blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

Genomic profiling of Mycosis Fungoides identifies patients at high risk of disease progression

Tracking no: ADV-2023-012125R1

Léa Fléchon (Canther, INSERM UMR-S1277, CNRS UMR9020, Lille University, Lille, France) Inès Arib (Canther, INSERM UMR-S1277, CNRS UMR9020, Lille University, Lille, France) Ankit Dutta (Dana-Farber Cancer Institute, United States) Lama Hasan Bou Issa (Canther, France) Romanos Sklavenitis-Pistofidis (Dana-Farber Cancer Institute, United States) Remi Tilmont (CHRU Lille, France) Chip Stewart (Broad Institute, United States) Romain Dubois (CHRU de Lille, France) Stéphanie Poulain (CHRU de Lille, France) Marie-Christine Copin (CHU Angers, France) Sahir Javed (CH Valenciennes, France) Morgane Nudel (CHRU Lille, France) Doriane Cavalieri (CHU Lille, France) Guillaume Escure (CHU Lille, France) Nicolas Gower (CHU Lille, France) Paul Chauvet (CHU Lille, France) Nicolas Gazeau (CHU-Lille, France) Cynthia Saade (CHU Lille, France) Marietou Binta Thiam (CHU Lille, France) Aïcha Ouelkite-Oumouchal (Canther, France) Silvia Gaggero (Canther, France) Émeline Cailliau (CHU Lille, France) Sarah Faiz (CHRU de Lille, France) Olivier Carpentier (CH Roubaix, France) Nicolas Duployez (CHU Lille, France) Thierry Idziorek (Inserm, France) Laurent Mortier (Claude Huriez University Hospital, France) Martin Figeac (Lille University, France) Claude Preudhomme (chru of Lille, France) Bruno Quesnel (CHU Lille, France) Suman Mitra (UMR 9020-UMR-S 1277-Canther, Institut de Recherche contre le Cancer de Lille, University Lille, CNRS, Inserm, CHU Lille, F-59000 Lille, France, France) Franck Morschhauser (University of Lille, France) Gad Getz (Broad Institute, Mass General Hospital, United States) Irene Ghobrial (Dana-Farber Cancer Institute, United States) Salomon Manier (Canther, INSERM UMR-S1277, CNRS UMR9020, Lille University, Lille, France)

Abstract:
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 low-risk patients. 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.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: Conceptualization, Methodology, Writing - Original Draft and Visualization, L.F., I.A. and S.Ma.; Software, L.F., R.S.-P., C.St.; Formal Analysis, L.F., I.A, A.K.D., R.S.-P., C.St., E.C., M.F, S.Mi, G.G, I.M.G and S.Ma.; Investigation, L.F., I.A., A.K.D., L.H.B.I., and S.Ma.; Resources, I.A., R.D., S.P., M.-C.C., S.J., M.N., A.O.-O., S.F., O.C., L.M., T.I., F.M., G.G, I.M.G, S.Ma.; Data Curation, L.F. and S.Ma.; Writing - Review & Editing, All authors; Supervision, G.G, I.M.G and S.Ma.; Funding Acquisition, S.Ma.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The WES samples used in this study were deposited in the SRA database (https://www.ncbi.nlm.nih.gov/sra) with the 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 [S.Ma] (salomon.manier@chu-lille.fr).

Clinical trial registration information (if any):

Genomic profiling of Mycosis Fungoides identifies patients at high risk of disease progression

Short title: Genomic evolution of Mycosis Fungoides

Léa Fléchon[1,*], Inès Arib[2,*], Ankit K. Dutta[3,4,5,*], Lama Hasan Bou Issa[1], Romanos Sklavenitis-Pistofidis[3,4,5], 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], Salomon Manier[1, 2, ‡]

[1]Canther, INSERM UMR-S1277, CNRS UMR9020, Lille University, Lille, France. [2]Department of Hematology, Lille Hospital, Lille, France. [3]Center for Prevention of Progression of Blood Cancers, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Medical Oncology, Harvard Medical School, Boston, MA, USA. [5]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Institute of Pathology, Lille Hospital, Lille, France. [7]Hematology Laboratory, Biology and Pathology Center, Lille Hospital, Lille, France. [8]Department of Pathology, Angers University, Angers Hospital, INSERM, CRCI2NA, Angers, France. [9]Department of Medical Oncology, Valenciennes Hospital, Valenciennes, France. [10]Department of Biostatistics, Lille Hospital, Lille, France. [11]Department of Pathology and Dermatology, Lille Hospital, Lille, France. [12] OncoThai, INSERM UMR-S1189, Lille University. [13]Lille University, Lille Hospital, CNRS, INSERM, Institut Pasteur de Lille, US 41 – UAR 2014 - PLBS, Lille, France. [14]Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [15]Harvard Medical School, Boston, MA, USA. [*]L.F, I.A and A.K.D. contributed equally. ‡ Lead contact and corresponding author: salomon.manier@chu-lille.fr

