Transcriptomic Changes During Stage Progression of Mycosis Fungoides

M Xiao[1], *D Hennessey*[1], *A Iyer*[1], *S O'Keefe*[1], *F Zhang*[2], *A Sivanand*[1], *R Gniadecki*[1*].
[1]Division of Dermatology, University of Alberta, Edmonton, AB, Canada
[2]Faculty of Medicine, University of Ottawa, ON, Canada

***Address for correspondence:** Dr. Robert Gniadecki, 8-112 Clinical Science Building, 11350-83 Ave, Edmonton, AB, T6G 2G3, Canada Tel: (780) 407-1555, Fax: (780) 407-2996, email: r.gniadecki@ualberta.ca

Keywords: cutaneous T-cell lymphoma (CTCL), mycosis fungoides (MF), transcriptome analysis, expression profiling

Word count: 3555 (excludes abstracts)
Table count: 0
Figure count: 5

Short title (64/70 characters): Gene Expression Patterns in Early and Advanced Mycosis Fungoides

Conflict of Interest: RG received advisory board honoraria from Kyowa Kirin, Sanofi, Recordati Rare Diseases, and Mallinckrodt and obtained research funding from Sanofi and Sun Pharma.

Funding: This study was supported by grants from the following sources: Canadian Dermatology Foundation (CDF RES0035718), University Hospital Foundation (University of Alberta), Bispebjerg Hospital (Copenhagen, Denmark), Danish Cancer Society (Kræftens Bekæmpelse R124-A7592 Rp12350), and an unrestricted research grant to R Gniadecki from the Department of Medicine, University of Alberta. M Xiao was supported by an Alberta Innovates Summer Research Studentship. A Sivanand was supported by scholarships from the Canadian Institutes of Health Research (CIHR), Alberta Innovates, and the University of Alberta.

These funding organizations had no role in the design or conduct of the research.

What is known:

- Mycosis fungoides (MF) is the most common cutaneous T-cell lymphoma characterized by a favourable prognosis in the early patch/plaque stage.
- Development of tumors heralds progression to the advanced stage and a significant increase in mortality.

What's new:

- Tumor progression is associated with recurrent mutations which can be linked to the upregulation of signaling pathways controlling cell proliferation, survival, mitosis, and DNA repair.
- Percolation of malignant cells between lesions and tumor self-seeding is likely to mediate stage progression in MF.

Translational message:

The development of cutaneous tumors in MF heralds stage progression and increased mortality. MF tumors show upregulation in several targetable signaling pathways involved in cell proliferation and survival. Ectopic expression of cancer-testis genes may explain mitotic aberrations in MF tumors. We also propose that high-grade malignant cells can spread hematogenously and seed other skin lesions (tumor self-seeding). Early, aggressive treatment of tumors may prevent tumor self-seeding and improve patient prognosis.

Abstract

Background: Mycosis fungoides (MF) is the most common cutaneous T cell lymphoma, which in the early patch/plaque stages runs an indolent course. However, ~25% of MF patients develop skin tumors, a hallmark of progression to the advanced stage and associated with high mortality. The mechanisms involved in stage progression are poorly elucidated.

Methods: We performed whole-transcriptome and whole-exome sequencing of malignant MF cells from skin biopsies obtained by laser-capture microdissection. We compared three types of MF lesions: early-stage plaques (ESP, n=12), and plaques and tumors from patients in late-stage disease (late-stage plaques, LSP, n=10, and tumors, TMR, n=15). Gene Ontology (GO) and KEGG analysis were used to determine pathway changes specific for different lesions which we linked to the recurrent somatic mutations overrepresented in MF tumors.

Results: The key upregulated pathways during stage progression were those related to cell proliferation and survival (MEK/ERK, Akt-mTOR), Th2/Th9 signaling (IL4, STAT3, STAT5, STAT6), meiomitosis (CT45A1, CT45A3, STAG3, GTSF1, and REC8) and DNA repair (PARP1, MYCN, OGG1). Principal coordinate clustering of the transcriptome revealed extensive gene expression differences between early (ESP) and advanced-stage lesions (LSP and TMR). LSP and TMR showed remarkable similarities at the level of the transcriptome, which we interpreted as evidence of cell percolation between lesions via hematogenous self-seeding.

Conclusion: Stage progression in MF is associated with Th2/Th9 polarization of malignant cells, activation of proliferation, survival, as well as increased genomic instability. Global transcriptomic changes in multiple lesions are probably caused by hematogenous cell percolation between discrete skin lesions.

Introduction

Mycosis fungoides (MF) is the most prevalent form of primary cutaneous T-cell lymphoma (CTCL), a heterogeneous group of extranodal lymphoproliferative disorders that involve the skin.[1] Early-stage (IA-IIA) disease is characterized by erythematous papules, patches, and plaques, and the majority of patients (80-90%) experience an indolent disease course with slow progression over years or even decades.[2–5] However, ~25% of MF patients progress to advanced-stage (IIB-IV) disease, distinguished by the appearance of skin tumors, extracutaneous dissemination to lymph nodes, blood, and viscera, as well as a dramatic reduction in survival (decrease in 5-year survival from approximately 80% to 26%).[2–5] The molecular mechanisms underpinning stage progression in MF remain largely unknown. Moreover, it has not been explained how the emergence of even a single tumor in a relatively small area of the skin impacts the global evolution and prognosis of MF. It would be logical to propose that the more aggressive cancer cells in skin tumors spread hematogenously to other skin areas, increasing tumor load and the clinical impact of the malignancy. However, the putative hematogenous percolation of tumor cells between different lesions has not been investigated because MF cells have been considered to represent mutated tissue-resident T-cells which tend to remain arrested in the defined areas of the skin and do not recirculate via lymphatics and the lymph nodes like the transient passenger T-cells.[6] We reasoned that the hypothesis that MF cells traffic between skin lesions can be addressed by comparing transcriptomic profiles of skin lesions in different clinical stages of MF.

Previous analysis of the mutational landscape of CTCL revealed the emergence of numerous subclones and very complex patterns of clonal driver mutations, with little consistency among patients.[7–11] Over the past decade, studies have focused on the analysis of the transcriptome in an attempt to characterize cancer driver genes and signaling pathways, but the results have been discordant.[12] This is in part due to the high degree of intratumoral heterogeneity as well as the lack of standard approaches in both sample selection and experimental design that hinders direct comparison between different transcriptome studies.[12] Many studies pooled samples from MF and the leukemic CTCL (Sézary syndrome, SS) in spite of the data indicating that these entities may represent separate diseases[6,13] and few transcriptome studies account for the stage of disease and the morphology of skin lesions (e.g., patch, plaque, or tumor). Moreover, the samples used for transcriptomics were usually crude skin biopsies, in which the proportion of neoplastic cells is quite low, often below 30%.[7]

In this study, we compared transcriptome profiles of early and advanced MF to reveal potential mechanisms of disease progression. To capture stage-dependent expression changes, we classified MF skin lesions according to the patient's clinical disease stage (early [stage I] vs advanced [stage ≥ IIB]) and

the morphology of the lesion (plaque vs tumor). To circumvent earlier limitations, we used laser capture microdissection of neoplastic cell clusters to eliminate signal from normal cells. We were able to confirm some of the previously reported pathways associated with advanced-stage MF and identified new transcriptomic changes, such as upregulation of pseudo-meiotic processes (meiomitosis) that may be involved in tumor progression. Moreover, we found that stage progression is associated with global changes in transcriptomic profiles affecting multiple MF lesions which we interpret as evidence of cancer cell percolation proposed by us previously[14] and analogous to tumor self-seeding.[15,16]

Methods

Samples, whole transcriptome, and whole-exome sequencing

Ethics approval was obtained under the application HREBA.CC-19-0435 from the Institutional Review Board at the University of Alberta. 37 laser-microdissected skin lesions were collected from 23 adult patients with a diagnosis of MF in stages IA to IVB (**Supplementary Table S1**), as previously described[7]. RNA was isolated from the microdissected tissues using AllPrep DNA/RNA Micro kit from Qiagen (Hilden, Germany). For samples MF4_2, MF4_3, MF5_1, MF7_1, MF7_2, MF11, MF11_1, MF19_1, MF19_2, and MF19_3 rRNA was inhibited using NEBNext rRNA Depletion kit (NEB, Ipswich, Massachusetts, USA) and later processed using NEBNext Ultra II Directional RNA Library Prep kit for Illumina (Ipswich, Massachusetts, USA). The rest of the samples in the study were processed using Ovation Solo RNA-seq system (Nugen, Redwood city, CA, USA). The manufacturer's instructions were followed and no modifications were made to the protocols. The RNA libraries were sequenced on an Illumina HiSeq 1500 sequencer using paired-end 150 kits (cat# PE-402-4002) (Hiseq PE rapid cluster kit V2) or NovaSeq 6000 S4 reagent kit 300 cycles (cat# 20012866). Previously published whole-exome sequencing from the same samples was used for the analysis of single nucleotide variants, as described.[17]

Skin biopsies were categorized into three lesion types according to the clinical stage and morphology of the lesion at the time of collection. Plaque lesions sampled from patients in early-stage MF (IA-IIA) were designated as ESP (early-stage plaque, n=12). Patients with advanced-stage disease (≥IIB) have two types of skin lesions, the TMR (tumors, n=15) and the plaques which either persist from the early stages (late-stage plaques, LSP, n=10). There was no significant difference in tumor cell fraction between microdissected samples from different lesion types.

