)¹⁴⁹, contenant 8 334 échantillons normaux utilisé dans l'analyse de routine du BI. L'application du filtre PoN réduit le taux de variants constitutionnels en augmentant la profondeur de séquençage dans les régions à faible couverture. Le PoN est également efficace pour éliminer les artefacts courants survenant lors du séquençage.

Pour chaque position génomique, le fichier PoN encode la distribution empirique du compte de lecture des allèles alternatifs et des AFs alternatives dans tous les échantillons normaux du PoN. Pour chaque variant appelé dans l'échantillon tumoral, le filtre modélise la AF du variant (VAF) comme une distribution *bêta* paramétrée par ses comptes de lectures alternatives et de références. Une somme pondérée de la distribution *bêta* et de la distribution empirique du PoN à la position génomique du variant sert de score pour cette position. Si le score est supérieur à un certain seuil, le site est signalé comme cohérent avec la distribution empirique observée dans le PoN et le variant est considéré comme constitutionnel. Si toutefois, le nombre de lectures pour l'allèle dans l'échantillon de tumeur était significativement plus élevé que ce qui

a été observé dans le PoN, son score de vraisemblance serait plus faible et le variant serait retenu comme somatique.

3.3.4 Contamination *Tumor in Normal*

En cas de contamination significative des cellules tumorales dans l'échantillon normal, les algorithmes de détection de mutations ont tendance à rejeter les véritables variants somatiques en raison de la présence de lectures dérivées de la tumeur dans le tissu sain apparié. Le logiciel DeTiN (*Decontaminating Tumor-in-Normal*) estime le niveau de contamination des mutations tumorales dans l'échantillon normal apparié¹⁵⁰. Il utilise les SNVs et SCNVs détectés dans l'échantillons tumoral pour établir des distributions de probabilité de contamination *Tumor-in-Normal* (TiN). Un modèle statistique bayésien est utilisé pour tester ces distributions et estimer le score final TiN dans l'échantillon normal. DeTiN améliore ensuite la sensibilité en récupérant les variants somatiques initialement écartés dans les cas touchés par la contamination. De ce fait, ce logiciel est utilisé uniquement lorsque nous possédons l'échantillon normal provenant du même patient que l'échantillon tumoral. Cette étape de récupération fonctionne moins bien lorsque le score de TiN $> 0,3$ et la notice de l'algorithme recommande d'exclure les paires d'échantillons supérieure à ce score. Dans notre cas, l'analyse DeTiN, lancée uniquement sur nos échantillons appariés TN, a révélé un score très faible avec une médiane de 0,01, et aucun échantillon n'a été exclu de l'analyse. De plus, aucun variant n'a été récupéré par DeTin dans nos trois paires échantillons TN.

3.3.5 Annotation des Mutations Somatiques

Comme nous ne connaissons pas l'importance des variants, nous avons besoin de les annoter pour connaître leur sens biologique. Les variants somatiques (SNVs et

InDels) ont été annotés pour leur effet oncogène à l'aide des outils VEP et Oncotator¹⁵¹. VEP est un outil qui permet d'agglomérer les annotations provenant de différentes BdDs dans le but de déterminer l'effet des SNVs et des InDels sur les gènes, les transcrits et les séquences protéiques. Oncotator associe les SNVs et les InDels avec des annotations de protéines génomiques et des variants associés au cancer.

Les annotations proviennent de diverses BdDs, notamment GENCODE pour les annotations génomiques¹⁵², UniProt pour les annotations protéiques spécifiques aux sites¹⁵³, dbNSFP pour les prédictions d'impact fonctionnel¹⁵⁴. De plus, les mutations somatiques identifiées sont comparées avec celles répertoriées dans les BdD COSMIC (*Catalogue of Somatic Mutations in Cancer database*)¹⁵⁵, CCLE (*Cancer Cell Line Encyclopedia*) du BI¹⁵⁶, ClinVar pour les annotations cliniques¹⁵⁷, et enfin dbSNP¹⁵⁸ et 1000Genomes¹⁵⁹ qui recensent les variants non somatiques.

3.3.6 Identification des Gènes Fréquemment Mutés

Les gènes mutés plus fréquemment que par hasard ont été identifiés à l'aide de MutSig2CV¹⁴⁹. Tout d'abord, ce logiciel commence par établir une estimation du bruit de fond mutationnel, qui représente la probabilité qu'un gène accumule des mutations de manière aléatoire, sans lien avec un effet cancérigène spécifique. Pour cela, MutSig2CV prend en compte les caractéristiques intrinsèques des gènes, comme leur longueur et leur composition en bases, ainsi que la fréquence des mutations dans différentes régions du génome. L'estimation de ce bruit de fond va permettre de modéliser le taux mutationnel de fond attendu pour chaque gène. À partir de ce modèle, MutSig2CV mesure l'excès de mutations somatiques observées dans chaque gène par rapport au taux de fond. Il vérifie ensuite si les mutations sont localisées,

pour connaître leur impact fonctionnel. Le logiciel identifie des « *hotspots* », c'est-à-dire, des sites où les mutations se produisent plus fréquemment, suggérant une pression de sélection sélective. À partir de ces informations (bruit de fond, conservation évolutive et présence de hotspots), MutSig2CV calcule pour chaque gène un score de significativité globale, associé à une q-value, une correction de la p-value en prenant en compte le taux de FPs (FDR pour *False Discovery Rate*). Le seuil par défaut de MutSig2CV pour la q-value est fixé à 0,1, ce qui signifie qu'au maximum 10% gènes identifiés comme significatifs peuvent être des FPs. Ce seuil est bien adapté aux analyses TN pour récupérer des gènes qui auraient pu être écartés par erreur par manque de significativité. Cependant, dans notre cas, avec l'utilisation d'un échantillon normal non-apparié, il a entraîné la validation de trop nombreux gènes (plusieurs centaines), réduisant la spécificité de l'analyse. Sur les conseils d'un ingénieur du BI, nous avons donc décidé d'abaisser ce seuil de la q-value à 0,01, limitant ainsi à 1% la proportion maximale de FPs.

3.3.7 Retrait des *Fishy Genes*

Les variants présents sur des gènes reportés pour être des « *fishy genes* », c'est-à-dire des gènes souvent identifiés à tort comme significativement mutés dans les analyses de cancers, ont été retirés manuellement en suivant la liste établie par Lawrence *et al.*, publiée dans *Nature* en 2013¹²². Ces gènes sont généralement considérés comme FPs, car ils sont fréquemment mutés en raison de leur caractéristiques intrinsèques, comme leur taille, leur faible expression ou leur réplication tardive, ce qui les rend susceptibles d'accumuler des mutations de manière aléatoire. Parmi ces *fishy genes* figurent des gènes codant pour de très grandes protéines comme *TTN*, *MUC4* ou *MUC16*, des récepteurs olfactifs souvent muté pour leur faible expression comme *OR2L13* et des gènes à forte variabilité comme *RYR2*,

RYR3, *CSMD1*, *CSMD3* et *PCLO*. Bien que ces gènes reviennent souvent dans les analyses de large envergure, ils sont souvent retirés car leurs mutations n'ont pas de lien direct avec le développement du cancer.

3.3.8 Estimation des Signatures Mutationnelles

Les processus mutationnels produisent des empreintes distinctives appelées signatures mutationnelles dans le génome du cancer, reflétant à la fois des dommages causés à l'ADN et des mécanismes de réparation. Pour estimer les signatures mutationnelles de nos échantillons, le logiciel SignatureAnalyzer^{160,161} a été utilisé sur la liste des gènes fréquemment mutés issue de MutSig2CV.

SignatureAnalyzer identifie le contexte de séquence pour chaque SNV en analysant les nucléotides situés à 5' et 3' du site muté et définit la substitution trinuécléotidique. Plus précisément, le logiciel analyse six types principaux de substitutions nucléotidiques : C→A, C→G, C→T, T→A, T→C et T→G tout en tenant compte du contexte mutationnel (les nucléotides situés avant et après la base mutée dans le sens 5' et 3'). Au total, il existe seize combinaisons possibles de séquences pour chaque mutation, ce qui donne $16 \times 6 = 96$ triplets nucléotidiques uniques. Le logiciel va ensuite visualiser la distribution de toutes les substitutions trinuécléotidiques des échantillons.

SignatureAnalyzer utilise, dans un second temps, l'algorithme de factorisation de matrice non négative (NMF pour *Non-negative Matrix Factorization*)¹²⁷, une analyse bayésienne de décomposition des matrices qui permet d'associer la distribution des substitutions trinuécléotidiques à des signatures mutationnelles sous-jacentes. Ces signatures mutationnelles spécifiques sont appelées « signatures de substitution d'une seule base » (SBS, pour *Single Base Substitution*) sont recensées dans la BdD de COSMIC et correspondent aux mécanismes causaux responsables des mutations.

3.4 Pureté, Ploïdie et Fraction de Cellules Cancéreuses et Filtres Internes

3.4.1 Estimation de la Pureté, la Ploïdie et la Fraction de Cellules Cancéreuses

Le logiciel ABSOLUTE¹⁶² a estimé la pureté et la ploïdie de chaque échantillon en intégrant les données SNVs, SCNVs et les AFs. ABSOLUTE évalue la proportion de cellules tumorales dans l'échantillon et calcule la fraction de cellules cancéreuses (CCF, pour *Cancer Cell Fraction*) associée à chaque altération génomique. Grâce à cette approche, la pureté moyenne de nos échantillons tumoraux a été évaluée à 86%, un niveau élevé qui indique une détection plus efficace des mutations *drivers* et autres événements pathogéniques. En ajustant la ploïdie et la pureté, ABSOLUTE détecte également la sous-clonalité des mutations et le CN absolu des segments d'ADN dans les cellules cancéreuses. Une mutation est considérée comme clonale si sa $CCF \geq 0,9$, ce qui signifie qu'elle est présente dans au moins 90% des cellules tumorales. En revanche, une mutation avec une $CCF < 0,9$, la mutation est dite sous-clonale, c'est-à-dire que sa présence est limitée à une sous-population tumorale.

Pour les échantillons TO, nous avons appliqué un filtre *Log odds Tumor-Only Germline Somatic*, tel que décrit par Chapuy *et al.* en 2018¹⁶³. En résumé, ce filtre utilise plusieurs paramètres, notamment la CCF, la ploïdie et la pureté de l'échantillon et le CN local, afin de calculer la probabilité qu'un variant soit compatible avec un événement de la lignée germinale ou somatique. Le modèle compare la VAF à celle attendue pour un événement hypothétique constitutionnel ou somatique. Ensuite, il applique un seuil sur le rapport de vraisemblance (*log odds*) pour différencier les variants somatiques des variants constitutionnels. Dans notre analyse, nous avons fixé ce seuil à « 0 », comme recommandé par Chapuy *et al.* 2018, afin d'éliminer des FPs supplémentaires provenant de variants susceptibles d'être originaires de la lignée germinale dans nos échantillons tumoraux.

3.4.2 Filtres Internes

3.4.2.1 Filtrage avec GnomAD

Pour exclure les mutations potentiellement constitutionnelles, nous avons utilisé la BdD d'agrégation génomique (*Genome Aggregation Database*) gnomAD v2.1¹⁶⁴. Cette BdD remplace l'ancienne, appelée ExAC (*Exome Aggregation Consortium*)¹⁶⁵ qui était à la base d'un filtre utilisée par Chapuy *et al.* 2018¹⁶³, et qui est maintenant intégrée dans gnomAD. L'utilisation de cette BdD répertorie des variants connus provenant des génomes et exomes d'individus non apparentés issus de diverses populations, dont 125 748 exomes alignés sur le génome GRCh37. Ce grand nombre d'individus permet de filtrer en utilisant la AF des variants pour identifier des mutations germinales, souvent rares mais potentiellement confondues avec des mutations somatiques dans les analyses de cancer.

Pour notre analyse, l'identifiant génomique de la *Human Genome Variation Society* (HGVS) a été utilisé pour extraire la AF des mutations présentes sur gnomAD. Cet identifiant se trouvait dans colonne « HGVS_genomic_change » de nos fichiers MAF, un format de fichier qui regroupe toutes les mutations identifiées par échantillon. En utilisant la BdD *Variant Effect Predictor* (VEP)¹⁶⁶ d'Ensembl du génome de référence GRCh37, nous avons pu accéder aux données gnomAD et appliqué une valeur seuil de AF de 0,0001 pour écarter les fréquences germinales fréquentes.

3.4.2.2 Filtrage avec les Fichiers FASTQ Normaux

D'après les auteurs de la publications Chapuy *et al.* 2018, l'utilisation de l'outil *Tumor-Only* et de gnomAD n'est pas suffisante pour filtrer tous les FPs. En effet, ils ont noté une fréquence moyenne de 3,3/Mb de plus dans les échantillons TO en ce qui concerne les variants constitutionnels après le filtrage¹⁶³. Comme critère final pour

filtrer les variants potentiels de la lignée germinale dans les échantillons TO, nous avons utilisé des scripts bash pour compter l'occurrence des variants mutés dans notre cohorte de treize échantillons du PoN au format FASTQ. Nous avons d'abord converti le fichier MAF en une liste de deux motifs par variant, composée de 11 nucléotides avant et après le variant. Le premier motif représentait le *wild-type* (variant de référence), tandis que le second était la forme mutée (variant alternatif). Ensuite, nous avons utilisé le script bash « `do_it.sh` » du dépôt GitHub « *HotCount* » pour compter le nombre d'occurrences des motifs *wild-type* et mutés dans le PoN. Les variants dont les motifs mutés apparaissaient plus d'une fois dans le PoN ont été exclues. Ce seuil a été choisi pour compenser le nombre réduit d'échantillons dans le PoN, qui avec treize échantillons, reste relativement bas pour supprimer efficacement tout variants constitutionnels et artefacts. Dans une petite cohorte, les variants qui apparaissent plus d'une fois ont plus de chance d'être des constitutionnels récurrents. Ce filtrage strict supplémentaire révèle non seulement des artefacts potentiels supplémentaires, mais permet également être plus rigoureux dans le processus de filtrage, complétant ainsi l'utilisation du PoN dans le pipeline.

3.5 Analyses Statistiques

Les chiffres et les estimations statistiques ont été obtenus à l'aide de R v.3.6.3 et de MATLAB. Les fichiers MAF ont été analysés avec le paquet R « *maftools* » v2.12.05 en utilisant les fonctions « *oncoplot* » (Figure 21), « *maf_compare* » (Forestplot, Figure 29) et *somaticInteractions* (*corrplot*, Figure 33). Les autres figures ont été générées à l'aide des paquets R v3.4.0 « *survival* » v3.4-0 et « *survminer* » v0.4.9. (graphiques de Kaplan-Meier, Figure 28 et Figure 30), ainsi que « *chrisamiller/fishplot* » v0.5.1 (Fishplots, Figure 35) et *ggplot2*.

3.6 Paysage Génomique

3.6.1 Mutations Nucléotidiques Somatiques Putatives

3.6.1.1 Nombre de Mutations Somatiques à Chaque Étape du Pipeline

Notre pipeline et les filtres utilisés nous ont permis de filtrer les variants de la lignée germinale et les artefacts dans les échantillons TO. Ils ont permis d'identifier en moyenne 135 SNVs/InDels dans les 67 échantillons TO (entre 14 et 451) et 178 dans les 3 échantillons TN (entre 83 et 284). Ces résultats sont comparables à d'autres études WES sur les CTCLs et le MF (respectivement 63, 102 et 107,2), bien que nos valeurs soient légèrement supérieures^{89,92,96}. Les détails des variants somatiques (SNVs et InDels) éliminés à chaque étape sont présentés dans la Figure 19.

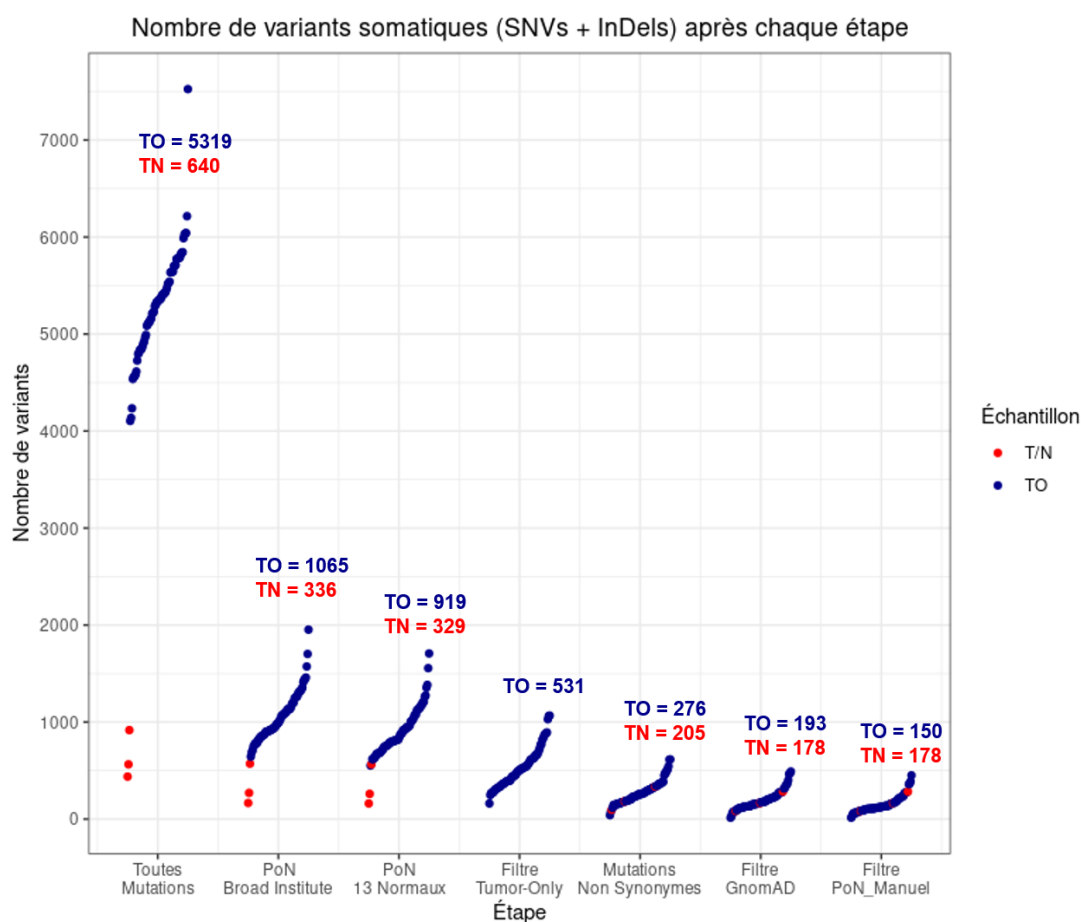


Figure 19 : Nombre de mutations somatiques (SNVs et InDels) par échantillon après chaque étape du pipeline. Les moyennes d'événements présents dans les cohortes TO (67 échantillons) et TN (3 échantillons) sont respectivement représentées en bleues et en rouges

au-dessus de chaque étape. Les trois paires TN proviennent de deux patients dont l'un (Patient n°7) a fourni deux échantillons tumoraux qui ont été comparés avec le même échantillon normal. Les mutations issues du pipeline (« Toutes Mutations ») ont été filtrées à partir du panel d'échantillons normaux (PoN, pour *Panel of Normal*) du Broad Institute (« PoN Broad Institute ») et celui de treize échantillons cutanés et sanguins qui ont été séquencés en même temps que nos échantillons tumoraux (« PoN 13 Normaux »). Un filtre spécifique du logiciel ABSOLUTE n'a été utilisé que sur les échantillons en TO (« Filtre *Tumor-Only* ») et nous avons mis en place deux filtres supplémentaires qui s'appuient sur la fréquence allélique (AF) des mutations grâce à la base de données GnomAD (« Filtre GnomAD ») et sur la recherche manuelle de mutations germinales dans les fichiers FASTQ (« Filtre_PoN_Manual »). Les mutations non-synonymes conservées sont les suivantes : *missense*, *nonsense*, *nonstop*, délétion et insertion simples, *frameshift* délétion et insertion, *In frame* délétion et insertion et *Splice site*.

L'étape « Toutes Mutations » correspond à la sortie brute des logiciels MuTect, Strelka et Mutect2, incluant l'ensemble des SNVs et des InDels détectés dans les cohortes TO (en bleu, Figure 19) et TN (en rouge, Figure 19). La différence marquée entre les deux jeux de données TO et TN été attendue, car l'échantillon Normal apparié dans le mode TN a permis d'éliminer de nombreux variants constitutionnels spécifiques à chaque individu, qui ne sont pas nécessairement repérés par les BdDs de variants constitutionnels intégrées aux logiciels détecteurs.

Ensuite, le logiciel MAFPoNFilter a été utilisé pour filtrer les artefacts et les variants constitutionnels présents dans les deux PoNs : « PoN *Broad Institute* », et « PoN 13 Normaux ». Le PoN du BI, bien qu'établi à partir d'échantillons WES assez anciens, générés avec des protocoles standards de la plateforme génomique du BI, reste une référence dans les analyses de routine pour retirer de nombreux artefacts, notamment ceux associés à des erreurs d'alignement. En complément, l'ajout du PoN de treize échantillons normaux, séquencés en parallèle de nos échantillons tumoraux, nous a permis de supprimer des artefacts techniques propres aux conditions de préparation en paillasse et de séquençage actuelles.

Un effet batch peut exister, car les protocoles de séquençage et de préparation des WES du PoN du BI ne sont pas entièrement connus et ont probablement évolué entre 2014 (année estimée de la création du PoN du BI) et 2020 (année de notre séquençage), ce qui peut expliquer la détection de certains FPs. Toutefois, l'utilisation d'un PoN spécifique à notre cohorte nous a été fortement recommandé, par l'équipe du Pr. Getz, pour augmenter la précision du filtrage et supprimer les FPs propres à notre procédure expérimentale.

Le filtrage par les deux PoNs a entraîné une réduction attendue du nombre d'événements observés dans les deux groupes d'échantillons : dans les échantillons TO, le nombre événements est passé de 5319 à 919 (- 83%), tandis que dans les échantillons TN, ce nombre est passé de 640 à 329 (- 48,6%). Bien que cette diminution soit attendue après un filtrage rigoureux, il peut être surprenant que près de la moitié des événements soient filtrés en TN, alors que la présence d'un échantillon normal apparié devrait déjà éliminer la majorité des variants constitutionnels. En fait, le rôle d'un PoN est un peu différent selon le mode d'analyse. En mode TO, le PoN est essentiel pour corriger l'absence de normal apparié et de filtrer les variants constitutionnels. En mode TN, le PoN sert principalement à affiner l'analyse en retirant des variants constitutionnels rares, des sites polymorphiques connus dans la population, des artéfacts techniques liés au séquençage et à l'alignement et des mutations faiblement supportées.

Le « Filtre *Tumor-Only* » du logiciel ABSOLUTE, appliqué spécifiquement à nos échantillons TO, a permis d'exclure les variants susceptibles d'être d'origine germinale selon les critères de discrimination du filtre. Cependant, l'utilisation de la macrodissection, qui vise à augmenter la pureté tumorales des échantillons, a limité l'efficacité de ce filtre. En effet, le filtre *Tumor-Only* fonctionne mieux lorsque la pureté

tumorale est modérée, car il peut alors modéliser une distribution somatique et une distribution germinale afin de calculer la probabilité qu'un variant appartienne à l'une ou à l'autre. Dans notre cas, l'augmentation de la pureté tumorale a réduit la proportion d'ADN normal dans les échantillons, ce qui a rendu difficile de distinguer la distribution germinale de la tumorale par le filtre. Ce fait a eu pour conséquence qu'il y a eu moins de variants qui ont pu être filtrés qu'initialement prévu.

Pour prioriser notre analyse, nous avons fait le choix de nous concentrer uniquement sur les « Mutations Non-Synonymes », éliminant ainsi les variants synonymes qui ne modifient pas la séquence protéique et qui ont moins d'impact du point de vue fonctionnel. Cette étape a réduit près de moitié le nombre moyen de variants étudiés.

Un filtre supplémentaire, « Filtre gnomAD », a ensuite été appliqué pour retirer des mutations constitutionnelles rares sur l'ensemble des échantillons de la cohorte. Enfin, le « Filtre *PON_Manual* » a été utilisé exclusivement sur les échantillons TO afin d'éliminer des mutations supplémentaires identifiées dans les fichiers FASTQ de nos échantillons normaux.

3.6.1.2 Charge Tumorale Mutationnelle

La TMB représente le nombre de mutations non synonymes (SNVs et InDels) par Mb et peut être utilisée comme biomarqueur dans de nombreux cancers¹⁶⁷. Dans une étude qui compare différents types de cancers, la TMB est considérée comme basse entre 1 et 5 mutations/Mb, intermédiaire entre 6 et 19, et élevée au-delà de 20 mutations/Mb¹⁶⁸. Après les étapes de filtrage des variants, la TMB a été estimée en faisant le rapport du nombre de SNVs/InDels détectés par échantillon à la taille de la région exomique ciblée lors du séquençage. La TMB médiane était de 3,52 (intervalle : 0,39 - 12,60) pour les échantillons TO, et de 4,64 (intervalle : 2,32 - 7,93) pour les

échantillons TN. Bien que cette médiane n'ait été calculée que pour trois paires d'échantillons TN, les valeurs de TMB entre les deux cohortes restent globalement proches. Cependant, la légère différence observée suggèrent que nos filtres TO ont éliminé en moyenne 1,12 mutations somatiques/Mb.

L'analyse statistique montre une faible corrélation positive, mais statistiquement significative entre l'âge au moment du diagnostic et la TMB par échantillon (test de Spearman, $\rho = 0,26$; $p\text{-value} = 0,03$). De plus, nous avons observé une TMB significativement plus élevée dans les échantillons HR, avec une médiane de 3,95 mutations/Mb, contre 3,02 mutations par Mb dans les échantillons LR (test de la somme des rangs de Wilcoxon, $W = 737,5$; $p\text{-value} = 0,02$) (Figure 20A).

Certains cancers cutanés, comme le mélanome malin, présentent des TMB différentes selon le sexe, avec une TMB est plus élevée chez les hommes¹⁶⁹. Cependant, en ce qui concerne notre étude, aucune différence significative de TMB n'a été observée entre les sexes (Figure 20B).

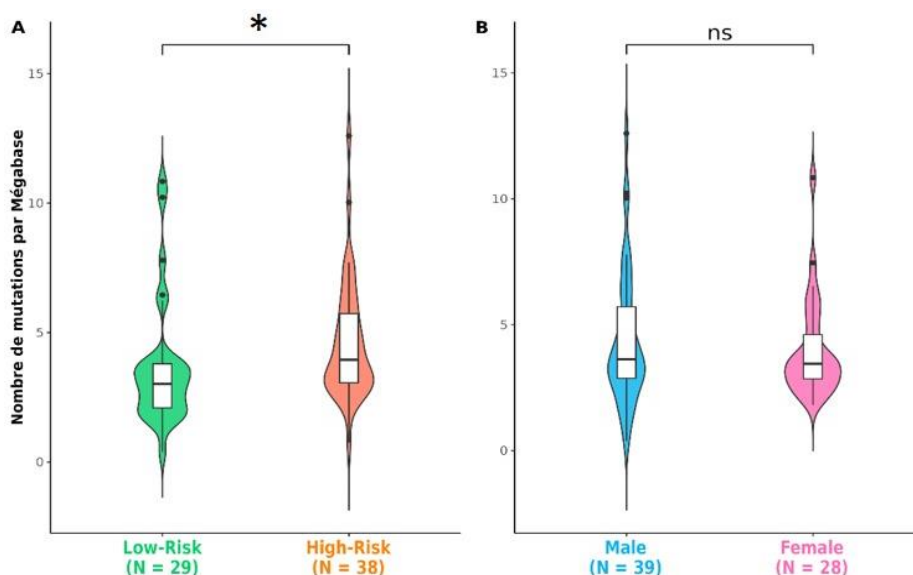


Figure 20 : Corrélation entre la TMB et le stade du MF et le genre du patient.

Diagrammes en violon de la TMB en fonction du stade (A ; Faible Risque, à gauche et Haut Risque, à droite) et du genre (B ; Homme, à gauche et Femme, à droite). Les distributions ont été comparées avec le test de la somme des rangs de Wilcoxon.

3.6.1.3 Gènes Fréquemment Mutés

En tout, treize gènes fréquemment mutés et impliqués dans le cancer ont été identifiés à l'aide du logiciel MutSig2CV. Cependant, il n'était pas possible d'utiliser directement MutSig2CV sur notre cohorte de 67 échantillons TO, en raison du biais lié aux échantillons séquentiels provenant d'un même patients, risquant de surestimer la fréquence de certaines mutations. Pour contrer ce problème, nous avons décidé de diviser les analyses en deux cohortes distinctes : une première cohorte regroupant uniquement les échantillons prélevés au moment du diagnostic (N = 48), et une seconde composée des échantillons tumoraux les plus avancés par patient (N = 48). Cette approche nous a permis d'identifier une liste complète des gènes significativement mutés dans chacune des deux cohortes. Enfin, les résultats des deux analyses ont été fusionnés en une liste unique de gènes, après suppression des doublons.

Les treize gènes mutés identifiés incluent des *drivers* mutationnels déjà signalés dans les CTCLs^{17,70,89,92,94,97}, ainsi que ceux répertoriés dans la plateforme pan-cancer « TumorPortal », qui regroupe des gènes mutationnels significatifs identifiés dans diverses tumeurs¹⁴⁹. Environ 51% des patients présentaient une mutation dans au moins un de ces *drivers* (Figure 21).

Parmi ces treize gènes se trouvent dix gènes *drivers* connus :

- *JUNB* (p.A282V ; Figure 22A), un TF de la différenciation des lymphocytes T, muté dans 13% des cas^{17,104,116,170}.
- *TET2* (Figure 22B), un facteur épigénétique muté dans 9% des cas^{91,97,171,172}.
- *MAPK1* (Figure 22C), un composant de la voie de signalisation des MAP kinases, muté dans 6% des cas.

- *FOXA1* (Figure 22D), un TF impliqué dans la signalisation de l'IFN et la suppression de la réponse immunitaire, muté dans 4% des cas^{172,173}.
- *FLT4* (Figure 22E), un récepteur tyrosine kinase.
- *PLCG1* (Figure 22F), une phospholipase impliquée dans la signalisation NF- κ B et NFAT^{90,174}.
- Les facteurs de la voie de signalisation JAK/STAT : *STAT3*, *STAT5A* et *STAT5B* (Figure 22G-I).
- *TP53* (Figure 22H).

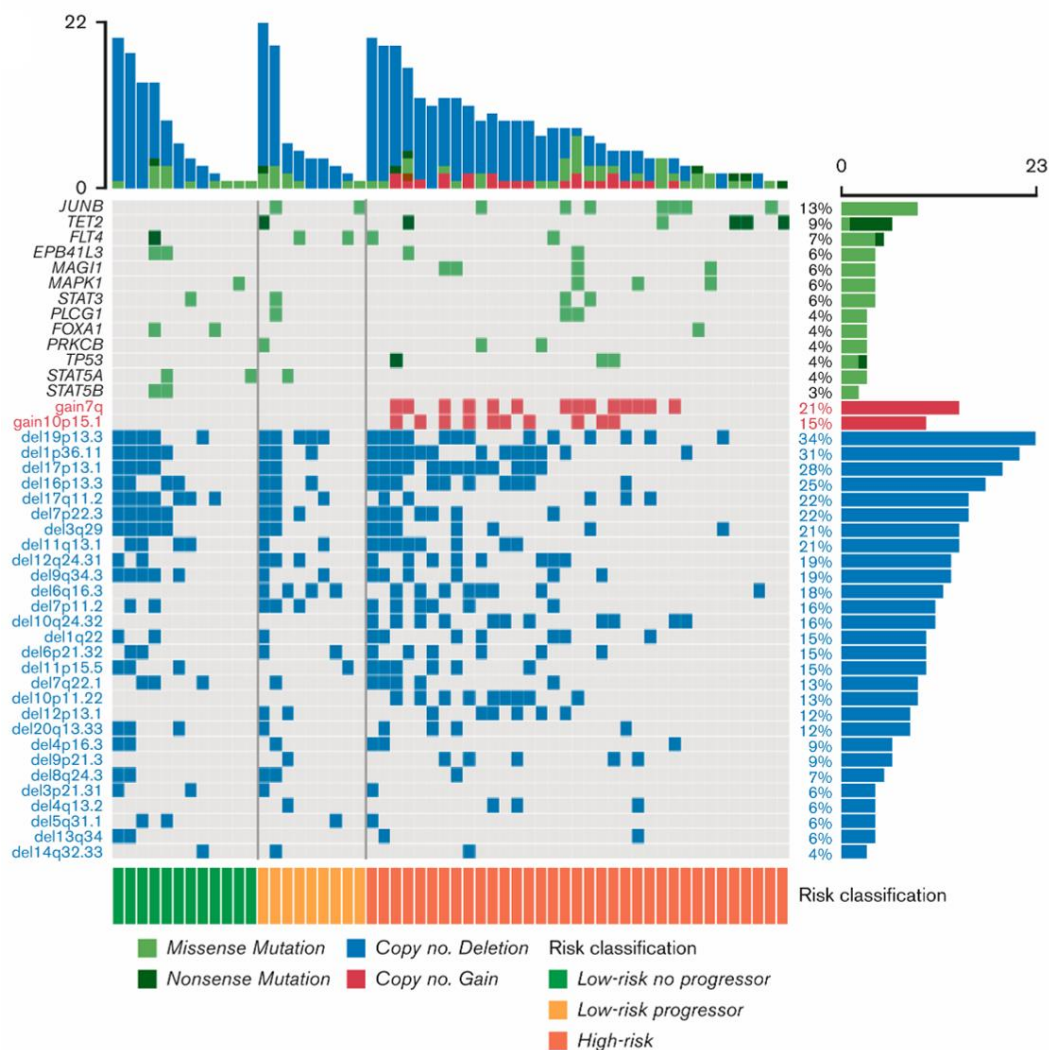


Figure 21 : Paysage des altérations génomiques les plus récurrentes dans le MF.

Oncoplot représentant le paysage des altérations génomiques dans les 67 échantillons tumoraux du MF divisés en stades LR et HR sur la base de la classification TNMB. Le stade « LR progressor » correspond aux échantillons provenant de patients LR au diagnostic qui

ont par la suite progressés en HR. Les gènes fréquemment mutés ont été sélectionnés par le logiciel MutSig2CV au moment du diagnostic et à un stade avancé et les SCNVs ont été détectés par le logiciel GISTIC2.0. La figure a été générée grâce à l'outil maftools sur R¹⁷⁵.

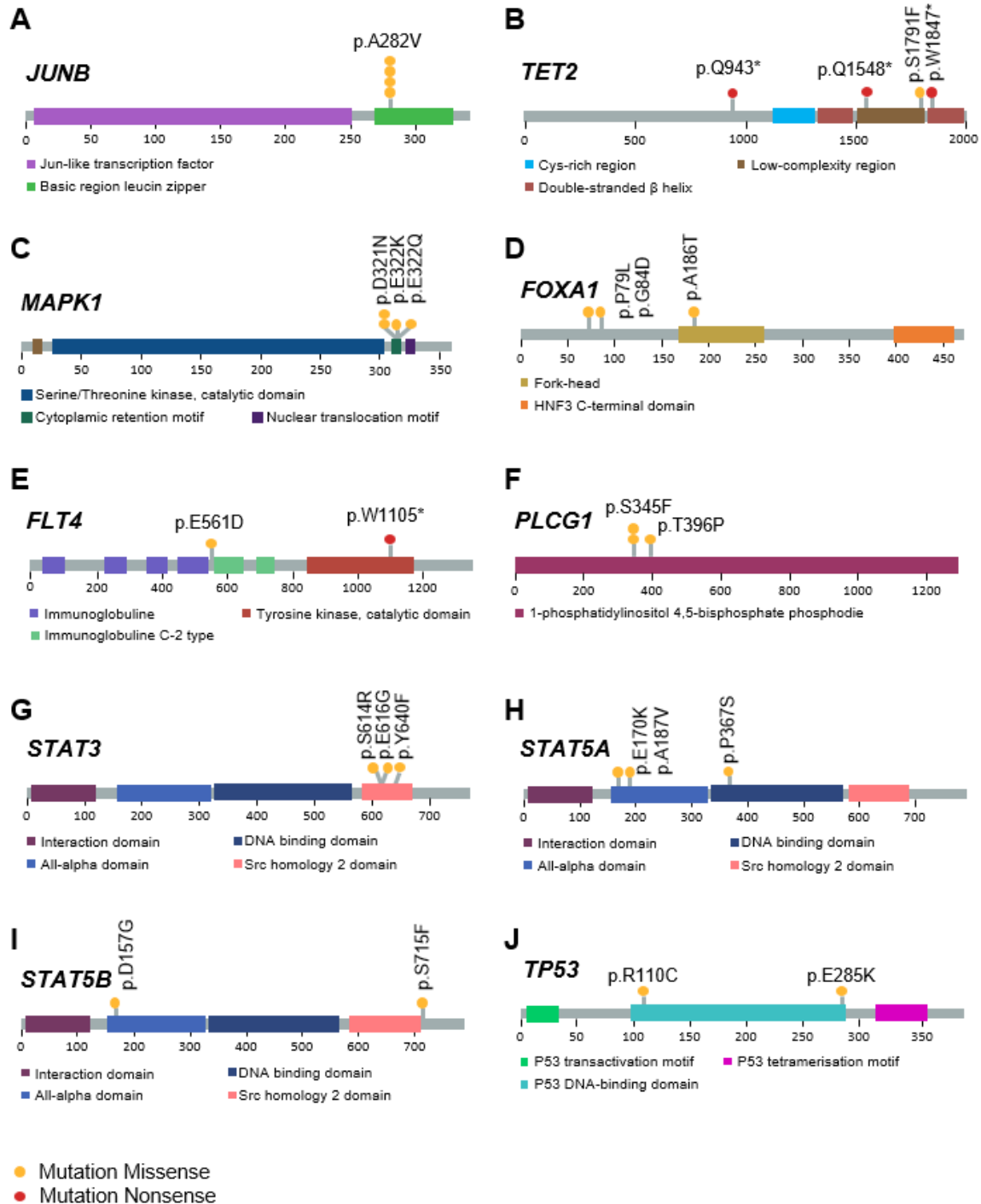


Figure 22 : Lollipop plots représentant toutes les mutations non-synonymes pour dix gènes significativement mutés. Représentation schématique des gènes *JUNB*, *TET2*, *MAPK1*, *FOXA1*, *FLT4*, *PLCG1*, *STAT3*, *STAT5A*, *STAT5B* et *TP53* avec leurs domaines

fonctionnels et les coordonnées des mutations *missenses* et *nonsenses* détectées. Les domaines ont été annotés grâce à l'outil SMART avec la BdD Pfam¹⁷⁶.

En particulier, *JUNB*, un TF de la famille des gènes JUN, est l'un des composants clés du complexe de transcription de la protéine d'activation-1 (AP-1 pour *Activating Protein-1*)^{17,170}. Le gène *JUNB* est fréquemment retrouvé dans les études concernant le MF et d'autres LCPs et participe très probablement dans la pathogenèse de ces maladies^{103,104,116}. La mutation p.A282V a d'ailleurs été rapportée dans le MF dans deux études récentes^{17,170}.

Le gène *TET2* (p.Q943*, p.S1791F, p.Q1548*, p.W1847*) joue un rôle dans l'oxydation de la méthylcytosine dans l'ADN. Les oxydations répétées, accompagnées de réparation de l'ADN, conduisent à la production de 5-hydroxyméthylcytosine ce qui provoque la déméthylation passive de l'ADN. L'inactivation de *TET2* est impliquée dans l'initiation des cancers hématologiques¹⁷¹. De nombreuses mutations de *TET2*, notamment des mutations *nonsenses*, ont été identifiées dans les CTCLs⁹¹, ainsi que dans les lymphocytes T tumoraux du derme chez MF avec huit autres gènes^{97,172}.

Le gène *MAPK1* (p.D321N, p.E322K, p.E322Q) code pour une protéine de la famille ERK (*Extracellular-signal-Regulated Kinase*) et joue un rôle dans la prolifération et la différenciation cellulaire. La mutation E322K, précédemment notée par da Silva Almeida *et al.* en 2015, conduit à une activation accrue des kinases ERK1/2⁹¹.

Parmi les treize gènes détectés par MutSig2CV, il y a trois autres gènes fréquemment mutés : *EPB41L3*, *MAGI1* et *PRKCB*. Les gènes *EPB41L3* et *MAGI1* sont connus comme ayant un rôle dans le cancer et ont été relevés dans une analyse WGS/WES sur cinq patients atteints d'un MF⁹². Quant au gène *PRKCB*, il participe à de nombreuses voies de signalisation, notamment NFkB, MAPK et celle du calcium. Une étude a retrouvé que ce gène était muté exclusivement dans la couche de l'épiderme du MF avec trente-six autres gènes *drivers*⁹⁷ (Figure 23).

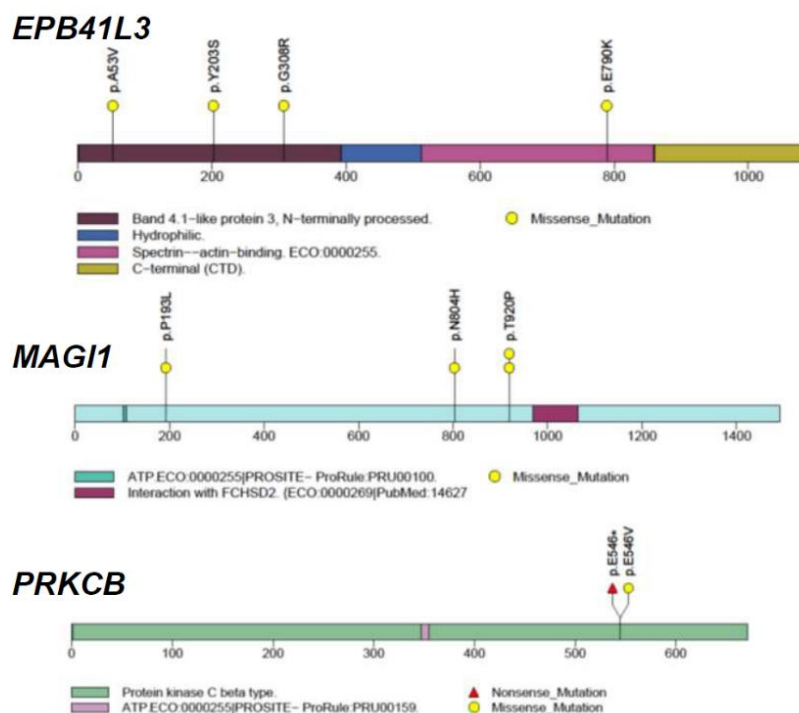


Figure 23 : Diagramme de mutation pour les gènes *EPB41L3*, *MAGI1* et *PRKCB* pour toutes les mutations non-synonymes mutées significativement. Représentation schématique des gènes *EPB41L3*, *MAGI1* et *PRKCB* avec leurs domaines fonctionnels et les coordonnées des mutations *missenses* et *nonsenses* détectées.

3.6.2 Variations du Nombre de Copies Somatiques Récurrentes

3.6.2.1 Identification des Altérations Significativement Récurrentes

Nous avons identifié les SCNVs récurrentes significativement à l'aide du logiciel GISTIC2.0. Comme pour les gènes fréquemment mutés, l'analyse a été menée sur la cohorte d'échantillons au diagnostic par patient (48 échantillons). En effet, compter un échantillon par patient évite une surestimation de la fréquence des SCNVs présents dans les échantillons séquentiels. Globalement, les SCNVs étaient les altérations génomiques les plus fréquentes et étaient présentes dans 84% des cas (Figure 21). Plus précisément, nous avons détecté des altérations significativement récurrentes, notamment 8 pertes de CN larges à l'échelle d'un bras chromosomique et 29 pertes

de CN focales, ainsi que 2 gains de CN larges et 1 gain de CN focal (valeur $q \leq 0,1$; Figure 24A et B).

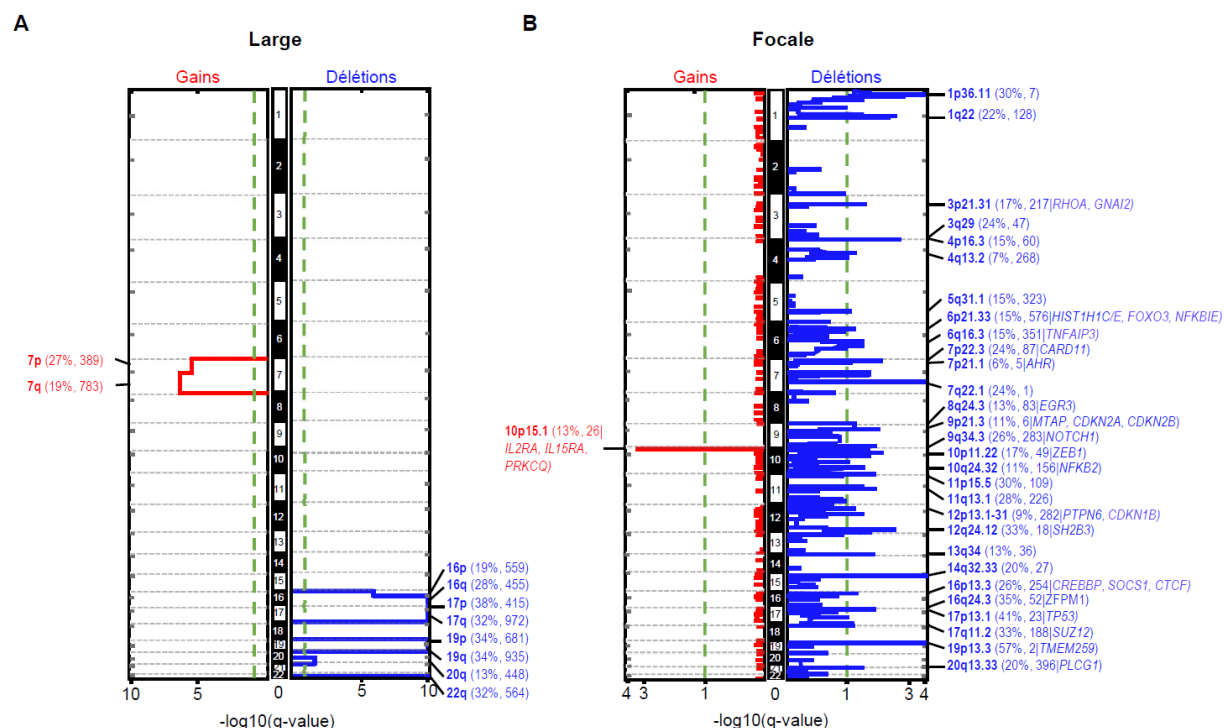


Figure 24 : Variations somatiques du nombre de copies récurrentes dans le MF.

Les SCNVs récurrentes larges/à l'échelle d'un bras chromosomique (A) ou focales (B) sont visualisés sous forme de diagrammes de type GISTIC en miroir. Les gains de copies sont représentés en rouge (partie gauche des diagrammes) et des pertes de copies en bleu (partie droite). Les chromosomes sont indiqués sur l'axe vertical. La ligne verte indique une q-value de 0,1. Les SCNVs sont étiquetés avec leur bande chromosomique (cytoband)/bras associé, suivi entre parenthèses par la fréquence de l'altération et le nombre de gènes totaux. Pour les événements focaux, les gènes *drivers* importants identifiés par le logiciel GISTIC2.0 sont étiquetés.

