Published in final edited form as:

Clin Cancer Res. 2017 August 15; 23(16): 4855–4864. doi:10.1158/1078-0432.CCR-17-0066.

Genomic Analysis of Thymic Epithelial Tumors Identifies Novel Subtypes Associated with Distinct Clinical Features

Hyun-Sung Lee¹, Hee-Jin Jang¹, Rohan Shah¹, David Yoon¹, Masatsugu Hamaji², Ori Wald¹, Ju-Seog Lee³, David J. Sugarbaker¹, and Bryan M. Burt¹

¹Division of Thoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA

²Department of Thoracic Surgery, Kyoto University Hospital, Kyoto, Japan

³Department of Systems Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

PURPOSE—To reconcile the heterogeneity of thymic epithelial tumors (TETs) and gain deeper understanding of the molecular determinants of TETs, we set out to establish a clinically relevant molecular classification system for these tumors.

EXPERIMENTAL DESIGN—Molecular subgrouping of TETs was performed in 120 patients from The Cancer Genome Atlas using a multidimensional approach incorporating analyses of DNA mutations, mRNA expression, and somatic copy number alterations (SCNA), and validated in two independent cohorts.

RESULTS—Four distinct molecular subtypes of TETs were identified. The most commonly identified gene mutation was a missense mutation in General Transcription Factor II-I (GTF2I group), which was present in 38% of patients. The next group was identified by unsupervised mRNA clustering of GTF2I wild type tumors and represented TETs enriched in expression of genes associated with T cell signaling (TS group; 33%). The remaining 2 groups were distinguished by their degree of chromosomal stability (CS group; 8%) or instability (CIN group; 21%) based upon SCNA analyses. Disease-free survival and overall survival were favorable in the GTF2I group and unfavorable in the CIN group. These molecular subgroups were associated with TET histology and clinical features including disease-free survival. Finally, we demonstrate high expression of PD1 mRNA and correlation of PD1 and CD8A in the TS subgroup.

CONCLUSIONS—Molecular subtyping of TETs is associated with disease-free and overall survival. Classification of TETs by a molecular framework could aid in the refinement of staging, and in discovery and development of rational treatment options for patients with TETs.

Correspondence: Bryan M. Burt, M.D., Division of Thoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Bryan.Burt@bcm.edu, Tel: 713-798-8266.

Data from this manuscript was presented at the 2016 Annual Meeting of the Western Thoracic Surgical Association and the 7th International Thymic Malignancy Interest Group Annual Meeting.

Conflict of Interest: The authors declare no potential conflicts of interest.

ACCESSION CODES

Data deposited into NCBI Gene Expression Omnibus (GEO) Gene expression microarrays GSE57892 and GSE29695 were utilized.

Keywords

thymoma; thymic epithelial tumor; molecular subtyping; mutation; systems biology

INTRODUCTION

Thymic epithelial tumors (TETs) are enigmatic tumors comprised of variable proportions of epithelial and lymphocytic components. TETs are rare tumors with an incidence of 0.32 per 100,000 people every year worldwide, and a simple extrapolation of this would estimate that approximately only 390 TET cases are diagnosed in the United States every year (1). Given the rarity of these tumors and the wide spectrum of their underlying cellular composition, our understanding of TET biology continues to evolve.

Two of the most important factors that bridge the biology and clinical behavior of these tumors are stage and histology. Stage describes of the invasiveness and metastatic capability of these tumors, and the current Masaoka-Koga staging system has proven useful for estimating prognosis in TET patients (2–4). The World Health Organization (WHO) classification system is a description of tumor histology, particularly as it relates to the distribution of epithelial and lymphocytic cellular components of these tumors, and has proven to be an independent prognostic factor for patients with TETs (5,6). The biology of TETs is complex, however, and can result in heterogeneous clinical outcomes even amongst patients of similar stage or histology (7,8). As a result, at least 14 staging systems have been proposed during the last 4 decades, and 5 of these systems have been validated, to some degree, within multiple patient cohorts (4,9,10). Whereas the Masaoka-Koga system is the most common staging system used in clinical practice, tumor heterogeneity exists within the stages defined in this system. For example, capsular invasion which separates Masaoka-Koga stages I and II does not significantly affect survival, and stage III tumors currently includes tumors with heterogeneous prognosis (8). Similarly, with the WHO histological classification system, previous reproducibility studies (7,11) and meta-analyses showed poor agreement on the separation between A and AB TETs, between B1, B2, and B3 TETs, and between B3 thymomas and thymic carcinoma. The molecular determinants of TETs are just beginning to be unraveled (12,13) and will provide additional insight into the stratification of prognosis and treatment of patients with these tumors.

Surgery is a pillar in the treatment of patients with TETS, but effective systemic therapy is critical for patients with for unresectable or metastatic tumors which presently have only 19% 10-year overall survival (OS) (14). Chemotherapy is rarely curative in this patient population and the development of targeted therapies has been hampered by the insufficient characterization of the genetic abnormalities of TETs (6). In addition to stratifying prognosis, molecular classification of TETs might encompass more accurate representation of tumor biology, and as our knowledge of targeted therapeutics advances, molecular subtyping will be important for guiding therapy.

Efforts of The Cancer Genome Atlas (TCGA) research network have provided a deeper understanding of the molecular alterations associated with a variety of tumors (15,16). Only recently, however, has information on the genomic changes associated with TETs become

available. With the advent of newer and more sensitive techniques in molecular biology and bioinformatics, there has been an incremental increase in the availability of high-quality data that will continue to expand our understanding of the molecular pathogenesis of TETs. Such deeper insight into the molecular determinants of TETs has potential to transform the current way that these tumors are staged and how TET patients are treated. To reconcile the heterogeneity of TETs and to gain deeper understanding of the molecular underpinnings of these tumors, we set out to establish a clinically relevant molecular classification system to categorize these tumors.