Bioinformatic analysis of gene expression data

The bioinformatics pipeline involved a series of aligners and statistical methods shown in **Figure 1.** All programs were used with default or developer-supplied settings unless otherwise specified. Code is available at https://github.com/d-henness/deseq2_project. Paired fastq files were preprocessed using fastp (v 0.19.7) with the --detect_adapter_for_pe flag enabled.[18] The resulting fastq files were then aligned and quantified using RSEM (v. 1.3.3).[19] RSEM was used with Bowtie 2 (v. 2.4.1)[20] as its backend aligner and an index constructed using Ensembl's 98 human genome annotation.[21] This yielded gene expression levels in units of transcripts per million (TPM).

Unsupervised Principal Coordinates Analysis (PCoA) was performed using edgeR (R package, v. 3.30.3)[22] to obtain an overview of the overall structure of the RNA-Seq data in a low-dimensional subspace. PCoA provides insights into the association between samples based on gene expression patterns and detects the formation of clusters among individual specimens. The DESeq2 program, implemented as a package in R (v. 1.28.1)[23] was used to identify differentially expressed genes (DEGs) that differed significantly between pairwise disease stages (log2 fold change (log2FC)>|1|). Benjamini-Hochberg correction was used to correct for multiple comparisons, with a standard false discovery cut-off (FDR) of <0.05.

To gain further insight into the functional processes differing between lesions, we used all of the available gene expression data (cutoff-free) and performed gene set analysis using the Generally Applicable Gene Set Enrichment (GAGE) method implemented as a package in R (v. 2.37.0).[24] We focused on Gene Ontology (GO) Biological Process (BP) terms[25] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.[26] In both databases, Benjamini-Hochberg correction was used to correct for multiple comparisons (FDR <0.05). Visual data representations were created using the R packages “ggplot2” (v. 3.3.1), "pheatmap" (v. 1.0.12), and 'EnhancedVolcano' (v. 1.6.0).[27]

Data availability

The exome sequencing data is available on dbGaP under accession number phs001877.v1.p1. The RNA sequencing fastq files are available on Sequence Read Archive (SRA) under accession number….[1]

[1] The accession number will be provided upon acceptance of the paper. For review purposes the RNA sequencing data can be requested via contacting the corresponding author.

Results

Changes in transcriptome signatures during stage progression of MF

Principal coordinates analysis (PCoA) was used to examine the associations between our samples (12 ESP, 10 LSP, 15 TMR) and revealed three clusters (**Figure 2A**). Cluster A contained the majority of ESP samples (11/12, or 91.7%) whereas the majority of advanced-stage samples (LSP and TMR) were distributed across clusters B1 and B2 (18/25, or 72% consisting of 8/10 LSP samples and 10/15 TMR samples). This suggests the transcriptome profile of ESP is distinct from the late-stage disease whereas the expression patterns among advanced lesions (LSP and TMR) are similar. Thus, the MF transcriptomic profile reflects the stage of the disease (advanced vs early) rather than the morphology of the lesion (plaque vs tumor) and indicates that significant global alterations in gene expression occur with progression to advanced-stage disease.

Next, we performed pairwise comparative analyses of the expression profiles between ESP, LSP, and TMR samples to identify differentially expressed genes (DEGs) between different lesion types. Differential expression analysis of TMR vs ESP (reference) lesions identified 1,154 DEGs (fold change>2; adjusted $P<0.05$), of which 908 DEGs were significantly up-regulated and 246 DEGs were downregulated in TMR compared to ESP (**Figure 2B)**. Fewer DEGs were detected for the comparisons of LSP to ESP (26 DEGs, 18 upregulated in LSP, **Figure 2C**) and TMR to LSP (29 DEGs, 11 upregulated in TMR, Figure 2D). A list of all differentially regulated genes across all comparisons is presented in **Supplementary Tables S2-S4**.

To further illustrate the relationship of genes found to be differentially expressed in each pairwise comparison, we used Venn diagram analysis and overlapped the 1,154 DEGs in TMR vs. ESP with the 26 DEGs in LSP vs. ESP (**Figure 3**). This revealed 1,137 DEGs unique to TMR (895 upregulated, 242 downregulated), 9 DEGs unique to LSP (5 upregulated, 4 downregulated), and a signature of 17 intersecting DEGs common to both LSP and TMR samples (13 upregulated, 4 downregulated), all concordantly altered (i.e., increased in both, or decreased in both). The large overlap in DEGs shared between LSP and TMR (both advanced lesions) compared with ESP corroborates the PCoA results and suggests that progression to advanced-stage affected the global transcriptomic profile of MF skin lesions.

GO and KEGG Functional Enrichment Analysis

We performed pairwise gene set analysis (GSA) using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases to identify enriched terms and pathways at

different MF stages. We used all of the available gene expression data instead of prefiltering for a list of strong DEGs in order to capture coordinated, low-level changes across genes that may belong to common pathways and are regulated by the same transcriptional network.

Our analyses identified 58 enriched GO biological processes and 21 KEGG pathways in TMR compared to ESP (**Supplementary Figure S1**). The results of GO analyses revealed that TMR lesions were significantly enriched in proliferative signaling ('Ras protein signal transduction', 'activation of MAPK kinase kinase activity', 'interleukin-2 mediated signaling') and immunosuppression processes ('response to interleukin-4' and 'TOR signaling') (**Figure 4A**). The KEGG pathway analysis results revealed TMR was highly enriched in 'Notch signaling pathway', 'base excision repair', and 'DNA replication'. (**Figure 4B**). While the heatmaps show a degree of heterogeneity in gene set enrichment scores across the TMR samples, the overall trend in expression of these annotations is upregulated in TMR. The significantly upregulated and downregulated members of each enriched gene set are presented in **Supplementary Table S5-S6.**

Next, we conducted functional enrichment analyses on LSP vs ESP (reference, **Figure 4B**). Although only a few individual genes were statistically significant between LSP and ESP, GSA identified 29 GO Biological Processes and 14 KEGG pathways overrepresented in LSP (**Supplementary Figure 2**). Among the GO and KEGG hits were five overlapping annotations also captured in the comparison between TMR and ESP (**Figure 4C**). These include three GO terms: 'respiratory burst' (inflammatory response), 'response to interleukin-4' (Th2 response), and 'TOR signaling' (growth and metabolism) as well as two KEGG pathways related to chromatin regulation: 'DNA replication' and 'base excision repair'.

We also identified GO terms and KEGG pathways uniquely enriched in the LSP group compared with ESP (**Figure 4C**), including three DNA damage removal annotations ('nucleotide-excision repair' [GO], 'mismatch repair' [GO], 'interstrand cross-link repair' [GO]), two T-cell function annotations ('T-cell receptor signaling' [KEGG], 'Activation of JUN kinase activity' [GO]), as well as hits in 'cell cycle' [KEGG] and 'p38MAPK cascade' [GO]. Enrichment of major DNA repair mechanisms in LSP is consistent with the widespread structural genomic alterations including fusion transcripts, insertions, deletions, and copy number variation described in advanced MF.[10]

Lastly, we performed GSA between TMR and LSP (reference). As expected, expression differences were smaller between LSP and TMR than in any other pairwise comparison, with only four enriched KEGG pathways and seven GO terms (**Supplementary Figure S3**). Compared with LSP, TMR

lesions were significantly upregulated in primary metabolic processes including 'ribosome biogenesis' 'biosynthesis of amino acids', and 'hexose metabolic process'.

Other Significant DEGs

Annotation databases like KEGG and GO allow us to collapse complex gene expression data into more manageable biological pathways and functional annotations, but the mapping of genes to pathways is limited by current knowledge of known molecular interaction networks and how gene products interact. To identify potential DEGs that could have been overlooked by functional analysis methods, we manually searched for individual genes that were significantly over-represented in each lesion type to provide a more nuanced understanding of transcriptional networks.

We confirmed overexpression of several genes reported previously in CTCL, such as *KIR3DL2* (>14 fold) and *KIR3DL4* (>6 fold), the members of the killer cell Ig-like receptor family that inhibit natural killer cell-mediated antitumor cytotoxicity. *KIR3DL2* has been implicated in SS[28] and transformed MF[29] and the anti-KIR3DL2 monoclonal antibody IPH4102 has recently demonstrated promising clinical activity in patients with refractory CTCL.[30] LSP and TMR had increased levels of *GNLY* (granulysin, an immune alarmin) and *NCR1* (natural cytotoxicity triggering receptor 1) whose expression was previously found to increase with the progression of MF[31] and in SS patients with high tumor burden[32].

Among the 29 DEGs between TMR and LSP, we found a >6 fold decrease in fractalkine Receptor (CX3CR1). CX3CL1 was shown to control effector T cell retention in inflamed skin and in the lungs in atopic dermatitis and asthma[33,34] and the decrease in its expression may facilitate escapement of malignant cells from the skin to the circulation in late stages MF.