3.6.2.2 Gènes *Drivers* Identifiés dans les SCNVs Focaux

Les fréquences de ces SCNVs varient de 6% à 57%, et le nombre de gènes dans les pics focaux allait de 1 (MUC12 en del7q22.1) à 576 (del6p21.33). Dans le CN focal gain10p15.1 (26 gènes ; fréquence de 13%) se trouvent les récepteurs *alpha* de l'interleukine-2 et les récepteurs *alpha* de l'interleukine-15 (*IL2RA* et *IL15RA*) ainsi que la protéine kinase *PRKCQ* de la voie NF- κ B. En particulier, *IL2RA* et *IL15RA* jouent un

rôle crucial dans la phosphorylation des gènes *STAT3* et *STAT5* dans la voie JAK/STAT, contribuant probablement à l'oncogenèse du MF¹⁷⁷⁻¹⁸⁰.

Parmi les gènes suppresseurs de tumeurs affectés par les pertes de CN les plus fréquentes, figurent *TMEM259* (19p13.3, fréquence 57%), *TP53* (17p13.1, 41%), *SUZ12* et *NF1* (17q11.2, 33%), *NOTCH1* (9q34.3, 26%), *CARD11* (7p22.3, 24%), *ZEB1* (10p11.22, 17%), *TNFAIP3* (6q16.3, 15%), *CDKN2A*, *CDKN2B* et *MTAP* (9p21.3, 11%) et *CDKN1B* (12p12.2, 9%).

Le gène *TMEM259* code pour la membraline, une protéine impliquée dans le contrôle de la qualité structurale des protéines mal repliées et la dégradation associée au réticulum endoplasmique¹⁸¹.

Le gène suppresseur de tumeur *TP53* est fréquemment muté dans la plupart des cancers chez l'humain¹⁸². Cependant, il a été noté que le ratio entre les délétions et les mutations de *TP53* est significativement plus élevé pour le MF/SS que dans les autres cancers⁸⁹. Les mutations somatiques de *TP53* sont plus présentes dans les stades tumoraux du MF que dans les stades précoces, même si elles sont moins répandues que dans le SS⁹². Salgado *et al.* ont fait en 2010 une étude bilan de 41 biopsies cutanées de patients atteints de MF provenant de différentes publications. Au stade tumoral, 27,5% des patients présentent une perte de 17p13.1 et la délétion de *TP53* a été suggérée comme un *driver* du MF¹⁰⁶.

Le gène *ZEB1* est un répresseur de transcription qui inhibe l'expression du gène *IL-2* et participe au remodelage du cytosquelette¹²⁹.

Le gène *TNFAIP3* (*Tumor Necrosis Factor- α Induced Protein 3*) inhibe la voie NF κ B¹¹⁷. Des altérations SNVs et SCNVs impactant *TNFAIP3* ont été détectées dans une cohorte de 49 échantillons WES de MF obtenus à partir de biopsies cutanées de

macules et de papules, suggérant son rôle potentiel en tant que mutation *driver* dans le MF¹⁷².

Les gènes *CDKN2A*, *CDKN2B* et *MTAP* sont habituellement co-délétés dans d'autres cancers et parfois dans le MF⁷⁰. Les deux gènes *CDKN2A* et *CDKN2B* sont des suppresseurs de tumeur qui codent pour les protéines p16-INK4a et p15 qui jouent un rôle essentiel dans la différenciation, la progression du cycle cellulaire, la sénescence et l'apoptose^{183,184}. Il a été montré que la perte de *CDKN2A/CDKN2B* dans le MF est associée avec un mauvais pronostic^{105,114}. Quant au gène *MTAP*, il est un élément majeur dans le métabolisme des polyamines et encode pour la méthylthioadénosine phosphorylase, une enzyme intervenant dans le métabolisme des purines¹⁸⁵.

Le gène *SUZ12* code pour l'une des protéines du complexe polycomb répressif 2 (PRC2) qui, lorsqu'il est dérégulé, induit un phénotype tumoral. Ce gène joue un rôle de répression génique par le biais de modifications épigénétiques des histones, favorisant le maintien des capacités de renouvellement et de dédifférenciation cellulaire¹⁸⁶.

3.6.2.3 État des connaissances des Bandes Cytogénétiques dans le MF

Nous notons que la délétion de la région 1p36.11 qui est très significative (7 gènes, $q\text{-value} = 6,489.10^{-6}$), bien qu'aucun des gènes de cette région ne soit considéré comme *driver*. Cependant, cette bande cytogénétique (cytoband) se situe à proximité du gène suppresseur de tumeur et modificateur de chromatine *ARID1A*, reconnu pour son rôle dans plusieurs cancers et rapporté comme fréquemment altéré dans d'autres études SCNVs sur le MF^{17,89,95,108}. Les coordonnées génomiques de la cytoband identifiée par *GISTIC2.0* sont très resserrées (chr1: 25,347,408 -25,766,839)

par rapport aux délétions typiquement relevées dans la littérature (qui incluent *ARID1A*)^{89,95}. Par contre, le gène *ARID1A* est situé un peu plus loin sur le chromosome (chr1: 27,022,022 -27,108,601). Cette observation pourrait suggérer que notre étude récente possède une sensibilité plus précise dans la détection des SCNVs. Nous pouvons avoir détecté une région régulatrice potentielle à proximité du gène *ARID1A* qui pourrait influencer son expression ou sa fonction dans le MF.

La délétion 3p21.31 (17%, q-value = 0,070), n'a pas été rapportée précédemment dans le MF ou d'autres CTCLs, mais elle a déjà été identifiée dans d'autres cancers comme le carcinome épidermoïde de la tête et du cou¹⁸⁷ ou le cancer du poumon à petites cellules¹⁸⁸. La délétion de cette région est suggérée comme ayant un impact sur le développement de lésions dysplasiques précoces. Le gène *driver* *RHOA*, qui régule le remodelage du cytosquelette et code pour une petite GTPase, est généralement muté et non délété dans le MF ou le SS^{89,96,99,102}. Le gène *GNAI2* encode pour une sous-unité Gai2 d'une protéine G inhibitrice et l'orientation des lymphocytes B, et n'est pas habituellement délété dans le MF. Il a toutefois été trouvé fréquemment muté dans un autre lymphome T, celui du cancer intestinal épithéliotrope¹⁸⁹, ainsi que dans le lymphome diffus à grande cellules B¹⁹⁰.

La délétion 6p21.33 (15%, p-value = 0,070) a été identifiée dans la publication de Park *et al.*, en 2021, qui se concentrait sur des échantillons MF/SS. Le TF *FOXO3* avait été proposé comme le gène *driver* de la délétion pour son rôle de suppresseur de tumeur. En effet, des études sur le SS ont proposé que lorsque *FOXO3* est dérégulé (notamment par le gène *PTEN* souvent muté dans le MF), il contribue au développement de la tumeur¹¹⁵.

Les gènes *HIST1H1C* et *HIST1H1E* codent pour des histones de liaison H1, qui jouent un rôle clé dans la compaction et l'organisation de la chromatine. Ces gènes sont fréquemment mutés ou délétés dans certains lymphomes, notamment les lymphomes B. Leur altération perturbent l'architecture épigénétique, entraînant des remaniements à grande échelle de la chromatine qui modifient l'accessibilité aux TFs et favorisent l'expression aberrante de gènes oncogéniques. Des études sur le modèle murin ont montré que la perte des allèles *HIST1H1C* et *HIST1H1E* conduit une prolifération excessive des lymphocytes B et à la formation d'un lymphome plus agressif¹⁹¹.

Des délétions de l'inhibiteur de NF- κ B I *epsilon* (*NFKBIE*) ont déjà été observées dans les lymphomes B médiastinaux primitifs, mais elles sont plus rares dans les lymphomes T¹⁹².

Un gain de la cytoband 7p22.2 (24%, q-value = 0,016) qui contient le gène *driver* *CARD11*, a été rapporté dans l'étude de Park *et al.*, Blood 2021¹⁷. Généralement, *CARD11* est plus souvent retrouvé muté plutôt qu'amplifié dans les CTCLs, tels que le SS et le MF^{89–91,94,97,102}. Dans les lymphomes T, *CARD11* et *PRKCQ* participent activement dans la voie canonique NF- κ B qui joue un rôle dans la prolifération et la survie des cellules tumorales (Figure 36)¹²⁰.

La délétion 7p21.1 (6%, q-value = 0,487) est retrouvée dans Park *et al.* 2021 avec le gène participant à la différenciation des lymphocytes T, *AHR*, et un autre gène *driver* candidat, *NFE2L3*¹⁷. Ce dernier est également identifié dans la publication de Choi *et al.* en 2015 comme un gène putatif suppresseur de tumeur dans de nombreux cancers de la liste du TCGA⁸⁹.

La délétion 10q24.32 (11%, q-value = 0,038) inclut le gène *driver* candidat *NFKB2* qui intervient dans la voie de signalisation NFκB. Ce gène et cette délétion génomique ont été identifiés dans plusieurs publications sur le MF^{17,89} et serait également un gène significativement plus délété dans les CTCLs leucémiques (dont le MF leucémique et le SS) par rapport au MF classique¹⁷. De plus, *NFKB2* est proposé comme un intervenant majeur dans la transition d'un MF leucémique vers un SS. La délétion de la cytoband est aussi observée dans Wang *et al.* 2015¹⁰².

La perte du gène régulateur du cycle cellulaire *CDKN1B* sur la cytoband 12p13.1 (9%, q-value = 0,065) est particulièrement documentée dans le MF et le SS^{89,193}. La perte du gène suppresseur *PTPN6* a aussi été identifié et est associée à une perturbation de la voie JAK/STAT. En absence de *PTPN6*, *STAT3* subit une phosphorylation excessive, ce qui entraîne une activation anormale de la signalisation JAK/STAT¹⁹⁴. Cette dérégulation est associée à un mauvais pronostic dans les lymphomes T périphériques¹⁹⁴.

La délétion de la cytoband 12q24.12 (33%, q-value = 0,004) a été notée dans les données WGS et RNAseq publiées par Bastidas Torres *et al.* en 2022 avec le gène *driver* candidat *SH2B3*, un intervenant de la voie JAK/STAT et qui serait l'un des deux gènes (avec *JAK2*) qui déterminent le rare lymphome T cutané primitif CD8⁺ épidermotrope cytotoxique agressif¹¹¹. Bien que la délétion de ce gène n'ait jamais été décrite pour le MF, nous avons choisi de l'associer à la del12q24.12, car ce gène était tronqué dans un des huit échantillons de MF dans la publication de da Silva Almeida *et al.* en 2015⁹¹.

La délétion 16p13.13-3 (26%, q-value = 0,053) porte trois gènes *drivers* candidats. Les deux premiers sont les modificateurs épigénétiques de chromatine

CTCF^{17,89} et *CREBBP*. Le dernier est *SOCS1*, intervenant dans la voie JAK/STAT et l'un des événements les plus fréquents après la délétion de *CDKN2A*. Le gène *SOCS1* serait même l'un des événements essentiels dans le développement du MF⁹⁵.

La délétion de la cytoband 16q24.3 (35%, p-value = 0,017), où se situe le gène *driver* *ZFPM1*, impliqué dans la différenciation des lymphocytes T, a été observée par Park *et al.* 2021¹⁷. Le gène *CBFA2T3*⁹⁵, également repéré, est impliqué dans une autre hémopathie, la leucémie aiguë mégacaryoblastique¹⁹⁵.

3.6.3 Signatures Mutationnelles

La découverte des signatures mutationnelles repose sur le nombre de mutations somatiques détectées et leur contexte (les nucléotides environnant avant et après le site muté) pour ainsi définir un ensemble de signatures caractéristiques selon la méthode NMF. Ensuite, l'activité de l'ensemble de ces signatures est évaluée dans chaque échantillon de la cohorte grâce à un modèle bayésien. Les signatures SBS sont répertoriées dans la BdD COSMIC, qui associe diverses signatures SBS à des cancers spécifiques et propose des étiologies pour chacune d'entre elles.

La substitution la plus fréquente est la transition cytosine-thymine (C>T) qui représentait en moyenne 70% des mutations dans l'ensemble de la cohorte (Figure 25A). Le contexte de cette mutation est plus souvent dipyrimidique (CC>TT) et correspondrait à une mauvaise réparation des liaisons covalentes induite par les UVs entre les pyrimidines adjacentes^{122,196,197}.

L'analyse a détecté deux signatures mutationnelles principales dans COSMIC : la signature SBS7b, qui est l'une des quatre signatures associées aux dommages causés par l'exposition directe aux UVs (de SBS7a à SBS7d), et la signature SBS15, liée à la réparation défectueuse de l'ADN (Figure 25A et B). La signature SBS1, liée à

la désamination avec l'âge, a également été observée dans notre cohorte. Cependant, son faible signal ne permettait pas de la distinguer de la signature UV.

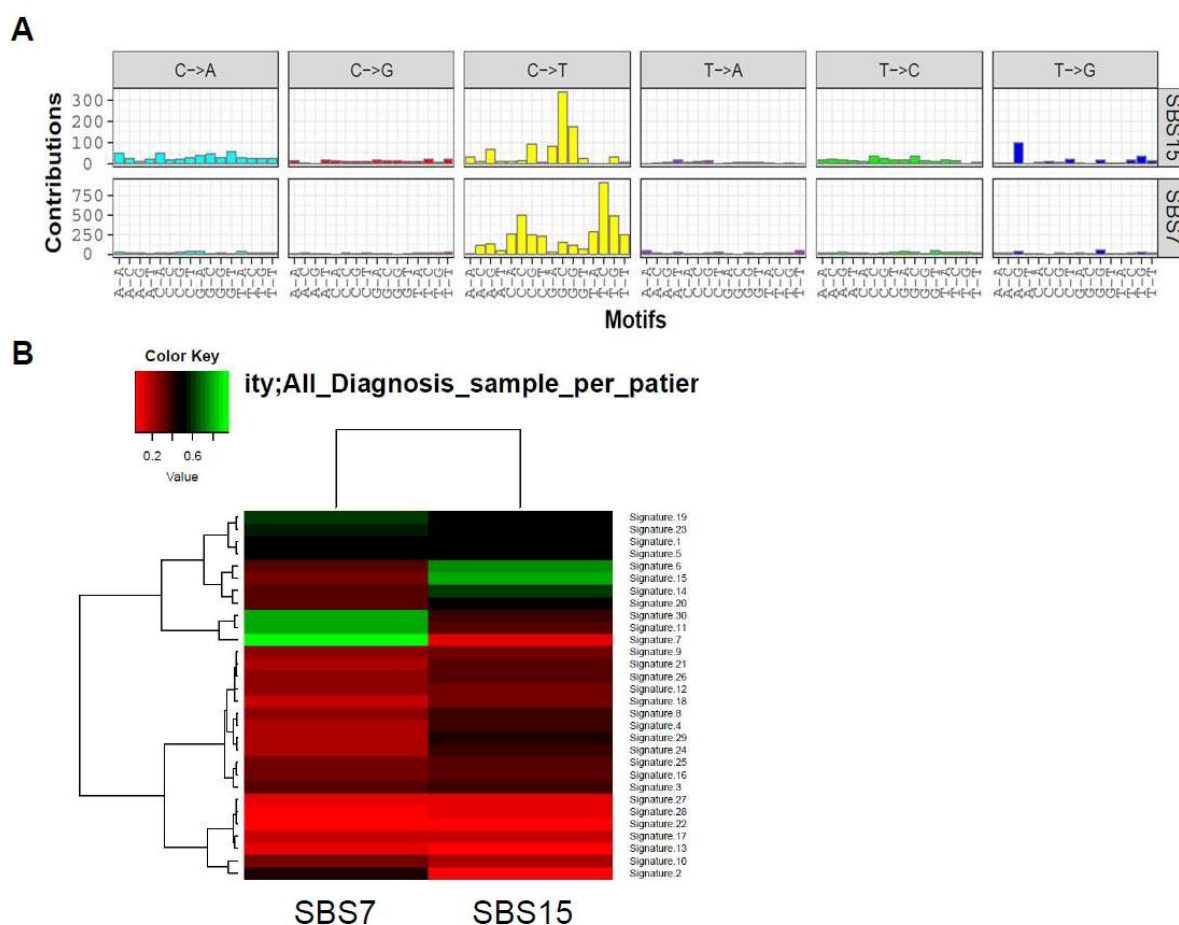


Figure 25 : Extraction des signatures mutationnelles. A. L'analyse des signatures mutationnelles pour 67 échantillons de MF a identifié deux signatures *Single Base Substitution* (SBS7b et SBS15) grâce aux regroupements des mutations quantifiées par la méthode de distance de la mutation la plus proche (*Nearest Mutation Distance*). **B.** Figure heatmap sur la similarité cosinus des deux signatures parmi les échantillons de MF.

L'enrichissement de la signature SBS7b est cohérent avec les études antérieures sur les CTCLs et est également significativement associé aux stades HR (Figure 26A)^{17,92,198,199}. De plus, l'augmentation de la TMB dans les échantillons HR est principalement associée à la signature UV (test de Wilcoxon ; $P = 5,735e-12$) (Figure 27A). Une première hypothèse pour expliquer la prédominance de SBS7b dans les HR serait que l'exposition directe aux UVs pourrait jouer un rôle dans la

progression du MF et contribuerait à l'accumulation de mutations dans les stades avancés (Figure 26A). Une autre hypothèse serait que les tumeurs HR, phylogénétiquement plus anciennes, auraient eu plus de temps pour accumuler des *drivers* et la signature UV (Figures 26B et 27A).

Toutefois, une étude comparative des signatures mutationnelles UVs entre le mélanome malin et le MF indique que les mutations liées à SBS7 jouent un rôle pathogène mineur dans le développement du MF. En effet, dans cette étude, les mutations attribuées aux signatures SBS7 ne représentent que 4% de la TMB du MF, contre plus de 90% dans le mélanome. Ces résultats suggèrent que les mutations *passengers* observées dans le MF seraient le résultat de l'exposition environnementale (soleil) ou thérapeutique (photothérapie) des lésions cutanées¹⁹⁹. Aussi, les données épidémiologiques montrent que le MF a tendance à se développer dans des régions cutanées protégées du soleil, ce qui renforce l'hypothèse selon laquelle le traitement par UV n'augmente pas le risque de développer ultérieurement un lymphome cutané¹⁹⁹.

La signature SBS15, l'une des sept signatures mutationnelles liées à la réparation défectueuse des mésappariements de l'ADN et à l'instabilité des microsatellites, est présente dans différents types de cancer, y compris le MF^{132,200}. La contribution de SBS15 est similaire dans les échantillons LR et HR, ce qui suggère que cette signature est liée à un événement plus fondateur et était déjà présente lors des premiers stades du développement tumoral (Figure 27B).

Ensuite, nous avons déterminé la contribution relative des signatures SBS7b et SBS15 pour chaque gène *driver* (Figure 26B). Cinq gènes étaient enrichis par les mutations associées à SBS15, à savoir *JUNB*, *STAT3*, *MAML2*, *FLT4* et *FOXA1*,

tandis que neuf gènes étaient principalement associés à SBS7b : *TET2*, *MAPK1*, *TP53*, *EPB41L3*, *MAGI1*, *PRKCB*, *HCK*, *L3MBTL4* et *MECOM*.

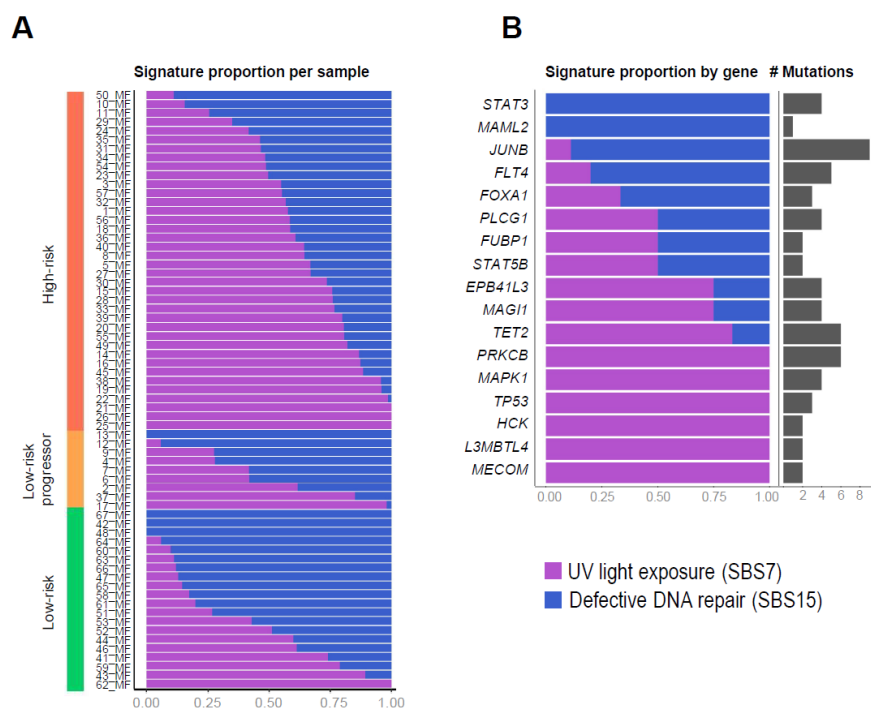


Figure 26 : Signatures mutationnelles identifiées dans les échantillons de MF.

A. Enrichissement relatif de l'activité des deux signatures *Single Base Substitution* (SBS) mutationnelles SBS7 (violet) et SBS15 (bleu) par échantillons, divisés en LR (vert), LR *progressors* qui ont progressés du stade LR vers HR (en orange) et HR (rouge). **B.**

Enrichissement relatif de l'activité des deux signatures mutationnelles par gène muté de manière récurrente. Le nombre de mutations par gène est indiqué à droite. Les gènes sont triés par importance de la signature SBS15, mécanismes de réparation de mésappariement de l'ADN.

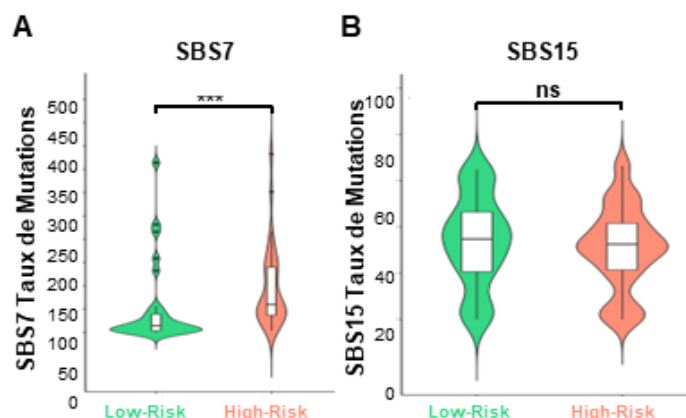


Figure 27 : Corrélation entre le taux de mutations des signatures mutationnelles et le stade du MF. Diagrammes en violon du taux de mutations dans les signatures mutationnelles dans les échantillons LR et HR pour SBS7b (A) et SBS15 (B). Les distributions ont été comparées avec le test de la somme des rangs de Wilcoxon.

3.6.4 Biomarqueurs Impactant la Survie Globale et la Progression

3.6.4.1 Survie Globale de la Cohorte

Nous avons tout d'abord comparé l'OS des patients atteints d'un MF type LR (stade IA à IIA) à celle des patients avec un HR (stades IIB à IVB). La durée médiane du suivi pour le groupe LR était de 4,8 années, tandis que celle pour le groupe HR est de 3,6 années. Comme attendu, les patients LR avaient une OS plus élevée que les patients HR, avec une mOS de 8,9 années, significativement supérieure à celle de 3,6 années des patients de HR (test du Log-Rank, $p = 0,0056$, Figure 28). Ces données confirment l'importance de distinguer les deux groupes de risque initialement établis.

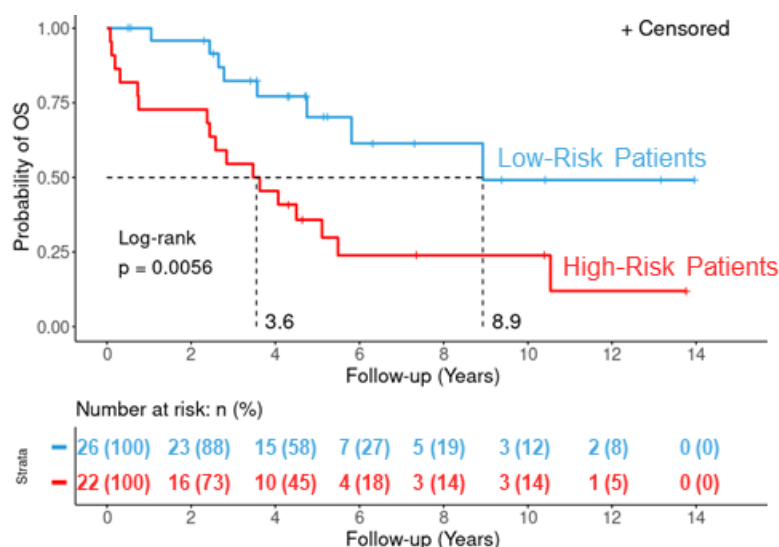


Figure 28 : Courbe Kaplan-Meier pour l'analyse de l'OS entre les patients LR et HR chez les 48 patients au diagnostic. Les critères d'évaluation ont été estimés à l'aide de la méthode de Kaplan-Meier. Les différences dans les courbes de survie ont été évaluées à l'aide du test du Log-rank. La médiane de suivi a été calculée à l'aide de la méthode inverse de Kaplan-Meier.

3.6.4.2 Association entre Biomarqueurs avec Stade et Pronostic

Nous avons ensuite comparé les profils génomiques des patients atteints d'une maladie de type LR à ceux des patients atteints d'une maladie de type HR. Les gains de copies gain7q (*odds ratio* [OR] = 0,07 ; p-value = 0,002) et gain10p15.1 (*IL2RA* et *IL15RA* ; OR = 0,11 ; p-value = 0,019) étaient associés aux patients HR, tout comme les délétions del10q24.32 (*NFKB2* ; OR = 0,09 ; p-value = 0,010) et del10p11.22 (*ZEB1* ; OR = 0,12 ; p-value = 0,038 ; Figure 29). À l'inverse, la délétion del17q11.2 (*SUZ12* et *NF1*), était significativement plus fréquente chez les patients LR (OR = 4,84 ; p-value = 0,014).

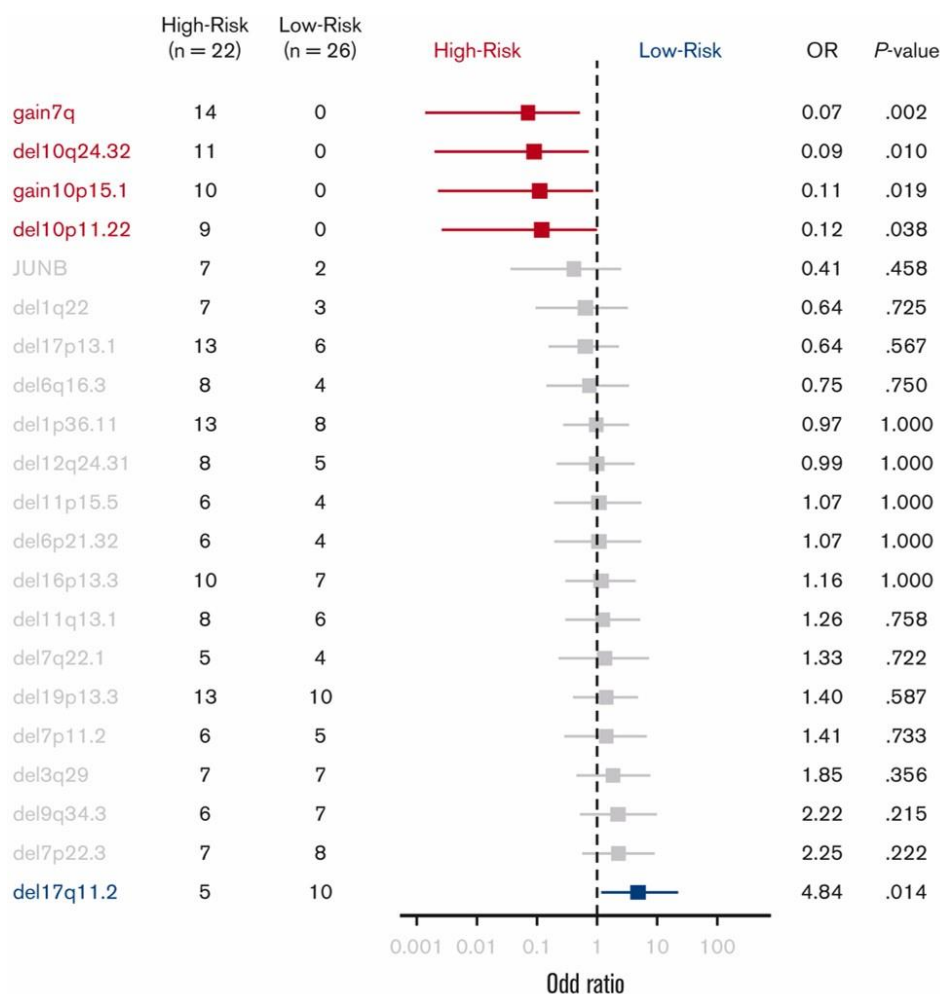


Figure 29 : Corrélation entre les événements génomiques et le stade du MF.

Forest plot montrant l'association entre l'altération de chaque gène et chaque SCNV et le stade clinique du MF selon leur *odd ratios* (OR). La figure a été générée grâce à la fonction « maf_compare » de l'outil Maftools sous R. La fonction maf_compare effectue un test de Fisher sur tous les gènes entre les cohortes LR et HR afin de détecter les événements mutés de manière différentielle. Dans le Forest plot généré, les événements ayant une faible fréquence inférieure à ~20% (c'est-à-dire pour neuf patients) n'ont pas été analysés.

Nous avons ensuite évalué l'impact des mutations *drivers* candidates détectées au diagnostic sur la OS des patients. La mOS de la cohorte était de 4,3 ans (intervalle, 0,3-14). Parmi les altérations détectées, quatre SCNVs ont été significativement associées à une OS plus courte (Figure 30A-D) : del10p11.22 (*ZEB1* ; mOS de 2,4 vs 8,9 années ; p-value = 0,00029) ; gain10p15.1 (*IL2RA* et *IL15RA* ; mOS de 2,4 vs 5,8 années ; p-value = 0,00029) ; gain7q (mOS de 2,6 vs 8,9 années ; p-value = 0,011) ; et del6q16.3 (*TNFAIP3* ; mOS de 1,9 vs 5,8 années ; p-value = 0,021).

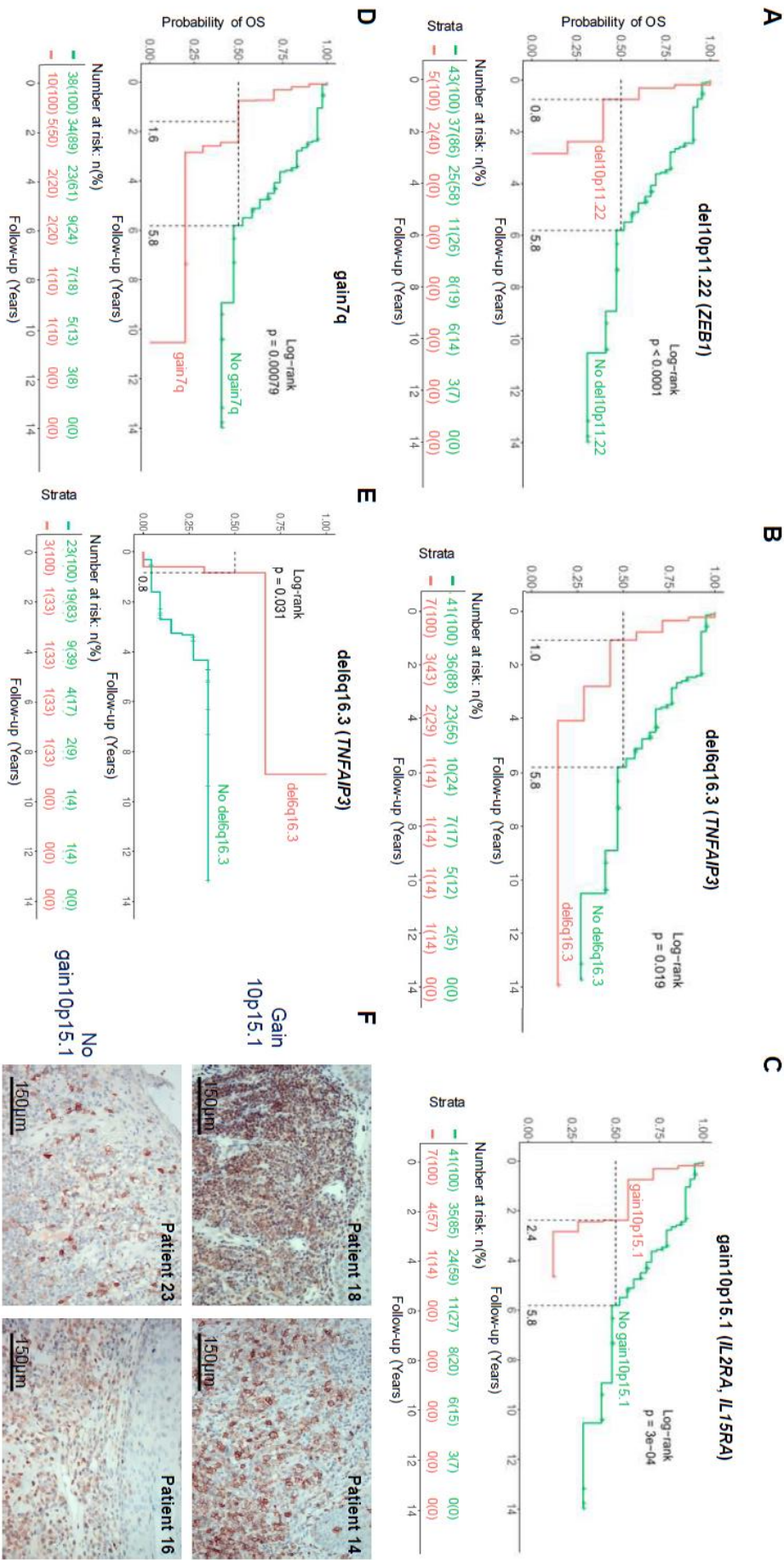
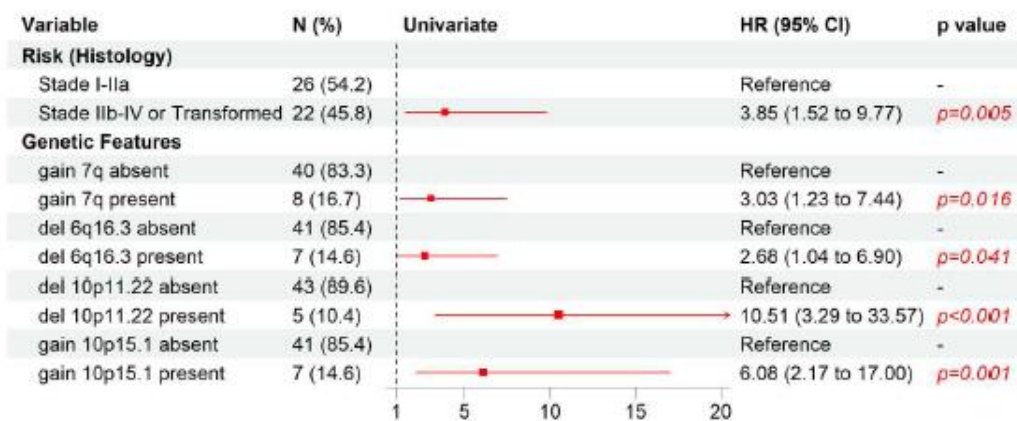


Figure 30 : Association du pronostic avec les différents facteurs génétiques dans le MF. Courbes Kaplan-Meier d'OS des patients ayant reçu un diagnostic récent du MF par rapport aux différents facteurs génétiques : **A.** del10p11.22 ; **B.** del6q16.3 ; **C.** gain10p15.1 ; **D.** gain7q ; **E.** Courbes de Kaplan-Meier pour l'analyse du temps jusqu'à la progression chez les patients de LR atteints de MF pour le facteur génétique del6q16.3. Le TTP a été mesuré à partir de la date du diagnostic jusqu'à la date de progression documentée vers la HR. Les p-values ont été obtenues par un test de Log-rank. **F.** IHC CD25 au diagnostic des biopsies cutanées MF chez les patients avec gain10p15.1 (panneaux supérieurs, échantillon MF pour les patients 18 et 14 avec présence de cellules tMF) ou sans gain10p15.1 (panneaux inférieurs, échantillon MF pour les patients 23 et 16). Les échelles indiquent 150 µm.

La présence des SCNVs focaux del10p11.22, gain10p15.1 et del6q16.3 est un facteur de risque significatif dans une analyse multivariée par étapes, en tenant compte du stade clinique du patient (Figure 31A et B). Il est intéressant de noter qu'aucun SNV n'a eu d'effet sur l'OS, peut-être trop rares par rapport à la taille de notre cohorte.

A



B

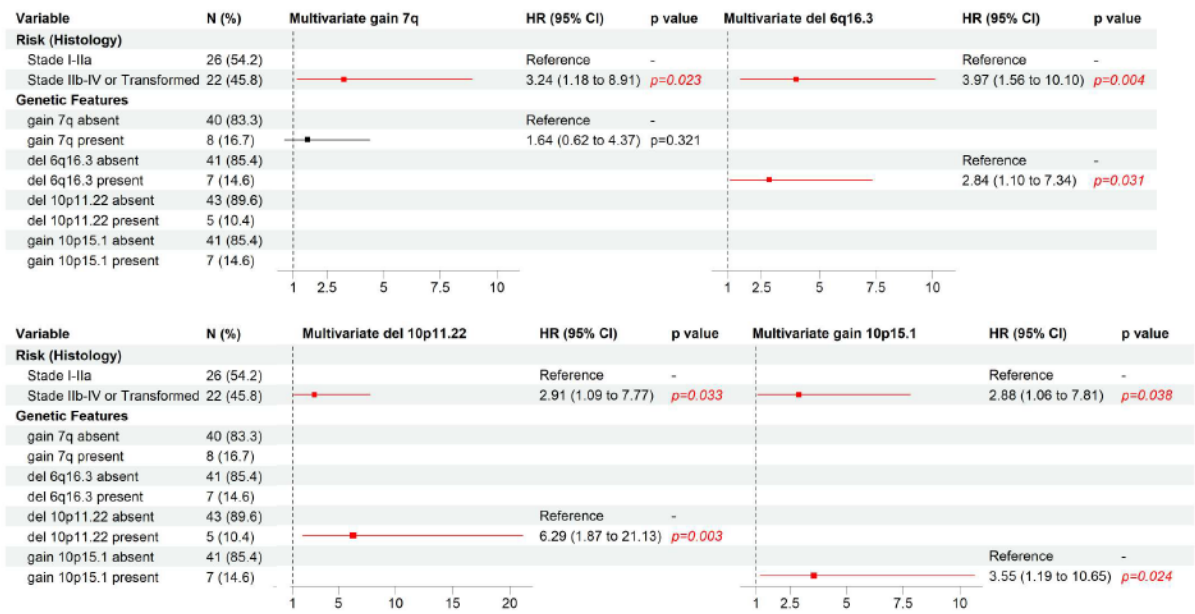


Figure 31: Analyses univariées et multivariées des biomarqueurs impactant le MF.

A. Analyses univariées et **B.** multivariées avec une régression pas à pas de Cox des événements génomiques avec l'utilisation d'*Hazard-ratio (HR)*. La modélisation des *HR* a été réalisée pour évaluer l'impact des altérations génétiques sur le risque de progression du MF.

CI pour Intervalle de confiance

Parmi les patients dans le groupe LR, sept ont progressé au cours du suivi, tandis que dix-neuf sont restés stables. Les patients LR présentant une del6q16.3 (*TNFAIP3*) avaient un TTP plus court (médiane = 0,8 an vs. NR ; p -value = 0,0031 ; Figure 30E). Cette altération est présente chez 15% des patients LR au diagnostic et pourrait représenter un marqueur pronostique pertinent dans la pratique clinique.

En ce qui concerne le gain10p15.1, nous avons observé une plus forte expression de CD25 par IHC dans les échantillons provenant de patients avec gain10p15.1 que dans ceux provenant de patients sans gain10p15.1. Ceci est cohérent avec un lien entre le gain d'*IL2RA* et l'expression de surface de CD25, suggérant un rôle de la signalisation *IL2/IL2RA* dans la progression de la maladie (Figure 30F).

3.6.5 L'Architecture Clonale et la Phylogénie du MF

Nous avons estimé la CCF pour chaque mutation *driver* candidate et déterminé si les altérations étaient clonales ($CCF \geq 0,9$) ou sous-clonales ($CCF < 0,9$). Nous avons observé une hétérogénéité dans la clonalité des altérations génomiques chez les patients atteints de MF. Certaines altérations étaient fréquemment clonales, telles que del9p21.3 (*CDKN2A*, *CDKN2B* et *MTAP*), del10p11.22 (*ZEB1*), gain7q, et des mutations dans *PRKCB* ou *STAT5B*. D'autres altérations étaient plus fréquemment sous-clonales, comme les mutations dans *TET2*, *TP53*, *PLCG1* ou del17p, qui représentent probablement des événements plus tardifs dans l'évolution de la maladie (Figure 32).

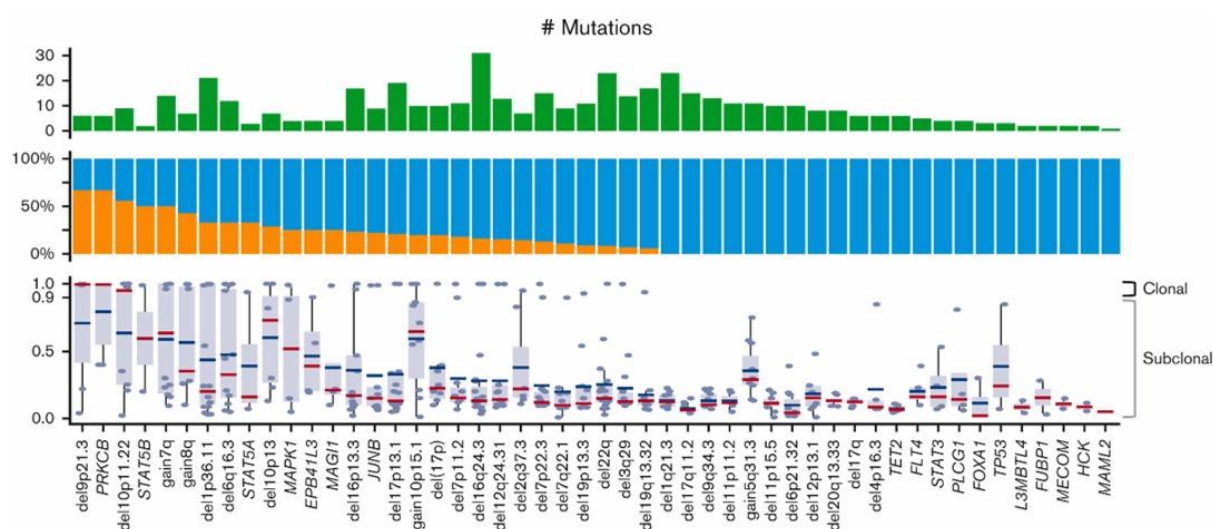


Figure 32 : Hétérogénéité clonale dans le MF. La partie supérieure représente la fréquence des événements *drivers* récurrents parmi les 67 échantillons (histogramme vert).

Le panneau du milieu indique leur proportion selon leur classification comme mutation clonales (orange) ou sous-clonales (bleu). En bas, la distribution des fraction de cellules cancéreuses (CCF) est visualisée pour chaque *driver*. Les valeurs moyennes (bleu foncés) et médianes (en rouge) des CCF sont indiquées. Les boîtes bleu clair représentent les interquartiles, tandis que les barres verticales noires indiquent l'étendue des valeurs extrêmes (hors valeurs aberrantes) pour chaque *driver*.

Nous avons ensuite analysé la cooccurrence des gènes *drivers* et des SCNVs. Nous avons constaté que la del9p21.3 (*CDKN2A*, *CDKN2B* et *MTAP*) coexiste

significativement avec le gain7q et la del10p11.22 (*ZEB1*). Le gain 10p15.1 (*IL2RA* et *IL15RA*) était significativement associé à la del6q16.3 (*TNFAIP3*) et à la mutation *TP53*. La del16p13.3 (*CREBBP*) était associée à la del17p et à la del19p13.3 (*TMEM259* ; Figure 33).

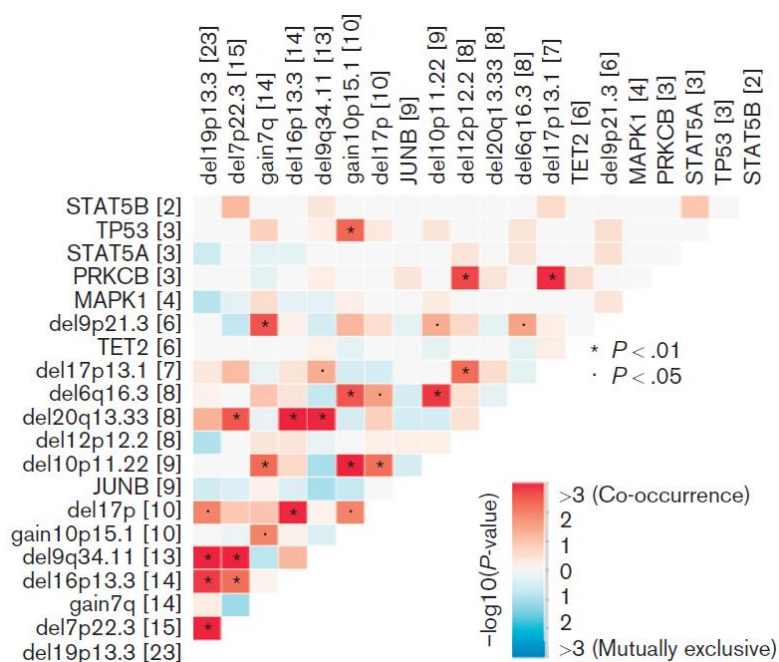


Figure 33 : Analyse des co-occurrences et des exclusions mutuelles des altérations génomiques récurrentes chez les 48 patients atteints de MF. Cette matrice (*Corrplot*) a été générée avec le package « *maftools* » sous R. Les paires d'altérations significatives ont été testées à l'aide du test exact de Fisher. Les cases en rouges indiquent qu'il existe une co-occurrence entre les altérations, tandis que les cases en bleu montrent une exclusion mutuelle des altérations. L'intensité de la couleur est proportionnelle à la $-\log_{10}(p\text{-value})$, avec des valeurs plus élevée signalant des relations plus significatives. Les p-values sont les suivantes : « . » pour une p-value < 0,05 ; « * » pour une p-value < 0,01.

Nos résultats sur l'hétérogénéité clonale des *drivers* et l'analyse des corrélations entre ces altérations ont permis de proposer une chronologie des altérations précoces et tardives dans le MF. Nous avons ordonné les mutations en analysant les échantillons contenant des paires d'altérations clonales et sous-clonales²⁰¹. En supposant que les mutations clonales surviennent avant les événements sous-clonaux, nous avons ainsi pu définir la chronologie des principales

altérations. Nous avons observé deux schémas généraux définissant la phylogénie du MF (Figure 34). Le premier est caractérisé par un gain7q clonal qui coïncide souvent avec une del9p21.3 (*CDKN2A*, *CDKN2B* et *MTAP*) ou une del10p11.22 (*ZEB1*), avec une acquisition supplémentaire d'un gain10p15.1 (*IL2RA*, *IL15RA*) lors de la progression. Le second repose sur une mutation clonale de la tyrosine kinase *PRKCB*, suivie par une del17p13.1 (*TP53*) et une del12p12.1 (*CDKN1B*). Il convient de noter que le gain7q et la del10p11.22 (*ZEB1*) sont souvent clonaux et associés à une évolution défavorable. Leur présence dès les premiers stades de la tumeur suggère que certains MF suivent un chemin évolutif vers une forme plus agressive. Le gain10p15.1 tardif (*IL2RA* et *IL15RA*), qui est également associé au HR, représente un événement moléculaire transformateur dans l'évolution du MF.

