Oncogenomic analysis identifies novel biomarkers for tumor stage mycosis fungoides

Zhengbang Dong, MD<a,b>, Xiaomei Zhu, MD<a>, Yang Li, MD<c>, Lu Gan, MD<a>, Hao Chen, MD<a>, Wei Zhang, MD<a>, Jianfang Sun, MD<a>.

Abstract
Patients with mycosis fungoides (MF) developing tumors or extracutaneous lesions usually have a poor prognosis with no cure has so far been available. To identify potential novel biomarkers for MF at the tumor stage, a genomic mapping of 41 cutaneous lymphoma biopsies was used to explore for significant genes.

The gene expression profiling datasets of MF were obtained from Gene Expression Omnibus database (GEO). Gene modules were simulated using Weighted Gene Co-expression Network Analysis (WGCNA) and the top soft-connected genes (hub genes) were filtered with a threshold (0.5). Subsequently, module eigengenes were calculated and significant biological pathways were enriched based on the KEGG database.

Four genetic modules were simulated with 3263 genes collected from the whole genomic profile based on cutoff values. Significant diseases genetic terminologies associated with tumor stage MF were found in black module. Subsequently, 13 hub genes including CFLAR, GCNT2, IFNG, IL17A, IL22, MIP, PLCG1, PTH, PTEN, REG1A, SNAP25, SUPT7L, and TP63 were shown to be related to cutaneous T-cell lymphoma (CTCL) and adult T-cell lymphoma/leukemia (ATLL).

In summary, in addition to the reported genes (IL17F, PLCG1, IFNG, and PTH) in CTCL/ATLL, the other high instable genes may serve as novel biomarkers for the regulation of the biological processes and molecular mechanisms of CTLT (MF/SS).

Abbreviations: ATLL = adult T-cell lymphoma/leukemia, CTCL = cutaneous T-cell lymphoma, GEO = gene expression omnibus, IFN-γ = interferon gamma, iNOS = inducible nitric oxide synthase, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = mycosis fungoides, PLCG1 = phospholipase Cγ1, SEA = staphylococcal enterotoxin-A, SS = Sezary syndrome, WGCNA = Weighted Gene Co-expression Network Analysis.

Keywords: gene co-expression, Gene Expression Omnibus database, mycosis fungoides, tumor stage.

1. Introduction
Mycosis fungoides (MF), also known as granuloma fungoides, is the most common form of cutaneous T-cell lymphoma (CTCL) and represents a complex series of diseases with various manifestations and treatment considerations.<sup>[1]</sup> MF has the long-term natural progression, a few years or sometimes decades of characteristics and the development of more infiltration plaques and eventually the formation of tumors, also known as tumor stage MF.<sup>[2]</sup> It is a relatively rare non-Hodgkin’s lymphoma with a stable incidence of 0.36/10⁵ from 1973 to 1992.<sup>[3]</sup> The rate of progression of the disease is variable, so patients may express patches, plaques, and tumors simultaneously in different areas of their skin and initially exhibit extracutaneous involvement.<sup>[4,5]</sup>

Previous studies of MF have shown that the most important prognostic indicators of survival are the type and extent of tumor involvement and the manifestation of extracutaneous disease.<sup>[6,7]</sup> However, due to the low resolution of comparative genomic hybridization techniques reported in previous studies, identification of specific genes involved in disease progression and prognosis remains very limited.<sup>[8,9]</sup> The development of whole genome analysis techniques has enabled the more complete and accurate characterization of human tumors with the goal of providing prognostic markers and targets for directed therapeutic intervention.<sup>[10]</sup>

In this study, a genetic array profile of numerous tumor-stage MF samples was analyzed in order to analyze genetic abnormalities in MF patients and to explore potential key genes. In addition, specific genetic changes were investigated that may help to understand disease progression and prognostic judgment.