Text word count: 3,997 words

Abstract word count: 183 words

Number of figures and tables: 5 Figures

Number of references: 50 references

Scientific Category: Lymphoid Neoplasia

Data Sharing Statement. The WES samples used in this study were deposited in the SRA database (https://www.ncbi.nlm.nih.gov/sra) with the 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 [S.Ma].

Key points:

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

Abstract

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 low-risk patients. 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 over 50% of all cases[1]. It typically exhibits an indolent disease course with slow progression 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 non-specific lesions that are difficult to diagnose. Patients may progress from low-risk (LR) clinical stages (TNMB IA to IIA) to high-risk (HR) stages (TNMB IIB to IVB), which involve tumors or generalized erythroderma[2,3]. At the cellular level, approximately 20% of patients with HR show histological transformation with transformed MF cells (tMF), a feature associated with poor prognosis[4–6]. Transformation is defined by the presence of more than 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, where 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 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 high risk 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 (GFELC) 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 patients samples with or without gain10p15.1. The diagnosis of MF was confirmed by PCR to detect clonal recombination of the T-cell receptor gene. We defined two 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 biopsies stored in the Biology-Pathology-Genetics unit's tumor bank (certification NF 96900-2014/65453-1). After macro-dissection 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 (IRB) of Lille University Hospital and Nord Ouest IV (protocol #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, 100ng of genomic DNA was fragmented on the Covaris ultrasonicator to target a base pair peak of approximately 150-200bp. 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 qPCR using the Kapa Library Quantification kit (Roche). Libraries were pooled and sequenced using 2 × 100bp 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 FastQC[14] 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. Post-alignment cleanup consisted of the removal of duplicate reads with MarkDuplicate[18] (Picard, BI), local realignment around 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 CalculateContamination[20] and CrossCheckLaneFingerprints[18], respectively, and no errors were identified. Finally, the potential oxidation of guanine to 8-oxoguanine (OxoG) 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 cut offs had an average coverage of 99.72% on the GRCh37 assembly a with mean target depth of coverage of 231.85X.

Copy number analysis from WES data. SCNAs and genome-wide allelic variations were detected using ModelSegments[20] software (BI). 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 three tumor samples that had an associated normal sample. Frequent large and focal SCNAs in the cohort were highlighted by GISTIC2.0[21] using a q-value threshold of 0.01.

Mutation calling of recurrently mutated genes. SNVs and InDels were detected using a bioinformatics pipeline developed by the BI and called “CGA WES Characterization Pipeline”. SNVs were detected by MuTect[22] and InDels by Strelka[23] and Mutect2[24] (BI). The results of these three software programs were then filtered using MAFPoNFilter[25] (BI). The latter uses two PoNs to segregate somatic from germline variants: the one from our cohort (from 13 WES normal samples) and the controlled access PoN used in the routine analysis of the BI (8,334 normal samples from the TCGA database). Somatic variants (SNVs and InDels) were annotated for their oncogenic effect using Variant Effect Predictor[26] and Oncotator[27]. They were then validated using the MutationValidator tool (BI), 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 MutSig2CV[25] (BI). Variants carried by genes known to be "fishy genes" i.e., 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.* 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, Massachusetts, USA). These 13 samples formed a PoN to filter out possible errors during the preparation or sequencing steps, as well as germline genetic/genomic events. As a final criterion to filter potential germline variants, we employed 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 two-motive list per variant, consisting of 11 nucleotides before and after the variant. The

first motive represented the wildtype (WT), corresponding to the reference variant, whereas the second one was the mutated (MU) form, representing the alternate variant. Subsequently, we used the "do_it.sh" bash script from the GitHub repository mafouille/HotCount (https://github.com/mafouille/HotCount) to count the number of occurrences of both WT and MU motives in the PoN. Variants with MU 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 cancer cell fraction. ABSOLUTE[29] (BI) software determined the purity and ploidy of each sample from the SNVs and SCNAs (Supplementary Table S10), and thus served us to define the CCF of each genomic alteration. For tumor samples without associated normal samples, we applied a tumor-only Germline Somatic Log odds filter, as described in Chapuy *et al.* 2018[30]. To summarize, for each variant, its CCF, sample ploidy and purity, and local CN were used to calculate the probability that the allelic fraction of the variant was consistent with a modeled allelic fraction for either a germline or somatic hypothetical event. Thus, the filter set a threshold for each sample to remove additional germline variants. To eliminate additional germline events, we set the tumor-only threshold to -1.