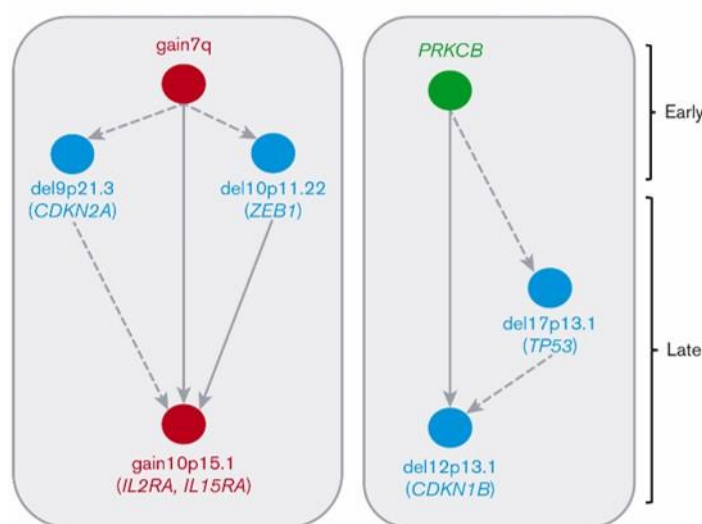


Figure 34 : Chronologie inférée des altérations génétiques dans le MF. Chronologie représentant l'ordre d'apparitions des altérations génomiques, avec les événements précoces en haut et les événements tardifs en bas. La couleur indique le type d'altération : en vert une mutation ponctuelle nucléotidique (SNV), en rouge les gains de copies et en bleu les délétions. Les flèches entre deux altérations ont été dessinées lorsque deux *drivers* ont été trouvés dans un échantillon avec un statut clonal vers sous-clonal. Si un *driver* d1 est clonal et un *driver* d2 est sous-clonal dans le même échantillon, ce modèle indique que d1 a été acquis avant d2 (noté d1 → d2). Les flèches pointillées indiquent n = 1 paire clonale/sous-clonale et les flèches pleines indiquent n ≥ 2 paires clonales/sous-clonales.

Nous avons analysé des échantillons séquentiels provenant de sept patients qui sont passés des stades précoces aux stades avancés du MF (LR *progressors*), avec des intervalles d'échantillonnage allant de 3 mois à 9 ans. Nous avons observé des preuves d'hétérogénéité clonale au moment du diagnostic dans les sept cas, ce qui indique que des ramifications clonales s'étaient déjà produites au stade précoce de la maladie. De plus, nous avons remarqué des cas dans lesquels des sous-clones initialement petits au moment du diagnostic présentaient une expansion substantielle à un moment ultérieur, ce qui indique un fort potentiel *driver* pour ces mutations/SCNVs à devenir clonales. Plus précisément, ce phénomène a été observé chez trois patients présentant de petits sous-clones porteurs de la mutation *JUNB* ou de la del12p13.1 (*CDKN1B*), qui ont été sélectionnés et sont devenus dominants lors de la progression (Figure 35B-D). Dans ces cas, toutes les altérations détectées dans les échantillons prélevés à un stade avancé étaient déjà présentes au départ, ce qui indique un modèle d'évolution clonale linéaire. Cependant, dans un autre cas, un sous-clone mutant *JUNB* a été identifié au moment de la progression de la maladie, ce qui suggère qu'une nouvelle ramification s'est produite et a entraîné la progression du stade LR vers le stade HR (Figure 35A). Après vérification manuelle avec le logiciel de visualisation des lectures, IGV, cette mutation sur *JUNB* n'était pas présente les sorties brutes du logiciel détecteur de SNVs, MuTect (profondeur moyenne des échantillons LR et HR : 160X et 185X), ce qui pourrait confirmer que cette mutation a été acquise au cours de la progression. Dans ce cas, la mutation de *JUNB* s'est ajoutée à une mutation clonale *VAV1*, impliquée dans la signalisation du TCR et rapportée dans diverses tumeurs malignes à lymphocytes T (Figure 36).

Une évolution clonale sous traitement a été observée chez un patient atteint de MF progressif, où un sous-clone portant une del17p13.1 résistant s'est étendu après

plusieurs lignes de traitement au cours des trois années suivant la progression de la maladie (Figure 35E). Ce phénomène suggère que cette altération pourrait être associée à une résistance thérapeutique et une pression sélective provoquée par le traitement. À l'inverse, chez un patient stable, dont deux échantillons ont été séquencés au stade LR à cinq ans d'intervalle, aucune nouvelle altération ni modification de la CCF initiale n'a été détectée (Figure 35F). Cette observation indique que le MF peut rester stable génétiquement sur plusieurs années. L'hétérogénéité clonale initiale ne suffit pas à elle seule pour expliquer la progression de la maladie. D'autres mécanismes, comme les altérations du microenvironnement, la dérégulation du système immunitaire ou des pressions thérapeutiques, pourraient aussi avoir une implication dans l'évolution du MF.

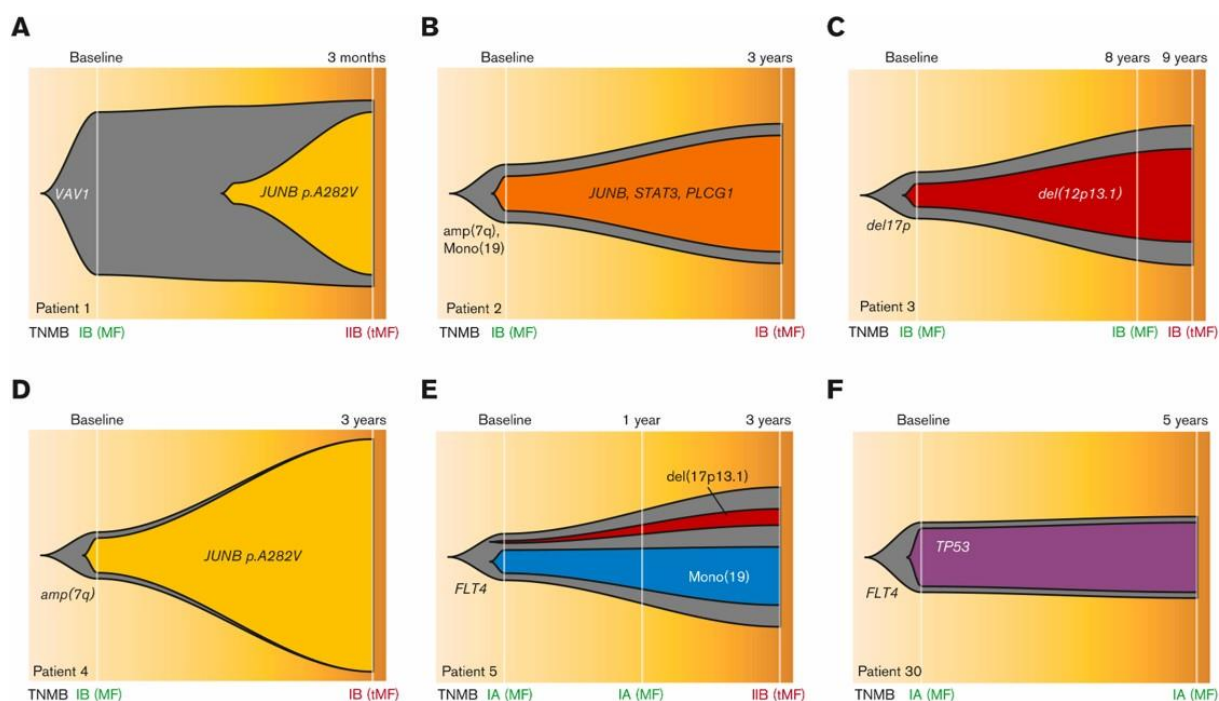


Figure 35 : Évolution clonale des échantillons séquentiels. A-E. Fish plots de cinq cas séquentiels d'échantillons aux stades LR qui ont progressé aux stades HR. **F.** Fish plot de deux échantillons séquentiels LR chez un patient qui n'a pas progressé. La classification

TNMB a été indiquée (en vert pour les échantillons LR et en rouge foncé pour les échantillons HR). tMF, MF transformé. Les fish plots ont été générés avec le paquet Github « [chrisamiller/fishplot](#) » sous R.

3.6.6 Modifications des Voies de Signalisations

Certains des SNVs et SCNVs *drivers* candidats participent dans de nombreuses voies de signalisation et sont en mesure de les altérer. C'est le cas des voies canoniques et non-canoniques TCR/NF- κ B (comme la délétion de *NFKB2* en 10q24.32, le gain de *PRKCQ* en 10p15.1, la mutation *PRCKB*, la délétion de *CARD11* en 7p22.3), la voie JAK/STAT (gain de *IL2RA* et *IL15RA* en 10p15.1, mutations de *STAT3*, *STAT5A* et *STAT5B*), la voie MAPK (mutation *JUNB* et *MAPK1*), la voie NOTCH (délétion *NOTCH1* et mutation *MAML2*), la régulation des CDK et du point de restriction G1/S du cycle cellulaire (délétion de *CDKN1B* en 12p13.1 et délétion *CDKN2A* en 9p21.3) et l'inhibition de l'apoptose (délétion et mutations de *TP53* et délétion de *TNFAIP3* en 6p16.3) (Figure 36).

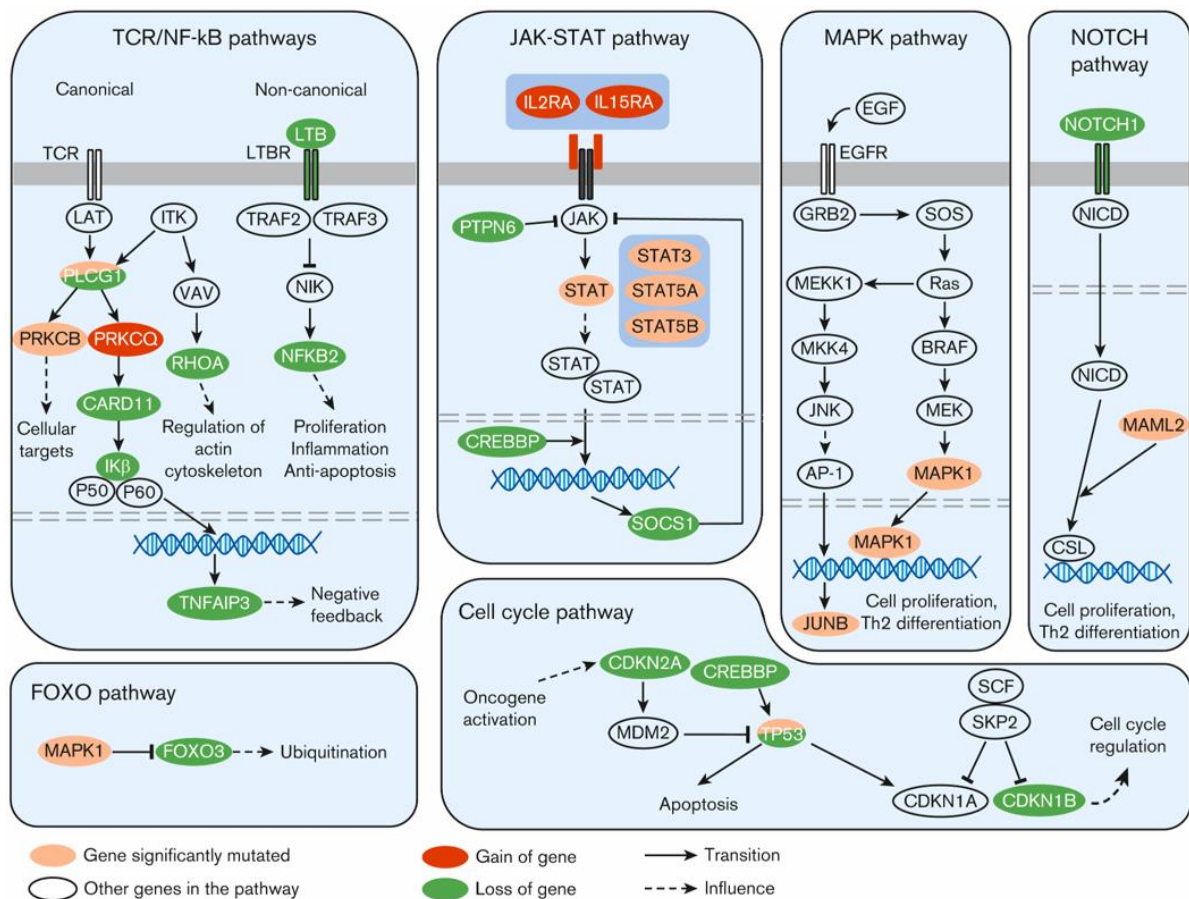


Figure 36 : Altérations génétiques récurrentes affectant les principales voies de signalisation impliquées dans le MF. Dans cette représentation schématique, les gènes fréquemment mutés dont le rôle est bien établi dans ces voies de signalisation ont été représentés avec les protéines qu'ils codent. TCR pour le récepteur des lymphocytes T.

Chapitre 4

Discussion

Sommaire

4.1	Cohorte Mycosis Fongoïde	126
4.1.1	Vue d'Ensemble sur la Cohorte	126
4.1.2	Échantillons <i>Tumor-Only</i> et <i>Tumor-Normal</i>	127
4.1.3	Cohorte PoN de Dix Échantillons de Sang	128
4.1.4	Amélioration de la Cohorte	129
4.2	Séquençage	130
4.2.1	Whole Exome Sequencing et Whole Génome Séquençage	130
4.2.2	Whole Exome Sequencing et le Séquençage Long-Read	132
4.2.3	Régions de Faible Complexité	133
4.2.4	Nombre de Lectures Séquencées et Profondeur de Séquençage	133
4.3	Filtres Supplémentaires dans la Réduction de Faux Positifs	135
4.3.1	Panel of Normal du Broad Institute et Filtre <i>Tumor-Only</i>	135
4.3.2	Retrait des Variants Synonymes	136
4.3.3	Filtres Internes	136
4.4	Paysage mutationnel	138
4.4.1	Survie Globale	138
4.4.2	Évolution Clonale	140
4.4.3	Épigénétique	141
4.4.4	Signatures Mutationnelles	142
4.4.5	Potentiels Biomarqueurs Identifiés	144
4.5	Perspectives	145

4.1 Cohorte Mycosis Fongoïde

4.1.1 Vue d'Ensemble sur la Cohorte

La cohorte étudiée dans cette thèse comprenait 48 patients atteints de MF, une maladie rare caractérisée par une grande hétérogénéité clinique et génomique. Un 49^{ème} patient avait fourni un échantillon tumoral présentant une très faible fraction tumorale (< 5%), rendant son analyse non exploitable. Par conséquent, l'échantillon a été reclassé comme normal et a été intégré dans le PoN. Le principal point fort de cette cohorte est la diversité des stades du MF, allant de formes indolentes LR, avec présence de macules et papules cutanées, à des formes plus agressives HR, marquées par des tumeurs et une érythrodermie. La répartition des patients entre LR (26 ; 54%) et HR (22 ; 46%) est relativement équilibrée. En revanche, le nombre total d'échantillons analysés est légèrement inférieur dans les stades LR (29 ; 43%) par rapport aux stades HR (38 ; 57%). Cependant, la présence de stades LR est une caractéristique peu courante dans l'étude du MF et des CTCLs en général^{17,92,98,99,101}. Parmi les 67 échantillons séquencés, 18 échantillons séquentiels ont permis de suivre l'évolution de la maladie et de différencier les patients qui ont progressé de ceux qui sont restés au stade LR. Cette diversité a permis l'analyse approfondie pour la détection des facteurs de progression du MF.

Toutefois, la taille relativement réduite de la cohorte et le nombre limité d'échantillons représentent une contrainte majeure. En effet, une étude avec un plan d'échantillonnage limité peut introduire des biais statistiques, ce qui réduit la robustesse des conclusions tirées¹²⁵. L'inclusion d'un plus grand nombre de patients provenant du CHU de Lille ou d'autres centres hospitaliers permettrait de renforcer la puissance des tests statistiques de survie et de détecter davantage d'associations entre les mutations et la progression de la maladie. De plus, lors des échanges avec

le département de biostatistiques du CHU de Lille, il a été déterminé que les événements génomiques apparaissant moins de sept fois dans la cohorte ne seraient pas inclus dans les tests statistiques de survie, afin d'assurer la fiabilité des résultats. Cette approche, bien qu'indispensable pour éviter des interprétations hasardeuses, a conduit à l'exclusion mutations rares qui auraient pu être pertinentes sur le plan biologique. Une cohorte plus large aurait ainsi permis d'intégrer ces événements dans l'analyse et d'obtenir une vision plus complète du paysage génomique du MF.

4.1.2 Échantillons *Tumor-Only* et *Tumor-Normal*

L'utilisation d'échantillons TO, une approche commune dans la pratique clinique, peut s'avérer problématique en ce qui concerne leur fiabilité. En effet, les études qui n'utilisent que des échantillons TO amènent à un risque accru de variants FPs pouvant aller jusqu'à 69% des mutations somatiques détectées²⁰². De plus, le nombre limité de trois échantillons TN est très faible pour contrôler la présence de ces FPs dans l'ensemble de la cohorte. Généralement, les études qui combinent des échantillons TO et TN ont des proportions plus équilibrées d'échantillons afin de comparer les fréquences des variants détectés entre les deux groupes et ainsi isoler et exclure de potentiels FPs^{163,202}. Dans l'ensemble de l'analyse bioinformatique, les échantillons tumoraux ont été associés avec un normal non apparié lors que sa présence été requise. Ils ont été utilisés seuls lorsqu'un logiciel pouvait fonctionner en mode TO. Les échantillons TN ont été analysés à la fois en mode appariés et en mode TO pour comparer le nombre de FPs détectés assez élevé, ce qui a justifié l'utilisation d'étapes de filtrages strictes.

La combinaison des échantillons tumoraux avec un échantillon normal non apparié est une bonne stratégie pour éliminer certains événements constitutionnels

connus. Il a été montré que cette combinaison était efficace pour identifier les variants somatiques sous-clonaux, dont beaucoup étaient hétérogènes entre les biopsies, mais il manquait les variants avec une VAF > 45%, en raison des filtres utilisés pour enlever ceux qui étaient constitutionnels²⁰².

Notre analyse, comme d'autres essais cliniques, ne possède pas beaucoup d'échantillons normaux qui peuvent servir à éliminer des variants constitutionnels propres aux individus. Les raisons évoquées sont généralement le coût du WES et des exigences complexes en matière de consentement éclairé des patients pour le prélèvement d'échantillons normaux. Dans notre cas, nous nous rapprochons plutôt dans la seconde catégorie où nos échantillons remontaient jusqu'en 2003 et où peu de demandes de prélèvements d'échantillons normaux avaient été réalisées.

4.1.3 Cohorte PoN de Dix Échantillons de Sang

L'origine de cette cohorte provient des échanges avec l'équipe de la Dr. Irene Ghobrial et du Dr. Gad Getz pour agrandir et améliorer le PoN utilisé dans l'analyse. Avoir plus d'échantillons dans le PoN est toujours utile afin d'éliminer des variants constitutionnels et FPs techniques récurrents causés par des types d'erreurs rares qui ne peuvent être détectés qu'avec l'ajout d'échantillons supplémentaires¹⁴⁶. L'utilisation d'échantillons extérieurs présente l'inconvénient d'être moins spécifique aux patients de notre cohorte, ce qui peut limiter la détection de mutations rares. C'est pourquoi avoir les trois échantillons normaux provenant de nos patients en plus dans le PoN peut permettre de pallier ce manque de spécificité.

D'après les recommandations du BI, le critère de sélection principale des échantillons à mettre dans le PoN est qu'ils doivent avoir été générés de la manière la plus identique possible. Les échantillons de sang frais provenant du *Research Blood*

Components ont été conservés sur de l'anticoagulant EDTA, n'ont pas dû passer par l'étape de macrodissection, ils ont subi une étape d'extraction avec un kit spécifique pour le sang. À part ces différences, les échantillons de sang et de peau sont passés par les mêmes étapes à partir de la construction des bibliothèques.

Il est généralement conseillé de réaliser un PoN à partir d'échantillons sains de sang. Il existe des études qui mélangent des échantillons de tissus solides et de sang pour la construction du PoN pour à la fois augmenter son efficacité et recouvrir un plus grand nombre de variants constitutionnels¹⁴⁶. Cependant, il existe un risque de contamination quand les échantillons normaux sont adjacents à la tumeur et que des cellules tumorales sont prélevées. Nos trois échantillons normaux se trouvaient bien à proximité du site de la tumeur. Bien qu'une vérification au microscope ait été faite pour s'assurer de l'absence de cellules tumorales, une contamination minime par ce biais est toujours possible. Malgré cela, aucune contamination de cellules tumorales dans les échantillons normaux cutanés n'a significativement été relevée par la mesure du logiciel DeTiN.

Il n'existe pas de recommandations précises sur la taille minimale du PoN. Le fait d'avoir un PoN reste une meilleure option que de ne pas en avoir. Le BI recommande un chiffre minimal arbitraire dans la pratique de 40 échantillons. Ce chiffre peut servir de cap pour de prochaines analyses.

4.1.4 Amélioration de la Cohorte

Il existe plusieurs points d'amélioration de la cohorte et de la robustesse de l'analyse :

- L'augmentation de la taille de la cohorte et du nombre d'échantillons à analyser. Si possible, il faut garder l'équilibre entre le nombre

d'échantillons LR et HR et continuer à avoir des échantillons séquentiels de patients qui sont les deux points forts de l'analyse. Un plus grand nombre d'échantillons et de patients permettrait des tests biostatistiques plus solides et la découverte de variants sous-clonaux à faible fréquence.

- La présence d'une cohorte indépendante de validation d'au moins une dizaine d'échantillons permettrait de confirmer les résultats obtenus et la robustesse des analyses.
- La création d'un PoN d'environ 40 échantillons normaux provenant des patients internes ou externes de la cohorte pour les raisons évoquées précédemment.

4.2 Séquençage

4.2.1 Whole Exome Sequencing et Whole Génome Séquencing

Le WES permet d'obtenir une profondeur importante dans les régions codantes du génome, ce qui est un avantage pour identifier des mutations somatiques dans des gènes. Le séquençage par la technologie Illumina NovaSeq 6000 dans notre analyse nous a permis d'obtenir une profondeur de séquençage moyenne de 231,85X. En pratique, il n'est pas possible d'atteindre ce niveau de profondeur avec la technologie WGS. Théoriquement, le WGS pourrait atteindre une profondeur similaire, cependant cela générerait une quantité de données considérablement plus élevée, entraînant des contraintes majeures en termes de capacité de stockage, de temps d'analyse et de puissance de calcul. De plus, le coût d'un WGS à une telle profondeur serait exponentiellement plus élevé, rendant cette approche actuellement irréaliste en recherche clinique.

L'apport du WGS est de pouvoir explorer l'oncogenèse du MF à l'échelle du génome entier, notamment dans la recherche de mutations dans les régions non-codantes et l'étude des variants structuraux. Il existe quelques études du WGS dans le MF et chez les CTCLs^{89,92,95,111,193}. Dans l'étude de Bastidas Torres *et al.* en 2018, une moyenne de 39 réarrangements génomiques inter et intra-chromosomiaux et de type chromothripsis, chez trois patients, ont été détectés dans le MF, dont certains ont un impact sur des gènes connus dans le MF ou le SS (*CDKN2A*, *CDKN2B*, *DLEU2*, *KDM6A*, *TP53*, *TP63*, and *VAV1*) ou d'autres cancers⁹⁵. Chez deux patients, les gènes *ARID1A*, *CDKN2A/B*, *PTPRC*, *SOCS1*, et *STK11* sont délétés en conséquence d'un réarrangement génomique⁹⁵. Également, la même analyse a pu confirmer d'existence de 24 transcrits de fusion dans le MF causées par les réarrangements et les délétions interstitielles, telles que *ATXN1-TP63* et *TP53-GPR3*^{95,203}. Un autre transcrit de fusion concerne les gènes *CTLA4:CD28* dans le SS^{94,102,204}. Le gène *CD28* est impliqué dans la prolifération cellulaire, mais aussi dans la transcription de *CTLA4* qui, à son tour, contrecarre les effets de *CD28* et inhibe la prolifération. Par conséquent, ce transcrit de fusion peut être un mécanisme aberrant qui favorise la prolifération et la survie des cellules oncogènes²⁰³.

Malgré les avantages du WGS, le WES reste une méthode privilégiée dans de nombreuses études exploratoires pour son rapport coût-efficacité, la détection de variants somatiques dans des gènes impliqués dans le cancer et une meilleure sensibilité dans la détection des mutations clonales et sous-clonales dans les tumeurs hétérogènes^{89,91,92,126,193,198,199}.

4.2.2 Whole Exome Sequencing et le Séquençage Long-Read

Le séquençage de 3^{ème} génération « long-read » est une technologie NGS qui, contrairement au WES, permet de séquencer de longues lectures : entre 20 et 100kbp avec la technologie SMRT de PacBio et jusqu'à 800kbp avec celle ONT de Oxford Nanopore. Le WES et les autres technologies « short reads » ne sont pas adaptées pour capturer des altérations étendues, des régions répétées ou riches en GC. Le long-read est en mesure d'évaluer la taille réelle de ces régions répétées, mais aussi d'avoir les séquences avant et après celles-ci ; ce qui améliore le séquençage, l'assemblage et la détection des variants dans des régions de faible complexité²⁰⁵. L'un des autres avantages principaux est la détection de variants structuraux de plus de 50bp, dans les génomes cancéreux, comme des inversions et des translocations, ou des délétions de grande taille²⁰⁶. Une étude récente de 2022 combinant WES et séquençage long-read ONT sur le MF précoce a permis de détecter les points de cassures (*breakpoints*) des variants structuraux impliqués dans des SCNVs classiques : les délétions de *CDKN2A* (chr9) et *ARID1A* (chr1) et l'amplification de *STAT5* (chr17). Cette découverte apporte une compréhension approfondie sur les changements génomiques ayant lieu dans les stades précoces du MF²⁰⁷.

Cependant, la précision du séquençage long-read reste aujourd'hui plus faible par rapport au WES, avec un taux d'erreur d'appel des bases d'environ 10% contre 1%²⁰⁸. Le taux d'erreur est tout de même en nette diminution, s'approchant des 1%²⁰⁹. Le long-read reste aussi plus cher que d'autres technologies NGS avec la préparation des échantillons et l'automatisation de certaines tâches. Cependant ce coût de séquençage a aussi considérablement diminué dans les environs de 600 dollars aujourd'hui²⁰⁷. Bien que le séquençage long-read semble être une technologie

prometteuse à l'avenir pour la détection de variants, le WES reste une méthode plus efficace pour la détection des mutations somatiques dans les régions codantes.

4.2.3 Régions de Faible Complexité

Le WES présente des limitations dans les régions de faible complexité, qui sont des séquences ADN caractérisée par une composition nucléotidique répétitive ou peu diversifiée. En effet, les lectures courtes d'environ 150 pb générés par WES ne sont pas suffisantes pour couvrir ces zones complexes du génome²⁰². L'aligneur BWA présente des difficultés à placer les lectures sur ces régions du génome de référence à cause de leur ambiguïté et les limitations du séquenceur¹⁴⁰. Les régions de faible complexité sont sources de FPs lors de la détection de variants où les algorithmes interprètent ces difficultés d'alignement comme des mutations. Des études comme celle de Shi *et al.* 2018 proposent d'exclure ces régions des analyses pour réduire le nombre de FPs à l'aide d'un filtre spécifique qui s'appuie sur une BdD qui recense ces zones.

De notre côté, nous avons vérifié manuellement chaque variant significatif en utilisant le logiciel IGV. Cette approche, bien que plus subjective et coûteuse en temps, a permis de visualiser l'alignement des lectures dans les régions autour de chaque variant et d'éliminer de potentiels FPs présents dans des régions répétitives ou riches en GC²¹⁰. Par exemple, l'un des filtres du logiciel aide à identifier les lectures avec une mauvaise qualité d'alignement sur le génome de référence au niveau de ces régions de faible complexité.

4.2.4 Nombre de Lectures Séquencées et Profondeur de Séquençage

Bien que nous ayons atteint une profondeur moyenne de 231,85X, il existe une certaine hétérogénéité dans le nombre de lectures séquencées parmi nos échantillons.

En effet, deux d'entre eux avec le moins de lectures séquencées possèdent une profondeur proche de 80X, tandis que ceux avec le plus de lectures ont une profondeur autour de 300-350X (Figure 16 et Figure 17). Nous observons une corrélation modérée, mais significative entre la profondeur de séquençage et le nombre de mutations validées par les filtres (test de Spearman, $\rho = 0,36$; $p\text{-value} = 0,003$). Il en résulte une capacité moindre à détecter des mutations somatiques, en particulier sous-clonales, présentes à des VAF plus basses. Nous avons observé que dans nos deux échantillons ayant une profondeur de séquençage plus faible, les algorithmes bioinformatiques ont rencontré plus de difficulté à identifier les SCNVs (1 et 5 SCNVs détectées ; test de Wilcoxon, $p\text{-value} < 0,001$ pour les deux comparaisons) par rapport aux SNVs (137 et 93 SNVs ; test de Wilcoxon, $p\text{-value} = 0,504$ et $p\text{-value} < 0,001$). Par ailleurs, un nombre de lectures insuffisant augmente le risque de FPs, car il devient plus difficile de générer un alignement consensus fiable des séquences contre le génome de référence.

Plusieurs facteurs peuvent expliquer une profondeur de séquençage inégale parmi les échantillons. Une faible quantité d'ADN au moment de l'extraction, des erreurs survenant lors de la préparation des bibliothèques ou durant le processus de séquençage des données. Bien qu'une profondeur de 80X puisse être acceptable pour la détection de variants somatiques les plus fréquents, elle est nettement insuffisante pour identifier des clones rares. De manière générale, une profondeur minimale de 100X est recommandée pour commencer à détecter des sous-clones¹⁴³. Ce constat requiert d'être extrêmement vigilant lors des différentes étapes de QC, en particulier pendant la sélection hybride, où la bonne correspondance entre les sondes et les cibles est cruciale. Une suggestion serait d'exclure les échantillons ayant une profondeur inférieure à 50X (même si aucun de nos échantillons ne présente une telle

profondeur), tout en surveillant attentivement ceux dont la couverture se situe entre 50 et 100X.

4.3 Filtres Supplémentaires dans la Réduction de Faux Positifs

4.3.1 Panel of Normal du Broad Institute et Filtre *Tumor-Only*

L'une des principales difficultés dans l'analyse bioinformatique des données de séquençage est de minimiser les FPs tout en maximisant la détection de mutations véritablement somatiques. Dans notre étude, plusieurs filtres ont été mis en place afin d'améliorer la spécificité des résultats et de réduire les erreurs. Leur application a joué un rôle clé dans l'élimination des artefacts techniques et des variants constitutionnels, nous permettant ainsi de concentrer notre attention sur les mutations plus probablement somatiques dans le cadre du MF.

L'absence d'échantillons normaux appariés pour retirer les variants constitutionnels explique en grande partie la différence importante du nombre d'événements relevés par les algorithmes de détection de SNVs et InDels (Figure 19).

Le PoN du BI a constitué un premier filtre essentiel dans notre analyse. Bien que ce PoN soit recommandé, sa version complète est à accès restreint et nous avons pu y accéder grâce à l'un des membres de l'équipe du Dr. Ghobrial. Ce filtre nous a permis de retirer les variants récurrents présents chez des individus sains, réduisant ainsi le nombre de FPs issus de problèmes techniques.

Le filtre interne *Tumor-Only*, mis en place par certains ingénieurs de l'équipe du Dr. Getz¹⁶³, a également été utilisé. Il faut noter que la version gratuite de ce filtre contient une erreur dans le calcul des CCF, ce qui nous a conduits à utiliser une version non officielle obtenue directement en contactant son créateur. Ce filtre nous a

permis d'ajuster la sévérité de l'élimination des variants des échantillons TO, en fonction d'un seuil choisi. Cependant, ce filtre a montré un meilleur fonctionnement dans les échantillons avec une pureté tumorale plus modérée, ce qui n'était pas totalement compatible avec une grande partie de nos échantillons où la pureté avait été améliorée par macrodissection.

De plus, bien que ce filtre soit efficace, il n'est pas suffisant pour retirer tous variants constitutionnels. En effet, les créateurs de ce filtre ont comparé des échantillons en utilisant à la fois des algorithmes TN et TO. Ils ont observé une persistance des variants germinaux à une fréquence moyenne de 3,3/Mb dans les échantillons TO après le filtrage. Malgré ces limitations, ce filtre s'est montré pertinent dans le retrait d'une grande partie des mutations constitutionnels et des FPs dans nos échantillons. Néanmoins, le nombre moyen de variants par échantillon TO reste élevé (531, Figure 19), ce qui a nécessité des étapes supplémentaires de filtrage pour affiner davantage les résultats.

4.3.2 Retrait des Variants Synonymes

Les variants synonymes ont été exclus de cette analyse, car ils sont traditionnellement considérés comme non-fonctionnels (mutations silencieuses). Cependant, depuis quelques années, leurs rôles potentiels sont mieux connus dans la vitesse de traduction qui influence la structuration des protéines et dans l'épissage alternatif qui régule l'expression des gènes^{211,212}. Ces mutations seront sûrement amenées à être conservées dans les analyses génomiques à venir.

4.3.3 Filtres Internes

Pour affiner davantage notre analyse, nous avons développé nos propres filtres internes afin de discriminer plus efficacement les variants pertinents.

Le filtre gnomAD a été mis en place en s'inspirant du filtre ExAC (qui a depuis été remplacé par gnomAD), utilisé dans l'article Chapuy *et al.* 2018¹⁶³. La BdD gnomAD est particulièrement utile pour éliminer les variants communs présents dans la population générale. Cependant, cet outil possède une couverture incomplète pour certaines populations moins bien représentées, ce qui peut entraîner la rétention de variants rares propres à certaines populations sous-étudiées.

Afin d'améliorer la performance de notre PoN, qui a été limité à treize échantillons, nous avons cherché à renforcer sa capacité à discerner et à retirer un variant constitutionnel. Au lieu de partir uniquement des fichiers post-alignement où plusieurs lectures ne sont pas alignées contre le génome de référence (1,4% par échantillon, soit environ 2,3 millions de lectures par échantillon), nous avons introduit un filtre s'appuyant sur les fichiers FASTQ normaux. Ce filtre supplémentaire vérifie si le variant est présent directement dans les fichiers FASTQ, qui concentrent toutes les lectures non alignées. Après plusieurs tests pour ajuster la sévérité du filtre, nous avons opté pour l'élimination de tous les variants apparaissant plus d'une fois dans ce PoN personnalisé. En résumé, ce filtre et son critère strict ont été créés pour pallier le manque d'échantillons dans notre PoN initial.

Le filtrage des *fishy genes* a été une étape finale visant à retirer certains gènes aberrants identifiés par MutSig2CV. Il n'existe pas beaucoup de descriptions sur ces gènes, mis à part dans la publication originale de Lawrence *et al.* 2013¹²². Ils apparaissent souvent dans les listes de gènes mutés de manière significative, bien que leur rôle dans le cancer soit douteux. De manière générale, la présence des *fishy genes* dans les listes de gènes fréquemment mutés proviendrait de la façon dont les tests statistiques sont appliqués sans tenir compte des facteurs comme la grande taille des gènes, les niveaux d'expression ou les caractéristiques génomiques spécifiques.

Par exemple, les gènes situés dans les zones de réplication tardive de l'ADN ont tendance à avoir des taux de mutation plus élevés¹²². Filtrer ces gènes permet d'améliorer la précision des analyses et d'éviter de fausses associations entre mutations et cancers.

Pour conclure, une grande partie de la thèse s'est concentrée sur le filtrage de plusieurs centaines de variants par échantillon, principalement des variants d'origine germinale et des artefacts, qui polluaient le signal mutationnel et nous empêchaient de tirer des conclusions claires. Nos critères ont peut-être pu être trop sévères, et il est possible que certains variants vrais-positifs aient été filtrés, mais c'est une conséquence de l'analyse d'échantillons TO : Faut-il être trop restrictif ou pas assez ? Dans le cadre d'une analyse exploratoire comme la nôtre, nous recommandons d'utiliser des critères stricts, avec des décisions binaires dans un premier temps. Ainsi, l'analyse bioinformatique et l'interprétation des variants sont facilitées. Par la suite, si une mutation candidate semble avoir été filtrée à tort, il est important d'analyser objectivement la raison de son exclusion et d'ajuster, si nécessaire, la sévérité des filtres. Cependant, dans notre analyse, nous n'avons pas rencontré de cas nécessitant un tel ajustement.

4.4 Paysage mutationnel

4.4.1 Survie Globale

L'analyse de l'OS et de la progression des patients atteints de MF a révélé des tendances importantes concernant l'impact des mutations somatiques identifiées sur le pronostic. Nous avons mis en évidence que SCNVS dans des gènes jouant un rôle dans le développement du MF, qui étaient significativement associées à une OS plus

courte : del10p11.22 (*ZEB1*), gain10p15.1 (*IL2RA* et *IL15RA*), gain7q et del6q16.3 (*TNFAIP3*).

Nos résultats indiquent qu'il existe une corrélation significative entre la TMB et le stade du patient, suggérant que le nombre de mutations pourrait jouer un rôle dans la résistance aux traitements et la progression tumorale. Ce taux mutationnel plus élevé chez les patients de HR est majoritairement associée à la signature mutationnelle liée à l'exposition aux UVs, ce qui souligne l'impact aggravant de cette exposition dans la progression de la maladie. Toutefois, il est important de noter que cette corrélation ne reflète pas nécessairement un lien de causalité direct, mais pourrait simplement refléter l'accumulation de mutations au fil du temps en raison de l'exposition solaire quotidienne. Ainsi, l'ajout progressif de mutations *drivers* pourrait être un effet de sélection évolutive sur le long terme. D'autant plus que nous avons vu une corrélation entre l'âge du patient et la charge tumorale, ce qui renforce l'idée d'un lien temporel entre l'accumulation de mutations et la progression de la maladie. Une idée d'amélioration future serait de poursuivre des études sur l'évolution clonale en suivant sur le long terme plus de patients et en renseignant avec précision les informations sur ceux ayant des antécédents d'exposition chronique aux UV à ceux qui n'en ont pas.

Il existe d'autres altérations génomiques associées aux stades HR du MF qui semblent également jouer un rôle dans la progression tumorale : le gain large du bras chromosomique 7q, le gain focal 10p15.1 (*IL2RA* et *IL15RA*), les délétions focales 10q24.32 (*NFKB2*) et 10p11.22 (*ZEB1*) Inversement, la délétion focale 17q25.3 (*SUZ12*) était associée aux stades LR du MF.

4.4.2 Évolution Clonale

L'analyse d'échantillons séquentiels a mis en évidence une hétérogénéité clonale avec des ramifications clonales déjà présentes à des stades précoces de la maladie. Typiquement, les sous-clones ont montré une expansion importante dans les échantillons séquentiels ultérieurs, ce qui indique qu'ils pourraient avoir un rôle de *driver* potentiel dans la progression. Par exemple, la mutation *JUNB* ou la délétion del12p13.1 (*CDKN1B*) ont leur CCF qui a augmenté avec la progression de la maladie. De plus, une évolution clonale a été observée pendant le traitement, avec l'expansion d'un sous-clone résistant portant la del17p13.1 chez un patient au cours des 3 années de suivi de la maladie. Cela confirme l'importance de surveiller la dynamique clonale au fil du temps, un aspect encore peu étudié dans le contexte du MF avec l'utilisation d'échantillons séquentiels, mais qui pourrait avoir des implications cliniques importantes pour le suivi des patients. Il est important que l'hétérogénéité clonale soit mieux connue, et ce, dès les stades précoces, afin d'avoir au plus tôt des indicateurs de la trajectoire future du MF.

Nous avons également pu établir des modèles de phylogénie tumorale du MF, mettant en évidence des événements précoces, tels que le gain7q, les del9p21.3 (*CDKN2A*) et del10p11.22 (*ZEB1*), ainsi que des mutations dans le gène *PRKCB*, qui semblent survenir dans les premiers stades de la maladie. Par la suite, l'acquisition ultérieure d'altérations, comme le gain10p15.1 (*IL2RA* et *IL15RA*), del17p13.1 (*TP53*) et del12p13.1 (*CDKN1B*) sont acquises, suggérant leur rôle dans la progression de du MF.

4.4.3 Épigénétique

Dans nos résultats, nous avons identifié des mutations impactant le gène *TET2* qui possède un rôle épigénétique dans la déméthylation de l'ADN dans le MF et les CTCLs^{91,98,126,171}. *TET2* est une enzyme qui, lorsqu'elle est dérégulée, va provoquer des oxydations répétées de la 5-méthylcytosine et provoquer la production de 5-hydroxyméthylcytosine. Des oxydations supplémentaires vont former la 5-formylcytosine et la 5-carboxylcytosine. L'ensemble de ces méthylcytosines oxydées sont des intermédiaires dans la déméthylation de l'ADN, perturbant ainsi la régulation des gènes¹⁷¹.

Une amélioration possible de notre analyse pour mieux comprendre l'impact de ces altérations sur la chromatine et la régulation épigénétique aurait été d'utiliser des techniques de séquençage comme le ChIP-seq ou l'ATAC-seq.

Le ChIP-seq (*Chromatin Immunoprecipitation*) permet d'étudier la liaison des protéines, comme les TFs ou les histones modifiées, sur l'ADN. Cette méthode peut aider à identifier les modifications post-traductionnelles des cytosines associées aux mutations génétiques observées dans des gènes comme *TET2*²¹³.

L'ATAC-seq (*Assay for Transposase-Accessible Chromatin*), une technologie plus récente, permet d'identifier les régions accessibles de la chromatine, indiquant les zones où l'ADN est plus susceptible d'interagir avec les TFs et les éléments régulateurs. Cette méthode nous aurait permis de savoir si les régions proches des mutations somatiques présentaient une accessibilité accrue ou réduite²¹⁴.

Il existe peu de publications sur l'utilisation du CHIP-seq et de l'ATAC-seq sur le CTCLs, souvent elles concernent le stade HR du MF ou le SS²¹⁵⁻²¹⁷. Par exemple, ces études montrent que l'accessibilité à la chromatine était différente entre les

échantillons tumoraux et les normaux. Il y a une activation de NF- κ B, associée à celle de l'un des trois motifs de TF : Jun-AP1, CTCF, ou un ensemble de motifs comprenant EGR, SMAD, MYC et KLF²¹⁷. Une autre étude CHIP-Seq combinée à de la RT-PCR a montré une diminution de l'expression du microARN miR-29b dans les lymphocytes T des patients atteints de CTCL, ce qui conduit à une augmentation de la liaison de la protéine BRD4 aux promoteurs de gènes oncogènes comme *NOTCH1* et *RBPJ*. En outre, BRD4 active des gènes associés à la signalisation autocrine de l'IL-15, renforçant la prolifération des lymphocytes malins²¹⁶. Une autre étude, utilisant la RT-PCR, a identifié et proposé trois autres microARN (miR-106b-5p, miR-148a-3p et miR-338-3p) comme marqueurs pour distinguer les groupes de patients HR et LR²¹⁸.

L'application de ces techniques au cours de la thèse aurait pu offrir une compréhension plus fine des interactions épigénétiques impactant les mutations identifiées et compléter ainsi nos résultats.

4.4.4 Signatures Mutationnelles

L'une des deux signatures mutationnelles identifiées dans notre étude correspond à SBS7b, qui appartient au groupe de quatre signatures recensées dans la BdD COSMIC associées à l'exposition directe aux UVs. Retrouver cette signature est cohérent avec la littérature, car les signatures SBS7a et SBS7b sont les deux plus relevées dans l'étude des CTCLs^{17,196,199}. Dans notre jeu de données, elle est significativement liée aux stades HR et enrichie dans les mutations touchant les gènes pancancer, tels que *TET2*, *MAPK1*, *TP53* et *PRKCB*.

Cependant, bien que cette signature soit à l'origine de plusieurs mutations, le rôle des UVs dans le développement des LCPs, tels que le MF, fait l'objet d'un débat¹⁹⁹. Tout d'abord, le MF tend à se développer principalement dans les régions

non-exposées à la lumière du soleil, contrairement à d'autres cancers UV-dépendants qui apparaissent généralement dans les zones exposées aux UVs ou qui ont souffert de brûlures solaires antérieures¹. Ensuite, il a été montré que les UVs jouent un rôle mineur dans l'apparition et le développement du MF, contrairement à d'autres cancers cutanés comme le mélanome malin¹⁹⁹. En effet, dans cet article de 2023, les deux signatures UVs, SBS7 et SBS38, sont retrouvées dans 92% des mutations des échantillons WES du mélanome, alors que SBS7 n'est retrouvée que dans environ 5% de la TMB du MF. Une autre étude de 2020 avait noté un taux de mutations UV de SBS7 dans le MF plus élevé, 34,8%, mais ce taux est à prendre avec prudence car il provient d'un séquençage ciblé sur 585 gènes⁹⁸. De plus, les mutations provoquées par les UVs sont concentrées dans la couche épidermique et non plus profondément dans le derme⁹⁷. Cet article suggère que ces mutations se sont accumulées dans l'épiderme à la suite d'une exposition environnementale et ne sont pas des événements précoces et causaux¹⁹⁹. Cette observation souligne l'importance future de systématiquement distinguer les couches cutanées dans les études sur le MF, afin de mieux mesurer et interpréter l'impact environnemental, comme cela a déjà été fait dans quelques études^{97,199}.

Toutefois, la cellule d'origine la plus probable pour le MF serait le lymphocyte T mémoire résident dans la peau⁶³. Contrairement à d'autres lymphocytes T migrants, ces cellules, une fois arrivées dans la peau, ne recirculent pas et elles resteraient dans les couches cutanées, depuis le derme jusqu'à l'épiderme. De ce fait, ces lymphocytes sont donc exposés aux UVs autant que d'autres cellules cutanées natives comme les kératinocytes ou les mélanocytes^{199,219}. Cela explique pourquoi des mutations C>T, caractéristiques de la signature SBS7, ont été retrouvées dans le MF, par exemple

dans le gène *TP53*^{220,221}, ce qui peut confirmer l'important rôle des UV dans cette maladie¹⁹⁸.

Tout comme SBS7b, la signature mutationnelle SBS15 est principalement caractérisée par des substitutions du type C>T. SBS15 est associée à des erreurs dans la réparation des mésappariements de l'ADN, ce qui conduit à une instabilité des microsatellites. Bien que le MF présente des perturbations au niveau des microsatellites, la signature SBS15 n'a pas encore été identifiée dans les CTCLs. De plus, cette signature n'est pas exclusivement associée à un type de cancer spécifique¹²⁷.

4.4.5 Potentiels Biomarqueurs Identifiés

Au cours de notre analyse, divers biomarqueurs potentiels pour le MF ont été détectés. Parmi les SCNVs, des gains chromosomiques larges et focaux ont été identifiés, tels que le gain7q et le gain10p15.1 (*IL2RA* et *IL15RA*), mais également des délétions focales, telles que les délétions del10q24.32 (*NFKB2*) et del10p11.22 (*ZEB1*). Ces quatre événements sont associés aux stades HR, ce qui suggère leur rôle dans la progression du MF. Ils pourraient donc être des biomarqueurs importants pour évaluer la sévérité de la maladie. À l'inverse, la délétion focale del17q25.3 (*SUZ12*) était spécifiquement associée aux stades LR, ce qui pourrait indiquer que cette altération participe d'une quelconque manière à limiter la progression du MF.