2. Materials and methods

2.1 Datasets collection
The Gene Expression Omnibus (GEO) database is an international public repository that archives and freely distributes high-throughput gene expression and genomics datasets, designed to facilitate the sharing of genomic and clinical data between researchers. Microarray expression profile of tumor stage MF, obtained from GEO database with access number GSE18098,<sup>[11]</sup> was used to identify high- or low-expression genes based on the
average cutoff value of the copy numbers. This data profile was composite of 41 patients (22 males and 19 females) with an age range of 17 to 84 years. This study follows the dissemination and application policy requirements of GEO public data and has been approved by the Ethics Committee of the Institute of Biomedicine Research of the Chinese Academy of Medical Sciences and Peking Union Medical College. All genetic material used in the dataset was collected from the peripheral blood of patients with tumor stage MF. And genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qagen, Hilden, Germany) following the protocol supplied by the manufacturer. The whole genome analysis of patient samples was processed via the Human Genomic CGH 44K microarray platform (Agilent Technologies, Palo Alto, CA).

2.2. Gene co-expression analysis and genome module detection

The gene co-expression network is a multidirectional relationship diagram, in which each node corresponds to a gene. If there is a significant relationship between 2 nodes, then the 2 nodes are connected to each other with an edge. In contrast to standard analysis of differential gene expression, which attempts to detect the association of individual genes with the disease, this strategy aims to distinguish higher-order relationships among gene products. The module eigengenes is defined as the important component of the corresponding module of the expression matrix. An important goal of co-expression network analysis is to detect subsets of modules that are highly linked to each other, which is considered as a group of closely co-regulated genes that focus on the coherence of gene network modules.

An important step in the detection of modular aggregates is the use of network proximity metrics to cluster genes into network modules. In short, if a pair of genes i and j are closely linked, they are functionally highly correlated. In general, the maximum expression similarity $s_{ij}$ between 2 genes is 1 and the minimum similarity is 0. Typically, Weighted Gene Co-expression Network Analysis (WGCNA) uses the topological overlap matrix (TOM) $a_{ij}$ as a measure of the proximity among genes and can also be defined using a weighted network: $d_{ij} = (s_{ij})^{\beta}$, where the power $\beta$ is the soft thresholding parameter. The method of dynamic branch cutting is used to define the branches of cluster tree as generated gene modules. Next, the gene within the given module is summarized with the module eigengene, which can be considered as the best summary of the standard module expression data and is defined as the first major component of the normalized expression pattern.

2.3. Metabolic function of gene modules

Enrichment analysis of each gene module is annotated, visualized, and further explored using a database, based on a background list of all the genes on the array. The exact Fisher’s test results were adjusted with Benjamini-Hochberg method to measure the significant level of enrichment terminologies. After that, the core net-map of enrichment results was constructed to explore important metabolic function due to a single gene may interact with more than one terms.

2.4. Hub genes identification

Hub genes were defined as those genes that are highly connected to others within a genetic module. In order to elucidate the importance of highly connected genes and to identify their molecular functions, several screened hub genes of each genetic module were searched in PubMed database. And potential novel progression and prognostic biomarkers that haven’t been reported in tumor stage MF would be found with this identify principles. Gene expression values and corresponding coefficient variations were calculated for each hub gene.

3. Results

3.1. Highly soft-connected genes

After processing of the expression profile dataset with the gene co-expression analysis procedure, a total of 3263 genes were collected with a cutoff value based on the R package WGCNA. Subsequently, each one of the genes was annotated and located in its own chromosome with a polar diagram (left part of Fig. 1).

![Figure 1](image-url). Location of collected genes on their own chromosome and the ideograph of chromosome 1. (Left part of Figure 1: polar ideograph includes the distribution characteristics of 3263 genes collected on 24 chromosomes. The first ring is the ideograph of chromosomes; the second ring represents the histogram of the gene counts; the third ring displays the location of genes corresponding to the distance from the neighboring regions (log-based); the forth ring calculates and adds the genomic density track of the genes; the inner density line represents the genomic instability. Right part of the Figure 1: ideograph and absolute copy number aberrations of chromosome 1 correspond to the P-value of each own genes.)
These results showed that most of our screening genes are located on chromosomes 1, 2, 3, 6, 11, 12, and 19 (Fig. 2A). However, the most unstable chromosomes are 15, 16, 19, 20, and 22. The ideogram of chromosome 1 with cytogenetic band information was shown in the right part of Figure 1.