MATERIALS AND METHODS

Integrative analysis from the TCGA dataset

We obtained available DNA sequencing, mRNA sequencing, DNA copy number alteration, and clinical data of the TET cohort (n=120) from The Cancer Genome Atlas (TCGA) portal (<https://gdc-portal.nci.nih.gov/>) and The Broad Institute TCGA GDAC Firehose (17). The samples of this dataset were obtained from multiple institutes in the United States, Brazil, Canada, France, Germany, and Italy from 2000 to 2013. DNA sequencing data from Illumina HiSeq 2000 DNA Sequencing in Canada's Michael Smith Genome Sciences Centre and RNA sequencing data from Illumina HiSeq2000 RNA Sequencing in the University of North Carolina at Chapel Hill were analyzed. Normalized reads per kilobase per million values in mRNA sequencing data were transformed logarithmically and centralized by subtracting the mean of each channel, and further normalized to equalize the variance. Genomic data were processed as described in previous studies (15,18). Molecular subgrouping of TETs was performed using a multidimensional approach incorporating DNA mutational analyses, unsupervised clustering of mRNA expression data, and somatic copy number alterations (SCNA). In SCNA analyses we regarded chromosomal instability solely on chromosomal deletions as the overwhelming majority of chromosomal aberrations were deletions of tumor suppressors (Supplementary Figure S1) (17).

The TCGA cohort includes 120 patients with primary TETs, in which 119 patients underwent surgical resection at the primary treatment modality and one patient with stage IV underwent cisplatin-based chemotherapy after anterior mediastinotomy. Time of median follow-up was 42.3 (0.5–150) months. Reverse phase protein arrays (RPPA) data on 90 human TETs from TCGA were analyzed.

Analysis of microarray data

Microarray data were analyzed with the Robust MultiArray Average algorithm and implemented quantile normalization with log2 transformation of gene expression intensities with BRB-Array Tools version 4.3.0 (Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA) and the R-script from the Bioconductor project (www.bioconductor.org). Then, we selected the human mRNAs and adjusted data with mean values for genes and arrays, respectively. An unsupervised hierarchical clustering algorithm was applied using the uncentered correlation coefficient as the measure of similarity and the method of average linkage (Cluster 3.0). Java Treeview 1.60 (Stanford University School of Medicine, Stanford, CA, USA) was used for tree visualization.

The 483 genes differentially expressed in TET subgroups between GTF2I mutation and wild type ($P < 10^{-7}$, False Discovery Rate (FDR) $< 10^{-7}$, and fold change > 5) were defined as GTF2I mutation gene set. Two hundred and five genes differentially expressed in TET subgroups between TS signature and others ($P < 10^{-7}$, FDR $< 10^{-7}$, and fold change > 5) were defined as TS gene set. The 124 genes differentially expressed in TET subgroups between CS and others ($P < 0.001$, FDR < 0.07 , and fold change > 2) were defined as CS gene set. Finally, the 381 genes differentially expressed in TET subgroups between CIN and others ($P < 10^{-7}$, FDR $< 10^{-7}$, and fold change > 2) were defined as CIN gene set (Supplementary Table S1). To determine mRNA signatures identifying the molecular subtypes of validation cohort 2 and melanoma cohort (19), we adopted a previously developed model (20). Briefly, gene-expression data in the training set were combined to form a series of classifiers according to the compound covariate predictor (CCP) algorithm and the robustness of the classifier was estimated by the misclassification rate determined during leave-one-out cross-validation (LOOCV) of the training set. After LOOCV, the sensitivity and specificity of the prediction models were estimated by the fraction of samples correctly predicted.

Independent validation

Two independent cohorts from the NCBI Gene Expression Omnibus were used to validate our results [validation cohort 1: GSE57892 (12) (n=22) and validation cohort 2: GSE29695 (21) (n=36)]. In cohort 1, multidimensional approach incorporating DNA mutational analyses, unsupervised clustering of mRNA expression data, and SCNA has been performed. In cohort 2, we validated the four subtypes on the basis of mRNA gene signatures that identified these groups. The WHO classification and Masaoka stage data were available from both cohorts.

Statistical analysis

The association of each subtype with disease-free survival (DFS) and OS in the TCGA TET cohort was estimated using Kaplan-Meier plots and log-rank tests. DFS was defined as the time from surgery to the first confirmed recurrence as well as death, and OS was defined as the time from surgery to death. Development of second primary cancer was excluded from the event of recurrence. Data were censored when a patient was alive without recurrence at last follow-up. Multivariable Cox proportional hazards regression analysis was used to evaluate independent prognostic factors associated with DFS. Covariates in these analyses included those with P-values less than 0.2 in univariable analysis. To assess the association of each molecular subtype with clinical phenotype, we used chi-square or Fisher exact test. Statistical significance was accepted as $P < 0.05$, and all tests were two-tailed. All statistical analyses were performed with SPSS 20.0 (SPSS, Inc., Chicago, Illinois) and R language and software environment (<http://www.r-project.org>). The Ingenuity® pathway analysis (www.ingenuity.com, Ingenuity, Redwood City, CA) was used for gene set enrichment analysis to identify enriched gene sets in all subtypes.