Nos analyses montrent également que plusieurs de ces mutations affectent des gènes impliqués dans des voies de signalisation importantes, notamment NF- κ B, JAK/STAT, et MAPK, qui contrôlent des processus essentiels, comme le contrôle de la prolifération cellulaire, l'inflammation et l'apoptose. L'activation de ces voies pourrait favoriser la survie et la prolifération des cellules tumorales et contribuer à l'agressivité

du MF dans les stades avancés. Par conséquent, ces biomarqueurs pourraient être utilisées dans le développement de thérapies ciblées.

Pour finir, afin de valider l'identification de ces biomarqueurs potentiels, il est indispensable de réaliser des analyses fonctionnelles supplémentaires. Par exemple, il serait important de vérifier si les mutations observées dans des gènes, tels que *JUNB*, peuvent effectivement être considérés comme des indicateurs associés à des sous-clones plus agressifs.

4.5 Perspectives

Notre étude a permis de mettre en évidence de nouvelles altérations génétiques et génomiques, ainsi que des mutations impactant des gènes liés à la progression du MF. Ces découvertes contribuent à approfondir la compréhension des mécanismes oncogéniques de la maladie. Le MF, bien que peu fréquent, est le CTCL le plus communément rencontré, et les connaissances sur son paysage mutationnel et son évolution clonale sont encore limitées. Les travaux de cette thèse ont contribué à identifier des biomarqueurs candidats dans la progression et l'agressivité de la maladie. Grâce à l'analyse d'échantillons séquentiels, nous avons pu démontrer l'hétérogénéité clonale dès les premiers stades de la maladie et la progression de sous-clones spécifiques, tels que ceux portant des mutations dans *JUNB* ou *TP53*.

Les échantillons et les données génomiques obtenus au cours de cette étude ont été intégrés dans la BdD publique SRA. Cette étape facilite non seulement l'accès aux données par la communauté scientifique, mais ouvre également la voie à de potentielles méta-analyses, qui pourraient confirmer ou affiner nos résultats.

Malgré nos résultats encourageants, il est important de procéder à la validation fonctionnelle des biomarqueurs identifiés. À ce jour, aucune étude sur le MF n'a

exploré l'utilisation des technologies d'extinction de gène (*Gene Silencing*), comme les ARN interférents (RNAi) ou Crisper-Cas9, pour évaluer l'impact fonctionnel de ces altérations. Il existe également des stratégies thérapeutiques ciblées, comme le traitement cellulaire CTX-130, s'appuyant sur la technologie CAR-T Cells, qui sont en cours de développement. Ce traitement cible spécifiquement la protéine CD70, exprimée à la surface des lymphocytes T tumoraux²²².

L'analyse des seules altérations génétiques ne suffit pas à capturer l'ensemble des mécanismes sous-jacents à la maladie. Une étude complémentaire intégrant l'analyse du transcriptome, du protéome et de l'épitranscriptome permettrait d'obtenir une vision plus globale des perturbations cellulaires dans le MF. En particulier, l'épitranscriptome qui représente l'ensemble des modifications chimiques post-transcriptionnelles apportées aux ARN sans modifier la séquence nucléotidique, est très pertinent pour comprendre l'impact fonctionnel des altérations épigénétiques détectées, telles que les mutations dans *TET2*. Par ailleurs, l'impact du microenvironnement, c'est à dire les interactions entre les cellules tumorales et les cellules immunitaires, demeure un domaine encore peu exploré dans le MF, notamment par l'utilisation de la technologie single-cell RNA-seq²²³.

Au final, cette étude a apporté une contribution importante à la compréhension du MF, tout en ouvrant la voie à des validations fonctionnelles et des analyses approfondies sur le microenvironnement tumoral et les voies de signalisation associées à la progression de la maladie.

Chapitre 5

Conclusion

Au cours de cette thèse, j'ai utilisé les données WES pour examiner les échantillons tumoraux prélevés au moment du diagnostic chez 48 patients atteints de MF, ainsi que des échantillons séquentiels récupérés ultérieurement pour sept de ces patients. Plus précisément, j'ai analysé les altérations génomiques les plus fréquentes, ainsi que l'hétérogénéité et l'évolution clonale dans le but d'ordonner de manière chronologique ces altérations afin de mieux comprendre la phylogénie du MF. Mes résultats soulignent la complexité du MF, avec une médiane de 135 altérations génomiques différentes (SNVs, InDels) et plusieurs SCNVs larges et focales par tumeur. Les altérations les plus récurrentes détectées dans le MF étaient cohérentes avec les observations précédentes faites dans les lymphomes T (MF, SS, maladies lymphoprolifératives T cutanées primitives, CTCL γ - δ)^{17,89,92,94,97}, avec des altérations dans les voies de signalisation NF- κ B, JAK/STAT, MAPK, du cycle cellulaire et de l'inhibition de l'apoptose (Figure 36).

J'ai identifié des altérations génomiques liées aux stades HR de la maladie, telles que le gain large du bras chromosomique 7q et le gain focal 10p15.1 (*IL2RA* et *IL15RA*), ainsi que les délétions focales 10q24.32 (*NFKB2*) et 10p11.22 (*ZEB1*), ce qui suggère leur rôle dans la progression du MF. Inversement, la délétion focale 17q25.3 (*SUZ12*) était associée aux stades LR de la maladie. La signature mutationnelle SBS7 correspondant à l'exposition à la lumière UV était également associée aux stades HR et enrichie dans le processus mutationnel de gènes tels que *TET2*, *MAPK1*, *TP53* et *PRKCB*. J'ai également décrit différents modèles de phylogénie du MF, mettant en évidence des événements précoces tels que le gain7q ou des mutations dans *PRKCB* et l'acquisition ultérieure d'altérations telles que le gain10p15.1 (*IL2RA* et *IL15RA*), del17p13.1 (*TP53*) ou del12p13.1 (*CDKN1B*). L'évolution clonale des stades LR à HR était soit linéaire, soit ramifiée, avec la sélection de sous-clones à fort potentiel évolutif.

J'ai également évalué la valeur pronostique de ces altérations génomiques au moment du diagnostic. J'ai identifié quatre SCNVs significativement associées à une OS plus courte : del10p11.22 (*ZEB1*), gain10p15.1 (*IL2RA* et *IL15RA*), gain7q et del6q16.3 (*TNFAIP3*). Ainsi, les multiples anomalies des voies de l'interleukine 2 et 15 observées pourraient constituer des cibles thérapeutiques potentielles dans le MF. L'identification de ces altérations génomiques HR peut aider à la gestion clinique du MF, en particulier pour les lésions focales précoces du MF qui peuvent être difficiles à diagnostiquer et à différencier histologiquement des dermatites eczémateuses ou psoriasiques^{22,23,133}. Des altérations génomiques spécifiques associées à des pronostics défavorables peuvent aider à orienter le diagnostic du MF.

D'un point de vue personnel, cette thèse m'a permis d'apporter mon expérience en bioinformatique dans un sujet d'étude fortement lié à la recherche clinique. Bien que j'étais familiarisée, avec mes études universitaires, sur le nettoyage et l'alignement de données exomiques, la recherche de variants a été une grande nouveauté très intéressante. Bien que la plupart des outils développés par le BI soient bien documentés sur le site du GATK, certains aspects techniques restaient lacunaires. Ce fait m'a parfois conduit à contacter directement leurs créateurs afin d'obtenir des précisions pour les faire fonctionner correctement, malgré l'ancienneté de certains logiciels, développés il y a plus d'une dizaine d'années. L'idée qu'un pipeline bioinformatique soit entièrement automatisé est souvent trompeuse. En réalité, bien que le pipeline du BI soit présenté comme étant adapté aux données TO, sa mise en pratique nécessitait des ajustements techniques importants de ma part pour faire fonctionner chaque logiciel.

La plupart du temps, chaque fichier de sortie d'un logiciel doit être réadapté en terme de format pour être utilisé en entrée pour le logiciel suivant. Je devais ajuster

les options pour chaque étape et également assurer une gestion régulière des ressources du cluster de calcul en terme d'espace et de mémoire vive. Le débogage était fréquemment nécessaire et impliquait une vérification régulière du code du pipeline ou de la version du logiciel utilisé. Après avoir standardisé le pipeline, l'implémentation des filtres a constitué la seconde partie qui a pris le plus de temps au cours de cette thèse. La création du filtre « FASTQ Normaux » a nécessité le codage de mon propre script de gestion de pipeline, m'apprenant la logique d'articulation de logiciels. J'ai beaucoup apprécié d'avoir eu ce type d'expérience.

En conclusion, l'analyse génomique d'une cohorte de patients atteints de MF, comprenant des échantillons séquentiels, m'a permis d'appréhender la complexité génomique du MF, la chronologie des événements somatiques, ainsi que les altérations associées au HR et à la progression de la maladie. Je pense que la réalisation d'une analyse NGS de certains biomarqueurs au moment du diagnostic du MF peut améliorer l'identification des patients présentant un HR de progression.

Références

1. Willemze, R. *et al.* WHO-EORTC classification for cutaneous lymphomas. *Blood* **105**, 3768–85 (2005).
2. Kempf, W., Zimmermann, A.-K. & Mitteldorf, C. Cutaneous lymphomas—An update 2019. *Hematological Oncology* **37**, 43–47 (2019).
3. Willemze, R. *et al.* The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood* **133**, 1703–1714 (2019).
4. Dobos, G. *et al.* Epidemiological changes in cutaneous lymphomas: an analysis of 8593 patients from the French Cutaneous Lymphoma Registry*. *British Journal of Dermatology* **184**, 1059–1067 (2021).
5. Willemze, R. Primary cutaneous lymphoma: the 2018 update of the WHO-EORTC classification. *La Presse Médicale* **51**, 104126 (2022).
6. Alaggio, R. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* **36**, 1720–1748 (2022).
7. Alibert, J. L. M. *Description des maladies de la peau: observées à l'hôpital St. Louis et exposition des meilleurs méthodes suivies pour leur traitement*. vol. 2 (Barois l'ainé et fils, 1806).
8. Willemze, R. Cutaneous T-Cell Lymphoma: Epidemiology, Etiology, and Classification. *Leukemia & Lymphoma* **44**, S49–S54 (2003).
9. Amorim, G. M., Niemeyer-Corbellini, J. P., Quintella, D. C., Cuzzi, T. & Ramos-e-Silva, M. Clinical and epidemiological profile of patients with early stage mycosis fungoides. *An. Bras. Dermatol.* **93**, 546–552 (2018).
10. Pimpinelli, N. *et al.* Defining early mycosis fungoides. *Journal of the American Academy of Dermatology* **53**, 1053–1063 (2005).

11. Lutzner, M. *et al.* Cutaneous T-Cell Lymphomas: The Sézary Syndrome, Mycosis Fungoides, and Related Disorders. *Ann Intern Med* **83**, 534–552 (1975).
12. Nielsen, P. *et al.* Role of B-cells in Mycosis Fungoides. *Acta Derm Venereol* **101**, adv00413 (2021).
13. Raghavan, S. S., Hong, E. K., Kim, Y. H. & Kim, J. Utility of CD30, Ki-67, and p53 in assisting with the diagnosis of mycosis fungoides with large cell transformation. *Journal of Cutaneous Pathology* **46**, 33–43 (2019).
14. Massone, C., Crisman, G., Kerl, H. & Cerroni, L. The prognosis of early mycosis fungoides is not influenced by phenotype and T-cell clonality. *British Journal of Dermatology* **159**, 881–886 (2008).
15. Martinez-Escala, M. E. *et al.* CD8 + mycosis fungoides: A low-grade lymphoproliferative disorder. *Journal of the American Academy of Dermatology* **77**, 489–496 (2017).
16. Kempf, W. *et al.* CD4/CD8 Double Negative Mycosis Fungoides With PD-1 (CD279) Expression—A Disease of Follicular Helper T-Cells? *The American Journal of Dermatopathology* **34**, 757–761 (2012).
17. Park, J. *et al.* Integrated genomic analyses of cutaneous T-cell lymphomas reveal the molecular bases for disease heterogeneity. *Blood* **138**, 1225–1236 (2021).
18. Wang, R. C. *et al.* Mycosis fungoides in Taiwan shows a relatively high frequency of large cell transformation and CD56 expression. *Pathology* **50**, 718–724 (2018).
19. Kim, M., Park, M. I., Lim, M. & Kim, J. Cytotoxic Variant of Mycosis Fungoides with CD8+ CD56+ Phenotype: A Case Report and Review of Literature. *Korean J Pathol* **48**, 390–393 (2014).
20. Cribier, B. J. The myth of Pautrier's microabscesses. *Journal of the American Academy of Dermatology* **48**, 796–797 (2003).

21. Elder, D., Massi, D., Scolyer, R. & Willemze, R. *WHO Classification of Skin Tumours*. (IARC, 2018).
22. Ahn, C. S., ALSayyah, A. & Sangüeza, O. P. Mycosis Fungoides: An Updated Review of Clinicopathologic Variants. *The American Journal of Dermatopathology* **36**, 933 (2014).
23. Kelati, A., Gallouj, S., Tahiri, L., Harmouche, T. & Mernissi, F. Z. Defining the mimics and clinico-histological diagnosis criteria for mycosis fungoides to minimize misdiagnosis. *International Journal of Women's Dermatology* **3**, 100–106 (2017).
24. Bunn, P. A. & Lamberg, S. I. Report of the Committee on Staging and Classification of Cutaneous T-Cell Lymphomas. *Cancer Treat Rep* **63**, 725–728 (1979).
25. Olsen, E. *et al.* Revisions to the staging and classification of mycosis fungoides and Sézary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC). *Blood* **110**, 1713–1722 (2007).
26. Olsen, E. A. *et al.* Primary cutaneous lymphoma: recommendations for clinical trial design and staging update from the ISCL, USCLC, and EORTC. *Blood* **140**, 419–437 (2022).
27. Agar, N. S. *et al.* Survival outcomes and prognostic factors in mycosis fungoides/Sézary syndrome: validation of the revised International Society for Cutaneous Lymphomas/European Organisation for Research and Treatment of Cancer staging proposal. *J Clin Oncol* **28**, 4730–4739 (2010).
28. Salhany, K. E. *et al.* Transformation of cutaneous T cell lymphoma to large cell lymphoma. A clinicopathologic and immunologic study. *Am J Pathol* **132**, 265–277 (1988).
29. Vergier, B. *et al.* Transformation of mycosis fungoides: clinicopathological and prognostic features of 45 cases. *Blood* **95**, 2212–2218 (2000).

30. van Doorn, R. *et al.* Mycosis Fungoides: Disease Evolution and Prognosis of 309 Dutch Patients. *Archives of Dermatology* **136**, 504–510 (2000).
31. Benner, M. F., Jansen, P. M., Vermeer, M. H. & Willemze, R. Prognostic factors in transformed mycosis fungoides: a retrospective analysis of 100 cases. *Blood* **119**, 1643–1649 (2012).
32. Mitteldorf, C., Stadler, R., Sander, C. A. & Kempf, W. Folliculotropic mycosis fungoides. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft* **16**, 543–557 (2018).
33. van Doorn, R., Scheffer, E. & Willemze, R. Follicular mycosis fungoides, a distinct disease entity with or without associated follicular mucinosis: a clinicopathologic and follow-up study of 51 patients. *Arch Dermatol* **138**, 191–198 (2002).
34. Stahly, S., Manway, M., Lin, C. C. & Sukpraprut-Braaten, S. Pagetoid Reticulosis: A Rare Dermatologic Malignancy Presenting in a Middle-Aged Female. *Cureus* (2021) doi:10.7759/cureus.18524.
35. Yamashita, T., Abbade, L. P. F., Marques, M. E. A. & Marques, S. A. Mycosis fungoides and Sézary syndrome: clinical, histopathological and immunohistochemical review and update. *An Bras Dermatol* **87**, 817–830 (2012).
36. Kempf, W. *et al.* Granulomatous Mycosis Fungoides and Granulomatous Slack Skin: A Multicenter Study of the Cutaneous Lymphoma Histopathology Task Force Group of the European Organization for Research and Treatment of Cancer (EORTC). *Archives of Dermatology* **144**, 1609–1617 (2008).
37. Levy, R. *et al.* Transformation of Mycosis Fungoides: T-Cell Receptor β Gene Analysis Demonstrates a Common Clonal Origin for Plaque-Type Mycosis Fungoides and CD30+ Large-Cell Lymphoma. *Journal of Investigative Dermatology* **101**, 296–300 (1993).

38. Diamandidou, E., Colome-Grimmer, M., Fayad, L., Duvic, M. & Kurzrock, R. Transformation of Mycosis Fungoides/Sezary Syndrome: Clinical Characteristics and Prognosis. *Blood* **92**, 1150–1159 (1998).
39. Cerroni, L., Rieger, E., Hödl, S. & Kerl, H. Clinicopathologic and Immunologic Features Associated With Transformation of Mycosis Fungoides to Large-Cell Lymphoma. *The American Journal of Surgical Pathology* **16**, 543 (1992).
40. Wolfe, J. T. *et al.* Large-cell transformation following detection of minimal residual disease in cutaneous T-cell lymphoma: molecular and in situ analysis of a single neoplastic T-cell clone expressing the identical T-cell receptor. *JCO* **13**, 1751–1757 (1995).
41. Vural, S. *et al.* Transformation of Mycosis Fungoides/Sezary Syndrome: Clinical Characteristics and Prognosis. *Turk J Haematol* **35**, 35–41 (2018).
42. Kadin, M. E., Hughey, L. C. & Wood, G. S. Large-cell transformation of mycosis fungoides—differential diagnosis with implications for clinical management: A consensus statement of the US Cutaneous Lymphoma Consortium. *Journal of the American Academy of Dermatology* **70**, 374–376 (2014).
43. Scarisbrick, J. J. *et al.* Cutaneous Lymphoma International Consortium Study of Outcome in Advanced Stages of Mycosis Fungoides and Sézary Syndrome: Effect of Specific Prognostic Markers on Survival and Development of a Prognostic Model. *JCO* **33**, 3766–3773 (2015).
44. Arulogun, S. O. *et al.* Long-term outcomes of patients with advanced-stage cutaneous T-cell lymphoma and large cell transformation. *Blood* **112**, 3082–3087 (2008).
45. Talpur, R. *et al.* CD25 Expression Is Correlated with Histological Grade and Response to Denileukin Difitox in Cutaneous T-Cell Lymphoma. *Journal of Investigative Dermatology* **126**, 575–583 (2006).

46. Dmitrovsky, E. *et al.* Cytologic transformation in cutaneous T cell lymphoma: a clinicopathologic entity associated with poor prognosis. *JCO* **5**, 208–215 (1987).
47. Greer, J. P. *et al.* Clinical features associated with transformation of cerebriform T-cell lymphoma to a large cell process. *Hematol Oncol* **8**, 215–227 (1990).
48. Barberio, E., Thomas, L., Skowron, F., Balme, B. & Dalle, S. Transformed mycosis fungoides: clinicopathological features and outcome. *British Journal of Dermatology* **157**, 284–289 (2007).
49. Herrmann, J. L. & Hughey, L. C. Recognizing large-cell transformation of mycosis fungoides. *Journal of the American Academy of Dermatology* **67**, 665–672 (2012).
50. Bittencourt, A. L., Oliveira, P. D., Carvalho-Andrade, A. & Araújo, I. Correlations between clinical and pathological features in 17 cases of mycosis fungoides before and after transformation. *International Journal of Dermatology* **54**, e327–e331 (2015).
51. Talpur, R., Sui, D., Gangar, P., Dabaja, B. S. & Duvic, M. Retrospective Analysis of Prognostic Factors in 187 Cases of Transformed Mycosis Fungoides. *Clinical Lymphoma Myeloma and Leukemia* **16**, 49–56 (2016).
52. Baykal Selcuk, L., Aksu Arica, D., Yayli, S., Seyman, Ü. & Bahadir, S. Mycosis Fungoides: A 10-year Turkish Experience. *Acta Dermatovenereol Croat* **27**, 153–158 (2019).
53. Lim, H. I. j. *et al.* Epidemiology and prognostic factors for mycosis fungoides and Sézary syndrome in a multi-ethnic Asian cohort: a 12-year review. *Journal of the European Academy of Dermatology and Venereology* **33**, 1513–1521 (2019).
54. Lansigan, F. *et al.* Outcomes of Patients with Transformed Mycosis Fungoides: Analysis from a Prospective Multicenter US Cohort Study. *Clin Lymphoma Myeloma Leuk* **20**, 744–748 (2020).

55. O'Donnell, M. *et al.* Disease characteristics, prognosis, and response to therapy in patients with large-cell transformed mycosis fungoides: A single-center retrospective study. *Journal of the American Academy of Dermatology* **86**, 1285–1292 (2022).
56. Battistella, M. *et al.* KIR3DL2 expression in cutaneous T-cell lymphomas: expanding the spectrum for KIR3DL2 targeting. *Blood* **130**, 2900–2902 (2017).
57. Ortonne, N. *et al.* CD158k/KIR3DL2 and NKp46 are frequently expressed in transformed mycosis fungoides. *Experimental Dermatology* **21**, 461–463 (2012).
58. Volk, A. L., Vannucci, S. A., Cook, W., Thompson, K. A. & Listinsky, C. M. Composite mycosis fungoides and B-cell chronic lymphocytic leukemia. *Annals of Diagnostic Pathology* **6**, 172–182 (2002).
59. Whitling, N. A. *et al.* Composite lymphoma of mycosis fungoides and cutaneous small B-cell lymphoma in a 73-year-old male patient. *Human Pathology* **44**, 670–675 (2013).
60. Chilom, D. S., Farcaș, S. S. & Andreescu, N. I. Primary Cutaneous B-Cell Lymphoma Co-Existing with Mycosis Fungoides—A Case Report and Overview of the Literature. *Life* **12**, 2067 (2022).
61. Rohan, T. Z. *et al.* Coexistence of large cell transformed mycosis fungoides and diffuse large B-cell lymphoma in one patient. *Journal of Cutaneous Pathology* **51**, 761–766 (2024).
62. CzEch Leukemia Study Group - for Life. CELL - Atlas of Haematological Cytology. <https://www.leukemia-cell.org/atlas/index.php>.
63. Campbell, J. J., Clark, R. A., Watanabe, R. & Kupper, T. S. Sézary syndrome and mycosis fungoides arise from distinct T-cell subsets: a biologic rationale for their distinct clinical behaviors. *Blood* **116**, 767–771 (2010).
64. Sokolowska-Wojdylo, M. *et al.* Circulating clonal CLA⁺ and CD4⁺ T cells in Sezary syndrome express the skin-homing chemokine receptors CCR4 and CCR10 as well

- as the lymph node-homing chemokine receptor CCR7. *Br J Dermatol* **152**, 258–264 (2005).
65. Novelli, M. *et al.* Blood Flow Cytometry in Sézary Syndrome. *American Journal of Clinical Pathology* **143**, 57–69 (2015).
 66. Vonderheid, E. C. *et al.* Update on erythrodermic cutaneous T-cell lymphoma: report of the International Society for Cutaneous Lymphomas. *J Am Acad Dermatol* **46**, 95–106 (2002).
 67. Diwan, A. H., Prieto, V. G., Herling, M., Duvic, M. & Jone, D. Primary Sézary syndrome commonly shows low-grade cytologic atypia and an absence of epidermotropism. *Am J Clin Pathol* **123**, 510–515 (2005).
 68. Larocca, C. & Kupper, T. Mycosis Fungoides and Sézary Syndrome. *Hematology/Oncology Clinics of North America* **33**, 103–120 (2019).
 69. Kirsch, I. R. *et al.* TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci Transl Med* **7**, 308ra158 (2015).
 70. van Doorn, R. *et al.* Oncogenomic analysis of mycosis fungoides reveals major differences with Sézary syndrome. *Blood* **113**, 127–136 (2009).
 71. Laharanne, E. *et al.* Genome-Wide Analysis of Cutaneous T-Cell Lymphomas Identifies Three Clinically Relevant Classes. *Journal of Investigative Dermatology* **130**, 1707–1718 (2010).
 72. Michel, L., Jean-Louis, F., Begue, E., Bensussan, A. & Bagot, M. Use of PLS3, Twist, CD158k/KIR3DL2, and NKp46 gene expression combination for reliable Sézary syndrome diagnosis. *Blood* **121**, 1477–1478 (2013).
 73. Wernham, A. G., Shah, F., Amel-Kashipaz, R., Cobbold, M. & Scarisbrick, J. Stage I mycosis fungoides: frequent association with a favourable prognosis but disease progression and disease-specific mortality may occur. *Br J Dermatol* **173**, 1295–1297 (2015).

74. Edinger, J. T., Clark, B. Z., Pucevich, B. E., Geskin, L. J. & Swerdlow, S. H. CD30 Expression and Proliferative Fraction in Nontransformed Mycosis Fungoides. *American Journal of Surgical Pathology* **33**, 1860–1868 (2009).
75. Latzka, J. *et al.* EORTC consensus recommendations for the treatment of mycosis fungoides/Sézary syndrome – Update 2023. *European Journal of Cancer* **195**, 113343 (2023).
76. Trautinger, F. *et al.* European Organisation for Research and Treatment of Cancer consensus recommendations for the treatment of mycosis fungoides/Sézary syndrome – Update 2017. *European Journal of Cancer* **77**, 57–74 (2017).
77. Gosmann, J., Stadler, R., Quint, K. D., Gutzmer, R. & Vermeer, M. H. Use of Pegylated Interferon Alpha-2a in Cutaneous T-cell Lymphoma: A Retrospective Case Collection. *Acta Derm Venereol* **103**, adv10306 (2023).
78. Duvic, M. *et al.* Phase 2 and 3 Clinical Trial of Oral Bexarotene (Targretin Capsules) for the Treatment of Refractory or Persistent Early-Stage Cutaneous T-Cell Lymphoma. *Archives of Dermatology* **137**, 581–593 (2001).
79. Piekarz, R. L. *et al.* Phase II Multi-Institutional Trial of the Histone Deacetylase Inhibitor Romidepsin As Monotherapy for Patients With Cutaneous T-Cell Lymphoma. *JCO* **27**, 5410–5417 (2009).
80. Falkenhain-López, D. *et al.* Single-centre experience of using pegylated liposomal doxorubicin as maintenance treatment in mycosis fungoides. *Br J Dermatol* **186**, 363–365 (2022).
81. Blazejak, C. *et al.* Clinical Outcomes of Advanced-Stage Cutaneous Lymphoma under Low-Dose Gemcitabine Treatment: Real-Life Data from the German Cutaneous Lymphoma Network. *Dermatology* **238**, 498–506 (2022).

82. de Masson, A. *et al.* Allogeneic transplantation in advanced cutaneous T-cell lymphomas (CUTALLO): a propensity score matched controlled prospective study. *The Lancet* **401**, 1941–1950 (2023).
83. Goyal, A., O’Leary, D. & Foss, F. Allogeneic stem cell transplant for treatment of mycosis fungoides and Sezary syndrome: a systematic review and meta-analysis. *Bone Marrow Transplant* **59**, 41–51 (2024).
84. Stratton, M. R. Exploring the Genomes of Cancer Cells: Progress and Promise. *Science* **331**, 1553–1558 (2011).
85. Greaves, M. & Maley, C. C. Clonal Evolution in Cancer. *Nature* **481**, 306.
86. Wong, H. K., Mishra, A., Hake, T. & Porcu, P. Evolving Insights in the Pathogenesis and Therapy of Cutaneous T-cell lymphoma (Mycosis Fungoides and Sezary Syndrome). *Br J Haematol* **155**, 150–166 (2011).
87. van Doorn, R. *et al.* A Novel Splice Variant of the Fas Gene in Patients with Cutaneous T-Cell Lymphoma. *Cancer Research* **62**, 5389–5392 (2002).
88. Caumont, C. *et al.* PLCG1 Gene Mutations Are Uncommon in Cutaneous T-Cell Lymphomas. *J Invest Dermatol* **135**, 2334–2337 (2015).
89. Choi, J. *et al.* Genomic landscape of cutaneous T cell lymphoma. *Nat Genet* **47**, 1011–1019 (2015).
90. Vaqué, J. P. *et al.* PLCG1 mutations in cutaneous T-cell lymphomas. *Blood* **123**, 2034–2043 (2014).
91. da Silva Almeida, A. C. *et al.* The mutational landscape of cutaneous T-cell lymphoma and Sézary syndrome. *Nat Genet* **47**, 1465–1470 (2015).
92. McGirt, L. Y. *et al.* Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood* **126**, 508–519 (2015).
93. Chang, L.-W. *et al.* An Integrated Data Resource for Genomic Analysis of Cutaneous T-Cell Lymphoma. *Journal of Investigative Dermatology* **138**, 2681–2683 (2018).

94. Ungewickell, A. *et al.* Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat Genet* **47**, 1056–1060 (2015).
95. Bastidas Torres, A. N. *et al.* Genomic analysis reveals recurrent deletion of JAK-STAT signaling inhibitors HNRNPK and SOCS1 in mycosis fungoides. *Genes, Chromosomes and Cancer* **57**, 653–664 (2018).
96. Khodadoust, M. S. *et al.* Pembrolizumab in Relapsed and Refractory Mycosis Fungoides and Sézary Syndrome: A Multicenter Phase II Study. *JCO* **38**, 20–28 (2020).
97. Iyer, A. *et al.* Independent evolution of cutaneous lymphoma subclones in different microenvironments of the skin. *Sci Rep* **10**, 15483 (2020).
98. Argyropoulos, K. V. *et al.* Targeted genomic analysis of cutaneous T cell lymphomas identifies a subset with aggressive clinicopathological features. *Blood Cancer J.* **10**, 1–5 (2020).
99. Park, J. *et al.* Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). *Blood* **130**, 1430–1440 (2017).
100. Kießling, M. K. *et al.* High-throughput mutation profiling of CTCL samples reveals KRAS and NRAS mutations sensitizing tumors toward inhibition of the RAS/RAF/MEK signaling cascade. *Blood* **117**, 2433–2440 (2011).
101. Scarisbrick, J. J., Woolford, A. J., Russell-Jones, R. & Whittaker, S. J. Loss of heterozygosity on 10q and microsatellite instability in advanced stages of primary cutaneous T-cell lymphoma and possible association with homozygous deletion of PTEN. *Blood* **95**, 2937–2942 (2000).
102. Wang, L. *et al.* Genomic profiling of Sézary Syndrome identifies alterations of key T-cell signaling and differentiation genes. *Nat Genet* **47**, 1426–1434 (2015).
103. Mao, X. *et al.* BCL2 and JUNB abnormalities in primary cutaneous lymphomas. *British Journal of Dermatology* **151**, 546–556 (2004).

104. Mao, X. *et al.* A genomic and expression study of AP-1 in primary cutaneous T-cell lymphoma: evidence for dysregulated expression of JUNB and JUND in MF and SS. *J Cutan Pathol* **35**, 899–910 (2008).
105. Laharanne, E. *et al.* CDKN2A–CDKN2B deletion defines an aggressive subset of cutaneous T-cell lymphoma. *Mod Pathol* **23**, 547–558 (2010).
106. Salgado, R. *et al.* Oligonucleotide Array-CGH Identifies Genomic Subgroups and Prognostic Markers for Tumor Stage Mycosis Fungoides. *Journal of Investigative Dermatology* **130**, 1126–1135 (2010).
107. Vermeer, M. H. *et al.* Novel and Highly Recurrent Chromosomal Alterations in Sézary Syndrome. *Cancer Research* **68**, 2689–2698 (2008).
108. Caprini, E. *et al.* Loss of the candidate tumor suppressor ZEB1 (TCF8, ZFHX1A) in Sézary syndrome. *Cell Death Dis* **9**, 1–15 (2018).
109. Caprini, E. *et al.* Identification of Key Regions and Genes Important in the Pathogenesis of Sézary Syndrome by Combining Genomic and Expression Microarrays. *Cancer Research* **69**, 8438–8446 (2009).
110. Lin, W. M. *et al.* Characterization of the DNA Copy-Number Genome in the Blood of Cutaneous T-Cell Lymphoma Patients. *Journal of Investigative Dermatology* **132**, 188–197 (2012).
111. Bastidas Torres, A. N. *et al.* Deregulation of JAK2 signaling underlies primary cutaneous CD8⁺ aggressive epidermotropic cytotoxic T-cell lymphoma. *haematol* **107**, 702–714 (2021).
112. Mao, X. *et al.* Molecular cytogenetic characterization of Sézary syndrome. *Genes, Chromosomes and Cancer* **36**, 250–260 (2003).
113. Böni, R. *et al.* Allelic Deletion at 9p21–22 in Primary Cutaneous CD30⁺ Large Cell Lymphoma. *Journal of Investigative Dermatology* **115**, 1104–1107 (2000).

114. Nicolae-Cristea, A. r. *et al.* Diagnostic and prognostic significance of CDKN2A/CDKN2B deletions in patients with transformed mycosis fungoides and primary cutaneous CD30-positive lymphoproliferative disease. *British Journal of Dermatology* **172**, 784–788 (2015).
115. Cristofolletti, C. *et al.* Comprehensive analysis of PTEN status in Sézary syndrome. *Blood* **122**, 3511–3520 (2013).
116. Mao, X. *et al.* Amplification and overexpression of JUNB is associated with primary cutaneous T-cell lymphomas. *Blood* **101**, 1513–1519 (2003).
117. Aksentijevich, I. & Zhou, Q. NF- κ B Pathway in Autoinflammatory Diseases: Dysregulation of Protein Modifications by Ubiquitin Defines a New Category of Autoinflammatory Diseases. *Frontiers in Immunology* **8**, (2017).
118. Damsky, W. E. & Choi, J. Genetics of Cutaneous T Cell Lymphoma: From Bench to Bedside. *Curr. Treat. Options in Oncol.* **17**, 33 (2016).
119. Man, K. *et al.* The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells. *Nat Immunol* **14**, 1155–1165 (2013).
120. García-Díaz, N., Piris, M. Á., Ortiz-Romero, P. L. & Vaqué, J. P. Mycosis Fungoides and Sézary Syndrome: An Integrative Review of the Pathophysiology, Molecular Drivers, and Targeted Therapy. *Cancers* **13**, 1931 (2021).
121. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* **22**, 105–113 (2016).
122. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
123. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

124. Assaf, C. *et al.* High detection rate of T-cell receptor beta chain rearrangements in T-cell lymphoproliferations by family specific polymerase chain reaction in combination with the GeneScan technique and DNA sequencing. *Blood* **96**, 640–646 (2000).
125. Eklund, Y., Aronsson, A., Schmidtchen, A. & Relander, T. Mycosis Fungoides: A Retrospective Study of 44 Swedish Cases. *Acta Dermato-Venereologica* **96**, 669–673 (2016).
126. Iyer, A. *et al.* Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Advances* **3**, 1175–1184 (2019).
127. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
128. Whittaker, S. Molecular genetics of cutaneous lymphomas. *Ann N Y Acad Sci* **941**, 39–45 (2001).
129. Williams, T. M. *et al.* Identification of a Zinc Finger Protein that Inhibits IL-2 Gene Expression. *Science* **254**, 1791–1794 (1991).
130. Hara, N. & Sawada, Y. Epigenetics of Cutaneous T-Cell Lymphomas. *IJMS* **23**, 3538 (2022).
131. van Doorn, R. *et al.* Epigenetic Profiling of Cutaneous T-Cell Lymphoma: Promoter Hypermethylation of Multiple Tumor Suppressor Genes Including BCL7a, PTPRG, and p73. *JCO* **23**, 3886–3896 (2005).
132. Scarisbrick, J. J. *et al.* Microsatellite Instability Is Associated with Hypermethylation of the hMLH1 Gene and Reduced Gene Expression in Mycosis Fungoides. *J Invest Dermatol* **121**, 894–901 (2003).
133. Hristov, A. C., Tejasvi, T. & Wilcox, R. A. Mycosis fungoides and Sézary syndrome: 2019 update on diagnosis, risk-stratification, and management. *American Journal of Hematology* **94**, 1027–1041 (2019).

134. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
135. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (2010).
136. 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).
137. Richterich, P. Estimation of Errors in “Raw” DNA Sequences: A Validation Study. *Genome Res.* **8**, 251–259 (1998).
138. Ewing, B. & Green, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res.* **8**, 186–194 (1998).
139. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <http://arxiv.org/abs/1303.3997> (2013).
140. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* **43**, 11.10.1-11.10.33 (2013).
141. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research* **41**, e67 (2013).
142. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
143. Van der Auwera, G. A. & O'Connor, B. D. (Brian D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. (O'Reilly Media, Sebastopol, CA, 2020).
144. Yousefi, S. *et al.* A SNP panel for identification of DNA and RNA specimens. *BMC Genomics* **19**, 90 (2018).

145. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology* **12**, R41 (2011).
146. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219 (2013).
147. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
148. Benjamin, D. *et al.* *Calling Somatic SNVs and Indels with Mutect2*. <http://biorxiv.org/lookup/doi/10.1101/861054> (2019) doi:10.1101/861054.
149. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
150. Taylor-Weiner, A. *et al.* DeTiN : Overcoming Tumor in Normal Contamination. *Nature methods* **15**, 531 (2018).
151. Ramos, A. H. *et al.* Oncotator: Cancer Variant Annotation Tool. *Human Mutation* **36**, E2423–E2429 (2015).
152. Frankish, A. *et al.* GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Research* **51**, D942 (2022).
153. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).
154. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Medicine* **12**, 103 (2020).
155. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Research* **52**, D1210–D1217 (2024).
156. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).

157. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* **46**, D1062 (2017).
158. Sherry, S. T. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308–311 (2001).
159. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
160. Kasar, S. *et al.* Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun* **6**, 8866 (2015).
161. Kim, J. *et al.* Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet* **48**, 600–606 (2016).
162. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413–421 (2012).
163. Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**, 679–690 (2018).
164. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
165. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
166. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
167. Sha, D. *et al.* Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. *Cancer Discovery* **10**, 1808–1825 (2020).
168. Goodman, A. M. *et al.* Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Molecular Cancer Therapeutics* **16**, 2598–2608 (2017).

169. Conforti, F. *et al.* Sex-based differences of the tumor mutational burden and T-cell inflammation of the tumor microenvironment. *Ann Oncol* **30**, 653–655 (2019).
170. Luchtel, R. A. *et al.* Recurrent MSCE116K mutations in ALK-negative anaplastic large cell lymphoma. *Blood* **133**, 2776–2789 (2019).
171. Lio, C.-W. J., Yuita, H. & Rao, A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood* **134**, 1487–1497 (2019).
172. Iyer, A. *et al.* Branched evolution and genomic intratumor heterogeneity in the pathogenesis of cutaneous T-cell lymphoma. *Blood Advances* **4**, 2489–2500 (2020).
173. He, Y. *et al.* FOXA1 overexpression suppresses interferon signaling and immune response in cancer. *J Clin Invest* **131**, (2021).
174. Patel, V. M. *et al.* Frequent and Persistent PLCG1 Mutations in Sézary Cells Directly Enhance PLCγ1 Activity and Stimulate NFκB, AP-1, and NFAT Signaling. *J Invest Dermatol* **140**, 380-389.e4 (2020).
175. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
176. Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research* **49**, D458–D460 (2021).
177. Zhang, Q. *et al.* Activation of Jak/STAT proteins involved in signal transduction pathway mediated by receptor for interleukin 2 in malignant T lymphocytes derived from cutaneous anaplastic large T-cell lymphoma and Sezary syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9148–9153 (1996).
178. Eriksen, K. W. *et al.* Constitutive STAT3-activation in Sezary syndrome: tyrphostin AG490 inhibits STAT3-activation, interleukin-2 receptor expression and growth of leukemic Sezary cells. *Leukemia* **15**, 787–793 (2001).

179. Qin, J.-Z. *et al.* Constitutive and Interleukin-7- and Interleukin-15-Stimulated DNA Binding of STAT and Novel Factors in Cutaneous T Cell Lymphoma Cells. *J Invest Dermatol* **117**, 583–589 (2001).
180. Waldmann, T. A. The biology of interleukin-2 and interleukin-15: implications for cancer therapy and vaccine design. *Nat Rev Immunol* **6**, 595–601 (2006).
181. Weijer, M. L. van de *et al.* Quality Control of ER Membrane Proteins by the RNF185/Membralin Ubiquitin Ligase Complex. *Molecular Cell* **79**, 768-781.e7 (2020).
182. Aubrey, B. J., Strasser, A. & Kelly, G. L. Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harb Perspect Med* **6**, a026062 (2016).
183. Collado, M., Blasco, M. A. & Serrano, M. Cellular Senescence in Cancer and Aging. *Cell* **130**, 223–233 (2007).
184. Wang, X. *et al.* p300 plays a role in p16INK4a expression and cell cycle arrest. *Oncogene* **27**, 1894–1904 (2008).
185. Marjon, K. *et al.* MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. *Cell Reports* **15**, 574–587 (2016).
186. Richly, H., Aloia, L. & Di Croce, L. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis* **2**, e204 (2011).
187. Ghosh, S. *et al.* Alterations of 3p21.31 tumor suppressor genes in head and neck squamous cell carcinoma: Correlation with progression and prognosis. *International Journal of Cancer* **123**, 2594–2604 (2008).
188. Roche, J., West, J., Gemmill, R. & Drabkin, H. Chromosome 3p, gènes suppresseurs de tumeurs et gènes de sémaphorines en 3p21.3. *Med Sci (Paris)* **14**, 283 (1998).
189. Nairismägi, M.-L. *et al.* JAK-STAT and G-protein-coupled receptor signaling pathways are frequently altered in epitheliotropic intestinal T-cell lymphoma. *Leukemia* **30**, 1311–1319 (2016).

190. Morin, R. D. *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122**, 1256–1265 (2013).
191. Yusufova, N. *et al.* Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. *Nature* **589**, 299–305 (2021).
192. Mansouri, L. *et al.* Frequent NFKBIE deletions are associated with poor outcome in primary mediastinal B-cell lymphoma. *Blood* **128**, 2666–2670 (2016).
193. Kiel, M. J. *et al.* Genomic analyses reveal recurrent mutations in epigenetic modifiers and the JAK–STAT pathway in Sézary syndrome. *Nat Commun* **6**, 8470 (2015).
194. Han, J. J. *et al.* Prognostic and therapeutic significance of phosphorylated STAT3 and protein tyrosine phosphatase-6 in peripheral-T cell lymphoma. *Blood Cancer Journal* **8**, 1–11 (2018).
195. Gruber, T. A. *et al.* An Inv(16)(p13.3q24.3)-Encoded CBFA2T3-GLIS2 Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell* **22**, 683–697 (2012).
196. Daya-Grosjean, L. & Sarasin, A. The role of UV induced lesions in skin carcinogenesis: an overview of oncogene and tumor suppressor gene modifications in xeroderma pigmentosum skin tumors. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**, 43–56 (2005).
197. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
198. Jones, C. L. *et al.* Spectrum of mutational signatures in T-cell lymphoma reveals a key role for UV radiation in cutaneous T-cell lymphoma. *Sci Rep* **11**, 3962 (2021).
199. Gniadecki, R., O’Keefe, S., Hennessey, D. & Iyer, A. Is Cutaneous T-Cell Lymphoma Caused by Ultraviolet Radiation? A Comparison of UV Mutational Signatures in Malignant Melanoma and Mycosis Fungoides. *Cells* **12**, 1616 (2023).

200. Querfeld, C. *et al.* Multicenter Phase II Trial of Temozolomide in Mycosis Fungoides/Sézary Syndrome: Correlation with O6-Methylguanine-DNA Methyltransferase and Mismatch Repair Proteins. *Clinical Cancer Research* **17**, 5748–5754 (2011).
201. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
202. Shi, W. *et al.* Reliability of Whole-Exome Sequencing for Assessing Intratumor Genetic Heterogeneity. *Cell Reports* **25**, 1446–1457 (2018).
203. Tensen, C. P., Quint, K. D. & Vermeer, M. H. Genetic and epigenetic insights into cutaneous T-cell lymphoma. *Blood* **139**, 15–33 (2022).
204. Sekulic, A. *et al.* Personalized treatment of Sézary syndrome by targeting a novel CTLA4:CD28 fusion. *Mol Genet Genomic Med* **3**, 130–136 (2015).
205. Mahmoud, M. *et al.* Structural variant calling: the long and the short of it. *Genome Biol* **20**, 246 (2019).
206. Sakamoto, Y., Zaha, S., Suzuki, Y., Seki, M. & Suzuki, A. Application of long-read sequencing to the detection of structural variants in human cancer genomes. *Comput Struct Biotechnol J* **19**, 4207–4216 (2021).
207. Hain, C., Stadler, R. & Kalinowski, J. Unraveling the Structural Variations of Early-Stage Mycosis Fungoides—CD3 Based Purification and Third Generation Sequencing as Novel Tools for the Genomic Landscape in CTCL. *Cancers* **14**, 4466 (2022).
208. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* **19**, 90 (2018).
209. Liu, Y. *et al.* DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* **22**, 295 (2021).

210. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer (IGV). *Cancer Res* **77**, e31–e34 (2017).
211. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683–691 (2011).
212. Sharp, N. Mutations matter even if proteins stay the same. *Nature* **606**, 657–659 (2022).
213. Nakato, R. & Sakata, T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* **187**, 44–53 (2021).
214. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nature protocols* **17**, 1518 (2022).
215. Geng, X. *et al.* GATA-3–dependent Gene Transcription is Impaired upon HDAC Inhibition. *Clinical Cancer Research* **30**, 1054–1066 (2024).
216. Kohnken, R. *et al.* Diminished microRNA-29b level is associated with BRD4-mediated activation of oncogenes in cutaneous T-cell lymphoma. *Blood* **131**, 771 (2017).
217. Qu, K. *et al.* Chromatin accessibility landscape of cutaneous T cell lymphoma and dynamic response to HDAC inhibitors. *Cancer cell* **32**, 27 (2017).
218. Lindahl, L. M. *et al.* Prognostic miRNA classifier in early-stage mycosis fungoides: development and validation in a Danish nationwide study. *Blood* **131**, 759–770 (2018).
219. Clark, R. A. Resident memory T cells in human health and disease. *Science translational medicine* **7**, 269rv1 (2015).
220. McGregor, J. M. *et al.* Spectrum of p53 Gene Mutations Suggests a Possible Role for Ultraviolet Radiation in the Pathogenesis of Advanced Cutaneous Lymphomas. *Journal of Investigative Dermatology* **112**, 317–321 (1999).

221. Wooler, G., Melchior, L., Ralfkiaer, E., Gjerdrum, L. M. R. & Gniadecki, R. TP53 Gene Status Affects Survival in Advanced Mycosis Fungoides. *Frontiers in Medicine* **3**, 51 (2016).
222. To, V. *et al.* CAR-T cell development for Cutaneous T cell Lymphoma: current limitations and potential treatment strategies. *Front. Immunol.* **13**, (2022).
223. He, X. *et al.* Non-coding RNAs in the spotlight of the pathogenesis, diagnosis, and therapy of cutaneous T cell lymphoma. *Cell Death Discov.* **10**, 1–15 (2024).