GnomAD filtering. We used the Genome Aggregation Database gnomAD to exclude potential germline mutations. The Human Genome Variation Society (HGVS) genomic identifier (HGVS_genomic_change column in maf file) was used in the VEP GRCH37 Ensembl database to obtain gnomAD allele frequency. A cut-off 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. The TTP 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 non-parametric Wilcoxon test. For the forest plot, events with a low frequency of less than ~20% (i.e. 9 patients) were not analyzed (Figure 2A). Figures and statistical estimations were obtained using R v.3.6.3 and MATLAB. R packages used are "maftools" v2.12.05 with *oncoplot* (Figure 1A), *maf_compare* (Figure 2A) and *somaticInteractions* (Figure 3B) functions, "survival" v3.4-0 and "survminer" v0.4.9. (Kaplan Meier graphs, Figure 2), "chrisamiller/fishplot" v0.5.1 (Fishplots, Figure 4). The maf_Compare function performs a Fisher test on all genes between the LR and HR cohorts to detect differentially mutated SNVs/SCNAs. Other figures were generated using the R ggplot2 v3.4.0 package.

This study was approved by the Institutional Review Board (IRB) of Lille University Hospital and Nord Ouest IV (protocol #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 (stage IIB to IVB) (Supplementary Figures S1A-D and Supplementary Tables S1 and S7). We detected mutations by WES and analyzed 67 tumor samples from 48 patients, using a validated pipeline to filter germline variants and artifacts from tumor-only samples[30] (Methods and Supplementary Figures S1 and S2). Thirteen patients had sequential sampling (Supplementary Figures S1C and D).

We found a median of 3.5 mutations/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 T-cell lymphoma, and 51% of patients had a mutation in at least one of these drivers (Figure 1A). These included the T-cell differentiation transcription factor *JUNB* (p.A282V, Figure 1B) in 13% of the cases[31–34]; the epigenetic factor *TET2* in 9% of the cases[35–38], the component of the MAPK pathway *MAPK1* in 6% of the cases (Figure 1B); the transcription factor *FOXA1* involved in interferon signaling and immune response suppression[37,39]; the tyrosine kinase receptor *FLT4*; the phospholipase *PLCG1* involved in NF-κB and NFAT signaling[7,40]; the JAK/STAT signaling pathway factors *STAT3*, *STAT5A*, and *STAT5B*; and *TP53* (Figure 1A and Supplementary Figure S4 and Supplementary Table S2).

Next, we identified significantly recurrent SCNAs by using GISTIC2.0[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 (CN) losses, and 2 arm-level and 1 focal copy gains (q-value $\leq$ 0.1; Supplementary Figures S3A and B and Supplementary Table S11). 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) (Supplementary Tables S12 and S13). 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 *PRKCQ*. 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 impacted 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 three-bases mutational sequence context and the clustering of the mutation in the genome to define specific COSMIC Single Base Substitution (SBS) signatures. Here, we detected two primary mutational signatures: the UV light exposure signature SBS7 and the defective DNA repair signature SBS15 (Supplementary Figures S5A, B, and E, Supplementary Table S8). 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 cutaneous T-cell lymphoma[32,45,46] and was also significantly associated with HR, suggesting a role for UV radiation in the progression of MF. It is also possible that HR tumors are phylogenically 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, $p = 5.735e^{-12}$). Signature SBS15 is one of the seven 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 (Supplementary Figure S5G). Next, we determined the relative contribution of SBS7 and SBS15 signatures to the mutational burden of driver genes (Supplementary Figure S5E). Five genes were enriched with mutations associated with SBS15, including *JUNB, STAT3, MAML2, FLT4*, and *FOXA1*, whereas nine 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 to those of patients with HR disease (Supplementary Tables S3 to S6 and S9). All CN gain7q (OR 0.07, p = 0.002) and gain10p15.1 (*IL2RA* and *IL15RA*, OR 0.11, p = 0.019) were found in patients at HR, as were del10q24.32 (*NFKB2*, OR 0.09, p = 0.010) and del10p11.22 (*ZEB1*, OR0.12, p = 0.038) (Figure 2A). Conversely, del17q11.2 (*SUZ12, NF1*) was significantly associated with LR (OR 4.84, p = 0.014). We observed a higher mutation rate in HR samples, with a median of 3.95 mutations/Mb compared to 3.02 mutations/Mb in LR samples (p = 0.019) (Supplementary Figure S5F). 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 and Supplementary Figure S5G).