We found ectopic expression of several meiosis-related cancer-testis antigens (meiCT) in TMR lesions. MeiCT antigens are normally present only in germ cells during oocyte development and spermatogenesis and become transcriptionally silent in normal somatic tissues. Ectopic activation of meiCT is a frequent phenomenon in cancer because global deregulation of epigenetic signaling causes unprogrammed gene activation.[35] While our results confirm upregulated expression of two previously described meiCT antigens (GTSF1 [gametocyte specific factor 1][36,37] and STAG3 [Stromal Antigen 3])[38] we also identified increased expression of seven putative CT antigens previously unreported in CTCL, including CT45A1 (Cancer/Testis Antigen Family 45 Member A1), its paralogue CT45A3 (Cancer/Testis Antigen Family 45 Member A3), BRDT (Cancer/Testis Antigen 9), PRAME (Cancer/Testis Antigen 130), PAGE5 (Cancer/Testis Antigen Family 16 Member 1), SPAG4 (Sperm Associated Antigen 4), and

PNMA5 (Paraneoplastic antigen-like 5). These results are in line with the notion that re-expression of a variety of CT antigens in malignant lymphocytes may contribute to increased genomic instability.

Linking gene mutations to transcriptomic changes

We examined whether the gene mutation pattern in MF could explain some of the observed alterations in the transcriptome. From the previously published catalogue of mutated genes obtained by whole-exome sequencing[17] we selected the genes which are frequently mutated in TMR and LSP and more expressed in those advanced-stage lesions compared to ESP. We assumed that those genes would be particularly relevant because mutations (especially damaging mutations) cause a compensatory overexpression of the gene.[39] We found that TMR/LSP lesions showed overexpression of mutated genes representing transcription factors and regulators (*MED12, MYCN, MYC, EGR3, CIC*), genes involved in genome integrity (*ERCC2, POLE*), the elements of the membrane receptor signalling network of integrated JAK-STAT, MEK/ERK, and PI3K/AKT/mTOR pathway (*PLCG1, INPPL1, VAV1, AKT1, PPP2R1A, NFKB2, CARD11, TRAF2, JAK3, STAT3*). Mutations in those genes have been linked to the signaling pathways in different cancers as summarized in **Figure 5A**.

Discussion

Our study addresses the poorly elucidated question of the mechanisms responsible for the development of tumors in MF, an event that defines the progression from the early to the advanced stage of the disease and which is associated with a dramatic increase in mortality. Among the studies where transcriptomic changes during stage progression were captured[37,40–42] only two reports[36,43] and a meta-analysis[41] compared DEGs directly between early and advanced lesions. The novel aspects of this study is that we not only compared skin lesions in early and late-stage patients, but we included different morphological types of MF lesions (the plaques and the tumors: ESP, LSP, TMR) and used cancer cell enriched, microdissected tissue of the high tumor cell fraction (mean purity [SD] 74.6 [20.4]%) which allowed us to largely limit the contribution of transcripts originating from non-neoplastic cells in the skin (epidermis, inflammatory cells, fibroblasts, and other stromal cells).

Compared with early-stage plaques (ESP), the tumor (TMR) samples showed an increased frequency of mutation in genes representing transcription factors and regulators, genes involved in genome integrity, JAK-STAT, MAPK/ERK, and PI3K/AKT/mTOR pathways **(Figure 5A)**. Although detailed mechanistic studies are needed to determine the relative importance of those mutations, overlaying the mutated genes on the known signaling pathway elements reveals their likely relevance in cancer progression. It was shown previously that mutated *PLCG1* increases the survival signaling via NF-

kB, NFAT and AP-1.[11,44] Mutated *STAT3* was also shown to have a pathogenic role in CTCL.[45,46]. Mutations in *CARD11*[9,11] and *JAK3*[2,47] are also supposed to increase the growth and survival of malignant T-cells in CTCL and are well-known driver mutations in other hematological malignancies.

Components of MEK/ERK pathway are reproducibly overexpressed in TMR and our mutation data revealed two potentially novel and previously unreported mechanisms of its activation **(Figure 5A)**. *MED12* is a tumor suppressor gene which in normal cells inhibits the transforming growth factor β (TGFβ), Akt/mTOR and MAPK/ERK signaling.[48–50] Interestingly, loss of function of mutated *MED12* was shown to be the main driver in certain tumors by activating RAS/MAPK/ERK[49–51]. *MED12* was also linked to drug resistance, which might be relevant in CTCL which is notoriously resistant to chemotherapy. Mutations in *Capicua* (*CIC*) may amplify MAPK/ERK signaling because CIC is a transcriptional repressor counteracting activation of genes downstream of ERK and contributes to the tumorigenesis of different solid tumors.[52–54]

VAV1 is a hematopoietic-specific RHO guanine exchange factor and its loss of function causes T-cell leukemias by increasing the Notch1-dependent ICN1 signaling due to inhibition of ICN1 ubiquitinylation and degradation.[55] Mutations in VAV1 were reported previously[56] but were linked to PI-3-kinase signaling rather than Notch signaling. However, previous reports from our lab documenting an increase in Notch1 and its signaling in advanced MF[57] confirmed by the Notch1 overexpression in TMR seen here, seem to indicate that VAV1-Notch1 axis may play in MF progression.

The notion that MF is a genetically stable tumor and does not hypermutate[58] has recently been refuted by genome and exome sequencing data showing numerous mutations and subclonal diversification even in early stages.[10,17,59] The high rate of mutations can be explained in part by our current findings showing dysregulation of major pathways controlling DNA repair mechanisms (nucleotide excision repair, mismatch repair and base excision repair). Although DNA repair mechanisms have not been studied in MF in detail, it is known that mutations affecting DNA repair genes such as *POLE*, *MYCN*, or *ATM* may contribute to genomic instability in CTCL.[17,60–62]

Another, novel mechanism of genomic instability might be related to abnormal regulation of mitosis. We identified ectopic expressions of five testis-specific genes (*CT45A1*, *CT45A3*, *STAG3*, *GTSF1*, and *REC8*) which are only expressed in testicular germ cells. Ectopic expression of cancer-testis (CT) antigens was previously found in carcinomas of the bladder,[63] lung,[64] liver[65] and suspected to play an important role in maintaining cell survival through inhibition of apoptosis (i.e. down regulating p53 and p21 tumor suppressor genes) and promotion of chromosomal instability (i.e. generation of double-strand breaks leading to loss of heterozygosity and chromosomal arrangements). CT genes (*SYCP1*,

SYCP3, REC8, SPO11, STAG3, GTSF1) have previously been reported in advanced CTCL and proposed to reflect the reactivation of meiotic and gametogenic programs in mitotic cells.[37,38] Our data documenting a dramatic increase in CT expression (10-30 fold up-regulation in MF tumors) support the putative role of meiomitosis[66,67] as a mechanism of genomic instability in MF.

MF is a spatially discontinuous neoplasm composed of circumscribed lesions of the skin. This calls for a question of whether those lesions evolve in isolation or whether malignant cells percolate between the lesions. Clinical observations support the latter notion; it is known that the development of the tumor in one area of the skin changes the clinical course to be more unfavourable and that those patients are more likely to develop additional tumors. We have therefore hypothesized that high-grade malignant cells in the tumors escape to the circulation and may metastatically seed the skin - either to the already existing lesions or to the new, previously uninvolved areas. The comparisons of global expression profiles of ESP, LSP and TMR seem to confirm our hypothesis. The principal component clustering of transcriptomes showed that LSP and TMR overlapped in two clusters (clusters B1 and B2, **Figure 2**) and separated well from ESP. The same was previously seen on the mutational level: LSP resembled tumors and differed from the ESP.[17,68] This shows that stage progression is associated with global genomic and transcriptomic changes in all skin lesions, in spite of the fact that the tumor formation is only present in a discrete skin area. Thus, cancer cells in MF are likely to percolate through the skin, occasionally seeding new areas and the existing lesions, the latter case resembling tumor self-seeding or cell exchange between distant metastases **(Figure 5B)**.[15,16,69,70] It is conceivable that early eradication of emerging tumors (e.g. by aggressive radiotherapy) would limit hematogenous percolation of malignant cells and retard the colonization of the skin with high-grade mutated cells.

Acknowledgements

We extend our gratitude to Dr. Thomas Salopek, Mrs. Rachel Doucet, and the nursing staff of Kaye Edmonton Clinic for their assistance with sample collection. This study was supported by grants from the following sources: Canadian Dermatology Foundation (CDF RES0035718), University Hospital Foundation (University of Alberta), Bispebjerg Hospital (Copenhagen, Denmark), Danish Cancer Society (Kræftens Bekæmpelse R124-A7592 Rp12350) and an unrestricted research grant to R.G. from the Department of Medicine, University of Alberta. M.X. was supported by an Alberta Innovates Summer Research Studentship.

Author's contributions

M.X. contributed to study design, conducted bioinformatics, analyzed the data, prepared the figures, and wrote the manuscript. RG designed the experiments, supervised data analysis, prepared the figures, and wrote the manuscript. DH and FZ conducted bioinformatics analysis. AI and SO performed wet-lab experiments. AS contributed to writing and editing the manuscript. All authors reviewed all versions of the paper and approved the final version of this manuscript.

References

1 Willemze R. Mycosis fungoides variants-clinicopathologic features, differential diagnosis, and treatment. *Semin Cutan Med Surg* 2018; **37**:11–7.

2 Johnson VE, Vonderheid EC, Hess AD, *et al.* Genetic markers associated with progression in early mycosis fungoides. *J Eur Acad Dermatol Venereol* 2014; **28**:1431–5.

3 Talpur R, Singh L, Daulat S, *et al.* Long-term outcomes of 1,263 patients with mycosis fungoides and Sézary syndrome from 1982 to 2009. *Clin Cancer Res* 2012; **18**:5051–60.

4 Bernengo MG, Quaglino P, Novelli M, *et al.* Prognostic factors in Sézary syndrome: a multivariate analysis of clinical, haematological and immunological features. *Ann Oncol* 1998; **9**:857–63.