3.2. Co-expression modules related to tumor stage MF

To identify the functional modules of patients with tumor stage MF, co-expression analysis of the 3263 genes was performed in WGCNA (Fig. 2B). Forty-nine hub genes were found in this co-expression network including ACBD6, ADAMTS7, ANKRD35,
Table 1

| Module | Correlation | P       | Top 5 Hub genes | Notable Pathways (P < .001) |
|--------|-------------|---------|-----------------|-----------------------------|
| Blue   | 0.96        | 0       | AKAP3, APD2, AQP8, ARVC2, ATOH8 | Diseases of signal transduction; class I MHC mediated antigen processing and presentation; signaling by WNT; antigen processing; ubiquitination and proteasome degradation; interleukin-20 family signaling |
| Black  | 0.73        | 5.80E-08| ADMT7, AVRIL, CEP98, CLPB, GCNT2 | RHO GTPase effectors; translation; transcriptional regulation by RUNX1; mitotic prophase; SUMOylation |
| Green  | 0.77        | 3.00E-09| ACTR1B, BTB09, DINA9, SLC30A11, ZNF689 | Intracellular signaling by second messengers; platelet activation, signaling and aggregation; class A1 phosphoinositide-like receptors; PP3 activates AKT signaling; Pten regulation |
| Yellow | −0.89       | 4.20E-15| NR4A, GRIN2B, FG6, ACS1, FBXW7 | Intracellular signaling by second messengers; PIP3AK activates AKT signaling; Pten regulation; negative regulation of the P13K/AKT pathway; P53/P52A and IER3 regulate P13K/AKT signaling |

Correlation of the first module eigengene with diseased state. This table characterizes the modules by Kyoto Encyclopedia of Genes and Genomes pathway enrichment, lists the genes involved in the categories, and by top 5 connected genes according to their K means, which connotes the genes correlation with the module eigengene.
Table 2
Diseases gene enrichment analysis results of black module.