A weighted permutation approach was implemented to test for mutual exclusivity and co-occurrence of alterations across a group of subgroups which is currently available as an additional step when sorting the subgroups by mutual exclusivity. The test is based on weighted permutations assessing the deviation of the observed coverage (number of columns

with a signal) compared to expected obtained by permuting events, maintaining the number of events per row and weighted permutations for columns. Mutual exclusivities of the subtypes were tested by Gitoools (22, 23). Principal component analysis (PCA) was performed on mRNA gene expression data and to analyze the association between the principal components and molecular subtypes relevant to TETs. We generated three-dimensional (3-D) plots by evaluating the association of the first three principal components (PC 1–3) with GTF2I mutation, T cell signaling, and SCNA (24).

RESULTS

Molecular classification of TETs

By integrating genomic data from multiple platforms including mutation data, mRNA expression, and copy number alterations, we used a decision tree categorizing approach to stratify 120 TETs into molecularly distinct subtypes (Figure 1A). General Transcription Factor II-I (**GTF2I**) is the most frequently mutated genes in TETs (missense mutation p.Leu424His in change to codon chromosome 7 c.74146970T>A). Interestingly, this mutation status is well reflected in mRNA expression patterns as all mutated tissues were clustered together in hierarchical clustering analysis ($P=1.17\times 10^{-15}$) (Supplementary Figure S2A). Therefore, we grouped tumors with GTF2I mutation ($n = 46$ [38.3 %]) as first distinct molecular subtype of TET. The second subtype was identified by unsupervised mRNA clustering of GTF2I wild type tumors and represented TETs significantly enriched in expression of genes associated with T cell signaling, hereafter referred to as the **TS** group ($n=39$, 33%) (Cluster C in Supplementary Figure S2B). Analysis of genome-wide SCNA data further identified two additional subtypes; chromosomal stability (**CS**; $n=10$, 8%) or instability (**CIN**; $n=25$, 21%) subtypes. In TETs, the foci significantly deleted were 9p21.3 containing strong tumor suppressors (CDKN2A and CDKN2B), 22q12.1, 6p25.2, 3p22.2, 2q37.1, 6q21, 10q26.3, 9p13.11, 22q13.32, 7q36.3, 11q22.1, and 9q22.33. CDKN2A at 9p21.3 was significantly deleted in the CIN group, with a deletion in this tumor suppressor gene occurring in 7 (28%) of patients in this group. Furthermore, differences in genomic stability in the CS and CIN group were also well reflected in mRNA expression patterns as they have distinct mRNA expression signature (Figure 1B). Mutual exclusivity and principal component analysis were implemented to justify decision-tree approach in supplementary Figure S3.

Molecular subtypes are associated with clinical phenotypes and outcomes

We correlated molecular subtypes with clinical covariates in the TCGA cohort (Figure 2A and Supplementary Table S2) and we observed four main trends: (i) The GTF2I mutation group was prevalent in WHO class A and AB tumors, and was associated with a decreased prevalence of myasthenia gravis (13%). (ii) The TS group occurred predominantly in B1 and B2 WHO classifications. (iii) The CIN group was enriched for B3 and thymic carcinoma histology. (iv) CS and CIN groups were associated with an increased prevalence of myasthenia gravis (40%).

Univariable analyses of DFS within the TCGA dataset demonstrated the expected correlation of Masaoka stage and WHO histology (Supplementary Figure S4). Univariable

Kaplan-Meier survival analyses demonstrated that patients with a GTF2I mutation and patients in the TS group demonstrated favorable DFS and OS and that patients in the CIN group demonstrated unfavorable DFS and OS (Figure 2B). Among 54 patients with advanced WHO classification (B2, B3, and thymic carcinoma), 7 patients (13.0%) had a GTF2I mutation and these patients did not experience recurrence (Figure 2C). Similarly, among 21 patients with Masaoka stage (III–IV), 5 patients (23.8%) had GTF2I mutation and these patients also did not suffer from recurrence (Figure 2D). In contrast, analyses of patients with early WHO classification (A, AB, and B1) and Masaoka stage (I–II) showed that patients characterized as CIN had unfavorable prognosis. Backward stepwise analysis in multivariable Cox regression models including WHO classification, Masaoka stage, and molecular subtyping revealed that molecular subtyping (CIN versus others) was an independent predictor of DFS (HR 3.27; 95% CI 1.37–7.79; $P=0.007$) (Supplementary Table S3).

Independent external validation of molecular subtypes

Two independent cohorts were utilized to validate our findings in the TCGA cohort: cohort 1 ($n=22$, GSE57892) (12) and cohort2 ($n=36$, GSE29695) (21). In these analyses, TET subtypes demonstrated a significant association with WHO classification and Masaoka stage in each independent cohort, consistent with our findings from the TCGA cohort (Figure 3). With respect to SCNA analyses, cohort 1 showed similar findings to the TCGA cohort, that the CIN group was particularly enriched for chromosomal deletions and specifically, deletions in CDKN2A. In cohort 2, DNA mutation and somatic copy number alteration data were not available. Therefore, molecular subtypes were identified on the basis of mRNA signature after proving statistically significant correlation between these mRNA signatures and their respective molecular subgroups within the TCGA cohort and in validation cohort 1 (Figure 4).

Therapeutic implications for molecular subtype-stratified TETs

To explore viable therapeutic strategies for TETs within each molecular subtype, deeper investigation of genomic data was performed in each subtype. mRNA expression of GTF2I and its isoforms in GTF2I mutation group was similar to that of adjacent normal tissue, and of TETs within the CS and CIN groups (Figure 5A) and GTF2I mutation and its mRNA expression was not correlated across all types of cancer in TCGA (Supplementary Figure S5). Using known transcription factor binding site motifs within the TRANSFAC® predicted transcription factor targets dataset (25) (Supplementary Table S4), we identified 220 target genes of the GTF2I transcription factor. Z-scores, which indicated the degree of target gene enrichment, and therefore GTF2I activity for the 220 target mRNAs were calculated and analyzed for each molecular subtype (Figure 5B). In these analyses, the TS subtype demonstrated the lowest GTF2I mRNA and isoform expression, with a distribution of 220 Z-scores that was shifted downward along the x-intercept as the 134th-ranked mRNA (indicating that 134 target mRNAs were depleted and only 86 mRNAs were transcribed). In the case of the CS subtype which demonstrated the highest GTF2I expression, an opposite pattern was observed: 90 mRNAs were depleted and 130 mRNAs were enriched. The curve of GTF2I subtype is located between TS and CIN subtypes, which implies GTF2I mutation may not affect the function of GTF2I gene as a transcription factor.