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 years). In total, 4 SCNAs were significantly associated with shorter OS, (Figure 2B to E): del10p11.22 (*ZEB1*, median OS [mOS] of 2.4 *vs.* 8.9 years, p = 0.00029); gain10p15.1 (*IL2RA*, *IL15RA*, mOS of 2.4 *vs.* 5.8 years, p = 0.00029); gain7q arm (mOS of 2.6 *vs.* 8.9 years, p = 0.011); and del6q16.3 (*TNFAIP3*, mOS 1.9 vs. 5.8 years, p = 0.021). 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 (Supplementary Figures S5C and 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 (TTP) (median 0.8 years *vs.* Not Reached [NR], p = 0.0031) (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 of samples from patients with gain10p15.1 as compared to 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 cancer cell fraction (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 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*, *IL15RA*) was significantly co-segregated with del6q16.3 (*TNFAIP3*) and *TP53* mutations. Del16p13.3 (*CREBBP*) co-occurred with del17p and de19p13.3 (*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 two 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*, *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*, *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 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 where subclones that were initially small at diagnosis exhibited substantial expansion at the 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, C, and 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 two 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 T-cell lymphoma in general[9,32,36,38,48,49] (Mycosis fungoides, Sézary syndrome, primary cutaneous $CD30^+$ T-cell lymphoproliferative disorders, primary cutaneous γδ T-cell lymphoma), with alterations in the NF-κB pathway (such as deletion *NFKB2* in 10q24.32, gain of *PRKCQ* in 10p15.1, mutation of *PRCKB*, deletion of *CARD11* in 7p22.3), JAK/STAT pathway (e.g. gain of *IL2RA* and *IL15RA* in 10p15.1, mutations in *STAT3*, *STAT5A* and *STAT5B*), MAPK pathway (mutations in *JUNB* and *MAPK1*), cell cycle pathway (such as deletion *CDKN1B* in 12p13.1, *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*, *NF1*) was associated with the LR stages of the disease. The mutational signature SBS7, associated with UV light, was 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 *PRCKB* 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, 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, as 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 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 high risk of disease progression and their clinical management.

Acknowledgements

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

Authorship Contributions

Conceptualization, Methodology, Writing – Original Draft and Visualization, L.F., I.A. and S.Ma.; Software, L.F., R.S.-P., C.St.; Formal Analysis, L.F., I.A, A.K.D., R.S.-P., C.St., E.C., M.F, S.Mi, G.G, I.M.G and S.Ma.; Investigation, L.F., I.A., A.K.D., L.H.B.I., and S.Ma.; Resources, I.A., R.D., S.P., M.-C.C., S.J., M.N., A.O.-O., S.F., O.C., L.M., T.I., F.M., G.G, I.M.G, S.Ma.; Data Curation, L.F. and S.Ma.; Writing – Review & Editing, All authors; Supervision, G.G, I.M.G and S.Ma.; Funding Acquisition, S.Ma.

Conflicts of Interests Disclosures

The authors declare no competing interests.

References

1. Willemze R. Cutaneous T-Cell Lymphoma: Epidemiology, Etiology, and Classification. *Leukemia & Lymphoma*. 2003;44(sup3):S49–S54.
2. Willemze R, Jaffe ES, Diaz-Perez JL, et al. WHO-EORTC classification for cutaneous lymphomas. *Blood*. 2005;105(10):18.
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, Duvic 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 and 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. *Journal of Investigative Dermatology*. 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 r., Benner M f., Zoutman W h., 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*. 2015;172(3):784–788.
14. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010;

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.journal*. 2011;17(1):10–12.

17. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;

18. Picard Tools (Broad Institute) Available online at: 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. *Current Protocols in Bioinformatics*. 2013;43(1):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. Sebastopol, CA: O'Reilly Media; 2020.

21. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology*. 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, et al. 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. Bioinformatics; 2019.

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 Biology*. 2016;17(1):122.

27. Ramos AH, Lichtenstein L, Gupta M, et al. Oncotator: Cancer Variant Annotation Tool. *Human Mutation*. 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. *Journal of Cutaneous Pathology*. 2008;35(10):899–910.

34. Mao X, Orchard G, Lillington DM, et al. 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 Advances*. 2020;4(11):2489–2500.

38. da Silva Almeida AC, Abate F, Khiabanian 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):.

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, 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*. 2001;117(3):583–589.