5 Diamandidou E, Colome M, Fayad L, *et al.* Prognostic factor analysis in mycosis fungoides/Sézary syndrome. *J Am Acad Dermatol* 1999; **40**:914–24.

6 Campbell JJ, Clark RA, Watanabe R, Kupper TS. Sezary syndrome and mycosis fungoides arise from distinct T-cell subsets: a biologic rationale for their distinct clinical behaviors. *Blood* 2010; **116**:767–71.

7 Iyer A, Hennessey D, O'Keefe S, *et al.* Clonotypic heterogeneity in cutaneous T-cell lymphoma (mycosis fungoides) revealed by comprehensive whole-exome sequencing. *Blood Adv* 2019; **3**:1175.

8 Chang L-W, Patrone CC, Yang W, *et al.* An Integrated Data Resource for Genomic Analysis of Cutaneous T-Cell Lymphoma. *J Invest Dermatol* 2018; **138**:2681–3.

9 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**:1465–70.

10 Choi J, Goh G, Walradt T, *et al.* Genomic landscape of cutaneous T cell lymphoma. *Nat Genet* 2015; **47**. doi:10.1038/ng.3356.

11 Wang L, Ni X, Covington KR, *et al.* Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat Genet* 2015; **47**:1426–34.

12 Dulmage BO, Geskin LJ. Lessons learned from gene expression profiling of cutaneous T-cell lymphoma. *Br J Dermatol* 2013; **169**:1188–97.

13 van Doorn R, van Kester MS, Dijkman R, *et al.* Oncogenomic analysis of mycosis fungoides reveals major differences with Sezary syndrome. *Blood* 2009; **113**:127–36.

14 Iyer A, Hennessey D, O'Keefe S, *et al.* Skin colonization by circulating neoplastic clones in cutaneous T-cell lymphoma. *Blood* 2019; **134**:1517–27.

15 Kim M-Y, Oskarsson T, Acharyya S, *et al.* Tumor self-seeding by circulating cancer cells. *Cell* 2009; **139**:1315–26.

16 Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. *Nature* 2016; **529**:298–306.

17 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**:2489–500.

18 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**:i884–90.

19 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; **12**. doi:10.1186/1471-2105-12-323.

20 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**:357–9.

21 Yates AD, Achuthan P, Akanni W, *et al.* Ensembl 2020. *Nucleic Acids Res* 2019; **48**:D682–8.

22 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**:139–40.

23 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**:550.

24 Luo W, Friedman MS, Shedden K, *et al.* GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* 2009; **10**:161.

25 The Gene Ontology Consortium, Ashburner M, Ball CA, *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25.

26 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**. doi:10.1093/nar/28.1.27.

27 K Blighe , S Rana, M Lewis. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. https://github.com/kevinblighe/EnhancedVolcano [accessed on 2 January 2021].

28 Michel L, Jean-Louis F, Begue E, *et al.* Use of PLS3, Twist, CD158k/KIR3DL2, and NKp46 gene expression combination for reliable Sézary syndrome diagnosis. *Blood* 2013; **121**:1477–8.

29 Ortonne N, Le Gouvello S, Tabak R, *et al.* CD158k/KIR3DL2 and NKp46 are frequently expressed in transformed mycosis fungoides. *Exp Dermatol* 2012; **21**:461–3.

30 Bagot M, Porcu P, Marie-Cardine A, *et al.* IPH4102, a first-in-class anti-KIR3DL2 monoclonal antibody, in patients with relapsed or refractory cutaneous T-cell lymphoma: an international, first-in-human, open-label, phase 1 trial. *Lancet Oncol* 2019; **20**:1160–70.

31 Shareef MM, Elgarhy LH, Wasfy RE-S. Expression of Granulysin and FOXP3 in Cutaneous T Cell Lymphoma and Sézary Syndrome. *Asian Pac J Cancer Prev* 2015; **16**:5359–64.

32 Bensussan A, Remtoula N, Sivori S, *et al.* Expression and function of the natural cytotoxicity receptor NKp46 on circulating malignant CD4+ T lymphocytes of Sézary syndrome patients. *J Invest Dermatol* 2011; **131**:969–76.

33 Staumont-Sallé D, Fleury S, Lazzari A, *et al.* CX□CL1 (fractalkine) and its receptor CX□CR1 regulate atopic dermatitis by controlling effector T cell retention in inflamed skin. *J Exp Med* 2014; **211**:1185–96.

34 Mionnet C, Buatois V, Kanda A, *et al.* CX3CR1 is required for airway inflammation by promoting T helper cell survival and maintenance in inflamed lung. *Nat Med* 2010; **16**:1305–12.

35 Rousseaux S, Debernardi A, Jacquiau B, *et al.* Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* 2013; **5**:186ra66.

36 Litvinov IV, Tetzlaff MT, Thibault P, *et al.* Gene expression analysis in Cutaneous T-Cell Lymphomas (CTCL) highlights disease heterogeneity and potential diagnostic and prognostic indicators. *Oncoimmunology* 2017; **6**:e1306618.

37 Litvinov IV, Cordeiro B, Huang Y, *et al.* Ectopic expression of cancer-testis antigens in cutaneous T-cell lymphoma patients. *Clin Cancer Res* 2014; **20**. doi:10.1158/1078-0432.CCR-14-0307.

38 Gantchev J, Martínez Villarreal A, Xie P, *et al.* The Ectopic Expression of Meiosis Regulatory Genes in Cutaneous T-Cell Lymphomas (CTCL). *Front Oncol* 2019; **9**. doi:10.3389/fonc.2019.00429.

39 El-Brolosy MA, Stainier DYR. Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet* 2017; **13**:e1006780.

40 Querfeld C, Leung S, Myskowski PL, *et al.* Primary T Cells from Cutaneous T-cell Lymphoma Skin Explants Display an Exhausted Immune Checkpoint Profile. *Cancer Immunol Res* 2018; **6**:900–9.

41 van Kester MS, Borg MK, Zoutman WH, *et al.* A meta-analysis of gene expression data identifies a molecular signature characteristic for tumor-stage mycosis fungoides. *J Invest Dermatol* 2012; **132**:2050–9.

42 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**:653–64.

43 Tracey L, Villuendas R, Dotor AM, *et al.* Mycosis fungoides shows concurrent deregulation of multiple genes involved in the TNF signaling pathway: an expression profile study. *Blood* 2003; **102**. doi:10.1182/blood-2002-11-3574.

44 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**:380–9.e4.

45 Nielsen M, Kaltoft K, Nordahl M, *et al.* Constitutive activation of a slowly migrating isoform of Stat3 in mycosis fungoides: tyrphostin AG490 inhibits Stat3 activation and growth of mycosis fungoides tumor cell lines. *Proc Natl Acad Sci U S A* 1997; **94**. doi:10.1073/pnas.94.13.6764.

46 Sommer VH, Clemmensen OJ, Nielsen O, *et al.* In vivo activation of STAT3 in cutaneous T-cell lymphoma. Evidence for an antiapoptotic function of STAT3. *Leukemia* 2004; **18**:1288–95.

47 McGirt LY, Jia P, Baerenwald DA, *et al.* Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood* 2015; **126**:508–19.

48 Guo X, Wang X-F. A mediator lost in the war on cancer. *Cell* 2012; **151**:927–9.

49 Huang S, Hölzel M, Knijnenburg T, *et al.* MED12 controls the response to multiple cancer drugs through regulation of TGF-β receptor signaling. *Cell* 2012; **151**:937–50.

50 Cani AK, Hovelson DH, McDaniel AS, *et al.* Next-Gen Sequencing Exposes Frequent MED12 Mutations and Actionable Therapeutic Targets in Phyllodes Tumors. *Mol Cancer Res* 2015; **13**:613–

9.

51 Guinney J, Ferté C, Dry J, *et al.* Modeling RAS phenotype in colorectal cancer uncovers novel molecular traits of RAS dependency and improves prediction of response to targeted agents in patients. *Clin Cancer Res* 2014; **20**:265–72.

52 Bunda S, Heir P, Metcalf J, *et al.* CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat Commun* 2019; **10**:661.

53 Weissmann S, Cloos PA, Sidoli S, *et al.* The Tumor Suppressor CIC Directly Regulates MAPK Pathway Genes via Histone Deacetylation. *Cancer Res* 2018; **78**:4114–25.

54 Eleveld TF, Schild L, Koster J, *et al.* RAS-MAPK Pathway-Driven Tumor Progression Is Associated with Loss of CIC and Other Genomic Aberrations in Neuroblastoma. *Cancer Res* 2018; **78**:6297–307.

55 Robles-Valero J, Lorenzo-Martín LF, Menacho-Márquez M, *et al.* A Paradoxical Tumor-Suppressor Role for the Rac1 Exchange Factor Vav1 in T Cell Acute Lymphoblastic Leukemia. *Cancer Cell* 2017; **32**:608–23.e9.

56 Park J, Yang J, Wenzel AT, *et al.* Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). *Blood* 2017; **130**:1430–40.

57 Kamstrup MR, Gjerdrum LMR, Biskup E, *et al.* Notch1 as a potential therapeutic target in cutaneous T-cell lymphoma. *Blood* 2010; **116**:2504–12.

58 Assaf C, Sanchez JAA, Lukowsky A, *et al.* Absence of microsatellite instability and lack of evidence for subclone diversification in the pathogenesis and progression of mycosis fungoides. *J Invest Dermatol* 2007; **127**:1752–61.