Canonical pathway analyses were performed to investigate in further detail the biological processes associated with molecular subgroups (Figure 5C). Ingenuity® Pathway Analysis based on differential mRNA expression analyses showed that the GTF2I group was enriched for genes related to human embryonic stem cell pluripotency and Wnt/ β -catenin signaling, which are involved in self-renewal of cells. The TS group had abundant genes related to T cell signaling including CD8A, CD8B, CD3D, as well as genes associated with CTLA4 signaling, T cell receptor signaling, and ICOS/ICOSL signaling. The CS group was associated with pathways involved in G-protein coupled receptor signaling, Toll-like receptor signaling, and regulation of the epithelial-mesenchymal transition (EMT) pathway, and the CIN group with NF- κ B, EGF, FAK, and telomerase signaling. Comparison of protein expression among molecular subgroups revealed findings consistent with the results of IPA analysis. The GTF2I group demonstrated high expression of β -catenin, PIK3CA, and YAP1 proteins that are important in the stem cell pathway. The TS group demonstrated high expression of GATA3, which is known as a master transcription factor for the differentiation of T helper 2 cells (26), and LCK playing the role of T lymphocyte activation and differentiation (27). Tumors in the CS group were associated with downregulation of TSC2, an inhibitory molecule that suppresses the mTOR pathway; and overexpression of collagen VI, which is related to the EMT pathway. Finally, the CIN group showed overexpression of SRC kinase and c-kit proteins (Figure 5D).

Given the upregulation of genes associated co-stimulatory and co-inhibitory T cell signaling in the TS group, we hypothesized that such a gene expression profile in TETs may be a reflection of an immunogenic tumor microenvironment. mRNA expression of immune checkpoint inhibitory genes including programmed cell death 1 (PD1), programmed cell death-ligand 1 (PD-L1), and cytotoxic T-Lymphocyte Associated Protein 4 (CTLA4) showed high expression of PD1 in TS group and high expression PD-L1 expression in CS and CIN groups (Figure 5E). mRNA analyses of CD8A and PD1 mRNA revealed CD8A and PD1 expression in some TET tumors of the GTF2I, CS, and CIN groups, but only TETs in the TS group demonstrated high expression of both CD8A and PD1 (Figure 5F), implying an abundance of PD1 expressing CD8 T cells that may respond favorably to immune checkpoint inhibitor therapy.

We therefore next examined whether a TS mRNA signature correlated with response to immune checkpoint inhibitors, in a cohort of patients with melanoma (n=54) who were treated with anti-PD1 or anti-CTLA4 antibodies (19). In this cohort of melanoma patients, individuals whose tumors demonstrated a TS pattern of mRNA expression in early on-treatment tumor biopsies demonstrated the best responses to anti-PD1 therapy (AUC=0.84, P=0.020). Additionally, patients with melanomas demonstrating a TS pattern of mRNA expression in tumor biopsies performed before treatment or early during treatment also showed good response to anti-PD1 or anti-CTLA4 therapy (AUC=0.668, P=0.034) (Figure 5G).

DISCUSSION

The molecular underpinnings of cancer have been delineated in a number of solid human tumors are the basis for personalized medicine approaches that have potential to

revolutionize patient care. The molecular landscape of TETs is just beginning to be uncovered and only a handful of molecular characterization studies have been conducted in these tumors (12,13,28–30). Petrini and colleagues firstly reported that next-generation sequencing in TETs firstly identified a missense mutation in GTF2I at high frequency in type A thymomas with better survival (12). Additionally, thymic carcinomas have been found to carry a higher number of mutations than thymomas with recurrent mutations of known cancer genes, including TP53, CYLD, CDKN2A, BAP1 and PBRM1 (12). Furthermore, the TCGA will be forthcoming with a deeper analysis of TET data. Thus far, however, a framework that assimilates multiple molecular platforms to account for the molecular heterogeneity of these tumors, and which categorizes clinically relevant molecular subtypes has not been reported.

Herein, we report a molecular classification system for TETs that is based upon distinct patterns of genomic alterations across multiple TET patient cohorts, and that is associated with clinical features. This molecular classification system is based upon 3 main platforms. A DNA sequencing platform first identified tumors with a GTF2I mutation, which have remarkably favorable prognosis. An mRNA sequencing or array platform was then utilized to distinguish TETs with molecular subtype with a T cell signaling gene profile (**TS** group), also a group of tumors also with relatively favorable prognosis. Finally, somatic copy number analyses were used to stratify chromosomally stable TETs (**CS**) from a chromosomal instability subtype (**CIN**), which was associated with the worst clinical prognosis. We show applicability of our molecular subtypes in two additional TET cohorts and their consistent and significant association with clinical phenotypes despite the various sources of heterogeneity and cohort differences. These molecular subtypes are therefore robust and discrete.