44. Kasar S, Kim J, Improgo 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, Degasperi 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):1–5.

50. Hristov AC, Tejasvi T, Wilcox RA. Mycosis fungoides and Sézary syndrome: 2019 update on diagnosis, risk-stratification, and management. *American Journal of Hematology*. 2019;94(9):1027–1041.

Figure legends

Figure 1. Landscape of most recurrent genomic alterations in MF. A. Landscape of genomic alterations in 67 tumor samples of MF divided into low-risk and high-risk disease based on the TNM stages. Alterations are divided into single nucleotide variants (SNVs) and somatic copy number alterations (SCNAs). **B.** Driver oncogene maps for *JUNB* and *MAPK1*. **C.** Relative enrichment of signature activities per samples divided into low-risk, low-risk progressors and high-risk.

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 low-risk and high-risk, as depicted by odds ratio. **B** to **E**: Kaplan Meier plots of individual genetic factors predictive of OS in univariate and multivariate models of 48 patients with a newly diagnosed MF: (**B.**) del10p11.22; (**C.**) gain of 10p15.1; (**D.**) gain of 7q; and (**E.**) del6q16.3. **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 tMF cells) or without gain of 10p15.1 (bottom panels, MF sample of patient 23 and patient 16). Scale bars indicate 150µm.

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 cancer cell fraction (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's 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 two alterations were drawn when two drivers were

found in one sample with an excess of clonal to subclonal events. Dashed arrows indicate n = 1 clonal-subclonal pair and solid arrows indicate n ≥ 2 clonal-subclonal pairs.

Figure 4. Clonal evolution of sequential samples. Fish plots of five serial cases of MF with samples at low-risk stages who progressed to high-risk stages (**A**-**E**). And one serial case with both time points at low-risk stage because the patient did not progress (**F**). TNMB classification has been indicated (green for Low-Risk and dark red for High-Risk samples) and tMF stands for transformed mycosis fungoides.