59 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**:15483.

60 Woollard WJ, Pullabhatla V, Lorenc A, *et al.* Candidate driver genes involved in genome maintenance and DNA repair in Sézary syndrome. *Blood* 2016; **127**:3387–97.

61 Rickman DS, Schulte JH, Eilers M. The Expanding World of N-MYC-Driven Tumors. *Cancer Discov* 2018; **8**:150–63.

62 Stagni V, Cirotti C, Barilà D. Ataxia-Telangiectasia Mutated Kinase in the Control of Oxidative Stress, Mitochondria, and Autophagy in Cancer: A Maestro With a Large Orchestra. *Front Oncol* 2018; **8**:73.

63 Fradet Y, Picard V, Bergeron A, LaRue H. Cancer-testis antigen expression in bladder cancer. *Prog Urol* 2005; **15**.URL https://pubmed.ncbi.nlm.nih.gov/16734221/ [accessed on 12 January 2021].

64 Djureinovic D, Hallström BM, Horie M, *et al.* Profiling cancer testis antigens in non-small-cell lung cancer. *JCI Insight* 2016; **1**:e86837.

65 Grizzi F, Franceschini B, Hamrick C, *et al.* Usefulness of cancer-testis antigens as biomarkers for the diagnosis and treatment of hepatocellular carcinoma. *J Transl Med* 2007; **5**:3.

66 Tsang M, Gantchev J, Netchiporouk E, *et al.* A study of meiomitosis and novel pathways of genomic

instability in cutaneous T-cell lymphomas (CTCL). *Oncotarget* 2018; **9**:37647–61.

67 Gantchev J, Martínez Villarreal A, Gunn S, *et al.* The ectopic expression of meiCT genes promotes meiomitosis and may facilitate carcinogenesis. *Cell Cycle* 2020; **19**:837–54.

68 Sivanand A, Hennessey D, Iyer A, *et al.* The Neoantigen Landscape of Mycosis Fungoides. *Front Immunol* 2020; **11**. doi:10.3389/fimmu.2020.561234.

69 Brown D, Smeets D, Székely B, *et al.* Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun* 2017; **8**:14944.

70 Reiter JG, Makohon-Moore AP, Gerold JM, *et al.* Reconstructing metastatic seeding patterns of human cancers. *Nat Commun* 2017; **8**:14114.

Figure legends

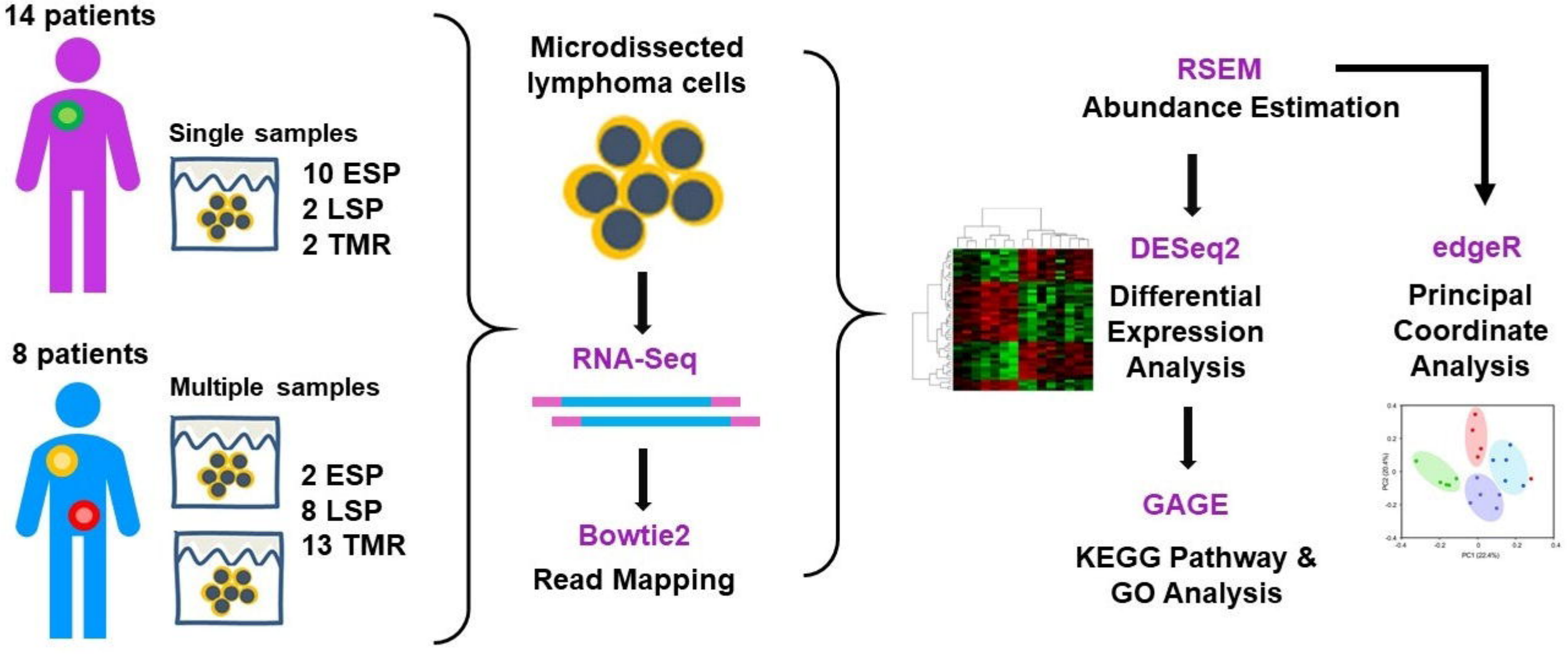
Figure 1. Summary of methods and study design. 32 biopsies of lesional skin were obtained from 23 patients, consisting of 14 patients that contributed single biopsies and 8 patients where multiple samples were taken. The lesions were categorized according to the both the clinical stage (advanced vs. early) as well as the morphological feature (plaque vs lesion): ESP (n=12 early stage plaque biopsies from patients in stage IA-IIA), LSP and TMR (n=10 late-stage plaques and 15 tumors biopsies from patients in stage ≥IIB, respectively). After RNA-Seq, reads were aligned to the reference sequence after removal of adapter sequences using Bowtie2. Transcript abundance was measured using RSEM. Two statistical methods were used for differential expression analysis and principal coordinate analysis (DESEq2 and edgeR, respectively). Generally applicable gene set enrichment (GAGE) analysis was performed to provide more information about the biological functions and pathways significantly enriched in each disease stage.

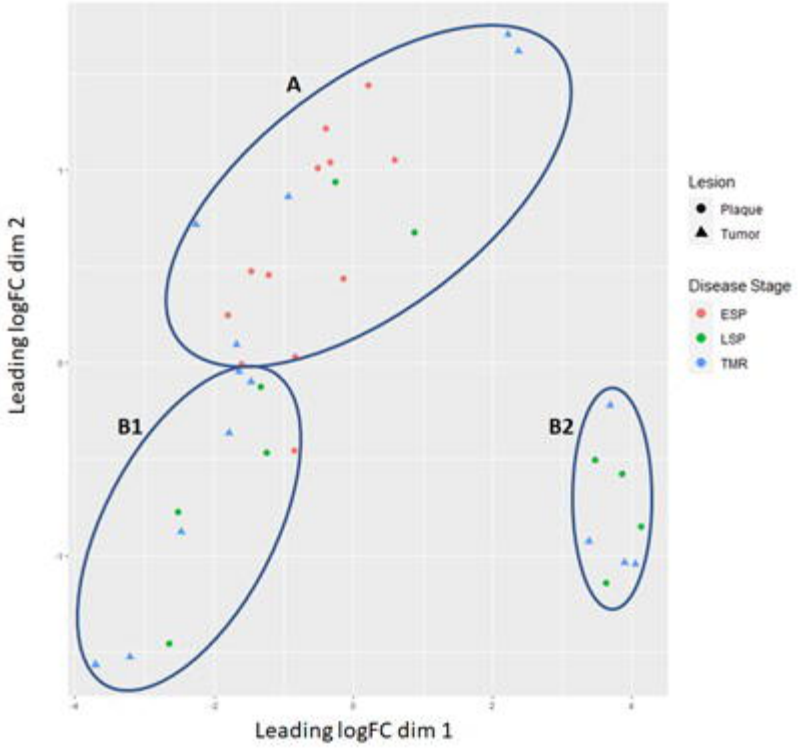
Figure 2. Comparative transcriptome analysis of ESP, LSP, and TMR lesions. A. Unsupervised Principal coordinate analysis (PoCA) showing the variation and clustering among ESP, LSP, and TMR lesion types, with ellipses drawn manually. LSP resembles TMR (clusters B1, B2), rather than ESP (cluster A). **B.** Volcano plot of differential expressed genes (DEGs) between TMR (comparison) and ESP (reference) group, generated using the EnhancedVolcano R package. The vertical y-axis) corresponds to the mean expression value of log10 (adjusted P value) after correction for multiple testing, and the horizontal x-axis displays the log2 fold change value. 1,154 transcripts with an adjusted p-value < 0.05 and |log2 (fold change)| ≥ 1 were mapped as DEGs (depicted as red dots). Genes that failed to reach statistical significance are present in green, while transcripts below fold-change cutoffs are presented in blue. Genes upregulated or downregulated in TMR compared to ESP are in the upper right and upper left quadrants, respectively. **C.** Volcano plot for the distribution of gene expression between LSP (comparison) and ESP (reference), mapping 26 DEGs (red dots). **D.** Volcano plot for the distribution of gene expression between TMR (comparison) and LSP (reference), mapping 29 DEGs (red dots).