6 ANNEXE : Articles

Genomic profiling of mycosis fungoides identifies patients at high risk of disease progression

Léa Fléchon,^{1,*} Inès Arib,^{2,*} Ankit K. Dutta,^{3-5,*} Lama Hasan Bou Issa,¹ Romanos Sklavenitis-Pistofidis,³⁻⁵ Rémi Tilmont,² Chip Stewart,⁵ Romain Dubois,⁶ Stéphanie Poulain,^{1,7} Marie-Christine Copin,⁸ Sahir Javed,⁹ Morgane Nudel,² Doriane Cavalieri,² Guillaume Escure,² Nicolas Gower,² Paul Chauvet,² Nicolas Gazeau,² Cynthia Saade,² Marietou Binta Thiam,² Aïcha Ouelkite-Oumouchal,¹ Silvia Gaggero,¹ Émeline Cailliau,¹⁰ Sarah Faiz,¹¹ Olivier Carpentier,¹¹ Nicolas Duployez,^{1,7} Thierry Idziorek,¹ Laurent Mortier,^{11,12} Martin Figeac,¹³ Claude Preudhomme,^{1,7} Bruno Quesnel,^{1,2} Suman Mitra,¹ Franck Morschhauser,² Gad Getz,^{5,14,15} Irene M. Ghobrial,^{3,4,15} and Salomon Manier^{1,2}

¹Canther, ONCOLille, INSERM UMR-S1277, CNRS UMR9020, and ¹²OncoThai unit, INSERM UMR-S1189, Lille University, Lille, France; ²Department of Hematology, ⁶Department of Pathology, ⁷Department of Hematology, Biology and Pathology Center, ¹⁰Department of Biostatistics, and ¹¹Department of Pathology and Dermatology, Lille Hospital, Lille, France; ³Center for Prevention of Progression of Blood Cancers, Dana-Farber Cancer Institute, Boston, MA; ⁴Department of Medical Oncology, Harvard Medical School, Boston, MA; ⁵Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA; ⁸Department of Pathology, Angers University, Angers Hospital, INSERM, CRCI2NA, Angers, France; ⁹Department of Medical Oncology, Valenciennes Hospital, Valenciennes, France; ¹³Lille University, Lille Hospital, CNRS, INSERM, Institut Pasteur de Lille, US 41 – UAR 2014 - PLBS, Lille, France; ¹⁴Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA; and ¹⁵Harvard Medical School, Boston, MA

Key Points

- Genomic analysis of MF identifies alterations associated with high risk of progression and shorter overall survival.
- Clonal evolution of MF shows acquisition of *JUNB*, gain of 10p15.1 (*IL2RA/IL15RA*), or del12p13.1 (*CDKN1B*) at progression.

Mycosis fungoides (MF) is the most prevalent primary cutaneous T-cell lymphoma, with an indolent or aggressive course and poor survival. The pathogenesis of MF remains unclear, and prognostic factors in the early stages are not well established. Here, we characterized the most recurrent genomic alterations using whole-exome sequencing of 67 samples from 48 patients from Lille University Hospital (France), including 18 sequential samples drawn across stages of the malignancy. Genomic data were analyzed on the Broad Institute's Terra bioinformatics platform. We found that gain7q, gain10p15.1 (*IL2RA* and *IL15RA*), del10p11.22 (*ZEB1*), or mutations in *JUNB* and *TET2* are associated with high-risk disease stages. Furthermore, gain7q, gain10p15.1 (*IL2RA* and *IL15RA*), del10p11.22 (*ZEB1*), and del6q16.3 (*TNFAIP3*) are coupled with shorter survival. Del6q16.3 (*TNFAIP3*) was a risk factor for progression in patients at low risk. By analyzing the clonal heterogeneity and the clonal evolution of the cohort, we defined different phylogenetic pathways of the disease with acquisition of *JUNB*, gain10p15.1 (*IL2RA* and *IL15RA*), or del12p13.1 (*CDKN1B*) at progression. These results establish the genomics and clonality of MF and identify potential patients at risk of progression, independent of their clinical stage.

Introduction

Primary cutaneous T-cell lymphoma (CTCL) is a clinically heterogeneous group of incurable extranodal lymphomas that target skin-resident mature T cells. Mycosis fungoides (MF) is the most common CTCL, accounting for >50% of all cases.¹ It typically exhibits an indolent disease course with slow progression

Submitted 12 December 2023; accepted 17 February 2024; prepublished online on *Blood Advances* First Edition 21 March 2024; final version published online 13 June 2024. <https://doi.org/10.1182/bloodadvances.2023012125>.

*L.F., I.A., and A.K.D. contributed equally to this study.

The WES samples used in this study were deposited in the SRA database (<https://www.ncbi.nlm.nih.gov/sra>; accession number: PRJNA925900).

Deidentified individual participant data are included as supplements in the online version of this article. Other data are available upon request from the corresponding author, Salomon Manier (salomon.maniar@chu-lille.fr).

The full-text version of this article contains a data supplement.

© 2024 by The American Society of Hematology. Licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

over several years, and patients present with a wide range of clinical symptoms and disease outcomes.² In the early stages, MF generally manifests as erythematous macules and plaques, which are nonspecific lesions that are difficult to diagnose. Patients may progress from low-risk (LR) clinical stages (tumor-node-metastasis-blood [TNMB] classification system IA to IIA) to high-risk (HR) stages (TNMB IIB to IVB), which involve tumors or generalized erythroderma.^{2,3} At the cellular level, ~20% of patients with HR show histological transformation with transformed MF cells, a feature associated with poor prognosis.⁴⁻⁶ Transformation is defined by the presence of >25% large cells (immunoblasts, large pleomorphic cells, or large anaplastic cells), which may or may not express CD30 within the infiltrate of the MF lesion.⁵ At the molecular level, MF is genetically heterogeneous, with no uniform single nucleotide variants (SNVs) and somatic copy number alterations (SCNAs). Previous microarray gene expression profiling studies of CTCLs, particularly MF, revealed deregulated expression of *TP53*, *PLCG1*,^{7,8} and *TNFR2*⁹ and deletions of the JAK-STAT signaling inhibitors *SOCS1* and *HNRNPK*.¹⁰ Upon histological transformation, loss of chromosomal region 9p21.3, in which the tumor suppressors *CDKN2A* and *CDKN2B* reside, is commonly observed.¹⁰⁻¹³

Nevertheless, the molecular underpinnings of disease progression in patients with MF have not been extensively studied; thus, clinicians rely on suboptimal clinical variables for risk stratification. To identify the molecular drivers of disease progression and aggressiveness, which may help improve clinical practice, we collected samples from patients with MF at different stages of the disease and performed deep whole-exome sequencing (WES) for the detection of somatic events. Specifically, we sequenced DNA from 67 skin samples obtained from 48 patients, including 18 sequential samples from patients who exhibited progression from LR to HR stages. This approach allowed for us to characterize the genomic landscape of MF across the stages of progression, identify putative driver genes contributing to the progression of patients with LR, and study the clonal evolution of tumor cells during MF progression. Overall, this study provides new insights into the genetic risk factors of disease progression in patients with MF, which may help improve clinical prognostication models and lead to the discovery of novel therapeutic approaches for patients with MF. Although these findings identify patients with a HR of progression, further validation in independent cohorts is needed to confirm their clinical utility.

Methods

Patient samples

We studied 67 tumor samples from a cohort of 48 patients diagnosed with MF at the Lille University Hospital between 2003 and 2017. All cases were reviewed by local specialists from the French Study Group on Cutaneous Lymphomas Network. Immunohistochemistry was performed using the following panel: CD20, PAX5, CD2, CD3, CD4, CD5, CD7, CD8, CD30, PD1, and Ki-67. Another immunohistochemistry analysis of CD25 was performed in patient samples with or without gain10p15.1. The diagnosis of MF was confirmed by polymerase chain reaction to detect clonal recombination of the T-cell receptor gene. We defined 2 prognostic groups: the LR group with TNMB stages IA, IB, and IIA and the HR group with either high TNMB stage (IIB, III, and IV) or transformed histology. Samples were extracted from 67 frozen skin

biopsy samples stored in the Biology-Pathology-Genetics unit's tumor bank (certification NF 96900-2014/65453-1). After macrodissection of the tumors, the percentage of tumor cells was visually estimated by microscopic observation. In total, the cohort included 67 tumor samples from 48 patients. For 13 of these patients, sequential samples were collected at different stages of disease progression. This study was approved by the Institutional Review Board of Lille University Hospital and Nord Ouest IV (protocol number: ECH18/03) in accordance with the Declaration of Helsinki, and all patients provided written informed consent.

DNA quality control and WES

The amount of extracted DNA was determined by spectrophotometry using the "Quant-iT Picogreen dsDNA Assay" kit (Thermo Fisher Scientific). For WES, 100 ng of genomic DNA was fragmented on the Covaris ultrasonicator to target a base pair peak of ~150 to 200 bp. DNA libraries were prepared using the Agilent SureSelect XT low input kit with target region capture using Agilent SureSelect Human All Exon V7 (Agilent Technologies) and the Illumina Dual-Index Adapter primer kit (Illumina). Libraries were generated in an automated manner using the Bravo NGS liquid handling robot in accordance with the supplier's recommendations. The final libraries were assessed using the Agilent High Sensitivity DNA Analysis kit on the Bioanalyzer 2100 and quantified by quantitative polymerase chain reaction using the Kapa Library Quantification kit (Roche). Libraries were pooled and sequenced using 2 × 100 bp paired-end reads on an Illumina NovaSeq 6000 platform on a S4 flowcell.

Alignment and quality control

Data were analyzed on the Terra computing platform (Broad Institute [BI], Cambridge, MA), which aggregates bioinformatics tools for genomic data analysis. The quality of the raw sequencing output in fastq format was obtained using FastQC¹⁴ software and visualized at the whole-cohort level with MultiQC.¹⁵ Illumina primers, very small reads (< 30 bp), and poor quality reads were removed with CutAdapt.¹⁶ Sequences were aligned with BWA-MEM,¹⁷ and alignment quality was estimated using Picard software suite tools. Postalignment cleanup consisted of the removal of duplicate reads with MarkDuplicate¹⁸ (Picard, BI), local realignment around insertion-deletions (InDels) with IndelRealigner, and base quality recalibration with BaseRecalibrator and ApplyBQSR.¹⁹ During sample preparation in the laboratory, there were risks of contamination or inversion, which were checked with CalculateContamination²⁰ and CrossCheckLaneFingerprints,¹⁸ respectively, and no errors were identified. Finally, the potential oxidation of guanine to 8-oxoguanine artifacts that can occur during the preparation of genomic libraries under the combined effect of heat, DNA cutting, and the introduction of metal contaminants were removed with CollectOxoGMetrics.¹⁸ Overall, tumor samples meeting all quality control cutoffs had an average coverage of 99.72% on the GRCh37 assembly, with a mean target depth of coverage of 231.85x.

Copy number analysis from WES data

SCNAs and genome-wide allelic variations were detected using ModelSegments²⁰ software. To summarize, the software will use the panel of normal (PoN) of 13 samples to detect copy and allele number variations in each tumor sample and model several

segments. The software was run in Tumor-Only mode for the entire set of samples and run again in “Matched-Normal” mode for the 3 tumor samples that had an associated normal sample. Frequent large and focal SCNAs in the cohort were highlighted by GIS-TIC2.0²¹ using a q-value threshold of 0.01.

Mutation calling of recurrently mutated genes

SNVs and InDels were detected using a Bioinformatics Cancer Genome Analysis (CGA) pipeline called “CGA WES Characterization Pipeline.” SNVs were detected by MuTect²² and InDels by Strelka²³ and Mutect2²⁴. The results of these 3 software programs were then filtered using MAFFPoNFilter²⁵. The latter uses 2 PoNs to segregate somatic from germ line variants: the 1 from our cohort (from 13 WES normal samples) and the controlled access PoN used in the routine analysis of the BI (8334 normal samples from The Cancer Genome Atlas Program database). Somatic variants (SNVs and InDels) were annotated for their oncogenic effect using Variant Effect Predictor²⁶ and Oncotator.²⁷ They were then validated using the MutationValidator tool, which establishes the minimum number of reads carrying the variant for it to be considered somatic. Genes more frequently mutated than chance were determined using MutSig2CV²⁵. Variants carried by genes known to be “fishy genes,” that is, known false-positive genes that are not plausible in the development of cancers, were manually removed according to the list established by Lawrence et al in 2013.²⁸

Panel of normal

In addition to the 3 normal samples from our cohort, 10 additional human blood samples were provided by the Research Blood Component (Watertown, MA). These 13 samples formed a PoN to filter out possible errors during the preparation or sequencing steps, as well as germ line genetic/genomic events. As a final criterion to filter potential germ line variants, we used bash scripts to count the occurrence of mutated variants in our cohort of PoN samples in the fastq format. First, we converted the maf file into a 2-motive list per variant, consisting of 11 nucleotides before and after the variant. The first motive represented the wildtype, corresponding to the reference variant, whereas the second one was the mutated form, representing the alternative variant. Subsequently, we used the “do_it.sh” bash script from the GitHub repository mafouille/HotCount to count the number of occurrences of both wildtype and mutated motives in the PoN. Variants with mutated patterns that appeared more than once in the PoN were excluded. This additional check not only reveals supplementary potential artifacts but also modulates the stringency of the filtering process, complementing the PoN in the pipeline.

Estimation of purity, ploidy, and CCF

ABSOLUTE²⁹ software determined the purity and ploidy of each sample from the SNVs and SCNAs (supplemental Table 10) and thus served us to define the cancer cell fraction (CCF) of each genomic alteration. For tumor samples without associated normal samples, we applied a tumor-only Germline Somatic Log odds filter, as described by Chapuy et al in 2018.³⁰ To summarize, for each variant, its CCF, sample ploidy and purity, and local copy number (CN) were used to calculate the probability that the allelic fraction of the variant was consistent with a modeled allelic fraction for either a germ line or somatic hypothetical event. Thus, the filter set a threshold for each sample to remove additional germ line

variants. To eliminate additional germ line events, we set the tumor-only threshold to -1 .

GnomAD filtering

We used the genome aggregation database gnomAD to exclude potential germ line mutations. The Human Genome Variation Society genomic identifier (HGVS_genomic_change column in maf file) was used in the VEP GRCH37 Ensembl database to obtain gnomAD allele frequency. A cutoff value of 0.0001 was applied.

Statistical analysis

Time-to-event end points were estimated using the Kaplan-Meier method. Differences in the survival curves were assessed using the log-rank test. The median follow-up was calculated using the reverse Kaplan-Meier method. Time to progression was measured from the date of diagnosis to the date of documented progression to HR. Cox proportional hazards modeling was performed to assess the impact of genetic alterations on the risk of disease progression. Violin plot conditions were compared using the nonparametric Wilcoxon test. Figures and statistical estimations were obtained using R v.3.6.3 and MATLAB. The maf files were analyzed using the R package “maftools” v2.12.05 with oncoplot (waterfall plot, Figure 1A), maf_compare (forest plot, Figure 2A), and somaticInteractions (corplot, Figure 3B) functions. The maf_compare function performs a Fisher test on all genes between the LR and HR cohorts to detect differentially mutated SNVs/SCNAs. In the generated forest plot, events with a low frequency of less than ~20% (ie, 9 patients) were not analyzed (Figure 2A). Other figures were generated using the R packages v3.4.0 “survival” v3.4-0 and “survminer” v0.4.9. (Kaplan-Meier graphs, Figure 2), and “chrismiller/fishplot” v0.5.1 (Fishplots, Figure 4) and ggplot2.

This study was approved by the Institutional Review Board of Lille University Hospital and Nord Ouest IV (protocol number: ECH18/03) in accordance with the Declaration of Helsinki, and all patients provided written informed consent.

Results

Genomic landscape of MF

The median age in our cohort was 62.5 years (range, 19-94), with a 60% male predominance, which is consistent with the known distribution of MF in the population.² Patients were stratified based on the TNMB classification into LR (stage IA, IB, or IIA) and HR (stages IIB to IVB) (supplemental Figure 1A-D; supplemental Tables 1 and 7). We detected mutations by WES and analyzed 67 tumor samples from 48 patients, using a validated pipeline to filter germ line variants and artifacts from tumor-only samples³⁰ (Methods; supplemental Figures 1 and 2). Thirteen patients had sequential sampling (supplemental Figure 1C-D).

We found a median of 3.5 mutations per Mb, corresponding to a median of 135 SNVs or insertion-deletions (InDels) per sample. The most recurrently mutated genes comprise previously reported mutational drivers in TCL, and 51% of patients had a mutation in at least 1 of these drivers (Figure 1A). These included the T-cell differentiation transcription factor *JUNB* (p.A282V; Figure 1B) in 13% of the cases³¹⁻³⁴; the epigenetic factor *TET2* in 9% of the cases³⁵⁻³⁸; the component of the MAPK pathway *MAPK1* in 6% of the cases (Figure 1B); the transcription factor *FOXA1* involved in

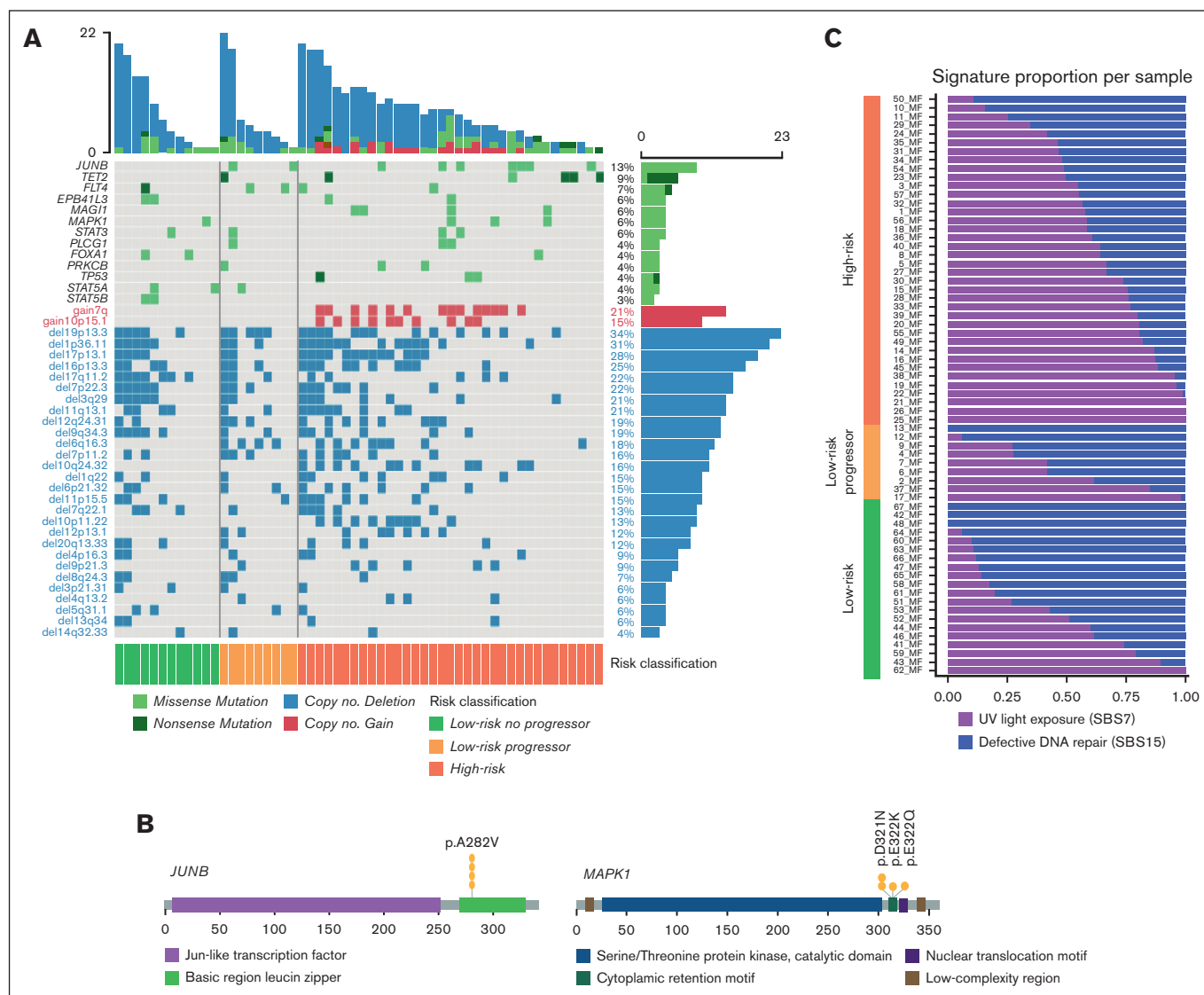


Figure 1. Landscape of most recurrent genomic alterations in MF. (A) Landscape of genomic alterations in 67 tumor samples of MF divided into LR and HR disease based on the TNMB stages. Alterations are divided into SNVs and SCNAs. (B) Driver oncogene maps of *JUNB* and *MAPK1*. (C) Relative enrichment of signature activities per samples divided into LR, LR progressors, and HR.

interferon signaling and immune response suppression^{37,39}; the tyrosine kinase receptor *FLT4*; the phospholipase *PLCG1* involved in NF- κ B and NFAT signaling^{7,40}; the JAK/STAT signaling pathway factors *STAT3*, *STAT5A*, and *STAT5B*; and *TP53* (Figure 1A; supplemental Figure 4; supplemental Table 2).

Next, we identified significantly recurrent SCNAs by using GIS-TIC2.0.²¹ Overall, SCNAs were the most common genomic alterations and were present in 84% of cases (Figure 1A). Specifically, we detected significantly recurrent alterations, including 8 arm-level and 29 focal copy number losses, as well as 2 arm-level and 1 focal copy gains (q-value ≤ 0.1 ; supplemental Figures 3A-B; supplemental Table 11). The frequencies of these SCNAs ranged from 6% to 57%, and the number of genes in the focal peaks varied from 1 (*MUC12* in del7q22.1) to 576 (del6p21.33) (supplemental Tables 12 and 13). In the focal CN gain *10p15.1* (26

genes; frequency of 13%) resides interleukin-2 receptor alpha and interleukin-15 receptor alpha (*IL2RA* and *IL15RA*) as well as the NF- κ B pathway protein kinase *PRKCB*. In particular, *IL2RA* and *IL15RA* play crucial roles in the phosphorylation of *STAT3* and *STAT5* in the JAK-STAT pathway, likely contributing to the pathogenesis of MF.⁴¹⁻⁴³ Among the tumor suppressor genes affected by the most frequent CN loss were *TMEM259* (19p13.3, frequency 57%), *TP53* (17p13.1, 41%), *SUZ12* and *NF1* (17q11.2, 33%), *NOTCH1* (9q34.3, 26%), *CARD11* (7p22.3, 24%), *ZEB1* (10p11.22, 17%), *TNFAIP3* (6q16.3, 15%), *CDKN2A* (9p21.3, 11%), and *CDKN1B* (12p12.2, 9%).

Mutational processes produce distinctive footprints called mutational signatures in the cancer genome that capture both DNA damages and repair mechanisms. We applied SignatureAnalyzer,⁴⁴ a tool that uses both the 3-bases mutational sequence

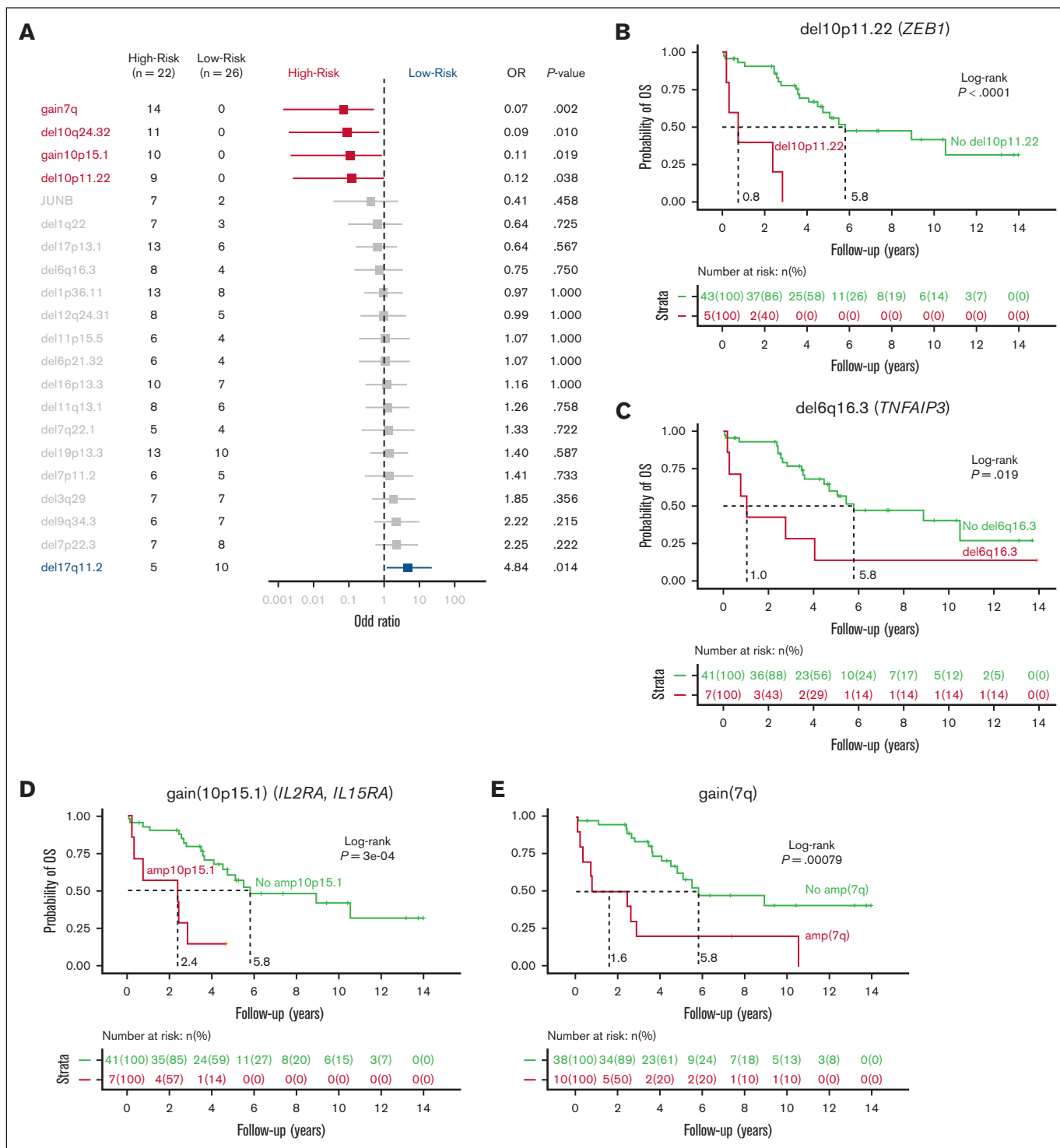


Figure 2. Correlation of genomic events to the disease stage. (A) Forest plot showing the association between individual genes alteration and clinical stage of MF divided into LR and HR, as depicted by OR. (B-E) Kaplan-Meier plots of individual genetic factors predictive of OS in univariate and multivariate models of 48 patients with a newly diagnosed MF: del10p11.22 (B); gain of 10p15.1 (C); gain of 7q (D); and del6q16.3 (E). (F) Kaplan-Meier curves for analysis of time to progression in patients with LR disease. P values were derived from log-rank test. (G) CD25 immunohistochemistry at diagnosis of MF skin biopsies in patients with gain of 10p15.1 (top panels, MF sample of patient 18 and patient 14 with presence of transformed MF cells) or without gain of 10p15.1 (bottom panels, MF sample of patient 23 and patient 16). Scale bars indicate 150 μ m.

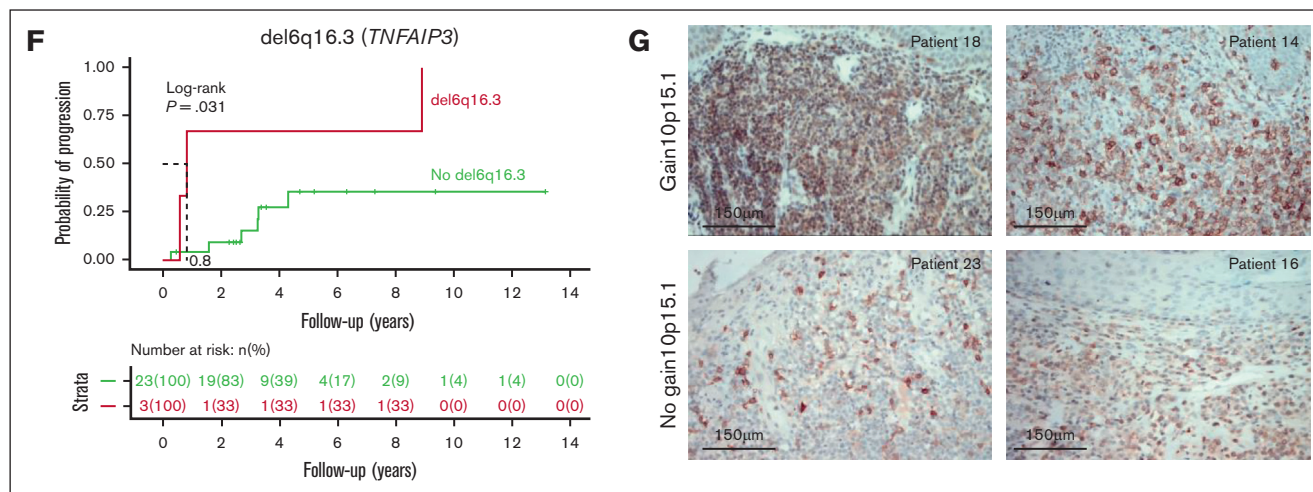


Figure 2 (continued)

context and the clustering of the mutation in the genome to define specific COSMIC (Catalogue of Somatic Mutations in Cancer database) single base substitution (SBS) signatures. Here, we detected 2 primary mutational signatures: the UV light exposure signature SBS7 and the defective DNA repair signature SBS15 (supplemental Figures 5A-B and E; supplemental Table 8). The age-related deamination signature, SBS1, was also frequently observed in our cohort but was not sufficiently prevalent to be separated from the UV signature. The enrichment of the SBS7 signature is consistent with previous studies on CTCL^{32,45,46} and was also significantly associated with HR, suggesting a role of UV radiation in the progression of MF. It is also possible that HR tumors are phylogenetically older and have more time to accumulate drivers and UV signature (Figure 1C). Additionally, there was an association between the total number of mutations and SBS7 signature (Wilcoxon test; $P = 5.735 \times 10^{-12}$). Signature SBS15 is 1 of the 7 mutational signatures related to defective DNA mismatch repair and microsatellite instability that are found in different cancer types. The contribution of SBS15 was similar in the LR and HR samples, suggesting a more founder event (supplemental Figure 5G). Next, we determined the relative contribution of SBS7 and SBS15 signatures to the mutational burden of driver genes (supplemental Figure 5E). Five genes were enriched with mutations associated with SBS15, including *JUNB*, *STAT3*, *MAML2*, *FLT4*, and *FOXA1*, whereas 9 genes were predominantly associated with SBS7, including *TET2*, *MAPK1*, *TP53*, *EPB41L3*, *MAGI1*, *PRKCB*, *HCK*, *L3MBTL4*, and *MECOM*.

Association of genetic features to disease stage and outcome

We further compared the genomic profiles of patients with LR disease with those of patients with HR disease (supplemental Tables 3-6 and 9). All CN gain7q (odds ratio [OR], 0.07; $P = .002$) and gain10p15.1 (*IL2RA* and *IL15RA*; OR, 0.11; $P = .019$) were found in patients at HR, as were del10q24.32 (*NFKB2*; OR, 0.09; $P = .010$) and del10p11.22 (*ZEB1*; OR, 0.12; $P = .038$; Figure 2A). Conversely, del17q11.2 (*SUZ12* and *NF1*) was significantly associated with LR (OR, 4.84; $P = .014$). We observed a higher mutation rate in HR samples, with a median of

3.95 mutations per Mb compared with 3.02 mutations per Mb in LR samples ($P = .019$) (supplemental Figure 5F). The increased mutation rate in HR was predominantly associated with the UV light exposure signature SBS7, suggesting either its role in disease progression or else reflecting time and addition of drivers (Figure 1C; supplemental Figure 5G).

Next, we assessed the impact of putative driver mutations at diagnosis on overall survival (OS) in patients with a median follow-up time of 4.3 years (range, 0.3-14). In total, 4 SCNAs were significantly associated with shorter OS, (Figure 2B-E): del10p11.22 (*ZEB1*; median OS [mOS] of 2.4 vs 8.9 years; $P = .00029$); gain10p15.1 (*IL2RA* and *IL15RA*; mOS of 2.4 vs 5.8 years; $P = .00029$); gain7q arm (mOS of 2.6 vs 8.9 years; $P = .011$); and del6q16.3 (*TNFAIP3*; mOS, 1.9 vs 5.8 years; $P = .021$). Furthermore, the presence of del10p11.22, gain10p15.1, and del6q16.3 remained a significant risk factor in a multivariate stepwise analysis, accounting for the patient's clinical stage (supplemental Figures 5C-D). Notably, no SNVs were shown to affect OS. In the LR group, 7 patients progressed, and 19 did not progress during follow-up. Patients with LR disease and del6q16.3 (*TNFAIP3*) had shorter time to progression (median, 0.8 years vs not reached; $P = .0031$; Figure 2F). This alteration is present in 15% of patients with LR at diagnosis and could serve as a prognostic marker in clinical practice for future publication with a validation cohort.

Regarding gain10p15.1, we observed a higher expression of CD25 by immunohistochemistry in samples from patients with gain10p15.1 than those of patients without gain10p15.1. This is consistent with a link between gain of *IL2RA* and the surface expression of CD25, suggesting a role of *IL2/IL2RA* signaling in the disease progression (Figure 2G).

The clonal architecture and phylogeny of MF

Next, we estimated the CCF for each putative driver and determined whether the alterations were clonal (≥ 0.9) or subclonal (< 0.9). We observed heterogeneity in the clonality of genomic alterations in patients with MF. Certain alterations were frequently clonal, such as del9p21.3 (*CDKN2A*), del10p11.22 (*ZEB1*), gain7q, and mutations in *PRKCB* or *STAT5B*. Other alterations

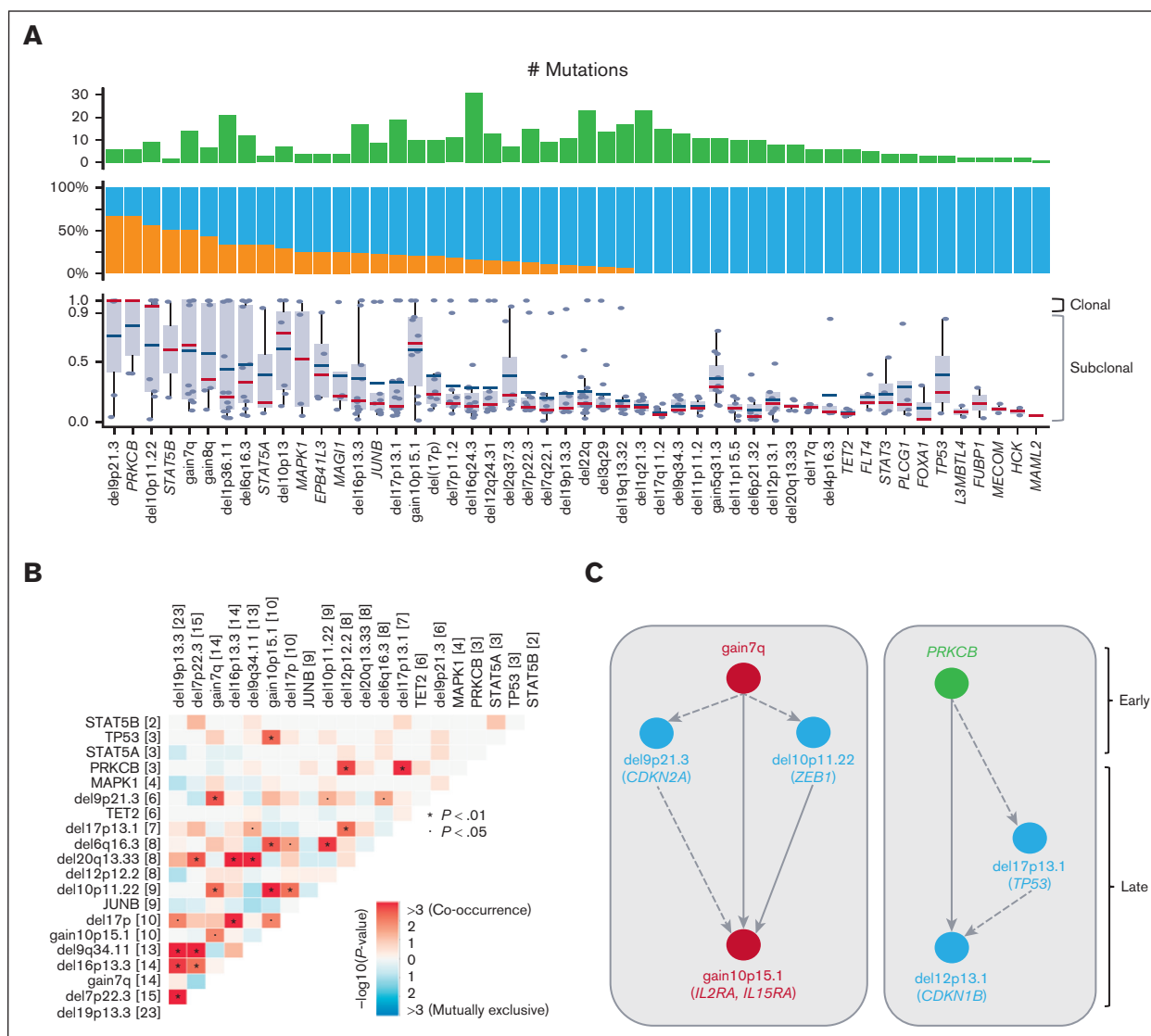


Figure 3. Clonal heterogeneity and inferred timing of genetic drivers. (A) Proportion in which recurrent drivers are found as clonal or subclonal across the 67 samples (top), along with the individual CCF values for each sample affected by a driver (bottom). Median CCF values are shown (bottom, bars represent the median and interquartile range for each driver). (B) Correlation matrix for the most recurrent genomic alterations for the 48 patients with MF, with Fisher exact test to detect such significant pair of mutations. (C) Timing of genomic alterations with early events at top and late events at bottom. Color indicates alteration types. Arrows between 2 alterations were drawn when 2 drivers were found in 1 sample with an excess of clonal to subclonal events. Dashed arrows indicate 1 clonal-subclonal pair and solid arrows indicate ≥ 2 clonal-subclonal pairs.

were more frequently subclonal, such as mutations in *TET2*, *TP53*, *PLCG1*, or *del17p*, which likely represent later events in the disease course (Figure 3A).

Subsequently, we analyzed the co-occurrence of driver genes and SCNAs. We found that *del9p21.3* (*CDKN2A*) significantly co-occurred with *gain7q* and *del10p11.22* (*ZEB1*). *Gain10p15.1* (*IL2RA* and *IL15RA*) was significantly cosegregated with *del6q16.3* (*TNFAIP3*) and *TP53* mutations. *Del16p13.3* (*CREBBP*) co-occurred with *del17p* and *del19p13.3* (*TMEM259*; Figure 3B). We applied a mutation-ordering method to samples with pairs of clonal and subclonal alterations.⁴⁷ Given that clonal mutations occur before subclonal events, we defined the timing of the main genetic alterations. We observed 2 general patterns

defining the phylogeny of the disease (Figure 3C): first, clonal *gain7q* that often co-occurs with clonal *del9p21.3* (*CDKN2A*) or *del10p11.22* (*ZEB1*), with further acquisition of *gain10p15.1* (*IL2RA* and *IL15RA*) at progression; second, the clonal mutation of the tyrosine kinase *PRKCB* followed by *del17p13.1* (*TP53*) and *del12p12.1* (*CDKN1B*). Of note, *gain7q* and *del10p11.22* (*ZEB1*) are often clonal and associated with an unfavorable outcome suggesting that the molecular path to aggressiveness is made at an early stage of the disease. *Gain10p15.1* (*IL2RA* and *IL15RA*), which is also associated with HR, represents a transforming molecular event in the disease course.

We further analyzed sequential samples from 7 patients who progressed from early stages (LR progressors) to advanced stages of

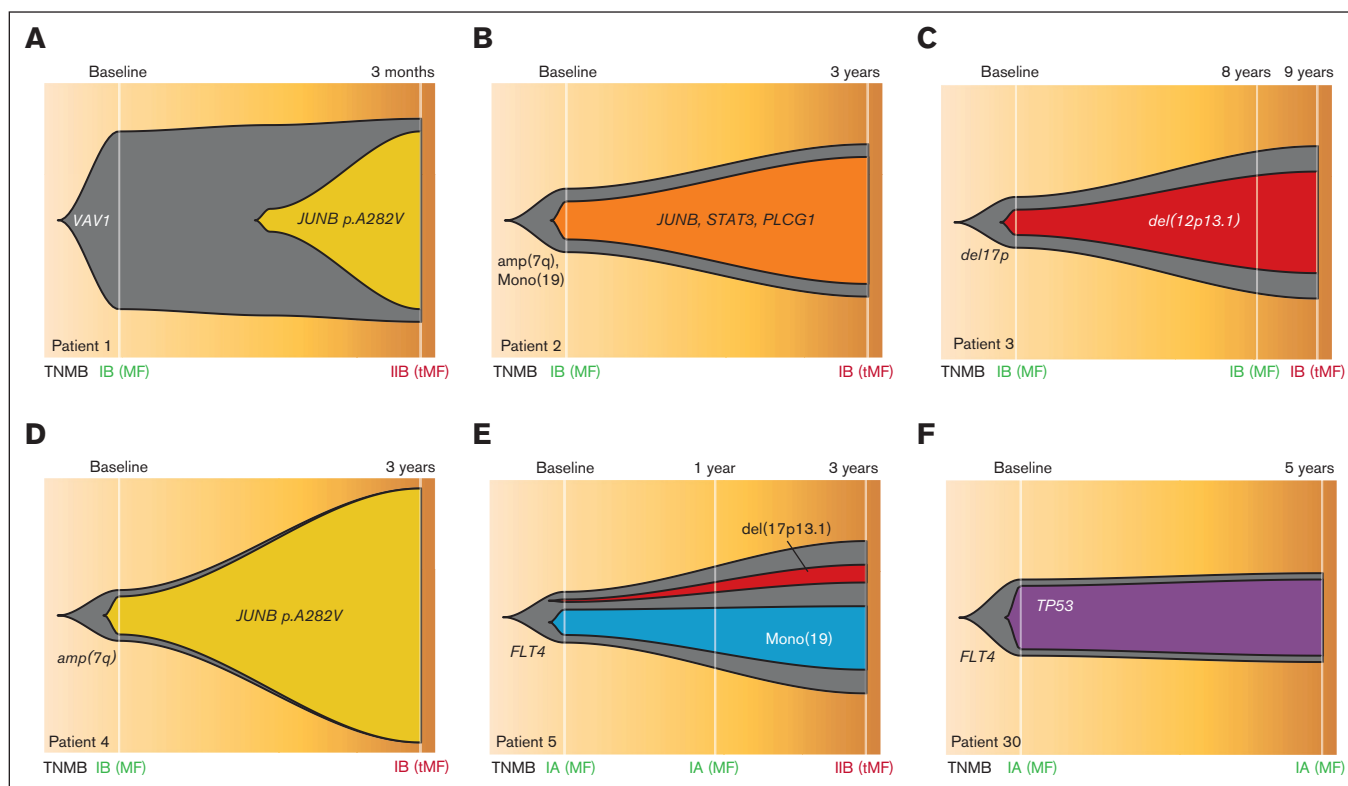


Figure 4. Clonal evolution of sequential samples. (A-E) Fish plots of 5 serial cases of MF with samples at LR stages who progressed to HR stages. (F) One serial case with both time points at LR stage because the patient did not progress. TNMB classification has been indicated (green for LR and dark red for HR samples). tMF, transformed MF.

the disease, with sampling intervals ranging from 3 months to 9 years. We observed evidence of clonal heterogeneity at diagnosis in all 7 cases, indicating that clonal branching had already occurred at the early stage of the disease. Moreover, we noticed cases in which subclones that were initially small at diagnosis exhibited substantial expansion at a later time point, indicating a strong driver potential for their mutations/SCNAs. Specifically, this was observed in 3 patients with small subclones harboring the *JUNB* mutation or *del12p13.1* (*CDKN1B*), which were selected for and became dominant at disease progression (Figure 4B-D). In these cases, all alterations detected in late-stage samples were already present at the baseline, pointing to a linear pattern of clonal evolution. In another instance, however, a *JUNB*-mutant subclone was first identified at the time of disease progression, suggesting that further branching had occurred and drove the progression from LR to HR disease (Figure 4A). In this case, the *JUNB* mutation was acquired on top of a clonal *VAV1* mutation, involved in T-cell receptor signaling, and has been reported in various T-cell malignancies (Figure 5). Clonal evolution was also observed during treatment, with the expansion of a resistant *del17p13.1* subclone in a patient receiving multiple lines of treatment within 3 years of disease progression (Figure 4E). Finally, in a patient who did not progress and whose 2 samples were sequenced at the LR stage but 5 years apart, no new alterations or increases in the CCF of baseline mutations were noticed at the later time point, implying that clonal heterogeneity at baseline may not be sufficient for disease progression (Figure 4F).

Discussion

In this study, we leveraged WES data to analyze tumor samples from 48 patients with newly diagnosed MF, together with sequential samples at later time points from 13 patients. We analyzed the most recurrent genomic alterations, as well as their clonal heterogeneity and clonal evolution, to temporally order these alterations and gain insight into the phylogeny of MF. Our results highlight the complexity of MF, with a median of 135 different genomic alterations per tumor. The most recurrent alterations in MF were consistent with previous observations in TCL in general^{9,32,36,38,48,49} (MF, Sézary syndrome, primary cutaneous CD30⁺ T-cell lymphoproliferative disorders, and primary cutaneous $\gamma\delta$ T-cell lymphoma), with alterations in the NF- κ B pathway (such as deletion *NFKB2* in 10q24.32, gain of *PRKCO* in 10p15.1, mutation of *PRCKB*, and deletion of *CARD11* in 7p22.3), JAK/STAT pathway (eg, gain of *IL2RA* and *IL15RA* in 10p15.1, and mutations in *STAT3*, *STAT5A* and *STAT5B*), MAPK pathway (mutations in *JUNB* and *MAPK1*), cell cycle pathway (such as deletion *CDKN1B* in 12p13.1 and *CDKN2A* in 9p21.3), and inhibition of apoptosis (deletion and mutations of *TP53* and deletion of *TNFAIP3* in 6p16.3; Figure 5).