The GTF2I mutation was the most commonly identified gene mutation in TETs, present in 38% of patients in the TCGA cohort and in 32% of patients in validation cohort 1. TFII-I is a multifunctional protein involved in the transcriptional regulation of several genes that control cell proliferation and developmental processes, and which is stimulated by the binding of TFII-I to the FOS promoter (31). It binds specifically to several DNA sequence elements and mediates growth factor signaling (32). Our analyses of unsupervised clustering of mRNA sequencing data of TET demonstrated that all patients with GTF2I mutation were distinctly clustered in one specific cluster, implying that the GTF2I mutation may be an important driving mutation in TETs even though we failed to show the association between GTF2I mutation and the alteration of transcription target mRNAs.

Our study has potentially important clinical implications, specifically related to estimating prognosis and to selecting treatment for patients suffering from TETs. When considered as a spectrum (ranging from GTF2I to TS to CS to CIN) this molecular classification system was associated with trends in clinical and histological TET features. For example, the GTF2I group was associated with more favorable WHO histology, less advanced Masaoka stage, and the absence of MG; and the CIN group was associated with less favorable WHO histology, more advanced Masaoka stage, and the presence of MG. Despite these relationships, the molecular classification system was independently associated with DFS when adjusted for these factors in multivariable analyses. Also highlighting the utility of this

molecular classification in TET prognosis was the ability of this system to resolve heterogeneous clinical outcomes among patients with tumors considered in the same clinical classification groups. Specifically, amongst patients with advanced stage thymoma (III–IV), the molecular system stratified two discrete cohorts by substantially different survival outcomes.

While targeted therapy for thymoma and thymic carcinoma is in its infancy, deeper studies into the molecular determinants of TETs could result in the development of new targeted therapeutic agents. Similarly, molecular classification of TETs could identify patients who would respond to existing immunotherapeutic agents. For example, a promising preliminary report of a phase II study of pembrolizumab (anti-PD1 inhibitor) in patients with recurrent thymic carcinoma revealed a response rate of 24% (33), and an early phase study of avelumab (anti-PD-L1 IgG1 antibody) demonstrated activity in thymoma with 4 of 7 patients having objective responses (34). We have shown that tumors in the TS group are enriched for genes related to co-stimulatory and co-inhibitory T cell signaling such as PD1. Such tumors may be poised to respond well to immune checkpoint inhibitor therapy, and these data are consistent with reports of PD1 positivity (46%) and programmed death ligand 1 (PD-L1) (23–80%) on TETs (35–38). TETs must be approached cautiously with checkpoint inhibitors, however, as 2 patients in the pembrolizumab study developed serious autoimmune disorders: one case of severe myositis/myocarditis and one case of type I diabetes (33), and all responders in the avelumab study experienced immune-related adverse effects including myositis in 3 patients and enteritis in 1 patient (34). Further, it has been shown in melanoma cohorts that patients with preexisting autoimmune disorders treated with checkpoint inhibitors are generally at increased risk of autoimmune adverse events (39). This is especially important in patients with TETs, a significant fraction of whom will have preexisting MG or other autoimmune disease. For example, our analyses have determined that MG is present in 13% in GTF2I group, 25% in TS group, 40% in CS, and 44% in CIN group. However, moderate to severe immune related adverse events generally can be managed with the interruption of the checkpoint inhibitor and the use of corticosteroid immunosuppression upon the severity of the observed toxicity.

Further, advanced TETs have demonstrated responses to epidermal growth factor receptor (EGFR) inhibitors, cetuximab (40,41) and erlotinib (42,43) in some reports. Our canonical pathway analyses have demonstrated that the CIN molecular subtype is associated with activation of EGF or SRC-FAK signaling, suggesting that tumors displaying this molecular phenotype could have a favorable response to EGFR or SRC-FAK inhibitors. It is reported that TETs responded to multi-targeted receptor tyrosine kinase (RTK) inhibitors, sunitinib (44), and Dasatinib, a novel, oral, multi-targeted kinase inhibitor of Bcr-Abl and SRC family kinases, as well as ephrin receptor kinases, platelet-derived growth factor receptor, and c-Kit (45), although SRC inhibition by saracatinib did not result in any clinical responses in patients with relapsed or refractory TETs (46). Lastly, our data demonstrate that tumors classified in the CS group were enriched for genes associated with EMT pathway and with Toll like receptor signaling, and therefore may be suited best for trials with mTOR inhibitor class drugs to which favorable responses of advanced TETs have been reported (47), and/or toll-like receptor antagonists. The characteristics of four molecular subtypes in TETs are summarized in Figure 6 with the caveat that the prognosis and potential avenues for

investigating treatment are based on limited follow-up data and the indirect association of treatment with pathway analysis and RPPA data, and require prospective study and validation. At the present time, this molecular staging system can serve as a springboard for deeper investigation for targeted molecular therapy in this disease, and with further discovery and experience with targeted molecular therapy of this disease, this system may potentially be useful for assigning prognosis and for stratifying patients to specific therapy.

Several limitations are inherent to our study. Although the number of TET samples in the TCGA cohort is smaller relative to TCGA cohorts of other tumor types, TETs are a rare tumor and a sample size of 120 patients could be considered respectable. Further, the significant association of molecular subtypes and clinical features such as WHO classification and Masaoka staging were validated in 2 additional datasets. Whereas these additional datasets are small compared to the TCGA set, they did not contain overlapping data. Because DNA mutation and SCNA data were unavailable in cohort 2, we constructed and validated mRNA signatures that could identify each of the molecular subtypes for further analysis. Thus, we could not fully validate the prognostic robustness of this molecular subtyping in the additional cohorts. However, despite the lack of validation in survival, the original study related to validation cohort 1 demonstrated that patients with tumors bearing GTF2I mutations (GTF2I subgroup in this manuscript) had a better prognosis than those bearing wild-type GTF2I (96% compared to 70% 10-year survival, respectively) in 204 patients (12). Lastly, clinical data from TCGA have report treatment modality as “surgical resection”, the completeness of resection could not be accounted for and may have biased our survival analyses.