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

Figure 1

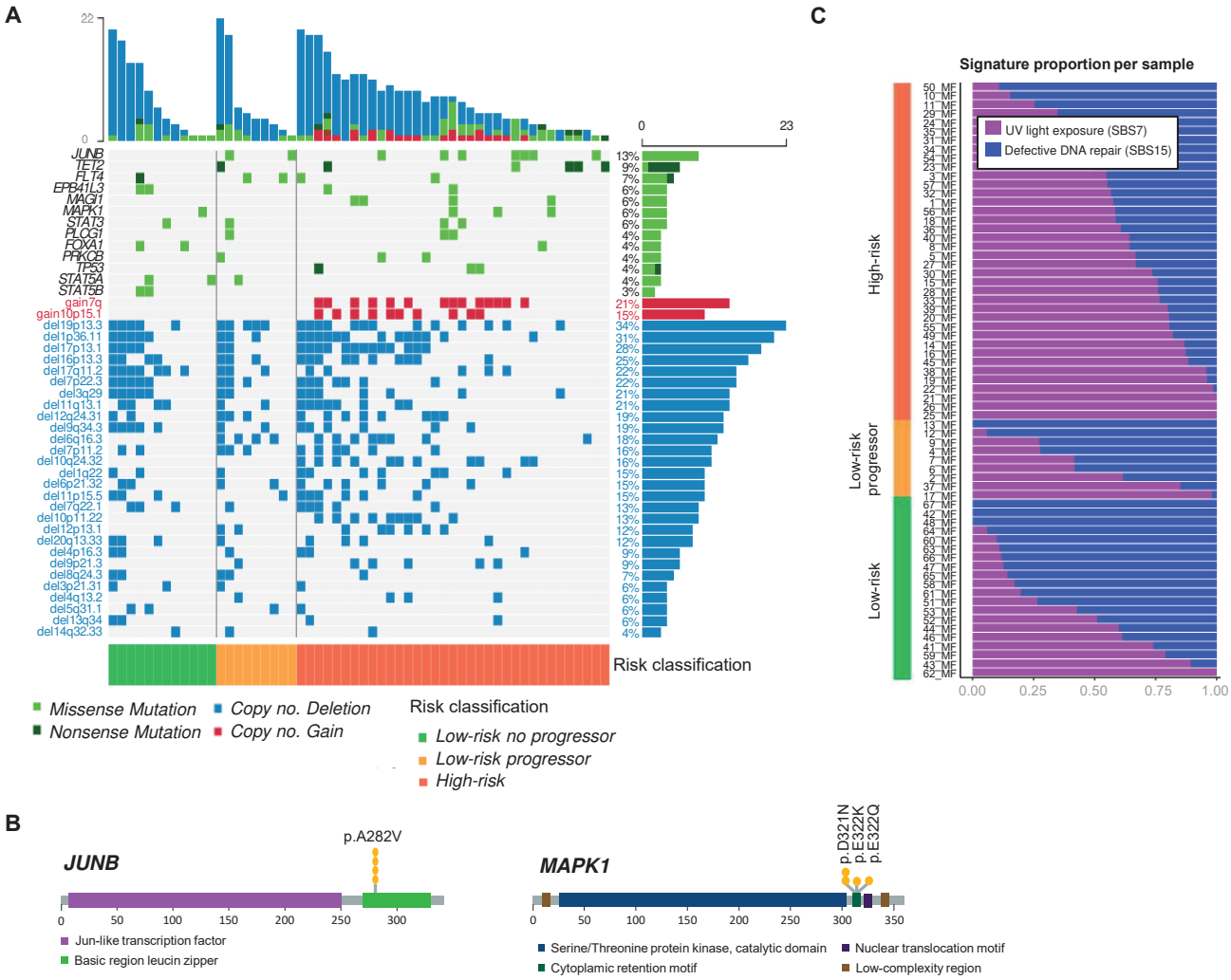