| ID      | Description                                      | P     | Genes                                                                 |
|---------|--------------------------------------------------|-------|----------------------------------------------------------------------|
| C0013080 | Down syndrome                                   | 0.04324| SPR/GNLY/REST/CHAT/COL6A1/FABP7/MAU/FNAR1/LTB/NFATC2/DOPYSLA/EN02/STP1/TCTN1/PRDX3/SNAP25/ACTB/RMY/RBM4/IFNG/AZGP1/CLDN8 |
| C0023493 | ATLL                                            | 0.0003| GCN72/TNFRSF1B/CLAR/C3/PTHTPTN6/SUPT7U/EN02/REG1A/SNAP25/PLCG1/     |
|          |                                                  |       | NRS1/OS3/RPSA/ATF7/IP/LRP1/ASPA/IFNG/IFG/PSMD/10/XIN/PMP            |
| C0014072 | Experimental autoimmune encephalomyelitis       | 0.01142| TRPM4/SLC25A37/GEN1/T1L12RA/ATNR6/1/L22FA/BF7/CLAR/L17A/IFNG/     |
|          |                                                  |       | C3/NFATC2/ETF/12/L2/SC/11/REG1A/VPS52/T2/REG1A/SNAP25/SLC1A6/NRS1/OS3/       |
|          |                                                  |       | RPSA/ATF7/IP/LRP1/TGL/STB/1/PRPG/1/FNG/TMP/MP                     |
| C0023492 | Leukemia, T-cell                                  | 0.00283| CDC20/FOX3/ELF2/PTN/PTP6/SUPT7L/RAC2/REG1A/SNAP25/SLC1A6/NRS1/OS3/ |
|          |                                                  |       | RPSA/ATF7/IP/LRP1/TGL/STB/1/PRPG/1/FNG/TMP/MP                     |
| C0021051 | Immunologic deficiency syndromes                | 0.04917| ADAM8/S/SLC25A37/CEACAM5/LTB/ZAP70/CD3/IG1/C3/EDO3C8/2/ZNF395/PTH/ |
|          |                                                  |       | APCS/MOD/1/RAC2/CDX/8/IRAK/VAMP/8/GFC/F1/FNG/EB3                  |
| C0026180 | Ki-1+ anaplastic large cell lymphoma            | 0.00330| CALM3/GNAL/MYH9/ATNR6/1/L22/CLAR/L17A/EPH4/H/TP63/ZNF395/NFATC2/   |
|          |                                                  |       | PTNP6/IFNG/ANAP25/ZNF395/FNG/FS                      |
| C0079773 | Lymphoma, T-cell, cutaneous                      | 0.02983| TBX2/1/HDAC1/L22/CLAR/L17A/TP63/PTNP6/SUPT7L/REG1A/PLCG1/PTK6/IFNG/FGS/ |
|          |                                                  |       | HST/14H                                     |
| C0035372 | Ret syndrome                                     | 0.0219 | HST/14H/NTNS2/RUNX2/PTTHN2/ZNF36/ACTB/HST/14HC/HST/4U/LUBE2/HST/14H |
| C0036920 | Sezary syndrome                                  | 0.02076| TBX2/1/GCN72/TNTR6/1/L2/17A/PTNP6/SUPT7L/PLCG1/KR3D3/LIFNG      |
| C0024291 | Lymphohistiocytosis, hemophagocytic             | 0.00888| LAM/1/IGLY/STXBP2/ZNF395/PTH/6/ACPS/AS1/FNG/MP                 |
| C0012236 | DIGeorge syndrome                               | 0.04164| TME/12A/HRA/SULT1E1/RAC2/ZNF4/DGCR8                          |
| C0054963 | X-linked lymphoproliferative disorder           | 0.01825| GCN72/L17A/PTPN6/ACPS/FNG                                     |
| C0751674 | Lymphangioleiomyomatosis                         | 0.01825| RPS/N8B1/CALM3/GCN72/TNFRSF1B/DES                          |
| C2316212 | Cryopyrin-associated periodic syndromes         | 0.01077| IL17A/IFNG/1/RAC/FNG                                     |
| C0221269 | Pseudolymphoma                                   | 0.00206| SPH/2/RANGAP1/TP63/PTPN                   |
| C0333909 | Lymphoid hyperplasia                             | 0.00826| IFNG/TNP/MP/1/CSL                      |
| C1619738 | Immune reconstitution inflammatory syndrome     | 0.01266| IL22/L17A/IFNG                                |
| C0265202 | Seckel syndrome                                  | 0.02444| MIP/H/NNMTG/1/GCN72                        |
| C00782196 | MF/S NS                                         | 0.03550| IL17A/SUPT7L/PLCG1                              |
| C0008599 | Polygalandular type-I autoimmune syndrome        | 0.04226| PTN/CYR1A/TFMRG/3                         |
| C0494261 | Combined immunodeficiency                        | 0.04226| GCN72/L17A/IFNG                                      |
| C1900272 | Latent autoimmune diabetes mellitus in adult     | 0.03738| GAD/1/FNG                                     |
| C1262117 | Fungal keratitis                                 | 0.04475| IL17A/IFNG                                     |

**Note:** ATLL = adult T-cell lymphoma/leukemia, MF = mycosis fungoides, NOS = nitric oxide synthase, SS = Sezary syndrome.
PLCG1 had been reported play a key role in promoting cancer metastasis, and blocking its expression could prevent cancer from spreading. And the research is underway and this gene may lead to the development of new anti-cancer drugs.[25]