In summary, these data support a GTF2I mutation as the most common gene mutation in TETs, and the presence of this mutation as a correlate of favorable outcome. Deeper investigation into the molecular mechanisms underlying TETs and the TET molecular stratification framework could hasten the clinical utility of this system as an adjunct to clinical staging and can support the discovery and development of rational treatment options for TET patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The results shown here are in part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. And, the authors would like to thank Michelle G. Almarez for her administrative assistance with this manuscript and Dr. Jungnam Joo in Biometric Research Branch, Research Institute and Hospital, National Cancer Center, Korea for her statistical assistance.

This project is supported by Dan L Duncan Cancer Center Pilot Project Grant, Baylor College of Medicine.

References

1. Engels EA. Epidemiology of thymoma and associated malignancies. *J Thorac Oncol.* 2010; 5:S260–S5. [PubMed: 20859116]

2. Masaoka A, Monden Y, Nakahara K, Tanioka T. Follow-up study of thymomas with special reference to their clinical stages. *Cancer*. 1981; 48:2485–92. [PubMed: 7296496]
3. Koga K, Matsuno Y, Noguchi M, Mukai K, Asamura H, Goya T, et al. A review of 79 thymomas: Modification of staging system and reappraisal of conventional division into invasive and non-invasive thymoma. *Pathol Int*. 1994; 44:359–67. [PubMed: 8044305]
4. Ruffini E, Detterbeck F, Van Raemdonck D, Rocco G, Thomas P, Weder W, et al. Thymic carcinoma: a cohort study of patients from the European society of thoracic surgeons database. *J Thorac Oncol*. 2014; 9:541–8. [PubMed: 24736078]
5. Zucali PA, Di Tommaso L, Petrini I, Battista S, Lee HS, Merino M, et al. Reproducibility of the WHO classification of thymomas: practical implications. *Lung Cancer*. 2013; 79:236–41. [PubMed: 23279873]
6. Kelly RJ, Petrini I, Rajan A, Wang Y, Giaccone G. Thymic malignancies: from clinical management to targeted therapies. *J Clin Oncol*. 2011; 29:4820–7. [PubMed: 22105817]
7. Suster S, Moran CA. Problem areas and inconsistencies in the WHO classification of thymoma. *Semin Diagn Pathol*. 2005; 22:188–97. [PubMed: 16711400]
8. Detterbeck FC, Stratton K, Giroux D, Asamura H, Crowley J, Falkson C, et al. The IASLC/ITMIG Thymic Epithelial Tumors Staging Project: proposal for an evidence-based stage classification system for the forthcoming (8th) edition of the TNM classification of malignant tumors. *J Thorac Oncol*. 2014; 9:S65–S72. [PubMed: 25396314]
9. Filosso PL, Ruffini E, Lausi PO, Lucchi M, Oliaro A, Detterbeck F. Historical perspectives: the evolution of the thymic epithelial tumors staging system. *Lung Cancer*. 2014; 83:126–32. [PubMed: 24377980]
10. Kondo K, Monden Y. Therapy for thymic epithelial tumors: a clinical study of 1,320 patients from Japan. *Ann Thorac Surg*. 2003; 76:878–84. [PubMed: 12963221]
11. Verghese ET, den Bakker M, Campbell A, Hussein A, Nicholson AG, Rice A, et al. Interobserver variation in the classification of thymic tumours—a multicentre study using the WHO classification system. *Histopathology*. 2008; 53:218–23. [PubMed: 18752504]
12. Petrini I, Meltzer PS, Kim IK, Lucchi M, Park KS, Fontanini G, et al. A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. *Nat Genet*. 2014; 46:844–9. [PubMed: 24974848]
13. Okumura M, Fujii Y, Shiono H, Inoue M, Minami M, Utsumi T, et al. Immunological function of thymoma and pathogenesis of paraneoplastic myasthenia gravis. *Gen Thorac Cardiovasc Surg*. 2008; 56:143–50. [PubMed: 18401674]
14. Hamaji M, Burt BM. Long-Term Outcomes of Surgical and Nonsurgical Management of Stage IV Thymoma: A Population-Based Analysis of 282 Patients. *Semin Thorac Cardiovasc Surg*. 2015; 27:1–3. [PubMed: 26074100]
15. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489:519–25. [PubMed: 22960745]
16. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014; 511:543–50. [PubMed: 25079552]
17. Broad Institute TCGA Genome Data Analysis Center. SNP6 Copy number analysis (GISTIC2). Broad Institute of MIT and Harvard; 2016.
18. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014; 513:202–9. [PubMed: 25079317]
19. Chen PL, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA, et al. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. *Cancer Discov*. 2016; 6:827–37. [PubMed: 27301722]
20. Jang HJ, Lee HS, Burt BM, Lee GK, Yoon KA, Park YY, et al. Integrated genomic analysis of recurrence-associated small non-coding RNAs in oesophageal cancer. *Gut*. 2017; 66(2):215–25. [PubMed: 27507904]
21. Badve S, Goswami C, Gökmen-Polar Y, Nelson RP Jr, Henley J, Miller N, et al. Molecular analysis of thymoma. *PLoS One*. 2012; 7:e42669. [PubMed: 22912720]