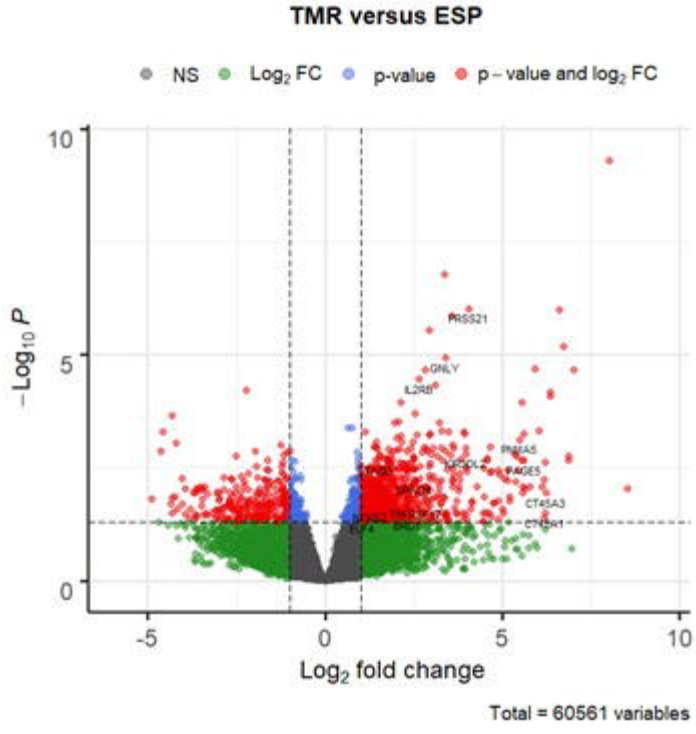
Figure 3. Venn diagrams analysis of differentially expressed genes. The yellow, red and green circles represent the number of up or down-regulated DEGs between each pairwise comparison (TMR vs. ESP, TMR vs. LSP, and LSP vs. ESP). The overlapping number represents mutual, concordantly changed DEGs common to each comparison and the non-overlapping number represents genes unique to each comparison.

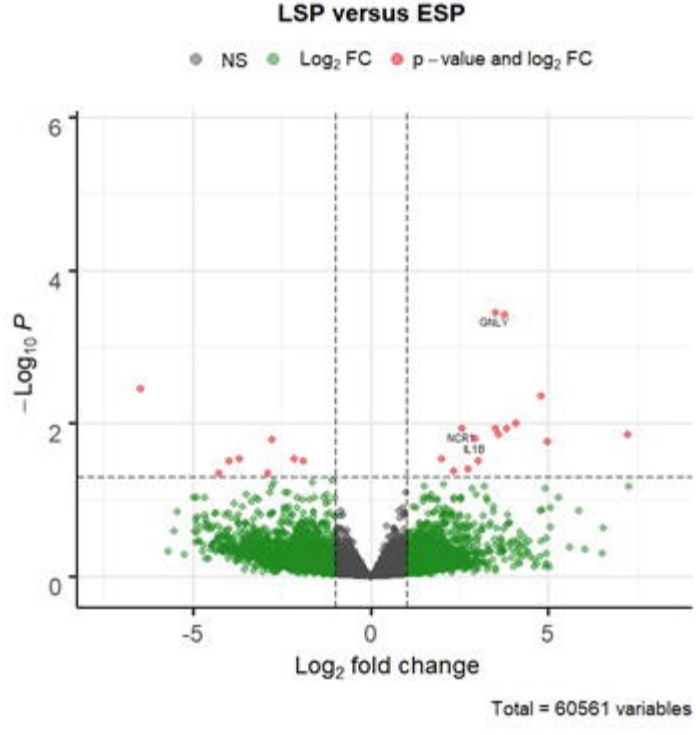
Figure 4. KEGG Pathway and GO enrichment analysis of early (ESP) and advanced staged MF (LSP or TMR) using all available gene expression data. A. Heatmaps of select enriched KEGG pathways and GO biological process terms (rows) that are significantly over-represented in TMR vs ESP. **B.** Heatmaps of select enriched KEGG pathways and GO biological process terms (rows) that are significantly over-represented in LSP vs ESP. **C.** Venn diagram showing GO and KEGG annotations common to both LSP and TMR relative to ESP, notably 'response to interleukin-4', 'respiratory burst', 'TOR signaling', 'DNA replication' and 'base excision repair'. KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology.

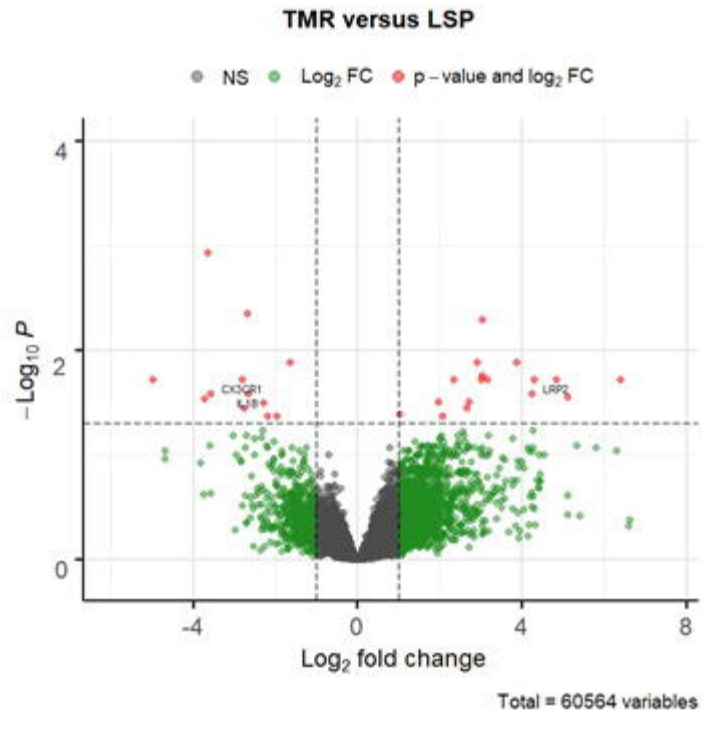
Figure 5. Proposed mechanism of stage progression in MF. A. Frequently mutated genes in advanced lesions of MF. The frequencies of non-synonymous mutations (missense or damaging: frameshift mutations, short read insertions and deletions (<6pm), stop gain or stop loss) are shown. Only genes showing compensatory overexpression (significantly increased Log2FC >1, adjusted P <0.05 mRNA levels in TMR vs ESP) are shown. **B.** Possible involvement mutated genes and upregulated signaling pathways during stage progression. The pathways are marked by red ellipses, mutated genes affecting the pathway are plotted in yellow rectangles. **C.** The proposed mechanism of propagation of mutated cells to skin lesions. Early stage plaques contain cells characterized by a low mutation burden[17] and a characteristic transcriptional signature as shown in *Figure 2.* Accumulation of mutations (colored dots in the "nuclei") changes the transcriptional signature which enables higher rate of proliferation and enhanced cell survival which lead to the formation of a tumor (high-grade malignant cells marked in red). Those cells can enter the circulation, e.g. due to loss of expression of CX3CR1 and can colonize the already existing plaques (cancer self-seeding) or new areas of the skin in a quasi-metastatic process as proposed by us previously[14].

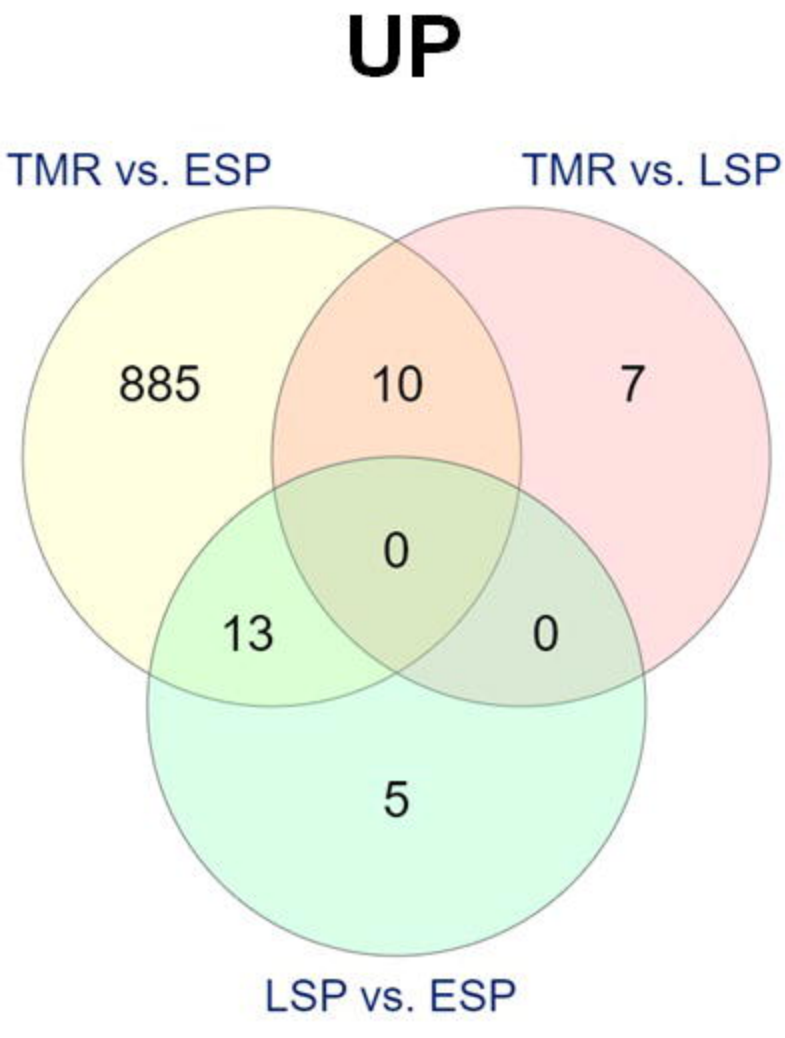

14 patients
Single samples
10 ESP
2 LSP
2 TMR
8 patients
Multiple samples
2 ESP
8 LSP
13 TMR
Microdissected lymphoma cells
RNA-Seq
Bowtie2
Read Mapping
RSEM
Abundance Estimation
DESeq2
Differential Expression Analysis
GAGE
KEGG Pathway & GO Analysis
edgeR
Principal Coordinate Analysis