Interleukin 17A, the same as IL-17A or IL-17, is a pro-inflammatory cytokine produced by a group of T helper 17 cells in response to the stimulation of IL-23.[26] IL-17 function is also essential to a subset of CD4+ T-cells, which results in their roles has been associated with many immune/autoimmune related diseases including rheumatoid arthritis, asthma, lupus, psoriasis,[27] and multiple sclerosis.[28] And recent study have shown that staphylococcal enterotoxin-A (SEA) from the skin of infected CTCL patients and recombinant SEA stimulate the activation of STAT3 and the up-regulation of IL-17, indicating that SEA-producing bacteria promote carcinogenesis as well as activation of previously involved oncogenic pathways.[29]

In addition, Interferon gamma (IFN-γ), known as an immune interferon, has antiviral, immunoregulatory and anti-tumor...
properties and is a product of lectin-stimulated human leukocytes and other antigen-stimulated lymphocytes. It changes a variety of genes that produce different physiological and cellular responses. Such as increasing the antigen presentation and lysosomal activity of macrophages, activating inducible nitric oxide synthase (iNOS), and promoting the adhesion and binding required for leukocyte migration. Besides that, IFN-γ also stimulates macrophages via T-helper cells, making them more powerful in killing intracellular organisms promoting granuloma formation.

However, there are some limitations in this study. First of all, we didn’t explore the differences between tumor stage MF with other subtypes due to limited sample size and datasets. This has affected the potential implications of our findings in tumor stage MF. Second, the limitations of these genes involved in experimental measurements have prevented us from validating their comprehensive predictive ability, and further validation of the results has not yet been applied. The last potential limitation is that potential biomarkers obtained from peripheral blood may not necessarily translate into useful specific tumor tissue biomarkers.

Taken together, despite several identified biomarkers have been reported in our results, the other important genes (e.g., CCLAR, GCNT2, IFNG, and so on) may also play important roles in biological processes and/or molecular functions. These findings may contribute to novel targets investigation for the diagnosis and therapy of tumor stage MF. In this study, we did not report additional comparisons because of the limited number of similar datasets and samples in published studies. Finally, further investigations and verification in the future needs to be compared with the reported biomarkers, such as PCLG1 and IFNG. Verifying the features of these genes reasonably represent the factors that influence the occurrence and development of the disease, so that an optimized risk-adapted diagnosis and treatment strategy can be designed.

Author contributions

Conceptualization: Zhengbang Dong, Jianfang Sun.
Data curation: Zhengbang Dong, Wei Zhang.
Formal analysis: Zhengbang Dong, Hao Chen.
Investigation: Wei Zhang, Yang Li.
Methodology: Yang Li.
Project administration: Jianfang Sun.
Software: Yang Li.
Supervision: Xiaomei Zhu, Hao Chen, Lu Gan.
Writing – original draft: Zhengbang Dong.
Writing – review & editing: Jianfang Sun.