We identified genomic alterations that are associated with HR stages of the disease: gain7q and gain10p15.1 (*IL2RA* and *IL15RA*), *del10q24.32* (*NFKB2*), and *del10p11.22* (*ZEB1*), suggesting their role in disease progression. Conversely, *del17q11.2* (*SUZ12* and *NF1*) was associated with the LR stages of the disease. The mutational signature SBS7, associated with UV light, was

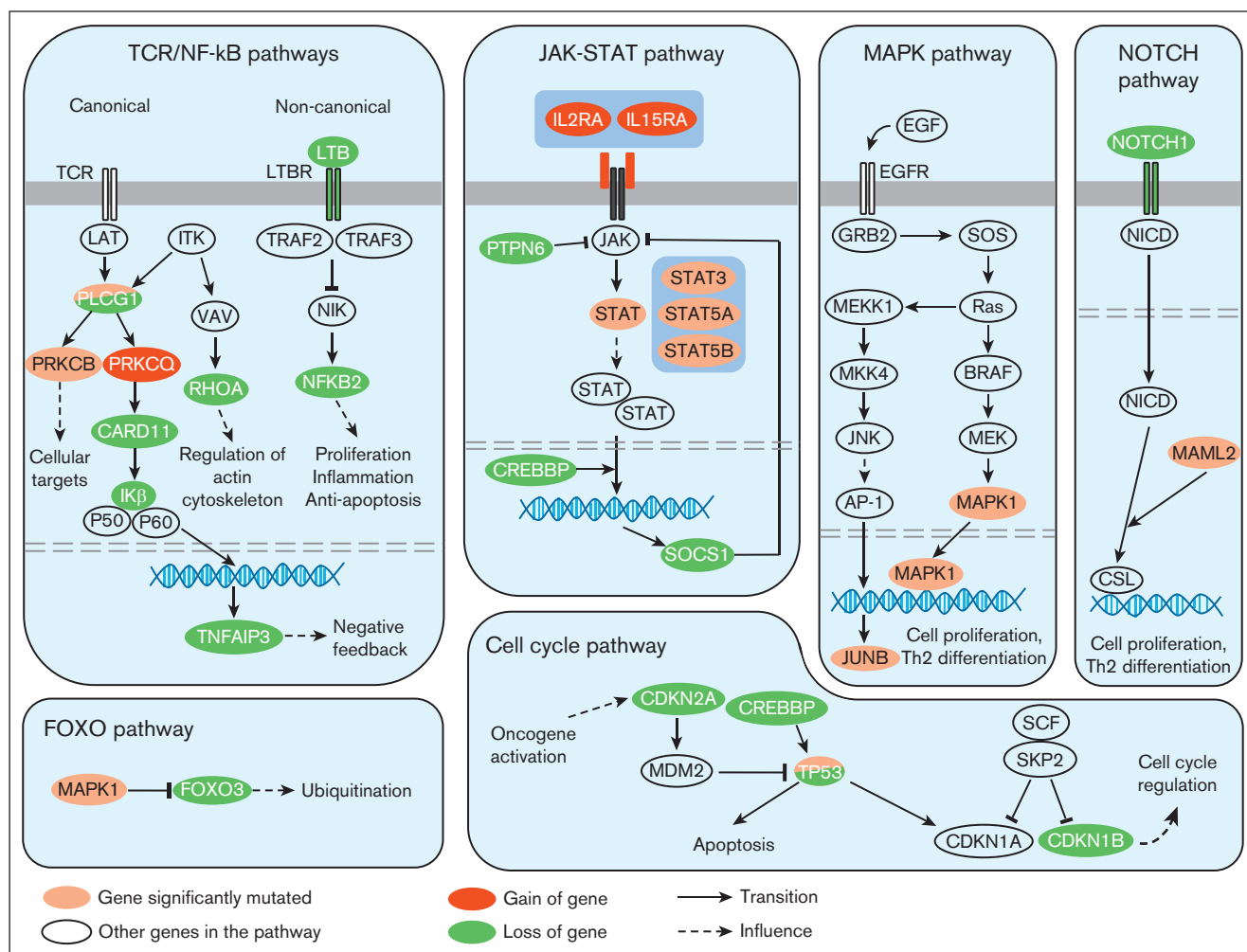


Figure 5. Recurrent genetic alterations affecting key signaling pathways involved in MF. In this diagram, frequently mutated genes with well-established roles in these signaling pathways have been depicted representing the proteins they encode. TCR, T-Cell Receptor.

also associated with HR stages and was enriched in the mutational process of genes such as *TET2*, *MAPK1*, *TP53*, and *PRKCB*. We also describe the different patterns of phylogeny of MF, with early events such as gain7q or mutations in *PRKCB* and further acquisition of alterations such as gain10p15.1 (*IL2RA* and *IL15RA*), del17p13.1 (*TP53*), or del12p13.1 (*CDKN1B*). This clonal evolution from the LR to HR stages was either linear or branched, with the selection of subclones with strong driving potential.

We also assessed the prognostic value of these genomic alterations during diagnosis. We found 4 SCNAs that were significantly associated with a shorter OS: del10p11.22 (*ZEB1*), gain10p15.1 (*IL2RA*, *IL15RA*), gain7q, and del6q16.3 (*TNFAIP3*). The identification of these HR genomic alterations can help in the clinical management of MF. Early focal lesions of MF are sometimes difficult to diagnose and histologically differentiate from eczematous or psoriasis dermatitis.⁵⁰ Specific genomic alterations associated with adverse outcomes can help to guide the diagnosis of MF. Our data illustrate the role of gain10p15.1 (*IL2RA* and *IL15RA*) in disease progression. Taken together, these genomic alterations represent potential early indicators that may be useful for MF prognostication but require validation in future studies. This can be therapeutically relevant, because

cytokine inhibitors such as bnz-1 are currently being tested in clinical trials.

In conclusion, sequencing a large cohort of patients with MF, including sequential samples, has allowed for us to understand the genomic complexity of MF, temporal ordering of genetic events, and biomarkers that are associated with HR and disease progression. We believe that introducing NGS evaluation at the time of MF diagnosis can improve the identification of patients at a HR of disease progression and their clinical management.

Acknowledgments

This study was supported by a grant from the GEFLUC Flandres Artois Foundation to (S. Manier). The authors acknowledge the Tumorotheque of CHU Lille, the Société Française d'Hématologie, and the I-SITE ULNE foundation for their assistance with this project.

Authorship

Contribution: L.F., I.A., and S. Manier contributed in conceptualization, methodology, and writing including original draft and

visualization; L.F., R.S.-P., C. Stewart and M.F. contributed in software; L.F., I.A., A.K.D., R.S.-P., C. Stewart, E.C., M.F., S. Mitra, G.G., I.M.G., and S. Manier contributed in formal analysis; L.F., I.A., A.K.D., L.H.B.I., and S. Manier contributed in investigation; I.A., R.D., S.P., M.-C.C., S.J., M.N., A.O.-O., S.F., O.C., L.M., M.F., T.I., F.M., G.G., I.M.G., and S. Manier contributed in resources; R.D. contributed in validation; L.F. and S. Manier contributed in data curation; all authors contributed in reviewing and editing; G.G., I.M.G., and S. Manier supervised the research; and S. Manier contributed in funding acquisition.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: L.F., [0000-0003-4140-475X](#); I.A., [0009-0000-7062-9814](#); A.K.D., [0000-0003-2535-3729](#); L.H.B.I., [0000-0002-2561-9980](#); R.S.-P., [0000-0002-8025-9850](#); R.T., [0000-0003-](#)

[1739-2162](#); C.S., [0000-0003-2245-9552](#); R.D., [0000-0001-6990-7354](#); S.P., [0000-0002-1792-7270](#); M.-C.C., [0000-0002-4729-7054](#); S.J., [0000-0002-1649-9577](#); M.N., [0000-0001-9100-2316](#); D.C., [0000-0002-4836-7237](#); G.E., [0000-0002-3151-9441](#); P.C., [0000-0001-6326-9283](#); N.G., [0000-0003-3449-6676](#); M.B.T., [0000-0002-1387-0510](#); A.O.-O., [0000-0002-7048-5555](#); S.G., [0000-0002-2685-1108](#); É.C., [0000-0001-5343-6471](#); O.C., [0000-0002-7788-6132](#); N.D., [0000-0002-3927-1022](#); T.I., [0000-0001-5893-9131](#); L.M., [0000-0002-2644-1790](#); M.F., [0000-0001-8523-3708](#); C.P., [0000-0002-1267-9546](#); B.Q., [0000-0002-6563-2709](#); S. Mitra, [0000-0002-3426-371X](#); F.M., [0000-0002-3714-9824](#); G.G., [0000-0002-0936-0753](#); I.M.G., [0000-0001-7361-3092](#); S. Manier, [0000-0001-7653-711X](#).

Correspondence: Salomon Manier, CHU Lille, Hospital Huriez, Rue Michel Polonovski, Lille 59000, France; email: salomon.manier@inserm.fr and salomon.manier@chu-lille.fr.

References

1. Willemze R. Cutaneous T-cell lymphoma: epidemiology, etiology, and classification. *Leuk Lymphoma*. 2003;44(Suppl 3):S49-S54.
2. Willemze R, Jaffe ES, Diaz-Perez JL, et al. WHO-EORTC classification for cutaneous lymphomas. *Blood*. 2005;105(10):3768-3785.
3. Olsen E, Vonderheid E, Pimpinelli N, et al. Revisions to the staging and classification of mycosis fungoides and Sézary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC). *Blood*. 2007;110(6):1713-1722.
4. Diamandidou E, Colome-Grimmer M, Fayad L, Ducic M, Kurzrock R. Transformation of mycosis fungoides/sezary syndrome: clinical characteristics and prognosis. *Blood*. 1998;92(4):1150-1159.
5. Vergier B, de Muret A, Beylot-Barry M, et al. Transformation of mycosis fungoides: clinicopathological and prognostic features of 45 cases. *Blood*. 2000;95(7):2212-2218.
6. Benner MF, Jansen PM, Vermeer MH, Willemze R. Prognostic factors in transformed mycosis fungoides: a retrospective analysis of 100 cases. *Blood*. 2012;119(7):1643-1649.
7. Vaqué JP, Gómez-López G, Monsálvez V, et al. PLCG1 mutations in cutaneous T-cell lymphomas. *Blood*. 2014;123(13):2034-2043.
8. Caumont C, Gros A, Boucher C, et al. PLCG1 gene mutations are uncommon in cutaneous T-cell lymphomas. *J Invest Dermatol*. 2015;135(9):2334-2337.
9. Ungewickell A, Bhaduri A, Rios E, et al. Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat Genet*. 2015;47(9):1056-1060.
10. Bastidas Torres AN, Cats D, Mei H, et al. Genomic analysis reveals recurrent deletion of JAK-STAT signaling inhibitors HNRNPK and SOCS1 in mycosis fungoides. *Genes Chromosomes Cancer*. 2018;57(12):653-664.
11. Böni R, Xin H, Kamarashev J, et al. Allelic deletion at 9p21-22 in primary cutaneous CD30+ large cell lymphoma. *J Invest Dermatol*. 2000;115(6):1104-1107.
12. Laharanne E, Chevret E, Idrissi Y, et al. CDKN2A-CDKN2B deletion defines an aggressive subset of cutaneous T-cell lymphoma. *Mod Pathol*. 2010;23(4):547-558.
13. Nicolae-Cristea A, Benner M, Zoutman W, et al. Diagnostic and prognostic significance of CDKN2A/CDKN2B deletions in patients with transformed mycosis fungoides and primary cutaneous CD30-positive lymphoproliferative disease. *Br J Dermatol*. 2015;172(3):784-788.
14. Andrews S, Fast QC. A quality control tool for high throughput sequence data. 2010. Accessed 22 November 2022. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
15. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048.
16. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10-12.
17. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013.
18. Picard Tools (Broad Institute, Cambridge, MA). Version 2.27.5 (October 7, 2022). Accessed 12 February 2023. <http://broadinstitute.github.io/picard/>
19. Van der Auwera GA, Carneiro MO, Hartl C, et al. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43(1110):11.10.1-11.10.33.
20. Van der Auwera GA, O'Connor BD, Brian D. *Genomics in the cloud : using Docker, GATK, and WDL in Terra*. O'Reilly Media; 2020.

21. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
22. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-219.
23. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* 2012;28(14):1811-1817.
24. Benjamin D, Sato T, Cibulskis K, et al. Calling somatic SNVs and indels with mutect2. *bioRxiv.* Preprint posted online 2 December 2019. <https://doi.org/10.1101/861054>
25. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014;505(7484):495-501.
26. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol.* 2016;17(1):122.
27. Ramos AH, Lichtenstein L, Gupta M, et al. Oncotator: cancer variant annotation tool. *Hum Mutat.* 2015;36(4):E2423-E2429.
28. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-218.
29. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012;30(5):413-421.
30. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med.* 2018;24(5):679-690.
31. Luchtel RA, Zimmermann MT, Hu G, et al. Recurrent MSCE116K mutations in ALK-negative anaplastic large cell lymphoma. *Blood.* 2019;133(26):2776-2789.
32. Park J, Daniels J, Wartewig T, et al. Integrated genomic analyses of cutaneous T-cell lymphomas reveal the molecular bases for disease heterogeneity. *Blood.* 2021;138(14):1225-1236.
33. Mao X, Orchard G, Mitchell TJ, et al. A genomic and expression study of AP-1 in primary cutaneous T-cell lymphoma: evidence for dysregulated expression of JUNB and JUND in MF and SS. *J Cutan Pathol.* 2008;35(10):899-910.
34. Mao X, Orchard G, Lillington DM, Russell-Jones R, Young BD, Whittaker SJ. Amplification and overexpression of JUNB is associated with primary cutaneous T-cell lymphomas. *Blood.* 2003;101(4):1513-1519.
35. Lio C-WJ, Yuita H, Rao A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood.* 2019;134(18):1487-1497.
36. Iyer A, Hennessey D, O'Keefe S, et al. Independent evolution of cutaneous lymphoma subclones in different microenvironments of the skin. *Sci Rep.* 2020;10(1):15483.
37. Iyer A, Hennessey D, O'Keefe S, et al. Branched evolution and genomic intratumor heterogeneity in the pathogenesis of cutaneous T-cell lymphoma. *Blood Adv.* 2020;4(11):2489-2500.
38. da Silva Almeida AC, Abate F, Khiabani H, et al. The mutational landscape of cutaneous T-cell lymphoma and Sézary syndrome. *Nat Genet.* 2015;47(12):1465-1470.
39. He Y, Wang L, Wei T, et al. FOXA1 overexpression suppresses interferon signaling and immune response in cancer. *J Clin Invest.* 2021;131(14):e147025.
40. Patel VM, Flanagan CE, Martins M, et al. Frequent and persistent PLCG1 mutations in sézary cells directly enhance PLCγ1 activity and stimulate NFκB, AP-1, and NFAT signaling. *J Invest Dermatol.* 2020;140(2):380-389.e4.
41. Zhang Q, Nowak I, Vonderheid EC, et al. Activation of Jak/STAT proteins involved in signal transduction pathway mediated by receptor for interleukin 2 in malignant T lymphocytes derived from cutaneous anaplastic large T-cell lymphoma and Sezary syndrome. *Proc Natl Acad Sci U S A.* 1996;93(17):9148-9153.
42. Eriksen KW, Kaltoft K, Mikkelsen G, et al. Constitutive STAT3-activation in Sezary syndrome: tyrphostin AG490 inhibits STAT3-activation, interleukin-2 receptor expression and growth of leukemic Sezary cells. *Leukemia.* 2001;15(5):787-793.
43. Qin J-Z, Kamarashev J, Zhang C-L, Dummer R, Burg G, Döbbeling U. Constitutive and interleukin-7- and interleukin-15-stimulated DNA binding of STAT and novel factors in cutaneous T cell lymphoma cells. *J Invest Dermatol.* 2001;117(3):583-589.
44. Kasar S, Kim J, Impropio R, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun.* 2015;6(1):8866.
45. Jones CL, Degasperis A, Grandi V, et al. Spectrum of mutational signatures in T-cell lymphoma reveals a key role for UV radiation in cutaneous T-cell lymphoma. *Sci Rep.* 2021;11(1):3962.
46. McGirt LY, Jia P, Baerenwald DA, et al. Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood.* 2015;126(4):508-519.
47. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature.* 2015;526(7574):525-530.
48. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet.* 2015;47(9):1011-1019.
49. Argyropoulos KV, Pulitzer M, Maura F, et al. Targeted genomic analysis of cutaneous T cell lymphomas identifies a subset with aggressive clinicopathological features. *Blood Cancer J.* 2020;10(11):116.
50. Hristov AC, Tejasvi T, Wilcox RA. Mycosis fungoides and Sézary syndrome: 2019 update on diagnosis, risk-stratification, and management. *Am J Hematol.* 2019;94(9):1027-1041.



Published in final edited form as:

Cancer Cell. 2022 November 14; 40(11): 1358–1373.e8. doi:10.1016/j.ccell.2022.10.017.

Immune biomarkers of response to immunotherapy in patients with high-risk Smoldering Myeloma

Romanos Sklavenitis-Pistofidis^{1,2,3,4}, Michelle P. Aranha^{1,2,4}, Robert A. Redd⁵, Joanna Baginska¹, Nicholas J. Haradhvala^{4,6}, Margaret Hallisey¹, Ankit K. Dutta^{1,2,3,4}, Alexandra Savell^{1,3}, Shohreh Varmeh^{1,4}, Daniel Heilpern-Mallory^{1,4}, Sylvia Ujwary^{1,2,3,4}, Oksana Zavidij^{1,2,3,4}, Francois Aguet⁴, Nang K. Su^{1,2,3,4}, Elizabeth D. Lightbody^{1,2,3,4}, Mark Bustoros^{1,2,3,4}, Sabrin Tahrir^{1,2,3,4}, Tarek H. Mouhieddine^{1,2,3,4}, Ting Wu⁴, Lea Flechon⁷, Shankara Anand⁴, Jacalyn M. Rosenblatt⁸, Jeffrey Zonder⁹, James J. Vredenburgh¹⁰, Adam Boruchov¹⁰, Manisha Bhutani¹¹, Saad Z. Usmani¹¹, Jeffrey Matous¹², Andrew J. Yee¹³, Andrzej Jakubowiak¹⁴, Jacob Laubach¹, Salomon Manier^{7,15}, Omar Nadeem^{1,3}, Paul Richardson^{1,2,3}, Ashraf Z. Badros¹⁶, Maria-Victoria Mateos¹⁷, Lorenzo Trippa⁵, Gad Getz^{2,4,13,18,19}, Irene M. Ghobrial^{1,2,3,4,19,20}

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

²Harvard Medical School, Boston, MA 02115, USA

³Center for Prevention of Progression (CPOP), Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA

⁴Broad Institute of MIT & Harvard, Cambridge, MA 02142, USA

⁵Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

⁶Harvard Graduate Program in Biophysics, Harvard University, Cambridge, MA 02138, USA

⁷INSERM UMRS1277, CNRS UMR9020, Lille University, 59000, France

⁸Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

Correspondence to: Irene M. Ghobrial, MD, Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215, Phone: (617) 632-4198, irene_ghobrial@dfci.harvard.edu, Gad Getz, PhD, Broad Institute of MIT and Harvard, 75 Ames St., Cambridge, MA 02142, Phone: (617)-714-7471, gadgetz@broadinstitute.org.

AUTHOR CONTRIBUTIONS

R.S.P., I.M.G., and G.G. conceived and designed the study; J.M.R., J.Z., J.J.V., A.B., M.B., S.U., J.M., A.J.Y., A.J., J.L., O.N., P.R., A.Z.B., I.M.G. provided patient care; R.S.P., J.B., M.H., A.K.D., S.V., D.H.M., O.Z., N.K.S., E.D.L., M.B., S.T., and T.H.M. acquired the data; R.S.P., M.P.A., R.A.R., J.B., M.H., S.U., T.W., L.F., S.A., and L.T. analyzed the data; R.S.P., N.J.H., L.T., and G.G. provided guidance in data analysis; R.S.P., M.P.A., R.A.R., J.B., N.J.H., A.K.D., S.U., A.S., F.A., L.T. interpreted the data; R.S.P. drafted the manuscript; all authors reviewed, edited and approved the manuscript.

DECLARATION OF INTERESTS

M.P.A., R.A.R., J.B., M.H., A.S., S.V., D.H.M., S.U., N.K.S., E.D.L., S.T., T.W., L.F., S.A., J.M.R., J.Z., J.J.V., A.B., S.U., J.M., A.J., J.L., O.N., P.R., A.Z.B., M.V.M., and L.T. declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic and/or gender minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. We support inclusive, diverse, and equitable conduct of research.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

⁹Barbara Ann Karmanos Cancer Institute, Detroit, MI 48201, USA

¹⁰St. Francis Hospital and Cancer Center, Hartford, CT 06105, USA

¹¹Levine Cancer Institute, Charlotte, NC 28204, USA

¹²Colorado Blood Cancer Institute, Denver, CO 80218, USA

¹³Massachusetts General Hospital Cancer Center, Boston, MA 02114, USA

¹⁴University of Chicago Cancer Center, Chicago, IL 60637, USA

¹⁵Department of Hematology, CHU Lille, Lille University, 59000, France

¹⁶University of Maryland Marlene and Stewart Greenebaum Cancer Center, Baltimore, MD 21201, USA

¹⁷University Hospital of Salamanca–Instituto de Investigación Biomédica de Salamanca, Salamanca, 37007, Spain

¹⁸Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA.

¹⁹These authors contributed equally

²⁰Lead contact

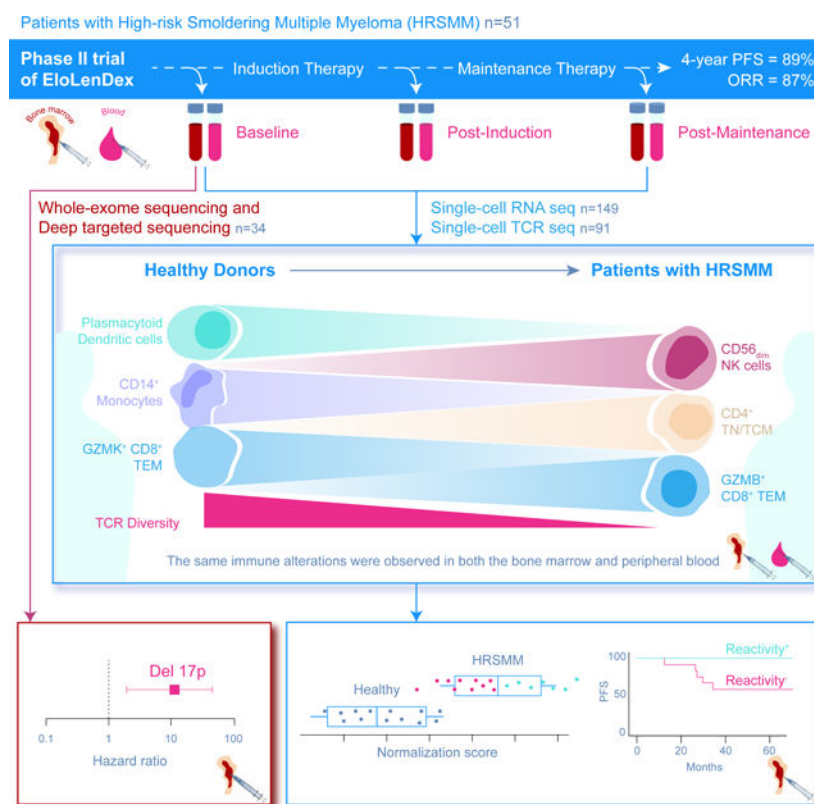
SUMMARY

Patients with Smoldering Multiple Myeloma (SMM) are observed until progression, but early treatment may improve outcomes. We conducted a Phase II trial of Elotuzumab, Lenalidomide, and Dexamethasone (EloLenDex) in patients with high-risk SMM and performed single-cell RNA and T cell Receptor (TCR) sequencing on 149 bone marrow (BM) and peripheral blood (PB) samples from patients and healthy donors (HD). We find that early treatment with EloLenDex is safe and effective and provide a comprehensive characterization of alterations in immune cell composition and TCR repertoire diversity in patients. Importantly, we show that the similarity of a patient's immune cell composition to that of HD may have prognostic relevance at diagnosis and post-treatment, and the abundance of GZMK⁺ CD8⁺ effector memory T cells may be associated with treatment response. Lastly, we uncover similarities between immune alterations observed in the BM and PB, suggesting that PB-based immune profiling may have diagnostic and prognostic utility.

eTOC blurb

Sklavenitis-Pistofidis et al. report results of a Phase II trial of EloLenDex in patients with high-risk smoldering multiple myeloma and use single-cell RNA sequencing to identify biomarkers of outcome. They show that immune cell composition impacts progression-free survival and that blood-based immune profiling can detect immune alterations observed in the marrow.

Graphical Abstract



Keywords

Smoldering Multiple Myeloma; Bone marrow; Peripheral blood; Immune profiling; Single-cell RNA sequencing; Single-cell TCR sequencing; Immune biomarkers; Immunotherapy; Granzyme K

INTRODUCTION

Multiple Myeloma (MM) is an incurable plasma cell malignancy that resides in the bone marrow (BM) and is preceded by an asymptomatic condition called Smoldering MM (SMM) (Weiss et al., 2009). Approximately half of patients with SMM will progress within 5 years from diagnosis, but not all patients with SMM progress (Kyle et al., 2007; Mateos et al., 2020). Therefore, patients with SMM are observed until end-organ damage occurs, which signifies progression to overt MM, warranting treatment (Visram et al., 2021). Two Phase III clinical trials, testing either Lenalidomide/Dexamethasone (LenDex) or Lenalidomide alone (Len) versus observation, demonstrate that early treatment can significantly prolong progression-free survival (PFS) in patients with high-risk SMM (HRSMM) and, in one instance, sustainably prolong overall survival (OS) (Mateos et al., 2013; Mateos et al., 2022; Lonial et al., 2020). Nevertheless, identifying the patients who will benefit the most from early intervention remains an unmet clinical need (Kim et al., 2020; Mateos et al., 2015). Currently, patients at high risk of progression are identified using prediction models based on clinical variables (Kyle et al., 2007; Pérez-Persona et al., 2007), such as the “20–2–20”

model developed by the International Myeloma Working Group (IMWG), which stratifies patients based on the BM plasma cell infiltration (>20%), the amount of M-protein in the serum (>2 g/dL), and the free light-chain ratio (>20) (Mateos et al., 2020). In recent years, it has become evident that genomic biomarkers detected through cytogenetics and sequencing, including translocations t(4;14) and t(14;16), gain or amplification of chr1q, del13q, del17p, translocations and copy number abnormalities (CNAs) involving *MYC*, and *KRAS* mutations, may be associated with faster progression in untreated patients with SMM (Mateos et al., 2020; Bustoros et al., 2020; Misund et al., 2019; Boyle et al., 2021; Bolli et al., 2018; Oben et al., 2021). Moreover, tumor-extrinsic microenvironmental factors are important for disease control at early stages, and the BM immune microenvironment is altered in both composition and functionality in patients with SMM (Zavidij et al., 2020; Bailur et al., 2019; Dhodapkar et al., 2003; Das et al., 2016). However, little is known about the role of BM-based immune profiling in patient prognostication and the utility of peripheral blood (PB) for the detection of immune alterations related to disease.

Here, we conducted a Phase II trial of the immunotherapeutic anti-SLAMF7 antibody, Elotuzumab, in combination with LenDex (EloLenDex, E-PRISM study; [ClinicalTrials.gov Identifier: NCT02279394](https://clinicaltrials.gov/ct2/show/study/NCT02279394)), to determine the utility and safety of early immunotherapy in patients with HRSMM. Moreover, we performed correlative DNA sequencing studies on 34 BM samples at baseline (BL), and single-cell RNA and T cell receptor (TCR) sequencing studies on 149 serial BM and PB samples from patients and healthy donors (HD) to identify genomic and immune biomarkers for optimal patient selection and monitoring of response to treatment.

RESULTS

Early treatment with Elotuzumab, Lenalidomide & Dexamethasone is safe and effective in patients with HRSMM

We conducted a Phase II trial of EloLenDex in patients with HRSMM ($n=51$). Enrollment in E-PRISM was based on the HRSMM criteria by Rajkumar et al. (Rajkumar et al., 2015). The primary objective was to determine the proportion of patients with HRSMM who were progression-free at 2 years post-treatment. Patients were initially randomized 1:1 to Arm A (EloLenDex, $n=40$) or Arm B (EloLen, $n=11$), but accrual to Arm B was halted early, as the difference in exposure to steroids was negligible between the two arms due to the premedication requirement for Elotuzumab infusion. Treatment was planned for 24 cycles followed by observation until progression to overt MM. Stem cell collection was allowed for eligible patients. The baseline characteristics of all patients are listed in Table 1, and the trial schema and CONSORT diagram are shown in Figures S1A and S1B.

Treatment-related Grade 3–5 adverse events with at least 10% frequency were hypophosphatemia ($n=19/51$; 37%), neutropenia ($n=13/51$; 26%), and lymphocytopenia ($n=11/51$; 22%). Thromboembolic events occurred in 6 patients (12%), and no secondary leukemias were observed. Treatment-related Grade 2 adverse events with at least 10% frequency and all Grade 3–5 events are summarized in Table 2 and Figure S1C. Overall, two deaths occurred: one due to uncontrolled diabetic ketoacidosis, bowel perforation, and septic

shock at cycle 19, and another due to hypertensive crisis and myocardial infarction at cycle 24 (Figure S1B).

Median follow-up for all 51 patients was 50 months (range 2–67). Six patients had PFS events defined as progression to myeloma-defining events ($n=4$) or death ($n=2$). Median PFS and OS were not reached for the 46 patients who were eligible for the study and received at least two full treatment cycles; PFS was 88.7% at 48 months (90% CI, 81.2–96.9%) and OS was 95.6% at 48 months (90% CI, 90.6–100%) (Figure S1D). The Kaplan-Meier curve of PFS for eligible patients who received at least two cycles is depicted in Figure 1A.

The overall response rate (Partial Response or better) for these patients was 87% ($n=40$). Three patients (7%) experienced a best overall response of Stringent Complete Response, 1 patient (2%) achieved Complete Response, 14 patients (30%) achieved Very Good Partial Response, 22 patients (48%) achieved Partial Response, and 6 patients (13%) achieved Minor Response (Table S1).

Overall, approximately 13% ($n=6$) of all patients and 20% ($n=3$) of patients defined as having HRSMM by the 20–2–20 criteria experienced progression or death within 4 years of follow-up (Figure S1E). Without treatment, approximately 80% of patients with HRSMM as defined by the 20–2–20 criteria are expected to progress within 5 years of follow-up (Mateos et al., 2020). Taken together, our data suggest that early treatment with EloLenDex was safe and effective in patients with HRSMM.

Genomic predictors of PFS in patients with HRSMM treated with EloLenDex

We observed three main categories of progression in our cohort: those who developed overt myeloma ($n=4$), those who were treated based on evolving biomarkers and the physician's decision ($n=4$), and those who presented biochemical progression ($n=19$). In two of the four individuals who developed overt MM and had samples available for testing, we performed whole-exome sequencing (WES) on CD138⁺ BM cells taken at BL and end of treatment (EOT), to assess whether early treatment had selected for aggressive subclones with high-risk genetic abnormalities. Both individuals had high-risk genomic features at BL: progressor patient #1 had del17p and a *TP53* single nucleotide variant (SNV), while progressor patient #2 had t(4;14), del1p, and *TP53* and *KRAS* SNVs. We did not observe selection of subclones carrying abnormalities associated with increased risk of progression (Figure 1B, 1C). In progressor patient #1, several large-scale CNAs increased in frequency with treatment, although none are known to be associated with an increased risk of progression. In both patients, we observed a post-treatment contraction of potentially aggressive subclones that harbor amp1q (progressor patient #1), and del1p and a *TP53* SNV (progressor patient #2), suggesting that in some instances, early treatment can alter the tumor's clonal dynamics.

A total of 15 patients showed no evidence of clinical or biochemical progression, suggesting that our patient selection can improve further to identify patients with HRSMM who are more likely to benefit from early treatment. Although we were underpowered, we did not observe a significant association between 20–2–20 risk staging and PFS in our cohort or the treatment arm in Lonial et al. (Figure S2A and S2B) (Lonial et al., 2020). To search

for genomic biomarkers of PFS, we performed WES ($n=31$) and deep targeted sequencing (DTS) ($n=3$) on BM samples collected at BL to detect coding genetic alterations and explore their association with patient outcomes (Figure 1D). Testing events present in at least 3 individuals, del17p, t(4;14), and del16q were significant predictors of PFS at the univariate level, while del17p was an independent predictor of progression at the multivariate level (Figure 1E, 1F). These results are consistent with our prior study showing that del17p is a risk factor for progression in untreated patients with SMM (Bustoros et al., 2020), and suggest that patients with SMM and del17p may need a different or longer regimen to prolong PFS, a hypothesis that requires validation in larger cohorts.

Comprehensive characterization of changes in the BM immune cell composition and TCR repertoire of patients with HRSMM

Next, we set out to explore whether immune profiling can help to identify patients with HRSMM who are more likely to benefit from EloLenDex. We performed single-cell RNA sequencing on 211,374 immune cells from 149 samples of 34 patients enrolled in E-PRISM [60 CD138⁺ BM mononuclear cell samples at BL ($n=28$), cycle 9 day 1 (C9D1, $n=16$) and EOT ($n=16$), and 68 PB mononuclear cell (PBMC) samples at BL ($n=33$), cycle 4 day 1 (C4D1, $n=1$), C9D1 ($n=13$), and EOT ($n=21$)], and 21 age-matched HD [BM samples ($n=11$) and PBMCs ($n=10$)]. We also integrated E-PRISM data with our cohort of non-trial patients with MGUS ($n=6$), low-risk SMM (LRSMM, $n=3$), HRSMM ($n=12$), newly diagnosed MM (NDMM, $n=9$), and HD ($n=11$) (Zavidij et al., 2020), reaching a total of 190 samples and 231,608 cells (Figure 2A–L). For 9 samples, we thawed two cell vials and found that the inter-replicate divergence of immune cell composition was significantly lower compared to cross-replicate estimates, suggesting that the quantification of immune cell composition is reproducible (Figure S3A). To assess whether RNA-based quantification is consistent with protein-based estimates, we performed cytometry by time-of-flight (CyTOF) on 17 BM samples (10 drawn at BL, and 7 drawn at EOT) with matched single-cell RNA sequencing data and observed significant positive correlation between protein-based and RNA-based abundance estimates (Figure S3B). By comparing the BM immune cell composition of patients with HRSMM at BL ($n=32$) and HD ($n=19$) with at least 100 cells, we observed a significant increase in the abundance of naïve and memory CD4⁺ T cells (including Tregs, Th1, Th2, and Th17 cells), granzyme B (GZMB)-expressing CD8⁺ T cells, and CD56_{dim} NK cells in patients, and a significant decrease in the abundance of CD14⁺ Monocytes, plasmacytoid DCs (pDCs), AXL⁺ SIGLEC6⁺ dendritic cells (AS-DCs), progenitor cells, and immature B cells (Figure 3A). Furthermore, we observed a relative decrease in the abundance of mucosa-associated invariant T cells (MAITs) in patients, a shift from CD14⁺ to CD16⁺ monocyte subpopulations, and a bias towards memory B cells (Figure S4A–C).

Next, we set out to characterize changes in the TCR repertoire of patients compared to HD and performed single-cell TCR sequencing on 91 samples with available single-cell RNA sequencing data. Specifically, we profiled 70 BM and PB patient samples drawn before and after therapy [BM: 14 at BL, 3 at C9D1, 2 at EOT; PB: 26 at BL, 8 at C9D1, 17 at EOT], and 21 samples from HD [BM:11; PB:10]. To control for variability in T cell numbers which can impact diversity estimates, we randomly sampled a range of T cell numbers ($n=75$ –300)

from each sample over 100 iterations, and averaged repertoire diversity estimates across iterations. We observed that patient BM TCR repertoires were significantly less diverse compared to those of HD, as assessed by the Chao index, independent of the number of T cells sampled (Figure 3B). The observed decrease in repertoire diversity appeared to be driven by broad low-level clonal expansion, rather than a targeted expansion of a few large clones dominating the repertoire (Figure 3C). Regulatory T cells and GZMB⁺ CD8⁺ TEM cells from patient BM were more likely to be part of an expanded clone compared to those from HD BM, where clonal expansion was contained within cytotoxic populations (Figure 3D, 3E). This finding is consistent with the increase in Treg and GZMB⁺ CD8⁺ TEM cell abundance observed in patient BM and suggests that these cells may be expanding in an antigen-specific manner. To gain more insight into the phenotype of clonally expanded GZMB⁺ CD8⁺ TEM cells in patients, we compared their gene expression to that of clonally expanded GZMB⁺ CD8⁺ TEM cells from HD (Figure 3F). Compared to HD, clonally expanded cytotoxic T cells from patients expressed higher levels of *ZNF683* (which encodes the tissue residence transcription factor Hobit), senescence markers *KLRG1* and *B3GAT1* (which encodes CD57), genes associated with effector function (*ITGB1*, *ITGA4*, *ITGB2*, *ITGB7*, *S100A4*, *S100A10*, *S100A6*, *TNF*, *PRF1*, *CCL5*, *GZMH*), and genes induced by interferon signaling (*STAT1*, *XAF1*, *IFITM1*, *IFITM2*), and lower levels of genes related to the induction of potent cytotoxic responses (*GZMB*, *FGFBP2*, *TBX21*, *PRDM1*, *EOMES*, *IFNGR1*) or granule formation and degranulation (*VAMP2*, *OCD1*, *SRGN*). These results suggest that, despite their increased abundance in patient BM, cytotoxic T cells may be dysfunctional. Further mechanistic studies are required to determine the impact of these changes on T cell functionality and tumor clearance.

Immune reactivity at BL and post-therapy immune normalization are associated with longer PFS in patients with HRSMM treated with EloLenDex

We hypothesized that the similarity of a patient's BM immune cell composition to that of HD could be harnessed for prognostication. We trained a Naïve Bayes classifier with 10-fold cross-validation and 100 iterations on a training set of 41 BM samples from patients and HD, to predict the presence of malignancy based on the composition of the BM immune microenvironment; the classifier achieved 94% accuracy in the testing set (n=16, Figure S5A). We then computed a weighted sum score of the product of each cell type's proportion and its corresponding signed importance to the classification ("normalization score"), and classified patients based on the median score at BL as reactive (least normal-like) or non-reactive (most normal-like) (Figure 4A). Patients who were classified as reactive at BL had significantly longer PFS (Figure 4B). Immune reactivity was independent of the patients' risk stage but was significantly associated with the presence of del17p, although patients with del17p were only a fraction of the non-reactive group (Figure 4C, 4D). Plasmacytoid DCs and the pro-inflammatory Cytokine⁺ CD14⁺ Monocytes, which expressed genes related to neutrophil chemotaxis and acute inflammation, were among the top 3 cell types with the highest importance for a sample to be classified as normal-like. Aside from monocytes, this pro-inflammatory signaling was also active in DCs and pDCs (gene expression signature GEX-6), and together with pDC signaling (GEX-13), they were significantly associated with shorter PFS (Figure 4E, S5B–E). As expected, reactive patients had significantly lower abundance of pDCs and Cytokine⁺ CD14⁺ Monocytes (Figure

S5F). By comparing gene expression levels between reactive and non-reactive patients, we observed a significant downregulation of exhaustion markers (*TOX*, *TNFRSF9*, *TNFSF9*, *PDCD1*, *NR4A2*, *NR4A3*) in GZMK⁺ CD8⁺ TEM cells of reactive patients, along with an upregulation of markers associated with long-lived memory effectors (*IL7R* and *CD27*), markers associated with terminal effectors (*GZMB*, *GZMH*, *FCGR3A*, *FGFBP2*, *NKG7*), and markers associated with functionality (*IFNG*, *TNF*), a profile consistent with long-lived memory effector cells of increased potency (Figure 4F). Moreover, in GZMB⁺ CD8⁺ TEM cells from reactive patients, we observed an upregulation of exhaustion markers (*TIGIT*, *LAG3*) and the senescence marker *KLRG1*, along with higher levels of *IFNG* and *TNF*, and downregulation of cytotoxicity markers (*GZMB*, *FCGR3A*, *PRF1*), a profile consistent with exhausted, senescent terminal effector cells (Figure 4G). Consistent with the hypothesis that these cells may be exhausted due to chronic antigenic stimulation, GZMB⁺ CD8⁺ TEM cells from reactive patients were more likely to be clonally expanded, compared to those from non-reactive patients (Figure 4H). Collectively, these results suggest that immune reactivity captures a patient subpopulation with decreased pro-inflammatory myeloid signaling, long-lived GZMK⁺ effector memory cells of increased potency, and clonally expanded, exhausted GZMB⁺ terminal effectors.

On average, patients showed significantly higher normalization scores at EOT, compared to BL (Figure 4I). Using a threshold of 0.1 based on the distribution of change in normalization scores at EOT, we defined a state of post-therapy immune normalization (PIN) and classified four patients who showed no or minimal increase in their normalization score at EOT as PIN⁻ (Figure 4J, 4K). Patients who achieved PIN at EOT (PIN⁺) had significantly longer PFS (Figure 4L), suggesting that PIN may be important for prognostication. Of note, both patients who had del17p out of the 12 patients with matched BL and EOT BM samples were classified as PIN⁻, so the effect of PIN on PFS needs to be deconvolved from that of del17p in larger cohorts.

Higher abundance of Granzyme K-expressing CD8⁺ T cells is associated with longer PFS in patients with HRSMM treated with EloLenDex

We previously showed that patients with HRSMM have decreased abundance of memory-like GZMK⁺ CD8⁺ TEM cells in favor of more mature GZMB⁺ effectors and that lower levels of memory T cells *in vivo* are associated with shorter OS (Zavidij et al., 2020). Therefore, we hypothesized that higher levels of GZMK⁺ CD8⁺ TEM cells in the BM may be associated with prolonged PFS in our patients. Consistent with our prior findings, patients with HRSMM showed a significant shift from GZMK⁺ to GZMB⁺ CD8⁺ TEM cells in the BM, with a significant decrease in the abundance of GZMK⁺ CD8⁺ TEM cells relative to all cytotoxic T cells at BL compared to HD (Figure 5A). However, the average levels of *GZMK* expression in patient T cells increased significantly at EOT, reflecting a significant increase in GZMK-expressing cells, which was validated by CyTOF (Figure 5B, S6A). In the PB, where we were powered to investigate changes in T cell clonality at EOT, we found that GZMK⁺ CD8⁺ TEM cells from EOT samples were more likely to belong to expanded clones (Figure 5C). Together, these results suggest that this subpopulation may play a role in response to immunotherapy. Compared to GZMB⁺ CD8⁺ TEM cells in our cohort, GZMK⁺ CD8⁺ TEM cells showed higher expression of *CXCR4* and lower

expression of S1P receptors, suggesting preferential retainment of GZMK⁺ CD8⁺ TEM cells within the BM (Figure 5D). Indeed, by comparing the abundance of these cells between matched BM and PB patient samples, we observed enrichment of GZMK⁺ CD8⁺ TEM cells in the BM (Figure 5E). This enrichment was also observed in BM samples from HD (data not shown), suggesting that this may be a physiological bias. Furthermore, GZMK⁺ CD8⁺ TEM cells had higher expression of canonical exhaustion markers *TOX*, *TIGIT*, *TNFRSF9* (encoding 4–1BB), and *TNFSF9* (encoding 4–1BBL), as well as markers of long-lived memory effector cells, including *TCF7* (encoding Tcf-1), *CD27* and *CD28*, a phenotype consistent with progenitor exhausted T cells (T_{PEX}) (Figure 5D). Indeed, at the protein level, GZMK⁺ CD8⁺ TEM cells were shown to be a major source of PD-1 expression within cytotoxic T cells in patients (Figure 5F, S6B), in accordance with the suspected T_{PEX} phenotype. Progenitor exhausted T cells maintain their proliferative capacity, can differentiate into terminal effector cells, and are thought to mediate the clinical benefits observed in patients treated with immune checkpoint inhibitors (Galletti et al., 2020). In our cohort, patients with a higher abundance of GZMK⁺ CD8⁺ TEM cells had significantly longer PFS in response to EloLenDex (Figure 5G). Similarly, patients with higher average *GZMK* expression across their T cells had significantly longer PFS (Figure S6C). Of note, no association was observed between the abundance of GZMK⁺ CD8⁺ TEM cells and age (Wilcoxon, p=0.55), while age had no effect on PFS (Log-rank, p=0.7).

Collectively, these results suggest that GZMK⁺ CD8⁺ TEM cells may be important for disease control and response to therapy in patients with SMM.

PB-based immune profiling accurately detects alterations in immune cell composition and TCR repertoire diversity observed in the BM

Next, we set out to determine whether PB can be used reliably for immune profiling of patients with HRSMM, as BM biopsies are invasive and carry risk, which precludes regular patient testing. First, we compared the Jensen-Shannon divergence of immune cell composition between matched and unmatched PB and BM samples and observed significantly lower divergence in matched pairs, which was also low overall, suggesting that matched PB samples reflect BM composition well (Figure 6A). In PCA space, where PC1 captured the compositional changes observed in patient BM (Figure 6B), HD PB and BM samples clustered together on the right end of the axis, while patient PB and BM samples clustered together on the opposite side of PC1 (Figure 6C). This observation suggests that the compositional differences between patient and HD PB are similar to those observed in patient BM compared to HD BM. Indeed, when we compared PB samples between patients and HD, we recovered compositional changes originally discovered in our BM-based analysis (Figure 6D). However, patient PB showed a significantly lower abundance of CD16⁺ monocytes, which are enriched in patient BM.

Contrary to HD where CD16⁺ monocytes are significantly enriched in the PB compared to the BM, in patients there was no significant difference between the PB and the BM, suggesting that these cells may be homing to the patient BM (Figure S7). Lastly, consistent with what we observed in patient BM, patient PB TCR repertoire showed significantly lower diversity compared to HD, suggesting that the observed effect on diversity may not be

constrained to the tumor microenvironment (Figure 6E). These results suggest that certain immune biomarkers may be measurable in the PB. Consistent with what we saw in the BM, higher BL levels of signature GEX-13 in patient PB were associated with significantly shorter PFS (Figure 6F). Of note, neither signature GEX-6 nor immune reactivity were prognostic in PB, suggesting that PB-based immune profiling may not be able to replace BM-based profiling entirely.

Lastly, given the concordance of compositional changes between patient BM and PB, we tested whether immune profiling of PB samples alone could enable the diagnosis of an individual with HRSMM. Remarkably, a classifier that we trained on BM composition data (this time excluding progenitor cells, which are primarily encountered in the BM) to infer the presence of SMM was able to correctly classify nearly all our patient and HD PB samples with an accuracy of 97% (Figure 6G). Collectively, our findings demonstrate that PB-based immune profiling accurately detects compositional changes observed in the BM and may hold diagnostic and prognostic potential.

DISCUSSION

While early treatment has been shown to delay progression in patients with HRSMM, identifying those who will benefit the most remains an unmet clinical need (Mateos et al., 2015; Kim et al., 2020). Despite the availability of clinical and genomic biomarkers of progression (Kyle et al., 2007; Mateos et al., 2020; Pérez-Persona et al., 2007; Bustoros et al., 2020; Misund et al., 2019; Boyle et al., 2021; Bolli et al., 2018; Oben et al., 2021), little is known about the role of BM immune dysregulation in disease progression and response to treatment or the diagnostic and prognostic utility of PB-based immune profiling. Here, we conducted a Phase II trial to determine the utility and safety of early treatment with EloLenDex in patients with HRSMM and performed correlative DNA sequencing and single-cell RNA and TCR sequencing studies to develop genomic and immune biomarkers for optimal patient selection and monitoring of response to treatment. The overall response rate was 87%, PFS was 88.7% at 48 months (90% CI, 81.2–96.9%), and OS was 95.6% at 48 months (90% CI, 90.6–100%). Although this trial did not have an observation arm, this progression rate (~13% at 4 years) is substantially lower than that determined for patients with HRSMM in large observational studies which showed progression rates of 50% or more within 2 years from diagnosis and 70–80% within 5 years from diagnosis (Rajkumar et al., 2015; Kyle et al., 2007; Mateos et al., 2020). This was also true for patients in our trial who were high-risk by the 20–2–20 criteria and who showed a 20% progression rate within 4 years of follow-up. Treatment was well-tolerated with minimal toxicity of Grade 3 or higher, including hypophosphatemia and cytopenias, which were expected. As this was a single-arm study, we cannot comment on how the response rate and PFS of EloLenDex compares to those of LenDex alone, and future clinical trials randomizing these 2 regimens would be required to answer this. Overall, EloLenDex was deemed safe and effective, with no evidence of aggressive subclone selection post-treatment in two patients who progressed. In fact, in those patients, we observed a regression of subclones harboring either amp1q or del1p and a *TP53* SNV. Although preliminary, these results suggest that in some instances early treatment could alter the clonal dynamics in patients who progress, which could potentially result in improved OS in the long-term. These findings would need

to be validated in larger trials with more available samples pre- and post-therapy. Aggressive tumor biology, as determined by the presence of del17p, was a significant predictor of shorter PFS in patients treated with EloLenDex in our cohort. Therefore, patients with HRSMM and del17p may perhaps need different regimens or longer duration of therapy.