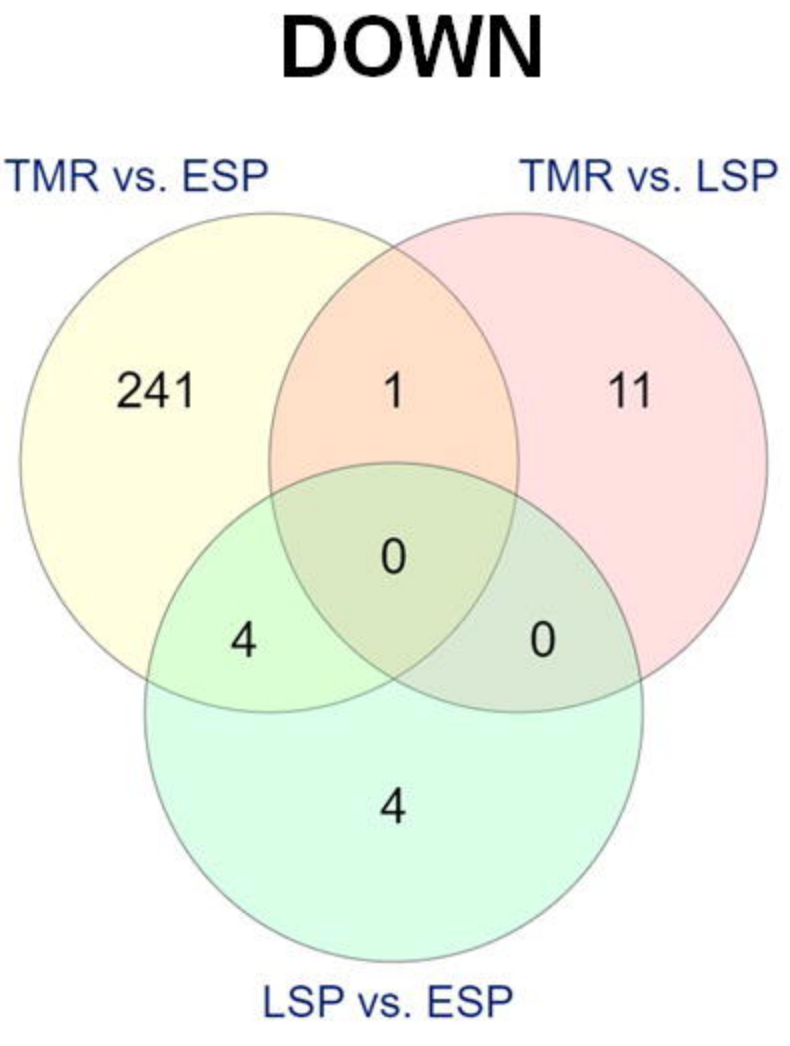

A
B1
B2
Leading logFC dim 2
Leading logFC dim 1
Lesion
Plaque
Tumor
Disease Stage
ESP
LSP
TMR


TMR versus ESP
NS
Log2 FC
p-value
p – value and log2 FC
−Log10 P
Log2 fold change
PRSS21
GNLY
IL2RB
PRAME
PAGE5
CT45A3
Total = 60561 variables

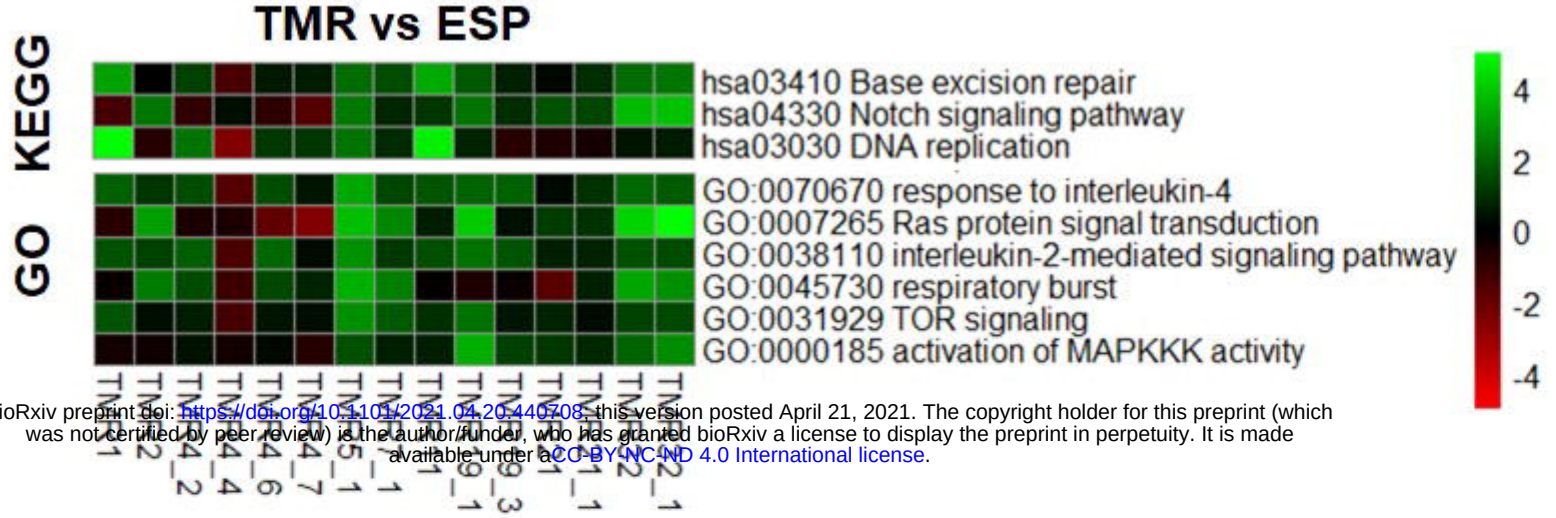

LSP versus ESP
NS
Log2 FC
p – value and log2 FC
−Log10 P
Log2 fold change
GNLY
NCR1
IL18
Total = 60561 variables


TMR versus LSP
NS
Log2 FC
p – value and log2 FC
−Log10 P
Log2 fold change
CXCR1
IL18
LRP2
Total = 60564 variables