References

[1] Kim YH, Liu H, Mrazekhard S, et al. Long-term outcome of 525 patients with mycosis fungoides and Sezary syndrome: clinical prognostic factors and risk for disease progression. Arch Dermatol 2003;139:857–66.
[2] Van Dam J, Swak MV, Catalano MF, et al. High-grade malignant structure is predictive of esophageal tumor stage risks of endosonographic evaluation. Cancer 2015;71:2910–7.
[3] Weinstock MA, Gardestad B. Twenty-year trends in the reported incidence of mycosis fungoides and associated mortality. Am J Public Health 1999;89:1240–4.
[4] Hwang ST, Jank J, Jaffe ES, et al. Mycosis fungoides and Sezary syndrome. Lancet 2008;371:945–57.
[5] Girardi M, Heald PW, Wilson LD. The pathogenesis of mycosis fungoides. N Engl J Med 2004;350:1978–88.
[6] Hurabille C, Michel L, Ram-Wolf F, et al. Expression of Sezary biomarkers in the blood of patients with erythrodermic mycosis fungoides. J Invest Dermatol 2016;136:317–20.
[7] Kim YH, Bishop K, Varghese A, et al. Prognostic factors in erythrodermic mycosis fungoides and the Sezary syndrome. Arch Dermatol 1995;131:1003–8.
[8] Karenko L, Sarna S, Kåkhön M, et al. Chromosomal abnormalities in relation to clinical disease in patients with cutaneous T-cell lymphoma: a 5-year follow-up study. Br J Dermatol 2003;148:55–64.
[9] Prochazkova M, Chevre T, Manhaguet G, et al. Common chromosomal abnormalities in mycosis fungoides transformation. Genes Chromosomes Cancer 2007;46:828–38.
[10] Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. Nat Genet 2005;37 Suppl:S11–7.
[11] Salgado R, Servitje O, Gallardo F, et al. Oligonucleotide array-CGH identifies genomic subgroups and prognostic markers for tumor stage mycosis fungoides. J Invest Dermatol 2010;130:1126–35.
[12] Stuart JM, Segal E, Koller D, et al. A gene-coexpression network for global discovery of conserved genetic modules. Science 2003;302:249–55.
[13] Oldham MC, Konopka G, Iwamoto K, et al. Functional organization of the transcriptome in human brain. Nat Neurosci 2008;11:1271–82.
[14] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4:1–45.
[15] Zhao W, Langfelder P, Fuller T, et al. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat 2010;20:281–300.
[16] Steve H, Yip AM. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 2007;8:22.
[17] Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008;24:719–20.
[18] Foroushani A, Agrahari R, Docking R, et al. Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigenogene package and its applications. BMC Med Genomics 2017;10:16.
[19] Vaque JP, Gómez-López G, Monsávez V, et al. PCLG1 mutations in cutaneous T-cell lymphomas. Blood 2014;123:2034–43.
[20] Krejsgaard T, Litvinov IV, Wang Y, et al. Elucidating the role of interleukin-17F in cutaneous T-cell lymphoma. Blood 2018;122:943–50.
[21] Suzan Obagi MD, Fred DeRuberto MD, MS MC, et al. Hypercalcemia and parathyroid hormone related protein expression in cutaneous T-cell lymphoma. Int J Dermatol 1999;38:855–62.
[22] Seo T, Tokura Y, Matsunoto T, et al. Tumour-specific cytotoxic T lymphocyte activity in Th2-type Sezary syndrome: its enhancement by interferon-gamma (IFN-γ) and IL-12 and fluctuations in association with disease activity. Clin Exp Immunol 1998;112:403–9.
[23] Nashan D, Faulhaber D, Staiger S, et al. Mycosis fungoides: a dermatological masquerader. Br J Dermatol 2007;156:1–0.
[24] Horvath S. Weighted network analysis: applications in genomics and systems biology. 2011;Springer.
[25] Sala G, Dittn F, Rammold C, et al. Phospholipase Cgamma I is required for metastasis development and progression. Cancer Res 2008;68:10187–96.
[26] Hymowitz SG, Filvaroff EH, Yin JP, et al. IL-17’s adopt a cytokine knot fold: structure and activity of a novel cytokine, IL-17F, and implications for receptor binding. EMBO J 2003;22:5332–41.
[27] Aggarwal S, Gurney AL. IL-17: prototype member of an emerging cytokine family. J Leukoc Biol 2002;71:1–8.
[28] Paul O, Bl EF, Massell BFT. Duckett Jones and his association with Paul Dudley White. Clin Cardiol 1990;13:567–9.
[29] Willerslevolsen A, Krejsgaard T, Lindahl LM, et al. Staphylococcus enterotoxin A (SEA) stimulates STAT3 activation and IL-17 expression in cutaneous T-cell lymphoma. Blood 2016;127:1287–96.
[30] Green JA, Cooperband SR, Kibrick S. Immune specificity and progesterone activity in cutaneous T-cell lymphoma. J Invest Dermatol 2016;136:317–20.
[31] Appelberg R, Moreira D, Barreira Silva P, et al. The Warburg effect in mycobacterial granulomas is dependent on the recruitment and activation of macrophages by interferon. Immunity 2015;45:495–507.