22. Perez-Llamas C, Lopez-Bigas N. Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS One*. 2011; 6:e19541. [PubMed: 21602921]
23. Babur Ö, Gönen M, Aksoy BA, Schultz N, Ciriello G, Sander C, et al. Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol*. 2015; 16:45. [PubMed: 25887147]
24. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015; 21:449–56. [PubMed: 25894828]
25. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database*. 2016; baw100. doi: 10.1093/database/baw100
26. Ho IC, Tai TS, Pai SY. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol*. 2009; 9:125–35. [PubMed: 19151747]
27. Nika K, Soldani C, Salek M, Paster W, Gray A, Etzensperger R, et al. Constitutively active Lck kinase in T cells drives antigen receptor signal transduction. *Immunity*. 2010; 32:766–77. [PubMed: 20541955]
28. Petrini I, Wang Y, Zucali PA, Lee HS, Pham T, Voeller D, et al. Copy number aberrations of genes regulating normal thymus development in thymic epithelial tumors. *Clin Cancer Res*. 2013; 19:1960–71. [PubMed: 23444221]
29. Inoue M, Starostik P, Zettl A, Ströbel P, Schwarz S, Scaravilli F, et al. Correlating genetic aberrations with World Health Organization-defined histology and stage across the spectrum of thymomas. *Cancer Res*. 2003; 63:3708–15. [PubMed: 12839963]
30. Wang Y, Thomas A, Lau C, Rajan A, Zhu Y, Killian JK, et al. Mutations of epigenetic regulatory genes are common in thymic carcinomas. *Sci Rep*. 2014; 4:7336. [PubMed: 25482724]
31. Grueneberg DA, Henry RW, Brauer A, Novina CD, Cheriya V, Roy AL, et al. A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes Dev*. 1997; 11:2482–93. [PubMed: 9334314]
32. Roy AL. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene*. 2001; 274:1–13. [PubMed: 11674993]
33. Giaccone G, Thompson J, Crawford J, McGuire C, Manning M, Subramaniam DS, et al. A phase II study of pembrolizumab in patients with recurrent thymic carcinoma. *J Clin Oncol*. 2016; 34(suppl) abstr 8517.
34. Rajan A, Heery CR, Perry S, Keen C, Mammen AL, Berman AW, et al. Safety and clinical activity of anti-programmed death-ligand 1 (PD-L1) antibody (ab) avelumab (MSB0010718C) in advanced thymic epithelial tumors (TETs). *J Clin Oncol*. 2016; 34(suppl) abstr e20106.
35. Padda SK, Riess JW, Schwartz EJ, Tian L, Kohrt HE, Neal JW, et al. Diffuse High Intensity PD-L1 Staining in Thymic Epithelial Tumors. *J Thorac Oncol*. 2015; 10:500–8. [PubMed: 25402569]
36. Katsuya Y, Fujita Y, Horinouchi H, Ohe Y, Watanabe S, Tsuta K. Immunohistochemical status of PD-L1 in thymoma and thymic carcinoma. *Lung Cancer*. 2015; 88:154–9. [PubMed: 25799277]
37. Yokoyama S, Miyoshi H, Nakashima K, Shimono J, Hashiguchi T, Mitsuoka M, et al. Prognostic value of programmed death ligand 1 and programmed death 1 expression in thymic carcinoma. *Clin Cancer Res*. 2016; 22:4727–34. [PubMed: 27166394]
38. Katsuya Y, Horinouchi H, Asao T, Kitahara S, Goto Y, Kanda S, et al. Expression of programmed death 1 (PD-1) and its ligand (PD-L1) in thymic epithelial tumors: impact on treatment efficacy and alteration in expression after chemotherapy. *Lung Cancer*. 2016; 99:4–10. [PubMed: 27565906]
39. Menzies AM, Johnson DB, Ramanujam S, Atkinson VG, Wong AN, Park JJ, et al. Anti-PD-1 therapy in patients with advanced melanoma and preexisting autoimmune disorders or major toxicity with ipilimumab. *Ann Oncol*. 2016 Sep 29. pii: mdw443. doi: 10.1093/annonc/mdw443
40. Palmieri G, Marino M, Salvatore M, Budillon A, Meo G, Caraglia M, et al. Cetuximab is an active treatment of metastatic and chemorefractory thymoma. *Front Biosci*. 2007; 12:757–61. [PubMed: 17127335]
41. Farina G, Garassino MC, Gambacorta M, La Verde N, Gherardi G, Scanni A. Response of thymoma to cetuximab. *Lancet Oncol*. 2007; 8:449–50. [PubMed: 17466903]

42. Christodoulou C, Murray S, Dahabreh J, Petraki K, Nikolakopoulou A, Mavri A, et al. Response of malignant thymoma to erlotinib. *Ann Oncol.* 2008; 19:1361–2. [PubMed: 18534960]
43. Pedersini R, Vattei E, Lusso MR, Mazzoleni G, Ebner H, Graiff C. Erlotinib in advanced well-differentiated thymic carcinoma with overexpression of EGFR: a case report. *Tumori.* 2008; 94:849–52. [PubMed: 19267104]
44. Thomas A, Rajan A, Berman A, Tomita Y, Brzezniak C, Lee MJ, et al. Sunitinib in patients with chemotherapy-refractory thymoma and thymic carcinoma: an open-label phase 2 trial. *Lancet Oncol.* 2015; 16:177–86. [PubMed: 25592632]
45. Chuah C, Lim TH, Lim AS, Tien SL, Lim CH, Soong R, et al. Dasatinib induces a response in malignant thymoma. *J Clin Oncol.* 2006; 24:e56–e8. [PubMed: 17135636]
46. Gubens MA, Burns M, Perkins SM, Pedro-Salcedo MS, Althouse SK, Loehrer PJ, Wakelee HA. A phase II study of saracatinib (AZD0530), a Src inhibitor, administered orally daily to patients with advanced thymic malignancies. *Lung Cancer.* 2015; 89(1):57–60. [PubMed: 26009269]
47. Wheler J, Hong D, Swisher SG, Falchook G, Tsimberidou AM, Helgason T, et al. Thymoma patients treated in a phase I clinic at MD Anderson Cancer Center: responses to mTOR inhibitors and molecular analyses. *Oncotarget.* 2013; 4:890–8. [PubMed: 23765114]