Figure 2

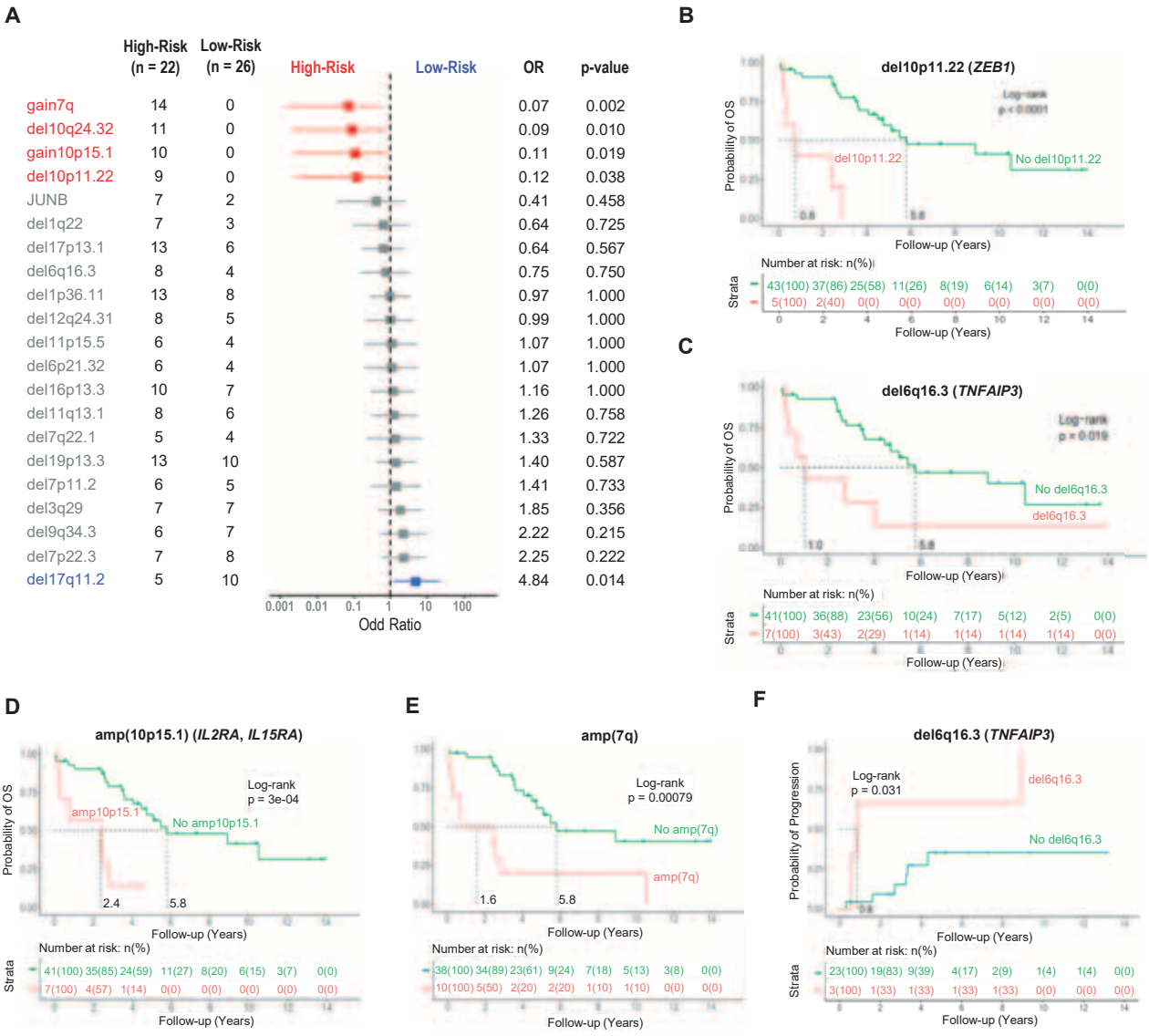

Figure 2

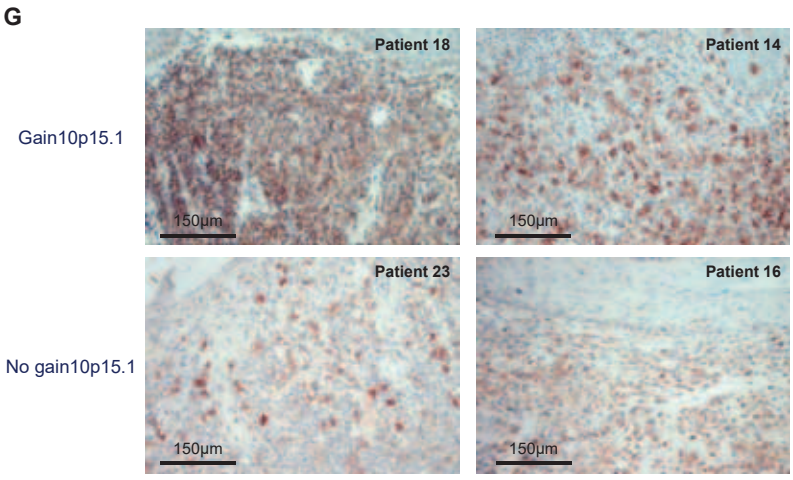