Patients with HRSMM showed increased abundance of naïve and memory CD4⁺ T cells, GZMB⁺ CD8⁺ TEM cells, and CD56_{dim} NK cells, and decreased abundance of CD14⁺ monocytes, and pDCs compared to HD. Furthermore, their TCR repertoire diversity was significantly lower compared to HD, due to broad low-level clonal expansion, which was mostly observed in GZMB⁺ CD8⁺ TEM cells and Tregs. Further studies are required to determine the functionality and antigen specificity of these immune cell populations and their role in controlling or promoting disease progression.

Importantly, we found that the similarity of a patient's BM immune microenvironment to that of HD may be useful for prognostication. At BL, patients with the least normal-like immune cell composition, presumed to be reactive to the tumor, had significantly longer PFS under treatment with EloLenDex. Reactive patients had lower proportions of pro-inflammatory myeloid cells and pDCs, and had long-lived GZMK⁺ CD8⁺ TEM cells of increased potency, and clonally expanded, exhausted GZMB⁺ CD8⁺ TEM cells. Granzyme K-expressing cells, which are normally enriched in the BM, had a phenotype compatible with progenitor exhausted cells, and clonally expanded post-treatment, suggesting that they may play a role in response to therapy. Patients with a higher abundance of GZMK⁺ CD8⁺ TEM cells in their BM showed significantly longer PFS. Furthermore, patients whose BM normalization scores increased post-therapy (Post-therapy Immune Normalization, PIN), had significantly longer PFS. This finding suggests that immune profiling may be useful to monitor response to treatment, although validation in larger cohorts and with different regimens is required.

Lastly, immune profiling currently requires a BM biopsy, which precludes its regular use for prognostication. Remarkably, by comparing PB from patients and HD, we identified nearly the same alterations in immune cell composition and TCR repertoire diversity as those seen in our BM-based comparisons. Moreover, a classifier trained to detect the presence of SMM based on BM immune profiling was able to correctly classify nearly all PB samples from patients with SMM and HD. Our results suggest that PB-based immune profiling may have diagnostic and prognostic utility, although it may not be able to entirely replace BM-based profiling.

Collectively, our results demonstrate the utility and safety of EloLenDex in patients with HRSMM, provide a comprehensive characterization of alterations in immune cell composition and TCR repertoire diversity in patient BM and PB, nominate immune biomarkers for optimal patient selection and assessment of response to treatment, and demonstrate that PB-based immune profiling may have diagnostic and prognostic utility in patients with SMM.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Irene Ghobrial (irene_ghobrial@dfci.harvard.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- Raw single-cell RNA and TCR sequencing data generated for this study have been deposited in dbGaP (phs002476.v1.p1). Raw single-cell RNA sequencing data generated in Zavidij et al. have been deposited in dbGaP (phs001323.v3.p1). Raw whole-exome sequencing data generated for this study have been deposited in dbGaP (phs001323.v3.p1).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical trial and patients

Study design: The primary objective of this trial was to determine the proportion of patients with high-risk SMM (HRSMM) who were progression-free at 2 years post-treatment. Secondary objectives included safety and toxicity, time to progression, overall response rate, duration of response, and overall survival.

From January 2015 to November 2016, patients were enrolled at Dana Faber Cancer Institute, Boston, Massachusetts; Beth Israel Deaconess Medical Center, Boston, Massachusetts; Carolina HC Levine Cancer Institute, Charlotte, North Carolina; Colorado Blood Cancer Institute, Denver, Colorado; Eastern Maine Medical Center, Bangor, Maine; Karmanos Cancer Institute, Detroit, Michigan; Marlene and Stewart Greenbaum Cancer Center, Baltimore, Maryland; Massachusetts General Hospital, Boston, Massachusetts; Newton-Wellesley Hospital, Newton, Massachusetts; St. Francis Hospital, Hartford, Connecticut; and University of Chicago, Chicago, Illinois.

Patients were eligible for this trial if they were 18 years of age or older, had an Eastern Cooperative Oncology Group performance status of 0 to 2, and were classified as HRSMM. HRSMM was defined according to Rajkumar et al. as the presence of BM clonal plasma cells $\geq 10\%$ (but less than 60%) and at least one of the following: serum M protein ≥ 3.0 mg/dL; IgA SMM; immunoparesis with reduction of two uninvolved immunoglobulin isotypes; serum involved/uninvolved free light chain ratio ≥ 8 (but less than 100); evolving type of SMM, i.e., progressive increase in M protein level; BM clonal plasma cells 50–60%; abnormal cell immunophenotype ($\geq 95\%$ of BM plasma cells are clonal) and reduction of one or more uninvolved immunoglobulin isotypes; t(4;14) or del17p or amp1q; increased

circulating plasma cells; MRI with diffuse abnormalities or 1 focal lesion ($\geq 5\text{mm}$); PET-CT with one focal lesion ($\geq 5\text{mm}$) with increased uptake without underlying osteolytic bone destruction (Rajkumar et al., 2015). Furthermore, patients had to show no evidence of myeloma-defining events as described by the IMWG (Rajkumar et al., 2021), and the following laboratory values had to be shown within 28 days prior to registration: Absolute Neutrophil Count $\geq 1000/\text{mm}^3$, Platelets (PLT) $\geq 50,000/\text{mm}^3$, Total Bilirubin $\leq 2.0\text{mg/dL}$, Aspartate and Alanine Aminotransferase (AST/ALT) $\leq 3 \times \text{ULN}$, and an estimated creatinine clearance $> 50\text{mL/min/1.73m}^2$. Patients with other concurrent chemotherapy, immunotherapy, radiotherapy, or any ancillary therapy considered investigational, serious medical or psychiatric illness likely to interfere with participation, a diagnosis or treatment for another malignancy within 2 years of enrollment were excluded.

All patients provided informed consent. The review boards of all participating centers approved the study in accordance with the Declaration of Helsinki and the International Conference of Harmonization Guideline for Good Clinical Practice. This study was registered as a phase II study with [ClinicalTrials.gov NCT02279394](https://clinicaltrials.gov/ct2/show/study/NCT02279394).

Study treatment: This phase II trial used a single-stage design to assess the efficacy of the combination of Elotuzumab and Lenalidomide with or without Dexamethasone in patients with HRSM. Patients were randomized 1:1 to Arm A (EloLenDex) or Arm B (EloLen), based on the stratification factors described below.

In the treatment schema a cycle is defined as 28 consecutive days. In cycles 1 and 2, 10mg/kg of intravenous push Elotuzumab was administered on days 1, 8, 15, and 22. In cycles 3–4, it was administered on days 1 and 15. A 25mg dose of oral Lenalidomide was administered on days 1–21 of each cycle. In Arm A, 40mg of oral Dexamethasone was administered on days 1, 8, 15, and 22 in cycles 1 and 2 and on days 1, 8, and 15 in cycles 3–8. Patients on maintenance treatment (cycles 9–24) were administered 20mg/kg of intravenous push Elotuzumab on day 1 and 15mg oral Lenalidomide on days 1–21. All patients received thromboprophylaxis.

In consideration of the Mateos et al. study, investigators were given the option to allow high-dose Dexamethasone to be resumed during maintenance due to biochemical progression (progressive increase in serum M-spike by at least 25% with an absolute increase of at least 0.5 g/dL or urine M-spike by at least 25% with an absolute increase of at least 200 mg/24 hours, on two successive evaluations, as determined by the IMWG response criteria or documented progression by serum free light chain criteria in the absence of serum or urine involvement; Mateos et al., 2013). A total of 5 patients were given Dexamethasone during maintenance. The dosing schedule for this followed that of cycles 3–8, with 8mg intravenous push and 32mg oral Dexamethasone on days 1 and 15 and 40mg oral Dexamethasone on day 8.

Stratification Factors: Originally, the study aimed to secondarily determine whether Dexamethasone is beneficial or detrimental in this immunotherapy regimen for SMM. Therefore, the study was designed with two arms, meant to serve as two independent phase II studies to determine the activity of EloLen with or without Dexamethasone in the patient

population. Patients were stratified by age (≥ 65 vs < 65) and high-risk cytogenetics [t(4;14), t(14;16), del17p, *TP53* mutation, amp1q] (high-risk, low-risk, FISH failure) before being randomly assigned to one of the two arms.

Arm B accrual: Arm B accrual was halted in April 2016 for two reasons: 1) treatment with Elotuzumab required premedication with high doses of steroids to avoid infusion reactions and the difference in Dexamethasone dosing between the two arms (approximately 4mg Dexamethasone per dosing day) for the first 2 cycles was not significant, and 2) a publication by Paiva et al. (Paiva et al., 2016). demonstrated that once weekly high-dose Dexamethasone does not have a detrimental effect on the immune system in patients with SMM in a clinical trial of Lenalidomide and Dexamethasone. The patients who were originally enrolled in Arm B remained in Arm B until the completion of therapy.

Cohort details: In total, 51 patients enrolled and received at least one dose of treatment. Three patients were deemed ineligible soon after they received treatment, due to misdiagnosis (overt myeloma with lesions on PET-CT, amyloidosis, and light chain deposition disease), and were excluded from further analysis. Two patients received less than 2 full cycles of treatment and were excluded from our overall response assessment and correlative studies but were included for toxicity assessment. Out of the 46 patients remaining in the cohort, 29 were female and 17 were male, while the median age at registration was 62 (range: 29–77).

Imaging: To assess the extent of disease involvement, imaging studies were performed at both BL and EOT. All patients had either an MRI with a skeletal survey or a PET/CT scan.

Efficacy and Safety Assessments: Toxicities were monitored throughout the study and for up to 30 days after the last dose of the study drug. Adverse events (AEs) were graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE; version 4.0; Bethesda, MD). In addition, neuropathy symptoms were assessed using the FACT/GOG-Neuropathy questionnaire (Version 4.0). Efficacy was measured using the IWMG criteria for response.

Treatment discontinuation and patient disposition: The median duration of treatment was 24 cycles (range= 1–24). In total, 40 of 51 patients received all 24 cycles. Of those, 30 (75%) were in arm A and 10 (25%) in arm B. Of the 10 patients in Arm A who did not complete all 24 cycles, 3 patients discontinued therapy during the induction cycles due to an incorrect initial diagnosis (one participant had myeloma lytic lesions on a PET scan that was originally misread at an outside institution and re-checked at DFCI afterward; another participant had amyloidosis that was missed at the time of enrollment but was discovered after 3 cycles of therapy—the patient achieved VGPR on therapy before being removed from the protocol and placed on active amyloidosis therapy with a proteasome inhibitor-based approach; one final patient had light chain deposition disease, which was identified on a kidney biopsy after one cycle of therapy—the patient was then removed from the protocol to start active therapy). Six other patients discontinued therapy due to toxicity and withdrawal of consent or physician decision. Patient disposition is visualized in Figure S1B.

Overall, two deaths occurred. One patient had Type I Diabetes and was hospitalized for Diabetic ketoacidosis (cycle 19). This was at a time when the patient was not receiving Dexamethasone. The patient developed diverticulitis, bowel perforation, and sepsis and died due to sepsis. The second patient was on the last cycle of therapy when they experienced uncontrolled hypertension due to non-compliance with their hypertension medications. They then developed heart failure and died of hypertensive crisis.

METHOD DETAILS

Sample processing—Bone marrow (BM) aspirate and peripheral blood (PB) samples were collected in EDTA tubes, and mononuclear cells (BMMCs & PBMCs) were isolated using Red Blood Cell Lysis Buffer (ThermoFisher). Tumor and immune cells were then isolated by magnetic enrichment for CD138 (Miltenyi Biotec), according to the manufacturer's instructions, and cryopreserved in FBS with 10% DMSO.

Whole-Exome and Deep Targeted Sequencing—We performed WES on baseline BM aspirate specimens from 37 study participants. Tumor DNA was extracted from CD138⁺ cells from the patients' BM. For germline control, DNA was obtained from PBMCs. Genomic DNA was extracted using the QIAamp DNA mini kit (QIAGEN) according to the manufacturer's protocols, and the DNA concentration was quantified using PicoGreen dsDNA Assay kit (Life Technologies). WES libraries were prepared by Agilent SureSelect XT2 Target Enrichment kit (V5+UTR probes). For three individuals, libraries for deep targeted sequencing (DTS) were prepared using SureSelect XT Reagent Kits (Agilent) and an in-house bait set targeting 117 genes, comprising pan-cancer driver genes and frequently mutated genes in MM (Bustoros et al., 2020). Lastly, for two out of four patients who progressed to overt myeloma, tumor samples were available at the time of progression and were also studied by WES (Roche MedExome+) to characterize the landscape of clonal evolution under treatment. All sequencing was performed on the Illumina HiSeq 4000 platform at the Genomics Platform of the Broad Institute of MIT & Harvard. Six WES samples were excluded from analysis due to low tumor purity (<15%). Our final baseline cohort comprised 31 WES samples and 3 DTS samples.

Single-cell RNA sequencing—We performed single-cell RNA sequencing on 149 CD138[−] BMMC and PBMC samples from 34 patients enrolled in the E-PRISM trial. Specifically, we sequenced 60 CD138[−] BMMC samples at BL ($n=28$), C9D1 ($n=16$), and EOT ($n=16$), 68 PBMC samples at BL ($n=33$), C9D1 ($n=13$), C4D1 ($n=1$), and EOT ($n=21$), 11 CD138[−] BMMC samples from age-matched HD, and 10 PBMC samples from age-matched HD. For 23 patients, BL BM and PB samples were matched; for 8 patients, C9D1 BM and PB samples were matched; for 13 patients EOT BM and PB samples were matched (Figure S8A). Of note, in contrast to our patient samples, HD BM and PB samples were from different individuals.

For library preparation, we used Chromium Single Cell 3' Gene Expression ($n=40$) and Chromium Single Cell 5' Gene Expression kits ($n=109$) by 10X Genomics. Specifically, we performed 3' single-cell RNA sequencing on 40 CD138[−] BMMC samples drawn at BL ($n=11$), C9D1 ($n=13$) or EOT ($n=9$), and PBMC samples drawn at BL ($n=4$) and C9D1

($n=3$). We performed 5' single-cell RNA sequencing on 98 CD138⁺ BMMC samples drawn at BL ($n=17$), C9D1 ($n=3$) or EOT ($n=7$), and PBMC samples drawn at BL ($n=29$), C4D1 ($n=1$), C9D1 ($n=10$) or EOT ($n=21$). We also performed 5' single-cell RNA sequencing on 11 CD138⁺ BMMC samples and 10 PBMC samples from HD. All 3' libraries were sequenced on a HiSeq2500 instrument and all 5' libraries were sequenced on a NovaSeq 6000 instrument at the Genomics Platform of the Broad Institute of MIT & Harvard.

Reproducibility of cell type abundance quantification via single-cell RNA

sequencing: To assess the reproducibility of our cell type quantification, we thawed two cell vials for each of 9 samples; for five of them, both replicate libraries were prepared using the same technology (3'- or 5'-end sequencing) (technical replicates); for four of them, each replicate library was prepared with a different technology (technology replicates).

Single-cell TCR sequencing—We performed single-cell TCR sequencing on 91 samples with available single-cell RNA sequencing data (Figure S8A). Specifically, we profiled 70 BM and PB patient samples drawn before and after therapy [BM: 14 at BL, 3 at C9D1, 2 at EOT; PB: 26 at BL, 8 at C9D1, 17 at EOT], and 21 samples from HD [BM:11; PB:10]. Complementary DNA generated from barcoded CD138⁺ immune cells using Chromium Single Cell 5' Gene Expression and V(D)J enrichment kits by 10X Genomics was subjected to V(D)J enrichment and library preparation and sequenced on a NovaSeq 6000 instrument at the Genomics Platform of the Broad Institute of MIT & Harvard.

Cytometry by Time-of-Flight (CyTOF)—Metal-tagged antibodies used for the mass cytometry panel are listed in the Key Resources table. In-house conjugation of Granzyme K antibody was performed with Maxpar labeling kit per manufacturer instructions (Fluidigm). All antibodies were used per the manufacturer's recommendation (Fluidigm).

Cryopreserved patient CD138⁺ BM Mononuclear Cell (BMMC) samples were thawed, counted using AO/PI, and pelleted by centrifugation at 400g for 10 minutes. Cells were then incubated in 103Rh viability stain for 15 minutes, washed in CyFACS, and incubated with undiluted Human TruStain FcX for 10 min for Fc receptor blocking. Antibody master mix was applied to the cell suspension for 30 minutes, washed, and fixed/permeabilized with FoxP3 Fixation/Permeabilization Concentrate and Diluent following manufacturer's guidelines (eBioscience). A mix of intracellular antibodies prepared with 1X Perm Wash was added to each sample and incubated for 30 minutes. Next, cells were washed with 1X Perm Wash, and incubated overnight at 4°C in FoxP3 Fixation/Permeabilization Concentrate and Diluent, containing 191/193Ir DNA Intercalator. Prior to acquisition, samples were transferred to 5 mL round-bottom polystyrene tubes with cell strainer caps, washed and filtered with Cell Staining Buffer (CSB), Cell Acquisition Solution (CAS), and resuspended in CAS supplemented with EQ Four Element Calibration Beads (1:10).

All mass cytometry data was collected on a HeliosTM Mass Cytometer (Fluidigm). The instrument was tuned using CyTOF Tuning Solution according to the Helios User Guide (Fluidigm, p. 60–68). A brief overview of tuning steps includes Pre-XY Optimization, Mass Calibration, XY Optimization, DV Calibration, Dual Calibration, Gases/Current Calibration,

and QC report. EQ Four Element Calibration Beads (1:10 in CAS) were used according to the manufacturer protocol before and during acquisition.

QUANTIFICATION AND STATISTICAL ANALYSIS

Clinical trial statistical analysis & biomarker discovery—Patient baseline characteristics were summarized as the number of patients or median values. Responses to study treatment were reported as proportions with 90% exact binomial confidence intervals.

We assessed the proportion of patients alive and progression-free at 2 years, compared to the previously published rate for HRSMM—a median time to progression of only 1.9 years according to the Kyle et al. model (Kyle et al., 2007). Therefore, we decided that for this study, a 2-year progression-free survival rate of 50% would not be considered promising; instead, a progression-free survival rate of 70% or higher would be a promising result.

This phase II study used a single-stage binomial design. With a sample size of 39 patients, the probability of concluding that the treatment is effective if the true rate is 50% was 0.10 and was >0.9 if the true rate is 70%. Assuming an ineligibility rate of 5%, we aimed to accrue 41 patients per Arm to the trial. Time to response (TTR) was measured from treatment initiation to the date the response was first observed. Duration of response was measured from response to progressive disease (PD) or death, censored at the date of last disease assessment for those who did not progress. Time-to-progression (TTP), and progression-free survival (PFS) were measured from the time of treatment initiation to event (PD for TTP; PD or death for PFS). Patients without event were censored at the date of last disease assessment for both TTP and PFS. If non-protocol therapy, excluding erythropoietin, was added prior to an event, patients were censored in the time-to-event analyses at the initiation of non-protocol therapy. Six total patients had progression-free survival events defined as progressive disease to myeloma-defining events ($n=4$) or death ($n=2$). Another four patients ($n=4$) were treated based on worsening biomarkers and the physician's decision and were considered progressors for our correlative studies only. The Kaplan-Meier method was used to estimate time to event and the log-rank test was used to compare time to event. Cox proportional hazards regression was employed to assess the significance of variable associations with outcome. Continuous variables, such as cell type proportions, were tested as such and were dichotomized based on the median for Kaplan-Meier curves. Statistical analyses were performed using R version 4.1.3. All results with p -values < 0.05 were considered statistically significant. The Benjamini-Hochberg method was used to correct for multiple hypothesis testing, when appropriate. Our data cut-off was November 7, 2020.

ECOG Randomized Trial of Lenalidomide versus Observation in patients with

HRSMM: Due to the unavailability of patient-level information, survival data were recapitulated from published Kaplan-Meier (KM) plots in Lonial et al (Lonial et al., 2020). The curves and at-risk tables were extracted as raster images, and their x and y coordinates were digitized using the commercial software DigitizeIt (<https://www.digitizeit.xyz/>). Using the algorithm proposed by Guyot et al., the coordinates and number of patients at-risk during each time interval were used to estimate censored, time-to-event data (Guyot et al., 2012). With the patient-level data, we were able to re-generate the survival curves, at-risk

tables, log-rank tests, etc. In particular, we used the KM method to explore the effect of the 20–20 risk stratification system on PFS in the Lenalidomide arm of the ECOG trial and the log-rank test to compute the p-value.

WES analysis—The output from Illumina software was processed by the Picard data processing pipeline to yield BAM files containing well-calibrated, aligned reads. We utilized the Getz Lab CGA WES Characterization pipeline developed at the Broad Institute to call, filter, and annotate somatic mutations and copy number variation. The pipeline employs the following tools: MuTect, ContEst, Strelka, Orientation Bias Filter, DeTiN, AllelicCapSeg, MAFPoNFilter, RealignmentFilter, ABSOLUTE, GATK, PicardTools, Variant Effect Predictor, Oncotator (Cibulskis et al., 2013; Cibulskis et al., 2011; Saunders et al., 2012; Costello et al., 2013; Taylor-Weiner et al., 2018; Landau et al., 2013; Lawrence et al., 2014; Carter et al., 2012; McKenna et al., 2010; McLaren et al., 2016; Ramos et al., 2015). SNVs and CNVs were further cleaned with a custom PoN made of matched normal samples (Lawrence et al., 2014), and a bait bias filter was developed to correct for observed artifacts in the SNV data, as described before (Bustoros et al., 2020). ABSOLUTE was applied to estimate sample purity, ploidy, and absolute somatic copy numbers, which were used to infer the CCFs of point mutations and CNVs from the WES data, following the framework previously described (Carter et al., 2012).

Single-cell RNA sequencing analysis

Analysis workflow: CellRanger v.5.0.1. was used for FASTQ file extraction and generation of count matrices (Zheng et al., 2017). We integrated E-PRISM data with our cohort of non-trial patients with MGUS ($n=6$), low-risk SMM (LRSM, $n=3$), HRSM (HRSM, $n=12$), newly diagnosed MM (NDMM, $n=9$) and HD ($n=11$) (Zavidij et al., 2020). Non-trial patients with SMM were risk-stratified based on the original Mayo Clinic criteria (Kyle et al., 2007), and NDMM was diagnosed based on the CRAB criteria.

We performed ambient RNA correction with SoupX; doublet detection with Scrublet, scDblFinder, and SCDS; normalization with Scran; and integration with Harmony (correcting for sample ID and technology) (Young et al., 2020; Wolock et al., 2019; Germain et al., 2021; Bais et al., 2019; Lun et al., 2016; Korsunsky et al., 2019). Samples prepared with 3'-end or 5'-end technology mixed well in PCA, demonstrating successful integration across technologies (Figure S8B). Droplets were deemed to be doublets when at least 2 out of four methods classified them as such and they were only removed from consideration when they clustered together and co-expressed markers of multiple cell types. Droplets containing dying cells with more than 15% mitochondrial gene expression were removed before clustering. Clusters with higher mitochondrial and ribosomal gene expression that clearly separated from well-annotated clusters and lacked interpretable expression markers were removed downstream. In total, we profiled 211,374 immune cells (~1,419 cells/sample), comprising T cells, NK cells, B cells, Monocytes, Dendritic cells, and progenitor populations. For each cell population (T cells, NK cells, B cells, Monocytes, Dendritic cells, progenitor cells), feature selection, dimensionality reduction, and clustering were also performed separately, to identify cell subpopulations (see below for annotation strategy).

Cell type/subtype proportions were calculated as the fraction of cells belonging to said type/subtype out of all immune cells unless otherwise specified. For comparisons of cell type proportions, samples with fewer than 100 cells were removed from consideration. Proportion changes were assessed with Wilcoxon's rank-sum tests or paired t-tests when comparing serial timepoints within the same patients or matched BM and PB samples. When multiple hypotheses were tested, p-values were corrected using the Benjamini-Hochberg approach. For survival analysis, cell type abundance was discretized based on the median.

Differential expression was performed using Wilcoxon's rank-sum tests at the single-cell level and DESeq2 at the pseudobulk level (Love et al., 2014). Genes with more than 3 counts in at least 10 cells were considered for differential expression. For pseudobulk analyses, genes with more than 10 counts in at least 25% of samples were considered for differential expression. P-values were corrected for multiple hypotheses testing using the Benjamini-Hochberg approach.

Cell type and subtype annotation: Gene expression markers used for cell type and subtype annotation can be found in Figure 2. Scraper-normalized expression values were Z-scored and the mean Z-score per cluster was used to plot the heatmaps.

Within the T cell compartment, we identified naïve CD8⁺ (CD8⁺ TN) and CD4⁺ T cells (CD4⁺ TN) (*LEF1*, *TCF7*, *SELL*, *CCR7*), central memory CD8⁺ (CD8⁺ TCM) and CD4⁺ T cells (CD4⁺ TCM) (*FAS*, *IL7R*, *HNRNP1L*), helper type 1 T cells (Th1) (*CD4*, *CXCR3*, *GZMA*, *GZMK*, *LYAR*, *CCL5*), helper type 2 T cells (Th2) (*CD4*, *GATA3*, *KRT1*, *CCR4*), helper type 17 T cells (Th17) (*CD4*, *RORA*, *KLRB1*, *CCR6*, *TNFRSF4*), regulatory T cells (Treg) (*CD4*, *IL2RA*, *FOXP3*, *CTLA4*, *RTKN2*, *IKZF2*), mucosa-associated invariant T cells (MAIT) expressing a mixture of Th17 and Th1 markers, as well as the specific marker *SLC4A10* (MAIT annotation was confirmed by invariant TCR chain usage in overlapping TCR sequencing data), two effector memory CD8⁺ T cell subpopulations, GZMK⁺ CD8⁺ TEM cells (*GZMK*, *CMC1*, *XCL1*, *XCL2*) and terminal effector cells (GZMB⁺ CD8⁺ TEM) (*GZMB*, *GZMH*, *FGFBP2*, *PRF1*, *FCGR3A*, *GNLY*). Furthermore, we identified a subpopulation of CD4⁺ TN cells expressing genes related to response to interferon-gamma (IFN⁺ CD4⁺ TN) (*MX1*, *ISG15*, *IFI6*, *IFI44L*), two separate subpopulations expressing genes related to response to interferon type-I (IFN-I⁺ CD4⁺ TEM & IFN-I⁺ CD8⁺ TEM) (*IFIT1*, *IFIT2*, *IFIT3*, *TNF*), and two activated CD4⁺ subpopulations expressing genes of the AP-1 module and general activation markers (aCD4⁺ TN & aCD4⁺ TCM) (*JUN*, *FOS*, *FOSB*, *CD69*, *DUSP1*, *TSC22D3*). Lastly, we identified a second CD4⁺ regulatory subpopulation, clustering adjacent to Tregs and expressing high levels of *CTLA4* and MHC-II-encoding genes, a profile most consistent with effector Tregs, which we annotated as eTregs.

Within the NK cell compartment, we identified CD56^{bright} NK cells (CD56^{br} NK) (*NCAM1*, *CD2*, *XCL2*, *KLRC1*, *IL7R*, *SELL*, *GZMK*), CD160⁺ NK cells expressing markers of CD56^{bright} NK cells as well as *CD160*, *SPRY2*, and *TOX2*, CD56^{dim} NK cells (CD56^{dim} NK) (*GZMB*, *GZMH*, *FGFBP2*, *PRF1*, *NKG7*, *CX3CR1*, *GNLY*, *B3GAT1*) and gamma-delta T cells (T_{gd}) (*CD3D*, *TRGV9*, *TRDV2*). Furthermore, we identified a subpopulation of NK cells expressing genes related to response to interferon-gamma (IFN⁺

NK) (*MX1*, *ISG15*, *IFI6*, *IFI44L*), a separate subpopulation expressing genes related to response to interferon type-I (IFN-I⁺ NK) (*IFIT1*, *IFIT2*, *IFIT3*, *TNF*), and an activated subpopulation with characteristics of both CD56^{bright} and CD56^{dim} NK cells (aNK), expressing genes of the AP-1 module and general activation markers (*FOS*, *JUN*, *JUNB*, *CD69*, *CXCR4*, *DUSP1*, *NFKB1*, *NFKB2*).

Within the B cell compartment, we identified immature B cells (IBC) (*MME*, *CD19*, *CD38*, *IGHM*, *SOX4*, *TCL1A*, *RAG1*, *RAG2*, *VPREB1*, *IGLL1*), transitional B cells (TBC) (*CD19*, *MS4A1*, *CD38*, *IGHM*, *IGHD*, *TCL1A*, *IL4R*, *SELL*), naïve B cells (NBC) (*CD19*, *MS4A1*, *IGHM*, *IGHD*, *IL4R*, *SELL*, *CCR7*, *FCER2*, *ABCB1*), a population expressing higher levels of *CXCR5*, *CD83*, and *TNFSF9* (encoding 4-1BBL), called germinal center B cells (GBC), marginal zone B cells (MZB) (*CD1C*, *IGHM*, *PLD4*, *LINC01857*), atypical B cells (ABC) (*TBX21*, *ITGAX*, *FCRL5*, *ENC1*, *TNFRSF1B*, *SOX5*, *MPP6*), and two populations of class-switched memory B cells: one expressed higher levels of *TCF7* and was called resting memory B cells (BRM), while the other expressed higher levels of *FAS*, *CD86*, *ITGB1*, *S100A10*, and *TOX*, and was called effector memory B cells (BEM). Lastly, we identified a subpopulation of B cells expressing genes related to response to interferon-gamma (IFN⁺ BC) (*MX1*, *ISG15*, *IFI6*, *IFI44L*).

Within the monocyte compartment, we identified three main subpopulations of classical (CD14⁺) monocytes: a subpopulation expressing higher levels of *SELL*, *VCAN*, *S100A8*, *S100A9*, and *S100A12* (*SELL*⁺ CD14⁺ Mono), a subpopulation expressing higher levels of genes encoding MHC-II peptides (HLA-DR⁺⁺ CD14⁺ Mono), and a subpopulation expressing pro-inflammatory factors, such as *IL1B*, *CXCL8*, *CXCL2*, *CCL3*, *CCL4*, and *CEBPD* (Cytokine⁺ CD14⁺ Mono). A fourth subpopulation expressed *CD14*, *SELL*, *VCAN*, *S100A8*, *S100A9*, *S100A12*, as well as *MPO*, *AZU1*, *ELANE*, *PRTN3*, and *RNASE2*, and was called “Neutrophil-like Mono”. Furthermore, we identified non-classical CD16⁺ Monocytes (*FCGR3A*, *MS4A7*, *CSF1R*, *CDKN1C*, *RHOC*), intermediated CD14⁺ CD16⁺ Monocytes expressing *CD14*, *FCGR3A* (which encodes CD16), and high levels of MHC-II-encoding genes, a CD16⁺ subpopulation of macrophages expressing high levels of complement factor C1q (*C1QA*, *C1QB*, *C1QC*, *SELENOP*, *SDC3*), and a population of monocytes expressing genes related to response to interferon-gamma (IFN⁺ Mono) (*MX1*, *ISG15*, *IFI6*, *IFI44L*).

Within the dendritic cell compartment, we identified canonical dendritic cells type 2 (cDC2) (*FCER1A*, *CLEC10A*, *CD1C* and MHC-II encoding genes), canonical dendritic cells type 1 (cDC1) (*CLEC9A*, *C1orf54*, *IDO1*, *CADM1*, *CLMK*, *BATF3*), monocyte-derived dendritic cells (MoDC) (*FCER1A*, *CLEC10A*, *CD1C*, *CD14*), plasmacytoid dendritic cells (pDCs) (*LILRA4*, *IL3RA*, *GZMB*, *IRF4*, *SERPINF1*), and AXL⁺ SIGLEC6⁺ dendritic cells (AS-DCs) (*AXL*, *SIGLEC6*, *ADAM33*, *LTK*). Furthermore, we identified cycling subpopulations of pDCs and cDC2s (cpDC, ccDC2), activated subpopulations expressing the pro-inflammatory markers *IL1B*, *CXCL8*, and *CCL3* (acDC2, apDC), and a MoDC subpopulation expressing genes related to response to interferon-gamma (IFN⁺ MoDC) (*MX1*, *ISG15*, *IFI6*, *IFI44L*).

Lastly, within the progenitor cell compartment, we identified hematopoietic stem cells (HSC) (*CD34, CD164, BEX1, BEX2, AVP, CRHBP, HLF*), multi-potent progenitor cells (MPP) (*CD34, CD33, MPO, FLT3, MZB1*), granulocyte-monocyte progenitor cells (GMP) (*MPO, AZU1, PRTN3, ELANE, LYZ*), monocyte-dendritic cell progenitors (MDP) (*MPO, LYZ, IRF8, LY86, RUNX2, LILRB4*), megakaryocyte-erythroid progenitors (MEP) (*GATA2, FCER1A, ITAG2B, CSF2RB*), erythroid progenitors (EP) (*GATA1, EPOR, CA1, CA2, EPCAM, KLF1, BLVRB, APOC1, APOE*), megakaryocyte progenitors (MKP) (*GATA2, FCER1A, ITGA2B, PLEK, GP1BB, PPBP, PF4, GP9*), basophil-mast cell progenitors (BMP) (*MS4A2, MS4A3, TPSAB1, TPSB2, HDC, CLC, PRG2*), common lymphoid progenitors (CLP) (*FLT3, TRBC2, MZB1, LTB, JCHAIN, ADA, BCL2*), pro-B cells (Pro-B) (*DNTT, MME, PAX5, RAG1, RAG2*), and pre-B cells (Pre-B) (*VPREB1, IGLL1, JCHAIN, TCL1A*). Within this compartment, we also identified a small population of stroma cells (*CXCL12, LEPR, KITLG, VCAM1, COL1A1, COL1A2, CCL2*).

Gene expression signature discovery & annotation: A Bayesian version of Non-negative Matrix Factorization (NMF) was performed using SignatureAnalyzer on the count matrix of the top 5,000 highly variable genes for gene expression signature discovery (Kasar et al., 2015; Kim et al., 2016). Half-normal priors were imposed for both the W (gene by signature) and the H matrix (cell by signature), as we reasoned that a single gene could contribute towards multiple expression programs and a single cell could be described by multiple expression programs. The tool was run 30 times and the run that resulted in a K (i.e., number of signatures) equal to the mode of the K distribution and had the lowest objective was selected for downstream analysis (Figure S5B). Gene expression signature markers were nominated by (i) multiplying the W matrix by the sum of signature activity across all cells in the H matrix, (ii) calculating the fraction of each signature's activity for each gene (matrix F) and (iii) ranking genes based on the product of W (i.e., how strongly each gene contributes to the signature) and F (i.e., how strongly each signature contributes to the gene). A total of 26 gene expression (GEX) signatures were extracted (Figure S5C): GEX-1 corresponded to cytokine signaling associated with GZMK-expressing subpopulations (*XCL1, XCL2, CCL3, CCL4*); GEX-2 captured naïve T cells (*TCF7, LEF1, TXNIP, IL7R, LTB, NOSIP*); GEX-3 captured dendritic cells (*CLEC10A, FCER1A, CD74, HLA-DRA, HLA-DRB1*); GEX-4 captured B cells (*CD19, MS4A1, IGHM, IGHD, CD79A, CD79B*); GEX-5 was associated with ferritin heavy and light chain expression (*FTH1, FTL*) and was primarily active in myeloid cells; GEX-6 corresponded to pro-inflammatory signaling (*IL1B, CXCL8, CXCL2, CEBPD*) and was primarily active in myeloid cells; GEX-7 corresponded to genes induced by interferon-gamma (*MX1, STAT1, XAF1, IFI6, IFI44L, ISG15*) and was active in T cells, B cells, NK cells, and monocytes; GEX-8 was associated with thymosin expression and the alpha and beta chains of the TCR (*TMSB4X, TMSB10, TRAC, TRBC1*), and was primarily active in T cells; GEX-9 was associated with histone 1 gene expression (*HIST1H1C, HIST1H1D, HIST1H1E*) and was primarily active in dendritic cells; GEX-10 was associated with metabolic programming (*GLUL, SLC43A2, SLC25A37*) and hypoxia-related signaling (*HIF1A, TIMP1, NEAT1*), and was primarily active in myeloid cells; GEX-11 corresponded to CD14⁺ monocytes (*CD14, MS4A6A, FCN1, LYZ*); GEX-12 corresponded to immature B cells (*ACSM3, IGLL1, VPREB1, TCL1A, MME, SOX4*); GEX-13 corresponded to pDCs (*LILRA4, JCHAIN, SERPINF1*,

PTGDS, ITM2C, IRF4, IRF8); GEX-14 corresponded to a different program within classical monocytes (*S100A8, S100A9, S100A12, VCAN*); GEX-15 corresponded to Th17 cells and MAITs (*RORA, KLRB1, CCR6*); GEX-16 was associated with interferon type-I signaling (*IFIT1, IFIT2, IFIT3, PMAIP1, OASL, TNF, HERC5, ZC3HAV1*); GEX-17 was associated with T cell surface antigens (*CD5, CD6, CD7*); GEX-18 and GEX-19 were associated with genes of the AP-1 module (*JUN, FOS, FOSB*) and general activation markers (*DUSP1, TSC22D3, NFKBIA*); GEX-20 corresponded to CD16⁺ monocytes (*FCGR3A, CDKN1C, SERPINA1, MS4A7*); GEX-21 corresponded to cytotoxic T cells (*GZMK, GZMH, NKG7, CD8A, CD8B*); GEX-22 corresponded to GZMB⁺ cytotoxic T cells and NK cells (*GZMB, GNLY, SPON2, CLIC3, FGFBP2*); GEX-23 was associated with genes frequently expressed by myeloid cells (*AHNAK, TLR2, CYBB*) and was primarily active in monocytes; GEX-24 was associated with memory CD4⁺ T cell expression markers (*GATA3, CD28, PRDM1, ITGB1, ANXA1*) and was primarily active in those populations; GEX-25 and GEX-26 were associated with cytoskeleton-related genes (*ACTB, ACTG1, PFN1, CFL*) and were primarily active in myeloid cells.

For survival analysis, each sample was assigned its mean signature activity across cells, and mean activity was discretized based on the median.

Naïve Bayes Classifier and Normalization Scores: We trained a Naïve Bayes classifier using the R package caret on a training set (n=41) comprising BM samples from patients with HRSMM and HD, using 10-fold cross-validation and 100 iterations, following a parameter sweep (Kuhn et al., 2008). The input to the classifier was the composition matrix of cell type proportions and included data on 63 subpopulations. The accuracy in the testing set (*n*=16) was 94% (95% CI: 69.8%–99.8%, *p*=0.0057) (Figure S5A). For each cell type, we compute its importance to the classification using the varImp() function, which employs an ROC approach, and signed it based on whether the cell type had a higher mean abundance in disease (minus) or HD (plus). We then computed the weighted sum of the product of each cell type's proportion and its signed importance (normalization score), and classified patients as reactive (least normal-like) or non-reactive (most normal-like), based on the median normalization score of patients at BL.

Single-cell TCR sequencing analysis—CellRanger v5.0.1 was used to extract FASTQ files and produce clonotype matrices (Zheng et al., 2017). When multiple alignments were called for a single chain, the alignment with the most UMIs was selected, and when multiple chains were called for a single cell barcode, the top two chains in terms of UMI counts were selected.

To control for variability in T cell numbers across samples, we downsampled a range of T cell numbers (n=75–500) from each sample with 100 iterations and averaged repertoire diversity estimates across iterations. Repertoire diversity was assessed using the Chao index, as implemented in the vegan R package. To estimate the proportion of clonotypes belonging to a given clone size (Rare: ≤1%; Small: >1% and ≤5%; Medium: ≥5% and <10%; Large: ≥10%) in a given sample, we downsampled 100 T cells from each sample with 100 iterations, converted clonotype counts into frequencies, computed the frequency of each clone size, and subsequently averaged the frequency of each clone size across iterations

and renormalized those frequencies for plotting. To estimate the probability of a given T cell subtype to be clonally expanded in patients or HD, we first computed stable clone sizes for each clonotype in each sample, by downsampling 100 T cells with 100 iterations, converting clonotype counts into frequencies, computing each clonotype's clone size, and subsequently assigning each clonotype the mode of its clone size distribution across all iterations. Next, for each T cell subtype, we randomly sampled 100 cells across all patients (or HD), and for each iteration, we computed the frequency of rare (proportion of singletons) and expanded (1-rare) clonotypes. For visualization purposes, the average frequencies were renormalized and plotted. P-values were computed by bootstrapping with 10,000 iterations.

CyTOF analysis—The data was normalized using the FCS Processing tab of the Fluidigm CyTOF Software 7.0.8493. Data analysis was manually performed using FlowJo 10.7.1. For initial data clean-up, cell events were gated to remove dead cells and debris through biaxial plots of Time vs. Event Length, Beads (for removal of the EQ Calibration Beads), and Gaussian-derived parameters (Residual, Width, Offset). The viability stain 103Rh was used to gate out dead cells on PBMC populations. All viable cells were backgated on both DNA parameters (191Ir and 193Ir) to ensure no doublets were included. Please refer to Hallisey et al. for the detailed protocol (Hallisey et al., 2022).

ADDITIONAL RESOURCES

The clinical trial study design and record can be found at: <https://clinicaltrials.gov/ct2/show/NCT02279394>

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors would like to thank the patients who participated in the E-PRISM trial. Additionally, the authors would like to thank Anna V. Justis, PhD, for medical editing support, and Sarah Nersersian, MSc, and Tiana Issa for illustration support. This research was supported by a Stand Up To Cancer-Bristol-Meyers Squibb Catalyst Research Grant (Grant Number: SU2C-AACR-CT05-17). Stand Up To Cancer is a division of the Entertainment Industry Foundation. Research Grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. Furthermore, this research was supported in part by the National Institutes of Health (NIH R01 CA205954), the Leukemia and Lymphoma Society (LLS, Career Development Program 2316-17), the Multiple Myeloma Research Foundation (MMRF), and the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation. R.S.P. is supported by the MMRF Research Fellowship Award, the International Waldenström's Macroglobulinemia Foundation's Robert A. Kyle Award, the Claudia Adams-Barr Award for Innovative Basic Cancer Research, and the International Myeloma Society. G.G. is partially supported by the Paul C. Zamecnik Chair in Oncology at Massachusetts General Hospital Cancer Center. S.Z.U. was supported by a career development award from the LLS.

N.J.H. is a consultant for Constellation Pharmaceuticals. F.A. is an employee of Illumina Inc. O.Z. is an employee of Ikena Oncology and a stockholder in Ikena Oncology and Morphosys AG. G.G. receives research funds from IBM and Pharmacyclis and is an inventor on patent applications filed by the Broad Institute related to MSMuTect, MSMutSig, POLYSOLVER, SignatureAnalyzer-GPU, and MSIDetect. He is also a founder, consultant, and holds privately held equity in Scorpion Therapeutics. I.M.G. has a consulting or advisory role with AbbVie, Adaptive, Amgen, Aptitude Health, Bristol Myers Squibb, GlaxoSmithKline, Huron Consulting, Janssen, Menarini Silicon Biosystems, Oncopeptides, Pfizer, Sanofi, Sognef, Takeda, The Binding Site, and Window Therapeutics; has received speaker fees from Vor Biopharma and Veeva Systems, Inc.; and her spouse is the CMO and equity holder of Disc Medicine. S.M. has a consulting role with Abbvie, Adaptive Biotechnology, Amgen, Celgene/BMS, GlaxoSmithKline, Janssen, Novartis, Oncopeptides, Regeneron, Roche, Takeda, and has received research funding from Abbvie, Adaptive Biotechnology, Amgen, Celgene/BMS, GlaxoSmithKline, Janssen, Novartis, Oncopeptides,

Regeneron, Roche, Takeda. A.J.Y. has a consulting role with Adaptive Biotechnologies, Amgen, BMS, Celgene, GSK, Janssen, Karyopharm, Oncopeptides, Sanofi, Takeda, and has received research funding from Amgen, Janssen, and Takeda. M.B. is a consultant for Sanofi, Genzyme, and Janssen, and has received research funding from MedImmune, Janssen, Legend Biotech, Amgen, Celularity, Bristol Myers Squibb, Celgene, Bluebird bio, Millennium, Takeda, Cerecor (currently Avalo therapeutics), and C4 Therapeutics. M.B has an advisory role and received honoraria from Bristol Myers Squibb, Takeda, Janssen, and Menarini. T.H.M. received advisory board fees from Legend Biotech. R.S.P., G.G., and I.M.G. are co-inventors on a patent application related to this work (PCT/US22/74839).