STATEMENT OF TRANSLATIONAL RELEVANCE

Thymic epithelial tumors (TETs) are enigmatic tumors comprised of epithelial and lymphocytic components. Subgrouping of TETs with a multidimensional molecular approach identified 4 molecular subtypes of TETs. The first group was identified by the most commonly identified gene mutation, a missense mutation in General Transcription Factor II-I (GTF2I group; 38%). The next group was identified by unsupervised mRNA clustering demonstrating enrichment in the expression of genes associated with T cell signaling (TS group; 33%). The remaining 2 groups were distinguished by their degree of chromosomal stability (CS group; 8%) or instability (CIN group; 21%), based upon somatic copy number alteration. Classification of TETs by a molecular framework could aid in the refinement of staging, and in discovery and development of rational treatment options for patients with TETs.

Figure 1. Four Distinct Molecular Subtypes of TETs

A. Illustration of the molecular subtype classification tree.

B. TETs are divided into four subtypes by a multidimensional approach incorporating DNA mutational analyses, unsupervised clustering of mRNA expression data, and somatic copy number alterations (SCNA). GTF2I mutation-positive (GTF2I, blue), T cell signaling (TS, green), chromosomally stable (CS, orange), and chromosomal instability (CIN, red). Clinical (top) and molecular data (bottom) in 120 tumors from the TCGA cohort profiled with mRNA expression and somatic copy number alteration (SCNA) are depicted.

Figure 2. Clinical and histological characteristics among molecular subtypes of TETs

A. World Health Organization (WHO) histologic classification, Masaoka-Koga stage, and presence of myasthenia gravis are shown for each of the molecular subtypes of TETs in the TCGA dataset.

B. Disease-free survival (DFS) and overall survival (OS) were examined in each of the 4 molecular subtypes of TETs in the TCGA cohort. C. DFS was examined in subgroups of patients with early stage (WHO classification A, AB, and B1) and advanced stage (WHO classification B2, B3, and thymic carcinoma) tumors. D. DFS was examined in subgroups of

patients with early stage (Masaoka-Koga stage I & II) and advanced stage (Masaoka-Koga stage III & IV) tumors.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Validation of TET molecular subtypes in external independent cohorts
Cohort 1 included DNA mutation, mRNA expression, and somatic copy number alteration data, and molecular subtyping was performed in the same manner as in the TCGA cohort. In cohort 2, DNA mutation and somatic copy number alteration data were not available and therefore molecular subtypes were identified on the basis of mRNA signature.

Figure 4. Validation of mRNA signature of molecular subtypes through TCGA cohort and validation cohort 1

A. A schematic overview of the strategy used for constructing prediction models and evaluating predicted outcomes based on gene expression signatures. CCP Compound covariate predictor, LOOCV leave-one-out cross-validation.

B. Receiver Operating Characteristic (ROC) curve between each molecular subtype and the probability of molecular subtypes predicted by mRNA signature. ROC curve in CS subgroup was not generated due to $n=2$.

Figure 5. Therapeutic Implication through Pathway Analysis and Proteomic Patterns of Molecular Subtypes of TETs

A. mRNA expression of GTF2I and its isoforms in four subtypes and adjacent normal tissues. A asterisk (*) denotes P<0.05 compared with other three subtypes.

B. A rank distribution plot of GTF2I mRNA target gene expression. Z-scores of 220 target mRNA targets from TRANSFAC[®] were sorted from the smallest to the largest values and plotted against an anonymous x-axis. The average of Z score in TS subtype was significantly lower than that in others (P=0.002).

C. The Top Canonical Pathways derived from Ingenuity[®] Pathway Analysis (IPA) in Molecular Subtypes of TET. We performed the Ingenuity[®] pathway analysis to find major canonical pathway of each molecular subtype. These pathways in Y-axis emerged following the core analysis in the IPA. The X-axis indicates the significance level scored as -log(P-value) from Fisher's exact test.

D. Proteomic patterns associated with the molecular subtypes of TETs. Reverse phase protein arrays (RPPA) data on 90 human TETs from TCGA in signaling pathways were analyzed according to the molecular subtypes of TETs. Critical values for two-sided significance level of 0.1 are Z=-1.28 and Z=1.28.

E. mRNA expression of immune checkpoint inhibitory genes. The log₂-transformed values of RPKM of PD1, PD-L1, and CTLA4 were shown.

F. A scatter plot of log₂-transformed values of RPKM of PD1 and CD8A. TS subtype was significantly deviated towards high expression of both CD8A and PD1 mRNA (Chi-square test).

G. Receiver Operating Characteristic curve to suggest the potential role of TS signature to predict the response to immune checkpoint inhibitors in melanoma cohort.

	GTF2I	TS	CS	CIN
Prognosis	Best	Moderately good	Moderately poor	Worst
WHO classification	A, AB	B1, B2	B2	B2, B3, C
Masaoka stage	I, II	I, II	II	II, III, IV
Myasthenia Gravis	Low	Mid	High	High
Copy number alterations	Low	Low	Low	High
Pathway Analysis	Wnt/ β-catennin	T cell signaling	mTOR EMT	SRC/EGF
Potential Treatment	Surgery Only	Immune Checkpoint Inhibitors	mTOR Inhibitors	Receptor Tyrosine Kinase Inhibitors

Figure 6.
Summary of characteristics of the four molecular subtypes of TETs.