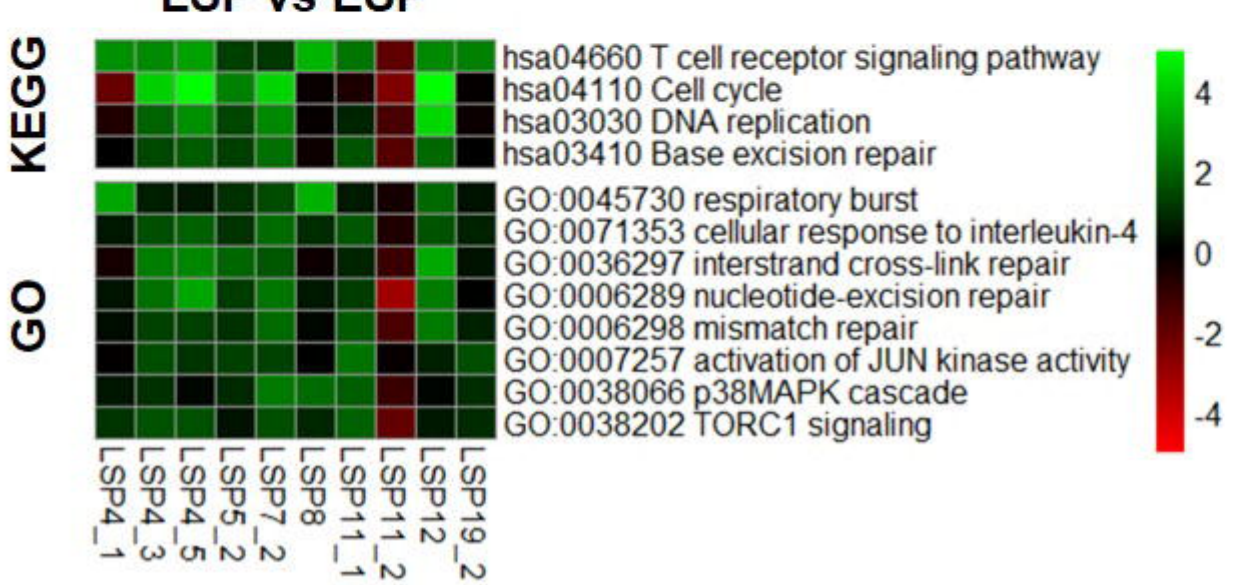

UP
TMR vs. ESP
TMR vs. LSP
885
10
7
0
13
0
5
LSP vs. ESP


DOWN
TMR vs. ESP
TMR vs. LSP
241
1
11
0
4
0
4
LSP vs. ESP

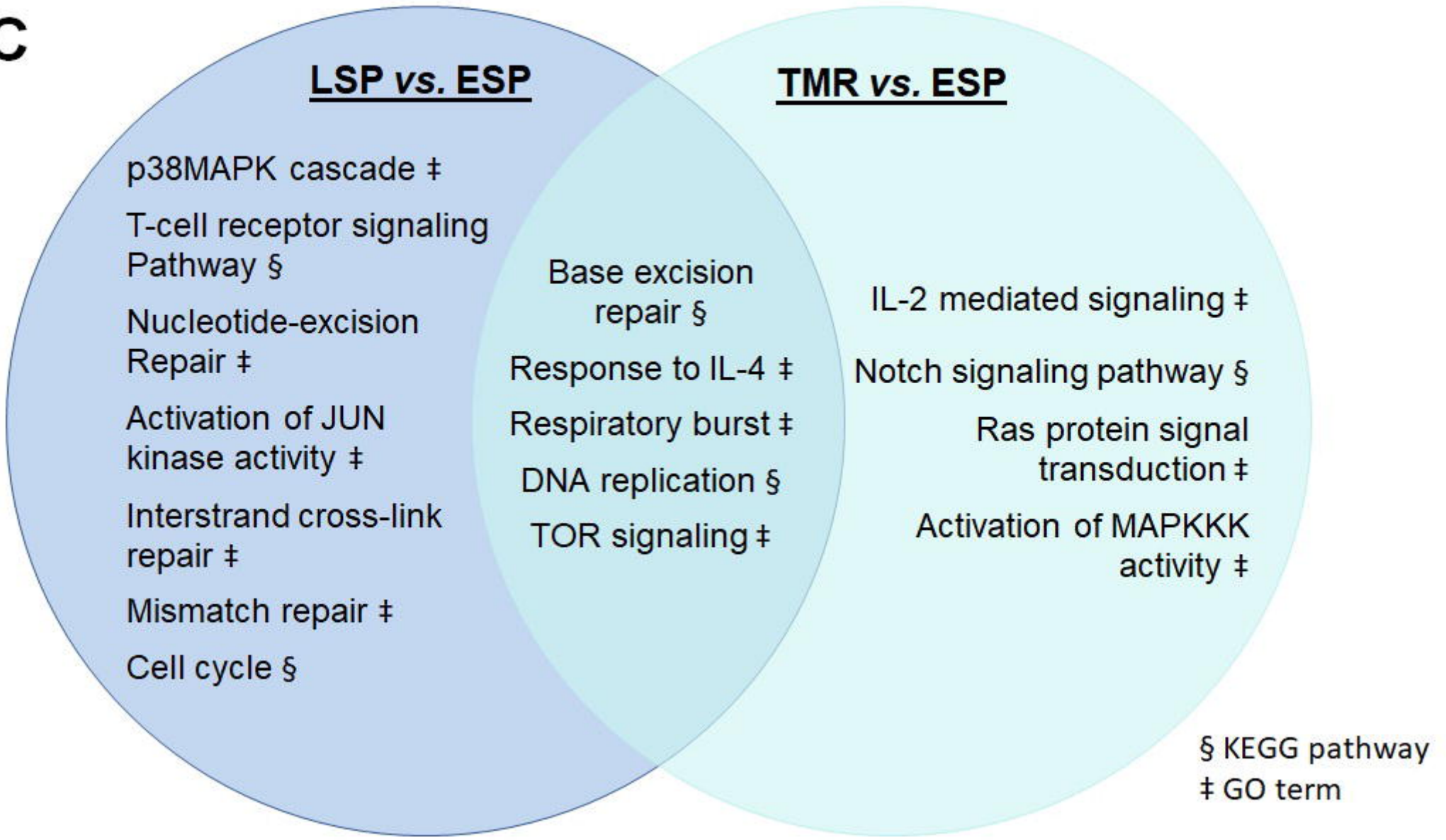
A

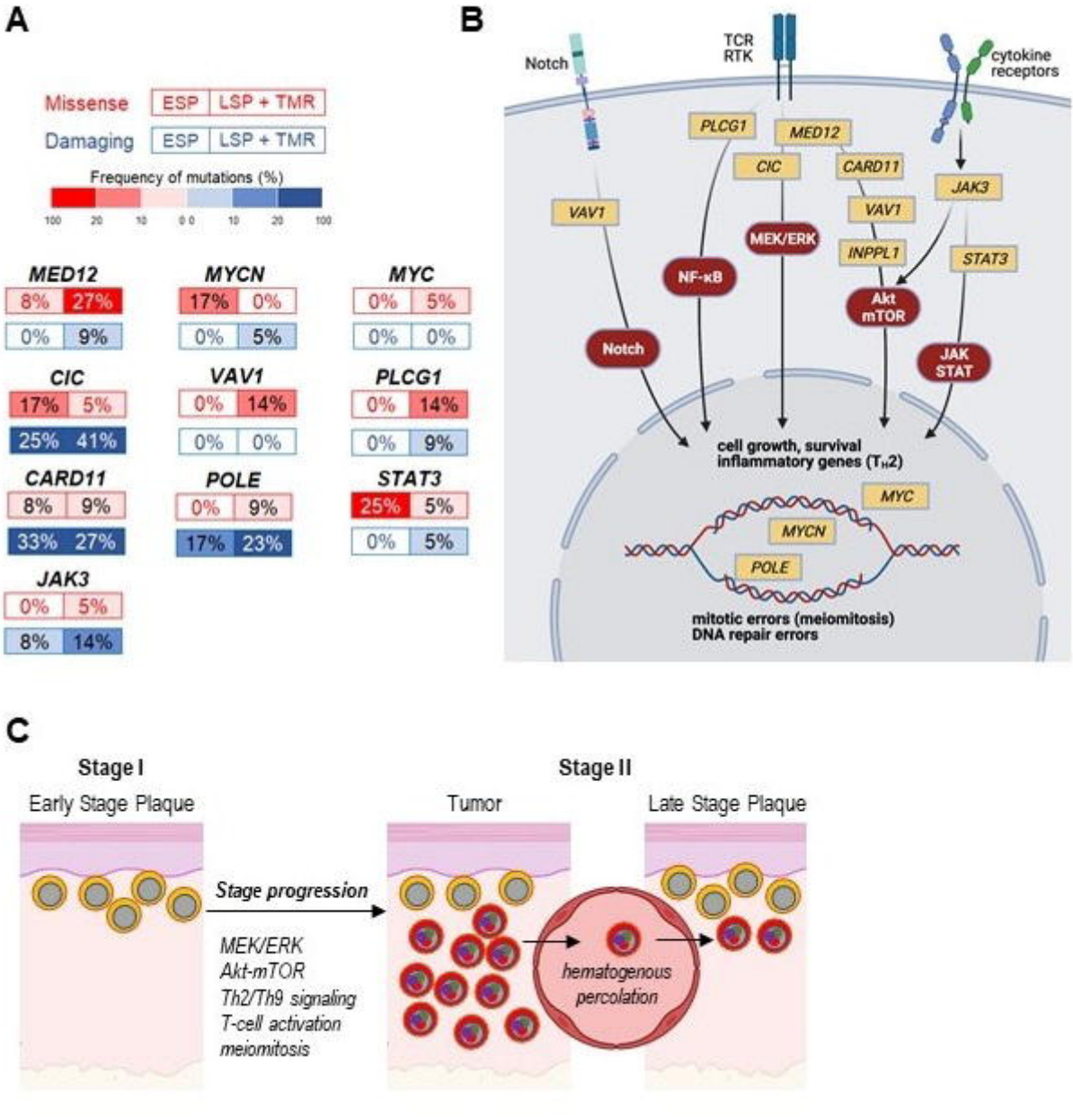


B



C




A
Missense
ESP
LSP + TMR
Damaging
ESP
LSP + TMR
Frequency of mutations (%)
100
20
10
0 0
10
20
100
MED12
8%
27%
0%
9%
MYCN
17%
0%
0%
5%
MYC
0%
5%
0%
0%
CIC
17%
5%
25%
41%
VAV1
0%
14%
0%
0%
PLCG1
0%
14%
0%
9%
CARD11
8%
9%
33%
27%
POLE
0%
9%
17%
23%
STAT3
25%
5%
0%
5%
JAK3
0%
5%
8%
14%
B
Notch
TCR
RTK
cytokine
receptors
PLCG1
MED12
CIC
CARD11
JAK3
VAV1
VAV1
MEK/ERK
INPPL1
STAT3
NF-κB
Akt
mTOR
Notch
JAK
STAT
cell growth, survival
inflammatory genes (TH2)
MYC
MYCN
POLE
mitotic errors (meiomitosis)
DNA repair errors
C
Stage I
Early Stage Plaque
Stage II
Tumor
Late Stage Plaque
Stage progression
MEK/ERK
Akt-mTOR
Th2/Th9 signaling
T-cell activation
meiomitosis
hematogenous
percolation