Figure 3

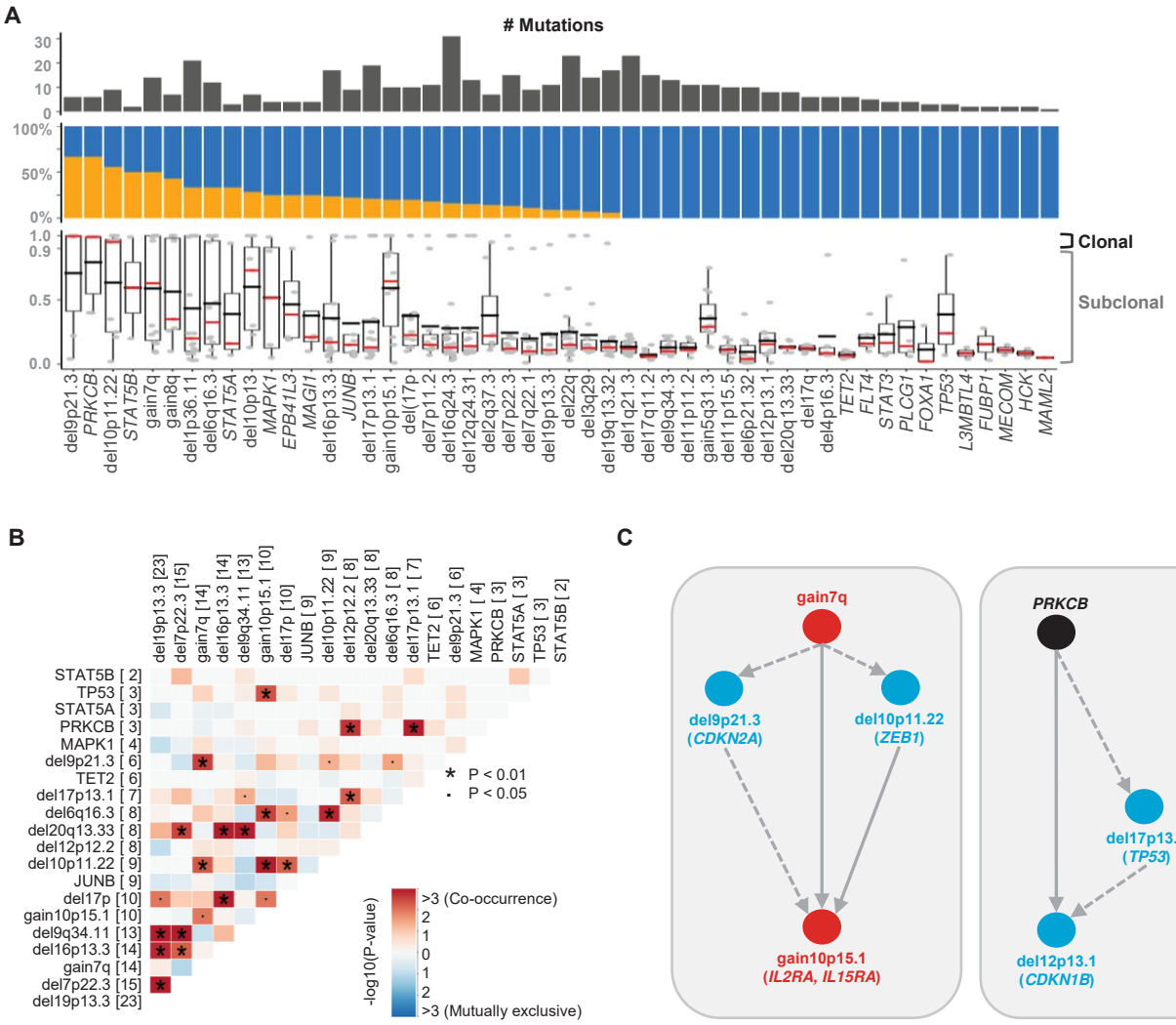

Figure 4

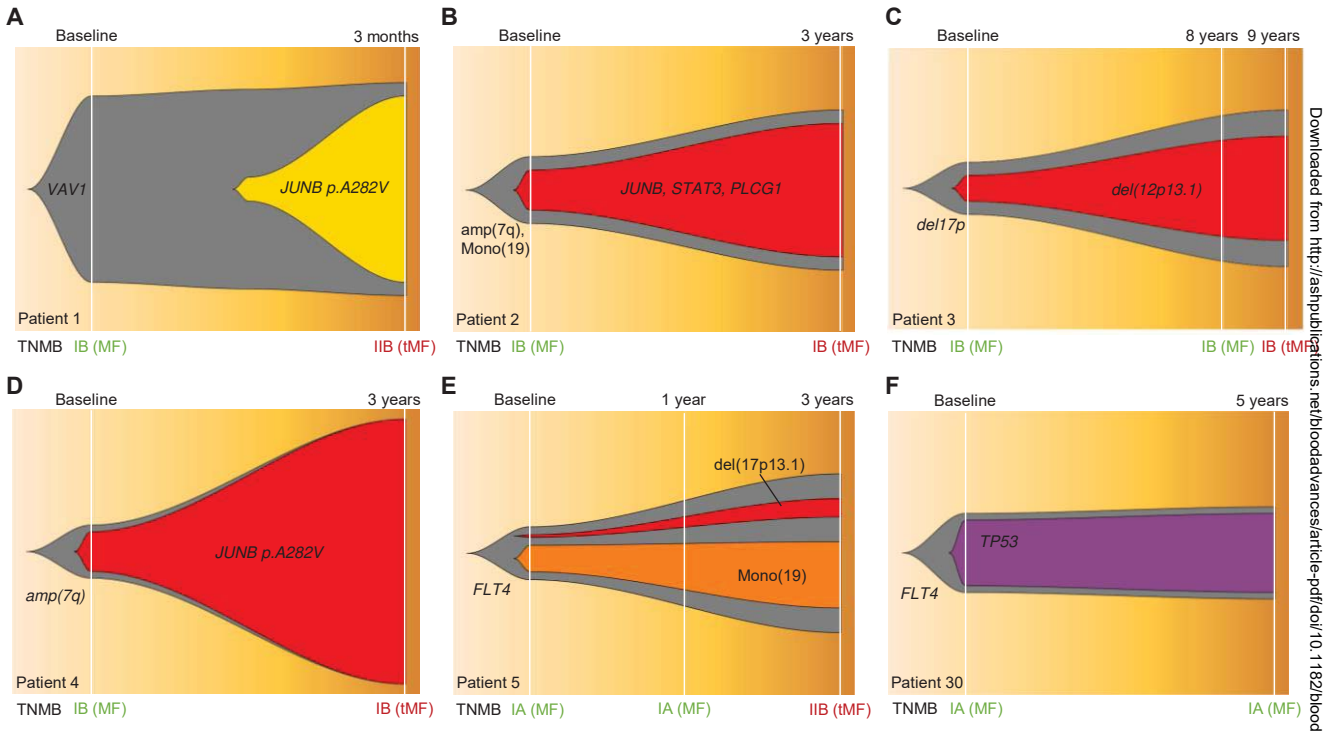

Figure 5