REFERENCES

- Weiss BM, Abadie J, Verma P, Howard RS, and Kuehl WM (2009). A monoclonal gammopathy precedes multiple myeloma in most patients. *Blood* 113, 5418–5422. 10.1182/blood-2008-12-195008. [PubMed: 19234139]
- Kyle RA, Remstein ED, Therneau TM, Dispenzieri A, Kurtin PJ, Hodnefield JM, Larson DR, Plevak MF, Jelinek DF, Fonseca R, et al. (2007). Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N Engl J Med* 356, 2582–2590. 10.1056/NEJMoa070389. [PubMed: 17582068]
- Mateos MV, Kumar S, Dimopoulos MA, Gonzalez-Calle V, Kastritis E, Hajek R, De Larrea CF, Morgan GJ, Merlini G, Goldschmidt H, et al. (2020). International Myeloma Working Group risk stratification model for smoldering multiple myeloma (SMM). *Blood Cancer J* 10, 102. 10.1038/s41408-020-00366-3. [PubMed: 33067414]
- Visram A., Cook J., and Warsame R. (2021). Smoldering multiple myeloma: evolving diagnostic criteria and treatment strategies. *Hematology Am Soc Hematol Educ Program* 2021, 673–681. 10.1182/hematology.2021000304. [PubMed: 34889380]
- Mateos MV, Hernandez MT, Giraldo P, de la Rubia J, de Arriba F, Lopez Corral L, Rosinol L, Paiva B, Palomera L, Bargay J, et al. (2013). Lenalidomide plus dexamethasone for high-risk smoldering multiple myeloma. *N Engl J Med* 369, 438–447. 10.1056/NEJMoa1300439. [PubMed: 23902483]
- Mateos MV, Hernandez MT, Salvador C, Rubia J, de Arriba F, Lopez-Corral L, Rosinol L, Paiva B, Palomera L, Bargay J, et al. (2022). Lenalidomide-dexamethasone versus observation in high-risk smoldering myeloma after 12 years of median follow-up time: A randomized, open-label study. *Eur J Cancer* 174, 243–250. 10.1016/j.ejca.2022.07.030. [PubMed: 36067617]
- Lonial S, Jacobus S, Fonseca R, Weiss M, Kumar S, Orlowski RZ, Kaufman JL, Yacoub AM, Buadi FK, O'Brien T, et al. (2020). Randomized Trial of Lenalidomide Versus Observation in Smoldering Multiple Myeloma. *J Clin Oncol* 38, 1126–1137. 10.1200/JCO.19.01740. [PubMed: 31652094]
- Kim EB, Yee AJ, and Raje N.(2020). Treatment of Smoldering Multiple Myeloma: Ready for Prime Time? *Cancers (Basel)* 12. 10.3390/cancers12051223.
- Mateos MV, and San Miguel JF (2015). Smoldering multiple myeloma: when to observe and when to treat? *Am Soc Clin Oncol Educ Book*, e484–492. 10.14694/EdBook_AM.2015.35.e484. [PubMed: 25993213]
- Perez-Persona E., Vidriales MB., Mateo G., Garcia-Sanz R., Mateos MV., de Coca AG., Galende J., Martin-Nunez G., Alonso JM., de Las Heras N., et al. (2007). New criteria to identify risk of progression in monoclonal gammopathy of uncertain significance and smoldering multiple myeloma based on multiparameter flow cytometry analysis of BM plasma cells. *Blood* 110, 2586–2592. 10.1182/blood-2007-05-088443. [PubMed: 17576818]
- Bustoros M, Sklavenitis-Pistofidis R, Park J, Redd R, Zhitomirsky B, Dunford AJ, Salem K, Tai YT, Anand S, Mouhieddine TH, et al. (2020). Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *J Clin Oncol* 38, 2380–2389. 10.1200/JCO.20.00437. [PubMed: 32442065]
- Misund K, Keane N, Stein CK, Asmann YW, Day G, Welsh S, Van Wier SA, Riggs DL, Ahmann G, Chesi M, et al. (2020). MYC dysregulation in the progression of multiple myeloma. *Leukemia* 34, 322–326. 10.1038/s41375-019-0543-4. [PubMed: 31439946]
- Boyle EM, Deshpande S, Tytarenko R, Ashby C, Wang Y, Bauer MA, Johnson SK, Wardell CP, Thanendrarajan S, Zangari M, et al. (2021). The molecular make up of smoldering myeloma highlights the evolutionary pathways leading to multiple myeloma. *Nat Commun* 12, 293. 10.1038/s41467-020-20524-2. [PubMed: 33436579]

- Bolli N, Maura F, Minvielle S, Gloznik D, Szalat R, Fullam A, Martincorena I, Dawson KJ, Samur MK, Zamora J, et al. (2018). Genomic patterns of progression in smoldering multiple myeloma. *Nat Commun* 9, 3363. 10.1038/s41467-018-05058-y. [PubMed: 30135448]
- Oben B, Froyen G, Maclachlan KH, Leongamornlert D, Abascal F, Zheng-Lin B, Yellapantula V, Derkach A, Geerdens E, Diamond BT, et al. (2021). Whole-genome sequencing reveals progressive versus stable myeloma precursor conditions as two distinct entities. *Nat Commun* 12, 1861. 10.1038/s41467-021-22140-0. [PubMed: 33767199]
- Zavidij O., Haradhvala NJ., Mouhieddine TH., Sklavenitis-Pistofidis R., Cai S., Reidy M., Rahmat M., Flaifel A., Ferland B., Su NK., et al. . (2020). Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat Cancer* 1, 493–506. 10.1038/s43018-020-0053-3. [PubMed: 33409501]
- Bailur JK, McCachren SS, Doxie DB, Shrestha M, Pendleton K, Nooka AK, Neparidze N, Parker TL, Bar N, Kaufman JL, et al. (2019). Early alterations in stem-like/resident T cells, innate and myeloid cells in the BM in preneoplastic gammopathy. *JCI Insight* 5. 10.1172/jci.insight.127807.
- Dhodapkar MV, Krasovsky J, Osman K, and Geller MD (2003). Vigorous premalignancy-specific effector T cell response in the BM of patients with monoclonal gammopathy. *J Exp Med* 198, 1753–1757. 10.1084/jem.20031030. [PubMed: 14638846]
- Das R, Strowig T, Verma R, Koduru S, Hafemann A, Hopf S, Kocoglu MH, Borsotti C, Zhang L, Branagan A, et al. (2016). Microenvironment-dependent growth of preneoplastic and malignant plasma cells in humanized mice. *Nat Med* 22, 1351–1357. 10.1038/nm.4202. [PubMed: 27723723]
- Rajkumar SV, Landgren O, and Mateos MV (2015). Smoldering multiple myeloma. *Blood* 125, 3069–3075. 10.1182/blood-2014-09-568899. [PubMed: 25838344]
- Galletti G, De Simone G, Mazza EMC, Puccio S, Mezzanotte C, Bi TM, Davydov AN, Metsger M, Scamardella E, Alvisi G, et al. (2020). Two subsets of stem-like CD8(+) memory T cell progenitors with distinct fate commitments in humans. *Nat Immunol* 21, 1552–1562. 10.1038/s41590-020-0791-5. [PubMed: 33046887]
- Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, Kumar S, Hillengass J, Kastiris E, Richardson P, et al. (2014). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol* 15, e538–548. 10.1016/S1470-2045(14)70442-5. [PubMed: 25439696]
- Paiva B., Mateos MV., Sanchez-Abarca LI., Puig N., Vidriales MB., Lopez-Corral L., Corchete LA., Hernandez MT., Bargay J., de Arriba F., et al. . (2016). Immune status of high-risk smoldering multiple myeloma patients and its therapeutic modulation under LenDex: a longitudinal analysis. *Blood* 127, 1151–1162. 10.1182/blood-2015-10-662320. [PubMed: 26668134]
- Guyot P, Ades AE, Ouwens MJ, and Welton NJ (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 12, 9. 10.1186/1471-2288-12-9. [PubMed: 22297116]
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, and Getz G.(2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31, 213–219. 10.1038/nbt.2514. [PubMed: 23396013]
- Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, and Getz G.(2011). ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* 27, 2601–2602. 10.1093/bioinformatics/btr446. [PubMed: 21803805]
- Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, and Cheetham RK (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817. 10.1093/bioinformatics/bts271. [PubMed: 22581179]
- Costello M., Pugh TJ., Fennell TJ., Stewart C., Lichtenstein L., Meldrim JC., Fostel JL., Friedrich DC., Perrin D., Dionne D., et al. . (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* 41, e67. 10.1093/nar/gks1443. [PubMed: 23303777]
- Taylor-Weiner A, Stewart C, Giordano T, Miller M, Rosenberg M, Macbeth A, Lennon N, Rheinbay E, Landau DA, Wu CJ, and Getz G.(2018). DeTiN: overcoming tumor-in-normal contamination. *Nat Methods* 15, 531–534. 10.1038/s41592-018-0036-9. [PubMed: 29941871]

- Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152, 714–726. 10.1016/j.cell.2013.01.019. [PubMed: 23415222]
- Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, and Getz G.(2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501. 10.1038/nature12912. [PubMed: 24390350]
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30, 413–421. 10.1038/nbt.2203. [PubMed: 22544022]
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, and DePristo MA (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297–1303. 10.1101/gr.107524.110. [PubMed: 20644199]
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, and Cunningham F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122. 10.1186/s13059-016-0974-4. [PubMed: 27268795]
- Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, Meyerson M, and Getz G.(2015). Oncotator: cancer variant annotation tool. *Hum Mutat* 36, E2423–2429. 10.1002/humu.22771. [PubMed: 25703262]
- Zheng GX., Terry JM., Belgrader P., Ryvkin P., Bent ZW., Wilso R., Ziraldo SB., Wheeler TD., McDermott GP., Zhu J., et al. . (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049. 10.1038/ncomms14049. [PubMed: 28091601]
- Young MD, and Behjati S.(2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* 9. 10.1093/gigascience/giaa151.
- Wolock SL, Lopez R, and Klein AM (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* 8, 281–291 e289. 10.1016/j.cels.2018.11.005. [PubMed: 30954476]
- Germain PL, Lun A, Garcia Meixide C, Macnair W, and Robinson MD (2021). Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* 10, 979. 10.12688/f1000research.73600.2. [PubMed: 35814628]
- Bais AS, and Kostka D.(2020). scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* 36, 1150–1158. 10.1093/bioinformatics/btz698. [PubMed: 31501871]
- Lun AT, McCarthy DJ, and Marioni JC (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 5, 2122. 10.12688/f1000research.9501.2. [PubMed: 27909575]
- Korsunsky I, Millard N., Fan J., Slowikowski K., Zhang F., Wei K., Baglaenko Y., Brenner M., Loh PR., and Raychaudhuri S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16, 1289–1296. 10.1038/s41592-019-0619-0. [PubMed: 31740819]
- Love MI, Huber W, and Anders S.(2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. 10.1186/s13059-014-0550-8. [PubMed: 25516281]
- Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun* 6, 8866. 10.1038/ncomms9866. [PubMed: 26638776]
- Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea A, and Getz G.(2016). Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet* 48, 600–606. 10.1038/ng.3557. [PubMed: 27111033]
- Kuhn M.(2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. 10.18637/jss.v028.i05 [PubMed: 27774042]

Hallisey M, Dennis J, Abrecht C, Pistofidis RS, Schork AN, Lightbody ED, Heilpern-Mallory D, Severgnini M, Ghobrial IM, Hodi FS, and Baginska J.(2022). Mass cytometry staining for human BM clinical samples. STAR Protoc 3, 101163. 10.1016/j.xpro.2022.101163. [PubMed: 35243367]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HIGHLIGHTS

- EloLenDex is associated with 4-year PFS of 89% in patients with high-risk SMM.
- Immune reactivity and post-therapy immune normalization (PIN) are associated with longer PFS.
- Higher abundance of GZMK⁺ cytotoxic T cells is associated with longer PFS.
- Blood-based immune profiling detects immune dysregulation associated with disease.

p-values were computed using Cox proportional hazards regression. F) KM curve of PFS in the E-PRISM cohort, stratified based on the presence of del17p ($n=33$). See also Figures S1, S2, and Table S1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

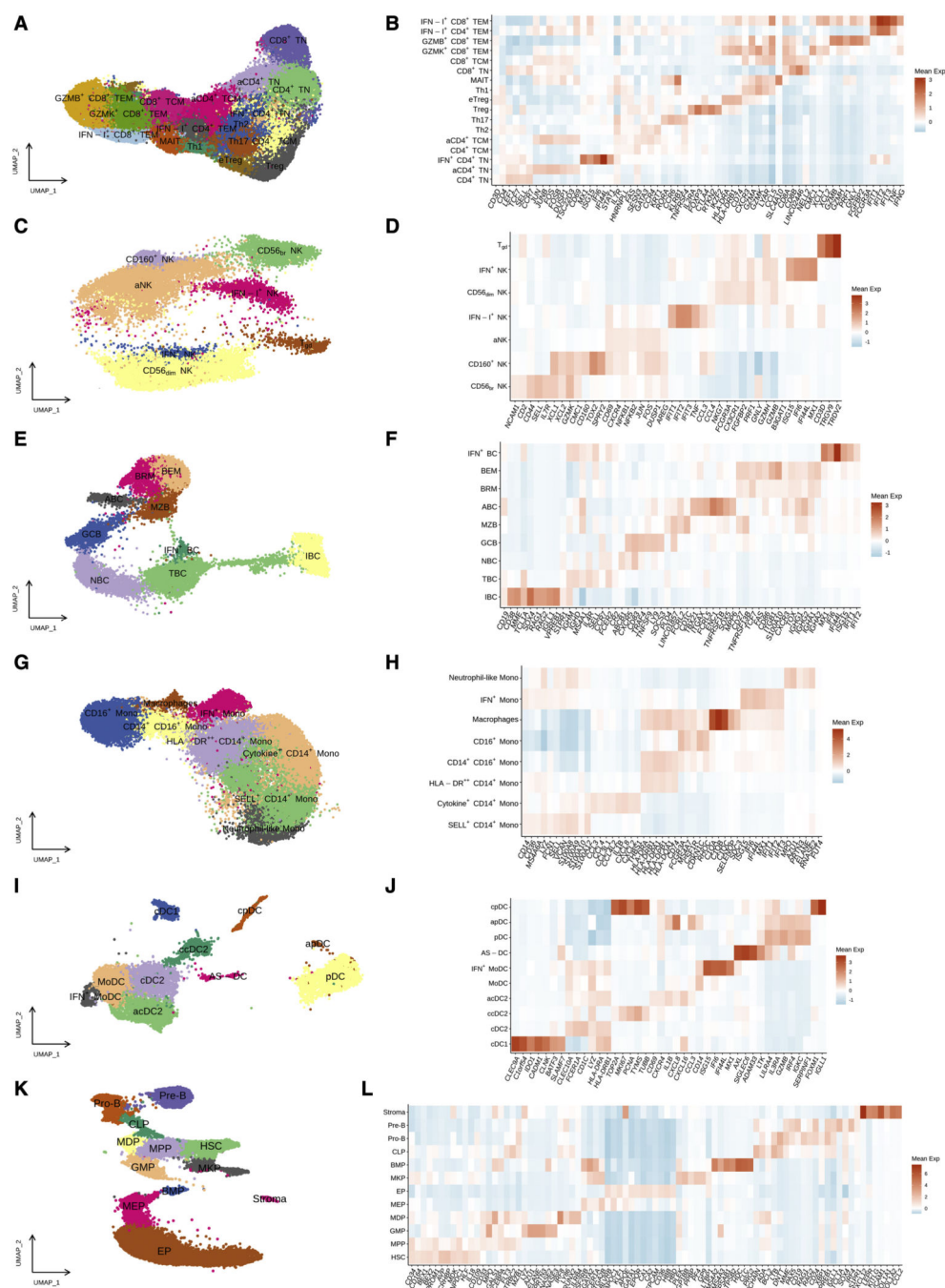
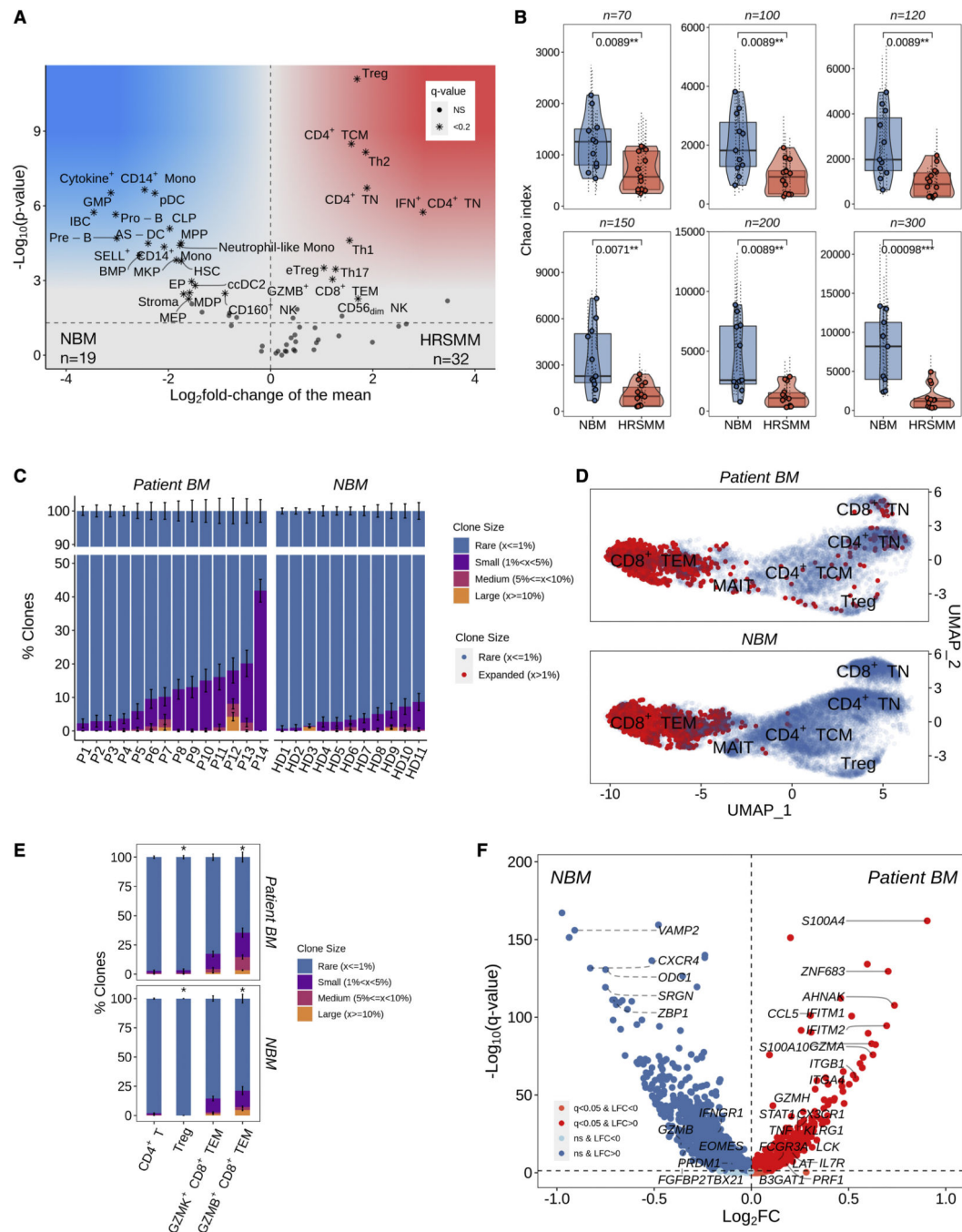
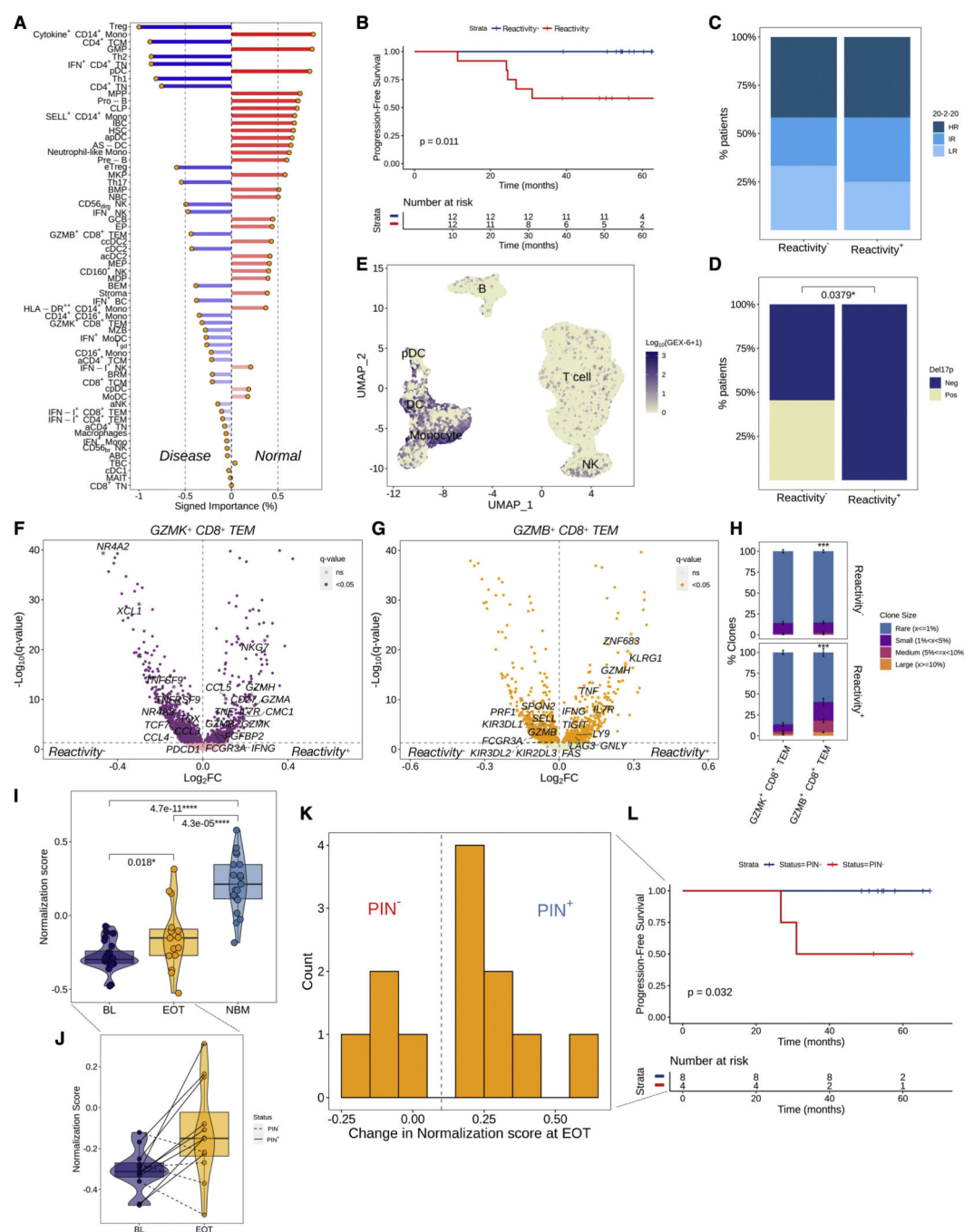


Figure 2. BM and PB immune cell populations in the E-PRISM cohort. UMAP embeddings and heatmaps of gene expression markers (Mean Z-score of normalized expression) in T cells (A, B), NK cells (C, D), B cells (E, F), Monocytes (G, H), Dendritic cells (I, J), and Progenitor cells (K, L).

**Figure 3.**

Comprehensive profiling of changes in BM immune cell composition and TCR repertoire in patients with HRSMM. A) Volcano plot of proportion changes in the BM of patients with HRSMM (n=32) compared to HD (NBM, n=19) with at least 100 cells overall. P-values were computed with Wilcoxon's rank-sum test and corrected using the Benjamini-Hochberg approach. Cell types with a q-value of <0.2 were marked with stars. B) Boxplots, violin plots, and scatter plots comparing BM TCR repertoire diversity, as assessed by the Chao index, between patients with HRSMM (n=14) and healthy individuals (NBM, n=11) given

different numbers of downsampled cells. Each data point represents the average diversity estimate across 100 random samples of the given size for a single sample; the range of diversity estimates across all iterations for each sample is visualized in error bars (dotted line). Violin outline width represents density. P-values were computed with Wilcoxon's rank-sum tests. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). C) Barplots showing the proportion of T cells in a given BL patient sample (P, n=14) or sample from a HD (n=11) that were determined to belong to one of four clone size categories (Rare: $\leq 1\%$; Small: $>1\%$ and $<5\%$; Medium: $\geq 5\%$ and $<10\%$; Large: $\geq 10\%$) through iterative (n=100) downsampling of 100 cells. The average proportion per clone size category was visualized and the standard deviation across iterations was depicted in solid-line error bars. D) UMAP embedding of HD BM (NBM) and patient BM T cells at BL with matched TCR data. T cells belonging to rare clonotypes (with a frequency of $\leq 1\%$) were colored in blue, while T cells belonging to expanded clonotypes (with a frequency of $>1\%$) were colored in red. E) Barplots showing the proportion of clonotypes in a given T cell subtype across all patients (n=14) or HD (n=11) that belonged to one of the four clone size categories. For each T cell subtype, 100 cells were randomly sampled 100 times from all patients or HD, and the proportion of expanded (1-Rare) clonotypes was compared between patients and HD using bootstrapping with 10,000 iterations. The average proportion per clone size category was visualized and the standard deviation across iterations was depicted in solid-line error bars. F) Volcano plot highlighting genes that are highly expressed in expanded GZMB⁺ CD8⁺ TEM cells from patients (n=743, in red) compared to expanded GZMB⁺ CD8⁺ TEM cells from HD (n=1,074, in blue). See also Figures S3 and S4.

**Figure 4.**

Immune reactivity at BL and post-therapy immune normalization are associated with significantly longer PFS in patients with HRSM under treatment. A) Barplot visualizing the signed importance of each cell type towards the classification. B) Kaplan-Meier (KM) curve of PFS in the E-PRISM cohort, stratified based on the median normalization score (Reactivity⁺: > median) (n=24). C, D) Barplots visualizing the frequency of 20-20 risk stages (C) and del17p (D) by reactivity status (LR: low-risk, IR: intermediate-risk, HR: high-risk). The p-value was computed with a Fisher's exact test. E) UMAP embedding of

lymphocytes and myeloid cells colored by the log-scaled activity of signature GEX-6. F, G) Volcano plots visualizing genes that are differentially expressed in GZMK⁺ CD8⁺ TEM (F) and GZMB⁺ CD8⁺ TEM (G) cells of reactive patients (n=12) compared to non-reactive (n=12). H) Barplot visualizing the proportion of clonotypes belonging to each of four clone size categories per cytotoxic T cell subtype in BM samples from reactive (n=7) or non-reactive patients (n=7). For each T cell subtype, 100 cells were randomly sampled 100 times from patients with or without reactivity, and the proportion of expanded (1-Rare) clonotypes was compared between the two using bootstrapping with 10,000 iterations. The average proportion per clone size category was visualized and the standard deviation across iterations was depicted in solid-line error bars. I) Boxplots, violin plots, and scatter plots visualizing normalization scores in patient BM samples drawn at BL (n=28) or EOT (n=16), and in HD BM samples (NBM, n=22). Violin outline width represents density. P-values were computed using a paired t-test for paired patient samples or Wilcoxon's rank-sum test for comparisons between patients and HD. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). J) Boxplots, violin plots, and scatter plots visualizing paired normalization scores at BL and EOT from patients with HRSMM (n=12). Violin outline width represents density. Solid lines connect samples from patients classified as PIN⁺; dashed lined connect samples from patients classified as PIN. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). K) Histogram of the distribution of change in normalization scores between paired BL and EOT BM patient samples (n=12). The dashed line corresponds to the threshold used to determine the presence of PIN. L) KM curve of PFS in the E-PRISM cohort, stratified based on PIN status (n=12). See also Figure S5.

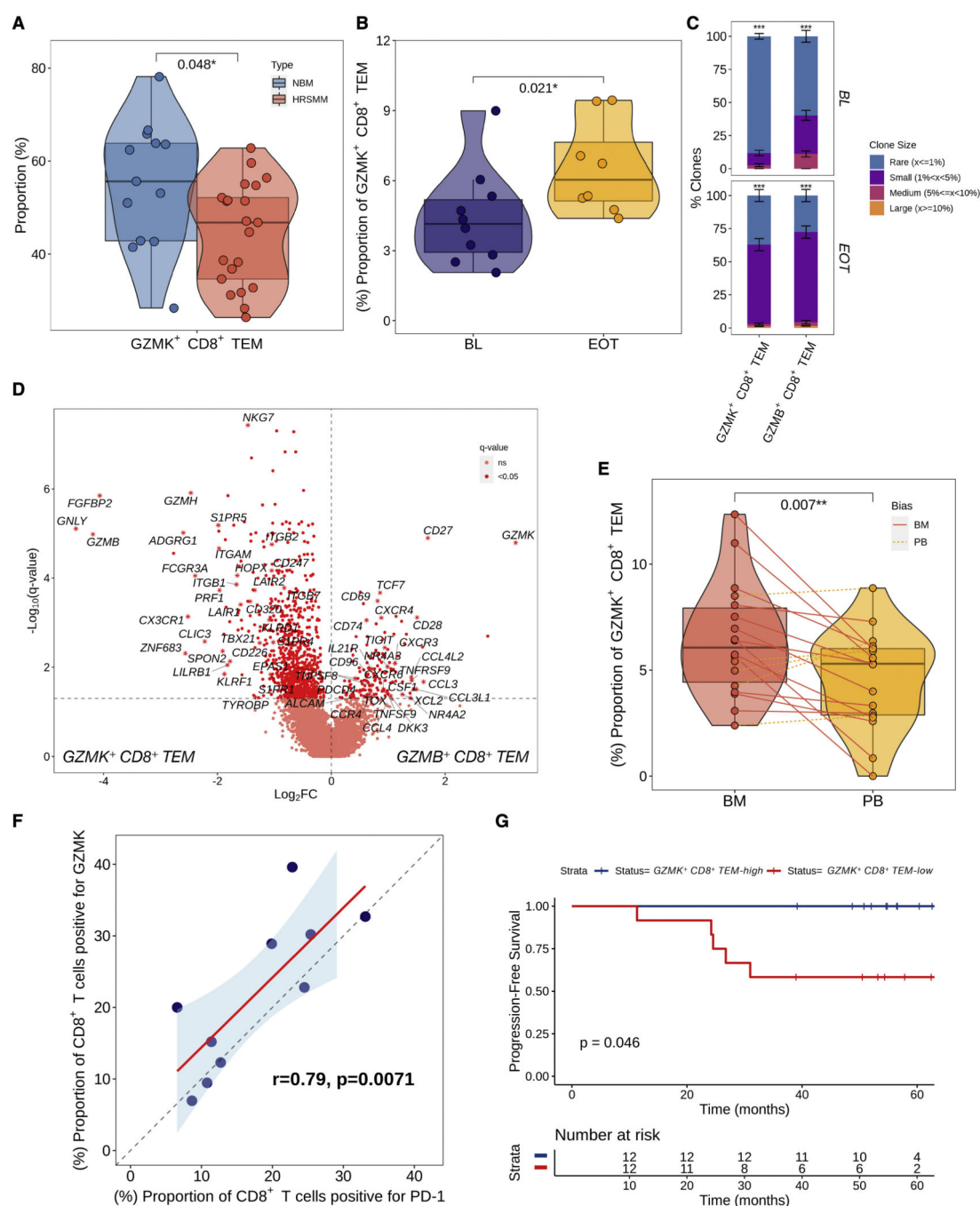
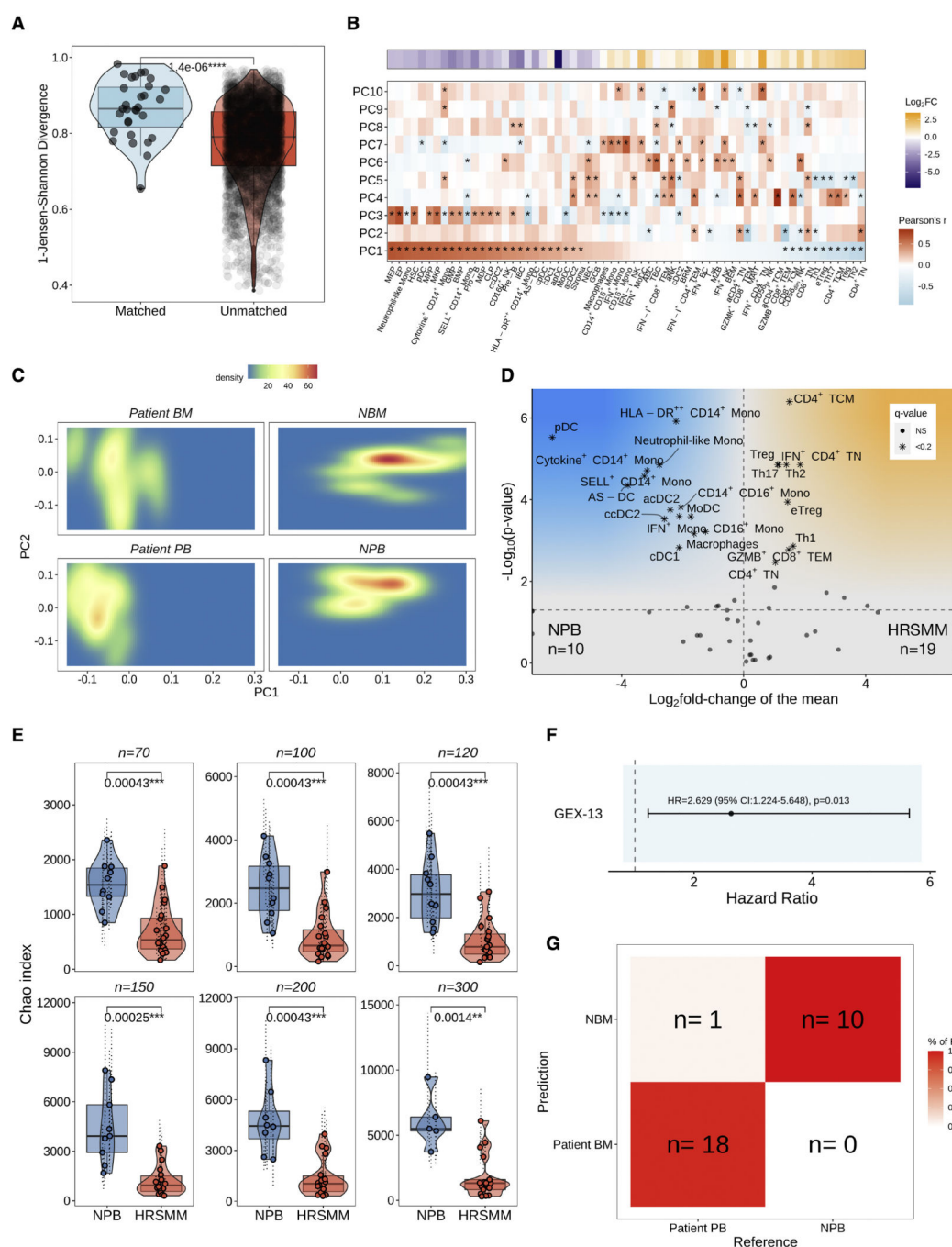


Figure 5. A higher abundance of Granzyme K-expressing CD8⁺ T cells is associated with longer PFS in patients with HRSMM treated with EloLenDex. A) Boxplots, violin plots, and scatter plots visualizing the relative abundance of GZMK⁺ CD8⁺ TEM and GZMB⁺ CD8⁺ TEM out of all cytotoxic T cells in patient BM (n=33) at BL compared to samples from HD (NBM, n=22). Violin outline width represents density. P-values were computed with Wilcoxon's rank-sum test. (Box: 1st quartile, median, 3rd quartile; whiskers: +/- 1.5*IQR). B) Boxplots, violin plots, and scatter plots visualizing the abundance of GZMK⁺ CD8⁺

TEM cells by CyTOF in patient BM samples drawn at BL (n=10) and EOT (n=7). Violin outline width represents density. The p-value was computed with Wilcoxon's rank-sum test, as these samples were not paired. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). C) Barplot visualizing the proportion of clonotypes belonging to each of four clone size categories per cytotoxic T cell subtype in patient PB samples drawn at BL (n=22) or EOT (n=17). For each T cell subtype, 100 cells were randomly sampled 100 times from patients at BL and EOT, and the proportion of expanded (1-Rare) clonotypes was compared between the two using bootstrapping with 10,000 iterations. The average proportion per clone size category was visualized and the standard deviation across iterations was depicted in solid-line error bars. D) Volcano plot visualizing genes that are differentially expressed between GZMK⁺ CD8⁺ TEM and GZMB⁺ CD8⁺ TEM cells from patient BM samples drawn at BL (n=28). E) Boxplots, violin plots, and scatter plots comparing the abundance of GZMK⁺ CD8⁺ TEM cells between paired BM and PB samples drawn at BL from patients with HRSMM (n=22). Violin outline width represents density. Solid red lines connect samples that are enriched in the BM, while dashed orange lines connect samples that are enriched in the PB. The p-value was computed using a paired t-test. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). F) Scatter plot visualizing the correlation between the BL proportion of CD8⁺ T cells expressing PD-1 and those expressing GZMK by CyTOF (n=10). A regression line was fitted (in red) and the correlation coefficient and p-value were computed using Pearson's approach. G) Kaplan-Meier curve of PFS in the E-PRISM cohort, stratified based on the median abundance of GZMK⁺ CD8⁺ TEM cells in patient BM samples at BL (n=26). See also Figure S6.

**Figure 6.**

PB-based immune profiling accurately detects alterations in immune cell composition and TCR repertoire diversity observed in the BM. A) Boxplots, violin plots, and scatter plots showing the Jensen-Shannon divergence between matched BM and PB samples (n=22), compared to unmatched samples. Violin outline width represents density. The p-value was computed using Wilcoxon's rank-sum test. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). B) Heatmap of Pearson's correlation coefficient (r) between BM immune cell abundance and the first 10 principal components. The x-axis was sorted by

decreasing PC1 value. The top panel shows the $\log_2$ fold-change in abundance between BM samples from patients with HRSMM (n=32) and those from HD (n=19). P-values were corrected using the Benjamini-Hochberg approach and stars correspond to pairs with significant ($q < 0.05$) correlation. C) Two-dimensional density plot of BM and PB samples from patients (BM: n=26, PB: n=29) or HD (BM: n=22, PB: 10) according to PC1 and PC2. D) Volcano plot of proportion changes in the PB of patients with HRSMM (n=19) compared to HD (NPB, n=10) with at least 100 cells overall. P-values were computed with Wilcoxon's rank-sum test and corrected using the Benjamini-Hochberg approach. Cell types with a q-value of < 0.2 were marked with stars. E) Boxplots, violin plots, and scatter plots comparing PB TCR repertoire diversity, as assessed by the Chao index, between patients with HRSMM (n=22) and healthy individuals (NPB, n=10) given different numbers of downsampled cells. Each data point represents the average diversity estimate across 100 random samples of the given size for a single sample; the range of diversity estimates across all iterations for each sample is visualized in error bars (dotted line). Violin outline width represents density. P-values were computed with Wilcoxon's rank-sum tests. (Box: 1st quartile, median, 3rd quartile; whiskers: $\pm 1.5 \times \text{IQR}$). F) Forest plot showing the effect of mean BL GEX-13 activity in the PB on PFS. Hazard ratio, 95% confidence interval, and p-value were computed using Cox proportional hazards regression. G) Confusion matrix visualizing the accuracy of a Naïve Bayes classifier trained on BM samples from patients and HD (training set, n=41) and tested in PB samples (n=29). See also Figure S7.

Table 1:

Baseline characteristics of patients enrolled in the E-PRISM study.

	Treatment arm			
	Total	A: Elo/Len/Dex	B: Elo/Len	<i>p-value</i>
	n = 51 (%)	n = 40 (78)	n = 11 (22)	
<u>Aae at registration</u>				
Median (range)	62 (29 – 79)	62 (29 – 79)	62 (44 – 75)	0.65
<u>Sex</u>				
Female	32 (63)	26 (65)	6 (55)	0.73
Male	19 (37)	14 (35)	5 (45)	
<u>Race</u>				
Black or African American	7 (14)	7 (18)	-	0.59
White	42 (82)	31 (78)	11 (100)	
<u>Ethnicity</u>				
Non-Hispanic	48 (94)	37 (92)	11 (100)	
<u>ECOG PS</u>				
00 - Fully Active	34 (67)	27 (68)	7 (64)	0.94
01 - Restricted	16 (31)	12 (30)	4 (36)	
02 - Ambulatory	1 (2)	1 (2)	-	
<u>Heavy chain</u>				
IgA	14 (27)	10 (25)	4 (36)	> 0.99
IgG	29 (57)	22 (55)	7 (64)	
<u>Liaht chain</u>				
Kappa Light Chain	2 (4)	2 (5)	-	> 0.99
Lambda Light Chain	3 (6)	3 (8)	-	
<u>Mayo criteria 20-2-20</u>				
Low	21 (41)	17 (42)	4 (36)	0.47
Intermediate	15 (29)	10 (25)	5 (45)	
High	15 (29)	13 (32)	2 (18)	
<u>Cytogenetics, FISH</u>				
t(4;14)				
Yes	6 (12)	4 (10)	2 (18)	
t(14;16)				
Yes	2 (4)	1 (2)	1 (9)	
17p deletion				
Yes	4 (8)	3 (8)	1 (9)	
+1q amplification				
Yes	11 (22)	9 (22)	2 (18)	

	Total	Treatment arm		<i>p-value</i>
		A: Elo/Len/Dex	B: Elo/Len	
	n = 51 (%)	n = 40 (78)	n = 11 (22)	
<u>High-risk cytogenetics</u>				
No	30 (59)	25 (62)	5 (45)	0.70
Yes	17 (33)	13 (32)	4 (36)	

Table 2:

Highest grade toxicity experienced on treatment for n=51 patients. AEs grade 2+ in at least 10% of patients and all grade 4+.

	Total N=51	Worst grade			
		2 - Moderate	3 - Severe	4 - Life Threatening	5 - Fatal
Upper respiratory infection	22 (43)	21 (41)	1 (2)	-	-
Fatigue	21 (41)	18 (35)	3 (6)	-	-
Hypophosphatemia	21 (41)	2 (4)	16 (31)	3 (6)	-
Neutropenia	15 (29)	2 (4)	12 (24)	1 (2)	-
Lymphocytopenia	14 (27)	3 (6)	11 (22)	-	-
Hyperglycemia	12 (24)	9 (18)	2 (4)	1 (2)	-
Constipation	10 (20)	9 (18)	1 (2)	-	-
Cytopenia	10 (20)	8 (16)	2 (4)	-	-
Diarrhea	10 (20)	8 (16)	2 (4)	-	-
Anemia	8 (16)	7 (14)	1 (2)	-	-
Lung infection	8 (16)	3 (6)	5 (10)	-	-
Hypertension	7 (14)	4 (8)	3 (6)	-	-
Insomnia	7 (14)	6 (12)	1 (2)	-	-
Thrombocytopenia	7 (14)	4 (8)	2 (4)	1 (2)	-
Rash maculopapular	6 (12)	4 (8)	2 (4)	-	-
Cataract	1 (2)	-	-	1 (2)	-
Cholecystitis	1 (2)	-	-	1 (2)	-
Diabetic ketoacidosis	1 (2)	-	-	1 (2)	-
Lymphocytosis	1 (2)	-	-	1 (2)	-
Myocardial infarction	1 (2)	-	-	-	1 (2)
Sepsis	1 (2)	-	-	-	1 (2)

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Human CD45 (HI30)-89Y—100 Tests	Fluidigm	Cat#3089003B
Anti-Human CD196/CCR6 (G034E3)-141Pr—50 Tests	Fluidigm	Cat#3141003A
Anti-Human CD19 (HIB19)-142Nd—100 Tests	Fluidigm	Cat#3142001B
Anti-Human CD45RA (HI100)-143Nd—100 Tests	Fluidigm	Cat#3143006B
Anti-Human CD69 (FN50)-144Nd—100 Tests	Fluidigm	Cat#3144018B
Anti-Human CD4 (RPA-T4)-145Nd—100 Tests	Fluidigm	Cat#3145001B
Anti-Human IgD (IA6-2)-146Nd—100 Tests	Fluidigm	Cat#3146005B
Anti-Human CD20 (2H7)-147Sm—100 Tests	Fluidigm	Cat#3147001B
Anti-Human CD14 (RMO52)-148Nd—100 Tests	Fluidigm	Cat#3148010B
Anti-Human CD56/NCAM (NCAM16.2)-149Sm—100 Tests	Fluidigm	Cat#3149021B
Anti-Human CD138/Syndecan-1 (DL-101)-150Nd—25 tests	Fluidigm	Cat#3150012C
Anti-Human TNFa (mab11)-152Sm—100 Tests	Fluidigm	Cat#3152002B
Anti-Human TIGIT (MBSA43)-153Eu—100 Tests	Fluidigm	Cat#3153019B
Anti-Human CD3 (UCHT1)-154Sm—100 Tests	Fluidigm	Cat#3154003B
Anti-Human CD27 (L128)-155Gd—100 Tests	Fluidigm	Cat#3155001B
Anti-Human IFNg (B27)-158Gd—100 Tests	Fluidigm	Cat#3158017B
Anti-Human CD161 (HP-3G10)-159Tb—100 Tests	Fluidigm	Cat#3159004B
Anti-Human/Mouse MIP1beta (D21-1351)-160Gd—100 Tests	Fluidigm	Cat#3160013B
Anti-Human/Mouse Tbet (4B10)-161Dy—100 Tests	Fluidigm	Cat#3161014B
Anti-Human FoxP3 (259D/C7)-162Dy—50 Tests	Fluidigm	Cat#3162024A
Anti-Human CD183/CXCR3 (G025H7)-163Dy—100 Tests	Fluidigm	Cat#3163004B
Anti-Human IL-17A (N49-653)-164Dy—100 Tests	Fluidigm	Cat#3164002B
Anti-Human CD279/PD-1 (EH12.2H7)-165Ho—100 Tests	Fluidigm	Cat#3165042B
Anti-Human CD197/CCR7 (G043H7)-167Er—50 Tests	Fluidigm	Cat#3167009A
Anti-Human CD8 (SK1)-168Er—100 Tests	Fluidigm	Cat#3168002B
Anti-Human CD25 (2A3)-169Tm—100 Tests	Fluidigm	Cat#3169003B
Anti-Human Granzyme B (GB11)-171Yb—100 Tests	Fluidigm	Cat#3171002B
Anti-Human CD57 (HCD57)-172Yb—100 Tests	Fluidigm	Cat#3172009B
Anti-Human HLA-DR (L243)-173Yb—100 Tests	Fluidigm	Cat#3173005B
Anti-Human CD94 (HP-3D9)-174Yb—100 Tests	Fluidigm	Cat#3174015B
Anti-Human Perforin (B-D48)-175Lu—100 Tests	Fluidigm	Cat#3175004B
Anti-Human CD127/IL-7Ra (A019D5)-176Yb—100 Tests	Fluidigm	Cat#3176004B
Anti-Human CD16 (3G8)-209Bi—100 Tests	Fluidigm	Cat#3209002B
Anti-Granzyme K antibody [GM-24C3] (ab3771)	Abcam	Cat#ab3771
Chemicals, peptides, and recombinant proteins		
1X Phosphate-Buffered Saline without Calcium & Magnesium	Mediatech	Cat #21040CV

REAGENT or RESOURCE	SOURCE	IDENTIFIER
10X RBC Lysis Buffer	BioLegend	Cat #420301
0.4% Trypan Blue Solution	Life Technologies	Cat#15250061
autoMACS Running Buffer - MACS Separation Buffer	Miltenyi Biotec	Cat#130091221
CD138 MicroBeads, Human	Miltenyi Biotec	Cat#130051301
DMSO	Sigma Aldrich	Cat #D2650
PBS, pH 7.4 (1X)	Gibco	Cat#10010023
Bovine Serum Albumin (BSA), 30% ± 2% in 0.85% NaCl	Sigma Aldrich	Cat#A7284
Sodium Azide, 10% (w/v) solution in Ultra-Pure H ₂ O	Teknova	Cat#S0209
RPMI 1640 (1X) Medium	Gibco	Cat#11875093
Fetal Bovine Serum, heat inactivated, qualified (FBS)	Gibco	Cat#10438026
Antibiotic-Antimycotic (100X)	Gibco	Cat#15240096
UltraPure™ DNase/RNase-Free Distilled Water	Invitrogen	Cat#10977023
ViaStain™ AOPI Staining Solution	Nexcelom Biosciences	Cat#CS2-0106-5mL
Cell-ID™ Intercalator-Rh—500 µM	Fluidigm	Cat#201103A
Cell-ID™ Intercalator-Ir—125 µM	Fluidigm	Cat#201192A
Maxpar® Cell Acquisition Solution	Fluidigm	Cat#201237
Maxpar® Cell Staining Buffer j?	Fluidigm	Cat#201068
EQ™ Four Element Calibration Beads	Fluidigm	Cat#201078
eBioscience™ Permeabilization Buffer (10X)	Invitrogen	Cat#00833356
eBioscience™ Fixation/Permeabilization Concentrate	Invitrogen	Cat#00512343
eBioscience™ Fixation/Permeabilization Diluent	Invitrogen	Cat#00522356
Human TruStain FcX™ (FcR Blocking Solution)	BioLegend	Cat#422302
Pierce™ 16% Formaldehyde Solution (w/v), Methanol-free	Thermo Scientific	Cat#28906
Critical commercial assays		
Maxpar® X8 Antibody Labeling Kit, 166Er—4 Rxn	Fluidigm	Cat#201166A
Software and algorithms		
FlowJo	Becton Dickinson & Company	https://www.flowjo.com/solutions/flowjo/downloads/
CyTOF Software v7.0.8493	Fluidigm	https://www.fluidigm.com/software
Biological Samples		
Peripheral blood and bone marrow samples collected	This paper; E-PRISM trial	ClinicalTrials.gov: NCT02279394
Deposited Data		
Single-cell RNA sequencing (n=149) and TCR sequencing (n=91)	This paper; E-PRISM trial	phs002476.v1.p1
Single-cell RNA sequencing	Zavidij et al., 2020	phs001323.v3.p1
Whole-exome sequencing	Bustoros et al., 2020	phs001323.v3.p1